

Regional scale inversion of chlorophyll content of *Dendrocalamus giganteus* by multi-source remote sensing

Cuifen Xia¹, Wenwu Zhou¹, Qingtai Shu^{1,*}, Zaikun Wu¹, Li Xu¹, Hunfen Yang¹, Zhen Qin¹, Mingxing Wang¹ and Dandan Duan²

(1. College of Forestry, Southwest Forestry University, Kunming 650224, China; 2. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences; Beijing 100097, China)

Abstract: **【Objective】** The spectrophotometer method is costly, time-consuming, laborious, and destructive to the plant. It will be lost during the transportation process, and can only obtain sample point data, which is a challenge to the estimation of chlorophyll content at the regional level. **【Method】** In this study, in order to improve the estimation accuracy, a new method of collaborative inversion of chlorophyll using Landsat 8 and Global Ecosystem Dynamics Investigation (GEDI) is proposed. Specifically, the chlorophyll content data set is combined with the preprocessed two remote sensing (RS) factors to construct three regression models of support vector machine (SVM), BP neural network (BP) and random forest (RF), and the better model is selected for inversion. In addition, the Kriging method is used to interpolate the GEDI point attribute data into the surface attribute data for modeling. **【Result】** The results showed that: (1) The chlorophyll model of single plant was $y = 0.1373x^{1.7654}$. (2) The optimal semi-variance function models of pai, pgap_theta and pgap_theta_a3 are exponential models. (3) The top three correlations between the two RS data and chlorophyll content were B2_3_SM, B2_3_HO, B2_5_EN and pai, pgap_theta, pgap_theta_a3. (4) The combination of Landsat 8 image and GEDI had the highest modeling accuracy, and RF had the best performance, with R^2 , RMSE and P of 0.94, 0.18 g/m² and 83.32%, respectively. **【Conclusion】** This study shows that it is reliable to use Landsat 8 images and GEDI to retrieve the chlorophyll content of *Dendrocalamus giganteus*, and reveals the potential of multi-source RS data in the inversion of forest ecological parameters.

Keywords: RS; Machine learning; Chlorophyll content; *Dendrocalamus giganteus*; Inver-sion

中图分类号 文献标识码: 文章编号:

Chlorophyll is an important biochemical index of plants. Chlorophyll content can directly reflect the photosynthetic capacity, nitrogen fixation capacity, and health status of vegetation, which is an important index to evaluate the growth and stand yield of forest [0]. Accurately estimating the forest chlorophyll content is helpful to further understand the forest ecosystem functions and the forest health status. Bamboo forest is very important in maintaining the forest ecosystem balance, mitigating global warming, promoting carbon sequestration [0]. Among the bamboo forests, *Dendrocalamus giganteus* is one of the largest bamboos in the world. Therefore, the chlorophyll content retrieval of *D. giganteus* at regional scale is the key to grasp the growth status and to ecosystem functions of *D. giganteus*. Traditionally, spectrophotometer is used to measure chlorophyll content at laboratory, which destructs plants and losses chlorophyll content during transportation [0]. Moreover, the traditional method operates at specific leaf scales with sample point data, which is limited in forestry application. Nowadays, non-destructive, real-time, efficient and accurate RS technology, can greatly enhance the efficiency and improve the precision of vegetation chlorophyll content retrieval at regional scale by combining with ground survey data. Meanwhile, numerous studies have shown that using RS technology combined with ground survey data to estimate vegetation chlorophyll content at a regional scale is both effective and feasible [0-6].

At present, the most used for chlorophyll content inversion is optical RS data (such as the Landsat series of the United States; [0-9], while there are few studies on the inversion of chlorophyll content with spaceborne LiDAR GEDI. The reflectance and spectral index of the optical sensor are sensitive to the horizontal structure of chlorophyll content. Chlorophyll mainly absorbs red light and blue-violet light, and the maximum absorption wavelength range is 420 nm ~ 663 nm, which has a strong reflection effect on green light [0-11]. However, the acquisition of Landsat

8 RS images is susceptible to climate and saturation problems [0-14], and previous studies have mainly focused on extracting spectral features from RS images [0-16], and exploring the relationship between vegetation index and plant chlorophyll a concentration, ignoring texture features that are conducive to improving interpretation accuracy [0]. Based on multi-spectral RS satellite Landsat 5 and Landsat 8 data, Yang Y et al. [0] used RF regression to construct an inversion model of texture features and spectral indices with lake chlorophyll a concentration. The research shows that texture features have a significant correlation with lake chlorophyll a concentration. At the same time, it also reflects that RS image texture features have good potential in estimating biochemical indicators. Compared with ICESat-2 (Ice Cloud Land Elevation Satellite-2), GEDI has the advantage of wide coverage. Compared with the above RS satellites, the advantage is that the emitted laser beam has a certain penetration ability to the forest canopy. The forest canopy's three-dimensional structure and the underlying topography are determined using echo waveform data. The deficiency is that the space-borne LiDAR is difficult to obtain the horizontal structure parameters of the forest, and due to the discreteness of the data, the forest information is incomplete. At present, GEDI data are mostly used in the inversion of forest canopy height, leaf area index, biomass, etc. Xu L et al. [0] used GEDI spaceborne LiDAR data to estimate the biomass of oak forests in Shangri-La. The research shows that modis_treecover, rv, sensitivity and oak tree biomass have obvious correlation, and the oak forests biomass estimation model established by random forest has the best accuracy. At the same time, it shows that the series of indicators contained in GEDI L2B product data can also show good model interpretation accuracy with forest structure parameters and vegetation biochemical parameters.

From the perspective of domestic and foreign research, it is rare to use optical RS data and spaceborne LiDAR data to estimate the chlorophyll content of bamboo. Therefore, in this study, Landsat 8 data and spaceborne LiDAR GEDI data were used to work together. Using the unique optical properties of chlorophyll, the statistical relationship between chlorophyll content and RS characteristic bands, as well as the statistical relationship between GEDI L2B product data indicators and chlorophyll content were analyzed. The inversion model was established by RF to improve the inversion accuracy of chlorophyll content.

Currently, the inversion methods of chlorophyll content based on RS data mainly involve empirical models, physical models, and coupling models. The empirical model method (including parametric model and non-parametric model) is convenient, fast, easy to operate, high efficiency and ideal accuracy, but the optimization of feature combination needs further study [11,0]. The coupling model is to couple the empirical model and the physical model, which can maximize the advantages of the statistical model, but the operation is time-consuming and inefficient [0]. Machine learning algorithms, as a novel modeling approach, are not constrained by a fixed model framework. They have the capability to iteratively learn from feedback errors during the model correction process, enhancing the understanding of the intricate relationship between independent and dependent variables. [0]. The chlorophyll content estimation based on machine learning algorithm can be divided into the following two processes. (1) Analysis the relationships between chlorophyll concentration and characteristic variables. (2) Calculate the chlorophyll concentration by using the function relationship [0]. The machine learning algorithm usually shows good chlorophyll concentration inversion effect because it can solve high-dimensional nonlinear problems [0-23]. For example, models such as RF [0], neural network [0], and genetic algorithm-optimized simplified support vector

machine (GA-SVM) have been proven to perform well in chlorophyll a estimation [0]. RF is one of the most popular machine learning methods, and its model performs better than other machine learning methods in forest biochemical parameter estimation. [0-27]. Although there have been many studies on RS monitoring of chlorophyll concentration using machine learning algorithms [0], a general method has not been proposed to achieve long-term monitoring of chlorophyll concentration, and the inversion model has poor universality [0]. However, so far, based on the cooperative operation of RS satellite and spaceborne LiDAR, and the use of more mature RF algorithm to estimate the chlorophyll content of *D. giganteus* is rare in forestry applications. Real-time, fast and accurate inversion of chlorophyll content has become one of the urgent problems for forestry researchers.

Therefore, the goal of this study is to use Xinning County, Yunnan Province, which has a large number of large clumped *D. giganteus*, as the primary test area, to use machine learning technology to estimate chlorophyll content, and to evaluate the potential of multi-source RS data collaboration in chlorophyll content inversion. The specific objectives of this study are: (1) To derive a model to retrieve chlorophyll content from a single *D. giganteus* plant. (2) An optimal model for inverting the chlorophyll content of *D. giganteus* was established. (3) To create a distribution map of *D. giganteus* plants in the study area, utilizing the optimal attributes derived from multispectral Landsat 8 and LiDAR GEDI satellite data. The feasibility of estimating the chlorophyll content of *D. giganteus* by multi-RS data collaborative operation was evaluated, which provided a feasible reference for the inversion of chlorophyll content of *D. giganteus* at medium and large regional scales.

1 Materials and Methods

1.1 Study Area

Xinning Yi and Dai Autonomous County (Xinning County, 23°38'15"-24°26'05" N, 101°16'30"-102°16'50" E) is located at the eastern foot of the middle of Ailao Mountain (Fig.1), belonging to Yuxi City in Yunnan Province [0]. The main terrain is mountainous, and the terrain is high in the northwest and low in the southeast. The highest altitude in the territory is 3165.9 m and the lowest is 422 m. Due to the influence of altitude difference, Xinning County has formed three climate types: dry-hot valley high temperature area, semi-mountain warm temperature area and alpine cold temperature area [0]. The annual precipitation is 869 mm, the annual maximum temperature and minimum temperature are 32.8 °C and 1.3 °C respectively, and the annual average temperature is 18.1 °C. It belongs to the subtropical low-latitude plateau monsoon climate zone. It is suitable for growing in an area with an altitude of 300 ~ 1200 m, and its daily average surface temperature is between 18 °C ~ 26 °C [0]. The forest coverage rate of Xinning County reached 61.99%, and the forest land area reached 3.18×10^5 hm². In the forest land, the arbor forest land is 2.78×10^5 hm², accounting for 87.47%; 1.49×10^4 hm² of bamboo forest land, accounting for 4.67%; 2.27×10^4 hm² of shrub land, accounting for 7.14%; other forest land is 2.29×10^3 hm², accounting for 0.72%. *D. giganteus* is one of the largest bamboos in the world. Yunnan is one of the main distribution areas of *D. giganteus*. It belongs to the genus of *Poaceae* [0], the average *D. giganteus* height is about 30 m, the diameter at breast height (DBH) is about 15 m, and the wall thickness is about 14 mm. The area of bamboo forest in China is about 7.02×10^6 hm². Therefore, accurate estimation of chlorophyll content of *D. giganteus* at regional scale can help monitor vegetation and ecological environment changes, and promote ecological protection, resource management and disaster monitoring. In addition, it can provide a scientific basis for local governments to guide forestry production and land planning and promote sustainable development.

