

Generalized Fine-Resolution FPAR Estimation Using Google Earth Engine: Random Forest or Multiple Linear Regression

Yiting Wang , Yinggang Zhan , Guangjian Yan , *Senior Member, IEEE*, and Donghui Xie 

I. INTRODUCTION

Abstract—Accurate estimation of fine-resolution fraction of absorbed photosynthetically active radiation (FPAR) is urgently needed for modeling land surface processes at finer scales. While traditional methods can hardly balance universality, efficiency, and accuracy, methods using coarse-resolution products as a reference are promising for operational fine-resolution FPAR estimation. However, current methods confront major problems of underrepresentation of FPAR-reflectance relations within coarse-resolution FPAR products, particularly for densely vegetated areas. To overcome this limitation, this article has developed an enhanced scaling method that proposes an outlier removal procedure and a method weighting the selected samples and models FPAR through weighted multiple linear regression (MLR) between the coarse-resolution FPAR product and the aggregated fine-resolution surface reflectance. Meanwhile, a random forest regression (RFR) method has also been implemented for comparison. Both methods were particularly applied to Landsat 8 OLI and moderate resolution imaging spectroradiometer (MODIS) FPAR data on the Google earth engine. Their performance was tested on a regional scale for an entire year. The results of the enhanced scaling method were closer to the in situ measurements (RMSE = 0.058 and $R^2 = 0.768$) and were more consistent with the MODIS FPAR (RMSE = 0.091 and $R^2 = 0.894$) than those of the RFR, particularly over densely vegetated pixels. This indicates that a well-designed simple MLR-based method can outperform the more sophisticated RFR method. The enhanced scaling method is also less sensitive to the number of training samples than the RFR method. Moreover, both methods are insensitive to land cover maps, and their computation efficiency depends on the number of images to be estimated.

Index Terms—Fraction of absorbed photosynthetically active radiation (FPAR), Google earth engine (GEE), Landsat, moderate resolution imaging spectroradiometer (MODIS), remote sensing trend surface, scaling.

Manuscript received 18 October 2022; revised 4 December 2022; accepted 14 December 2022. Date of publication 19 December 2022; date of current version 5 January 2023. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 41901301, in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JQ-739, in part by the NSFC under Grant 41971306, and in part by the Open Research Fund Program of State Key Laboratory of Ecohydraulics in Northwest Arid Region under Grant 2020KFKT-7. (*Corresponding author: Yiting Wang.*)

Yiting Wang and Yinggang Zhan are with the College of Geomatics, Xi'an University of Science and Technology, Xi'an 710054, China (e-mail: wyt_rs@163.com; zygrs@stu.xust.edu.cn).

Guangjian Yan and Donghui Xie are with the State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China, and also with the Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China (e-mail: gjyan@bnu.edu.cn; xiedonghui@bnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3230456

THE fraction of absorbed photosynthetically active radiation (FPAR) is defined as the fraction of photosynthetically active radiation (PAR) absorbed by vegetation, where PAR refers to the incoming solar radiation within the spectral range of 0.4–0.7 μm [1]. It provides critical information on vegetation photosynthesis for modeling the energy-and-carbon exchange between the biosphere and atmosphere. Numerous land surface process models, e.g., soil-vegetation-atmosphere transfer models [2] and production efficiency models [3], require spatially explicit FPAR data as input to calculate the vegetation productivity, energy budget, and carbon transfer [4]. As the only efficient approach to acquire large-scale spatially continuous FPAR [5], accurate estimation from satellite data is increasingly important [6]. While many global satellite FPAR products are currently available at coarse spatial resolutions (250 m–1 km) [7], [8], [9], fine-resolution FPAR products are urgently needed for modeling finer-scale land surface processes and impacts [6], [10], [11]. However, despite the many methods developed for FPAR estimation from satellite data, they still have a long way to go from algorithm development to operational fine-resolution FPAR estimation in terms of algorithm generalization, computation efficiency, accuracy, and robustness.

Traditional methods for FPAR estimation include two main groups: physical and empirical methods. Physical methods use radiative transfer equations [7], gap probability [12], or recollisional probability [2], [13] to model the relations between FPAR and canopy reflectance through a physical description of the light transfer process. Physical methods have been successfully used in generating global operational FPAR products, e.g., moderate resolution imaging spectroradiometer (MODIS) leaf area index (LAI)/FPAR products [7], but they can hardly adapt to fine-resolution satellite data because they require many input parameters, such as the clumping index and soil albedo. Empirical methods model the linear or nonlinear relations between reflectance and FPAR from radiative transfer simulations [14], [15], [16] or in situ measurements [17]. They are simple and efficient and have been widely used at local scales, but they may experience major difficulties in extending an established empirical model to wider areas or longer time periods.

Recently, an alternative approach has been to model the complex relations between reflectance and vegetation parameters using operational coarse-resolution products as a reference through

statistical learning methods. Multiple linear regression (MLR) was initially used to estimate Landsat LAI [18], and then the random forest regression (RFR) method was also used to estimate LAI from Landsat data [19] or FPAR from sentinel-2 data [20] using MODIS LAI/FPAR products as a reference. These methods inherited the sound physical foundations of the operational coarse-resolution products [7], [8], [9], but their major challenge is to accurately represent the complex FPAR-reflectance relations contained in the operational coarse-resolution FPAR product that are highly variable in space and time and may not adapt to fine resolutions due to the scale effect [21], [22]. Both the MLR in [18] and RFR in [19], [20] selected homogeneous MODIS pixels from multiple images to capture the LAI-reflectance relations and avoid scaling issues. Considering the difference in FPAR-reflectance relations between different solar zenith angles (SZAs) on different dates, Wang et al. [23] proposed a scaling method to estimate fine-resolution FPAR via MLR learning from each pair of MODIS FPAR and Landsat images. However, these methods still face the problem of underrepresentation of FPAR-reflectance relations over densely vegetated areas, where fewer homogeneous MODIS pixels can be selected and the fine-resolution reflectance saturates with increasing vegetation density. This underrepresentation results in the underestimation of high FPAR values and prevents the further application of these methods to the operational estimation of fine-resolution FPAR. Therefore, a generalized method for fine-resolution FPAR estimation is urgently needed.

In addition, although both the RFR and MLR methods have been successfully used in retrieving fine-resolution vegetation parameters, few studies have compared their performance in generating operational fine-resolution FPAR products. The RFR method is advantageous over other traditional machine learning methods, e.g., artificial neural networks, for its capability of dealing with high dimensional problems and high computation efficiency, while the MLR method is simple, efficient, and can work with very few samples. The RFR method was found to outperform the simpler MLR method in downscaling MODIS LST products [24], [25], but the performance of a statistical learning algorithm depends on many factors. Whether the more complex RFR method outweighs the simpler MLR method remains an unsolved issue.

Therefore, this article aims to develop a generalized method for the operational estimation of fine-resolution FPAR based on the scaling method proposed by Wang et al. [23], hereafter referred to as the enhanced scaling method. To overcome the underrepresentation of FPAR-reflectance relations over densely vegetated pixels, the enhanced scaling method proposed a method weighting the selected samples, and modeled the FPAR-reflectance relations through a weighted MLR method. Other improvements include the self-adaptive selection of representative samples and an outlier removal procedure. The RFR method was also developed for comparison, and the retrieval chain was modified from the RFR method used for LAI estimation in [19]. Both the MLR-based enhanced scaling method and RFR method were implemented on the Google earth engine (GEE) to provide operational fine-resolution FPAR products. The sensitivity of the two methods to the number of training

samples and to different land cover maps and their computation efficiency were investigated in detail. The performance of the two methods was compared over a large area covering more than 50 000 km² and during the entire year from September 2020 to August 2021. The results were evaluated against the MODIS FPAR and in situ FPAR measurements. This article expects to promote the operational estimation of fine-resolution FPAR, enhance our understanding of the performance of statistical learning methods, and provide a reference for retrieving other land surface parameters using a similar approach.

II. STUDY AREA AND DATA SET

A. Study Area

The study area is the Guanzhong area in Shaanxi Province, China, covering a total area of 55 623 km² (see Fig. 1). The Guanzhong area features complex terrain and spans a wide range of elevations. The northern part belongs to the southern edge of the Loess Plateau, the southern part is the steep Qinling Mountains, and the Weihe River runs from west to the east in the central parts. The land cover types in the study area vary, including forests, croplands and built areas. Thus, the study area is representative for testing the performance of the two methods on a regional scale.

