

Reconstruction of 500 m, 8-day Historical MODIS Fractional Vegetation Cover (FVC) Dataset (1982–2000) in China

Xinyu Ding, Qunming Wang, Haoxuan Yang, Peter M. Atkinson

Abstract—Fractional vegetation cover (FVC) is a critical component of ecosystems, global climate change and the carbon cycle. Several FVC products have been released, the most widely used of which are the GLASS FVC products (including the GLASS-MODIS and GLASS-AVHRR FVC products). Specifically, the GLASS-MODIS FVC product covers the period from 2000 to present with a 500 m spatial resolution, whereas the GLASS-AVHRR FVC product is available from 1982 to present with a coarser spatial resolution of 5 km. For local monitoring of patterns of change in vegetation, however, there is a great need for fine spatial resolution (e.g., 500 m in this paper) and long-term time-series FVC datasets. To this end, we proposed to reconstruct a 500 m, 8-day historical MODIS FVC dataset (1982–2000) by making full use of the advantages of the existing GLASS-MODIS FVC (fine spatial resolution of 500 m) and GLASS-AVHRR FVC (long-term coverage from 1982 to the present) products covering China in this paper. The known GLASS-AVHRR FVC product was first used to fit the relationship between the FVC data after 2000 and before 2000, based on a random forest (RF) model. The trained relationship was migrated to the GLASS-MODIS FVC product, that is, predicting the MODIS FVC before 2000 based on the input of MODIS FVC after 2000. The validation using 64 scenes of Landsat FVC reference data revealed that the predicted historical MODIS FVC dataset has a reliable accuracy with a correlation coefficient (CC) value of 0.84, root mean square error (RMSE) of 0.14, Bias of 0.04 and unbiased RMSE (ubRMSE) of 0.12. Moreover, an accuracy evaluation in seven different regions in 1999 suggested that the historical MODIS FVC is closer to the Landsat FVC than the GEOV2 FVC product. Overall, the 500 m, 8-day MODIS FVC dataset (1982–2000) in China can provide important historical data for long-term, local monitoring of vegetation, which has great potential in supporting studies in a range of applications areas including ecology, hydrology and climatology. This dataset is available at <https://doi.org/10.6084/m9.figshare.24616446.v1>.

Index Terms—Fractional vegetation cover (FVC), historical, MODIS, dataset, random forest (RF).

I. INTRODUCTION

Vegetation plays an essential role in the global carbon cycle by absorbing carbon dioxide from the atmosphere in the form of carbohydrates through photosynthesis and fixing it in terrestrial ecosystems [1–4], which can protect terrestrial ecosystems by regulating climate, conserving water and preventing soil erosion [5–7]. Moreover, vegetation is an effective way to alleviate the urban heat island (UHI) effect by decreasing the temperature through evapotranspiration [8]. Therefore, monitoring vegetation dynamics has become an important topic in ecological and environmental assessments. Fractional vegetation cover (FVC) is an important indicator of surface vegetation conditions and is defined as the proportion of the vertical projection area of vegetation canopy (including vegetation components such as leaves, stems and branches) per unit area [9]. FVC is also a key factor in transpiration, photosynthesis, global climate change and other terrestrial processes and climate patterns, and it has been used widely in applications such as agriculture, forestry, drought monitoring and related fields [10], [11].

Remote sensing is the main tool used to estimate FVC at a large scale and has the advantages of a large monitoring range (e.g., global coverage) and regular revisit frequency. Generally, the FVC extracted by remote sensing technology refers to green vegetation cover. Currently, there are three classical methods for estimating FVC [12]. First, empirical methods construct the relationship between true FVC and vegetation index [13]. The methods can be divided into linear and nonlinear regression. Secondly, physical methods aim to establish the physical relationships between FVC and the spectral reflectance of the vegetation canopy, such as via a radiative transfer model [14] and geometric-optical model. For example, as a widely used radiative transfer model, PROSAIL is composed of the leaf optical properties model PROSPECT [15] and the canopy bidirectional reflectance model SAIL. The vegetation canopy reflectance and the corresponding FVC simulated based on the PROSAIL model can be used as training data to train a learning model, and then the learning model can be applied to estimate FVC from satellite observations. Finally, spectral unmixing methods assume that remote sensing image pixels contain different surface feature information [16]. In this case, the linear spectral mixture model (LSMM) has been the most widely used choice [17], [18]. Owing to the complexity of endmember extraction, several machine learning methods were developed recently [19], [20].

Up to now, several FVC products have been published, including the ENVISAT [21], GEOV [22], MuSyQ [23] and

Manuscript received XXX; revised XXX; accepted XXX. This work was supported by the National Natural Science Foundation of China under Grants 42222108 and 42171345 (Corresponding author: Q. Wang).

X. Ding is with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China.

Q. Wang is with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China and also with Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing 100081, China (e-mail: wqm11111@126.com).

H. Yang is with Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK, 74078, USA.

P.M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK and also with Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

GLASS FVC products [12], [24]. Information regarding the sensors, resolution and available time for these products is listed in Table 1. Amongst them, the GLASS FVC is a widely used product, due to the advantages in terms of spatial resolution (i.e., 500 m for MODIS) or temporal coverage (i.e., from 1982 to the present for AVHRR). Originally, the GLASS FVC product was derived from 500 m moderate resolution imaging spectroradiometer (MODIS) data (denoted as the GLASS-MODIS FVC product) using the general regression neural networks (GRNNs) approach [25], [26], where Thematic Mapper (TM) and Enhanced Thematic Mapper plus (ETM+) data were used as labels to construct training samples [12]. Considering that the computational efficiency of GRNNs was unsatisfactory for long-term products, alternative algorithms were considered in later studies, such as the multivariate adaptive regression splines (MARS) [27–29]. Generally, the 500 m GLASS-MODIS FVC product can provide more detailed spatial information than other FVC products at kilometer resolution. However, as MODIS surface reflectance data were acquired after 2000, GLASS-MODIS FVC product data are only available from 2000 to the present. That is, a long-term time-series FVC product that extends to before 2000 is not available. Therefore, to further expand the temporal coverage of GLASS FVC, a GLASS FVC product derived from advanced very high resolution radiometer (AVHRR) data (denoted as the GLASS-AVHRR FVC product) was developed, where the relation between the GLASS-MODIS FVC product and the AVHRR reflectance data was identified in the training process [24]. As the AVHRR data have been available from 1982, the GLASS-AVHRR FVC product can span from 1982 to the present, which is generally longer than other FVC products at the global scale.

Table 1. Information on some FVC products.

Product name	Sensor	Spatial resolution	Temporal resolution	Available time
ENVISAT	MERIS	1.2 km	10-day	2002-2012
GEOV	VGT	1 km	10-day	1999-2020
	https://land.copernicus.eu/global/products/fcover			
MuSyQ	MODIS, FY-3	1 km	5-day	2010-now
	http://www.doi.org/10.11922/sciencedb.j00001.00266			
GLASS FVC	AVHRR	5 km	8-day	1982-now
	http://www.glass.umd.edu/FVC/AVHRR/			
	MODIS	500 m	8-day	2000-now
	http://www.glass.umd.edu/FVC/MODIS/500m/			

The GLASS-MODIS FVC product, with a 500 m spatial resolution and an 8-day temporal resolution, has good spatial and temporal continuity characteristics as well as satisfactory accuracy at the global scale. Therefore, it has been used as ancillary data in conjunction with other data types for vegetation change monitoring and downscaling. For example, Mu *et al.* [30] used the GLASS-MODIS FVC product from 2001–2018 to analyze the vegetation change trend in China and quantified the CO₂, temperature, shortwave radiation, precipitation, and land

cover change effects on changes in vegetation cover using a generalized linear model. Hu *et al.* [31] used the GLASS-MODIS FVC product as a fine spatial resolution independent variable to assist in downscaling soil moisture data in the Tibetan Plateau region. Therefore, the GLASS-MODIS FVC product at fine spatial resolution provides important data supporting a wider range of researches.

As mentioned earlier, 500 m FVC data are absent before 2000. Thus, presently, the existing GLASS-MODIS FVC product cannot satisfy the requirement for long time-series vegetation monitoring at fine spatial resolution, and it would be of great interest to extend the time span of the GLASS-MODIS FVC product at 500 m spatial resolution. For existing FVC products, the GLASS-AVHRR FVC has the longest historical time span (i.e., beginning from 1982). Consequently, the GLASS-AVHRR FVC could be used as base data to extend the temporal length of the GLASS-MODIS FVC back to 1982. In contrast to some historical, coarse spatial resolution FVC data (e.g., GLASS-AVHRR FVC and GEOV), the produced 500 m historical FVC data would provide more refined and abundant spatial information. Moreover, based on these advantages, the constructed data will contribute to vegetation change trend analysis over a longer period and can also be employed to study vegetation-climate interactions and the relationships between ecosystems over historical periods.

Currently, most spatio-temporal reconstruction methods aim to fill missing data using partial available data that are spatially adjacent (known as gap filling), which is essentially an interpolation task. However, this task is substantially different from the task of reconstructing completely missing data (i.e., an extrapolation task). In this paper, the 500 m historical MODIS FVC data to be predicted is completely missing at the time of interest. Thus, the gap filling methods are not applicable to the reconstruction task in this paper.

In this paper, we reconstructed a 500 m, 8-day historical MODIS FVC dataset from 1982–2000 in China, based on the GLASS-MODIS FVC with fine spatial resolution and GLASS-AVHRR FVC with a longer period (from 1982 to the present). First, a training model was constructed using the GLASS-AVHRR FVC product at different times in China, based on the assumption that the FVC temporal change pattern in the GLASS-AVHRR FVC product is similar to that of the GLASS-MODIS FVC product. Specifically, the relation between the GLASS-AVHRR FVC data after 2000 (i.e., input) and GLASS-AVHRR FVC data at any time within 1982–2000 (i.e., label) was fitted using a random forest (RF) model. Then, the trained RF model was applied to estimate the historical 500 m MODIS FVC dataset from 1982–2000, with the GLASS-MODIS FVC after 2000 as input. The accuracy of the estimated 500 m, 8-day FVC data from 1982 to 2000 was evaluated by referring to Landsat FVC data. The reconstructed FVC dataset not only maintains the fine spatial resolution of the GLASS-MODIS FVC product, but also further extends the temporal length (i.e., from 1982). Thus, the reconstructed 500 m, 8-day MODIS FVC dataset (1982–2000) can provide important sources for long-term, local monitoring of vegetation, which can support studies in a range of applications areas including ecology, hydrology and climatology.

The remainder of this paper is organized as follows: Section II introduces the used data, including the GLASS-MODIS and

GLASS-AVHRR FVC products and Landsat data. Workflow details for reconstructing the historical MODIS FVC dataset in China are elaborated in Section III. Section IV presents validation of the reconstructed FVC results and the corresponding spatio-temporal analysis. Section V discusses further issues with the produced FVC dataset, including FVC change analysis, advantages in spatial resolution and the handling of uncertainty. Finally, Section VI concludes the paper.

II. DATA

A. GLASS-MODIS FVC product

The GLASS product suite [32] provides global MODIS FVC data (i.e., GLASS-MODIS FVC) with an 8-day temporal resolution and a 500 m spatial resolution [9]. The GLASS-MODIS FVC product was generated using a machine learning method applied to the MODIS surface reflectance data (i.e., MOD09A1), where the MARS model was chosen as the most applicable algorithm compared with the other three models [27]. Landsat data from global samples were used as labels for the training samples. Direct validation using 44 ground measurements from the Validation of Land European Remote Sensing Instruments (VALERI) sites revealed that the GLASS-MODIS FVC product trained by MARS has a reliable performance for estimating FVC ($R^2=0.836$, root mean square error (RMSE)=0.149). Considering the MODIS surface reflectance data were acquired from 24 February 2000, the corresponding GLASS-MODIS FVC product is not available before 2000. This paper aimed to reconstruct the historical MODIS FVC dataset for China from 1982–2000. The GLASS-MODIS FVC product was downloaded from <http://www.glass.umd.edu/FVC/MODIS/500m/>.

B. GLASS-AVHRR FVC product

The GLASS-AVHRR FVC product is available as a GLASS product suite with a 5 km spatial resolution and 8-day temporal resolution. Jia *et al.* [24] used the GLASS-MODIS FVC product to generate a long-term GLASS FVC product from AVHRR data dating back to 1982. More precisely, the AVHRR surface reflectance data [33], including the red and near-infrared bands, were used as input data and the corresponding GLASS-MODIS FVC product (reprojected to 5 km) was used as a label for training samples in training the MARS model. Based on the trained model, the 5 km FVC product starting in 1982 was estimated using AVHRR reflectance data as input. To ensure consistency with the GLASS-MODIS FVC results, the AVHRR FVC results were corrected linearly using the GLASS-MODIS FVC product. Direct validation of the ground measurements demonstrated the reliability of the GLASS-AVHRR FVC product ($R^2=0.834$, RMSE=0.145). Furthermore, compared to the GLASS-MODIS FVC product from 2013, satisfactory spatial and temporal consistencies were found between the two products. The data can be freely downloaded from <http://www.glass.umd.edu/FVC/AVHRR/>.

C. Landsat data

The Landsat-5 TM reflectance data, which can be acquired from Earth Explorer (<https://earthexplorer.usgs.gov/>), were used to validate the estimated historical MODIS FVC results in this paper. The spatial and temporal resolution of Landsat data are 30 m and 16-day, respectively. To match by location, the GLASS-MODIS and GLASS-AVHRR FVCs and Landsat data were unified into a single Geographic Coordinate System (i.e., WGS84). The reference FVC data at 500 m spatial resolution were derived by applying the dimidiate pixel model to the Landsat data. Further details are provided in Section III-D.

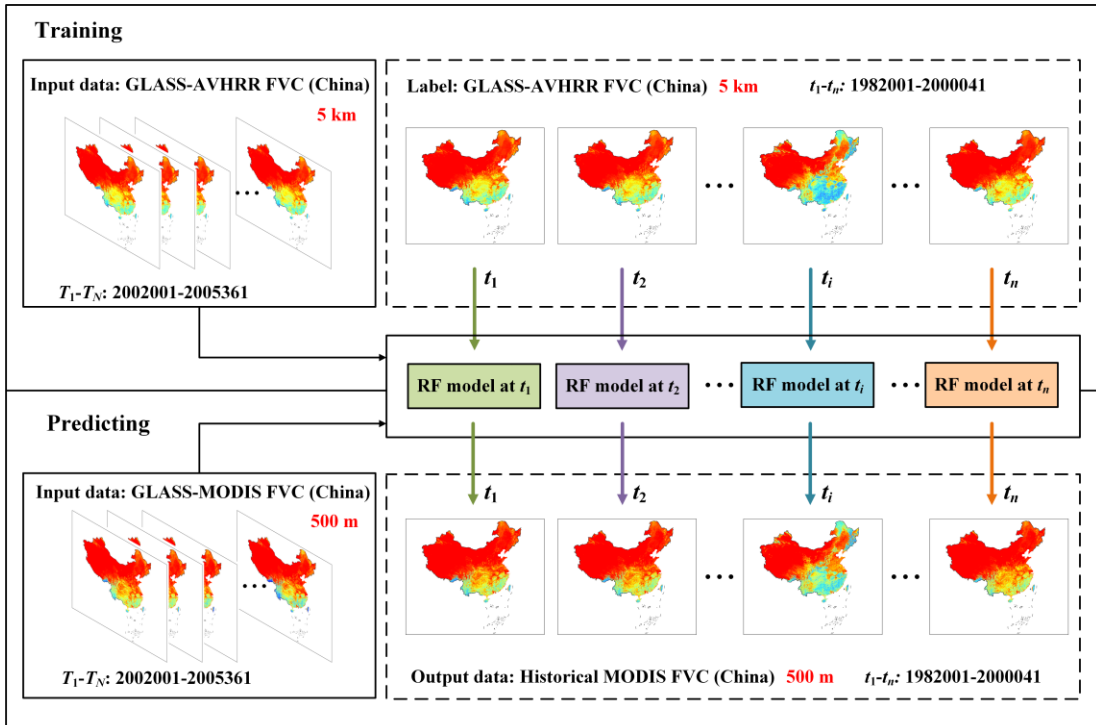


Fig. 1. Flowchart showing generation of historical MODIS FVC in China from 1982–2000.

