



OPEN ACCESS

EDITED BY

Ram P. Sharma,
Tribhuvan University, Nepal

REVIEWED BY

Fugen Jiang,
Central South University Forestry and
Technology, China
Suwit Ongsomwang,
Suranaree University of Technology, Thailand

*CORRESPONDENCE

Qingtai Shu

✉ shuqt@swfu.edu.cn

RECEIVED 15 May 2025

ACCEPTED 10 July 2025

PUBLISHED 12 August 2025

CITATION

Zhou W, Shu Q, Xia C, Xu L, Xiang Q, Fu L,
Yang Z and Wang S (2025) Forest canopy
closure estimation in mountainous
southwest China using multi-source
remote sensing data.
Front. Plant Sci. 16:1629146.
doi: 10.3389/fpls.2025.1629146

COPYRIGHT

© 2025 Zhou, Shu, Xia, Xu, Xiang, Fu, Yang and
Wang. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Forest canopy closure estimation in mountainous southwest China using multi-source remote sensing data

Wenwu Zhou^{1,2}, Qingtai Shu^{2*}, Cuifen Xia², Li Xu², Qin Xiang²,
Lianjin Fu³, Zhengdao Yang² and Shuwei Wang²

¹Guangyuan Forestry Workstation, Guangyuan, China, ²College of Forestry, Southwest Forestry University, Kunming, China, ³Faculty of College of Soil and Water Conservation, Southwest Forestry University, Kunming, China

Forest canopy closure (FCC) is an important biological parameter to evaluate forest resources and biodiversity, and the use of multi-source remote sensing synergy to achieve high-accuracy estimate regional FCC at low cost is a current research hotspot. In this study, Shangri-La City, a mountainous area in southwest China, was considered as the research area. The satellite-borne LiDAR ICESat-2/ATLAS data were used as the main information source. Combined with 54 measured plot data, the improved machine learning model of the Bayesian optimization (BO) algorithm was used to obtain the FCC in the footprint-scale ATLAS footprint. Then, the multi-source remote sensing image Sentinel-1/2 and terrain factors were combined to perform regional-scale FCC remote sensing estimation based on the geographically weighted regression (GWR) model. The research results showed that (1) among the 50 extracted ATLAS LiDAR feature indices, the best footprint-scale modeling factors are Landsat_perc, h_dif_canopy, asr, h_min_canopy, toc_roughness, and n_touc_photons after random forest (RF) feature variable optimization; (2) among the BO-RFR, BO-KNN, and BO-GBRT models developed at the footprint scale, the FCC results estimated by the BO-GBRT model were the best ($R^2 = 0.65$, RMSE = 0.10, RS = 0.079, and $P = 79.2\%$), which was used as the FCC estimation model for 74,808 footprints in the study area; (3) taking the FCC value of ATLAS footprint scale in forest land as the training sample data of the regional-scale GWR model, the model accuracy was $R^2 = 0.70$, RMSE = 0.06, and $P = 88.27\%$; and (4) the R^2 between the FCC estimates from regional-scale remote sensing and the measured values is 0.70, with a correlation coefficient of 0.784, indicating strong agreement. Additionally, the average FCC is 0.50, predominantly distributed between 0.3 and 0.6, comprising 68.43%. These findings highlight the advantages of mountain FCC estimation using ICESat-2/ATLAS high-density,

high-precision footprints and the fact that small-sample estimation results at the footprint scale can serve as training data for the regional-scale GWR model, offering a reference for low-cost, high-precision FCC estimation from footprint scale to regional scale.

KEYWORDS

ICESat-2/ATLAS, Bayesian optimization algorithm, machine learning method, geographically weighted regression, multi-source remote sensing data, forest canopy closure

1 Introduction

Forest canopy closure (FCC) refers to the ratio of crown projection area to forest area in the stand (Meng, 2006) and is a basic parameter of stand structure and stand environment, as well as an evaluation index of forest tending and cutting (Meng, 2006; Chen et al., 2019; Li and Mao, 2020; Hua and Zhao, 2021; Pu et al., 2021); timely and non-destructive high-precision estimation of FCC is of great significance for understanding and monitoring the impact of human activities and climate change on forest ecosystems (Wang et al., 2015). The traditional direct measurement methods of FCC mainly include canopy closure measuring instrument, crown projection, measuring line, and artificial visual (Meng, 2006; Wang et al., 2015), which rely on manual and small-scale accurate measurement, but these are time-consuming, labor-intensive, and ineffective (Chen et al., 2019); furthermore, they cannot meet the research standards of the spatial distribution and variation of FCC at large spatial scales (Wang et al., 2015). The development and application of remote sensing technology combined with small sample standard measured data for regional-scale FCC inversion is a common method due to the low-cost, high-efficiency, and global coverage of remote sensing data resources (Lee and Lucas, 2007; Yang et al., 2022).

At present, there are many studies on estimating FCC using optical remote sensing data or airborne light detection and ranging (LiDAR) data combined with different methods (Lee and Lucas, 2007; Hua and Zhao, 2021; Xu et al., 2022; Yang et al., 2022; Duan et al., 2023). However, optical remote sensing data are susceptible to spectral saturation, and airborne LiDAR data are expensive and difficult to obtain. Using synthetic aperture radar (SAR) data (Neumann et al., 2009; Varghese and Joshi, 2015) and spaceborne LiDAR [ice, cloud, and land elevation satellite/geoscience laser altimeter system (ICESat/GLAS)] data to estimate FCC is relatively inadequate (Wang et al., 2015; Cui et al., 2021), but the GLAS footprint is larger (footprint diameter is 70 m and footprint interval is 170 m), and it is susceptible to terrain, especially in complex alpine regions, which will lead to an improvement in FCC estimation accuracy. Compared with ICESat-2/ATLAS (ice, cloud, and land elevation satellite/advanced topographic laser altimeter system), ATLAS has a smaller footprint size (footprint diameter is

only 17 m and footprint interval is 0.7 m), which greatly reduces the influence of terrain on spot echo (Moudry et al., 2022). In the studies conducted by Hua and Zhao (2021); Li and Mao (2020), and Yang et al. (2022), data from over 70 measured samples combined with remote sensing data were used to quantitatively invert the FCC. However, this study investigated only 54 samples, which not only meets the principle of a large sample size (50) and the accuracy requirements of field investigations (Song et al., 2022a), but also reduces experimental costs. Currently, numerous studies focus on the inversion of forest vertical structure (e.g., forest canopy height and forest height), forest biomass, and understory topography using the latest generation of photon-counting spaceborne LiDAR and ICESat-2/ATLAS (Narine et al., 2019; Lin et al., 2020; Zhu et al., 2020; Song et al., 2022a, b). However, there are few studies focusing on the estimation of forest horizontal structure parameters (e.g., leaf area index and FCC) (Xi et al., 2023). Because the ATLAS footprint data show a spatial discontinuous distribution in the form of strips, it cannot meet the requirements for full coverage in the study area (Narine et al., 2019; Zhu et al., 2020). Therefore, the parameter indicators need to be predicted by choosing the spatial interpolation method or spatial regression method in geostatistics in order to obtain the faceted attribute data covering the continuity of the whole study area and then realize the remote sensing mapping of the FCC (Wang et al., 2015; Zhu et al., 2020; Zhou et al., 2023; Yu et al., 2024; Zhou et al., 2024). For example, it combines ground target information from continuous points with continuous surface remote sensing data (e.g., Sentinel-1/2) to achieve multi-source integration and assimilation of multi-sensor and auxiliary data. This approach aims to improve the accuracy of FCC estimation (Zhao et al., 2016; Zhou et al., 2023, 2024) and enable FCC remote sensing mapping at the regional scale.

ICESat-2/ATLAS and Sentinel-1 are active remote sensing technologies. ATLAS employs advanced photon-counting LiDAR technology, featuring more sensitive single-photon detectors and a higher pulse repetition frequency (Neumann et al., 2019), enabling the acquisition of photon point cloud data with smaller footprints and higher sampling density (Lin et al., 2020). The C-band of SAR possesses penetrability and dielectric properties, making it resilient to factors such as region, time, and climate during imaging, thus capturing the structural characteristics of forests (Zhang et al., 2022). However, it

lacks rich spectral information. The multi-spectral sensor of Sentinel-2 can capture electromagnetic radiation information outside the canopy, providing rich canopy data. Its red-edge band enhances FCC estimation accuracy (Hua and Zhao, 2021), but it is limited in acquiring information about tree trunks and branches. Characteristic variables significantly impact the estimation accuracy and inversion results of the model (Zhang et al., 2022). Therefore, in order to reveal the explanatory and contribution of multiple variable factors to FCC, reduce the influence of spectral saturation on vegetation (Zhao et al., 2016), and improve the prediction accuracy of the model, the independent variable factors at the spot scale in this study were determined by 50 parameter values extracted by ICESat-2/ATLAS after feature variable optimization. At the regional scale, the commonly used remote sensing factors such as SAR factor, texture feature, vegetation index and single band reflectivity (Chen et al., 2019; Zhang et al., 2022), and the necessary auxiliary data terrain factor were selected to construct the FCC extrapolation model.

Based on the measured sample size and remote sensing dataset, an appropriate model is selected to achieve the best estimation results (Shu et al., 2022). Machine learning methods offer greater advantages in model fitting accuracy and inversion results compared to traditional statistical methods (Shu et al., 2022). In the canopy closure studies by Hua and Zhao (2021); Wang et al. (2015), and Xu et al. (2022), nonparametric models demonstrated the highest accuracy, and the estimation results were verified. However, the lack of optimization algorithms suggests that model accuracy and estimation results can be further improved. In this study, machine learning methods such as K-NN, RFR, and GBRT are selected as the basic models at the footprint scale. The Bayesian optimization (BO) algorithm is then employed to enhance the performance of these basic models, aiming to construct the optimal FCC estimation model. The BO algorithm leverages prior knowledge to approximate the posterior distribution of the unknown objective function and then selects the next best hyperparameter combination based on this distribution, thereby quickly reducing the computational load while optimizing model performance and improving estimation accuracy (Cui and Yang, 2018; Zhang et al., 2021b). As a sequential optimization method, BO effectively explores and balances the known parameter space and the unknown parameter space through surrogate models and acquisition functions; it is capable of obtaining a globally approximate optimal solution with minimal evaluation costs, thereby avoiding the pitfalls of local optima (Cui and Yang, 2018; Zhang et al., 2021b). In the same nonparametric optimization model, the BO algorithm can reduce simulations for model optimization and improve the model operation rate, increase model estimation accuracy, and provide forecast uncertainty more than particle swarm optimization (PSO), genetic algorithm (GA), and differential evolution (DE) (Zhang et al., 2021b). It is one of the commonly used algorithms to optimize the performance of nonparametric models. This study employed the geographically weighted regression (GWR) model to construct an FCC estimation model for the study area at a regional scale. The GWR model is a local spatial regression technique that predicts unknown spatial variables using known data (Guo et al., 2012; Nazeer and Bilal, 2018;

Zhang et al., 2019b; Song et al., 2022c). Although widely used in urban geography, forestry, and other disciplines (Guo et al., 2012; Zhang et al., 2019b), its application to medium-scale and large-scale FCC estimation remains uncommon.

At present, there are few studies on the use of ICESat-2/ATLAS data for estimating FCC, particularly in combination with spaceborne LiDAR and multi-source remote sensing data for cost-effective, regional-scale canopy closure inversion. In this study, ICESat-2/ATLAS data were employed to extract the modeling parameters. Using the BO-RFR, BO-KNN, and BO-GBRT models, the optimal FCC estimation model for footprints was constructed from 54 measured plot datasets. Sentinel-1/2 imagery and digital elevation model (DEM) data were used as sources to extract remote sensing factors. After conducting an OLS (ordinary least squares) test and normal transformation, an FCC extrapolation model was constructed using the GWR model to obtain continuous spatial distribution of FCC information across the study area. The aims of this study are to construct a portable footprint canopy closure estimation model and to explore and verify the feasibility and reliability of the regional FCC estimation method based on the GWR model and multi-source remote sensing data. At the same time, the optimization ability of the BO algorithm to the machine learning model is explored, and a low-cost and high-precision method of estimating FCC is proposed.