B. Data Set

- 1) *Landsat 8*: A total of 86 scenes of Landsat 8 OLI images with cloud coverage less than 10% from September 2019 to August 2021 were selected on the GEE as atmospherically corrected surface reflectance data in the UTM/WGS 84 projection (see Fig. 2). To implement the RFR method, Landsat data from September 2019 to August 2020 were used for training, and those from September 2020 to August 2021 were used for validation. The enhanced scaling method was only implemented from September 2020 to August 2021.
- 2) *MODIS LAI/FPAR Product*: The MODIS Version 6 LAI and FPAR product (MCD15A3H) at four-day intervals and 500 m spatial resolution [26] was used in this article. To match the selected Landsat images spatially and temporally, the MODIS FPAR images at the dates of available Landsat images were selected and resampled to 480 m resolution on the GEE.
- 3) *Land Cover Maps*: According to [19], the land cover maps used in the RFR method were supposed to be similar to the MODIS land cover product (MCD12Q1) used in MODIS FPAR estimation. For this reason, the finer resolution observation and monitoring of global land cover (FROM-GLC) product at 30-m resolution in 2015 [27] was used as input for the RFR method. To address the discrepancy in classification schema and biomes between the FROM-GLC and MCD12Q1 data, the land cover categories of the FROM-GLC data were analyzed and recoded carefully in comparison to the MCD12Q1 data using the method described in [19]. Finally, we defined eight biomes

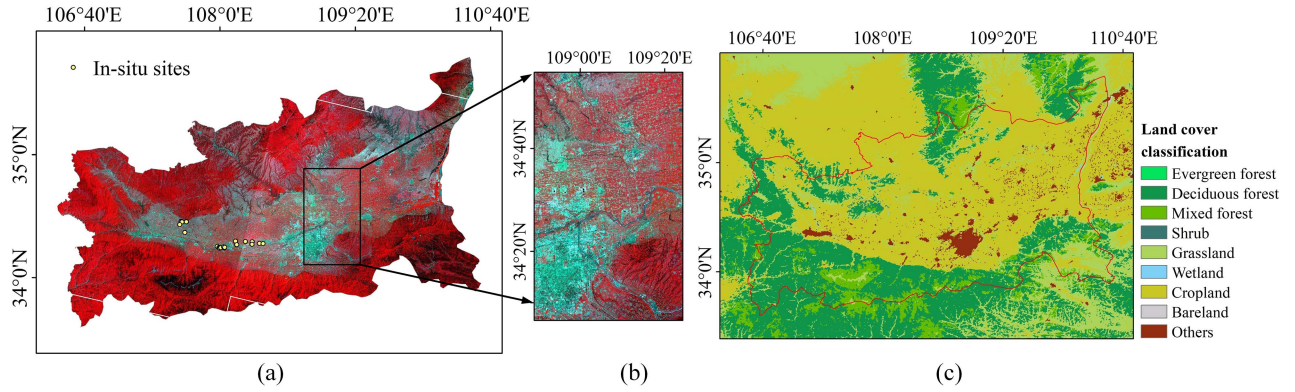


Fig. 1. Map of the study area. (a) Visible-NIR composite Landsat images. (b) Subarea with high spatial heterogeneity. (c) MCD12Q1 land cover map.

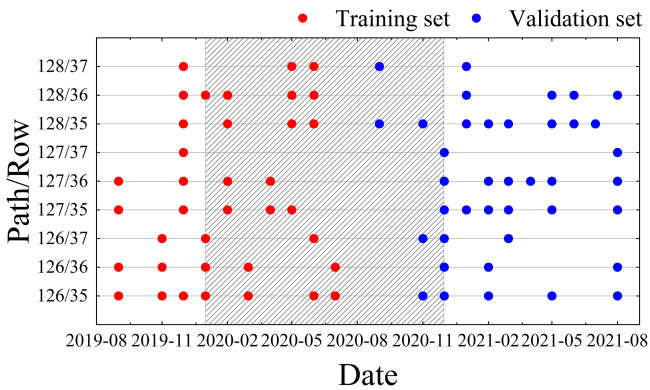


Fig. 2. Temporal distributions of the Landsat images used for training and validation in the RFR method.

TABLE I
MAPPING BETWEEN LAND COVER TYPES FROM FROM-GLC AND MODIS
FPAR BIOME CLASSIFICATION

Land cover	FROM-GLC code	MCD12Q1
Cropland	10	Cropland
Evergreen Forest	21,23	Evergreen Broadleaf Forest (EBF) Evergreen Needleleaf Forest (ENF)
Deciduous Forest	22,24	Deciduous Broadleaf Forest (DBF) Deciduous Needleleaf Forest (DBF)
Mixed Forest	25,26	Mixed Forest
Cropland	31,32,33	Cropland
Shrub	41,42	Shrub
Wetland	51,52,53	Wetland
Bareland	90	Bareland

for FROM-GLC land cover types to maintain consistency with the MODIS land cover map (see Table I).

- 4) *In situ Data*: Field campaigns were implemented in the study area to collect in situ FPAR data on April 16, April 17, and April 30, 2021. In situ sites were selected as homogeneous areas covering at least one Landsat pixel (i.e., 30 m \times 30 m). Within each site, sampling points were selected uniformly at 3–5 m intervals along the diagonal direction, and FPAR was measured sequentially.

The SunScan Canopy Analyzer [28], [29] was used to measure FPAR at each sampling point by measuring the incident and reflected PAR above and below the canopy layer [30] for crops (i.e., maize and wheat) in the study area. For high trees, the fraction of intercepted PAR was measured as the approximation of FPAR by measuring the incident PAR in nearby open areas and under the canopy. To match the satellite data in time and space, the measured FPAR was temporally normalized to the time of satellite overpass using the method in [30], and those within the same Landsat pixel were averaged to the 30-m resolution. Finally, a total of 63 in situ FPAR measurements were acquired to validate the Landsat FPAR retrievals.

III. METHODOLOGY

A. Enhanced Scaling Method

The core of the MLR-based scaling method [23] is to explicitly linearize the complex relations between MODIS FPAR and Landsat surface reflectance. The FPAR is modeled as a linear combination of multiband reflectance, which varies with land cover types and time, but is consistent across different spatial resolutions. As a result, a specific MLR model can be trained for each land cover type from each pair of MODIS FPAR and Landsat images as follows:

$$F_{c,t} = A_{c,t}^T P_{c,t} + \varepsilon \quad (1)$$

where $F_{c,t}$ is a $1 \times l$ matrix that includes MODIS FPAR values of l selected samples for land cover type c on date t , $P_{c,t}$ is a $(n_b+1) \times l$ vector matrix that includes the reflectance of n_b bands plus a constant term of l selected samples, $A_{c,t}$ is a $(n_b+1) \times 1$ vector matrix that represents the algorithm coefficients (including the constant term) for a specific linear retrieval model, and ε is the residual error between the prediction and true value.

The implementation of the scaling method includes four major steps: classifying the fine-resolution images automatically using the K-means algorithm; selecting the best-quality and homogeneous MODIS FPAR retrievals and the corresponding aggregated fine-resolution surface reflectance to compose the training samples; determining the coefficients of the linear model at 480 m resolution through the MLR for each land cover type on

each date; and estimating the fine-resolution FPAR using the established MLR model.

However, applying the existing scaling method to operational estimation of fine-resolution FPAR confronts a major problem of underrepresentation of densely vegetated pixels that have high FPAR values and relatively high heterogeneity due to growth differences. This problem leads to the underestimation of FPAR for densely vegetated pixels. Particularly when MODIS FPAR is very high (>0.8), the retrieved Landsat FPAR rarely exceeds 0.8. This underestimation problem can be attributed to a series of factors, including: using only the best-quality MODIS FPAR retrievals derived from the main algorithm under unsaturated conditions; the constant standard for selecting homogeneous MODIS pixels without considering seasonal variations; interference from invalid samples that have disproportionate MODIS FPAR and Landsat normalized difference vegetation index (NDVI) values [19]; and the saturation of Landsat reflectance for dense vegetation.

To address the underrepresentation problem, the enhanced scaling method made improvements in the following aspects:

First, the enhanced scaling method utilizes high-quality MODIS FPAR retrievals from the main algorithm under both saturated and unsaturated conditions, which have quality flags lower than 50 ($QC \leq 50$). This would include more saturated samples of dense vegetation.

Second, the enhanced scaling method proposes a self-adaptive standard for selecting homogeneous MODIS pixels. The mean band-average coefficient of variation (CV) of the surface reflectance is used as the threshold. A MODIS pixel with a CV value less than the mean CV is regarded as pure. Since spatial heterogeneity varies with vegetation growth, the mean CV represents the dominant spatial heterogeneity at a given time, which can help select adequate samples to include the growth variations within the same land cover type.