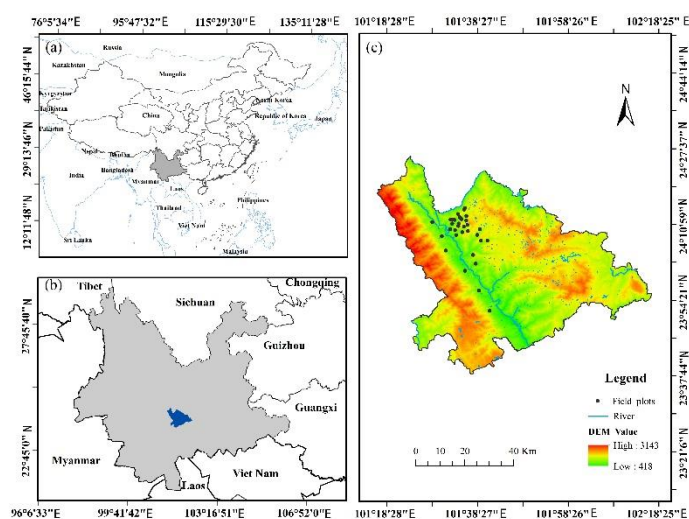


Figure. 1 Location map of the study area (Xinping County) in China (a) and in Yunnan Province (b). The digital elevation model (DEM) and location of field sampling plots in the Xinping County are shown in (c).

1.2 Acquisition and processing of measured data

The 35 measured data of chlorophyll content of *D. giganteus* used in this study were obtained from the standard sample survey of 30 m × 30 m (about 0.09 km²) in Jiasa Town, Shuitang Town and Laochang Township in Xinping County (Fig. 1). Since May to October of each year is the rainy season in Yunnan, November to April of the following year is the dry season. Among them, the rainfall from mid-June to mid-August is the most, accounting for about 60% of the annual rainfall. During this period, the rainfall is abundant and the climate is suitable. It is the main growth period of *D. giganteus*, and it is also the most representative period to reflect the change of chlorophyll content of *D. giganteus*. Considering the comprehensive factors such as weather, road safety, and the representativeness and typicality of the field sample setting, after screening, it was found that the weather was sunny from July 11 to July 16. Therefore, we finally chose July 12 to July 14. The field survey was carried out in the study area and the experimental data were collected. The DBH, coordinates and other factors were measured, and the materials required for the experiment (including standard *D. giganteus* and standard leaves) were collected. Among them, the diameter of the DBH is set to 5cm, and the *D. giganteus* with a DBH of 10cm is used as the standard plant. The coordinates of each square sample plot were measured by the differential locator of the southern surveying and mapping Real-Time Kinematic (RTK) in the fixed solution state, and the latitude, longitude and altitude information of the sample center were recorded after the error was less than 2 cm.

1.2.1 Subsubsection

In the 35 sample plots within the distribution range of *D. giganteus*, 137 single plants of *D. giganteus* with good growth and no pests, diseases and mechanical damage were randomly selected as standard sample *D. giganteus* (about 4 plants were selected on average in each sample plot). The selected standard *D. giganteus* was cut from the base, and its DBH was measured. The collected 137 leaf preservations were brought back to the laboratory for weighing. The measurement index of individual *D. giganteus* is shown in table 1.

Table 1. The statistics of DBH and leaf weight of single *D. giganteus* plant.

Name	Sample Size	Minimu m	Maximu m	Avera ge	SD
DBH (cm)	137	3.2	12.8	8.5	2.37
Total leaf fresh weight (kg)	137	0.05	6.54	1.90	1.34

1.2.2 Standard leaf selection and sample determination

49 strains were randomly selected from 137 standard sample *D. giganteus*, and different sizes of new, old and

young leaves were selected from the upper, middle and lower parts as standard sample leaves. The surface of the sampling leaves was wiped clean, and the upper, middle and lower parts on both sides of the middle vein (remove the middle vein) were evenly sampled, weighed, cut into pieces, and added liquid nitrogen to grind to powder. Add 80% C₃H₆O, grind to homogenate, filter, and constant volume. The chloroplast pigment extract was poured into a colorimetric dish with a light diameter of 1cm. With 80% C₃H₆O as blank control, the absorbance was measured by spectrophotometer at 663 nm and 645 nm, respectively. The statistics of chlorophyll content in leaves of *D. giganteus* is shown in Table 2. The specific method of chlorophyll determination in standard leaves of *D. giganteus* is based on the principle and technology of plant physiological and biochemical experiments [0]. The calculation formula is as follows:

$$C_a = 12.72A_{663} - 2.59A_{645} \quad (1)$$

$$C_b = 22.88A_{645} - 4.67A_{663} \quad (2)$$

$$C_T = C_a + C_b = 20.29A_{645} + 8.05A_{663} \quad (3)$$

where, A_{663} and A_{645} are the absorbance of chlorophyll solution at 663 nm and 645 nm, C_a , C_b and C_T are chlorophyll a, chlorophyll b and total chlorophyll content, respectively, in mg/g.

Table 2. The statistics of chlorophyll content of bamboo leaves.

Chl(mg/g)	Sample Size	Minimum	Maximum	Average	SD
C_a	49	1.78	2.67	2.36	0.26
C_b	49	0.35	2.73	1.27	0.54
C_T	49	2.13	5.34	3.63	0.78

1.2.3 Measurement of chlorophyll of *D. giganteus* at plot scale

The chlorophyll content of 137 individual plants of *D. giganteus* was used as the dependent variable and DBH as the independent variable to establish a power function relationship between the two. Based on the allometric growth equation, the chlorophyll-DBH model was constructed as the basic model of chlorophyll of *D. giganteus*, and then the chlorophyll content of *D. giganteus* in 35 plots was calculated by this model (Table 3).

Table 3. The statistics of chlorophyll content of *D. giganteus* in sample plot.

Name	Sample Size	Minimum	Maximum	Average	SD
Chl-ab (g/m ²)	35	0.16	1.21	0.51	0.27

1.3 RS image data acquisition and information extraction

1.3.1 Landsat 8 OLI image data

In February 2013, NASA launched the Landsat 8 satellite, which is equipped with a land imager (OLI) and a thermal infrared sensor (TIRS). It is the most widely used civilian satellite in the world. Landsat 8 OLI has a total of 11 bands, of which 1-7, 9-OLI is a multi-spectral band (30 m), the 8-OLI is a panchromatic band (15 m), the 10, 11-TIRS is a thermal infrared band (100 m). In this study, visible light band (B2~B4) and near-infrared band (B5) were used for research. The Landsat 8 OLI image (September 24, 2021, orbit No.130/43) was obtained from the geospatial data cloud (<http://www.gscloud.cn/search>) covered the entire study area. The RS image processing software ENVI 5.6 was utilized to preprocess the satellite image by performing radiometric calibration, atmospheric correction using FLAASH, and geometric correction. Subsequently, 73 RS factors were extracted [0]. Three topographic factors were extracted from DEM data [0], a total of 76 characteristic variable factors, which were used to construct the initial characteristic variable set at the regional scale (Table 4). Among them, this study uses the gray level (GLCM) co-occurrence matrix method to extract texture features, and starts the probability statistics tool to obtain ME, VA, HO, CO, DI, EN, SM, CR. The processing window sizes are 3 × 3 and 5 × 5, respectively. The

step size is 1×1 , and the gray-scale quantization level compression level is 64.

Table 4. Summary of RS factor information extraction from Landsat 8 OLI data.

Variables	Amount	Description
ρ_{B_i}	4	The original incidence of the i -th band ($i = 2, 3, 4, 5$)
DVI	1	Difference vegetation index: $DVI = \rho_{NIR} - \rho_R$, ρ_{NIR} , ρ_R the reflectance of the near-infrared band and the red band respectively.
RVI	1	Ratio vegetation index: $RVI = \rho_{NIR} / \rho_R$
NDVI	1	Normalized vegetation index: $NDVI = (\rho_{NIR} - \rho_R) / (\rho_{NIR} + \rho_R)$
SAVI	1	Soil Adjusted Vegetation Index: $SAVI = \frac{1.5(\rho_{NIR} - \rho_R)}{\rho_{NIR} + \rho_R + 0.5}$
EVI	1	Enhanced vegetation index $EVI = 2.5 \left[\frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + 6.0\rho_R - 7.5\rho_B + 1)} \right]$, ρ_B is the reflectivity of the blue band.
$B_{i_N_T}$	64	Texture features is the texture filtering T under the i -band $N \times N$ window. $i = 2, 3, 4, 5$; $N = 3, 5$; T is texture filtering. It is divided into: mean ME, variance VA, synergy HO, contrast CO, dissimilarity DI, information entropy EN, second moment SM, correlation CR.
Elevation	1	Elevation
Slope	1	Slope factor extracted by DEM
Aspect	1	Slope aspect factor extracted by DEM

1.3.2 GEDI data

GEDI was successfully launched by the International Space Station (ISS) of the United States on December 5, 2018, sampling and collecting data between 51.6° north and south latitudes around the world. The GEDI system consists of three lasers, generating a total of eight beam ground cross sections, including about 25 m footprint samples, with an interval of about 60 m along the orbit. The GEDI beam's cross section is spaced approximately 600 m apart along the Earth's surface orbit, with a transverse channel width of around 4.2 km. The GEDI consists of four product-level datasets. L1 provides geographically located return energy waveform data, while L2 includes geographically located surface elevation and canopy height information. Additionally, L3 offers gridded vegetation structure details, and L4 consists of footprint-level and gridded aboveground biomass data.