2 Research materials

2.1 Study area

Shangri-La City is located in the northwest of Yunnan Province, China (latitude 26°52'11.44"–28°50'59.57" N, longitude 99°23'6.08"–100°18'29.15" E), which belongs to a typical alpine terrain of the Yunnan, Sichuan, and Tibet triangle region (Shu et al., 2022; Song et al., 2022b), as shown in Figure 1. The general trend of terrain in the area is high in the northwest and low in the southeast, with a relative elevation difference of 4,042 m; moreover, the average altitude is 3,459 m, the average temperature is 4.7°C–16.5°C, and the average annual rainfall is 649.4 mm; in summary, it belongs to the mountain cold temperate monsoon climate. The total land area of Shangri-La city is 1,141,739 ha², of which forestry land area is 950,911.7 ha², which accounts for 83.3% of the total area; the forest coverage rate reaches 76%, which is an important protection forest area in Yunnan Province of China. Typically, there are 10 vegetation types distributed in the city and the main vegetation type is cold temperate coniferous forests including *Picea asperata*, *Abies fabri*, *Pinus densata*, *Quercus semecarpifolia*, and *Larix gmelinii* (Shu et al., 2022; Song et al., 2022b; Xi et al., 2023).

2.2 Sample plot design and data preprocessing

The 54 sample plot data used in the study were sampled in November 2021 in Shangri-La City. The experimental design selects 54 sample circles with a radius of 8.5 m and an area of

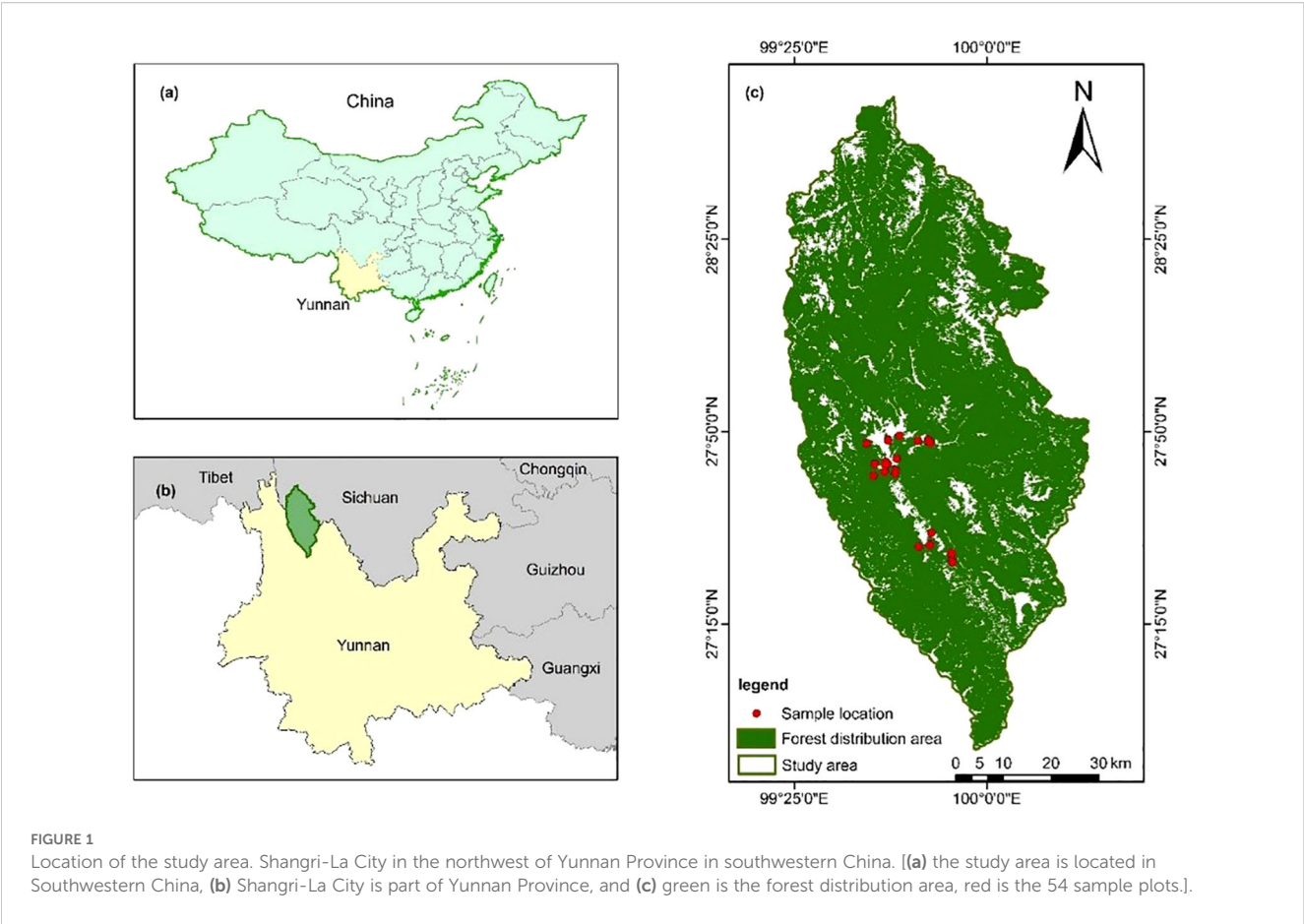


TABLE 1 Descriptive analysis of FCC statistics.

Statistical	Maximum	Minimum	Mean	SD	Variance	Median
54	0.83	0.20	0.50	0.176	0.031	0.5

approximately 0.023 ha², which are consistent with the footprint size emitted by the ATLAS sensor mounted on ICESat-2; it covers the main forest vegetation types at different slopes and altitudes in the study area and records the coordinate information of the samples, tree species, diameter at breast height, tree height, and the measurement of the FCC. Among them, the latitude and longitude coordinates of the center point of the sample circle are consistent with the coordinates of the footprint center of ICESat-2/ATLAS, and using the southern mapping T66 Real-Time Kinematic (RTK) in the fixed solution state, the mean value of five consecutive point lofting was taken by the differential positioning instrument of Thousand Seeker SR3 (Pro version), the error between the center point coordinates of all sample sites and those of the footprints was less than 0.02 m, and, finally, the latitude and longitude coordinates of the center point of the sample circle are determined. In this paper, the measuring line method (Meng, 2006) is used to calculate the FCC of 54 sample circles (Table 1). The definition of the measuring line method is as follows: to select a representative section in the sample circle, set up a certain length of the

measuring line, along the line to observe the projection of the crown of each tree, and measure the projection length, the crown of the projected length of the sum of the length of the measuring line, and the length of the measuring line for the value of the degree of canopy closure (Meng, 2006).

2.3 ICESat-2/ATLAS data products acquisition and preprocessing

2.3.1 ICESat2/ATLAS data acquisition

The ICESat-2 satellite was the first to be equipped with a photon-counting LiDAR payload on a spaceborne platform. In September 2018, it was successfully launched by NASA (National Aeronautics and Space Administration) at the Vandenberg Space Force Base in the United States. The ATLAS system laser on board launched a total of six laser beams at a time. The photon point cloud data with a footprint diameter of 17 m and a sampling interval of 0.7 m were obtained (Neumann et al., 2019; Moudry et al., 2022); its 22 standard

data products are divided into four levels and stored in the US Ice and Snow Data Center (<https://nsidc.org/data/icesat-2/data-sets>) in HDF5 format (Neumann et al., 2019; Lin et al., 2020; Zhu et al., 2020). The ATL03 global positioning photon data contain six laser beam bands, which are evenly segmented at a distance of 20 m along the track, labeled as gt11–gt3r, and record the time, latitude and longitude of all photon events. Geospatial location information, as well as information such as the number of photons, belongs to the secondary product data. Based on this, the source data can generate more advanced products (Zhu et al., 2020). The ATL08 product is a geophysical data product containing ground elevation information and vegetation height information generated in 100-m segments along the orbital direction based on ATL03 data after further noise removal and signal photon classification (Zhu et al., 2020; Moudry et al., 2022). This study utilized free ICESat2/ATLAS data obtained from the earthdata website (<https://search.earthdata.nasa.gov/>). All ATL03 and ATL08 data products from January 2020 to June 2021 in Shangri-La were selected, and each dataset comprises a total of 118 data points, 354 tracks, and 708 photon trajectory beams.

2.3.2 Photon point cloud denoising and classification algorithm

Because ATLAS is a more sensitive single-photon detector compared to GLAS and has a higher pulse repetition frequency and a weak signal emission, it also captures a significant amount of noise photons when receiving reflected photons from specific ground targets (Neumann et al., 2019; Zhu et al., 2020). Therefore, to use these data for quantitative remote sensing inversion, noise photons must be removed to improve the accuracy of model estimation. In this paper, we employ a combination of the different densities-based spatial clustering of applications with noise (DDBSCAN) and k-nearest neighbors-based (KNNB) algorithms (Zhang et al., 2021a) for denoising. It is demonstrated that this combined approach outperforms the use of either the DDBSCAN or KNNB algorithm alone (Nie et al., 2018; Zhang et al., 2021a). Additionally, the final measurement parameter is replaced by the maximum density difference in DDBSCAN to address the impact of photon density inconsistency on algorithm performance.

The signal photons after denoising need to be accurately classified, which are mainly divided into ground photons and canopy top photons. The classification results will affect the inversion and mapping accuracy of forest parameters (Zhu et al., 2020). The progressive triangular irregular network (TIN) densification (PTD) method was used to distinguish the photon point cloud data into ground photons and canopy photons (Nie et al., 2017, 2018; Zhang et al., 2021a). This method has high ground photon recognition accuracy in complex terrain areas such as large altitude drop. In order to further improve the classification accuracy, the ground point is set to the lowest elevation point under the farthest point from TIN.

2.3.3 Footprint-scale parameter extraction and forest footprint distribution map

After further photon point cloud denoising and classification, the number of effective photons of ATL03 reaches tens of millions.

According to the ATL08 product, 94,039 effective footprints in the study area were obtained by thinning sampling in a 100-m section. The latest sub-compartment attribute data of forest resources survey in Shangri-La City (2016) were used for overlay analysis. In the study area, 74,808 effective forest footprints (Figure 2) and 19,231 non-forest footprints were obtained, a total of 50 parameters (including 54 standard footprints data) in the effective forest footprints were extracted, and the parameter introduction is detailed in the literature (Nie et al., 2018; Zhang et al., 2021a).

2.4 Preprocessing and feature variable extraction of regional-scale remote sensing data

2.4.1 Preprocessing of regional-scale remote sensing data

The study utilized Sentinel-1/2 images captured in October 2021, which were freely downloaded from the European Space Agency (ESA) (<https://scihub.copernicus.eu/dhus/#/home>) in November 2021. The SAR data include C-band dual-polarization (VV and VH) single-look complex data from the Sentinel-1A satellite, acquired in ground range detected (GRD) Level 1 product interferometric wide (IW) mode. The sensor operates at a center frequency of 5.405 GHz, with a swath width of 250 km and a spatial resolution of 15 m × 15 m after resampling. To extract the backscattering coefficient from dual-polarization backscatter images, SNAP (sentinel application platform) software was employed for data preprocessing steps including precise orbit determination, thermal noise removal, radiometric calibration, multi-looking, speckle filtering, geocoding, and dB conversion (Figures 3a, b). The Sentinel-2 L2A-level multispectral data used is a product of L1C-level images after Sen2cor atmospheric correction. Using SNAP software, each band is resampled to 15 m by three convolutions, and then 5-m SPOT-5 high-precision images are used for geometric correction, and the SCS + C model is used for topographic correction (Figure 3c).

