



OPEN Mapping forest aboveground carbon stock of combined stratified sampling and RFRK model with mean annual temperature and precipitation

Min Peng¹, Mingrui Xu², Jialong Zhang^{1✉}, Bo Qiu¹, Chenkai Teng¹ & Chaoqing Chen^{2,3}

Accurately estimating forest aboveground carbon stock (ACS) is essential for achieving carbon neutrality. At present, most non-parametric models still have errors in estimating carbon stock in regions. Given the autocorrelation inherent in spatial interpolation, combining non-parametric models with spatial interpolation offers significant potential. In this study, we combined the random forest (RF) with the ordinary kriging and co-kriging of the mean annual temperature, precipitation, slope, and elevation to establish the random forest residual kriging (RFRK) model. Meanwhile, we also developed the multiple linear regression residual kriging (MLRRK) model and the random forest residual kriging (RFRK) model. Finally, we selected the optimal model for the estimation and mapping of the ACS. The results indicate that: (1) the model achieves an R^2 of 0.871, P of 90.4%, and RMSE of 3.948 t/hm²; (2) the RFCK model with mean annual precipitation (RFCKpre) outperforms the one with mean annual temperature (RFCKtem), while the RFOK model exhibits the lowest accuracy; (3) the RFCKpre exponential model has the highest accuracy, with the highest R^2 of 0.63 and RI (0.23), the lowest RMSE of 9.3 and SSR (41,612). These findings suggest that the RFRKpre model has improved the accuracy of estimating the ACS of regional forests.

Keywords Random forest residual kriging, Forest aboveground carbon stock, Mapping, Stratified sampling

Forests are the largest terrestrial ecosystems and play a crucial role in the global carbon cycle^{1–4}. In 2020, China announced its goal to peak carbon emissions by 2030 and achieve carbon neutrality by 2060 for the first time^{5,6}. Accurately estimating forest aboveground carbon stock (ACS) is essential for carbon budgeting and trading carbon sinks, making it a critical component of global climate change and carbon cycle research^{7,8}. Forest ACS refers to the process by which forest plants absorb CO₂ from the atmosphere and store it in vegetation or soil, thereby reducing atmospheric CO₂ concentration⁶. It is a key parameter for analyzing carbon cycle processes, assessing carbon sink capacity, and understanding carbon distribution within forest ecosystems. It also serves as an essential indicator of ecosystem integrity and community structure^{9,10}. Therefore, access to accurate forest carbon stock data is critical for developing effective carbon emission strategies^{11,12}. Additionally, National Forest Continuous Inventory and Forest Management Inventory data are widely used to estimate large-scale forest biomass and ACS¹³.

Stratified sampling is the most commonly used method in forest resource surveys¹⁴. The Forest Management Inventory data includes a large number of samples, which may introduce errors in carbon stock estimation. Remote sensing images provides macro-level forest information and is not constrained by terrain or other conditions, whereas Forest Management Inventory data includes detailed field survey information. The integration of both can facilitate complementary information, offering more accurate foundational data for stratified sampling. Research results indicate that this approach can significantly enhance the accuracy of forest

¹The Key Laboratory of Forest Resources Conservation and Utilization in the Southwest Mountains of China Ministry of Education; Key Laboratory of National Forestry and Grassland Administration On Biodiversity Conservation in Southwest China; Yunnan Provincial Key Laboratory for Conservation and Utilization of In-Forest Resource, Southwest Forestry University, Kunming 650224, Yunnan, China. ²College of Soil and Water Conservation, Southwest Forestry University, Kunming 650224, China. ³Panzhuhua City Natural Resources Survey and Statistics Center, Panzhuhua 617000, China. ✉email: jialongzhang@swfu.edu.cn

ACS estimates^{15–17}. While this method is well-established in remote sensing, its implementation remains crucial for the specific sample selection and analysis process in our study.