This study uses GEDI-derived L2B data to have a richer source of information than L2A data, such as: `pai`, `leaf_on_doy`, `cover`, etc. The GEDI data in this study were downloaded from the Earthdata website (<https://search.earthdata.nasa.gov/>). The access date was June 2023, and the data acquisition time range was from January 1, 2022 to December 31, 2022. All 35 orbital data beams in the study area are selected based on Xinping County's vector boundary. In order to obtain light spots with high quality and complete information, this study filtered out invalid light spots through 5 indicators of GEDI (Table 5), namely: `lon_lowestmode`, `lat_lowestmode`, `qualit_flag`, `sensitivity`, `degraded_flag` and previous research experience [0].

Table 5. GEDI footprint quality filtering criteria.

Parameters	Retention value	Retention basis
<code>lon_lowestmode</code>	101-103°E	The geographical coordinate range of the study area.

lat_lowestmode	23-25°N	The geographical coordinate range of the study area.
quality_flag	1	It indicates that the waveform meets the specific highquality standard.
Sensitivity	≥ 0.9	It indicates high spot quality.
degrade_flag	0	A value of 1 means the satellite is in a descending orbital state, resulting in inaccurate data. Consequently, the retention value is 0 light spots.

After the target points were screened out by the above 5 indicators, there remained 58,421 valid footprint points in the study area, of which 1,689 are *D. giganteus* woodlands and 56,732 non-*D. giganteus* woodlands (Fig.2). GEDI L2B product data information can be queried from GEDI L2B product data dictionary, containing 35 modelling alternatives and 5 quality screening indicators.

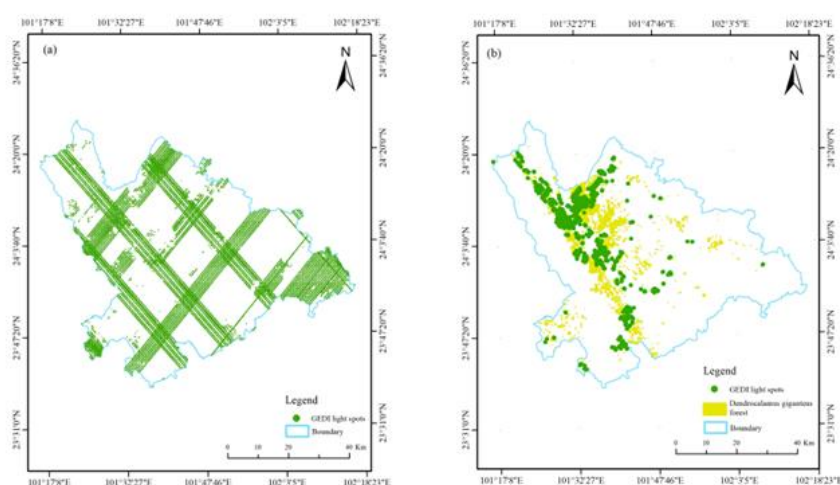


Figure 2. (a) Distribution of all footprint locations. (b) Distribution of light spots following filtration.

In this study, we also used the data of the secondary inventory of forest resources in Xiping County in 2016, which was used for mask extraction of *D. giganteus* forest land.

1.4 Research method

The main steps of constructing a single *D. giganteus* chlorophyll model based on the traditional model chlorophyll estimation method and conducting region-scale chlorophyll inversion are as follows: (1) Construction of chlorophyll model of single *D. giganteus*; (2) Selection and accuracy evaluation of interpolation model; (3) Correlation analysis and importance ranking of feature parameters; (4) Construction and accuracy evaluation of the chlorophyll content estimation model (Fig.3). The specific process is to fit the basic model of chlorophyll content of *D. giganteus* by the basic information of individual *D. giganteus*. The chlorophyll content at the plot level is calculated by the basic model, which is used as the training sample (dependent variable) for modeling. The characteristic value of the characteristic variable factor on the corresponding sample site is extracted from the remote sensing data as the modeling sample (independent variable), and the chlorophyll content prediction model of *D. giganteus* at the regional scale is constructed by using the dependent variable and the independent variable, and then the chlorophyll content of *D. giganteus* in the study area is inverted.

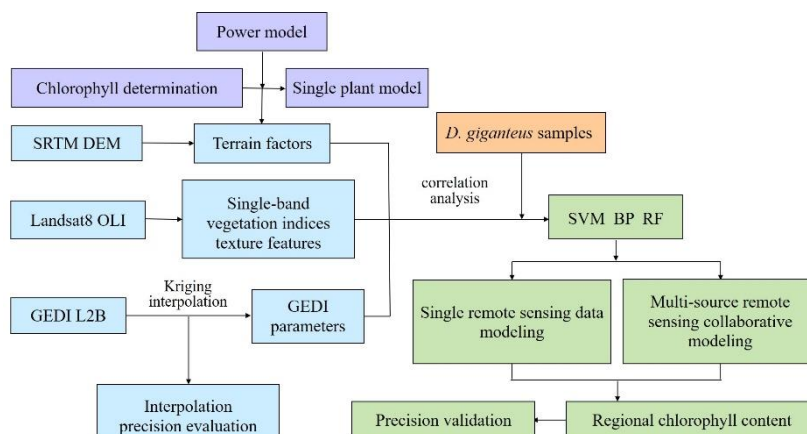


Figure 3. Technical route.

1.4.1 Construction of chlorophyll model of single *D. giganteus* plant

The relative growth equation formula (also known as "allometric growth equation") method is used to establish a chlorophyll model of a single bamboo plant, and this model is taken as the basic model of chlorophyll of a single *D. giganteus* plant in the study area. The total chlorophyll of 35 plots was calculated and then used for regional scale chlorophyll inversion. The relative growth equation is calculated as follows:

$$Chl_T = aDBH^b \quad (4)$$

where, Chl_T represents the chlorophyll content of a single *D. giganteus* plant, DBH represents the diameter of a single *D. giganteus* plant, a and b represent the parameters to be estimated for a single plant model.

1.4.2 Geostatistics method

• Semi-variogram

Semi-variogram, also known as semi-variance function, is a special tool in geostatistics. It was first proposed by Matheron in 1963, who defined it as the mathematical expectation of the square of the increment of a regionalized variable. The research object of this paper is chlorophyll content. The chlorophyll content of each pixel can be considered as a regionalized variable affected by its characteristic factors. The functional relationship between independent variables and chlorophyll content can be used to describe the spatial distribution characteristics of chlorophyll [0-36]. Therefore, in this study, GS + 9.0 software was used to fit the optimal semi-variance model of each parameter, defined as the following formula:

$$r(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z(x_i) - z(x_i + h)\}^2 \quad (5)$$

where, $r(h)$ represents the value of semi-variance function; $N(h)$ represents the number of point pairs with an interval distance of h . $z(x_i)$ represents the variable of the region, that is, the pixel chlorophyll; h is the distance between two sample points.

• Simple Kriging interpolation method

The simple Kriging interpolation technique, frequently referred to as the 'Ordinary Kriging' (OK) interpolation approach, stands out as one of the most prevalently employed methods within geostatistical analysis. In the study, ArcGIS Pro 2.8 software was used to provide unbiased estimation by combining the structural characteristics of sampling data and the semi-variance function of regionalized variables. According to the principle of minimum variance, the weighting coefficient of unsampled points is determined, and the data of unsampled points are obtained by interpolation method [0]. That is, using different sample data in the study area to estimate the unmeasured data in the study area, defined as the formula:

$$Z_{\theta}^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (6)$$

where, $Z_{\theta}^*(x_0)$ represents the Kriging estimation result, n is the number of GEDI footprints, λ_i denotes the weight factor, x_i represents the location of any point in the study area, $Z(x_i)$ denotes the sample point of a known value.

- Interpolation accuracy evaluation

In this study, GS + 9.0 software was utilized to derive the optimal semi-variance model for the parameters. The evaluation of the model's fitting effect first considered the R^2 and the sum of residual squares, followed by the nugget effect value, and finally assessed the range and sill. The larger the R^2 , the smaller the RSS, the larger the range value, the smaller the nugget value and the nugget effect value, the better the fitting effect of the semi-variance function model [0].

Then, under the ArcGIS pro 2.8 software geostatistics module, the OK interpolation is used to interpolate the GEDI point attribute data into the area attribute data. The interpolation effect was evaluated by mean error (ME), root-mean-square error (RMSE), standard mean error (MSE), standardized root-mean-square error (RMSSE) and mean standard error (ASE) [0]. Each evaluation index was defined as a formula:

$$ME = \frac{\sum_{i=1}^n [\hat{Z}(x_i) - Z(x_i)]}{n} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\hat{Z}(x_i) - Z(x_i)]^2}{n}} \quad (8)$$

$$MSE = \frac{\sum_{i=1}^n [\hat{Z}(x_i) - Z(x_i)] / \hat{\sigma}(x_i)}{n} \quad (9)$$

$$RMSSE = \frac{\sum_{i=1}^n \{[\hat{Z}(x_i) - Z(x_i)] / \hat{\sigma}(x_i)\}^2}{n} \quad (10)$$

$$ASE = \sqrt{\frac{\sum_{i=1}^n \hat{\sigma}(x_i)}{n}} \quad (11)$$

where, $\hat{Z}(x_i)$ is the predicted value of x at i position, $Z(x_i)$ is the observed value of x at i position, and n is the number of footprint points.