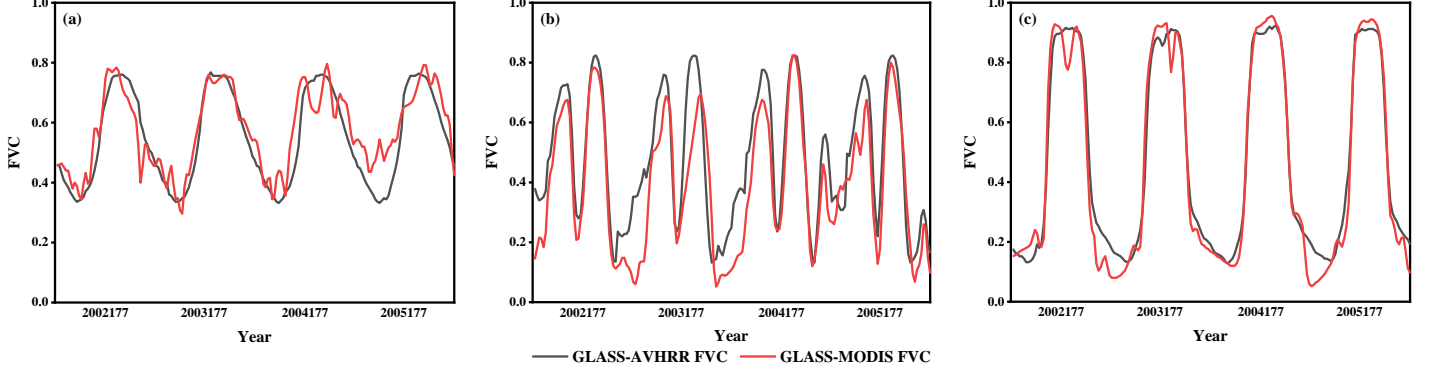


Fig. 2. The GLASS-AVHRR FVC (black line) and GLASS-MODIS FVC (red line) products temporal curves during 2002–2005 at three different geographical locations in China. (a) Lat: 26°12'N, Lon: 102°24'E; (b) Lat: 33°46'N, Lon: 115°25'E; (c) Lat: 43°31'N, Lon: 129°19'E.

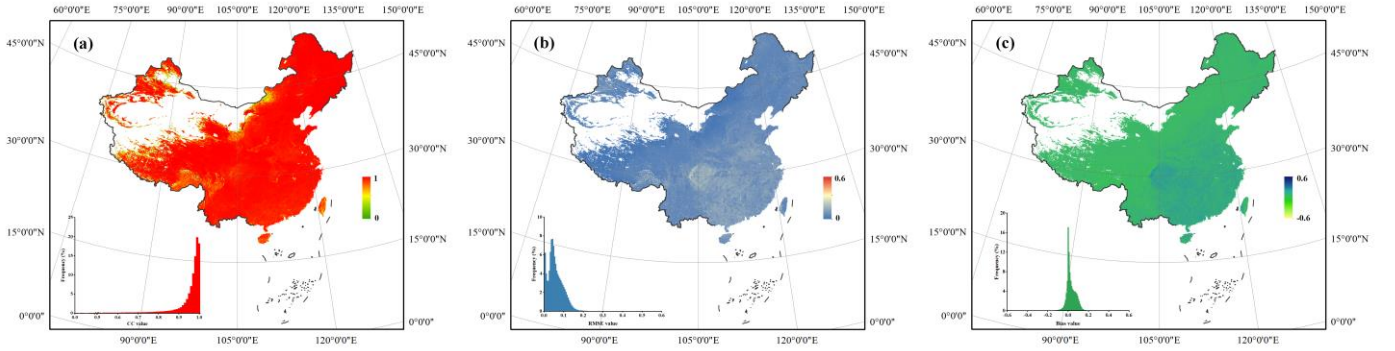


Fig. 3. The relation (in terms of three indicators) between the FVC temporal profiles (from 2002–2005) of the GLASS-AVHRR FVC and GLASS-MODIS FVC products (upscaled to 5 km). (a) CC; (b) RMSE; (c) Bias (using the GLASS-AVHRR FVC as reference).

III. METHODS

In this paper, we considered exploiting known FVC products (i.e., GLASS-AVHRR FVC) as training samples, in which the RF model was used as the training model. A strategy flowchart is shown in Fig. 1. The training samples were constructed using the GLASS-AVHRR FVC product (China) time-series data, with the data from day of year (DOY) 2002001–2005361 as input and known historical data (i.e., within DOY 1982001–2000041) as labels. The trained RF model was used to estimate the historical MODIS FVC dataset from DOY 1982001–2000041. Finally, the estimated historical MODIS FVC dataset was validated using Landsat FVC data. Details of this process are described in the following subsections.

A. Construction of training samples

Based on the GLASS-AVHRR and GLASS-MODIS FVC products, three groups of FVC change curves from 2002–2005 in China are shown in Fig. 2, with geographical locations selected randomly in China. We found that, for all three groups, the overall change trend of GLASS-AVHRR and GLASS-MODIS FVC is similar, despite the local difference in FVC values between the two products. Moreover, the intra-year variation in FVC is regular, and the inter-year variation also remains largely constant, indicating that the time-series of both the GLASS-AVHRR and GLASS-MODIS FVC products exhibit a regular change. In addition, we calculated the correlation coefficient (CC), RMSE and Bias between the FVC

temporal profiles for each pixel of the GLASS-AVHRR FVC and GLASS-MODIS FVC products (upscaled to 5 km) from 2002–2005, respectively, as shown in Fig. 3. It can be seen that most regions of China present CC values above 0.90, RMSE values below 0.1 and Bias values ranging from -0.1 to 0.1. Thus, the two FVC products have a large correlation and small differences between the temporal profiles. Based on the similarity in the two products, we assume that the pattern of temporal changes in the GLASS-AVHRR FVC can be transferred to the GLASS-MODIS FVC. Therefore, to reconstruct the historical MODIS FVC before 2000, the longer time-series GLASS-AVHRR FVC product for China was taken to construct the training samples for the predicted model. Specifically, the training model was constructed for each date within 1982–2000, in turn, in which the GLASS-AVHRR FVC data from 2002–2005 were used as the input for training, whereas the GLASS-AVHRR FVC data for a date within 1982–2000 were used as labels to train the corresponding learning model.

B. RF model for training

The RF model is a decision tree-based machine learning algorithm proposed by Breiman [34]. The RF model can quantify complex nonlinear relationships that are not affected by outliers or redundant data and there is no significant over-fitting risk [35]. Therefore, the RF has been utilized widely in fields such as ecology [36], [37] and agriculture [38]. This paper exploited the RF model to estimate the historical MODIS FVC dataset. In addition, the temporal length of the GLASS-AVHRR

FVC data (China) used as input data was evaluated to determine the appropriate temporal length for historical MODIS FVC estimation (Section IV-A). In addition, the number of decision trees (ntree) was set to 200 and the number of variables at each node (mtry) was set to 5.

C. Estimation of the historical MODIS FVC dataset (1982–2000)

The GLASS-AVHRR FVC product in China was utilized as input data, and the optimal temporal length for estimating the historical MODIS FVC was determined as 4 years (illustrated in Section IV-A). The RF model was trained using GLASS-AVHRR FVC data (China) using the following expression:

$$\mathbf{FVC}_{t_i}^{\text{AVHRR}} = f_{t_i}(\mathbf{FVC}_{T_1}^{\text{AVHRR}}, \mathbf{FVC}_{T_2}^{\text{AVHRR}}, \dots, \mathbf{FVC}_{T_N}^{\text{AVHRR}}) \quad (1)$$

where $\mathbf{FVC}_{T_1}^{\text{AVHRR}}, \mathbf{FVC}_{T_2}^{\text{AVHRR}}, \dots, \mathbf{FVC}_{T_N}^{\text{AVHRR}}$ are the fixed time-series GLASS-AVHRR FVC data (China) with 5 km spatial resolution based on the optimal temporal length (i.e., 4 years in this paper) as the input data for training. Specifically, $T_1, T_2, \dots, \text{and } T_N$ represent the period from 2002001–2005361. Note that the data of two dates, 2002305 and 2005169, are missing from the GLASS-MODIS FVC product. To maintain a consistent number of training samples, the corresponding GLASS-AVHRR FVC data constructing the training samples do not include the scenes of these two dates, resulting in 182 scenes. That is, $N=182$ in this paper. $\mathbf{FVC}_{t_i}^{\text{AVHRR}}$ represents the GLASS-AVHRR FVC data (5 km) at a time t_i ($i = 1, 2, \dots, n$) in the historical period of 1982001–2000041 (i.e., $n=834$), which was used as a label for training the RF model. f_{t_i} is the nonlinear relationship operator at the corresponding time t_i .

In line with the assumption mentioned in Section III-A, the trained RF model was used to produce a 500 m, 8-day historical MODIS FVC dataset from 1982–2000 (i.e., DOY from 1982001–2000041). The specific expression is as follows:

$$\mathbf{FVC}_{t_i}^{\text{MODIS}} = f_{t_i}(\mathbf{FVC}_{T_1}^{\text{MODIS}}, \mathbf{FVC}_{T_2}^{\text{MODIS}}, \dots, \mathbf{FVC}_{T_N}^{\text{MODIS}}) \quad (2)$$

where f_{t_i} is the function fitted using Eq. (1) and $\mathbf{FVC}_{T_1}^{\text{MODIS}}, \mathbf{FVC}_{T_2}^{\text{MODIS}}, \dots, \mathbf{FVC}_{T_N}^{\text{MODIS}}$ are known time-series GLASS-MODIS FVC data (China) with 500 m spatial resolution from 2002001–2005361 as input data. Note that the input data are also fixed, and the temporal length of the input data is the same as that in Eq. (1). $\mathbf{FVC}_{t_i}^{\text{MODIS}}$ is the predicted MODIS FVC (500 m) at time t_i in the historical period (i.e., within 1982001–2000041). For each prediction time t_i , the training model in Eq. (1) and predicting model in Eq. (2) were constructed in turn. The specific processes are summarized as follows:

- 1) The GLASS-AVHRR FVC data (China) from 2002001–2005361 were used as the fixed input of the training data and the known historical GLASS-AVHRR FVC (China) at the prediction time t_i (i.e., in the historical period of 1982001–2000041) was used as the label.

- 2) The RF model was trained based on the training data in step 1).
- 3) The corresponding GLASS-MODIS FVC (China) from 2002001–2005361 were input to the trained RF model, producing the prediction of the historical MODIS FVC at time t_i .
- 4) The above steps were repeated for all time points in the period from 1982001–2000041. Then, a 500 m, 8-day historical FVC dataset from 1982–2000 in China was reconstructed.

D. Validation with Landsat data

Owing to the lack of in-situ FVC data for validation in the predicted time interval, the processed Landsat FVC data were selected as references based on Landsat-5 surface reflectance images, which were acquired from 1984–2000 with a 30 m spatial resolution. The classic dimidiate pixel model was used to calculate the acquired Landsat image FVC values [10], [39], [40]. The dimidiate pixel model is a simple and widely used model that assumes that a pixel is composed only of vegetation and soil [41], [42]. The normalized difference vegetation index (NDVI) reflects vegetation growth during different periods as well as vegetation coverage at different locations [43]. Therefore, based on the dimidiate pixel model, the 30 m FVC was estimated using the NDVI [44]. The mathematical expression is as follows:

$$\text{FVC}^{\text{Landsat}} = \frac{\text{NDVI} - \text{NDVI}_s}{\text{NDVI}_v - \text{NDVI}_s} \quad (3)$$

where $\text{FVC}^{\text{Landsat}}$ is the estimated 30 m FVC; NDVI is the NDVI value of the pixel, which is calculated by bands 3 (i.e., Red band) and 4 (i.e., NIR band) from Landsat-5 surface reflectance data (i.e., $\text{NDVI} = \frac{\text{Band4} - \text{Band3}}{\text{Band4} + \text{Band3}}$); NDVI_s is the NDVI value of a full soil pixel, and NDVI_v is the NDVI value of a full vegetation pixel.

The NDVI_s and NDVI_v values remain constant in the ideal case; however, these two parameters were affected by other environmental factors. Despite this, the NDVI_s and NDVI_v values can be confirmed by NDVI statistical analysis of the entire Landsat image, which assumes that full soil and vegetation pixels exist for each Landsat image. In this paper, the NDVI of each Landsat image was calculated and then the cumulative probability distribution was analyzed statistically. The NDVI values at 2% and 98% cumulative percentage were determined as NDVI_s and NDVI_v , respectively [45], [46]. After estimating the 30 m FVC according to Eq. (3), the result was upscaled to 500 m to validate directly the corresponding 500 m historical MODIS FVC.

IV. RESULTS

A. Evaluation of the RF model

The influence of the temporal length of the time-series data for the RF model was evaluated to ensure the reliability of training sample selection. Five temporal lengths were evaluated in this study, with a 1-year minimum and a 5-year maximum temporal length. The model performance was evaluated by predicting the MODIS FVC in China for 2001 (i.e., from DOY 2001001–2001361, a total of 46 scenes) and comparing it with

the existing GLASS-MODIS FVC product at the corresponding time [12]. The CC, RMSE, Bias and unbiased RMSE (ubRMSE) were utilized as statistical indicators. Fig. 4 shows the accuracy for each temporal length in the RF model. According to Fig. 4, it is not difficult to find that the CC increases while the RMSE decreases with temporal length increasing from 1-year to 5-year. For the 4-year temporal length, the CC value can reach 0.95, but does not increase obviously as the temporal length increases to 5-year. Therefore, we take the 4-year (i.e., from 2002–2005) as the desirable temporal length.