The MLR model is a widely used tool for describing existing data and serves as an essential predictive model¹⁸. Its intuitive nature, combined with low technical requirements when processing remote sensing data, makes it suitable for a variety of forest types^{19,20}. In practical applications, the MLR model demonstrates high computational efficiency. Zhang²¹ used a regression residual model combined with a stepwise multiple regression model to build an estimation model, improving accuracy and reducing estimation errors. However, the MLR model is sensitive to outliers and prone to overfitting, which can result in discrepancies between the model's predictions and the measured values²². Interpolating the residuals of the MLR model and constructing a Multiple Linear Regression Residual Kriging (MLRRK) model can effectively reduce errors and improve the estimation accuracy of the model. Li, et al.²³ interpolated the residuals to reduce the prediction error.

In recent years, Random Forest (RF) has been widely used in research on estimating forest ACS using remote sensing^{7,24}. RF is an effective method for addressing unique ecological problems in various scenarios. It can also be used for classification, regression, and evaluating the importance of variables^{25,26}. Comparisons between RF and other machine learning techniques, as well as traditional statistical regression, have shown that RF offers higher prediction accuracy due to its lower sensitivity to noise in training samples²⁷. Zimbres, et al.²⁸ found that optical data contributed more to predictive models than radar data, with the RF model demonstrating slightly higher accuracy compared to the Classification and Regression Tree (CART) model. An Thi Ngoc, et al.²⁹ showed that combining Sentinel-2 images with RF regression algorithms can effectively predict the spatial distribution of aboveground biomass (AGB) in forests. However, it is important to note that prediction errors can occur when the distance between observations is either too large or too small.

The above studies have overlooked the spatiotemporal heterogeneity of AGB and other forest parameters, as well as the ambiguity in the geographical spatial information of AGB and the physical modeling mechanisms. Geostatistical methods can effectively capture spatial heterogeneity and correlations, and construct spatial prediction models for precise estimation. Sekulic, et al.³⁰ introduced a new spatial interpolation method, Random Forest Spatial Interpolation (RFSI), which selects the optimal combination of spatial proximity observations for making predictions at unknown locations, thereby increasing prediction accuracy. The combination of Random Forest and Kriging to build the Random Forest Kriging model effectively addresses the deficiencies mentioned above^{31,32}. The Kriging method is based on the variable itself, accounting for its spatial variability, and provides a theoretical estimation of interpolation errors. In the random forest residual kriging (RFRK) model, Kriging interpolation is applied to model the residuals predicted by the RF model. Since the residuals contain components reflecting spatial autocorrelation, the RFRK model isolates these components and incorporates them into the RF model's predictions. As a result, mapping accuracy is improved. This approach is widely used for estimating ACS³³. Several studies have utilized the random forest/kriging framework for forest ACS mapping, leading to improved accuracy^{34–36}. Du, et al.¹³ used kriging interpolation to study forest ACS at a small-scale county level and produced a thematic map at this fine scale. Zhou, et al.³¹ employed the RFRK model and included elevation data as a covariate. Demonstrating that the RFRK model significantly improved the performance of AGB prediction. However, the accuracy improvement was limited, as only elevation was considered as a covariate. Currently, the majority of studies primarily focus on predicting ACS using models such as machine learning, while neglecting the impact of spatial autocorrelation. Therefore, utilizing model residuals for kriging interpolation to enhance the accuracy of ACS estimation has become a critical task³⁷.

As one of the dominant tree species in the region, *Pinus densata* holds significant research value. Most studies on forest ACS in Shangri-La focus on the impact of terrain factors, using terrain factors for modeling and estimation³⁸. Since most of the estimations of the ACS of *Pinus densata* in Shangri-La City do not take into account the impacts of precipitation and temperature, this paper combines the residuals of non-parametric models with these two factors for research. This study uses the MLRRK and RFRK models to model and map the ACS of *Pinus densata*. We evaluated the models using the coefficient of determination (R^2), relative Root Mean Square Error (rRMSE), Root Mean Squared Error (RMSE), prediction accuracy (P), and Relative Improvement (RI) to identify the most accurate models. The study's specific objectives were to (1) utilize stratified sampling to draw representative samples and increase precision. (2) Analyze and compare the interpolation results of RFRK and MLLRK. (3) Study and analyze the interpolation results of the ordinary Kriging interpolation method and cokriging interpolation method. (4) Produce thematic maps of carbon stock. (5) Present the shortcomings of this study as well as future opportunities and challenges.