Third, the enhanced scaling method proposes a 3σ -based method to remove invalid samples, which have Landsat NDVI values falling outside the normal range of MODIS FPAR values. Specifically, the selected samples are sorted ascendingly by MODIS FPAR and divided into different groups by 0.02 intervals. For each group, the mean (μ) and standard deviation (σ) of Landsat NDVI are calculated. The samples having Landsat NDVI values outside the 95% confidence interval $[\mu - 2\sigma, \mu + 2\sigma]$ are removed as outliers. As a result, invalid samples with disproportionate Landsat NDVI and MODIS FPAR values, especially those with low MODIS FPAR but high Landsat NDVI values, are removed, allowing the underrepresentation of FPAR over densely vegetated pixels to be improved. This 3σ -based outlier removal method can remove exceptional values while maintaining the normal values. The method requires sufficient samples (i.e., >10) within each MODIS FPAR group. When the number of samples within each group is ≤ 10 , the neighboring groups can be combined or the 0.02 interval can be increased.

Fourth, saturation of Landsat surface reflectance is a major cause of the underrepresentation of the FPAR-reflectance relations in densely vegetated areas [19], [20]. This saturation is represented by the very small increase in Landsat NDVI with increasing vegetation density after Landsat NDVI > 0.8 . Due to

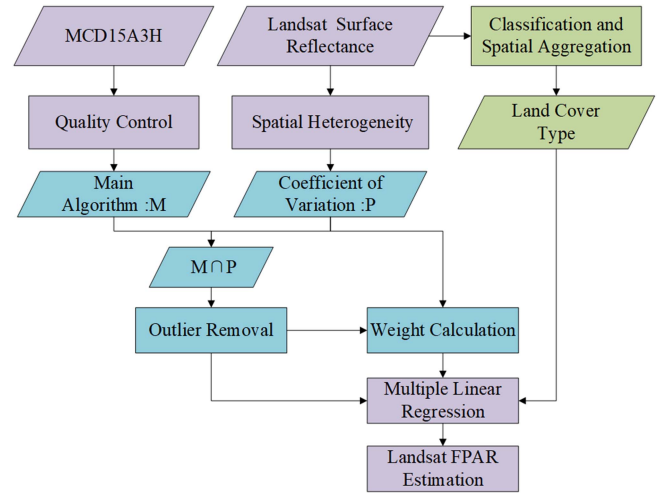


Fig. 3. Flow chart of the enhanced scaling method.

the saturation problem, the Landsat FPAR values estimated by the scaling method rarely exceed 0.9 [23].

To overcome this saturation problem, the enhanced scaling method proposes a weighting method to increase the weights of samples with very high MODIS FPAR values. The weights are calculated as follows:

$$W_t = \begin{cases} 1, & (\theta_t = 0) \\ 1 + (1 - \theta_t), & (\theta_t > 0) \end{cases} \quad (2)$$

where W_t represents the weight for each sample and θ_t is the ratio of samples with MODIS FPAR ≥ 0.9 occupying the samples with MODIS FPAR ≥ 0.8 . When $\theta_t = 0$ or $\theta_t = 1$, indicating that no samples have MODIS FPAR ≥ 0.9 or all samples with MODIS FPAR ≥ 0.8 are also ≥ 0.9 , the weights remain at 1. When $\theta_t > 0$, the value of W_t ranges from 1 to 2; the lower θ_t is, the greater the weight of the samples with MODIS FPAR ≥ 0.9 .

Based on the weighting procedure, the enhanced scaling method employs a weighted MLR approach to model the FPAR-reflectance relations.

The enhanced scaling method inherited the four major steps of the scaling method and is primarily modified in selecting, screening, and weighting samples. The input bands of the model include green, red, near-infrared (NIR), shortwave infrared 1 (SWIR1) and shortwave infrared 2, while the blue band was excluded for its low signal-to-noise ratio after atmospheric correction [32], [33], [34], [35]. The enhanced scaling method is self-adaptive and implemented as an automated process on GEE (sample codes are available at <https://code.earthengine.google.com/aec5d7e00b8b98d358b803ad0ed8fb2b>), as shown in Fig. 3.

B. Random Forest Regression

The RFR method was initially proposed by Breimans [36] as a nonlinear statistical ensemble method to prevent overfitting [37]. The RFR builds multiple unrelated decision trees through a random selection of features and bootstrap sampling from the

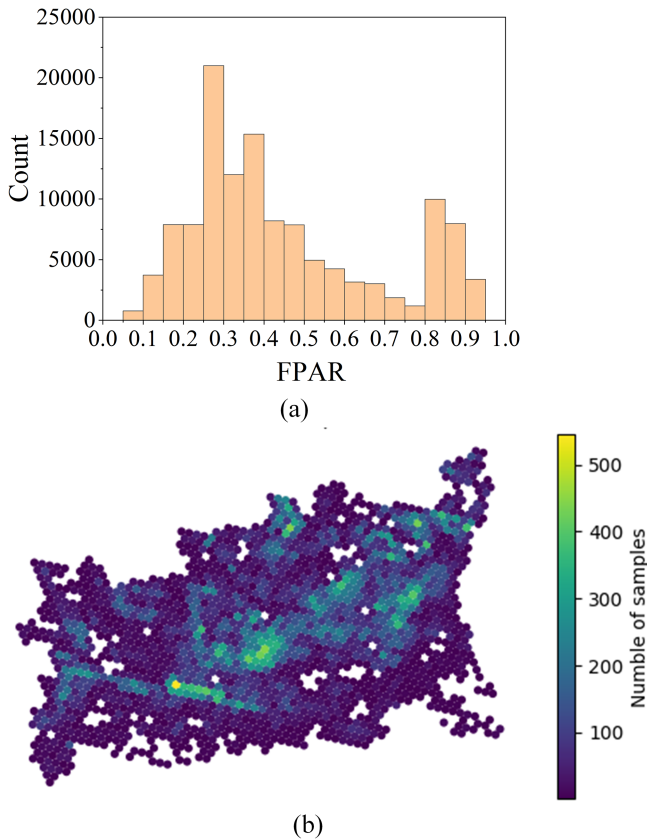


Fig. 4. (a) Frequency histogram and (b) spatial distribution of the screened samples for the RFR.

entire training set [20]. Each decision tree can yield a prediction, and the final output of an RFR is the mean of all predictions from all individual decision trees. The RFR method used for fine-resolution FPAR estimation was inherited from the method used for fine-resolution LAI estimation in [19] (sample codes are available at: <https://github.com/yanguhuikang/Landsat-LAI>) and implemented on the GEE because it is more generalized for regional applications than that designed for cropland in [20].

To ensure the representativeness of the samples, the RFR in [19] designed a complete workflow of the sample generation process, which was also used in this article and involved five major steps as follows.

- 1) Spatially representative MODIS pixels were selected as those within which more than 90% of the 30-m FROM-GLC pixels belonged to a single land cover type, hereafter referred to as “spatial samples.”
- 2) For each spatial sample, high-quality FPAR retrievals for homogeneous MODIS pixels were screened using the criteria of a quality flag below 50 ($QC \leq 50$) and a band-average CV value below 0.15.
- 3) Within the homogeneous samples, outliers were removed as the samples whose NDVI fell outside the 1.5 interquartile range (IQR) from the first and third quartile for each FPAR interval.
- 4) All the selected samples were combined, and the sample balance among different FPAR values was controlled. We calculated the sample density in space and down-sampled

densely populated areas to make the sample distribution more uniform.

- 5) The samples were further screened to remove confounding unsaturated and saturated samples that share similar spectral signatures from Landsat surface reflectance but contrasting FPAR values, hereafter referred to as “saturation screening process.” A random classifier was trained to identify saturated and unsaturated samples based on Landsat surface reflectance in green, red, NIR, SWIR bands, NDVI, normalized difference water index (NDWI), and solar illumination angles. Separate models were built for different land cover types. Based on ten-fold cross-validation, the misclassified samples were removed from the samples.

Finally, a total of 124 509 samples were finally selected, which were distributed evenly across the entire study area and across the range of MODIS FPAR values, as shown in Fig. 4.

Considering the differences in the FPAR-reflectance relations between different land cover types, we built random forest (RF) models for each land cover type. Model inputs include surface reflectance in the green, red, NIR, and SWIR1 bands, NDVI, NDWI [38], geographic coordinates (longitude, latitude) of the pixel center, and solar zenith and azimuth of the scene center. In the RFR training, the number of regression trees was set to 100, the minimum leaf population was 50, and the number of variables for each segmentation was 5.

C. Validation and Comparison

Both the MLR-based enhanced scaling method and the RFR method were implemented on the GEE for estimating Landsat FPAR from September 2020 to August 2021. The results were validated using in situ FPAR measurements and compared with MODIS FPAR products. The root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were used to evaluate the accuracy with in situ data and the consistency with MODIS FPAR data [20], [23], [39]. The performance of the two methods was also evaluated for homogeneous and heterogeneous pixels.

Moreover, the sensitivity of both methods to the number of training samples was investigated. For the RFR method, varying numbers of images were inputted for training, and the Landsat FPAR images were estimated and evaluated for the entire year from September 2020 to August 2021. As the enhanced scaling method only requires one single pair of Landsat-MODIS images, the Landsat image with p127/r036 on May 30, 2021, was used, and the performance of the method was investigated by using a varying number of samples, which were randomly selected from the entire training set, for the MLR process.