1.4.3 Establish a regional chlorophyll estimation model

In order to use different model methods to compare the estimation accuracy of 2 different RS data and the synergy of two different RS data, this study selected the more traditional non-parametric SVM, BP and RF model, combined with RS characteristic variables and field sample survey data, to develop a chlorophyll content estimation model for *D. giganteus* in the study location.

Support vector machine regression (SVR) is a regression technique derived from SVM. The basic idea is to find a function so that the difference between the predicted value and the true value of the function for all training samples does not exceed a given threshold, and in the case of satisfying this condition, the function is smoothed as much as possible. This study uses the statistics of MATLAB R² 023a software and the fitrsvm function in the machine learning toolbox to train the SVM regression model. The kernel function selects the Gaussian kernel, and optimizes the hyperparameters through the random search optimizer [0]. Randomly sample within the predefined range of the hyperparameters and evaluate the performance of each sampling point. The performance is evaluated by leaving a cross.

BP neural network is a commonly used feed-forward artificial neural network. It trains parameters by back-propagation algorithm, so as to realize tasks such as pattern recognition and function approximation. The BP network consists of an input layer, a hidden layer, and an output layer, with each layer containing multiple neuron nodes that are connected to all nodes in the previous layer. The main training process includes two stages: forward propagation and back propagation. In this study, the MLPRegressor module under the Neural Network Toolbox

built in MATLAB 2023a software is used to realize the regression modeling of BP neural network.

RF is an ensemble learning algorithm based on decision tree, which averages or votes the results of multiple different and independent sub-trees as the final result of the model. The randomness in RF mainly comes from randomly selecting samples and feature variables [0]. In the training phase, the samples are sampled from the training set by the self-help method (Bootstrap) sampling, and several new training samples are obtained. The number of samples in the new training samples is the same as that of the original samples. Therefore, for a certain sample, it may be selected multiple times or may not be selected. The training data obtained from each tree is not all the data in the original training sample, which reduces the risk of overfitting [0]. In RF, there are two important parameters that control the model: one is the number of trees of subtrees ($n_estimators$), and the other is the maximum number of features ($max_features$). By controlling the number of sub-trees, the accuracy of the model is affected. When the amount of data is large, the more the number of sub-trees, the more reliable the recognition accuracy of the model. When the amount of data is small, too many subtrees will bring too much computation, which will reduce the efficiency of the model. Similarly, the maximum number of features should also be appropriately adjusted according to the feature dimension in the training samples. Therefore, the modeling parameters should be selected according to the number of samples [0]. In this study, the RF model mainly uses the RandomForestRegressor module in the sklearn, pandas, and numpy libraries of the Python3.10 version programming software to realize the construction of the estimation model and the inversion of the chlorophyll content of the bamboo in the study area.

1.4.4 Model evaluation

In order to verify the prediction accuracy and estimation results of the RS estimation model of chlorophyll content, the leave-one-out cross validation (LOOCV) method was used. Due to the limited number of ground survey samples, this method is used to reduce the potential error caused by the segmentation of training and validation samples. LOOCV participates in training modeling and verification one by one for small sample data, which solves the problem that the modeling set and the verification set are the same, and effectively avoids the problem of local optimum in the fitting model [0-45]. Compared with K-fold cross-validation, the verification results are reproducible and not affected by random factors, and have stronger robustness, which can effectively solve the problem of overfitting or under-fitting of the estimation model [0].

In this study, to better evaluate the feasibility of the RS estimation model of chlorophyll content in *D. giganteus* and verify the model's performance, three indicators were selected: the determination coefficient R-squared (R^2), root mean square error (RMSE), and overall prediction accuracy (P). Among them, the closer the R^2 value is to 1, the better the goodness of fit of the model; the smaller the RMSE value and the larger the P value, the higher the accuracy of the model. On the contrary, the model fitting effect is not ideal. Each evaluation index is defined as the formula:

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (13)$$

$$P = \left(1 - \frac{RMSE}{\bar{y}} \right) \times 100\% \quad (14)$$

where, n is the number of samples, y_i is the predicted value of the model, $\hat{y}_i$ is the measured value of chlorophyll, and $\bar{y}$ is the predicted average value of the model, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

2 Results and analysis

2.1 Single plant *D. giganteus* model

In this study, the power function model was constructed with the DBH of single *D. giganteus* as the independent variable and the chlorophyll as the dependent variable. According to the chlorophyll modeling data, the single plant chlorophyll-diameter scatter plot was drawn (Fig.4). Based on the R^2 value, the model exhibits a high degree of fitting; furthermore, the accuracy of the model is demonstrated to be very good through analysis of the RMSE and P values. The chlorophyll content of individual *D. giganteus* is positively correlated with the DBH, that is, under certain growth conditions, the chlorophyll of *D. giganteus* increases with increase of the DBH, which is in line with the actual growth situation of *D. giganteus*.

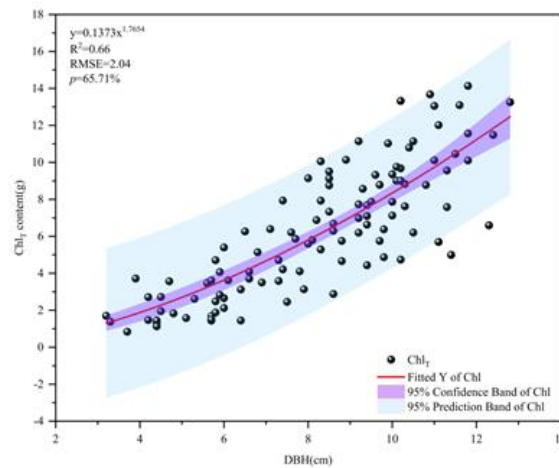


Figure 4. Chlorophyll model of a single *D. giganteus* plant.

2.2 Selection of Semivariogram Models

In this study, under the GS + 9.0 software, the linear, spherical, exponential and Gaussian models were used to fit the semi-variance function, and the nugget value, sill value, nugget effect and range of each parameter were calculated. To ensure model accuracy, we select the semi-variance function model with the highest R^2 and smallest residual (RSS) as optimal. The model parameters obtained by fitting the semi-variance are shown in Table 6. The extracted GEDI parameters have strong spatial autocorrelation and can be analyzed by OK interpolation. Among them, when the nugget effect is less than 25%, it indicates a strong spatial autocorrelation, and greater than 75% indicates a weak spatial autocorrelation [0]. In this study, except for the linear model, the spherical, exponential, and Gaussian models of pai, pgap_theta, and pgap_theta_a3 all satisfy the optimal range of nugget effect less than 25%.

Table 6. Four models fit semi-variance function result tables.

Parameter name	Model	Nugget	Sill	Nugget effect	Range	R^2	RSS
pai	Linear	3.45	3.71	0.93	61536.40	0.56	0.07
	Spherical	0.19	3.61	0.05	3700.00	0.62	0.06
	Exponential	0.42	3.61	0.12	3600.00	0.66	0.06
	Gaussian	0.59	3.61	0.16	3117.69	0.62	0.06
pgap_theta	Linear	0.06	0.07	0.94	61536.40	0.40	3.51×10^{-5}

	Spherical	0.00	0.07	0.05	3700.00	0.68	1.86×10^{-5}
	Exponential	0.01	0.07	0.12	3900.00	0.73	1.60×10^{-5}
	Gaussian	0.01	0.07	0.17	3290.90	0.68	1.90×10^{-5}
	Linear	0.06	0.06	0.96	61536.40	0.32	1.77×10^{-5}
pgap_theta_	Spherical	0.00	0.06	0.06	3600.00	0.75	6.57×10^{-6}
a3	Exponential	0.01	0.06	0.11	3300.00	0.77	5.90×10^{-6}
	Gaussian	0.01	0.06	0.18	3117.69	0.75	6.83×10^{-6}

Note: The pai parameter represents the Total Plant Area Index. The variable pgap_theta represents the Pgap(theta) under L2B calculated by L2A products. Similarly, the parameter pgap_theta_a3 is the same, where a3 represents the third algorithm of GEDI.

2.3 Verification of Kriging interpolation results

The Kriging interpolation results of pai, pgap_theta and pgap_theta_a3 were fully evaluated by cross-validation method, and the results showed that all indexes met the evaluation principles (Table.7). This shows that the OK interpolation method performs well in accuracy and spatial prediction, and reveals the feasibility of using the OK interpolation method to interpolate point data into polygon surface attribute information (Fig.5).

Table 7. OK interpolation cross-validation results table.

Parameter name	ME	MSE	ASE	RMSE	RMSSE	Model
Pai	0.00	0.00	1.70	1.69	1.00	Exponential
pgap_theta	0.00	0.00	0.25	0.27	1.08	Exponential
pgap_theta_a3	0.00	0.00	0.27	0.27	1.01	Exponential