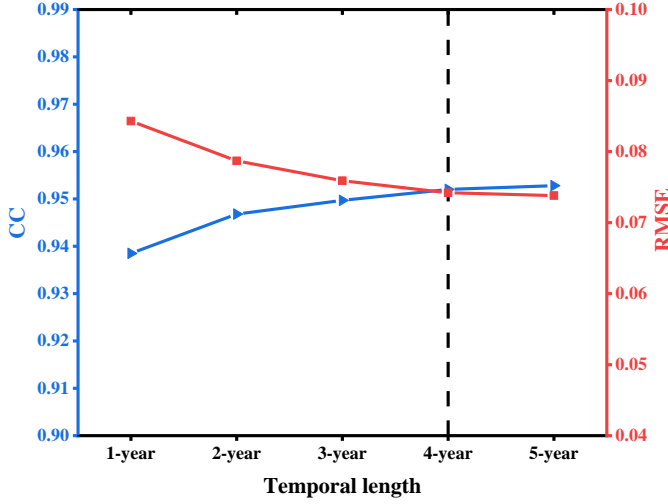


Fig. 4. Evaluation of the temporal length of the time-series used in the RF model.

We trained the RF model based on the optimal temporal length, and the four accuracy curves of the trained RF model for predicting the FVC of the whole China in 2001 are shown in Fig. 5. It can be seen that the CC value remains between 0.92 and 0.97 for the 46 predicted results (i.e., from DOY 2001001–

2001361) in 2001. Meanwhile, the ubRMSE remains between 0.04 and 0.08. The averages of the four accuracy indicators are also plotted in Fig. 5. We can clearly observe that the average CC of the predicted FVC can achieve 0.95 and the average Bias is only -0.02, and the corresponding average RMSE and ubRMSE values are 0.07 and 0.07, respectively. The results suggest that the RF model trained by the GLASS-AVHRR FVC product (China) with the optimal temporal length can reliably predict FVC.

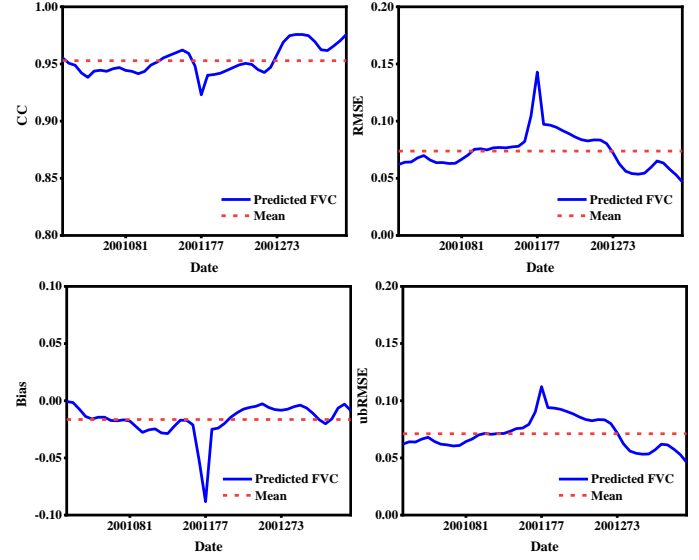
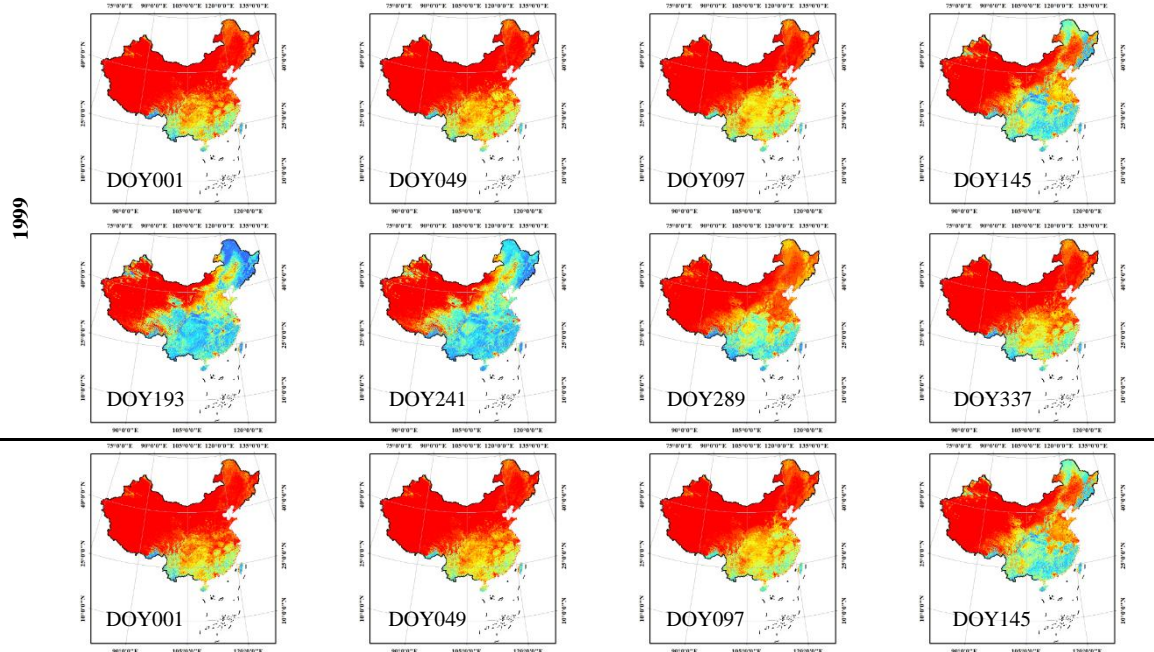


Fig. 5. Four statistical metrics (i.e., CC, RMSE, Bias and ubRMSE) of the predicted 500 m FVC in China (with the GLASS-MODIS FVC product in 2001 as reference).



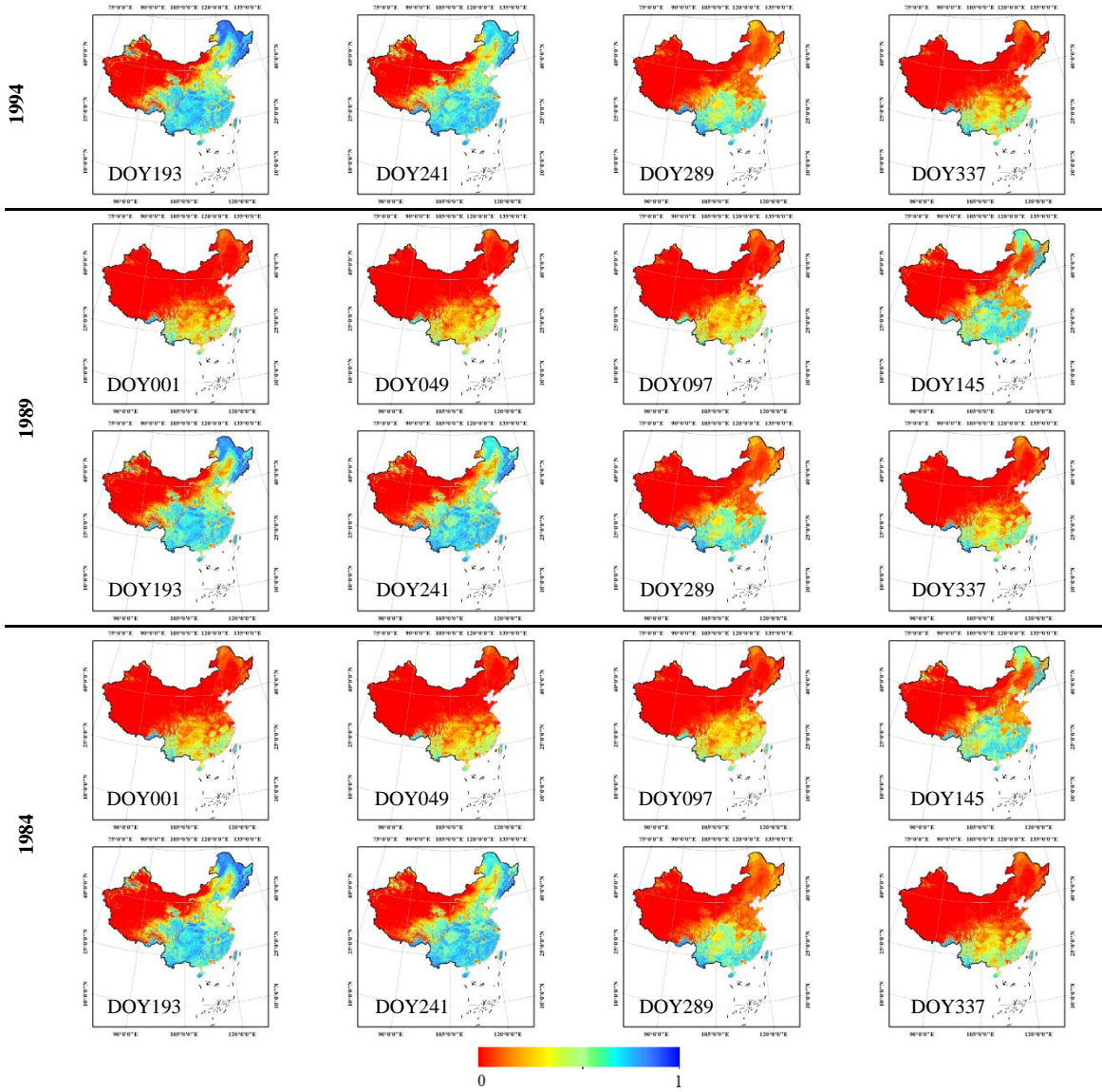


Fig. 6. Exhibition of the predicted historical MODIS FVC in China for several randomly selected days.

B. Validation of the 500 m FVC results using Landsat FVC

1) *Overall validation.* With the RF model, we obtained the 500 m 8-day FVC historical dataset (1982–2000) in China. Part of the data are shown in Fig. 6. Since Landsat-5 data were available from 1984, we acquired 64 Landsat tiles from 1984–2000. More precisely, for each year, the historical MODIS FVC results were validated with four Landsat tiles sampled randomly at different times. The spatial distribution map of Landsat tiles used for validation is displayed in Fig. 7. The spatial size of the validation region is 200×200 pixels, with a 500 m spatial resolution. Table 2 shows the accuracy evaluation results between the historical MODIS and Landsat FVCs at a 500 m resolution. To provide an intuitive accuracy assessment, we present scatterplots for four selected years (i.e., 1996, 1994, 1992 and 1990) of the validation results (Fig. 8). According to the validation analysis (Table 2), it is observed that the CC values are all above 0.71, and the RMSE values are below 0.22. Moreover, we calculated the averages of each indicator for validation results of all 48 scenes. The results show that the mean values of CC, RMSE, Bias and ubRMSE are 0.84, 0.14,

0.04 and 0.12, respectively. As shown in Fig. 8, the scatter of the MODIS FVC results is clustered around the line $y=x$ (black line), suggesting that the historical FVC estimated by the RF model is close to the reference data (that is, Landsat FVC). This reveals that the MODIS FVC dataset produced in the historical period has a reliable accuracy.

Table 2. Accuracy evaluation results (predicted MODIS FVC vs. Landsat FVC) at 500 m resolution.

Date	Path/Row	CC	RMSE	Bias	ubRMSE
1999/03/04	130/042	0.8686	0.1479	0.1135	0.0948
1999/05/18	127/035	0.9275	0.1129	0.0730	0.0861
1999/05/27	126/037	0.8360	0.1228	-0.0745	0.0976
1999/06/29	117/030	0.8808	0.1841	-0.1181	0.1412
1998/05/25	117/030	0.9019	0.1274	0.0360	0.1223
1998/05/26	124/033	0.8392	0.2100	0.1768	0.1134
1998/05/28	122/032	0.9095	0.1637	0.1348	0.0930
1998/10/24	125/044	0.7717	0.1555	0.1147	0.1051
1997/05/25	122/025	0.8047	0.1618	0.0488	0.1542
1997/05/28	127/035	0.9303	0.0910	0.0186	0.0891

1997/06/13	127/035	0.9387	0.0968	-0.0066	0.0966
1997/06/23	117/029	0.8051	0.1879	-0.1223	0.1427
1996/04/12	130/042	0.8830	0.1617	0.1287	0.0979
1996/05/22	122/032	0.8887	0.1361	0.0914	0.1008
1996/09/14	127/035	0.8029	0.1021	0.0002	0.1021
1996/12/14	124/044	0.8089	0.1410	0.1052	0.0939
1995/05/07	127/035	0.8587	0.1109	0.0038	0.1108
1995/05/09	125/041	0.7890	0.1128	-0.0305	0.1086
1995/06/08	127/035	0.9244	0.1126	0.0264	0.1095
1995/06/19	124/033	0.8221	0.1807	0.1336	0.1217
1994/05/17	122/032	0.8877	0.1246	0.0816	0.0941
1994/05/27	128/036	0.9352	0.0907	-0.0047	0.0905
1994/06/15	117/030	0.8660	0.1250	-0.0157	0.1240
1994/08/22	145/030	0.7354	0.1536	0.0608	0.1411
1993/04/04	130/042	0.8717	0.1556	0.1137	0.1061
1993/05/14	122/032	0.8877	0.1070	0.0478	0.0957
1993/09/22	127/035	0.8226	0.1205	0.0791	0.0909
1993/12/06	124/044	0.8072	0.1503	0.1115	0.1008
1992/02/13	130/042	0.7614	0.1667	0.1099	0.1254
1992/05/14	127/035	0.9189	0.1238	0.0575	0.1096
1992/05/27	122/032	0.8595	0.1192	0.0475	0.1093
1992/07/17	127/035	0.8741	0.1102	-0.0077	0.1100
1991/02/10	130/042	0.7687	0.1468	0.0789	0.1238
1991/05/25	122/025	0.7520	0.1683	0.0556	0.1588
1991/11/15	124/043	0.8162	0.1189	-0.0297	0.1151
1991/11/15	124/044	0.7967	0.1031	0.0133	0.1031
1990/03/11	130/042	0.8411	0.1695	0.1348	0.1027
1990/05/18	126/037	0.7973	0.1302	-0.0804	0.1024
1990/05/22	122/032	0.8963	0.1002	-0.0029	0.1001
1990/11/28	124/044	0.8055	0.1096	0.0116	0.1090
1989/05/03	122/032	0.8769	0.2075	0.1709	0.1178
1989/05/11	130/042	0.8972	0.1405	0.0990	0.0997
1989/06/01	117/028	0.8963	0.1341	0.0227	0.1322
1989/12/02	125/044	0.8131	0.1445	0.0962	0.1079
1988/02/02	130/042	0.7133	0.1556	0.0801	0.1335
1988/06/04	127/038	0.7326	0.2203	-0.1835	0.1234
1988/06/06	125/041	0.7561	0.1247	-0.0606	0.1090
1988/06/25	114/028	0.7764	0.1984	-0.0561	0.1903
1987/02/15	130/042	0.7579	0.1800	0.1316	0.1228