IV. RESULTS

A. Consistency With MODIS FPAR

Fig. 5 shows the mosaiced Landsat FPAR images of the Guanzhong area in different months in 2021. The estimated Landsat FPAR of both the RFR and enhanced scaling method were generally consistent with the MODIS FPAR. Spatial details

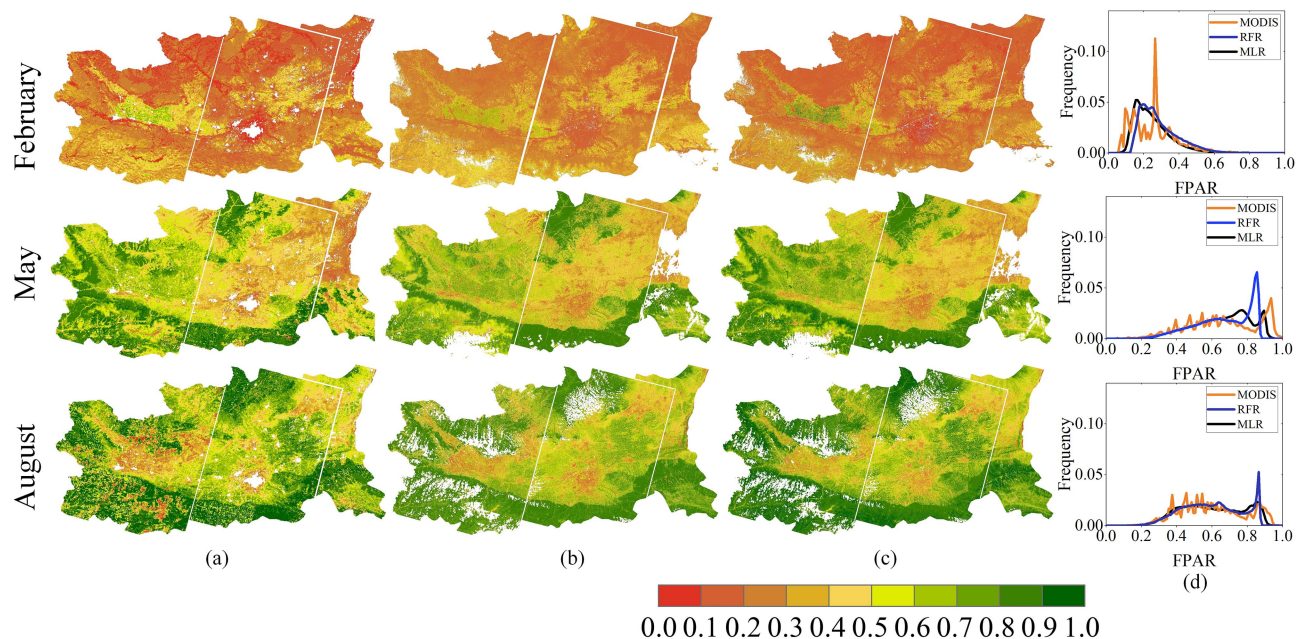


Fig. 5. Comparison of (a) MODIS FPAR images and Landsat FPAR images estimated by the (b) RFR and (c) MLR-based enhanced scaling methods in the Guanzhong area and (d) the corresponding histograms in different months in 2021.

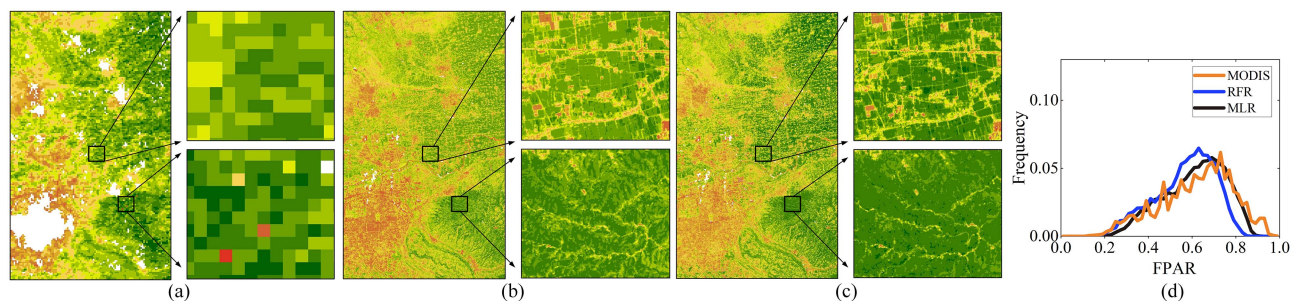


Fig. 6. Comparison of (a) MODIS FPAR and the Landsat FPAR estimated by the (b) RFR and (c) enhanced scaling methods in the subarea on August 2, 2021, and (d) the corresponding histograms.

and histograms reveal that the RFR underestimated high FPAR values in spring (May) and summer (August) and overestimated low FPAR values in winter (February), while the results of the enhanced scaling method were closer to the MODIS FPAR in spatial details and absolute values.

Fig. 6 compares Landsat FPAR retrievals using the two methods with the MODIS FPAR image in the test area on August 2, 2021. Investigation of spatial details over the subareas of cropland and woodland clearly shows that the enhanced scaling method accurately restored the spatial patterns of MODIS FPAR, while the RFR underestimated high FPAR values. The histograms also show that the results of the MLR-based enhanced scaling method were highly consistent with the MODIS FPAR, while the RFR underestimated high FPAR values. Entirely built upon the selected samples, the RFR may experience major problems in densely vegetated heterogeneous areas, where fewer samples could be selected and averaging over a large number of samples may underrepresent pixels with too little or too much vegetation.

Since both the RFR and enhanced scaling method rely on pure MODIS pixels to model the relations between FPAR and reflectance [22], [40], the estimated Landsat FPAR of both methods for the entire year from September 2020 to August 2021 were compared with MODIS FPAR in homogeneous MODIS pixels and heterogeneous MODIS pixels, as shown in Fig. 7. For both methods, the density contours of homogeneous pixels were generally more concentrated around the 1:1 line than those of heterogeneous pixels. The contour lines of the enhanced scaling method bulged out in the direction of the 1:1 line very well and were more concentrated than those of the RFR for all pixels. The contour lines show that the RFR overestimated low FPAR values (i.e., <0.5) and underestimated high FPAR values (i.e., >0.8) for all pixels. This indicates that the RFR model built on historical data may not adapt to new data. In Fig. 7(a) and (b), a few homogeneous pixels with MODIS FPAR values ranging from 0.4 to 0.7 derived high Landsat FPAR values of approximately 0.8. These pixels were removed from sample screening as outliers but were kept in the validation. Compared with the MODIS

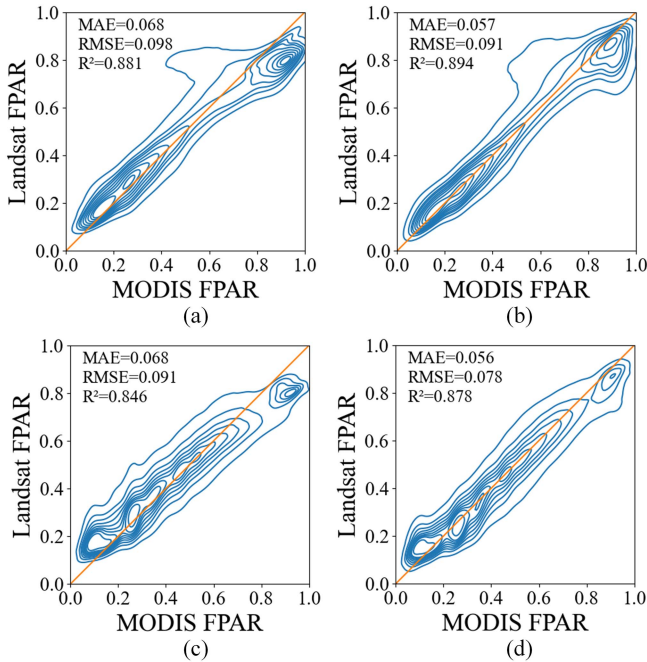


Fig. 7. 2D density contour of MODIS FPAR versus Landsat FPAR retrievals from the (a and c) RFR and (b and d) enhanced scaling methods for (a and b) homogeneous and (c and d) heterogeneous pixels.

FPAR products, the Landsat FPAR retrievals of the enhanced scaling method (overall RMSE = 0.076, MAE = 0.052, and $R^2 = 0.887$) were more accurate than those of the RFR method (overall RMSE = 0.091, MAE = 0.068, and $R^2 = 0.819$).