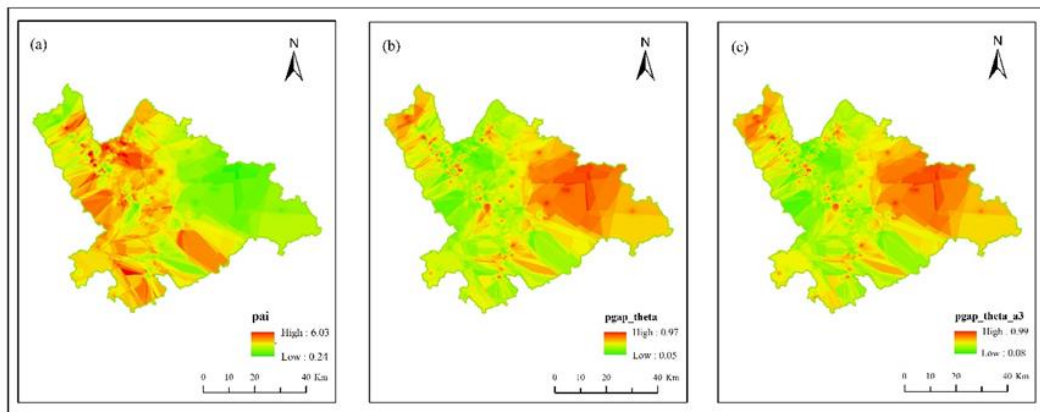


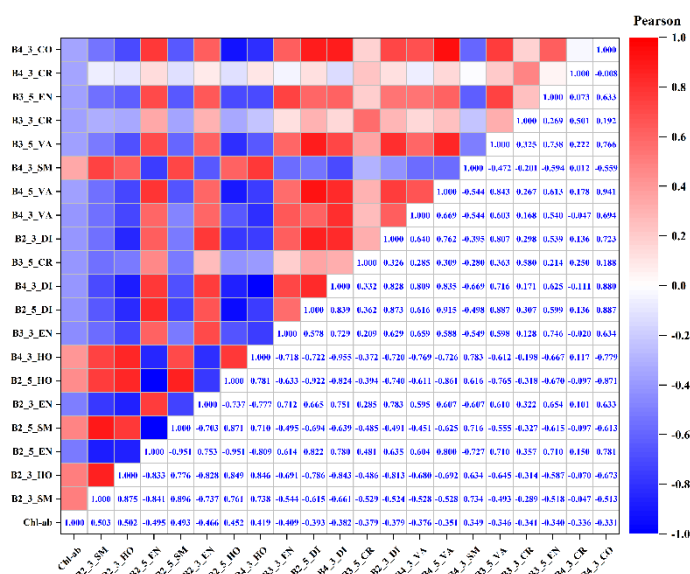
Figure 5. The spatial prediction distribution map of GEDI variables.

2.4 Variable screening results and importance

Considering that the sample size in this study is small, the characteristic parameters of the top 3 of the correlation with chlorophyll content from high to low in the two data sources are selected as the modeling variables. It can be seen from Fig.6 and Fig.7 that the correlation between Landsat 8 and GEDI characteristic parameters and chlorophyll content ranks the top three, and the significance level is above 0.01, that is, Landsat 8 selects B2_3_SM, B2_3_HO, B2_5_EN, GEDI selected pai, pgap_theta, pgap_theta_a3. Among them, the correlation between B2_3_SM extracted by Landsat 8 and chlorophyll content was the highest, which was 0.503. The correlation between pai extracted by GEDI and chlorophyll content

was the highest, which was 0.346.

After the optimal parameters were selected by Pearson correlation analysis, the feature importance of variables was evaluated by random forest. It can be seen from Fig.8 that the importance of variable characteristics using single Landsat 8 or GEDI data is different from that using two kinds of data. The contribution of B2_5_EN and B2_3_SM is the largest, and the contribution of pgap_theta is the lowest. The importance of the characteristics of the two data sources from high to low is B2_5_EN, B2_3_SM, pa, i, pgap_theta_a3, B2_3_HO, pgap_theta. The contribution rates were 29.503 %, 24.852 %, 16.873 %, 10.701 %, 9.389 % and 8.682 %, respectively.



Note: B2_3_SM and B2_3_HO represent the second-order moment and cooperativity texture information calculated under the 3×3 window in the second band of ladsat 8 image, respectively. B2_5_EN represents the information entropy texture information calculated in the 5×5 window of the second band of ladsat8 image.

Figure 6. Correlation matrix between Landsat 8 variables and chlorophyll content.

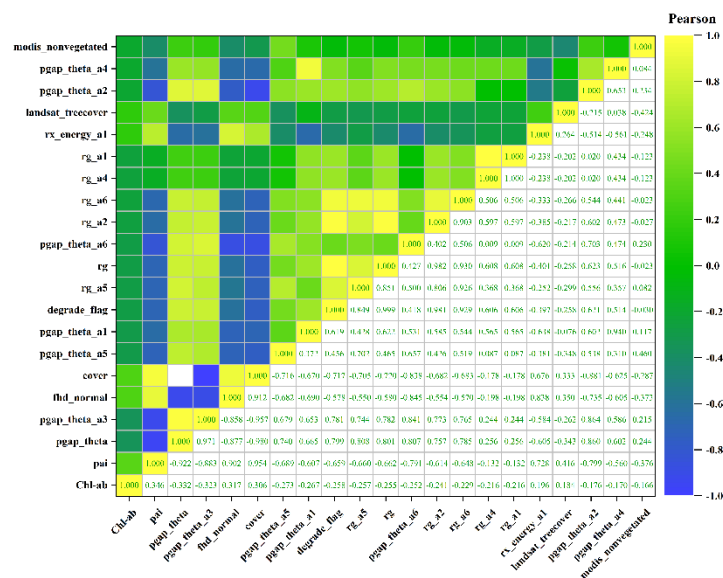


Figure 7. Correlation matrix between GEDI variable and chlorophyll content.

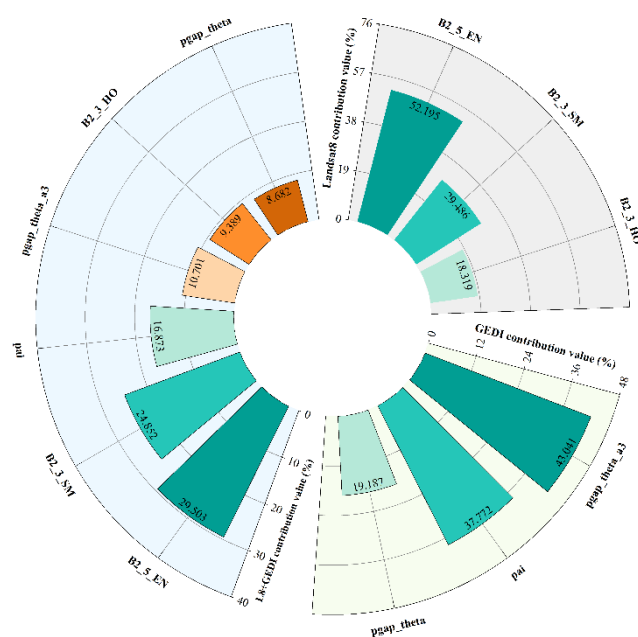
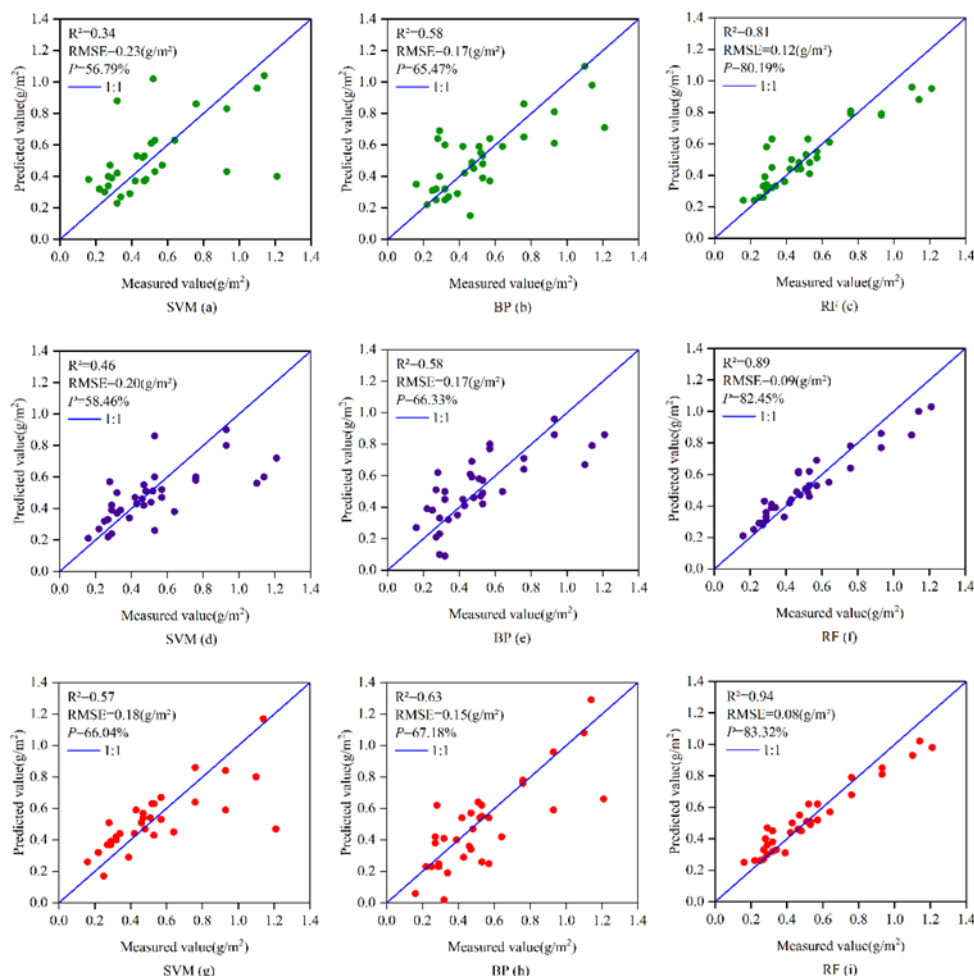


Figure 8. RF modeling parameter importance contribution ratio.

2.5 Accuracy evaluation of model

2.5.1 Analysis of regional scale chlorophyll content model estimation results

With the help of SVM, BP neural network and RF algorithms, the optimal chlorophyll content estimation model was constructed by using the characteristic variables selected by Landsat 8 and GEDI respectively (Fig.9). The modeling results showed that the RF regression model is better than the SVM and BP neural network models. The R^2 of RF (c), (f), and (i) are 0.81, 0.88 and 0.94, respectively. The RMSE are 0.12 g/m², 0.09 g/m², and 0.08 g/m², respectively. The P are 80.19%, 82.45%, and 83.32%, respectively. Among them, the model fitting effect and accuracy of GEDI feature variables participating in modeling are better than Landsat 8, and the results of Landsat 8 and GEDI collaborative participation in modeling are better than the first two. Therefore, this study verified the reliability of the collaborative operation of Landsat 8 and GEDI and the use of random forests to establish a chlorophyll estimation model.