2.4.2 Regional-scale feature variables extraction

A total of 91 feature variables were extracted, including remote sensing factors such as texture features, vegetation indices, single-band reflectivity, SAR factors, and three terrain factors (Table 2). All feature variables were extracted using ENVI 5.6 software. VV/VH represents the ratio of VV to VH, while VV–VH denotes the difference between VV and VH. Texture features were generated using the gray-level co-occurrence matrix (GLCM) method within the second-order texture algorithm. The window size was set to 5 × 5, the step size was set to 1, and the gray level was set to 64, resulting in the extraction of eight texture features.

2.5 Digital elevation model data

In this study, DEM data with a spatial resolution of 12.5 m were used to extract three topographic factors: slope, aspect, and

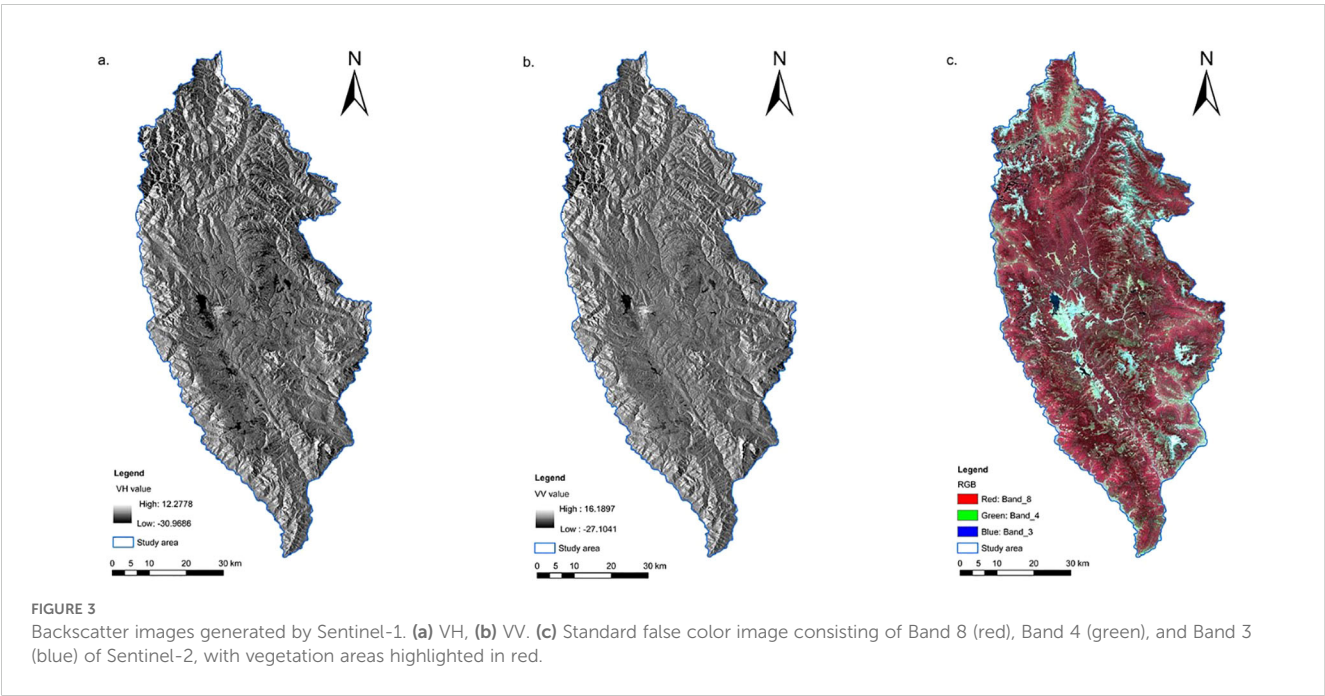
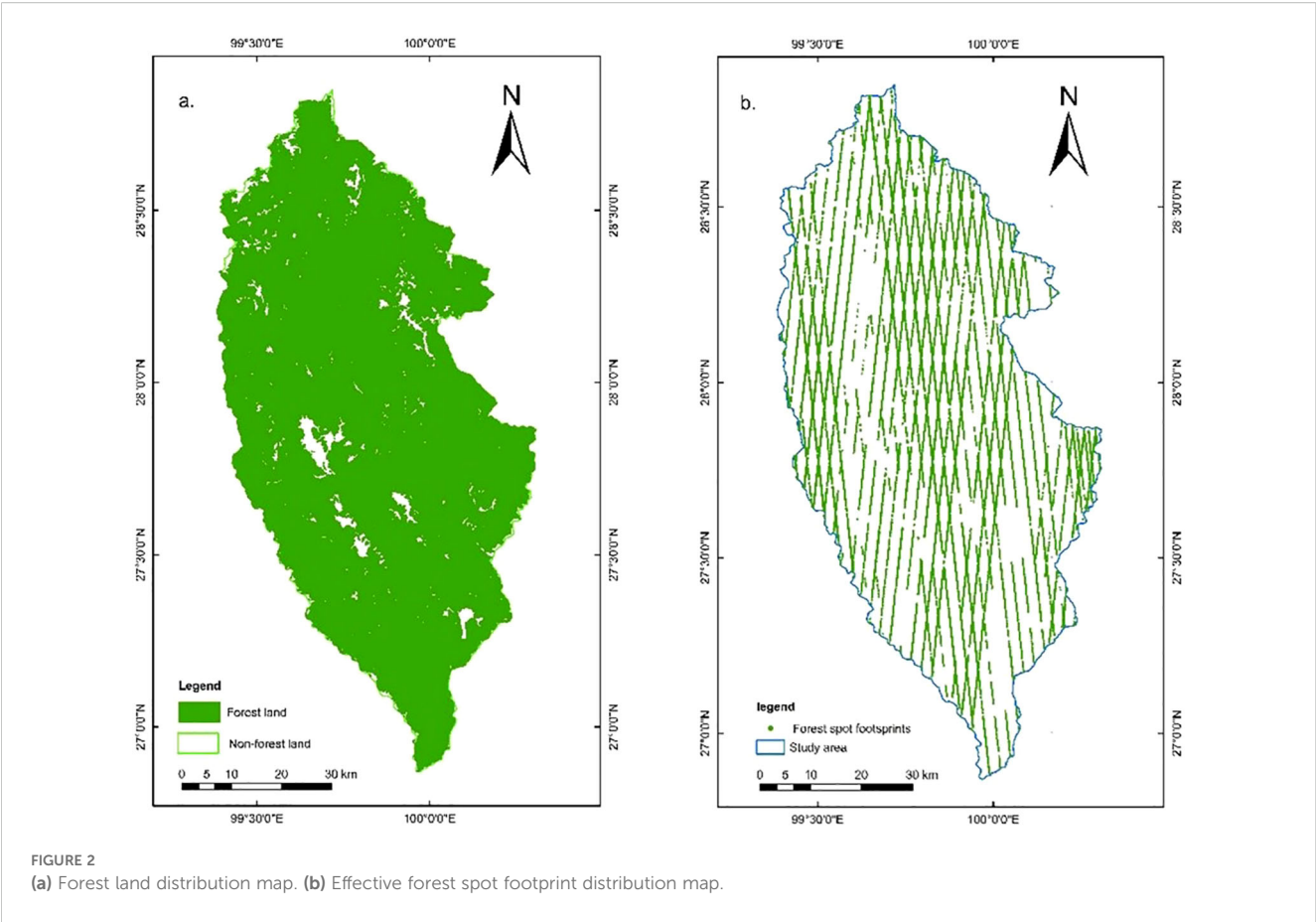


TABLE 2 Factor extraction of regional-scale feature variables.

Data sources	Variable type	Variable name	Number of variable
Sentinel-1	Backscattering coefficient	VV, VH, VV/VH, VV-VH	4
Sentinel-2	Single-band factor	Band 2, Band 3, Band 4, Band 5, Band 6, Band 7, Band 8, Band 8A	8
	Vegetation index	NDVI, DVI, SAVI, OSAVI, EVI, EVI2, RVI, MASVI, GNDVI, GRVI, RDVI, IDVI	12
	Texture feature	SE_ME, SE_VA, SE_HO, SE_CO, SE_DI, SE_EN, SE_SM, SE_CR	64
DEM	Topographic factors	Slope, aspect, elevation	3

SE_XXX represents the texture feature generated by the single band of Sentinel-2, SE represents the single band of B2-B8A, and XXX represents the texture feature.

elevation. The data were obtained from the polarimetric synthetic aperture radar (PolSAR) sensor aboard the ALOS satellite and were freely downloaded from the official earthdata website (<https://www.earthdata.nasa.gov>, accessed in November 2021).

3 Research methods

The methodology consists of four main steps (Figure 4): (1) dataset collection, preprocessing, and index extraction; (2) selection and modeling of footprint-scale characteristic variables and canopy closure estimation; (3) selection and modeling of regional-scale characteristic variables and FCC estimation for the study area; and (4) spatial mapping and analysis of FCC in the study area.

3.1 Bayesian optimization algorithm

The BO algorithm can obtain a global approximate optimal solution with little evaluation cost; in addition, the famous “Bayesian theorem” is used in the optimization process (Cui and Yang, 2018). The core is to use the probability model to represent the costly complex objective function of the original evaluation (Zhang et al., 2021b); the active selection strategy is constructed by using the posterior information of the surrogate model, that is, the acquisition function (Cui and Yang, 2018); this results in the probability model more accurately satisfying the behavior of the black box function and effectively reducing unnecessary sampling, thus theoretically ensuring the final convergence to the global optimal solution (Cui and Yang, 2018). In short, to reduce the model calculation amount and optimize the target model

parameters, the model estimation accuracy should be improved. The Bayesian formula is as follows:

$$p(f|D_{1:t}) = \frac{p(D_{1:t}|f)p(f)}{p(D_{1:t})}$$

where f represents the unknown objective function (parameters in the optimization model), $D_{1:t} = \{(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)\}$ represents the observed set, x_t represents the decision vector, $y_1 = f(x_t) + \varepsilon_t$ represents the observed value, ε_t represents the observation error; $p(D_{1:t}|f)$ represents the likelihood distribution of y , due to the error of the observed value, $p(f)$ represents the prior probability distribution of f , that is, the assumption of the unknown objective function state, $p(D_{1:t})$ denotes the marginal likelihood distribution of the marginalized f , which is mainly used in BO for hyperparameters, $p(f|D_{1:t})$ represents the posterior probability distribution of f , and the confidence of the unknown objective function after the prior is corrected by the observed dataset.

The BO process is an iterative process (Cui and Yang, 2018), and the optimization framework is shown in Table 3. There are three core steps: (1) Select the next evaluation points with the highest “potential” x_t according to the maximum acquisition function. (2) Calculate the objective function value $y_t = f(x_t) + \varepsilon_t$ according to the selected evaluation point x_t . (3) The newly obtained input–observation pair $\{x_t, y_t\}$ is added to the historical observation set $D_{1:t-1}$, and the probabilistic surrogate model is continuously updated to prepare for the next model iteration. The research mainly optimized the important parameters of random forest regression (RFR), gradient boosting regression tree (GBRT), and K-nearest neighbor (K-NN) models for 1,000 times to find the best parameters for modeling. The algorithm flow is shown in Figure 5.

3.2 Footprint-scale estimation model

In this study, RFR, GBRT, and K-NN models were selected as the basic models for footprint-scale FCC estimation. The introduction of each model is shown in various studies (Franco-Lopez et al., 2001; Coulston et al., 2016; Tian et al., 2021). The BO algorithm was used to optimize the three basic models to find the best kernel parameters for modeling and accurately estimate the footprint-scale FCC. The optimization parameters are shown in Table 4.