1987/05/14	122/032	0.8946	0.1068	0.0348	0.1009
1987/06/02	127/038	0.7917	0.2004	-0.1530	0.1295
1987/10/08	127/035	0.8863	0.1870	0.1435	0.1199
1986/05/27	122/032	0.8992	0.1008	0.0249	0.0977
1986/06/04	114/028	0.8184	0.1731	0.0134	0.1725
1986/06/09	117/030	0.8183	0.1820	0.1049	0.1488
1986/12/13	130/042	0.7403	0.1628	0.1037	0.1255
1985/05/14	116/030	0.7989	0.1741	-0.0716	0.1587
1985/05/16	114/028	0.8467	0.1704	0.0789	0.1510
1985/05/21	117/029	0.8925	0.1417	0.0426	0.1352
1985/05/21	117/030	0.8577	0.1338	0.0322	0.1299
1984/05/18	117/027	0.8093	0.1315	0.0122	0.1309
1984/05/18	117/028	0.8747	0.1555	0.0924	0.1250
1984/05/18	117/029	0.8653	0.1466	0.0659	0.1310
1984/05/18	117/030	0.8706	0.1710	0.1238	0.1180
Average		0.8403	0.1446	0.0416	0.1173

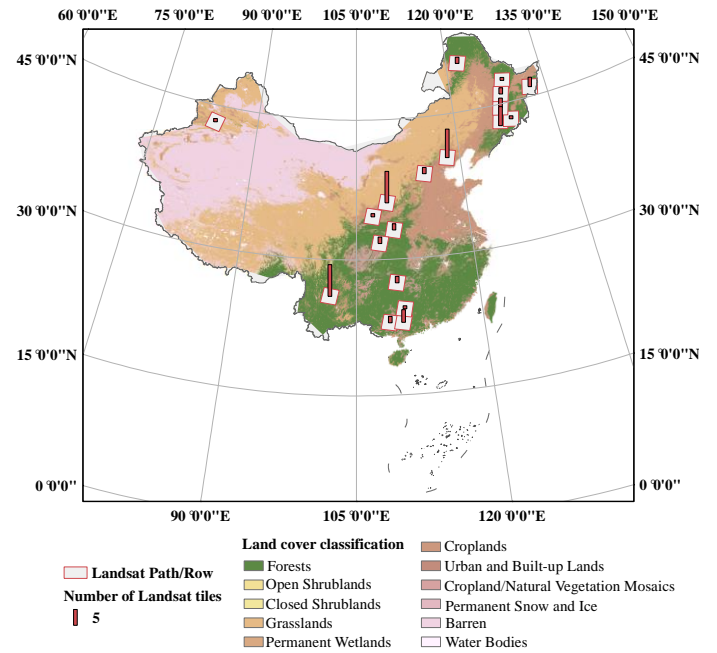
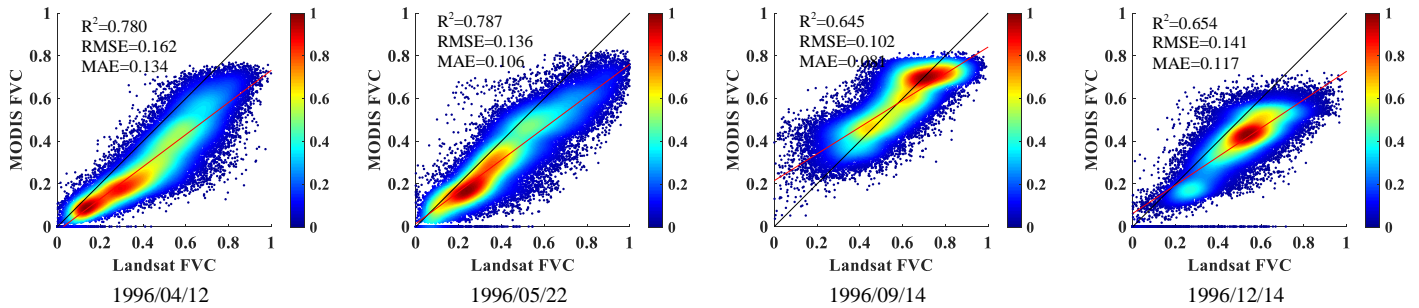


Fig. 7. The spatial distribution of Landsat tiles used for validation. The base map is the land cover type map derived from the MCD12Q1, where the different forest types are all categorized as forests.



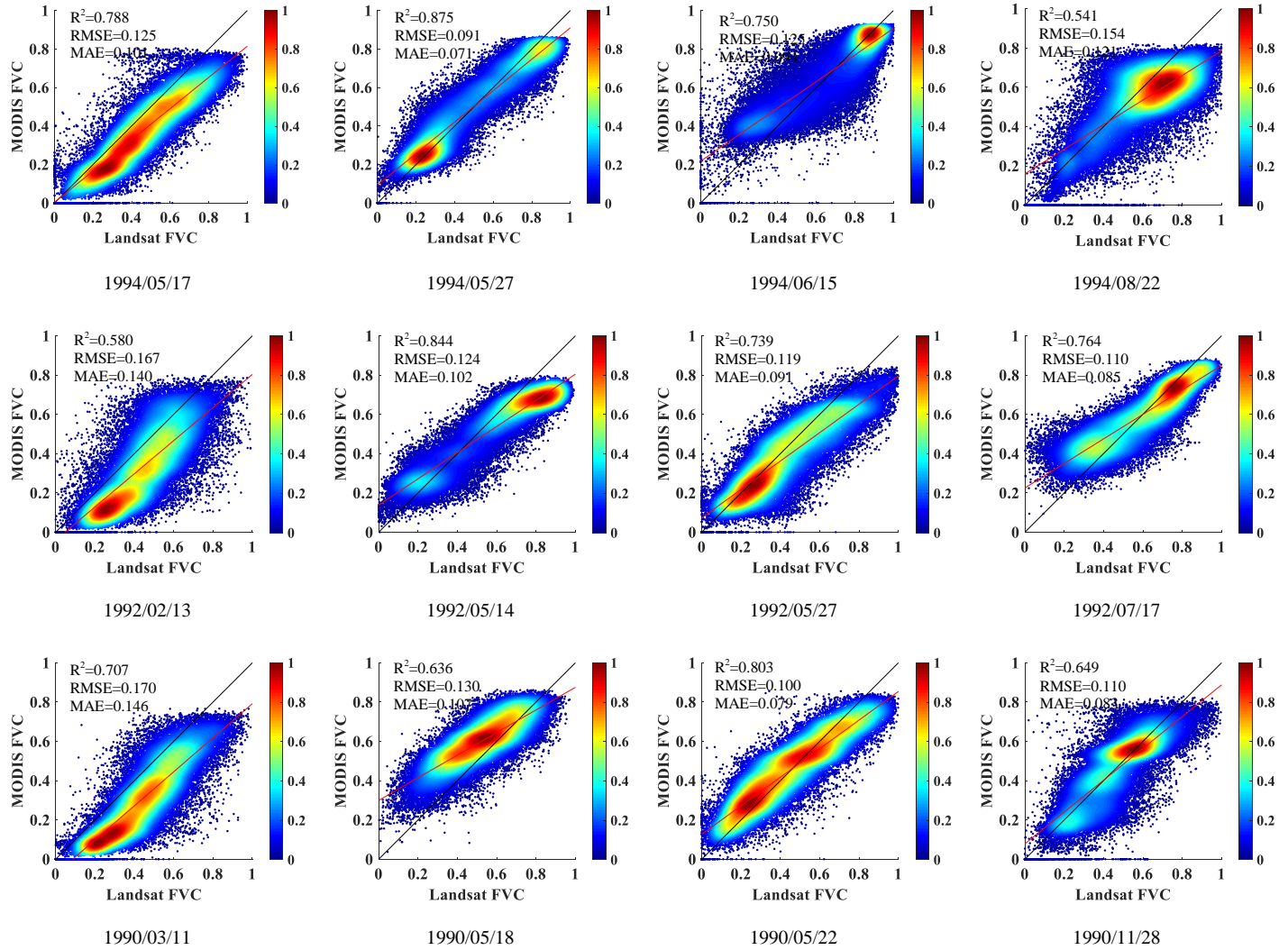


Fig. 8. Density scatterplots of Landsat FVC and historical MODIS FVC on several randomly selected days. The black line represents the $y=x$ line. The red line is the fitted linear line. Point colors indicate the probability density.

2) *Validation based on different land cover types.* To further validate the historical MODIS FVC dataset, we evaluated the accuracy under three main land cover types of vegetation in China (i.e., forests, grasslands and croplands), based on the MCD12Q1 land cover product. Specifically, a 200×200 pixel region of the historical MODIS FVC (at 500 m resolution) was selected, which corresponds to the Landsat tile with Path/Row of 122 and 032, as shown in Fig. 9. The accuracy of the historical MODIS FVC for the selected region was evaluated for all three land cover types at three different times, as listed in Table 3. It is observed that the average CC values for forests, grasslands and croplands are 0.86, 0.87 and 0.87, respectively, with the corresponding average RMSE values of 0.13, 0.12 and 0.10. This suggests that the 500 m MODIS FVC dataset generally presents stable accuracy under different land cover types.

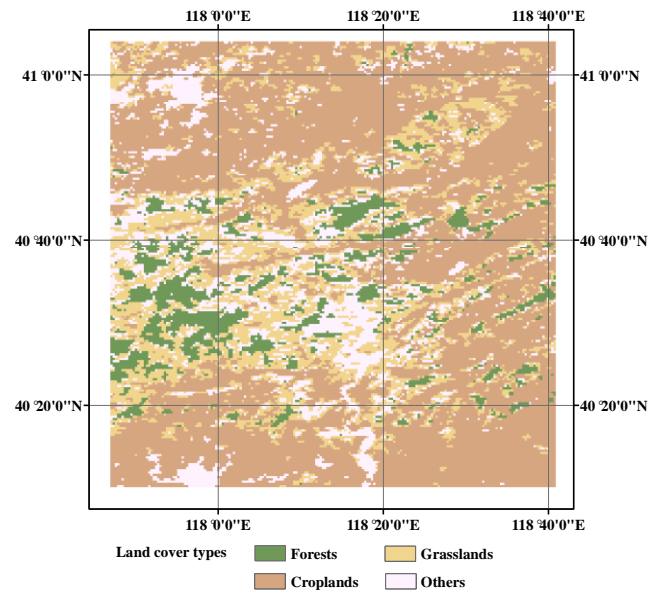


Fig. 9. The spatial location and the land cover types of the selected region (200×200 pixels; 500 m) for validation based on different land cover types.

Table 3. Accuracy evaluation results for different land cover types (predicted historical MODIS FVC vs. Landsat FVC) at 500 m resolution.

	Date	CC	RMSE	Bias	ubRMSE
Forests	1996/05/22	0.8461	0.1369	0.0777	0.1128
	1992/05/27	0.8478	0.1295	0.0472	0.1206
	1986/05/27	0.8852	0.1161	0.0456	0.1067
	Average	0.8597	0.1275	0.0568	0.1134
Grasslands	1996/05/22	0.8631	0.1320	0.0732	0.1099
	1992/05/27	0.8552	0.1208	0.0332	0.1162
	1986/05/27	0.8934	0.1062	0.0272	0.1027
	Average	0.8706	0.1197	0.0445	0.1096
Croplands	1996/05/22	0.8706	0.1115	0.0598	0.0941
	1992/05/27	0.8493	0.1078	0.0317	0.1031
	1986/05/27	0.8876	0.0944	0.0205	0.0921
	Average	0.8692	0.1046	0.0373	0.0964

C. Comparison with the GLASS-AVHRR FVC product

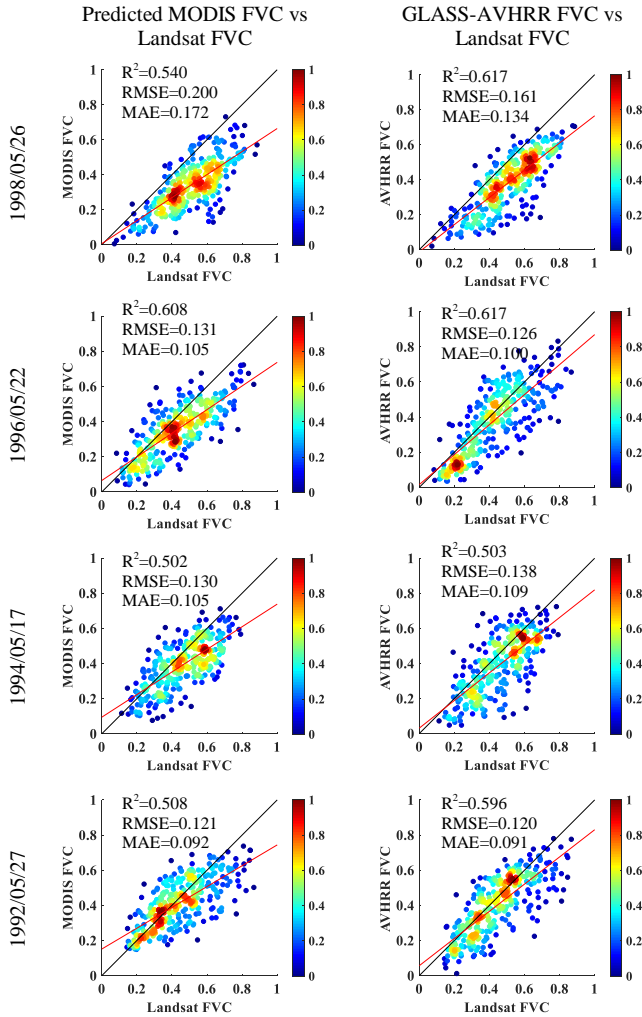


Fig. 10. The density scatterplots of predicted MODIS FVC (or GLASS-AVHRR FVC) versus Landsat FVC across multiple time points.

Four tiles in Table 2 (with dates marked in Fig. 10) were selected for comparison. Fig. 10 shows a comparison in terms of the density scatterplots between GLASS-AVHRR FVC and

predicted MODIS FVC across multiple time points within the historical period, using Landsat FVC as a reference. The spatial resolution of the three different datasets was unified to 5 km. From the density scatterplots, it can be observed that the scatter of GLASS-AVHRR FVC is closer to the line $y=x$ than that of predicted MODIS FVC. This reveals that uncertainties exist in the prediction model, making the predicted MODIS FVC deviates from the original GLASS-AVHRR FVC.

D. Comparison with the GEOV2 FVC product

The global biophysical products Version 2 (GEOV2) FVC product is defined as the fraction of ground covered by green vegetation, which has the same physical meaning as the GLASS-FVC product. To be specific, it was generated by a neural network using the Top-of-Canopy (TOC) input reflectance in visible, near-infrared and shortwave infrared [47]. The GEOV2 FVC product period was available from 1999 to June 2020 with a 10-day temporal resolution and a 1 km spatial resolution. The GEOV2 FVC product in 1999 was collected from the Copernicus Global Land Service (<https://land.copernicus.eu/global/products/fcover>) for comparison.