Fig. 8 shows the retrieved Landsat FPAR values and the MODIS FPAR values as a function of day of year (DOY) from September 2020 to August 2021. In general, the retrieved Landsat FPAR values from the enhanced scaling method were more consistent with the MODIS FPAR values than those of the RFR in terms of the mean values, variations and RMSE for all three vegetation types. The RFR overestimated the FPAR values during the pre-growth and senescence stages for all three vegetation types. During peak growth stages, the RFR underestimated the FPAR values for the woodlands but overestimated the FPAR values for the grasslands and croplands. Such bias can be attributed to the adaptability of the RFR model built on training samples to the new data. In contrast, the MLR-based enhanced scaling method performed best over croplands, slightly underestimated the FPAR values for the woodlands and overestimated the FPAR values for the grasslands during peak growth stages. Such bias might be related to the high spatial heterogeneity in the grasslands during the peak season, as shown by the large standard deviations in the MODIS FPAR and Landsat FPAR retrievals in Fig. 8.

B. Validation With in Situ FPAR

The scatter plots in Fig. 9 reveal that the FPAR values estimated by both the RFR and the enhanced scaling method were generally consistent with the in situ measurements. This can be attributed to the high accuracy of the MODIS FPAR data over croplands, as validated in previous studies [30], [41],

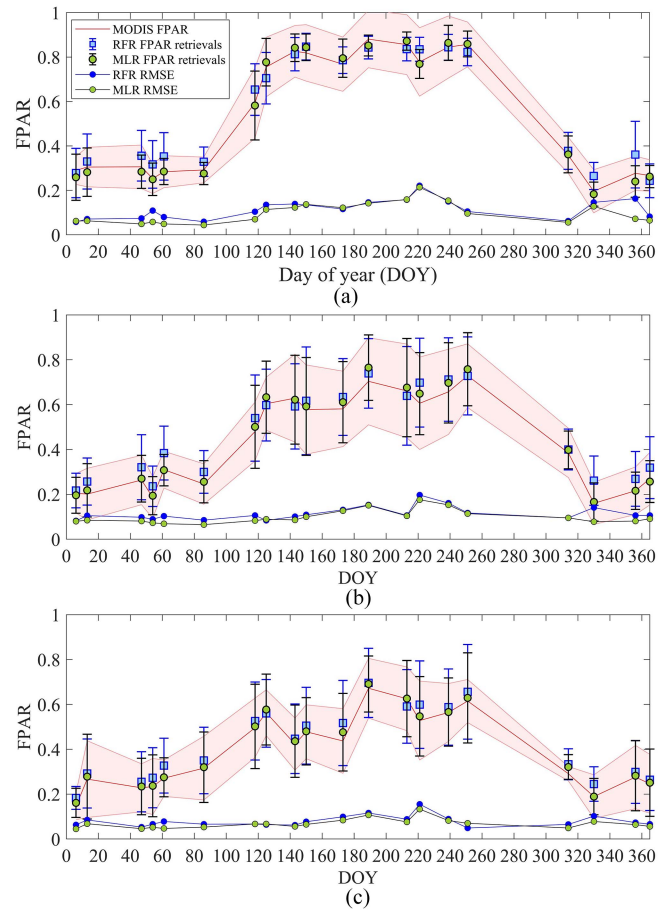


Fig. 8. Spatiotemporal patterns of Landsat FPAR retrievals using the RFR and enhanced scaling methods in the Guanzhong area from September 2020 to August 2021 compared with the MODIS FPAR values as a function of DOY for (a) woodland, (b) grassland, and (c) cropland.

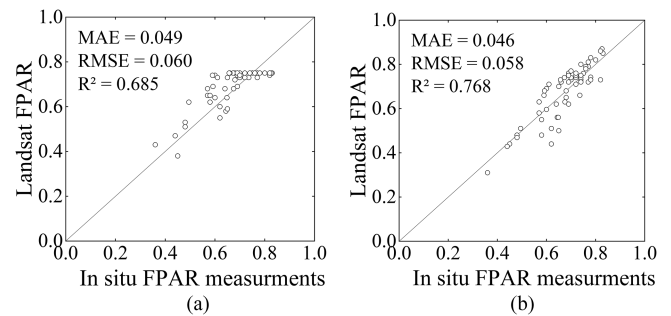


Fig. 9. Comparison of the in situ FPAR measurements and Landsat FPAR retrievals from the (a) RFR and (b) enhanced scaling methods.

[42]. The FPAR estimated by the RFR method saturated at 0.75 when the in situ FPAR exceeded 0.7, while the enhanced scaling method accurately estimated the FPAR values, which were very close to the 1:1 line from low to high values. This indicates that the MLR-based enhanced scaling method can estimate fine-resolution FPAR values more accurately.

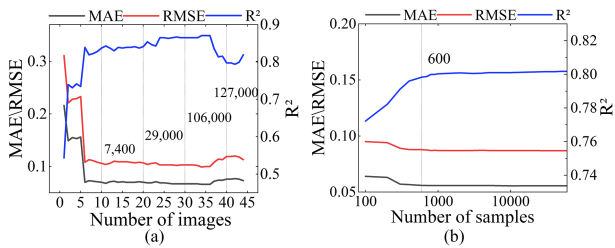


Fig. 10. Sensitivity of the (a) RFR and (b) enhanced scaling methods to the number of training images/samples.

C. Sensitivity to Training Samples

As a data-driven method, RFR relies on a large number of samples to model the FPAR-reflectance relations. The impacts of the number of training samples on the RFR were thus investigated by varying the number of input images for training. The trained RFR models were applied to the study area for the entire year from September 2020 to August 2021. Fig. 10(a) shows the change in the accuracy of the RFR with varying numbers of input images for training. A rapid increase in the accuracy was observed when the number of input images increased from 1 to 7, while the number of screened samples increased from 850 to 6000. The accuracy was then stabilized when the number of input images increased from 10 to 20, corresponding to at least 7400 training samples. For 20 to 35 input images, the accuracy remained at a relatively high level, corresponding to at least 29 000 training samples. However, when the number of input images was 35, a continual increase in the number of input images reduced the accuracy. This is because adding different training samples can cause different effects. Adding a peak-season image will add more saturated samples, which can improve the accuracy of saturated pixels but reduce the accuracy of unsaturated pixels. Consequently, the final effects of adding a peak-season image depend on the relative proportions of saturated and unsaturated pixels in the images to be predicted. Therefore, when the number of input images was 35, two peak-season images were added, and the decrease in the accuracy of unsaturated pixels was greater than the increase in the accuracy of saturated pixels, resulting in an overall decrease in the accuracy. This indicates that the RFR is very sensitive to training samples.

Fig. 10(b) shows the change in the accuracy of the MLR-based enhanced scaling method with varying numbers of samples. When the number of samples increased from 100 to 600, the RMSE decreased from 0.095 to 0.088, and the R^2 , increased from 0.772 to 0.798. When the number of samples was > 600 , the accuracy of the MLR was very stable. This indicates that MLR requires fewer samples and is less sensitive to the training samples.

V. DISCUSSION

A. Sensitivity to Land Cover Types

Both the RFR and MLR methods require land cover maps to establish land cover-specific models for estimating fine-resolution FPAR. Relevant research shows that the uncertainty of land cover maps will affect the retrieval accuracy [43], [44],

[45], [46]. Thus, we investigated the sensitivity of both methods to land cover maps by using different land cover maps. For both methods, the same Landsat image with p127/r036 on May 30, 2021, was used and classified into five land cover types using three different methods, including the unsupervised K-means clustering algorithm, classification and regression tree (CART) algorithm and RF algorithm. For the CART and RF supervised algorithms, two sets of training samples were used: manually selected training samples and randomly generated samples. Therefore, including the FROM-GLC land cover map, six land cover maps were available for both the RFR and MLR methods.

For the RFR method, the FROM-GLC land cover map was used as the benchmark land cover map. Using the benchmark land cover map as ground truth, the accuracy of the other five land cover maps ranged from 53% to 82%, as given in Table II. To maintain the same study area as the enhanced scaling method, the training samples for the RFR on the Landsat image with p127/r036 were selected from the entire set of training samples. Each land cover map was used to update the land cover types of the selected samples.

For the enhanced scaling method, the land cover map previously derived using the unsupervised K-means clustering algorithm was used as the benchmark land cover map. The accuracy of the other five land cover maps ranged from 53% to 91%, as given in Table II.

For both methods, each land cover map was used as input to generate a new FPAR image. The FPAR image corresponding to the benchmark land cover map was used as the benchmark FPAR image to evaluate the other FPAR images. The statistical metrics, including RMSE, MAE, and R^2 , were computed for the respective pairs of FPAR images estimated using the different land cover maps. The magnitude of the difference in FPAR (∂FPAR) between the new and benchmark FPAR images was examined.