The Landsat 8 OLI remote sensing images dataset for the study area was preprocessed using ENVI 5.6 and ArcGIS 10.8 software. Based on the Forest Management Inventory data, a stratified sampling method was employed, categorizing the samples by forest age to ensure that the selected samples were representative and accurately reflected the ACS in the study area. During the remote sensing image processing, the optimal set of features strongly correlated with ACS was extracted. These features were selected using the Recursive Feature Elimination (RFE) method, which removed the less important features and retained only the most critical variables for model construction. In the predictive modeling phase, both MLR and RF models were applied to derive the predicted ACS values and calculate the corresponding residuals. Subsequently, the MLRRK and RFRK models were used to construct more accurate ACS prediction models. Finally, the optimal model was selected by comparing the predictive accuracy of the different models. The flow of this study is illustrated in Fig. 1.

Results

Stratified sampling

As shown in Fig. 2, the sampling base is concentrated in the third stratum, with a smaller proportion in the first and second strata, but it is ensured that each age group is sampled so that the samples are evenly distributed and representative.

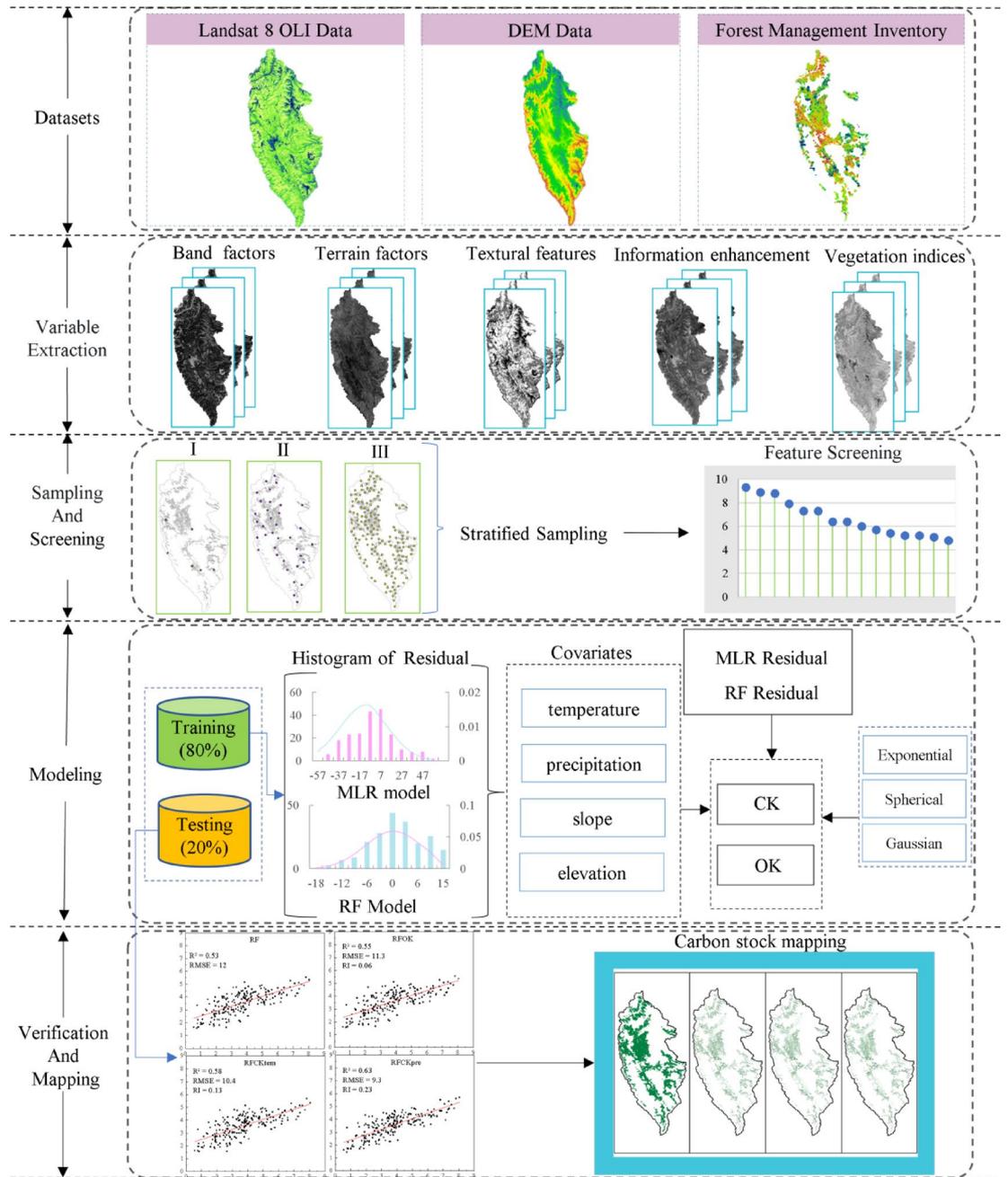


Fig. 1. Overall flow chart of this study.

Remote sensing features selection