3.3 Regional-scale estimation model of geographically weighted regression model

In GWR, the locally weighted least squares method is used to solve for local parameters (Guo et al., 2012; Zhang et al., 2019b), with weights calculated based on the spatial distance between the location to be estimated and the locations of other observation points (Nazeer and Bilal, 2018). As an extension of the linear

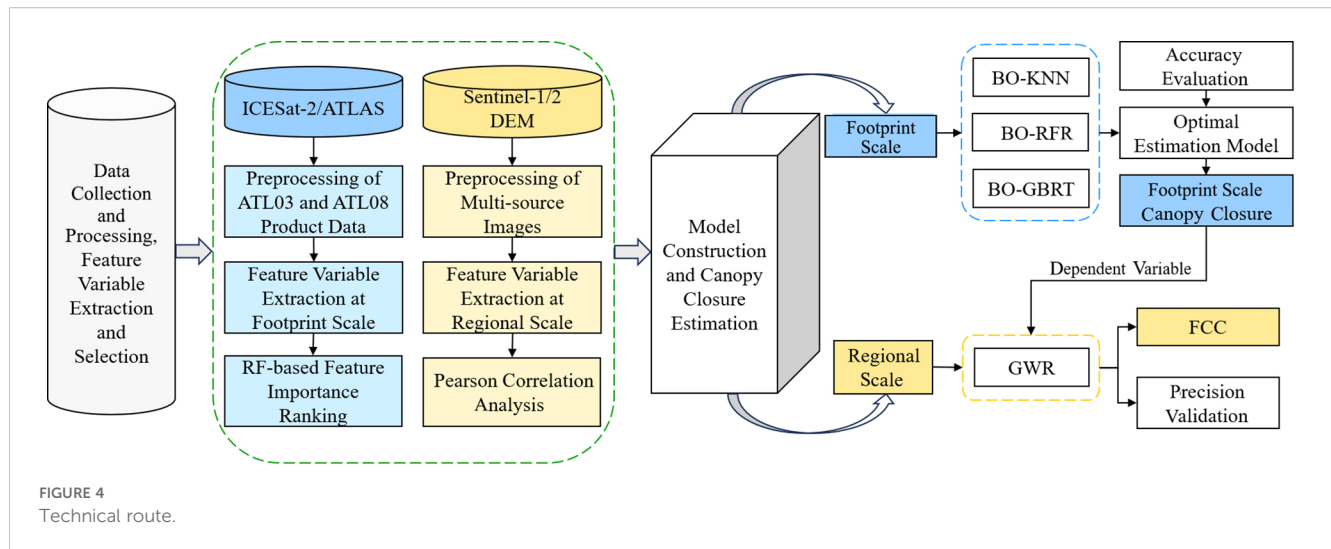


TABLE 3 Framework of the bayesian optimization algorithm.

Algorithm: Bayesian optimization algorithm
Input: proxy model f , acquisition function α
Output: hyperparameter result X^*
1: Initialize the hyperparameter x
2: for $t = 1, 2, \dots, T$ do
3: Maximize the acquisition function to get the next evaluation point: $x_t = \operatorname{argmax}_{x \in X} \alpha(x D_{1:t-1})$
4: Evaluate the objective function value $y_t = f(x_t) + \varepsilon_t$
5: Integrate data: $D_t = D_{t-1} \cup \{x_t, y_t\}$, and update the proxy model
6: end for

regression model, GWR has strong spatial variability and correlation in spatial position, which can effectively explain the influence of different independent variables on target variables in different spatial positions. The mathematical model of GWR is as follows (Song et al., 2022c):

$$y_i = a_0(u_i, v_i) + \sum_k a_k(u_i, v_i)x_{ik} + \varepsilon_i$$

where y_i is the target variable of point i , x_{ik} is the value of the k independent variable in i , k is the independent variable count, and i is the sample point count, ε_i is the residual error, (u_i, v_i) is the spatial coordinates of the i sample point, $a_k(u_i, v_i)$ is the local regression coefficient at point i , that is, the spatial location function.

Because the research data are continuous, the Gaussian kernel function model is selected to construct the spatial weight matrix, and the calculation formula is as follows (Nazeer and Bilal, 2018):

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta^2}\right)$$

where i, j denotes the spatial location of the regression point, w_{ij} is the weight value of the observation at the position j representing the coefficient at point i , d_{ij} is the Euclidean distance between i and j , θ is the fixed bandwidth size defined by the distance metric.

As GWR is weak in the diagnosis of independent variable factors; OLS is needed for the collinearity diagnosis and significance test of independent variable factors to judge the feasibility of constructing GWR, to select the independent variable factors that fit the GWR model, and finally to improve the accuracy of the GWR model and to construct a more realistic regression model.

3.4 Evaluation of model accuracy

In this study, the coefficient of determination (R^2), root mean square error (RMSE), mean absolute residual (MAR), and prediction accuracy (P) of the leave-one-out cross validation (LOOCV) method were used to verify the prediction accuracy of the estimation model (Shu et al., 2022; Song et al., 2022b). This method is applied to small sample data for sequential training and verification, addressing local optimization issues and enhancing model robustness (Shu et al., 2022). Additionally, compared to K-fold and holdout cross-validation, LOOCV is not influenced by random factors (Song et al., 2022b), thereby reducing the uncertainty of model estimation results. The calculation formula is as follows:

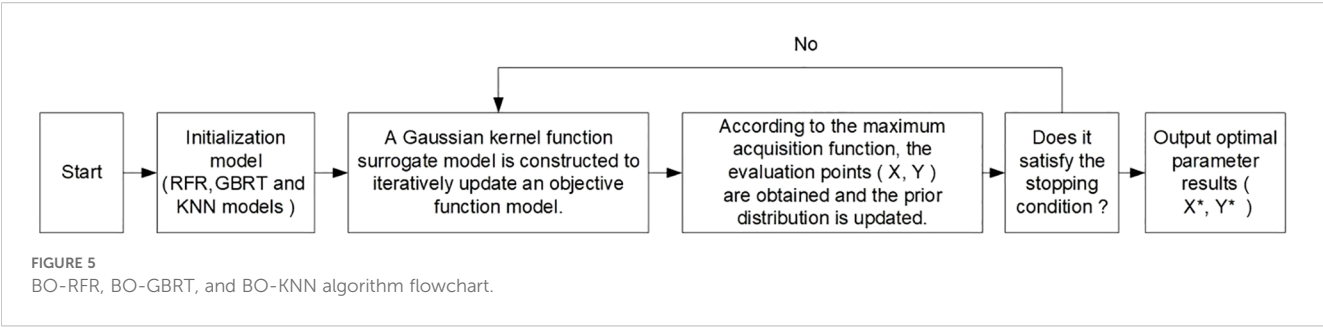
$$R^2 = \frac{\sum_i^N (\hat{y}_i - \bar{y})^2}{\sum_i^N (y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_i^N (y_i - \hat{y}_i)^2}{N}}$$

$$P = (1 - \frac{RMSE}{\bar{y}}) \times 100\%$$

$$MAR = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where \hat{y}_i is the model prediction value; \bar{y} is the average model prediction value; y_i is the canopy closure measured value; and N is the total number of verification samples.



4 Results and analysis

4.1 Optimization results of characteristic variables

4.1.1 Optimization results of footprint-scale characteristic variables

In this study, the extracted 50 parameters values were used to evaluate the importance of features using RF (Glenn et al., 2016; Chen et al., 2022). All parameters have a certain contribution rate. The 5% of the importance of features was set as a threshold, and six feature variables were selected as the best independent variables for modeling (Table 5). Among them, Landsat_perc has the highest feature contribution of 12.68%, and asr has the lowest feature importance of 5.03%.

4.1.2 Optimization results of regional-scale characteristic variables

The remaining 13 independent variables after the preferential selection of the 91 regional-scale characteristic variables extracted using Pearson correlation analysis (Duncanson et al., 2020) had correlations greater than 0.2 and significant at the 0.01 level (Table 6). Among them, the average correlation of the GLCM generated by the green edge band and the red edge band was strong, which may be related to the fact that the texture feature factors can describe more detailed forest structure information (Shu et al., 2022). This result is consistent with the results of Zhang et al. (2016) in the Daxing'anling area of Inner Mongolia. The maximum correlation coefficients of B3_DI and B3_HO were 0.338 and 0.338,

respectively, and the minimum correlation coefficient of GNDVI was 0.22. The VV–VH correlation coefficient based on the difference between the backscattering coefficient VV and VH in the SAR factor is −0.287. Among the terrain factors, only the slope meets the preferred standard, and the correlation coefficient was 0.336.

4.2 ICESat-2/ATLAS footprint-scale indefinite FCC modeling results

Six independent variables selected by ATLAS parameters were used to participate in the modeling to construct the best FCC estimation model of footprint scale. According to the model accuracy (Figures 6a, c, e), the overlap of multiple values in the same interval is caused by the fact that there are multiple measured values that are equal in different places in 54 sample plots. In the modeling results, there is a significant change in model accuracy for the nonparametric model before and after optimization using the BO algorithm. Before optimization (Table 7), the K-NN, RFR, and GBRT models had R^2 between 0.23 and 0.30, an RMSE range of 0.14–0.16, and a P range from 67.26% to 72.73%. After optimization (Table 7), the BO-KNN, BO-RF, and BO-GBRT models had R^2 between 0.41 and 0.65, with an average increase of 48.95% compared with before, an RMSE range of 0.10 to 0.14, with an average error reduction of 20.36% over the previous ones, and a P of 73.12% to 79.22% with an improved accuracy of 5.93% compared to the previous average. Among them, the BO-RFR and BO-GBRT models had the best fitting degree. The BO-GBRT model had the highest R^2 (0.65), a minimum RMSE (0.10), and the highest P (79.22%); hence, the comprehensive evaluation of the model

TABLE 4 Description of RF, GBRT, and K-NN model parameters.

Model type	Parameters	Description	Type
RF, GBRT	max_depth	The maximum depth of the tree	int
	n_estimators	The number of trees in the forest	int
	min_samples_split	The minimum number of samples required to split an internal node	int or float
	min_samples_leaf	The minimum number of samples required to be at a leaf node	int or float
K-NN	n_neighbors	The number of neighbors to use by default for k-neighbors queries	int
	weights	The weight function used in prediction	Str or callable

TABLE 5 Statistics of the results of ICESat-2/ATLAS feature variable preferences.

Variable name	Description	Value (%)
landsat_perc	Average percentage value of the valid Landsat Tree Cover Continuous Fields product for each 100-m segment	12.68%
toc_roughness	Standard deviation of the relative heights of all photons classified as top of canopy within the segment	6.47%
h_min_canopy	The minimum of relative individual canopy heights within segment. Relative canopy heights have been computed by differencing the canopy photon height from the estimated terrain surface	6.06%
h_dif_canopy	Difference between h_canopy and h_median_canopy	5.38%
n_toc_photons	The number of photons classified as top of canopy within the segment	5.37%
asr	Apparent surface reflectance	5.03%

was better. The optimized model residual diagram reflected the deviation between the measured value and the predicted value, and the fluctuation range of residual (Figures 6b, d, f) was between -0.3 and 0.4 . Figure 6b fluctuated greatly, and the mean absolute residual reached 0.116 . The change trend of Figures 6d, f was similar, and the minimum MAR of Figure 6f was 0.079 ; moreover, the error between the measured value and the predicted value was smaller. In summary, the BO-GBRT model had the best comprehensive fitting accuracy, which is selected as the best estimation model for footprint-scale FCC.