Table II compares the FPAR images using the different land cover maps. The table data show that large discrepancies existed between different land cover maps, as the overall Kappa coefficients between the respective pairs of land cover maps varied from 0.43 to 0.87. Generally, the differences in FPAR were small for both the RFR and enhanced scaling methods, with RMSE ranging from 0.034 to 0.037 and from 0.025 to 0.038, respectively. The R^2 values exceeded 0.966 for all cases, indicating strong correlations between the different FPAR images using different land cover maps. The pixels with $\partial\text{FPAR} < 0.05$ occupied more than 89.98% in the results of the RFR and 91.04% in the results of the enhanced scaling method. All these statistics reveal that the different land cover maps would result in limited differences in FPAR retrievals for both the RFR and enhanced scaling methods. Moreover, the differences in the FPAR images are positively related to the consistencies between the respective pairs of land cover maps. For example, in cases 4, 8, and 9, the land cover maps with an overall accuracy exceeding 80% resulted in very low RMSE values between the respective pairs of FPAR images.

The results of the enhanced scaling method were generally more accurate than those of RFR, while case 10 using the FROM-GLC land cover map in the enhanced scaling method is an exception that derived the lowest accuracy. This indicates

TABLE II
COMPARISON BETWEEN FPAR USING THE DIFFERENT LAND COVER MAPS

Information	Random forest regression					Enhanced scaling method				
	Case1	Case2	Case3	Case4	Case5	Case6	Case7	Case8	Case9	Case10
Land cover map	Compared to the benchmark land cover map (FROM-GLC)					Compared to the benchmark land cover map (K-means)				
Training samples	Manually selected		Randomly selected			Manually selected		Randomly selected		
Classification	Cart	RF	Cart	RF	K-means	Cart	RF	Cart	RF	FROM-GLC
Overall accuracy	57%	61%	75%	82%	53%	59%	64%	88%	91%	53%
Kappa Coefficient	0.45	0.48	0.66	0.79	0.43	0.47	0.54	0.84	0.87	0.43
FPAR images compared to the benchmark FPAR image										
MAE	0.021	0.021	0.018	0.017	0.022	0.015	0.014	0.012	0.011	0.023
RMSE	0.037	0.035	0.034	0.035	0.037	0.029	0.029	0.025	0.028	0.038
R ²	0.972	0.986	0.976	0.972	0.970	0.980	0.980	0.984	0.980	0.966
Area ratio(%)										
$\hat{\sigma}$ FPAR=0	30.56	31.07	38.26	39.11	24.61	41.32	44.42	45.65	54.23	21.98
$0 < \text{FPAR} \leq 0.05$	59.99	58.91	52.57	51.32	65.61	51.60	48.25	49.21	40.29	69.06
$0.05 < \text{FPAR} \leq 0.1$	5.39	6.56	5.72	5.23	5.77	5.59	5.54	4.11	3.82	6.29
$\hat{\sigma}$ FPAR >0.1	4.05	3.46	3.45	4.34	4.01	1.50	1.79	1.04	1.66	2.67

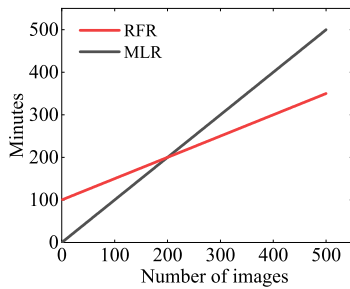


Fig. 11. Computation time varying with the number of Landsat images to be estimated.

that the use of a different land cover map from a different date can significantly reduce the accuracy of the enhanced scaling method.

B. Computation Efficiency

Both the RFR and enhanced scaling methods were automated and streamlined on the GEE, and their computation efficiency was further investigated. Fig. 11 compares the computation time for predicting varying numbers of images using the RFR and enhanced scaling methods. For the RFR method, the input images for training were kept the same as in Section III-B, and the time for predicting different numbers of images was recorded. The time used by the RFR includes the time for screening samples, training the model, and estimating the FPAR for new images. As shown in Fig. 11, the RFR used 100 min for screening samples, while it used a half minute for estimating the FPAR for one Landsat image. For the MLR-based enhanced scaling method, the processes of sample screening, MLR training, and FPAR estimation are completed on one single pair of MODIS-Landsat images, which in all took 1 min. Thus, the time used by the RFR and enhanced scaling method was equal when predicting approximately 200 images. This offers us insights into choosing an optimal method. MLR-based enhanced scaling is highly efficient for predicting fewer images, while RFR is optimal for predicting large numbers of images. However, although both methods were implemented on the GEE, the enhanced scaling method is more automated than the RFR by excluding all human interference.

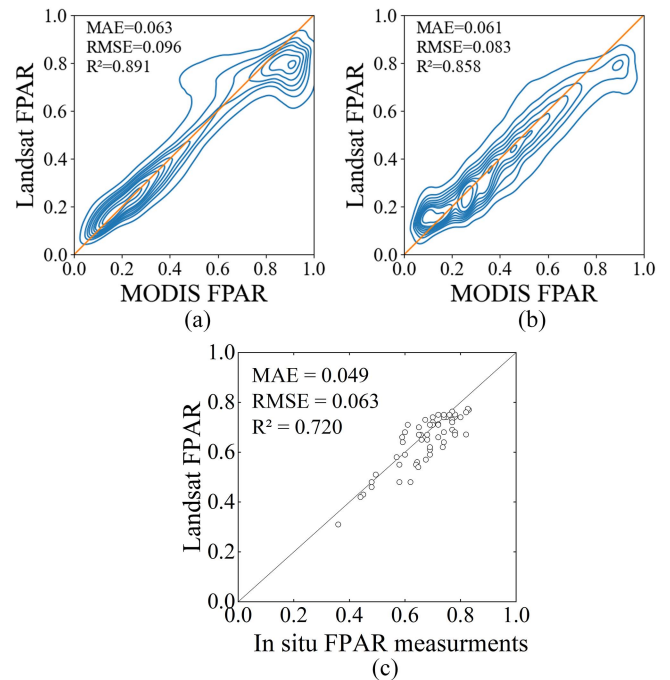


Fig. 12. Validation of the Landsat FPAR retrievals using the existing scaling method. Comparison with the MODIS FPAR for (a) homogeneous pixels and (b) heterogeneous pixels. (c) Comparison with the in situ FPAR measurements.

In contrast, the sample balancing in the RFR requires an overall analysis of the screened samples to decide whether and how the samples are balanced. The time of this process was not included in the comparisons above.

C. Improvement Over the Scaling Method

To validate the improvements over the existing scaling method, the results of the existing scaling method were also compared with the MODIS FPAR for the entire validation period and with the in situ FPAR measurements, as shown in Fig. 12. It clearly shows that the existing scaling method significantly underestimated FPAR for densely vegetated pixels with $\text{FPAR} > 0.75$ compared with MODIS FPAR and in situ FPAR measurements. The Landsat FPAR retrievals from the

existing scaling method were less consistent with the MODIS FPAR and less accurate than the enhanced scaling method.

D. Advantages and Disadvantages of Both Methods

Both the RFR and enhanced scaling methods have been successfully applied to FPAR estimation on the GEE, while both methods have their advantages and disadvantages. Both the RFR and enhanced scaling methods can inherit the high accuracy and wide applicability of the main algorithm of the MODIS FPAR product and are easily automated on the cloud computing platform.

To increase the representativeness of the FPAR-reflectance relations in the MODIS FPAR products, both the RFR and MLR methods used high-quality FPAR retrievals over homogeneous MODIS pixels and involved an outlier removal process. The difference is that the sample screening in the RFR is more rigorous and complex. First, the RFR adopted a higher standard (e.g., a lower CV threshold) for homogeneous MODIS pixels because the nonlinear RFR algorithm built over heterogeneous MODIS pixels cannot be applied to Landsat pixels due to scale effects [21], [22]. Second, the proposed 3σ -based outlier removal method in the enhanced scaling method is less rigorous. In our experiments, the 3σ -based method removed fewer outliers than the IQR method in the RFR method. In statistical learning models, a few outliers within the samples can enhance the generalization ability of the model [31]. Third, the RFR involved two additional processes of sample balancing and saturation screening because an unbalanced sample set and spectral similarities between saturated and unsaturated samples would undermine the skill of the RFR [19], [47]. In contrast, as a linear algorithm, the enhanced scaling method used self-adaptive CV thresholds to define homogeneous MODIS pixels, which can capture the major within-class spatial variations and seasonal variations. A weighted MLR procedure was proposed to increase the representativeness of densely vegetated samples.