The 'Random Forest Regressor' algorithm used in this study is available in the 'Sklearn' package of Python³⁹. To ensure the comprehensive representation of each variable, the RFE method was employed. The RFE was applied using Random Forest to iteratively remove the least important variables. After several iterations, the top 15 variables were selected based on their contribution to model accuracy. This method selected the top 15 most important the optimal set of features in determining the regional forest ACS estimates.

The top 15 variables deemed most influential were identified. Among the selected optimal set of features, there are 9 textural features and 4 vegetation indices, and the remaining 2 are terrain factors. This indicates that the optimal set of features is highly correlated with ACS. These key factors are visualized in Fig. 3 to illustrate their significance in our analysis.

Using the aforementioned 15 modeling parameters, an RF model was constructed and trained with the training set. The R² between the predicted ACS and the measured ACS based on the training set was 0.871, and the RMSE was 3.948 t/hm², indicating excellent model fitting performance. As shown in Table 1.

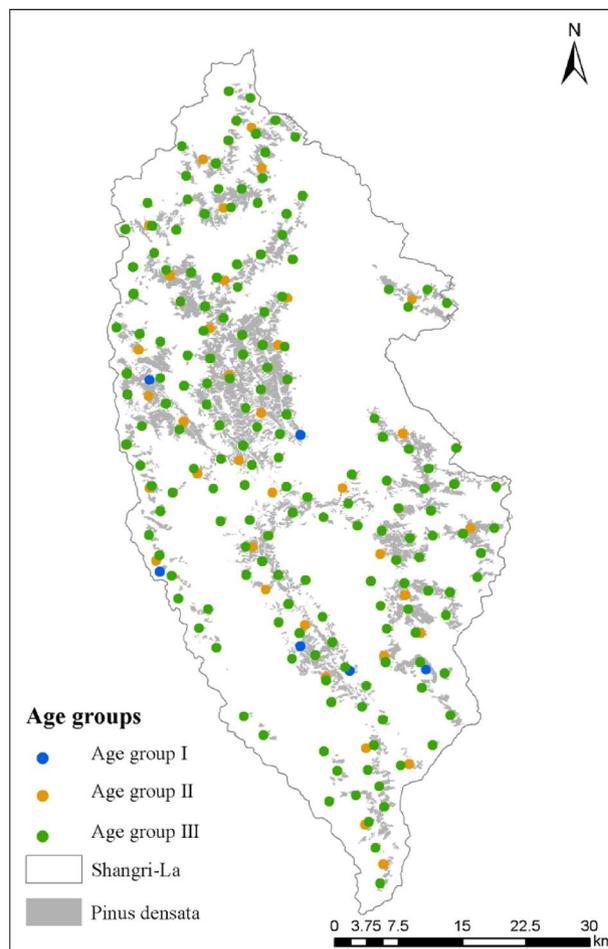


Fig. 2. Schematic diagram of stratified sampling results. The distribution map of *Pinus densata* was created by the author using ArcGIS 10.8 (<https://www.arcgis.com/>) based on Forest Management Inventory data.