4.3 The spatial distribution of FCC in footprint scale

The spatial distribution of FCC values within the ATLAS footprints in the study area was estimated using the BO-GBRT model (Figure 7). FCC was mainly concentrated in the range 0.3 – 0.6 , a few were distributed in the range 0.6 – 0.9 , and a very small part was between 0 and 0.3 . The overall spatial distribution of the depression FCC value in the footprint varied greatly, and the distribution in local areas was relatively uniform. The high-value footprints of depression FCC in the study area were distributed from southeast to north, and the northern

area was the main distribution area of FCC high value with relatively uniform distribution while the southern and south-central areas had more distribution of FCC low-value footprint. As the terrain of the study area is high in the northwest and low in the southeast, the climatic conditions in the southeast are more suitable than those in the northwest and north, where human settlements are found, with a lower FCC value (Song et al., 2022b; Yu et al., 2024). This illustrated the feasibility of using ICESat-2/ATLAS data for estimating FCC.

4.4 The spatial distribution of FCC in the study area

4.4.1 Diagnosis of independent variable factors based on OLS and normal test results

In order to eliminate the influence of multicollinearity among multivariate factors on the GWR model, the 13 preferred multi-source remote sensing factors were analyzed for covariance diagnosis using OLS. The independent variables with variance inflation factor (VIF) greater than 10 were deleted, and the remaining six independent variable factors (Table 8) were significant at the 0.01 level. Among them, when the VIF value of vegetation index was within the range of 6 – 7 , the VIF values of slope factor, SAR factor, and texture feature factor were close to 1 .

The premise of using the GWR model to construct a mathematical model is that the experimental data must conform to normal distribution (Guo et al., 2012; Zhang et al., 2019b). The frequency distribution histogram of the FCC and the six independent variable factors were tested by the data, showing a bell-shaped curve (Figure 8), which conformed to the normal distribution.

4.4.2 Prediction results and verification of the GWR model

NDVI, GNDVI, B8_SM, B8A_CR, VV-VH, and Slope were used as dependent variables; meanwhile, the GWR model tool provided by ArcGIS Pro was used to predict FCC at the regional scale. The bandwidth type is the number of neighbors, and the neighborhood method selects the golden search to find the minimum corrected Akaike information criterion (AICc) to determine the optimal bandwidth. Further AICc represents a measure of the model, and a smaller value indicates that the fitted mathematical model is better (Guo et al., 2012), and the local weight

TABLE 6 Multi-source remote sensing factor preference results statistics.

Variable name	Correlation coefficient	Variable name	Correlation coefficient	Variable name	Correlation coefficient
Slope	0.336**	B3_HO	-0.338^{**}	B7	0.251**
VV-VH	-0.287^{**}	B8_SM	-0.327^{**}	NDVI	0.286**
B3_SM	-0.329^{**}	B8_EN	0.293**	GNDVI	0.220**
B3_EN	0.307**	B6_ME	0.287**		
B3_DI	0.338**	B8A_CR	0.271**		

** is expressed significant at the 0.01 level.

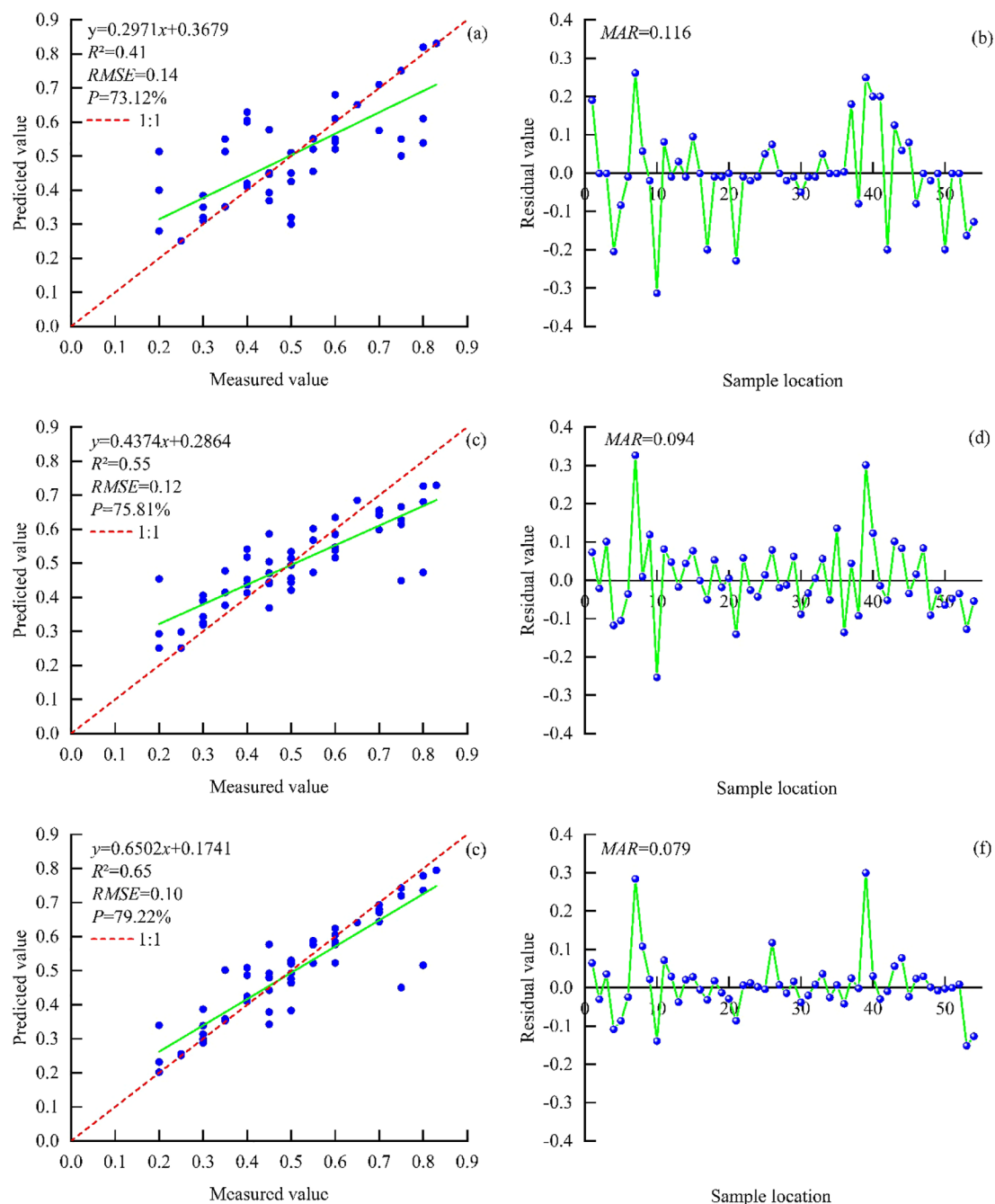


FIGURE 6

FCC shows the accuracy and residual footprint scale of the estimate model with the BO-KNN (a, b), BO-RFR (c, d), and BO-GBRT (e, f) models.

is double squared. According to Figure 9a, the FCC value of ATLAS footprints estimated by the GBRT model was used as the dependent variable of the GWR model, with a GWR model validation accuracy of $R^2 = 0.53$, $RMSE = 0.09$, and $AICc = 106,917.29$. While using the ATLAS footprints, the FCC value is estimated by the BO-GBRT model as the dependent variable of the GWR model (Figure 9b), and the GWR model validation accuracy is $R^2 = 0.70$, which is 32.08% better than before optimization. $RMSE = 0.06$ is reduced by 33.33% compared to that of the comparison before optimization, and the $AICc = 99,265.12$, which is 7.16% lower than that before

optimization. The optimized model residuals were mainly distributed in the range of -0.25 to 0.25 (Figure 9c), the MAR was 0.047 , and the local R^2 was mainly distributed in the range of 0.3 – 0.6 with an average of 0.50 , and the estimation model has high accuracy. At the same time, while handling the large number of modeling samples, it is feasible to use the GWR mathematical model fitted with six explanatory variables to predict the FCC value of the unknown spatial region (Figure 9d).

According to Figure 9d, the average value of FCC was 0.50 , and the values were mainly distributed between 0.3 and 0.6 , accounting

TABLE 7 Modeling results of the footprint-scale depression estimation model.

	Model type	R^2	RMSE	P (%)
Un-optimized	KNN	0.23	0.16	67.26
	RFR	0.28	0.15	70.31
	GBRT	0.30	0.14	72.73
Optimized	BO-KNN	0.41	0.14	73.12
	BO-RFR	0.55	0.12	75.81
	BO-GBRT	0.65	0.10	79.22

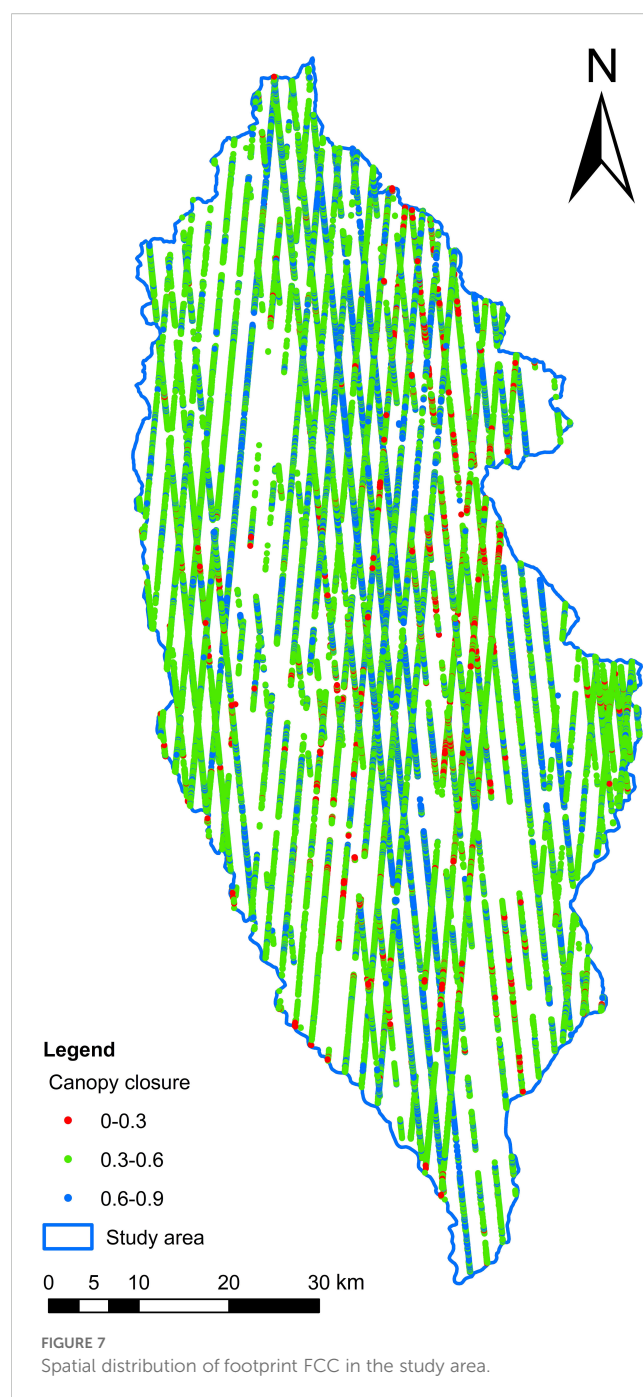
for 68.43%, followed by 0.6–1, accounting for 22.48%, and by 0–0.3, accounting for 9.09%. The areas with low estimates of depression FCC in the study area were mainly distributed at the margins, mostly in the river or perennial snow-covered areas, and in the southeastern urban areas where humans gathered (Song et al., 2022a). The area with high depression FCC runs through from northwest to southeast, and the northern area was the main distribution area with high depression FCC, mainly due to the increase in plantation forest area in the central and northern regions, while the northeast region was the distribution area of the Pudatso National Forest Park (Shi et al., 2015; Su et al., 2020), which confirmed the reliability of the results of forest depression in the study area estimated by the GWR model. Moreover, Shangri-La is an alpine mountain area in the study area, which belongs to the national ecologically fragile area, and the logging of natural forests has been prohibited since 1998, and there has been no relevant management and logging activities in the past 20 years; hence, the results had certain credibility.