Note: (a), (b) and (c) Landsat 8 feature variables are involved in modeling, (d), (e) and (f) Participate in modeling for GEDI characteristic variables, (g), (h) and (i) The Landsat 8 and GEDI feature variables are involved in the modeling at the same time.

Figure 9. Scatter plot of 3 different models.

2.5.2 Analysis of regional scale chlorophyll content model estimation results

The optimal model established by Landsat 8 and GEDI RS data was used to estimate the chlorophyll content of *D. giganteus* in Xinping County (Table.8). The distribution range of chlorophyll content in *D. giganteus* is 0.24 g/m² ~ 1.02 g/m², the average chlorophyll content is about 0.52 g/m², accounting for about 50.67 % of the total area of *D. giganteus* in the study area. The maximum chlorophyll content is about 1.02 g/m², and the minimum is about 0.24 g/m². Throughout the study area, the distribution of high and low chlorophyll content alternates, indicating significant regional variations (Fig.10). The high chlorophyll content area is mainly concentrated in the border area of three townships, namely, Shuitang Township, Laochang Township and Gasa Township, as well as in the area close to the Gasa River and the western part of the Mosha River, while the eastern area, which is inhabited by a large number of people and with a small distribution of *D. giganteus*, has a lower value of chlorophyll content. In addition, from the spatial distribution map, the chlorophyll content range corresponding to different chroma bands

can be observed to be mainly concentrated between 0.40 and 0.70. The number and proportion of pixels in this interval are the highest, reflecting that the growth status of *D. giganteus* in Xinping County is relatively good.

Table 8. OK interpolation cross-validation results table.

Chl-ab (g/m ²)	Number of pixels	Proportion (%)
0.24~0.40	36625	22.55
0.40~0.55	65814	40.52
0.55~0.70	42897	26.41
0.70~0.85	10128	6.23
0.85~1.02	6974	4.29

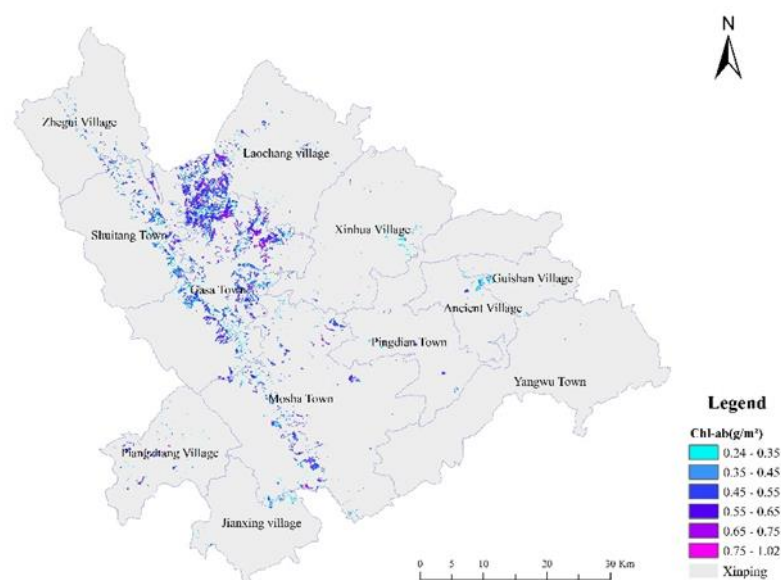


Figure 10. The spatial distribution map of chlorophyll content in *D. giganteus*.

3 Discussion

In this study, we studied the collaborative operation of multi-source RS data to estimate the chlorophyll content of *D. giganteus*. The power function model of chlorophyll content in single plant of *D. giganteus* was established, and then the chlorophyll content of each plot was calculated by the single plant model, and the chlorophyll content in the study area was estimated by RF machine learning algorithm. This can effectively reduce costs, improve efficiency and estimation accuracy, and provide a reference for long-term monitoring of vegetation chlorophyll content. Combining spaceborne (such as GEDI, ICESat-2) and optical (such as Landsat, Sentinel) multi-source RS data offers researchers a novel research approach to estimate vegetation chlorophyll content. The primary challenge encountered when estimating chlorophyll content with GEDI data and Landsat 8 OLI data is addressing the discreteness of GEDI footprint points and resolving the resolution mismatch between GEDI and Landsat 8 images. Therefore, this study aims to solve these difficulties to improve the accuracy and precision of chlorophyll content

inversion. Through the solution of the above problems, it shows that there is great potential for multi-source RS data to collaboratively estimate chlorophyll content at the county scale, and provides a reference for the application of medium and large regional scales.

3.1 The potential of multi-source RS data to estimate chlorophyll content

The single RS data can no longer meet the estimation requirements of medium and large-scale chlorophyll content retrieval. Optical RS is susceptible to light saturation effect [18-43]. In addition to the high cost of acquisition, the high-resolution images (such as GF, QuickBird, IKONOS) are quite different from the space-borne LiDAR GEDI data in resolution [0], that is, discrepancies in pixel scale among image data and between pixel scale and plot scale can introduce uncertainties in the estimation of chlorophyll content, thereby impacting the accuracy of the estimation. Although the spaceborne LiDAR is less affected by the light saturation effect and can obtain the vertical structure information of the forest, its light spot data is discrete and discontinuous in space [0-48]. In view of the above situation, this study used the satellite RS data Landsat 8 OLI and the spaceborne LiDAR data GEDI L2B to carry out the inversion of chlorophyll content in the study area, which made up for the limitations of using a single data source to invert chlorophyll content, and also solved the problem of large differences between pixel scales and between pixel scales and sample scales.

In view of the discrete and discontinuous distribution of spaceborne LiDAR GEDI L2B product data, this study uses a more mature geostatistical method for OK interpolation. In order to ensure the validity of the model fitting results, the effective light spots are divided according to the ratio of 8:2 before OK interpolation, that is, 80% of the light spot data is used as the training set, and the remaining 20% of the light spot data is used as the verification set of OK interpolation results. Firstly, the semi-variance function analysis is carried out on 80% of the training set data [0], and the most common linear, spherical, exponential and Gaussian models were used to fit the semi-variance function. In order to ensure the accuracy of the model, the model with strong spatial autocorrelation (nugget effect less than 25%), the highest coefficient of determination (R^2) and the smallest residual (RSS) was selected as the optimal semi-variance function model [0]. In this study, except for the linear model. The spherical, exponential and Gaussian models of pai , pgap_theta and pgap_theta_a3 all satisfy the optimal range of nugget effect less than 25%. According to the principle of maximum R^2 and minimum RSS, the exponential model of pai and pgap_theta is selected as the optimal semi-variance function model. In addition, considering the above factors, according to the principle of maximum R^2 , pgap_theta_a3 finally chooses the exponential model as the optimal semi-variance function model. Secondly, OK interpolation is performed under ArcGIS 10.8. Finally, the interpolation results are verified by cross validation method. According to the index evaluation principle of Bostan P A et al [0], the pai , pgap_theta and pgap_theta_a3 selected in this study all meet the values of ME and MSE close to 0. The values of RMSE and ASE are close to each other, and RMSSE is close to 1. This result is consistent with the verification results of Bargaoui et al [0] and Qiao et al [0], both of which are used to study biomass. Therefore, this study not only solves the discreteness of GEDI light spot data, but also confirms the feasibility of OK interpolation. From Fig.9, it can be seen that the collaborative modeling effect of two RS data sources is better and more accurate than the model estimated by the variables of a single data source. The R^2 of the model increased from 0.81 ~ 0.89 to 0.94, RMSE decreased from 0.09 ~ 0.12 g/m² to 0.08 g/m², and P increased from 80.19% ~ 82.45% to 83.32%, which

laid a foundation for accurately estimating the chlorophyll status of the forest area and further understanding the function of the forest ecosystem and the health status of the forest. In addition, this study also confirmed that GEDI L2B data can be used not only for the estimation and inversion of structural parameters such as closing degree, biomass, and carbon storage, but also for the estimation and inversion of biochemical parameters such as chlorophyll content. This provides a research case for the estimation and inversion of mesoscale and large-scale chlorophyll content, and provides a scientific basis for the health monitoring of global forest ecosystems.

3.2 Analysis of the influence of parameter selection on model accuracy

In this study, the parameter selection includes the independent variable selection of the chlorophyll content model of the single plant and the independent variable selection of the chlorophyll content estimation model of the regional scale. The selection of independent variables for the chlorophyll content model of single plant of *D. giganteus* is almost blank in previous studies, but scholars have done similar research in the study of tree biomass. There is a significant correlation between the above-ground biomass of trees and their diameter at breast height. For field measurements, there is usually a large error in the measurement results of tree height, so height is not a better modeling parameter [50,53]. In order to avoid the error caused by the introduction of *D. giganteus* height and improve the efficiency and feasibility of field measurement, it was found that the regression model with DBH as a single variable could more accurately reflect the trend of aboveground biomass of different bamboo varieties with DBH [53-55]. Therefore, in this study, the DBH of *D. giganteus* was used as a single variable to establish a chlorophyll content model of *D. giganteus*. Compared with the traditional destructive sampling and estimation of chlorophyll content, the model established by the allometric growth equation has better universality, which provides a favorable reference value for estimating the chlorophyll content of large clustered *D. giganteus* and forest health monitoring in the future.

Aiming at the selection of independent variables for the estimation model of chlorophyll content of *D. giganteus* at the regional scale, a large number of previous studies have been limited to the study of single-band reflectance and band combination information in chlorophyll inversion [7], while ignoring the texture feature information in RS images that is conducive to improving the interpretation accuracy, and the application of GEDI L2B RS data with rich feature information in chlorophyll content inversion. From the results of this study (Fig.6), the correlation between texture features and chlorophyll content is better than the relationship between vegetation index and single-band reflectance and chlorophyll content, and is consistent with the results of Yang Y et al [9]. The model established by GEDI L2B feature parameters is more accurate than the model established by Landsat 8 OLI feature parameters. As shown in Figs.9 (c) and (f), the model R^2 established by GEDI L2B feature parameters is 0.89, while the model R^2 established by Landsat 8 OLI feature parameters is 0.81, indicating that the optical RS data itself has light saturation effect. At the same time, this study also confirms that GEDI L2B product data is not limited to the study of tree biomass and carbon storage. It can also be used in the application of mesoscale and large-scale chlorophyll content inversion.