Comparative analysis of MLRRK and RFRK interpolation

From Tables 2 and 3, we found that the Nugget effect and sum of squared residuals (SSR) of the Exponential model are smaller than the Spherical and Gaussian models. Additionally, the R^2 value of the Exponential model surpasses that of the Spherical and Gaussian models, indicating superior fitting performance. Consequently, the Exponential model chosen in this study is adopted as the final model for residual interpolation.

The performance of the MLRCK and RFCK Exponential models in Co-Kriging interpolation, using various covariates (slope, elevation, temperature, and precipitation), was assessed. The results indicate that the RFCK Exponential model consistently achieves higher fitting accuracy than the MLRCK Exponential model. Notably, when temperature and precipitation are included, the R^2 value of the RFCK Exponential model increases, and the SSR value decreases, indicating enhanced precision and predictive capability. When slope and elevation were incorporated as covariates in interpolation, accuracy improved; however, the enhancement was less pronounced than when temperature and precipitation were included. The study suggests that incorporating mean annual temperature and precipitation into residual kriging interpolation helps reduce the occurrence probability of extreme values and, to some extent, mitigates the tendency to underestimate low values and overestimate high values.

Comparison of model accuracy

Figure 4b exhibited an RI value of 0.06, with an increase in R^2 from 0.53 to 0.55 and a decrease in RMSE from 12 to 11.3 t/hm². Figure 4c exhibited an RI value of 0.13, an increase in R^2 from 0.53 to 0.58, and a decrease in RMSE from 12 to 10.4 t/hm². Similarly, Fig. 4d exhibited an RI value of 0.23, with an increase in R^2 from 0.53 to 0.63 and a decrease in RMSE from 12 to 9.3 t/hm². Overall, all three models exhibited improved accuracy compared to Fig. 4a, with Fig. 4b demonstrating the most significant improvement.

Residual kriging interpolation

From the residual kriging interpolation results, it is evident that the RFCK model (Fig. 5b,c) exhibits a broader residual predictive range compared to the RFOK model (Fig. 5a). Specifically, the RFCK_{pre} (Fig. 5c), which demonstrates a larger residual predictive range than the RFCK_{tem} (Fig. 5b). The residual prediction ranges are as