To analyze the effect of sample screening on the RFR, we took a subarea covering the Landsat scene p127/r036 and used the sample screened by the enhanced scaling method to train the RFR. A total of 323362 training samples were screened from the 5 Landsat images from 2019 to 2020 using the samples screened by the enhanced scaling method, which is much more than the 124 509 samples screened by the RFR method. The trained RFR models were used to estimate Landsat FPAR images from 2020 to 2021, and the results were shown in Fig. 13. Compared to the results of the previously trained RFR, using the samples screened by the enhanced scaling method slightly improved the values of RMSE and R^2 for the RFR, while the overestimation of low FPAR values and the underestimation of high FPAR values were even more significant. This indicates that the sample screening strategy in the enhanced scaling method does not improve the performance of the RFR, which may be attributed to the following reasons.

First, sample balancing and saturation screening are important for the RFR to achieve optimal accuracy [19], but are not included in the enhanced scaling method. The sample set screened by the enhanced scaling method generally has a much larger size and a higher proportion of unsaturated samples (i.e., 50% and

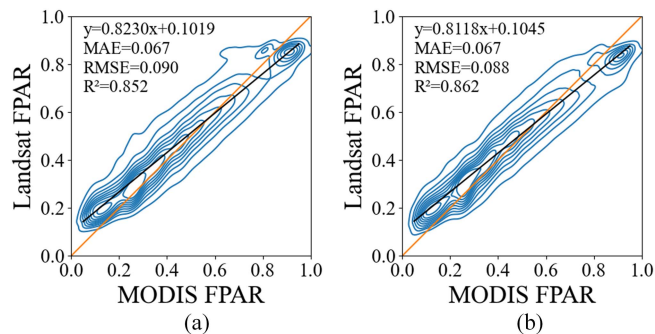


Fig. 13. 2D density contour of MODIS FPAR versus Landsat FPAR retrievals from the RFR using samples screened (a) by the RFR and (b) by the enhanced scaling method for estimating Landsat FPAR images (p127/r036) from 2020 to 2021.

65% for the RFR and enhanced scaling method for the woodland, respectively), which improves the overall prediction accuracy but does not improve the underestimation of high FPAR values.

Second, the definition of homogeneous MODIS pixels is more rigorous in the RFR than in the enhanced scaling method. Consequently, many samples screened by the enhanced scaling method have larger subpixel spatial variations, particularly for densely vegetated areas, where scale effects possibly occur [21], [22] and undermine the performance of the RFR.

Third, the algorithm design might be a major cause for the prediction bias in the RFR method. The RFR averages the predictions of all candidate decision trees as the final output, which tends to produce a moderate value instead of a very high or a very low value [48], [49]. Previous studies found similar biases for low and high values in downscaling LST satellite products and attempted to correct the residuals using the coarse-resolution product [50], [51]. However, such correction is performed at coarse resolution, thereby causing a block effect in the fine-resolution images, and was not considered in this article.

Therefore, the enhanced scaling method is even simpler, more accurate, and less sensitive to the training samples than the RFR method based on better representation of FPAR-reflectance relations in densely vegetated areas. The enhanced scaling method is more ready-to-use for requiring only one pair of MODIS FPAR and Landsat images, yet the FPAR-reflectance relations trained from one image pair can only apply to the image pair observed at similar SZAs. In contrast, the RFR requires many input images to cover various dates, locations, SZAs and land cover types to model the complex FPAR-reflectance relations. The major strength of the RFR method is that once built, it can quickly estimate fine-resolution FPAR from satellite images, while the major limitations include prediction biases and higher requirements for training samples. In practical applications when MODIS FPAR values are poorly retrieved on some dates, the RFR can still work, while enhanced scaling can use the MODIS FPAR data at adjacent dates in the same year or similar dates in another year as alternatives to establish the FPAR-reflectance relations.

Therefore, for a specific application, the choice of an optimal method depends on a series of factors. The enhanced scaling method is more ready-to-use with fewer requirements on input data, higher accuracy, and a higher degree of automation, while

the RFR method is suitable for continuous estimation of FPAR for hundreds of images in the same study area.

VI. CONCLUSION

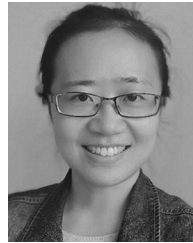
In this article, we developed an enhanced scaling method that integrates an outlier removal procedure and a method for weighting the samples in the MLR for operational fine-resolution FPAR estimation using the MODIS FPAR product as a reference. The proposed enhanced scaling method based on weighted MLR was compared to the RFR method on the GEE for regional application in the large Guanzhong area for predicting Landsat FPAR images for one entire year from September 2020 to August 2021. The main conclusions can be drawn as follows.

- 1) Both the enhanced scaling method and RFR method generate MODIS-consistent and spatially consistent fine-resolution FPAR. The RFR method overestimated low FPAR values and underestimated high FPAR values, while the enhanced scaling method outperformed the RFR method. In particular, the underestimation of high FPAR values for densely vegetated pixels was significantly improved in the enhanced scaling method.
- 2) Both the enhanced scaling method and RFR method were validated with in situ FPAR measurements. The enhanced scaling method performed better than the RFR method, particularly for densely vegetated pixels. Overall RMSE values of 0.049 and 0.046 and overall R^2 values of 0.685 and 0.768 were attained for the RFR and enhanced scaling methods, respectively.
- 3) The enhanced scaling method is less sensitive to the training samples, while the RFR method has a higher degree of reliance on the training samples. The enhanced scaling method is easier to implement with a minimum requirement of one pair of MODIS-Landsat images and is more accurate than the RFR method over both homogeneous and heterogeneous pixels.

REFERENCES

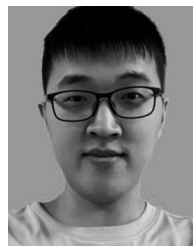
- [1] M. Weiss, F. Baret, S. Garrigues, and R. Lacaze, "LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: Validation and comparison with MODIS collection 4 products," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 317–331, 2007.
- [2] W. Fan, Y. Liu, X. Xu, G. Chen, and B. Zhang, "A new fAPAR analytical model based on the law of energy conservation: A case study in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3945–3955, Sep. 2014, doi: [10.1109/JSTARS.2014.2325673](https://doi.org/10.1109/JSTARS.2014.2325673).
- [3] X. Tao, S. Liang, T. He, and H. Jin, "Estimation of fraction of absorbed photosynthetically active radiation from multiple satellite data: Model development and validation," *Remote Sens. Environ.*, vol. 184, pp. 539–557, 2016.
- [4] A. Loew et al., "Do we (need to) care about canopy radiation schemes in DGVMs? Caveats and potential impacts," *Biogeosciences*, vol. 11, no. 7, pp. 1873–1897, 2014.
- [5] D. Tian, W. Fan, and H. Ren, "Progress of fraction of absorbed photosynthetically active radiation retrieval from remote sensing data," *J. Remote Sens.*, vol. 24, no. 11, pp. 1307–1324, 2020.
- [6] F. Castaldi, R. Casa, F. Pelosi, and H. Yang, "Influence of acquisition time and resolution on wheat yield estimation at the field scale from canopy biophysical variables retrieved from SPOT satellite data," *Int. J. Remote Sens.*, vol. 36, no. 9, pp. 2438–2459, 2015.
- [7] R. B. Myneni et al., "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data," *Remote Sens. Environ.*, vol. 83, no. 1–2, pp. 214–231, 2002.
- [8] Y. Knyazikhin, J. Martonchik, R. B. Myneni, D. Diner, and S. W. Running, "Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data," *J. Geophys. Res., Atmos.*, vol. 103, no. D24, pp. 32257–32275, 1998.
- [9] Z. Xiao, S. Liang, R. Sun, J. Wang, and B. Jiang, "Estimating the fraction of absorbed photosynthetically active radiation from the MODIS data based GLASS leaf area index product," *Remote Sens. Environ.*, vol. 171, pp. 105–117, 2015.
- [10] Y. Wang, G. Yan, D. Xie, R. Hu, and H. Zhang, "Generating long time series of high spatiotemporal resolution FPAR images in the remote sensing trend surface framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2021, Art. no. 4401915, doi: [10.1109/TGRS.2021.3067913](https://doi.org/10.1109/TGRS.2021.3067913).
- [11] Y. Wang, D. Xie, Y. Zhan, H. Li, G. Yan, and Y. Chen, "Assessing the accuracy of landsat-MODIS NDVI fusion with limited input data: A strategy for base data selection," *Remote Sens.*, vol. 13, no. 2, pp. 266, 2021.
- [12] J. M. Chen, "Canopy architecture and remote sensing of the fraction of photosynthetically active radiation absorbed by boreal conifer forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 6, pp. 1353–1368, Nov. 1996, doi: [10.1109/36.544559](https://doi.org/10.1109/36.544559).
- [13] T. Majasalmi, M. Rautiainen, and P. Stenberg, "Modeled and measured fPAR in a boreal forest: Validation and application of a new model," *Agricultural Forest Meteorol.*, vol. 189, pp. 118–124, 2014.
- [14] R. Myneni and D. Williams, "On the relationship between FAPAR and NDVI," *Remote Sens. Environ.*, vol. 49, no. 3, pp. 200–211, 1994.
- [15] R. Liu, H. Ren, S. Liu, Q. Liu, B. Yan, and F. Gan, "Generalized FPAR estimation methods from various satellite sensors and validation," *Agricultural Forest Meteorol.*, vol. 260, pp. 55–72, 2018.
- [16] K. Huemmrich and S. Goward, "Vegetation canopy PAR absorptance and NDVI: An assessment for ten tree species with the SAIL model," *Remote Sens. Environ.*, vol. 61, no. 2, pp. 254–269, 1997.
- [17] R. Fensholt, I. Sandholt, and M. S. Rasmussen, "Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements," *Remote Sens. Environ.*, vol. 91, no. 3–4, pp. 490–507, 2004.
- [18] F. Gao, M. C. Anderson, W. P. Kustas, and Y. Wang, "Simple method for retrieving leaf area index from Landsat using MODIS leaf area index products as reference," *J. Appl. Remote Sens.*, vol. 6, no. 1, 2012, Art. no. 063554.
- [19] Y. Kang et al., "A data-driven approach to estimate leaf area index for Landsat images over the contiguous US," *Remote Sens. Environ.*, vol. 258, 2021, Art. no. 112383.
- [20] Y. Sun, Q. Qin, H. Ren, and Y. Zhang, "Decameter cropland LAI/fPAR estimation from sentinel-2 imagery using google earth engine," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2021, Art. no. 4400614, doi: [10.1109/TGRS.2021.3052254](https://doi.org/10.1109/TGRS.2021.3052254).
- [21] Y. Wang, D. Xie, and X. Li, "Universal scaling methodology in remote sensing science by constructing geographic trend surface," *J. Remote Sens.*, vol. 18, no. 6, pp. 1139–1146, 2014.
- [22] H. Wu and Z.-L. Li, "Scale issues in remote sensing: A review on analysis, processing and modeling," *Sensors*, vol. 9, no. 3, pp. 1768–1793, 2009.
- [23] Y. Wang, G. Yan, R. Hu, D. Xie, and W. Chen, "A scaling-based method for the rapid retrieval of FPAR from fine-resolution satellite data in the remote-sensing trend-surface framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7035–7048, Oct. 2020, doi: [10.1109/TGRS.2020.2978884](https://doi.org/10.1109/TGRS.2020.2978884).
- [24] C. Hutengs and M. Vohland, "Downscaling land surface temperatures at regional scales with random forest regression," *Remote Sens. Environ.*, vol. 178, pp. 127–141, 2016.
- [25] H. Ebrahimi and M. Azadbakht, "Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels," *Comput. Geosci.*, vol. 124, pp. 93–102, 2019.
- [26] K. Yan et al., "Evaluation of MODIS LAI/fPAR product collection 6. Part 2: Validation and intercomparison," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 460.
- [27] P. Gong et al., "Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.
- [28] R. Mourad, H. Jaafar, M. Anderson, and F. Gao, "Assessment of leaf area index models using harmonized landsat and sentinel-2 surface reflectance data over a semi-arid irrigated landscape," *Remote Sens.*, vol. 12, no. 19, 2020, Art. no. 3121.
- [29] G. Yan et al., "Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives," *Agricultural Forest Meteorol.*, vol. 265, pp. 390–411, 2019.