Seven study areas were randomly selected to compare the predicted historical MODIS FVC dataset to the GEOV2 FVC product, with the Landsat FVC as reference. The specific dates and locations are listed in Table 4. It should be noted that there exists a difference in spatial resolution among the three FVC products. To ensure comparative fairness, we unified the spatial resolution of the three different datasets to 1 km. The scatterplots and error maps are shown in Fig 11. From the density scatterplots, the historical MODIS FVC scatter is more concentrated around the black line (i.e., the $y=x$ line) than GEOV2 FVC for each study region, indicating that the historical MODIS FVC is closer to the Landsat FVC. Moreover, the error maps (in absolute value) in Fig. 11 indicate that the error of the historical MODIS FVC is smaller than that of GEOV2 FVC in all seven regions. The accuracy evaluation in Table 5 also shows that the historical MODIS FVC results are more accurate than those of GEOV2 FVC. For the seven different regions, the average CC for the historical MODIS FVC is 0.03 larger than that of GEOV2 FVC, and the corresponding average RMSE is 0.02 smaller. More precisely, the CCs of the historical MODIS FVC are 0.01, 0.03, 0.02, 0.03, 0.07, 0.01 and 0.02 greater than those of GEOV2 FVC for Regions 1-7, respectively, and the corresponding RMSE values are 0.03, 0.02, 0.02, 0.01, 0.01, 0.02 and 0.04 smaller. In Region 1, the Bias and ubRMSE values of the historical MODIS FVC are 0.02 and 0.02 less, respectively, than those of GEOV2 FVC.

Table 4. The dates and locations for the data of the selected seven regions.

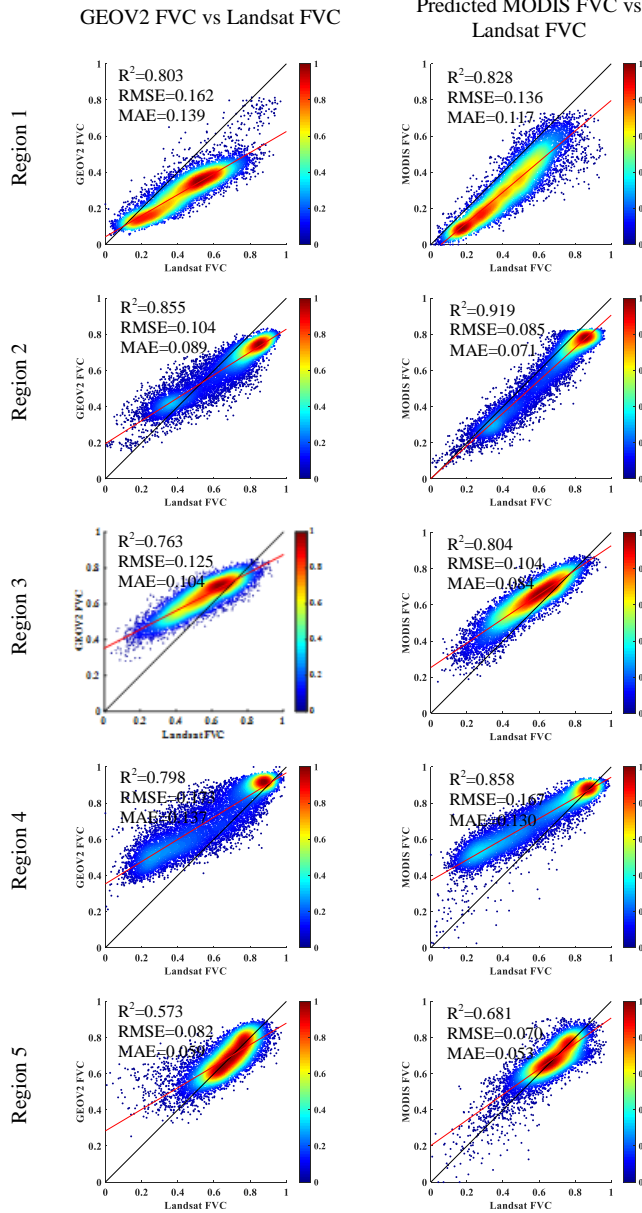
	Date	Location
Region 1	1999/03/04	25°45'N, 101°33'E
Region 2	1999/05/18	35°50'N, 108°54'E
Region 3	1999/05/27	33°13'N, 109°57'E
Region 4	1999/06/29	42°59'N, 126°40'E
Region 5	1999/09/25	23°11'N, 108°51'E
Region 6	1999/11/05	22°47'N, 110°37'E
Region 7	1999/12/23	22°56'N, 110°38'E

Table 5. Accuracy evaluation of GEOV2 FVC and predicted historical MODIS FVC in the seven different regions (the most accurate results are marked in bold).

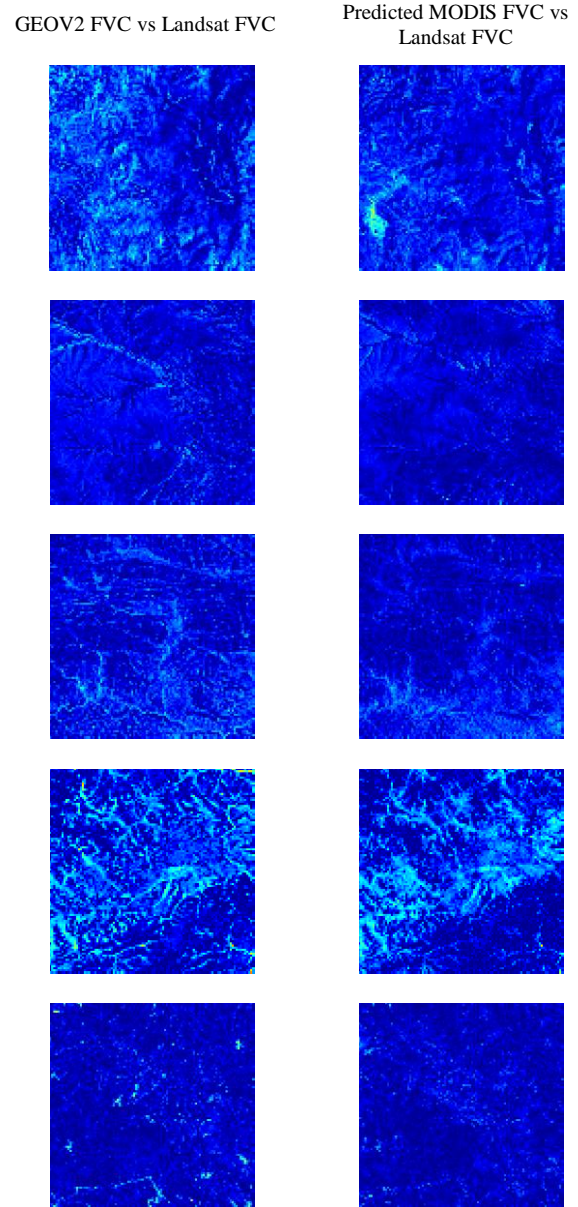
	Dataset	CC	RMSE	Bias	ubRMSE
Region 1	GEOV2 FVC	0.8961	0.1616	0.1336	0.0909
	Predicted MODIS FVC	0.9097	0.1357	0.1135	0.0744
Region 2	GEOV2 FVC	0.9246	0.1041	-0.0453	0.0938
	Predicted MODIS FVC	0.9589	0.0849	-0.0608	0.0592
Region 3	GEOV2 FVC	0.8737	0.1253	-0.0901	0.0871
	Predicted MODIS FVC	0.8965	0.1039	-0.0745	0.0725
Region 4	GEOV2 FVC	0.8935	0.1734	-0.1253	0.1199

	Predicted MODIS FVC	0.9263	0.1672	-0.1181	0.1184
Region 5	GEOV2 FVC	0.7570	0.0817	-0.0135	0.0806
	Predicted MODIS FVC	0.8253	0.0699	-0.0056	0.0696
Region 6	GEOV2 FVC	0.8126	0.0947	-0.0704	0.0634
	Predicted MODIS FVC	0.8228	0.0766	-0.0327	0.0693
Region 7	GEOV2 FVC	0.7913	0.1310	0.1103	0.0705
	Predicted MODIS FVC	0.8137	0.0900	0.0535	0.0723
Average	GEOV2 FVC	0.8498	0.1245	-0.0144	0.0866
	Predicted MODIS FVC	0.8790	0.1040	-0.0178	0.0765

Density scatterplots



Error maps



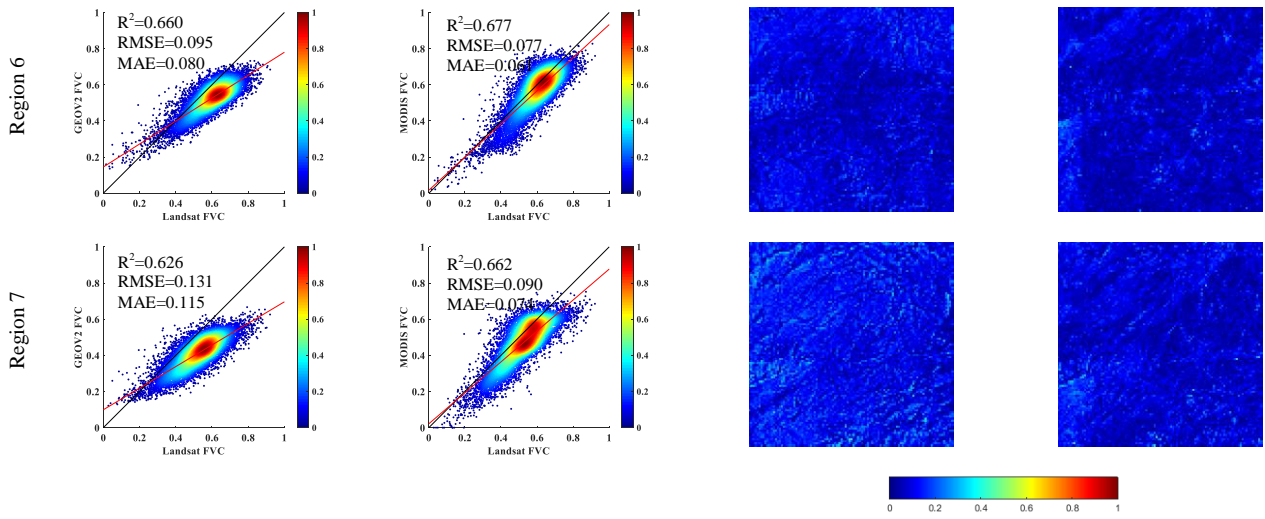
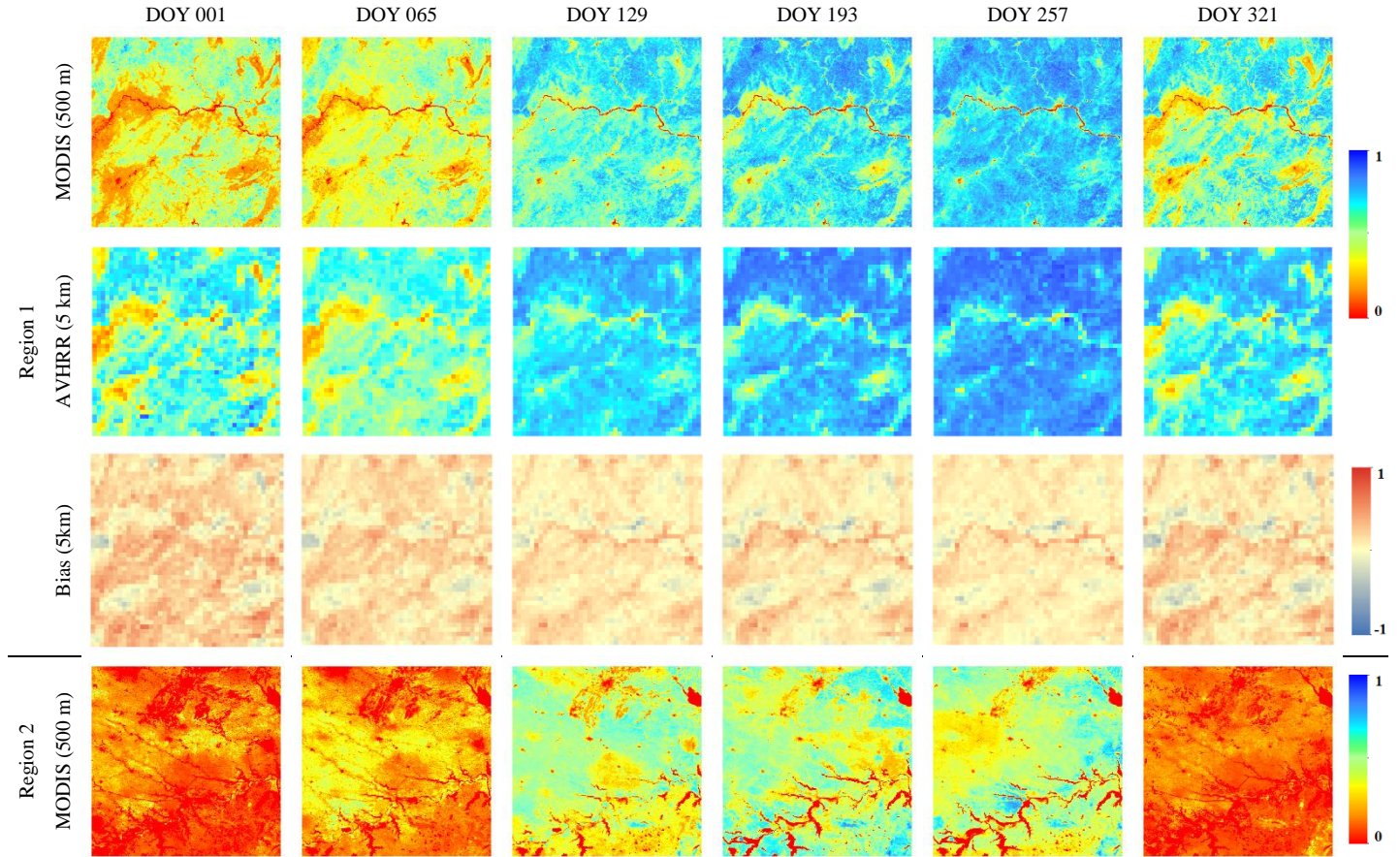


Fig. 11. The density scatterplots and error maps (in absolute value) of predicted MODIS FVC (or GEOV2 FVC) versus Landsat FVC for the seven regions.

E. Spatial and temporal analysis

1) *Spatial consistency with AVHRR FVC.* Examples of the predicted historical MODIS FVC (500 m), GLASS-AVHRR FVC (5 km) and the Bias maps for these two products (5 km) in three randomly selected regions on DOY 001, 065, 129, 193, 257 and 321 in 1989 are shown in Fig. 12. It can be found that

the FVC distribution between the historical MODIS and GLASS-AVHRR FVCs is basically consistent for each region. This is because the historical MODIS FVC results were estimated using an RF model trained by the GLASS-AVHRR FVC product.