The prediction results of the GWR model were verified using data from 40 sample plots (20 × 20 m) from the field survey in the study area in November 2016 (Figure 10). It is verified that the R^2 between the FCC predicted value and the measured value is 0.62, and the Pearson correlation coefficient is 0.784 (significant at the 0.01 level), which has high consistency. It indicated that the method of estimating FCC in the study area by using the ATLAS footprint FCC value as the training sample data of the GWR model and cooperating with multi-source remote sensing factors was feasible, and the estimation results were reliable.

5 Discussion

5.1 Model error propagation and Bayesian optimization algorithm

Fu et al. (2014) and Qin et al. (2017) showed the uncertainty of aboveground biomass and carbon storage in forests, respectively. The uncertainty of the remote sensing model was the main error source that causes the uncertainty of biomass and carbon storage estimation, and the accuracy of the model plays an important role in the estimation results. In this study, the FCC value estimated by the



footprint scale is used as the training sample of the regional-scale GWR model. In the scale-up process, there is model error transmission. In order to weaken the influence of this error on the regional-scale FCC estimation results, the BOA is used to further optimize the three initial machine learning models, so as to optimize the optimal FCC estimation model of footprint scale and improve the estimation accuracy of the model. This study shows that after using BO to optimize the initial machine learning model, the accuracy of the model has been significantly improved compared with that before optimization. However, the FCC value of ATLAS footprints estimated by the BO-GBRT model was used as the dependent variable of the GWR model. Compared with before

TABLE 8 OLS covariance diagnosis table for the independent variable factors.

Variable name	NDVI	GNDVI	B8_SM	B8A_CR	VV-VH	Slope
VIF	6.70	6.88	1.01	1.09	1.05	1.09

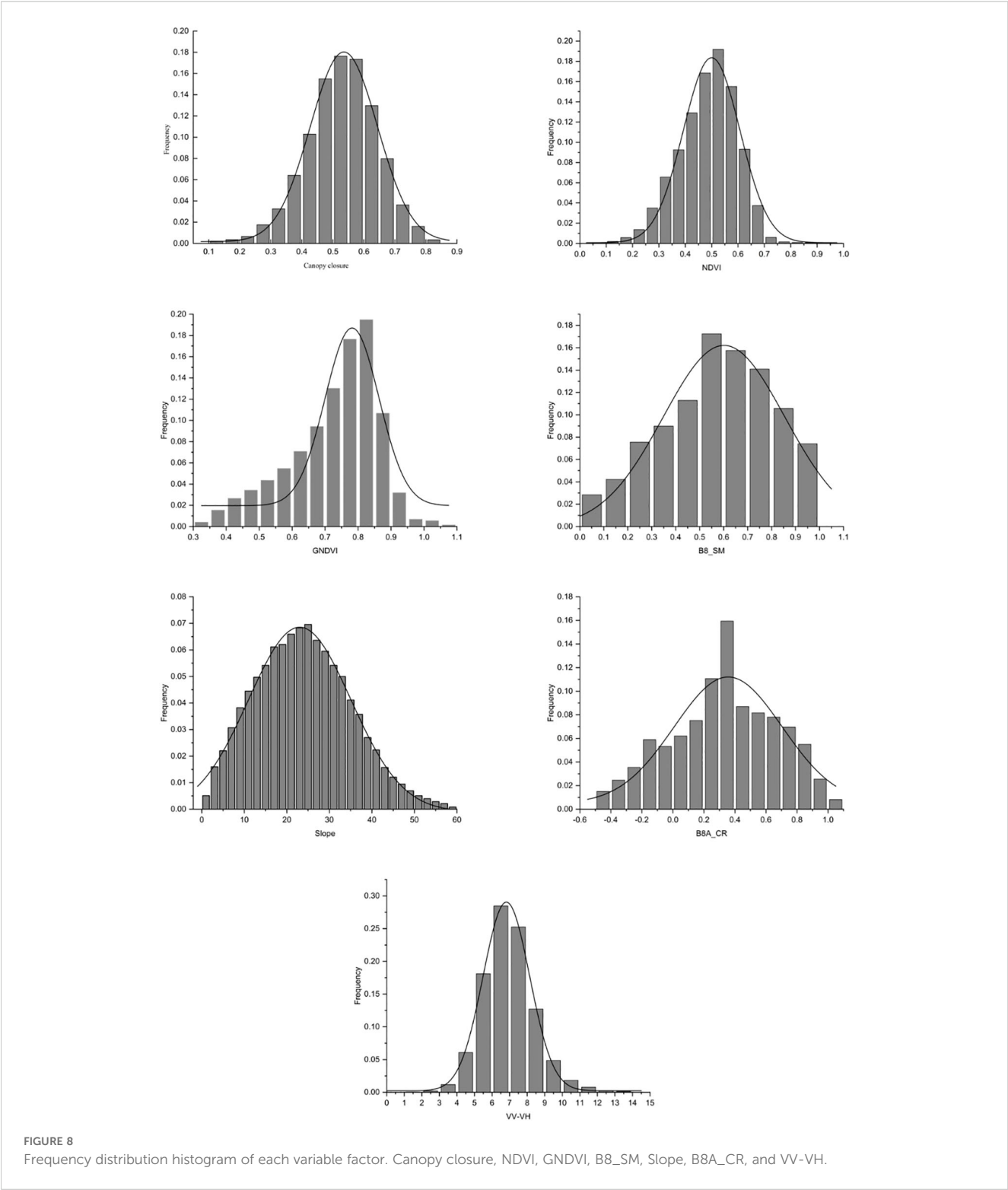


FIGURE 8 Frequency distribution histogram of each variable factor. Canopy closure, NDVI, GNDVI, B8_SM, Slope, B8A_CR, and VV-VH.

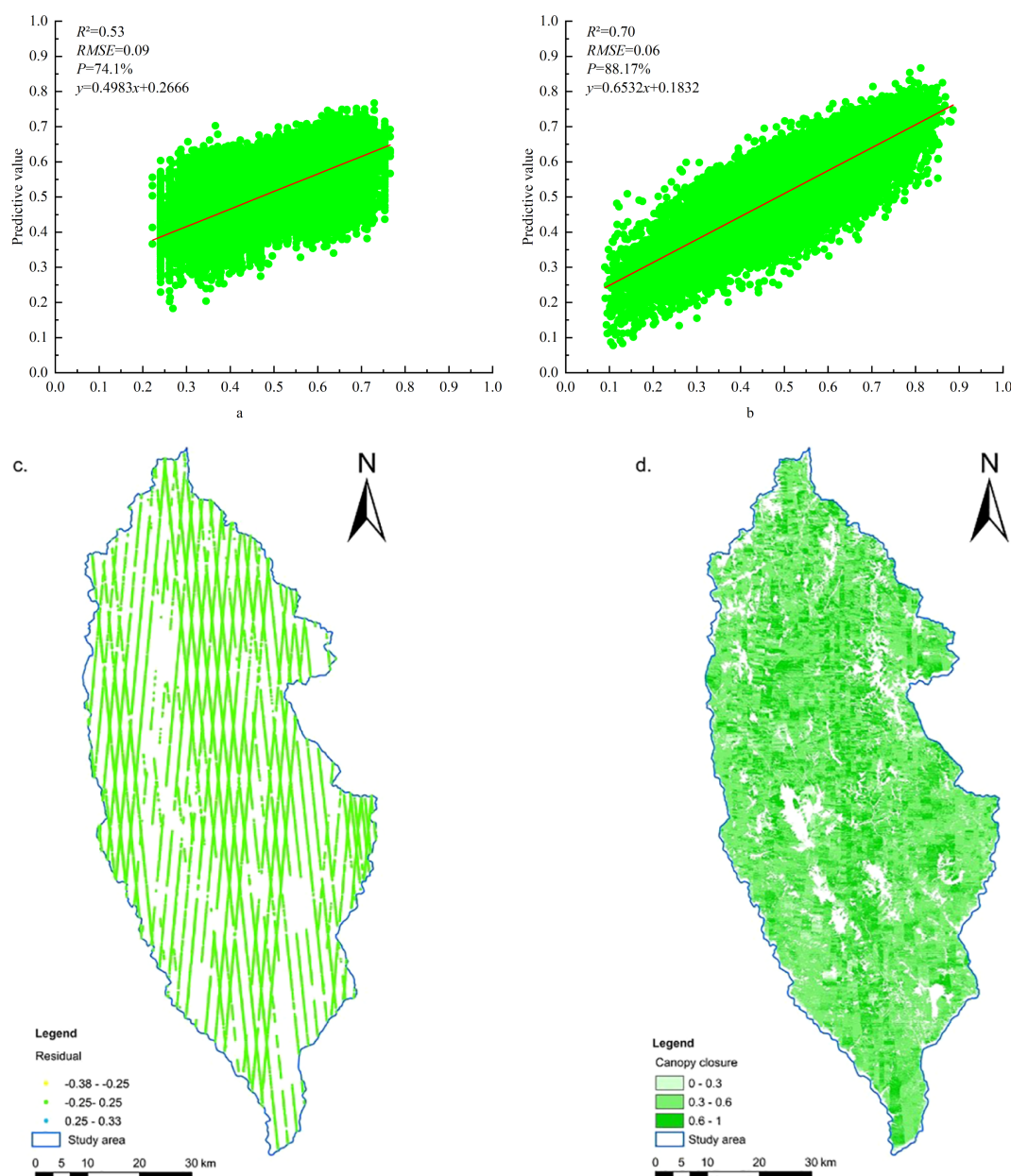


FIGURE 9

Accuracy of GWR model fit and estimation results. (a) Before optimization. (b) After optimization. (c) Residual distribution. (d) Spatial distribution of FCC in the study area.

optimization, R^2 increased by 32.08%, RMSE decreased by 33.33%, AICc decreased by 7.16%, and P increased by 14.07%. Therefore, BOA can effectively improve the accuracy of the estimation model and weaken the influence of model error transfer on the FCC estimation results. In this study, the model fitting accuracy is BO-GBRT > BO-RFR > BO-KNN. The reason is that the K-NN model was suitable for large samples because of non-assumptions on the data and non-sensitivity to abnormal samples (Shu et al., 2022), and the GBRT model was based on the iterative improvement of the original model, so that the next new model had a smaller error than the previous model, and a new combined model was established in the gradient direction of the residual reduction, which often has

higher fitting accuracy than RF (Zhang et al., 2019a; Yu et al., 2021). However, the BOA was only used to optimize the main parameters of the original model for 1,000 times. In order to further improve the estimation accuracy of the model, the comprehensive search algorithm can be introduced to optimize all the parameters of the model (Pedregosa et al., 2011), or introduce algorithms such as deep forest so that small sample data can also be fitted by neural network learning (Xia et al., 2022).