3.3 Model selection in uncertainty evaluation of chlorophyll content estimation accuracy

This study involves selecting a basic model for estimating chlorophyll content in individual *D. giganteus* plants and a regional-scale model for estimating chlorophyll content. The basic model of chlorophyll content per plant of

D. giganteus chooses the allometric growth equation as its basic model. In previous studies, the allometric growth equation was mostly used to estimate the biomass, net output productivity and biogeochemical cycle budget in forest ecosystem [0]. A small number of scholars used to establish the regression model of the relationship between leaf area index and DBH of single tree to predict the change process of productivity of *Ribinia pseudoacacia* forest [0]. From Fig.4, it can be seen that there is a significant allometric growth relationship between the independent variable DBH and the dependent variable chlorophyll content per plant. It is indicated that the allometric growth equation can also be used to estimate the chlorophyll content of single plant by the basic model of chlorophyll content of single plant.

Aiming at the selection of chlorophyll estimation models for regional scale *D. giganteus*, this study comprehensively considers whether the selected model matches the number of known samples. Because the representativeness of the selected model is related to the number of modeling samples, the more the number of modeling samples is, the more representative the estimation model is, and the uncertainty will also decrease. However, with the increase of the number of model samples, when the number of model samples reaches a certain critical value, the accuracy of the estimation model will no longer change significantly. Therefore, in order to save manpower, material and financial resources, and meet the small sample principle (30) and accuracy requirements of field investigation [54], this study investigated 35 measured sample plots for RS modeling research. According to the previous research on the estimation of chlorophyll content, the accuracy of the chlorophyll content estimated by the more mature RF algorithm model is higher than that of other common parameter models (such as partial least squares model, multiple linear regression model) and non-parametric models (such as SVM model, K nearest neighbor algorithm) [0,0]. The results of this study indicate that the RF model provides the most accurate estimation of chlorophyll content. Therefore, it was selected as a RS model to estimate the chlorophyll content of *D. giganteus* in Xinping County. From the results of this study, the RF model is also more accurate than the SVM model and the BP model to estimate the chlorophyll content. Therefore, RF model was finally selected as the estimation model of chlorophyll content of *D. giganteus* in Xinping County. The chlorophyll content of *D. giganteus* in the study area was $0.24 \sim 1.02 \text{ g/m}^2$. At present, there are few studies on the chlorophyll content of bamboo plants, especially on the chlorophyll content of *D. giganteus*. Therefore, compared with the results of Jin et al. 's [0] RS estimation of total chlorophyll content in wheat leaves, R^2 is 0.868 and RMSE is 0.384 g/m^2 , the estimation accuracy of this study is higher. Compared with the results of Richardson et al. [0] and Gitelson et al. [0], which only studied the chlorophyll content of single leaves of higher plants, this study extrapolated the chlorophyll content of single plant leaves to the RS estimation of chlorophyll content in the study area, which provided an important reference value for the assessment of forest health and the scientific management of forest resources. In addition, from the three levels of *D. giganteus* chlorophyll content from high to low, the number of samples was 14, 14 and 7, respectively, and the proportion of graded pixels to the total pixels was 22.55%, 66.93% and 10.62%, respectively (Table.8), indicating that the distribution of samples and pixels is relatively reasonable. At the same time, it reflects the representativeness of sampling and the rationality of modeling results, so as to reduce the uncertainty and error transmission caused by sampling.

3.4 Limitations of estimation of chlorophyll content in *D. giganteus*

In this study, the size of the sample plot was 30 m × 30 m, so the Landsat 8 data with a resolution of 30 m and the GEDI data with a spot footprint radius of 25 m were selected. In order to ensure its consistency, this study resampled the spatial resolution of GEDI data to 30 m by Kriging interpolation [0]. Although the GEDI L2B has chlorophyll-related leaf area index, canopy cover and waveform vegetation energy values, its spot footprint points are discontinuous and the amount of data is large, in order to solve this problem, the poor quality spots need to be eliminated before Kriging interpolation [0], and then interpolated into surface data to obtain the characteristic parameter information of the whole study area.

Furthermore, this study only uses three common estimation models to estimate the chlorophyll content of *D. giganteus*. In the future research, according to the different number of samples, the appropriate parameter model and the optimized non-parametric model can be selected as the estimation model of chlorophyll content, and the parameter selection can be taken into account. Whether there is anti-interference between the parameters of the two RS data sources, which makes the model better than the single RS data modeling effect, and then select the optimal model as the inversion model of chlorophyll content; for the demand of high-precision, high-resolution and large-scale chlorophyll content inversion, high-resolution GF, IKONOS, QuickBird, Sentinel-2 and hyperspectral data can be combined with GEDI or ICESat-2 data, and finally modeling and inversion with a unified resolution.

4 Discussion

The purpose of this study was to explore the feasibility of using the RS data of two different sensors, Landsat 8 and spaceborne LiDAR GEDI, to estimate the chlorophyll content of large sympodial *D. giganteus* in Xinping County. The results show that through the comprehensive utilization of two RS data sources for collaborative modeling, the model fitting effect is better and the accuracy is higher, compared with the single RS data source modeling. The spatial distribution map of chlorophyll content is consistent with the sub-compartment data of the second-class survey of forest resources in Xinping County in 2016, and is consistent with the distribution of large-scale clustered *D. giganteus* resources in the study area. This shows that on the one hand, the comprehensive utilization of multivariate RS data has more comprehensive and more accurate vegetation information than single RS data, and has more advantages in improving the accuracy and reliability of the estimation accuracy of chlorophyll content of *D. giganteus*. On the other hand, RS technology can quickly monitor large areas of forest vegetation around the world. Compared with traditional destructive sampling, it is more efficient, convenient and environmentally friendly, and the data is universal, which ensures the repeatability of estimating chlorophyll content. This provides a scientific basis for the monitoring and management of global vegetation ecosystems, promotes the application and development of ecological environment RS technology, and promotes the progress of global RS.

Qi, M.; Zhang, C. Research Progress on Hyper-spectral Remote Sensing Retrieval for Forest Physical and Chemical Parameters. *World Forestry Research*. 2016, 29, 52-57.

Ni, H.; Chu, H.; Su, W.; Fan, S. Effects of management intensities on soil aggregate stability and carbon, nitrogen, phosphorus distribution in *Phyllostachys edulis* forests. *Chinese Journal of Applied Ecology*. 2023, 34, 928-936.

Madeira, A.C.; Mentions, A.; Ferreira, M.E.; Taborda, M.d.L. Relationship between spectroradiometric and chlorophyll measurements in green beans. *Communications in Soil Science Plant Analysis*. 2000, 31, 631-643.

Ahmad, I.; Zhu, G.; Zhou, G.; Song, X.; Hussein Ibrahim, M.E.; Ibrahim Salih, E.G. Effect of N on growth, antioxidant capacity,

and chlorophyll content of sorghum. *Agronomy*. 2022, 12, 501.

Qiao, L.; Tang, W.; Gao, D.; Zhao, R.; An, L.; Li, M.; Sun, H.; Song, D. UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. *Computers electronics in agriculture*. 2022, 196, 106775.

Wu, Q.; Zhang, Y.; Zhao, Z.; Xie, M.; Hou, D. Estimation of relative chlorophyll content in spring wheat based on multi-temporal UAV remote sensing. *Agronomy*. 2023, 13, 211.

Houborg, R.; McCabe, M.; Cescatti, A.; Gao, F.; Schull, M.; Gitelson, A. Joint leaf chlorophyll content and leaf area index retrieval from Landsat data using a regularized model inversion system (REGFLEC). *Remote Sensing of Environment*. 2015, 159, 203-221.

Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote sensing of environment*. 2016, 185, 46-56.

Yang, Y.; Zhang, X.; Gao, W.; Zhang, Y.; Hou, X. Improving lake chlorophyll-a interpreting accuracy by combining spectral and texture features of remote sensing. *Environmental Science Pollution Research*. 2023, 30, 83628-83642.

Carder, K.L.; Chen, F.; Cannizzaro, J.; Campbell, J.; Mitchell, B. Performance of the MODIS semi-analytical ocean color algorithm for chlorophyll-a. *Advances in Space research*. 2004, 33, 1152-1159.

Sun, D.; Li, Y.; Wang, Q. A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and in situ hyperspectral data. *IEEE transactions on geoscience remote sensing*. 2009, 47, 2957-2965.

Barraza-Moraga, F.; Alcayaga, H.; Pizarro, A.; F  lez-Bernal, J.; Urrutia, R. Estimation of chlorophyll-a concentrations in Lanalhue Lake using Sentinel-2 MSI satellite images. *Remote Sensing*. 2022, 14, 5647.

Chen, L.; Ren, C.; Zhang, B.; Wang, Z.; Man, W.; Liu, M. Improved Object-Based Mapping of Aboveground Biomass Using Geographic Stratification with GEDI Data and Multi-Sensor Imagery. *Remote Sensing*. 2023, 15, 2625.

Liu, H.; Fan, W.; Xu, Y.; Lin, W. Research Progress in Forest Information Extraction Based on Multi-source Data Collaboration Operation. *World Forestry Research*. 2020, 33, 33-37.

Boucher, J.; Weathers, K.C.; Norouzi, H.; Steele, B. Assessing the effectiveness of Landsat 8 chlorophyll a retrieval algorithms for regional freshwater monitoring. *Ecological Applications*. 2018, 28, 1044-1054.