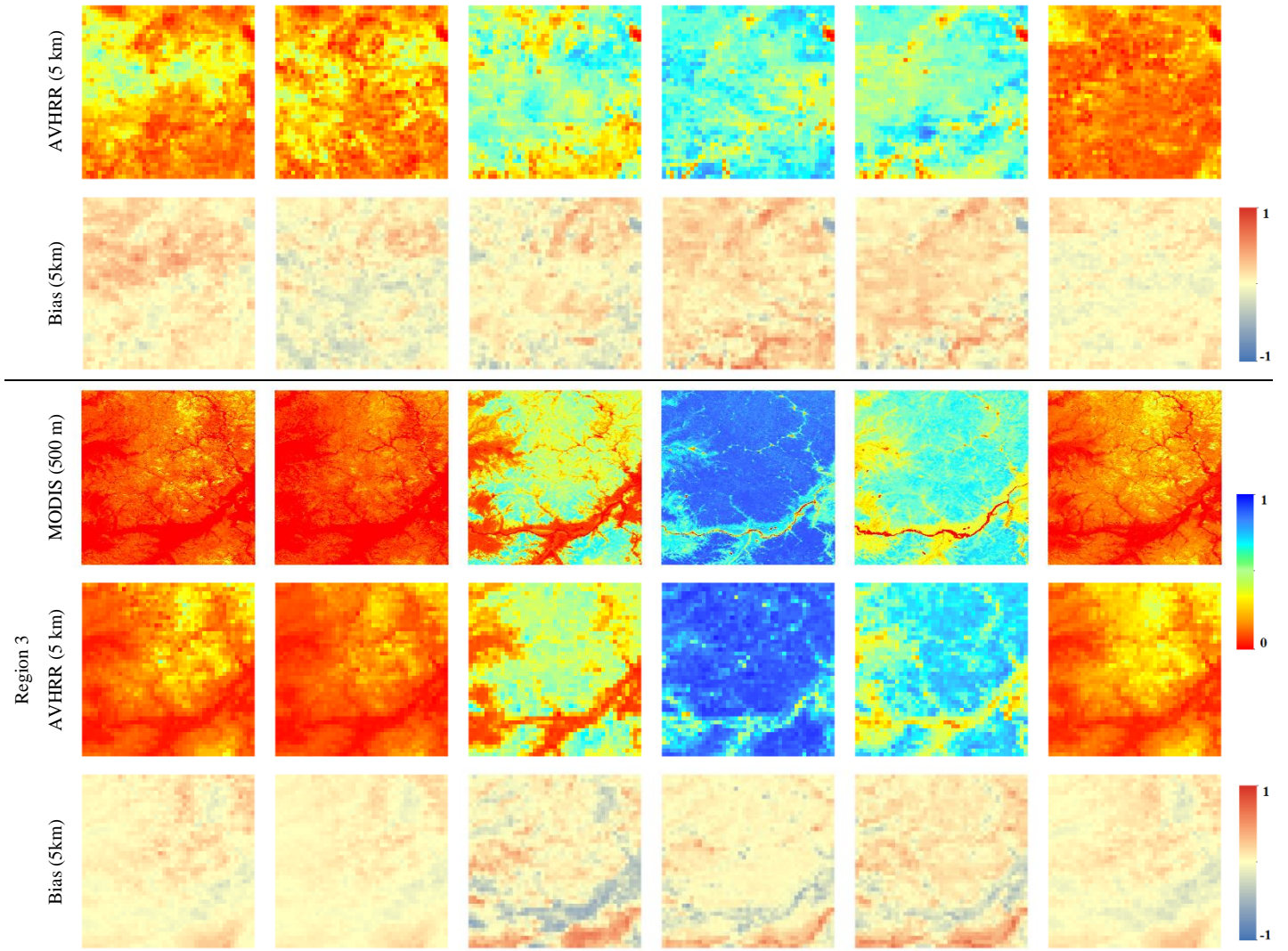
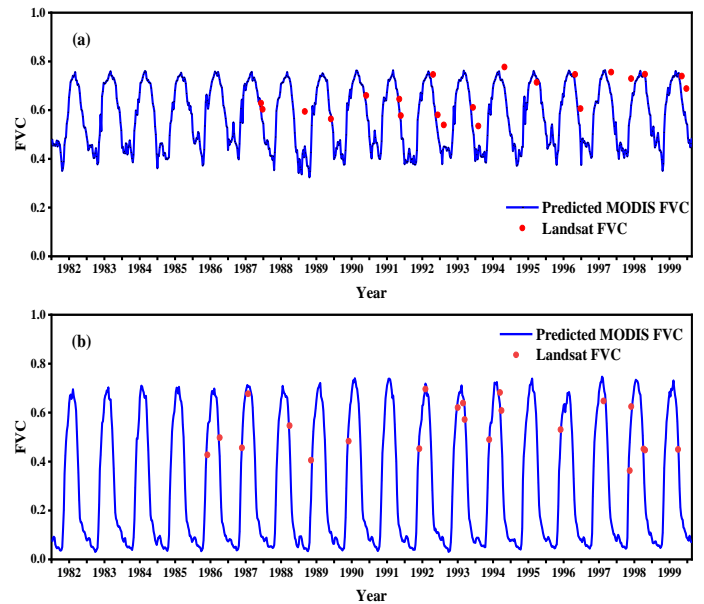


Fig. 12. The predicted historical MODIS FVC dataset, the corresponding GLASS-AVHRR FVC product and the Bias maps at 5 km resolution (using the GLASS-AVHRR FVC as a reference) in three different regions.

2) *Temporal consistency with Landsat FVC.* The time-series of the predicted historical MODIS FVC at three different sites (denoted as Sites A, B and C) from 1982–2000 is shown in Fig. 13, and the Landsat FVC was added as a reference. Due to cloud contamination and differences in temporal resolution, the number of Landsat data acquisitions was limited. Specifically, 21, 22 and 20 Landsat data points were acquired for the corresponding three sites. The three sites were based on MODIS pixels with a 500 m spatial resolution, and the aggregated 500 m Landsat FVCs were plotted. Fig. 13 shows that most of the red dots (i.e., Landsat FVC) are distributed near the blue line (i.e., historical MODIS FVC), suggesting that the historical MODIS FVC agrees with Landsat FVC for the period of interest. In addition, the FVC time-series curves show regular seasonal changes from 1982–2000 at each site. Table 6 shows the overall accuracy assessment for the three sites. For the site in Fig. 13(a), the CC value is 0.86, and the RMSE, Bias and ubRMSE values are 0.08, 0.07 and 0.05, respectively. Moreover, the CCs are 0.79 and 0.68 for the site in Fig. 13(b) and Fig. 13(c). Therefore, the temporal profile of the predicted historical MODIS FVC is in satisfactory agreement with the Landsat FVC.



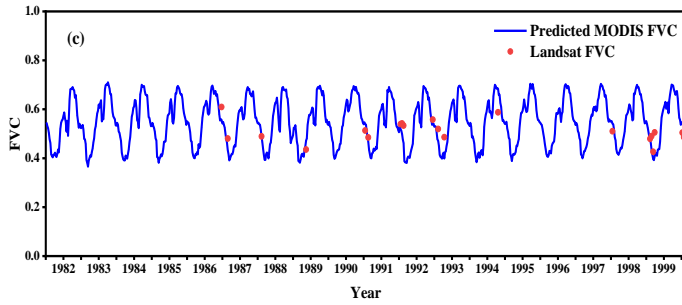


Fig. 13. The temporal profile of predicted historical MODIS FVC time-series at three different sites (at 500 m spatial resolution) from 1982–2000, where the corresponding Landsat FVC is added as reference. (a)–(c) are the results for Sites A–C.

Table 6. Accuracy assessment of the temporal profile of the predicted historical MODIS FVC at three different sites.

Site	The number of Landsat FVC	CC	RMSE	Bias	ubRMSE
A	21	0.8634	0.0821	0.0678	0.0464
B	22	0.7924	0.0790	0.0105	0.0783
C	20	0.6840	0.0507	0.0167	0.0479

V. DISCUSSION

A. FVC change trend analysis

The annual average FVC change trends in China, calculated from the historical MODIS FVC dataset predicted in this paper, are shown in Fig. 14(a). The red line represents the FVC linear fitted change trend. Although there are some fluctuations in the annual FVC average value in China from 1982–1999, the overall FVC change trend is increasing. This also shows that the area covered by vegetation in China has increased from 1982–1999. In addition, to further observe the relationship between trends in vegetation cover and changes in the corresponding vegetation policies, we used the example of China's Three North Region, where the Three-North Shelterbelt Forest Program (TNSFP) was implemented [48]. Specifically, the TNSFP is a large-scale artificial forestry ecological project started in 1978 and planned to end in 2025, which is a strategic program to improve the ecological environment, reduce natural disasters and maintain living spaces [49]. For the range of China's Three North Region, we referred to the Resource and Environment Science and Data Center (<https://www.resdc.cn/data.aspx?DATAID=138>). Considering summer is the time when vegetation productivity is at a maximum, this paper calculated the annual average FVC values of China's Three North Region in summer (i.e., June to August) from 1982–1999 and plotted them on a continuous curve in Fig. 14(b). It can be clearly observed that the general change trend of FVC values in the China's Three North Region is also upward. This correlates with the work of the TNSFP on afforestation and increasing forest cover during this period. Moreover, the slope of the linearly fitted line for the China's Three North Region in summer is larger than those for the entire China region. This suggests that the predicted historical MODIS FVC result is a reliable dataset that reflects true FVC changes.

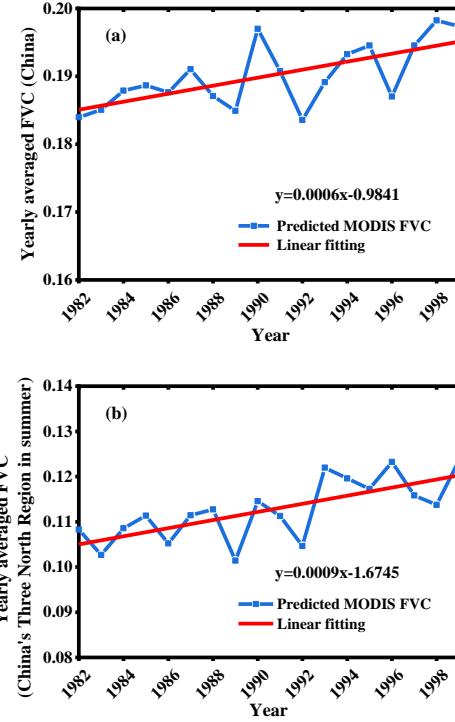


Fig. 14. Annual average FVC variations from 1982–1999 in different spatial extents. (a) China; (b) China's Three North Region.

B. The advantage of the produced historical MODIS FVC dataset

In this paper, we generated a historical MODIS FVC dataset with a 500 m resolution, which increases the spatial resolution of the GLASS-AVHRR FVC product by a factor of 10. To display the advantages of the historical MODIS FVC dataset in terms of spatial scales, we selected the data for two different areas (i.e., the Dongting Lake and the Yangtze River) in 1982 for comparison with the GLASS-AVHRR FVC product. The comparison results and the Bias maps (5 km) for the river and lake areas are shown in Fig. 15, where the black line outlines the specific extent. It is not difficult to find that the predicted MODIS FVC dataset is filled with red (i.e., representing that the FVC is zero) for the corresponding lake and river areas, whereas this is not the case for the AVHRR FVC dataset. For example, in DOY 257, 1982, the river area in the MODIS FVC dataset is completely dominated by red, but the corresponding area in the AVHRR FVC dataset is lighter than red. For the Bias maps, we can find that the differences are larger in the area of lakes and rivers. In addition, it is difficult to observe the shapes of lakes and rivers at a 5 km spatial resolution.

We calculated the FVC values for both datasets at three different sites in the selected lake (denoted as Sites D, E and F) and river areas (denoted as Sites G, H and I) for the entirety of 1982, as shown in Fig. 16. From Fig. 16(a1) and (b1), we can find that the FVC values for the predicted historical MODIS FVC dataset were consistently equal to zero throughout 1982, whereas the FVC values for the AVHRR FVC dataset are all greater than 0, whether in the lake or river areas. The reason is that the AVHRR FVC dataset contains a large amount of land cover information in one pixel with a 5 km resolution, and the produced MODIS FVC dataset can more accurately distinguish

different vegetation covers by increasing the resolution to 500 m. 500 m resolution can provide more detailed spatial information. It is, thus, demonstrated that the produced MODIS FVC with

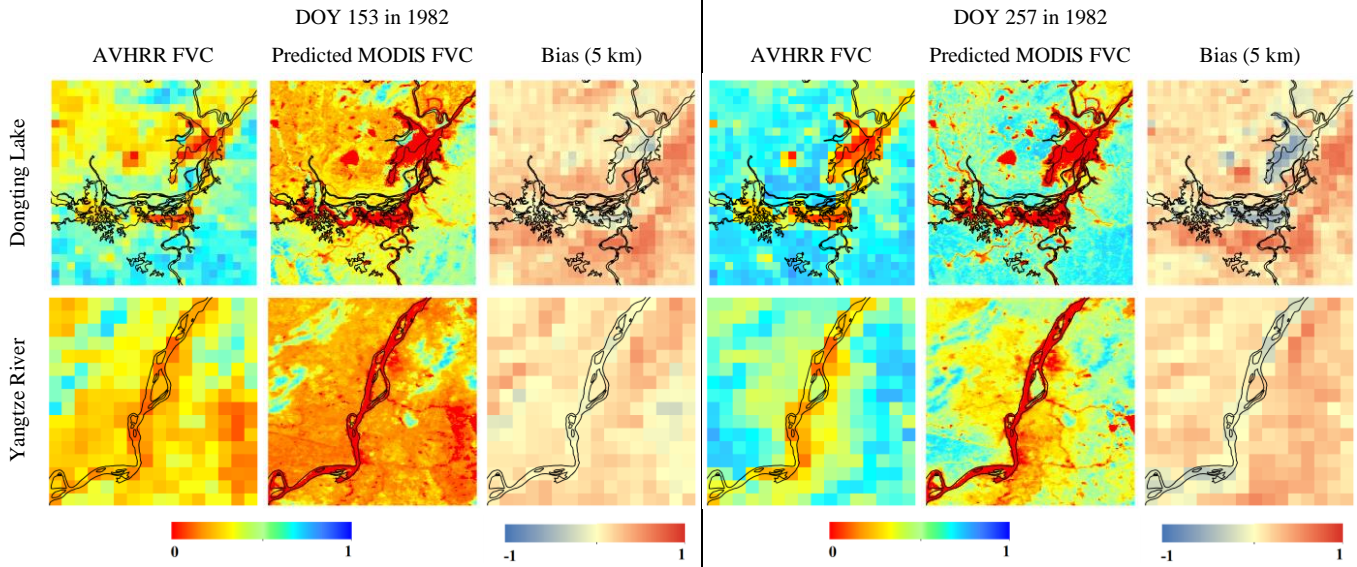


Fig. 15. Visual comparison and Bias maps (the AVHRR FVC as reference) between the AVHRR and predicted MODIS FVC datasets for a section of Dongting Lake and the Yangtze River at two time points (i.e., DOY 153 and 257) in 1982. The black line outlines the specific extent of Dongting Lake and the Yangtze River.