In the extrapolation model of FCC in the study area, the AICc value of the GWR model is too large before and after optimization as the sample size was too large, and the growth of the maximum likelihood estimation of the variance of the random error term may

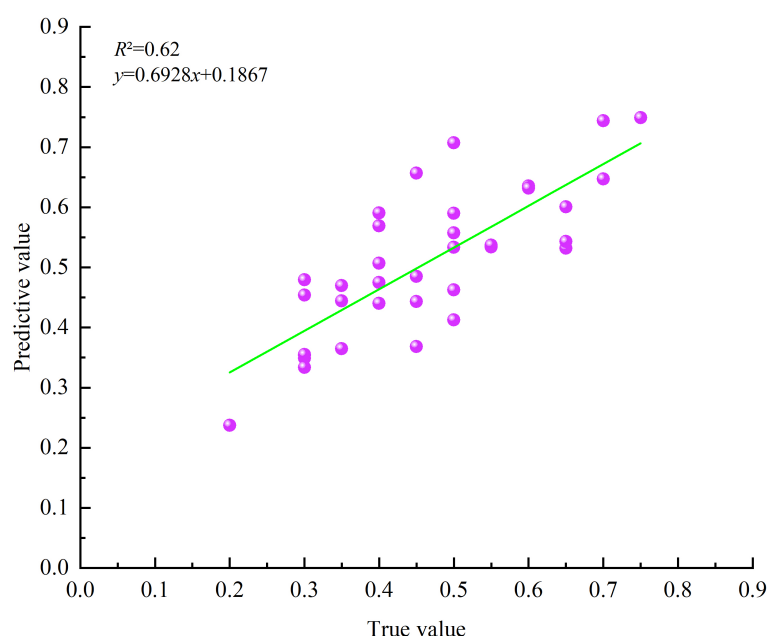


FIGURE 10
Linear model fitting diagram of measured value and predicted value.

slow down (Zhang et al., 2019b). The spatial distribution of the weight of the independent variable with the FCC will change due to the geographical location, showing the local spatial dependence and heterogeneity of the independent variable index; therefore, the GWR model can well combine the spatiality between the independent variable indicators to predict the canopy closure of the unknown space (Guo et al., 2012), specifically the canopy closure P that reached 88.27%, and the RMSE was 0.06.

5.2 Characteristic variable setting and selection

The optimization and combination of characteristic variable factors determine the accuracy of the prediction model and inversion results to a certain extent (Zhang et al., 2022). In the study, the more advanced photon-counting LiDAR data were used in the feature variable setting of the footprint scale (Lin et al., 2020; Zhu et al., 2020), and the spectral saturation point is higher. RF was used to select six characteristic variables with the largest contribution to construct the footprint-scale FCC estimation model. In the selection of regional-scale feature variables, optical remote sensing data are affected by the “light saturation” characteristics of forest vegetation to varying degrees. Among them, the single-band reflectance had the greatest impact, followed by the vegetation index, while the texture feature can represent ground object structure information in remote sensing images, reflecting the important information of spatial changes of land cover type in remote sensing images (Shu et al., 2022) and the forest structure information, which is least affected by “light saturation” (Zhao et al., 2016). In this paper, the integration of

multi-sensor and auxiliary data is realized by adding SAR factors and terrain factors to solve the problem of data saturation (Zhao et al., 2016). At the same time, considering the collinearity problem among the independent variables of the GWR model (Guo et al., 2012; Nazeer and Bilal, 2018; Zhang et al., 2019b; Song et al., 2022a), the same type of factors should not be too much, so the study only selected the sliding window of 5×5 texture feature factors. Finally, after correlation analysis, OLS test, and normal transformation, six explanatory variables were selected to construct the GWR mathematical model, which reduced the saturation problem of remote sensing data to a certain extent and improved the model estimation accuracy.

5.3 Transplantability of FCC estimation models in different study areas or forest types

In this paper, a process-oriented programming processing module was established for ATLAS data processing and parameter batch extraction, feature variable preference, optimization of the main parameters of the model, and optimization of multiple nonparametric models to fit the best model. Xie et al. (2022) used multispectral satellite images in Google Earth Engine to improve the estimation of FCC. The results showed that the calibration of model parameters needed to determine the range of values manually. The advantage of this study is that researchers only need to input ATLAS data and measured sample data for modeling, and select the best model according to their own needs to estimate all the FCC prediction values in the regional footprints. Based on the characteristics of ICESat-2/ATLAS

data (Neuenschwander and Pitts, 2019), full domain coverage of global regions can be basically achieved to meet the needs for flexible selection of different study areas and model portability testing. Zhu et al. (2020) analyzed the forest height based on different modes of spaceborne LiDAR data, and the results showed that the forest height consistency model established in different forest types or experimental area data was universal, and the model accuracy was consistent with the original model accuracy. The dataset used in the study was affected by complex terrain and high-altitude factors. However, the vegetation community in the study area was not distinguished and only relative sampling was carried out, and the sample size was 54, which had met the principle of large samples in field investigation (Shu et al., 2022). In future studies, the sample size of different slopes and different community compositions can be increased to explore their effects on the estimation results separately. At the same time, the same experimental design of different communities can be set up in low-altitude and relatively flat areas such as plains and hills to verify the portability of the model and obtain more accurate FCC estimation results. This provided not only a reliable reference for the study of energy transfer and microclimate changes in global forest ecosystems and forest tending evaluation but also a way of thinking for characterizing FCC maps on a global scale.

5.4 The influence of mountain terrain on FCC estimation results

The complex terrain area had a certain impact on the FCC estimation results: the larger the terrain slope, the greater the influence of the vegetation canopy on the laser echo. Compared with ICESat-1/GLAS data (Wang et al., 2015; Cui et al., 2021), the new-generation ICE-Sat-2/ATLAS data were used in this study; the footprint diameter is only 17 m and the footprint interval is 0.7 m, which greatly reduced the influence of terrain on footprint echo (Moudry et al., 2022) and improved the accuracy of model estimation. The data of 54 measured plots used in this study showed that the proportion of plots with a slope of 0°–10°, 10°–20°, and greater than 20° was 42.59%, 29.63%, and 27.78%, respectively, and the slope distribution was relatively uniform. Based on this, a forest ATLAS footprint FCC estimation model was constructed, in which the model had high verification accuracy ($R^2 = 0.65$, RMSE = 0.10, $P = 79.2\%$, MAR = 0.079) and can be used as a mountain FCC estimation model. This provides the possibility for footprint-scale low-cost ground plot survey to achieve accurate regional-scale FCC estimation.

6 Conclusions

In order to evaluate the ability of spaceborne photon-counting radar to estimate FCC, ICESat-2/ATLAS was used to obtain photon point cloud data. After denoising and classification of the photon point cloud, the parameter values (including 54 measured sample data) were extracted; meanwhile, six feature variables were selected

by RF, and the best footprint-scale FCC estimation model was constructed based on BO-RFR, BO-KNN, and BO-GBRT to obtain the FCC value of ATLAS footprints. Additionally, the training sample data of the GWR model and the continuous planar FCC products in the whole study area are predicted. The main conclusions are as follows:

1. After the pretreatment of ICESat-2/ATLAS data, the extracted parameters had an ideal way of estimating FCC. Among them, the BO-GBRT model had the best verification accuracy as the best footprint scale FCC estimation model ($R^2 = 0.65$, RMSE = 0.10, $P = 79.22\%$, MAR = 0.079). It showed that the BO algorithm can improve the model fitting accuracy, and the best estimation model compared with a variety of nonparametric models can reduce the influence of model error transfer on the FCC estimation results. The parameter with the largest contribution rate to the model was `landsat_perc` at 12.68%.
2. The ATLAS footprint FCC was used as the training sample of the regional-scale GWR model, and the FCC results were better in the study area with Sentinel 1/2 images and topographic factors. A total of 91 feature variables were extracted from DEM and multi-source remote sensing images for feature optimization, and 13 independent variables were retained. After OLS test and normal transformation, a total of six explanatory variables were involved in the mathematical model construction of GWR, and the final model verification accuracy was $R^2 = 0.70$, RMSE = 0.06, and $P = 88.27\%$. The VV–VH factor was calculated based on the difference between VV and VH, which had good adaptability to the model, and the correlation was -0.287 . The texture feature factors `B3_DI` and `B3_HO` had the strongest correlation with FCC.
3. The results of the study area estimated by the GWR model were used for spatial mapping, and the FCC distribution in the study area was consistent with the distribution of footprint-scale FCC. The R^2 between measured value and predicted value of the sample was 0.65, and the correlation coefficient was 0.784, which had high consistency. The average value of FCC was 0.50, and the values were mainly distributed between 0.3 and 0.6, accounting for 68.43%, followed by 0.6 to 1, accounting for 22.48%, and by 0 to 0.3, accounting for 9.09%. The high value area of FCC was distributed from northwest to southeast. The northern and southeastern regions were the main distribution areas of high and low FCC values, respectively, which is highly consistent with the forest distribution in the study area. The research showed that it is feasible to use ICESat-2/ATLAS data to predict the FCC value of ATLAS footprints based on the optimized machine learning method and to use such data as the training sample data of the GWR model, combined with multi-source remote sensing factors to estimate regional FCC. In summary, this provides a new scientific method for obtaining large-scale FCC with low cost and high precision.

Data availability statement

All satellite remote sensing data used in this study are publicly available and free of charge. ICESat-2/ATLAS data and ALOS can be obtained at <https://www.earthdata.nasa.gov> (accessed in November 2021), and Sentinel 1/2 data can be obtained at <https://developers.google.com/earth-engine/datasets> (accessed in November 2021). Further questioning can point to the corresponding author.

Author contributions

WZ: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. QS: Funding acquisition, Project administration, Supervision, Resources, Writing – review & editing. CX: Conceptualization, Data curation, Methodology, Visualization, Writing – review & editing. LX: Conceptualization, Data curation, Methodology, Visualization, Writing – review & editing. QX: Data curation, Validation, Visualization, Writing – review & editing. LF: Data curation, Validation, Visualization, Writing – review & editing. ZY: Conceptualization, Methodology, Software, Writing – original draft. SW: Conceptualization, Methodology, Software, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The work was supported by the Joint Agricultural Project of Yunnan Province (No. 202301BD070001-002), the National Natural Science Foundation of China (Nos. 31860205 and 31460194), and the Yunnan Province Education Department Scientific Research Fund Project (No. 2023Y0728), China, in 2023.

References

- Chen, G., Lou, T., Jing, W., and Wang, Z. (2019). Sparkpr: an efficient parallel inversion of forest canopy closure. *IEEE Access* 7, 135949–135956. doi: 10.1109/ACCESS.2019.2941966
- Chen, L., Ren, C., Bao, G., Zhang, B., Wang, Z., Liu, M., et al. (2022). Improved object-based estimation of forest aboveground biomass by integrating LiDAR data from GEDI and ICESat-2 with multi-sensor images in a heterogeneous mountainous region. *Remote Sens.* 14, 2743. doi: 10.3390/rs14122743
- Coulston, J. W., Blinn, C. E., Thomas, V. A., and Wynne, R. H. (2016). Approximating prediction uncertainty for random forest regression models. *Photogrammetric Eng. Remote Sens.* 82, 189–197. doi: 10.14358/PERS.82.3.189
- Cui, L., Jiao, Z., Zhao, K., Sun, M., Dong, Y., Yin, S., et al. (2021). Retrieving forest canopy elements clumping index using ICESat GLAS lidar data. *Remote Sens.* 13, 948. doi: 10.3390/rs13050948
- Cui, J., and Yang, B. (2018). Survey on Bayesian optimization methodology and applications. *J. Software* 29, 3068–3090. doi: 10.13328/j.cnki.jos.005607
- Duan, Z., Wu, L., and Jiang, X. (2023). Effect of point cloud density on forest remote sensing retrieval index extraction based on unmanned aerial vehicle LiDAR data. *Geomatics Inf. Sci. Wuhan Univ.* 48, 1923–1930. doi: 10.13203/j.whugis20210719
- Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., et al. (2020). Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sens. Environ.* 242, 111779. doi: 10.1016/j.rse.2020.111779
- Franco-Lopez, H., Ek, A. R., and Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens. Environ.* 77, 251–274. doi: 10.1016/S0034-4257(01)00209-7
- Fu, Y., Lei, Y., and Zeng, W. (2014). Uncertainty assessment in regional-scale above ground biomass estimation of Chinese fir. *Scientia Silvae Sinicae*.
- Glenn, N. F., Neuenschwander, A., Vierling, L. A., Spaete, L., Li, A., Shinneman, D. J., et al. (2016). Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sens. Environ.* 185, 233–242. doi: 10.1016/j.rse.2016.02.039
- Guo, L., Zhang, H., Chen, J., Li, R., and Qin, C. (2012). Comparison between CO-Kriging model and geographically weighted regression model in spatial prediction of soil attributes. *Acta Pedologica Sin.* 49, 1037–1042.
- Hua, Y., and Zhao, X. (2021). Multi-model estimation of forest canopy closure by using red edge bands based on sentinel-2 images. *Forests* 12, 1768. doi: 10.3390/f12121768
- Lee, A. C., and Lucas, R. M. (2007). A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests. *Remote Sens. Environ.* 111, 493–518. doi: 10.1016/j.rse.2007.04.018