Cao, Z.; Ma, R.; Duan, H.; Pahlevan, N.; Melack, J.; Shen, M.; Xue, K. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sensing of Environment*. 2020, 248, 111974.

Xu, L.; Lai, H.; Yu, J.; Luo, S.; Guo, C.; Gao, Y.; Zhou, W.; Wang, S.; Shu, Q. Carbon Storage Estimation of *Quercus aquifolioides* Based on GEDI Spaceborne LiDAR Data and Landsat 9 Images in Shangri-La. *Sustainability*. 2023, 15, 11525.

Sims, D.A.; Gamon, J.A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote sensing of environment*. 2002, 81, 337-354.

Xu, X.; Lu, J.; Zhang, N.; Yang, T.; He, J.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. *ISPRS Journal of Photo-grammetry Remote Sensing*. 2019, 150, 185-196.

Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*. 2016, 7, 3-10.

Blix, K.; Eltoft, T. Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval. *Remote Sensing*. 2018, 10.

Abba, S.; Hadi, S.J.; Abdullahi, J. River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia computer science*. 2017, 120, 75-82.

Jimeno-S  ez, P.; Senent-Aparicio, J.; Cecilia, J.M.; P  rez-S  nchez, J. Using machine-learning algorithms for eutrophication modeling: case study of Mar Menor Lagoon (Spain). *International Journal of Environmental Research Public Health*. 2020, 17, 1189.

Shah, S.H.; Angel, Y.; Houborg, R.; Ali, S.; McCabe, M.F. A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sensing*. 2019, 11, 920.

Lu, F.; Chen, Z.; Liu, W.; Shao, H. Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecological Engineering*. 2016, 95, 422-429.

Su, J.; Wang, X.; Zhao, S.; Chen, B.; Li, C.; Yang, Z. A structurally simplified hybrid model of genetic algorithm and support vector machine for prediction of chlorophyll a in reservoirs. *Water*. 2015, 7, 1610-1627.

Sonobe, R.; Sano, T.; Horie, H. Using spectral reflectance to estimate leaf chlorophyll content of tea with shading treatments. *Biosystems engineering*. 2018, 175, 168-182.

Kang, Q.; Xu, W.; Wang, L.; Hong, z.; Liu, Y. Extraction of Sugarcane Plantation in Mountainous Areas Based on Landsat-8 and Sentinel-2 Time-series Synthetic. *Chinese Journal of Tropical Crops*. 2023, 44, 1276-1287.

Azadeh A, Ghavami K, Garc  a J J. The influence of heat on mechanical properties of *Dendrocalamus giganteus* bamboo[J]. *Journal of Building Engineering*, 2021, 43: 102613.

- Li, X. The planting status and industrial development suggestions of cigar tobacco in Xinning County of Yunnan Province. *Agricultural Engineering Technology*. 2022, 42, 18+20.
- Zhao, X.; Hui, C.; Zhu, S.; Liu, W.; Cai, C.; Zhang, W.; Tu, D.; Zhu, L. Response of Soil Organic Matter Content and Its Infrared Spectral Characteristics to Different Planting Durations of *Dendrocalamus brandisii*. *Acta Agriculturae Universitatis Jiangxiensis*. 2022, 44, 1448-1456.
- Wang, X.; Huang, J. *Principles and Techniques of Plant Physiological Biochemical Experiment*. Beijing: higher education press. 2015.
- Mancino, G.; Ferrara, A.; Padula, A.; Nolè, A. Cross-comparison between Landsat 8 (OLI) and Landsat 7 (ETM+) derived vegetation indices in a Mediterranean environment. *Remote Sensing* 2020, 12, 291.
- Wu, Q.; Jin, Y.; Fan, H. Evaluating and comparing performances of topographic correction methods based on multi-source DEMs and Landsat-8 OLI data. *International Journal of Remote Sensing* 2016, 37, 4712-4730.
- Curran, P.J.; Atkinson, P.M. Geostatistics and remote sensing. *Progress in Physical Geography*. 1998, 22, 61-78.
- Ji, W.; Adamchuk, V.I.; Chen, S.; Su, A.S.M.; Ismail, A.; Gan, Q.; Shi, Z.; Biswas, A. Simultaneous measurement of multiple soil properties through proximal sensor data fusion: A case study. *Geoderma*. 2019, 341, 111-128.
- Hengl, T.; Heuvelink, G.B.; Rossiter, D.G.; geosciences. About regression-kriging: From equations to case studies. *Computers geosciences*. 2007, 33, 1301-1315.
- Li, Y.; Li, M.; Liu, Z.; Li, C. Combining kriging interpolation to improve the accuracy of forest aboveground biomass estimation using remote sensing data. *IEEE Access*. 2020, 8, 128124-128139.
- Bostan, P.; Heuvelink, G.B.; Akyurek, S. Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *International Journal of Applied Earth Observation Geoinformation*. 2012, 19, 115-126.
- Sun, D.; Li, Y.; Wang, Q. A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and in situ hyperspectral data. *IEEE transactions on geoscience remote sensing*. 2009, 47, 2957-2965.
- Breiman, L. Random forests. *Machine learning*. 2001, 45, 5-32.
- Li, J.; Mao, X. Comparison of canopy closure estimation of plantations using parametric, semi-parametric, and non-parametric models based on GF-1 remote sensing images. *Forests*. 2020, 11, 597.
- Cao, Y. Estimation of aboveground biomass of regional forest based on multi-source remote sensing. 2021.
- Marchetti, F. The extension of Rippa's algorithm beyond LOOCV. *Applied Mathematics Letters* 2021, 120, 107262.
- Ma, J.; Zhang, W.; Ji, Y.; Huang, J.; Huang, G.; Wang, L. Total and component forest aboveground biomass inversion via LiDAR-derived features and machine learning algorithms. *Frontiers in Plant Science* 2023, 14, 1258521.
- Tao, Y.; Xu, K.; Yi, Z.; Luo, X.; Gao, Y. A Semi-Variogram-based Analysis of Spatial Heterogeneity of Urban Heat Islands. *Journal of Southwest University(Natural Science)*. 2018, 40, 145-152.
- Dorado-Roda, I.; Pascual, A.; Godinho, S.; Silva, C.A.; Botequim, B.; Rodríguez-Gonzálvez, P.; González-Ferreiro, E.; Guerra-Hernández, J. Assessing the accuracy of GEDI data for canopy height and aboveground biomass estimates in Mediterranean forests. *Remote Sensing* 2021, 13, 2279.
- Liu, X.; Su, Y.; Hu, T.; Yang, Q.; Liu, B.; Deng, Y.; Tang, H.; Tang, Z.; Fang, J.; Guo, Q. Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2 data. *Remote Sensing of Environment* 2022, 269, 112844.
- Cai, C.; Cao, S.; Kong, F.; Hu, L.; Liu, T.; Sun, W.; Wang, L. A dataset of spatial distribution of spruce aboveground biomass in Western Tianshan Mountains, Xinjiang in 2014. 2022, 7, 250-263.
- Araújo, T.S.M.; Higuchi, N.; de Carvalho Júnior, J.A.; management. Comparison of formulae for biomass content determination in a tropical rain forest site in the state of Pará, Brazil. *Forest ecology*. 1999, 117, 43-52.
- Bargaoui, Z.K.; Chebbi, A. Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. *Journal of Hydrology* 2009, 365, 56-73.
- Qiao, P.; Lei, M.; Yang, S.; Yang, J.; Guo, G.; Zhou, X. Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environmental Science Pollution Research* 2018, 25, 15597-15608.
- Chave, J.; Riéra, B.; Dubois, M.-A. Estimation of biomass in a neotropical forest of French Guiana: spatial and temporal variability. *Journal of Tropical Ecology*. 2001, 17, 79-96.
- Ji, X.; Luo, Q.; Ding, Y.; Wang, Y.; Zhao, J.; Wang, S. A Study on the Aboveground Biomass Model of *Dendrocalamus brandisii*. *Journal of Bamboo Research*. 2015, 34, 49-53.
- Yang, Q.; Su, G.; Duan, Z.; He, K.; Guo, Y.; Wang, Z.; Sun, Q.; Peng, Z. Biomass structure and its regression models of *Dendrocalamus hamiltonii* Nees et Arn. et Munro population. *Journal of Northwest A & F University(Natural Science Edition)*. 2008, 127-134.
- Wang, C. Biomass allometric equations for 10 co-occurring tree species in Chinese temperate forests. *Forest Ecology Manage-*

ment. 2006, 222, 9-16.

Wang, T. Characteristics of leaf area index of artificial robinia pseudoacacia forest in loess hilly region. 2019.

Hua, Y.; Zhao, X. Multi-Model Estimation of Forest Canopy Closure by Using Red Edge Bands Based on Sentinel-2 Images. *Forests*. 2021, 12, 1768.

Lv, J. Hyperspectral Remote Sensing Inversion Models of Crop Chlorophyll Content Based on Machine Learning and Radiative Transfer Models. China University of Geosciences (Beijing). 2012.

Jin, X.-l.; Wang, K.-r.; Xiao, C.-h.; Diao, W.-y.; Wang, F.-y.; Chen, B.; Li, S.-k. Comparison of two methods for estimation of leaf total chlorophyll content using remote sensing in wheat. *Field Crops Research* 2012, 135, 24-29.

Richardson, A.D.; Duigan, S.P.; Berlyn, G.P. An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New phytologist* 2002, 153, 185-194.

Gitelson, A.A.; Merzlyak, M.N. Remote estimation of chlorophyll content in higher plant leaves. *International journal of remote sensing* 1997, 18, 2691-2697.

Li, Z.; Xuan, F.; Dong, Y.; Huang, X.; Liu, H.; Zeng, Y.; Su, W.; Huang, J.; Li, X. Performance of GEDI data combined with Sentinel-2 images for automatic labelling of wall-to-wall corn mapping. *International Journal of Applied Earth Observation Geoinformation* 2024, 127, 103643.