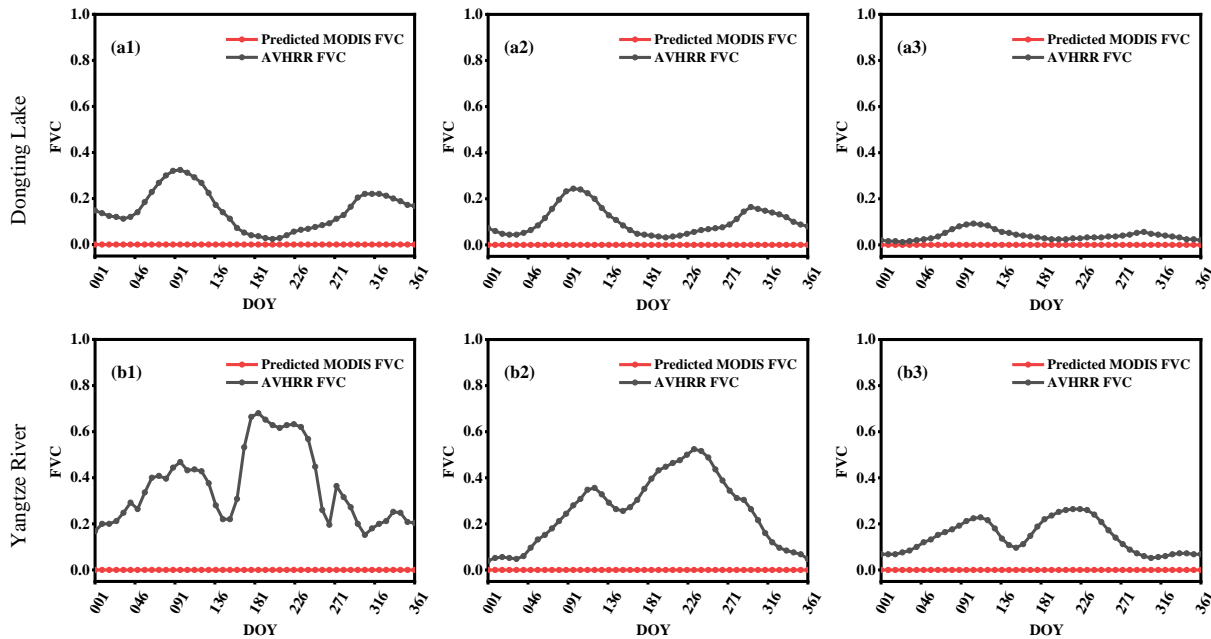


Fig. 16. The temporal profile of FVC at three different sites in the selected lake and river areas in 1982. (a1)-(a3) are the results for Sites D-F and (b1)-(b3) are the results for Sites G-I.

C. Uncertainty in the training data

In this paper, the GLASS-AVHRR FVC dataset was used as the training data to train the RF model. The quality of the training samples will influence the subsequent results. However, the GLASS-AVHRR FVC dataset, as an inversion product, also contains certain uncertainty. We selected the Xinjiang Uygur Autonomous Region in western China for further analysis. The main land cover types of the region are shown in Fig. 17. Additionally, a visual comparison between GLASS-AVHRR

FVC and predicted MODIS FVC in 1999 at different time points (i.e., DOY 001, 161 and 321) for the region is displayed in Fig. 18. From Fig. 17, it can be seen that while most of the area is dominated by barren land, a distinct cover of grasslands and croplands exists in the northern part. However, in Fig. 18, we can find that a large proportion of the GLASS-AVHRR FVC dataset in the grassland covered areas of the northern Xinjiang contains values of 0, which leads to an underestimation of FVC in this region. Based on this observation, the corresponding

predicted MODIS FVC inherits this uncertainty, that is, also with values of 0 in the northern part.

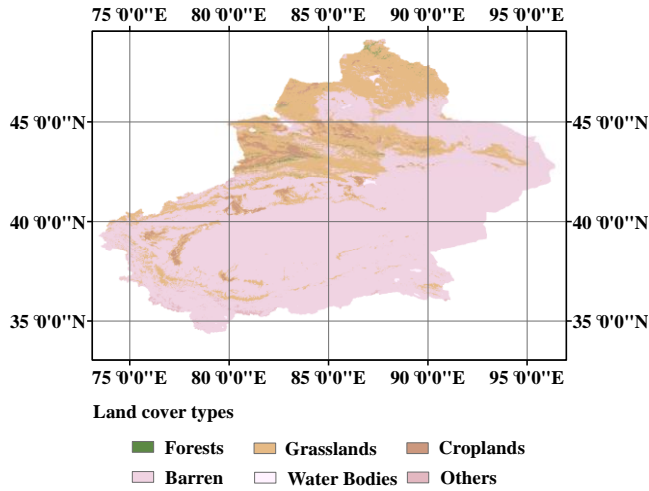


Fig. 17. The land cover types of the Xinjiang Uygur Autonomous Region.

D. Uncertainty in prediction

One source of uncertainty in the present research is the imperfect relationship between the GLASS-AVHRR and GLASS-MODIS FVC products. In this study, based on the similarity in the FVC temporal profiles between the GLASS-MODIS and GLASS-AVHRR FVC products in China (see Figs. 2 and 3), the GLASS-AVHRR FVC product change

pattern (China) was migrated to the GLASS-MODIS FVC product (China). Although the above assumption can be confirmed by the predicted FVC results, the differences between the GLASS-AVHRR and GLASS-MODIS FVC products cannot be ignored, not least because the acquisition platforms of the two products are different. In particular, the AVHRR and MODIS data differ in viewing angles and spectral wavelengths, which further contribute to the differences in the two types of observations. Moreover, due to the substantial difference in spatial resolution between MODIS and AVHRR data (500 m vs. 5 km), the transfer of GLASS-AVHRR FVC to GLASS-MODIS FVC can bring scale uncertainty, especially when the distribution of land cover types is complex and local heterogeneity is great (e.g., dominated by objects with spatial size smaller than 5 km, presenting mixed pixels in AVHRR data). Thus, the existence of the observed discrepancies will cause errors in the estimated historical MODIS FVC dataset during 1982–2000 to a certain extent. In addition, the RF model was selected as the training and prediction model. This is because RF can process a large number of training samples and performs well in fitting nonlinear relationships. With the development of machine learning, some deep learning networks have the advantages of processing complicated data, avoiding noise effects and improving computational efficiency [50–53]. Therefore, further studies may consider using deep learning model to increase the accuracy of predicting historical FVC time-series datasets.

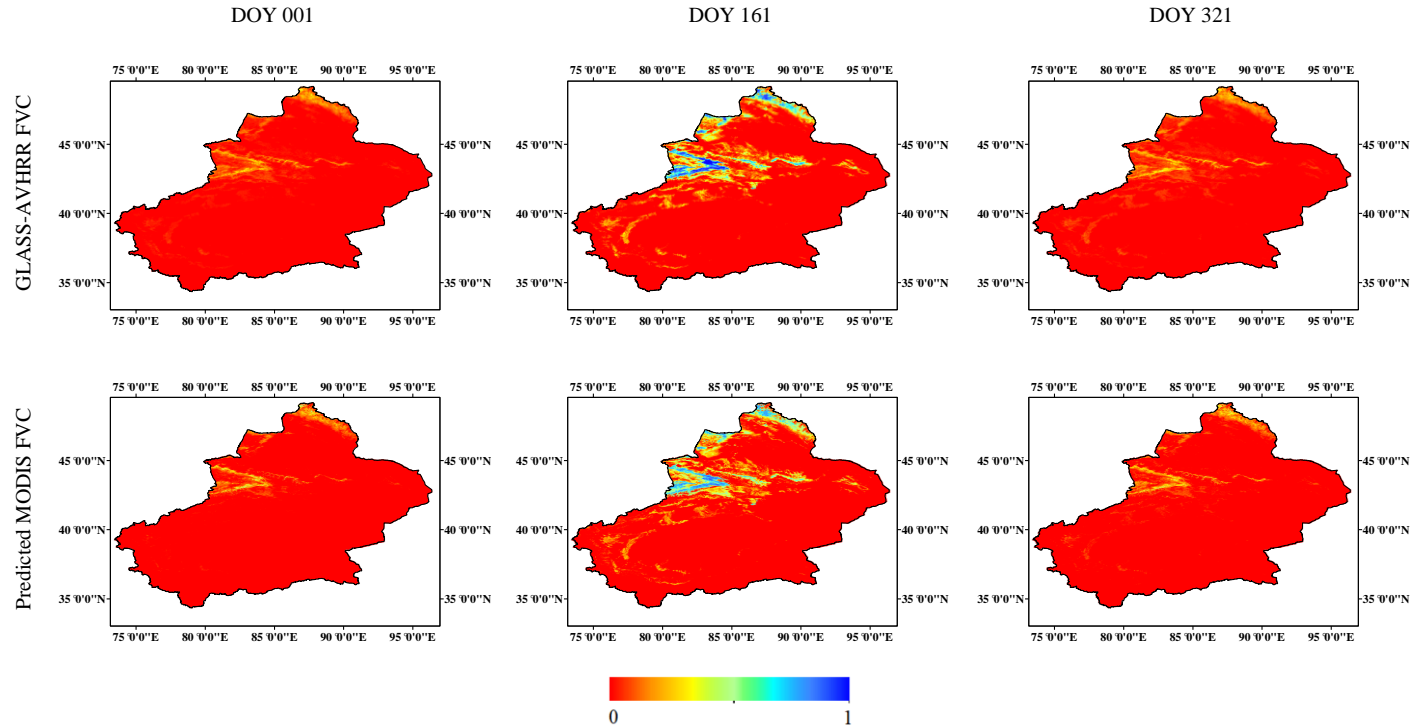


Fig. 18. Visual comparison between the GLASS-AVHRR FVC and predicted MODIS FVC datasets for the Xinjiang Uygur Autonomous Region at three different time points in 1999.

E. Limitations in validation

Due to the lack of ground measurement FVC data in the historical period, we used Landsat FVC data, which were calculated using the dimidiate pixel model, as reference data with which to validate the predicted MODIS FVC results from 1982–2000. First, the revisit interval of Landsat data is 16 days and is susceptible to cloud contamination. Moreover, Landsat data cannot guarantee complete cloud-free coverage in China and, therefore, the validation of reconstructed FVC data in China is also limited. Second, we calculated the NDVI values from the Landsat reflectance data and used the dimidiate pixel model to calculate the FVC. For the dimidiate pixel model (Eq. (3)), the key parameter definitions, that is $NDVI_s$ and $NDVI_v$, were based on the assumption that full soil and vegetation pixels existed for each Landsat image. However, the actual situations in different ecosystems do not exactly match these assumptions. Therefore, there is uncertainty in the definitions of $NDVI_s$ and $NDVI_v$. This has also been the subject of discussion by many scholars [44]. In the future, a more extensive validation of the predicted MODIS FVC dataset would be required, with more reliable reference data, including historical ground measurements and other finer spatial resolution remote sensing data.

VI. CONCLUSION

In this paper, the historical FVC dataset for China from 1982–2000, with a 500 m spatial resolution and 8-day frequency, was derived from the GLASS-AVHRR and GLASS-MODIS FVC products. We assumed that the FVC change pattern of the GLASS-AVHRR FVC product could be migrated to the GLASS-MODIS FVC product. Specifically, the time-series training samples were created by the existing GLASS-AVHRR FVC for the Chinese region, with the GLASS-AVHRR FVC from 2002–2005 as input and known historical GLASS-AVHRR FVC data (i.e., within 1982–2000) as the label. The historical MODIS FVC dataset was predicted using the trained RF model, with the MODIS FVC from 2002–2005 as input. The accuracy was evaluated using 48 Landsat tiles. The results indicated that the predicted historical MODIS FVC dataset performed satisfactorily, with a CC and RMSE of 0.84 and 0.14, respectively. In addition, the predicted historical MODIS FVC dataset was demonstrated to be more accurate than the GEOV2 FVC product (that is, the average values of CC and RMSE are increased by 0.03 and decreased by 0.02, respectively) for seven different regions in 1999. According to an analysis of spatio-temporal consistency, the historical MODIS FVC dataset reflects reliably the seasonal change in vegetation, and corresponds well with the Landsat FVC data in the temporal domain. Overall, the predicted historical MODIS FVC dataset expands the temporal length of the GLASS-MODIS FVC product in China, which begins in 1982, and can present much more spatial details at a finer spatial resolution of 500 m. To the best of our knowledge, this is the FVC time-series with the finest spatial resolution for 1982–2000. The 500 m FVC data will be of great value for various applications domains (e.g., climate change and carbon cycle) at regional scale. The dataset is available at <https://doi.org/10.6084/m9.figshare.24616446.v1>.

APPENDIX

Specific $NDVI_s$ and $NDVI_v$ values for each Landsat tile used for validation are listed in Table A1.

Table A1. The $NDVI_s$ and $NDVI_v$ values for each Landsat tile used for validation.

Date	Path/Row	$NDVI_s$	$NDVI_v$
1999/03/04	130/042	0.0592	0.2961
1999/05/18	127/035	0.0851	0.4032
1999/05/27	126/037	0.1215	0.4478
1999/06/29	117/030	0.1269	0.4597
1998/05/25	117/030	0.0758	0.4703
1998/05/26	124/033	0.0761	0.3453
1998/05/28	122/032	0.0508	0.4460
1998/10/24	125/044	0.0418	0.3739
1997/05/25	122/025	0.1030	0.3087
1997/05/28	127/035	0.0832	0.4223
1997/06/13	127/035	0.1025	0.4134
1997/06/23	117/029	0.1426	0.4877
1996/04/12	130/042	0.0629	0.3016
1996/05/22	122/032	0.0549	0.3619
1996/09/14	127/035	0.1185	0.3302
1996/12/14	124/044	0.0677	0.3377
1995/05/07	127/035	0.0610	0.3361
1995/05/09	125/041	0.0985	0.3794
1995/06/08	127/035	0.1052	0.3937
1995/06/19	124/033	0.0577	0.3425
1994/05/17	122/032	0.0558	0.3660
1994/05/27	128/036	0.1044	0.4648
1994/06/15	117/030	0.0580	0.4765
1994/08/22	145/030	0.0230	0.4291
1993/04/04	130/042	0.0547	0.2939
1993/05/14	122/032	0.0524	0.3644
1993/09/22	127/035	0.0957	0.3447
1993/12/06	124/044	0.0517	0.3490
1992/02/13	130/042	0.0426	0.2720
1992/05/14	127/035	0.0962	0.3684
1992/05/27	122/032	0.0655	0.3829
1992/07/17	127/035	0.0927	0.3857
1991/02/10	130/042	0.0379	0.3027
1991/05/25	122/025	0.0875	0.2296
1991/11/15	124/043	0.0832	0.3074
1991/11/15	124/044	0.1020	0.2831
1990/03/11	130/042	0.0430	0.2916
1990/05/18	126/037	0.1323	0.4504
1990/05/22	122/032	0.0813	0.4084
1990/11/28	124/044	0.0684	0.3298
1989/05/03	122/032	0.0521	0.3115
1989/05/11	130/042	0.0626	0.2945
1989/06/01	117/028	0.0713	0.4340
1989/12/02	125/044	0.0496	0.3043
1988/02/02	130/042	0.0304	0.2968
1988/06/04	127/038	0.1671	0.4617