Acknowledgments

All authors thank NASA NSIDC for the release of ICESat-2/ATLAS data and ALOS data (<https://search.earthdata.nasa.gov>, visited in November 2021), as well as the Copernicus series of satellite data released by the European Space Agency (ESA) (<https://developers.google.com/earth-engine/datasets>, visited in November 2021), and thank the reviewers and members of the editorial team for their constructive comments. We thank the reviewers and guest editors for their valuable comments and suggestions, which significantly improved the quality of the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, J., and Mao, X. (2020). Comparison of canopy closure estimation of plantations using parametric, semi-parametric, and non-parametric models based on GF-1 remote sensing images. *Forests* 11, 597. doi: 10.3390/f11050597
- Lin, X., Xu, M., Cao, C., Dang, Y., Bashir, B., Xie, B., et al. (2020). Estimates of forest canopy height using a combination of ICESat-2/ATLAS data and stereo-photogrammetry. *Remote Sens.* 12, 3649. doi: 10.3390/rs12213649
- Meng, X. (2006). *Forest Mensuration* (Beijing: China Forestry Publishing House).
- Moudrý, V., Gdulová, K., Gábor, L., Šárová, E., Barták, V., Leroy, F., et al. (2022). Effects of environmental conditions on ICESat-2 terrain and canopy heights retrievals in Central European mountains. *Remote Sens. Environ.* 279, 113112. doi: 10.1016/j.rse.2022.113112
- Narine, L. L., Popescu, S., Neuenschwander, A., Zhou, T., Srinivasan, S., and Harbeck, K. (2019). Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. *Remote Sens. Environ.* 224, 1–11. doi: 10.1016/j.rse.2019.01.037
- Nazeer, M., and Bilal, M. (2018). Evaluation of ordinary least square (OLS) and geographically weighted regression (GWR) for water quality monitoring: A case study for the estimation of salinity. *J. Ocean Univ. China* 17, 305–310. doi: 10.1007/s11802-018-3380-6
- Neuenschwander, A., and Pitts, K. (2019). The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sens. Environ.* 221, 247–259. doi: 10.1016/j.rse.2018.11.005
- Neumann, M., Ferro-Famil, L., and Reigber, A. (2009). Estimation of forest structure, ground, and canopy layer characteristics from multibaseline polarimetric interferometric SAR data. *IEEE Trans. Geosci. Remote Sens.* 48, 1086–1104. doi: 10.1109/TGRS.2009.2031101
- Neumann, T. A., Martino, A. J., Markus, T., Bae, S., Bock, M. R., Brenner, A. C., et al. (2019). The Ice, Cloud, and Land Elevation Satellite-2 Mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. *Remote Sens. Environ.* 233, 111325. doi: 10.1016/j.rse.2019.111325
- Nie, S., Wang, C., Dong, P., Xi, X., Luo, S., and Qin, H. (2017). A revised progressive TIN densification for filtering airborne LiDAR data. *Measurement* 104, 70–77. doi: 10.1016/j.measurement.2017.03.007
- Nie, S., Wang, C., Xi, X., Luo, S., Li, G., Tian, J., et al. (2018). Estimating the vegetation canopy height using micro-pulse photon-counting LiDAR data. *Optics express* 26, A520–A540. doi: 10.1364/OE.26.00A520
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pu, Y., Xu, D., Wang, H., An, D., and Xu, X. (2021). Extracting canopy closure by the CHM-based and SHP-based methods with a hemispherical FOV from UAV-LiDAR data in a poplar plantation. *Remote Sens.* 13, 3837. doi: 10.3390/rs13193837
- Qin, L., Zhang, M., Zhong, S., and Yu, X. (2017). Model uncertainty in forest biomass estimation. *Acta Ecologica Sin.* 37, 7912–7919. Available online at: <https://link.cnki.net/urlid/11.2031.Q.20170814.1502.032> (Accessed August 14, 2017).
- Shi, L., Zhao, H., Li, Y., Ma, H., Yang, S., Wang, H., et al. (2015). Evaluation of Shangri-La County's tourism resources and ecotourism carrying capacity. *Int. J. Sustain. Dev.* 22, 103–109. doi: 10.1080/13504509.2014.927018
- Shu, Q., Xi, L., Wang, K., Xie, F., Pang, Y., and Song, H. (2022). Optimization of samples for remote sensing estimation of forest aboveground biomass at the regional scale. *Remote Sens.* 14, 4187. doi: 10.3390/rs14174187
- Song, X., Mi, N., Mi, W., and Li, L. (2022c). Spatial non-stationary characteristics between grass yield and its influencing factors in the Ningxia temperate grasslands based on a mixed geographically weighted regression model. *J. Geographical Sci.* 32, 1076–1102. doi: 10.1007/s11442-022-1986-5
- Song, H., Shu, Q., Xi, L., Qiu, S., Wei, Z., and Yang, Z. (2022a). Remote sensing estimation of forest above-ground biomass based on spaceborne lidar ICESat-2/ATLAS data. *Trans. Chin. Soc. Agric. Eng.* 38, 191–199.
- Song, H., Xi, L., Shu, Q., Wei, Z., and Qiu, S. (2022b). Estimate forest aboveground biomass of mountain by ICESat-2/ATLAS data interacting cokriging. *Forests* 14, 13. doi: 10.3390/f14010013
- Su, T., Spicer, R. A., Wu, F.-X., Farnsworth, A., Huang, J., Del Rio, C., et al. (2020). A Middle Eocene lowland subtropical “Shangri-La” ecosystem in central Tibet. *Proc. Natl. Acad. Sci.* 117, 32989–32995. doi: 10.1073/pnas.2012647117
- Tian, Y., Huang, H., Zhou, G., Zhang, Q., Tao, J., Zhang, Y., et al. (2021). Aboveground mangrove biomass estimation in Beibu Gulf using machine learning and UAV remote sensing. *Sci. Total Environ.* 781, 146816. doi: 10.1016/j.scitotenv.2021.146816
- Varghese, A., and Joshi, A. (2015). Polarimetric classification of C-band SAR data for forest density characterization. *Curr. Sci.* 108, 100–106. Available online at: <https://www.jstor.org/stable/24216181> (Accessed January 10, 2015).
- Wang, R., Xing, Y., Wang, L., You, H., Qiu, S., and Wang, A. (2015). Estimating forest canopy cover by combining spaceborne ICESat-GLAS waveforms and multispectral Landsat-TM images. *Chin. J. Appl. Ecol.* 26, 1657–1664. doi: 10.13287/j.1001-9332.20150331.005
- Xi, L., Shu, Q., Sun, Y., Huang, J., and Song, H. (2023). Optimizing an ICESat2-based remote sensing estimation model for the leaf area index of mountain forests in southwestern China. *Remote Sens. Natural Resour.* 35, 160–169.
- Xia, H., Tang, J., and Qiao, J. (2022). Review of deep forest. *J. Beijing Univ. Technol.* 48, 182–196.
- Xie, B., Cao, C., Xu, M., Yang, X., Duerler, R. S., Bashir, B., et al. (2022). Improved forest canopy closure estimation using multispectral satellite imagery within Google Earth Engine. *Remote Sens.* 14, 2051. doi: 10.3390/rs14092051
- Xu, E., Shu, Q., Sun, Y., Chen, E., Li, Z., Zhao, L., and Liu, Q. (2022). An estimation model for regional forest canopy closure combined with UAV LiDAR and high spatial resolution satellite remote sensing data. *Geomatics Inf. Sci. Wuhan Univ.* 47, 1298–1308. doi: 10.13203/j.whugis.20210001
- Yang, X., He, P., Yu, Y., and Fan, W. (2022). Stand canopy closure estimation in planted forests using a Geometric-Optical Model based on remote sensing. *Remote Sens.* 14, 1983. doi: 10.3390/rs14091983
- Yu, J., Lai, H., Xu, L., Luo, S., Zhou, W., Song, H., et al. (2024). Estimation of forest canopy cover by combining ICESat-2/ATLAS data and geostatistical method/cokriging. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 17, 1824–1838. doi: 10.1109/JSTARS.2023.3340429
- Yu, R., Yao, Y., Wang, Q., Wan, H., Xie, Z., Tang, W., et al. (2021). Satellite-derived estimation of grassland aboveground biomass in the three-river headwaters region of China during 1982–2018. *Remote Sens.* 13, 2993. doi: 10.3390/rs13152993
- Zhang, L., Cheng, J., Jin, C., and Zhou, H. (2019b). A multiscale flow-focused geographically weighted regression modelling approach and its application for transport flows on expressways. *Appl. Sci.* 9, 4673. doi: 10.3390/app9214673
- Zhang, J., Lu, C., Xu, H., and Wang, G. (2019a). Estimating aboveground biomass of Pinus densata-dominated forests using Landsat time series and permanent sample plot data. *J. Forestry Res.* 30, 1689–1706. doi: 10.1007/s11676-018-0713-7
- Zhang, W., Tang, L., Chen, F., and Yang, J. (2021b). Prediction for TBM penetration rate using four hyperparameter optimization methods and random forest model. *J. Basic Sci. Engineering.* 29, 1186–1200. doi: 10.16058/j.issn.1005-0930.2021.05.009
- Zhang, L., Zeng, Y., Zhuang, R., Szabó, B., Manfreda, S., Han, Q., et al. (2021a). *In situ* observation-constrained global surface soil moisture using random forest model. *Remote Sens.* 13, 4893. doi: 10.3390/rs13234893
- Zhang, R., Li, Z., Pang, Y., and Bao, Y. (2016). Canopy closure estimation in a temperate forest using airborne LiDAR and LANDSAT ETM+ data. *Chinese J. Plant Ecol.* 40, 102–115. doi: 10.17521/cjpe.2014.0366
- Zhang, W., Zhao, L., Li, Y., Shi, J., Yan, M., and Ji, Y. (2022). Forest above-ground biomass inversion using optical and SAR images based on a multi-step feature optimized inversion model. *Remote Sens.* 14, 1608. doi: 10.3390/rs14071608
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., and Yu, S. (2016). Examining spectral reflectance saturation in Landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sens.* 8, 469. doi: 10.3390/rs060469
- Zhou, W., Shu, Q., Wang, S., Yang, Z., Luo, S., Xu, L., et al. (2023). Estimation of forest canopy closure in northwest Yunnan based on multi-source remote sensing data collaboration. *Chin. J. Appl. Ecol.* 34, 1806–1816. doi: 10.13287/j.1001-9332.202307.021
- Zhou, W., Shu, Q., Xu, L., Yang, Z., Gao, Y., Wu, Z., et al. (2024). Construction of forest canopy closure estimation model in the northwestern Yunnan based on global ecosystem dynamics investigation multi-beam LiDAR data. *Acta Ecologica Sin.* 44, 3525–3539. doi: 10.20103/j.stxb.202309212048
- Zhu, X., Nie, S., Wang, C., Xi, X., Li, D., Li, G., et al. (2020). Estimating terrain slope from ICESat-2 data in forest environments. *Remote Sens.* 12, 3300. doi: 10.3390/rs12203300