- [30] Y. Wang, D. Xie, S. Liu, R. Hu, Y. Li, and G. Yan, "Scaling of FAPAR from the field to the satellite," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 310.
- [31] J. W. Osborne and A. Overbay, "The power of outliers (and why researchers should always check for them)," *Practical Assessment, Res., Eval.*, vol. 9, no. 1, 2004, Art. no. 6.
- [32] T. Hwang, C. Song, P. V. Bolstad, and L. E. Band, "Downscaling real-time vegetation dynamics by fusing multi-temporal MODIS and Landsat NDVI in topographically complex terrain," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2499–2512, 2011.
- [33] J. G. Masek et al., "A Landsat surface reflectance dataset for North America, 1990–2000," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 68–72, Jan. 2006.
- [34] H. K. Zhang et al., "Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences," *Remote Sens. Environ.*, vol. 215, pp. 482–494, 2018.
- [35] J. Ju, D. P. Roy, E. Vermote, J. Masek, and V. Kovalsky, "Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods," *Remote Sens. Environ.*, vol. 122, pp. 175–184, 2012.
- [36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [37] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [38] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 257–266, 1996.
- [39] X. Tao, S. Liang, and D. Wang, "Assessment of five global satellite products of fraction of absorbed photosynthetically active radiation: Intercomparison and direct validation against ground-based data," *Remote Sens. Environ.*, vol. 163, pp. 270–285, 2015.
- [40] J. M. Chen, "Spatial scaling of a remotely sensed surface parameter by contexture," *Remote Sens. Environ.*, vol. 69, no. 1, pp. 30–42, 1999.
- [41] M. Claverie, E. F. Vermote, M. Weiss, F. Baret, O. Hagolle, and V. Demarez, "Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France," *Remote Sens. Environ.*, vol. 139, pp. 216–230, 2013.
- [42] F. Yang, H. Ren, X. Li, M. Hu, and Y. Yang, "Assessment of MODIS, MERIS, GEOV1 FPAR products over Northern China with ground measured data and by analyzing residential effect in mixed pixel," *Remote Sens.*, vol. 6, no. 6, pp. 5428–5451, 2014.
- [43] J. Hu et al., "Performance of the MISR LAI and FPAR algorithm: A case study in Africa," *Remote Sens. Environ.*, vol. 88, no. 3, pp. 324–340, 2003.
- [44] H. Fang et al., "Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties," *J. Geophys. Res., Biogeosci.*, vol. 118, no. 2, pp. 529–548, 2013.
- [45] H. Fang, W. Li, and R. B. Myneni, "The impact of potential land cover misclassification on MODIS leaf area index (LAI) estimation: A statistical perspective," *Remote Sens.*, vol. 5, no. 2, pp. 830–844, 2013.
- [46] L. Liu, X. Zhang, Y. Gao, X. Chen, X. Shuai, and J. Mi, "Finer-resolution mapping of global land cover: Recent developments, consistency analysis, and prospects," *J. Remote Sens.*, vol. 2021, vol. 2021, Art. no. 528697.
- [47] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.
- [48] G. Zhang and Y. Lu, "Bias-corrected random forests in regression," *J. Appl. Statist.*, vol. 39, no. 1, pp. 151–160, 2012.
- [49] J. Song, "Bias corrections for Random Forest in regression using residual rotation," *J. Korean Stat. Soc.*, vol. 44, no. 2, pp. 321–326, 2015.
- [50] H. Ebrahimi, H. Aghighi, M. Azadbakht, M. Amani, S. Mahdavi, and A. A. Matkan, "Downscaling MODIS land surface temperature product using an adaptive random forest regression method and Google Earth Engine for a 19-years spatiotemporal trend analysis over Iran," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2103–2112, 2021.
- [51] R. Wang, W. Gao, and W. Peng, "Downscale MODIS land surface temperature based on three different models to analyze surface urban heat island: A case study of Hangzhou," *Remote Sens.*, vol. 12, no. 13, 2020, Art. no. 2134.



Yiting Wang was born in Hanzhong city, Shaanxi Province, China, in 1986. She received the B.S. and Ph.D. degrees in geographic information systems from Beijing Normal University, Beijing, China, in 2008 and 2018, respectively, the M.S. degree in photogrammetry and remote sensing from Peking University, Beijing, China, in 2011, the M.S. degree in earth sciences from the University van Amsterdam, Amsterdam, The Netherlands, in 2012.

She is currently an Associate Professor with the College of Geomatics, Xi'an University of Science and Technology, Xi'an, China. Her primary interests include scale effects in remote sensing, remote sensing of vegetation, and scaling of remote sensing products.



Yinggang Zhan received the B.S. degree in geographic information science from Henan University of Economics and Law, Zhengzhou, China, in 2019. He is currently working toward the M.S. degree with the College of Geomatics, Xi'an University of Science and Technology, Xi'an, China.

His research interests include remote sensing of vegetation, scaling of remote sensing products and cloud computing techniques.



Guangjian Yan (Senior Member, IEEE) received the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 1999.

He is currently a Professor with the State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing, China. He has authored or coauthored more than 200 papers. His research interests include multiangular remote sensing, vegetation remote sensing, scale issues, and radiation budgets.



Donghui Xie received the Ph.D. degree in remote sensing and geographic information systems from Beijing Normal University, Beijing, China, in 2005.

From 2005 to 2007, she was a Postdoctoral Research Associate with Beijing Normal University. She is currently with the State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing, China. Her research interests include canopy radiative transfer modeling, biophysical parameter retrieval of vegetation, and remote sensing data fusion.