1988/06/06	125/041	0.1232	0.4389
1988/06/25	114/028	0.1209	0.5034
1987/02/15	130/042	0.0282	0.2968
1987/05/14	122/032	0.0587	0.3547
1987/06/02	127/038	0.1212	0.4237
1987/10/08	127/035	0.0803	0.2671
1986/05/27	122/032	0.0676	0.3790
1986/06/04	114/028	0.0705	0.4606
1986/06/09	117/030	-0.0121	0.4742
1986/12/13	130/042	-0.0033	0.5426
1985/05/14	116/030	0.0681	0.3481
1985/05/16	114/028	0.0541	0.3920
1985/05/21	117/029	0.0536	0.4399
1985/05/21	117/030	0.0609	0.4479
1984/05/18	117/027	0.0128	0.3642
1984/05/18	117/028	0.0484	0.3666
1984/05/18	117/029	0.0531	0.3769
1984/05/18	117/030	0.0725	0.2914

REFERENCES

- [1] W. Knorr, T. Kaminski, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P. P. Mathieu, "Carbon cycle data assimilation with a generic phenology model," *Journal of Geophysical Research*, vol. 115, G04017, 2010.
- [2] K. Wang, A. Bastos, P. Ciais, X. Wang, C. Rodenbeck, P. Gentine, F. Chevallier, V. W. Humphrey, C. Huntingford, M. O'Sullivan, S. I. Seneviratne, S. Sitch, and S. Piao, "Regional and seasonal partitioning of water and temperature controls on global land carbon uptake variability," *Nature Communications*, vol. 13, no. 1, p. 3469, 2022.
- [3] X. Jiang, A. D. Ziegler, S. Liang, D. Wang, and Z. Zeng, "Forest Restoration Potential in China: Implications for Carbon Capture," *Journal of Remote Sensing*, vol. 2022, p. 0006, 2022.
- [4] Z. Chang, L. Fan, J.-P. Wigneron, Y.-P. Wang, P. Ciais, J. Chave, R. Fensholt, J. M. Chen, W. Yuan, W. Ju, X. Li, F. Jiang, M. Wu, X. Chen, Y. Qin, F. Frappart, X. Li, M. Wang, X. Liu, X. Tang, S. Hobeichi, M. Yu, M. Ma, J. Wen, Q. Xiao, W. Shi, D. Liu, and J. Yan, "Estimating Aboveground Carbon Dynamic of China Using Optical and Microwave Remote-Sensing Datasets from 2013 to 2019," *Journal of Remote Sensing*, vol. 3, p. 0005, 2023.
- [5] S. Yang, Q. Feng, T. Liang, B. Liu, W. Zhang, and H. Xie, "Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region," *Remote Sensing of Environment*, vol. 204, pp. 448–455, 2018.
- [6] X. Lian, S. Jeong, C. E. Park, H. Xu, L. Z. X. Li, T. Wang, P. Gentine, J. Penuelas, and S. Piao, "Biophysical impacts of northern vegetation changes on seasonal warming patterns," *Nature Communications*, vol. 13, no. 1, p. 3925, 2022.
- [7] Y. Zhang, S. Piao, Y. Sun, B. M. Rogers, X. Li, X. Lian, Z. Liu, A. Chen, and J. Penuelas, "Future reversal of warming-enhanced vegetation productivity in the Northern Hemisphere," *Nature Climate Change*, vol. 12, no. 6, pp. 581–586, 2022.
- [8] R. Yao, L. Wang, X. Huang, Y. Liu, Z. Niu, S. Wang, and L. Wang, "Long-term trends of surface and canopy layer urban heat island intensity in 272 cities in the mainland of China," *Science of The Total Environment*, vol. 772, 2021.
- [9] A. A. Gitelson, Y. J. Kaufman, R. Stark, and D. Rundquist, "Novel algorithms for remote estimation of vegetation fraction," *Remote Sensing of Environment*, vol. 80, no. 1, pp. 76–87, 2002.
- [10] G. Jiapaer, X. Chen, and A. Bao, "A comparison of methods for estimating fractional vegetation cover in arid regions," *Agricultural and Forest Meteorology*, vol. 151, no. 12, pp. 1698–1710, 2011.
- [11] J. Bian, A. Li, Z. Zhang, W. Zhao, G. Lei, G. Yin, H. Jin, J. Tan, and C. Huang, "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 Sensing of Environment*, vol. 197, pp. 98–114, 2017.
- [12] K. Jia, S. Liang, S. Liu, Y. Li, Z. Xiao, Y. Yao, B. Jiang, X. Zhao, X. Wang, S. Xu, and J. Cui, "Global land surface fractional vegetation cover estimation using general regression neural networks from MODIS surface reflectance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4787–4796, 2015.
- [13] J. Xiao and A. Moody, "A comparison of methods for estimating fractional green vegetation cover within a desert-to-upland transition zone in central New Mexico, USA," *Remote Sensing of Environment*, vol. 98, no. 2-3, pp. 237–250, 2005.
- [14] D. Liu, K. Jia, H. Jiang, M. Xia, G. Tao, B. Wang, Z. Chen, B. Yuan, and J. Li, "Fractional vegetation cover estimation algorithm for FY-3B reflectance data based on random forest regression method," *Remote Sensing*, vol. 13, no. 11, 2021.
- [15] J. Sun, S. Shi, J. Yang, W. Gong, F. Qiu, L. Wang, L. Du, and B. Chen, "Wavelength selection of the multispectral lidar system for estimating leaf chlorophyll and water contents through the PROSPECT model," *Agricultural and Forest Meteorology*, vol. 266-267, pp. 43–52, 2019.
- [16] C. Quintano, A. Fernández-Manso, Y. E. Shimabukuro, and G. Pereira, "Spectral unmixing," *International Journal of Remote Sensing*, vol. 33, no. 17, pp. 5307–5340, 2012.
- [17] C. Small, "Estimation of urban vegetation abundance by spectral mixture analysis," *International Journal of Remote Sensing*, vol. 22, no. 7, pp. 1305–1334, 2010.
- [18] F. Xu, S. Heremans, and B. Somers, "Urban land cover mapping with Sentinel-2: a spectro-spatio-temporal analysis," *Urban Informatics*, vol. 1, no. 1, 2022.
- [19] Q. Wang, X. Ding, X. Tong, and P. M. Atkinson, "Spatio-temporal spectral unmixing of time-series images," *Remote Sensing of Environment*, vol. 259, 2021.
- [20] X. Ding, Q. Wang, and X. Tong, "Integrating 250 m MODIS data in spectral unmixing for 500 m fractional vegetation cover estimation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 111, 2022.
- [21] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau, "Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation," *Remote Sensing of Environment*, vol. 105, no. 4, pp. 313–325, 2006.
- [22] S. Garrigues, R. Lacaze, F. Baret, J. T. Morisette, M. Weiss, and J. E. Nickeson, "Validation and intercomparison of global Leaf Area Index products derived from remote sensing data," *Journal of Geophysical Research*, vol. 113, no. 0148-0227, G02028, 2008.
- [23] J. Zhao, J. Li, Q. Liu, B. Xu, W. Yu, S. Lin, and Z. Hu, "Estimating fractional vegetation cover from leaf area index and clumping index based on the gap probability theory," *International Journal of Applied Earth Observation and Geoinformation*, vol. 90, 2020.
- [24] K. Jia, L. Yang, S. Liang, Z. Xiao, X. Zhao, Y. Yao, X. Zhang, B. Jiang, and D. Liu, "Long-term global land surface satellite (GLASS) fractional vegetation cover product derived from MODIS and AVHRR Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 2, pp. 508–518, 2019.
- [25] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, pp. 568–576, 1991.
- [26] D. F. Specht, "The general regression neural network-Rediscovered," *Neural Networks*, vol. 6, no. 7, pp. 1033–1034, 1993.
- [27] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, pp. 1–67, 1991.
- [28] J. Stoklosa and D. I. Warton, "A generalized estimating equation approach to multivariate adaptive regression splines," *Journal of Computational and Graphical Statistics*, vol. 27, no. 1, pp. 245–253, 2018.
- [29] L. Yang, K. Jia, S. Liang, J. Liu, and X. Wang, "Comparison of four machine learning methods for generating the GLASS fractional vegetation cover product from MODIS Data," *Remote Sensing*, vol. 8, no. 8, 2016.
- [30] B. Mu, X. Zhao, D. Wu, X. Wang, J. Zhao, H. Wang, Q. Zhou, X. Du, and N. Liu, "Vegetation Cover Change and Its Attribution in China from 2001 to 2018," *Remote Sensing*, vol. 13, no. 3, 2021.
- [31] Z. Hu, L. Chai, W. T. Crow, S. Liu, Z. Zhu, J. Zhou, Y. Qu, J. Liu, S. Yang, and Z. Lu, "Applying a wavelet transform technique to optimize general fitting models for SM analysis: A case study in downscaling over the Qinghai-Tibet Plateau," *Remote Sensing*, vol. 14, no. 13, 2022.
- [32] S. Liang, X. Zhao, S. Liu, W. Yuan, X. Cheng, Z. Xiao, X. Zhang, Q. Liu, J. Cheng, H. Tang, Y. Qu, Y. Bo, Y. Qu, H. Ren, K. Yu, and J. Townshend, "A long-term Global Land Surface Satellite (GLASS) data-set for environmental studies," *International Journal of Digital Earth*, vol. 6, no. sup1, pp. 5–33, 2013.
- [33] Z. Xiao, S. Liang, X. Tian, K. Jia, Y. Yao, and B. Jiang, "Reconstruction of

- long-term temporally continuous NDVI and surface reflectance from AVHRR data,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5551–5568, 2017.
- [34] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [35] K. Fawagreh, M. M. Gaber, and E. Elyan, “Random forests: from early developments to recent advancements,” *Systems Science & Control Engineering*, vol. 2, no. 1, pp. 602–609, 2014.
- [36] D. F. Parkhurst, K. P. Brenner, A. P. Dufour, and L. J. Wymer, “Indicator bacteria at five swimming beaches-analysis using random forests,” *Water Research*, vol. 39, no. 7, pp. 1354–1360, 2005.
- [37] D. R. Cutler, T. C. Edwards, Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, “Random forests for classification in ecology,” *Ecology*, vol. 88, no. 11, pp. 2783–2792.
- [38] H. Wang, R. Magagi, K. Goita, M. Trudel, H. McNairn, and J. Powers, “Crop phenology retrieval via polarimetric SAR decomposition and Random Forest algorithm,” *Remote Sensing of Environment*, vol. 231, 2019.
- [39] X. Zhang, C. Liao, J. Li, and Q. Sun, “Fractional vegetation cover estimation in arid and semi-arid environments using HJ-1 satellite hyperspectral data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 506–512, 2013.
- [40] M. Zhang, J. Wang, and S. Li, “Tempo-spatial changes and main anthropogenic influence factors of vegetation fractional coverage in a large-scale opencast coal mine area from 1992 to 2015,” *Journal of Cleaner Production*, vol. 232, pp. 940–952, 2019.
- [41] G. Yang, R. Pu, J. Zhang, C. Zhao, H. Feng, and J. Wang, “Remote sensing of seasonal variability of fractional vegetation cover and its object-based spatial pattern analysis over mountain areas,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 77, pp. 79–93, 2013.
- [42] J. Li, J. Wang, J. Zhang, J. Zhang, and H. Kong, “Dynamic changes of vegetation coverage in China-Myanmar economic corridor over the past 20 years,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, 2021.
- [43] R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak, “The interpretation of spectral vegetation indexes,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 481–486, 1995.
- [44] L. Gao, X. Wang, B. A. Johnson, Q. Tian, Y. Wang, J. Verrelst, X. Mu, and X. Gu, “Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 364–377, 2020.
- [45] R. B. Myneni, S. Hoffman, Y. Knyazikhin, J. L. Privette, J. Glassy, and Y. Tian, “Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data,” *Remote Sensing of Environment*, vol. 83, pp. 214–231, 2002.
- [46] L. Feng, Z. He, X. Liu, and J. Zhang, “Dynamic monitoring of the fractional vegetation cover in Jilin province based on MODIS-NDVI data,” *Journal of University of Chinese Academy of Sciences*, vol. 31, no. 4, pp. 492–499, 2014.
- [47] A. Verger, F. Baret, and M. Weiss, “Near real-time vegetation monitoring at global scale,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 8, pp. 3473–3481, 2014.
- [48] X. Suo and S. Cao, “China’s three north shelter forest program: cost-benefit analysis and policy implications,” *Environment, Development and Sustainability*, vol. 23, no. 10, pp. 14605–14618, 2021.
- [49] B. Qiu, G. Chen, Z. Tang, D. Lu, Z. Wang, and C. Chen, “Assessing the Three-North Shelter Forest Program in China by a novel framework for characterizing vegetation changes,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 133, pp. 75–88, 2017.
- [50] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [51] L. Zhang, L. Zhang, and B. Du, “Deep learning for remote sensing data: a technical tutorial on the state of the art,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.
- [52] S. Shi, Y. Zhong, Y. Liu, L. Zhang, and D. Li, “Cross-temporal high spatial resolution urban scene classification and change detection based on a class-weighted deep adaptation network,” *Urban Informatics*, vol. 3, no. 1, 2024.
- [53] M. Hao, S. Chen, H. Lin, H. Zhang, and N. Zheng, “A prior knowledge guided deep learning method for building extraction from high-resolution remote sensing images,” *Urban Informatics*, vol. 3, no. 1, 2024.



Xinyu Ding received the B.S. degree from Jiangxi University of Science and Technology, Ganzhou, China, in 2020. She is currently pursuing the Ph.D. degree with Tongji University, Shanghai, China.

Her research interest includes remote sensing image processing, spectral unmixing, and time-series analysis.



Qunming Wang received the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018. His 3-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored over 100 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing*, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Professor Wang serves as Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering & Remote Sensing*, and was Associate Editor for *Computers and Geosciences* (2017–2020).



Haoxuan Yang received the B.S. and M.S. degrees in land resources management from Northeast Agricultural University, Harbin, China, in 2016 and 2020, respectively. He received the Ph.D. degree in civil and hydraulic engineering from Tongji University, Shanghai, China in 2024. Currently, he is a Postdoctoral fellow with the Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, USA.

His research interests focus on soil properties mapping, reconstruction, and remote sensing.



Peter M. Atkinson received the Ph.D. degree from the University of Sheffield (NERC CASE award with Rothamsted Experimental Station) in 1990. More recently, he received the MBA degree from the University of Southampton in 2012.

He is currently Distinguished Professor of Spatial Data Science and Dean of the Faculty of Science and Technology at Lancaster University, UK. He was previously Professor of Geography at the University Southampton, where he is

currently Visiting Professor. He is also Visiting Professor at the Chinese Academy of Sciences, Beijing. He previously held the Belle van Zuylen Chair at Utrecht University, the Netherlands, is a recipient of the Peter Burrough Award of the International Spatial Accuracy Research Association and is a Fellow of the Learned Society of Wales. The main focus of his research is in remote sensing, geographical information science and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems. He has published over 300 peer-reviewed articles in international scientific journals and around 50 refereed book chapters. He has also edited nine journal special issues and eight books.

Professor Atkinson is Editor-in-Chief of *Science of Remote Sensing*, a sister journal of *Remote Sensing of Environment*. He also sits on the editorial boards of several further journals including *Geographical Analysis*, *Spatial Statistics*, *International Journal of Applied Earth Observation and Geoinformation*, and *Environmental Informatics*.