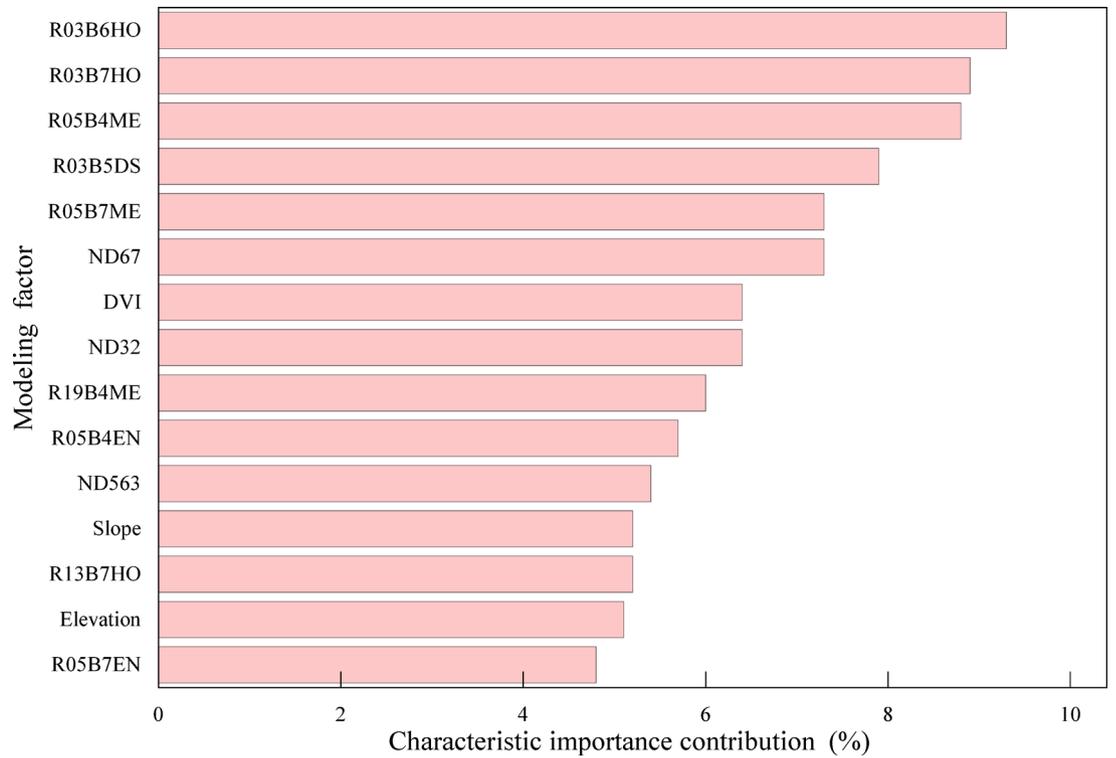


Fig. 3. Importance ranking of the optimal set of features for remote sensing. As in R03B6HO, R03 stands for a window size of 3×3 and B6 stands for the 7th band of Landsat.

Combination of characteristic variables	R ²	RMSE (t/hm ²)	P (%)	rRMSE (%)
Texture characterization (271)	0.856	4.106	90.2	10.4
Remote sensing element (24)	0.861	4.738	88.3	12.0
Remote sensing + Texture characteristics (48)	0.857	5.686	87.2	14.4
Remote sensing + Texture characteristics (15)	0.871	3.948	90.4	10.0

Table 1. Comparison of random forest regression evaluations modeled by multiple combinations of characteristic variables.

Kriging	Variational function model	Nugget effect	R ²	SSR
MLR Residual Ordinary Kriging, MLROK	Exponential	0.7584	0.477	52,600
	Spherical	0.8029	0.407	53,203
	Gaussian	0.8375	0.293	57,231
MLR Residual C0-Kriging _{pre} , MLRCK	Exponential	0.6973	0.586	45,172
	Spherical	0.7650	0.454	48,705
	Gaussian	0.8027	0.358	56,650
MLR Residual C0-Kriging _{tem} , MLRCK	Exponential	0.8544	0.505	46,264
	Spherical	0.9173	0.382	52,786
	Gaussian	0.9468	0.367	57,024
MLR Residual C0-Kriging _{slope} , MLRCK	Exponential	0.7851	0.487	47,595
	Spherical	0.8066	0.396	53,313
	Gaussian	0.8547	0.451	48,635
MLR Residual C0-Kriging _{elev} , MLRCK	Exponential	0.7991	0.497	46,698
	Spherical	0.8256	0.391	53,061
	Gaussian	0.859	0.435	53,261

Table 2. MLR residual kriging variational function fitting model and its parameters.

Kriging	Variational function model	Nugget effect	R ²	SSR
RF Residual Ordinary Kriging, RFOK	Exponential	0.4834	0.645	42,186
	Spherical	0.5718	0.594	49,674
	Gaussian	0.6625	0.505	55,873
RF Residual C0-Kriging _{pre} , RFCK	Exponential	0.4587	0.675	41,612
	Spherical	0.5825	0.599	49,185
	Gaussian	0.6357	0.539	55,724
RF Residual C0-Kriging _{tem} , RFCK	Exponential	0.4723	0.658	41,737
	Spherical	0.6178	0.559	49,379
	Gaussian	0.6760	0.491	55,870
RF Residual C0-Kriging _{slope} , RFCK	Exponential	0.4756	0.651	41,867
	Spherical	0.6035	0.614	50,291
	Gaussian	0.6712	0.501	55,943
RF Residual C0-Kriging _{elev} , RFCK	Exponential	0.4823	0.649	42,158
	Spherical	0.5970	0.584	49,662
	Gaussian	0.6627	0.504	55,896

Table 3. RF residual kriging variance function fitting model and its parameters.

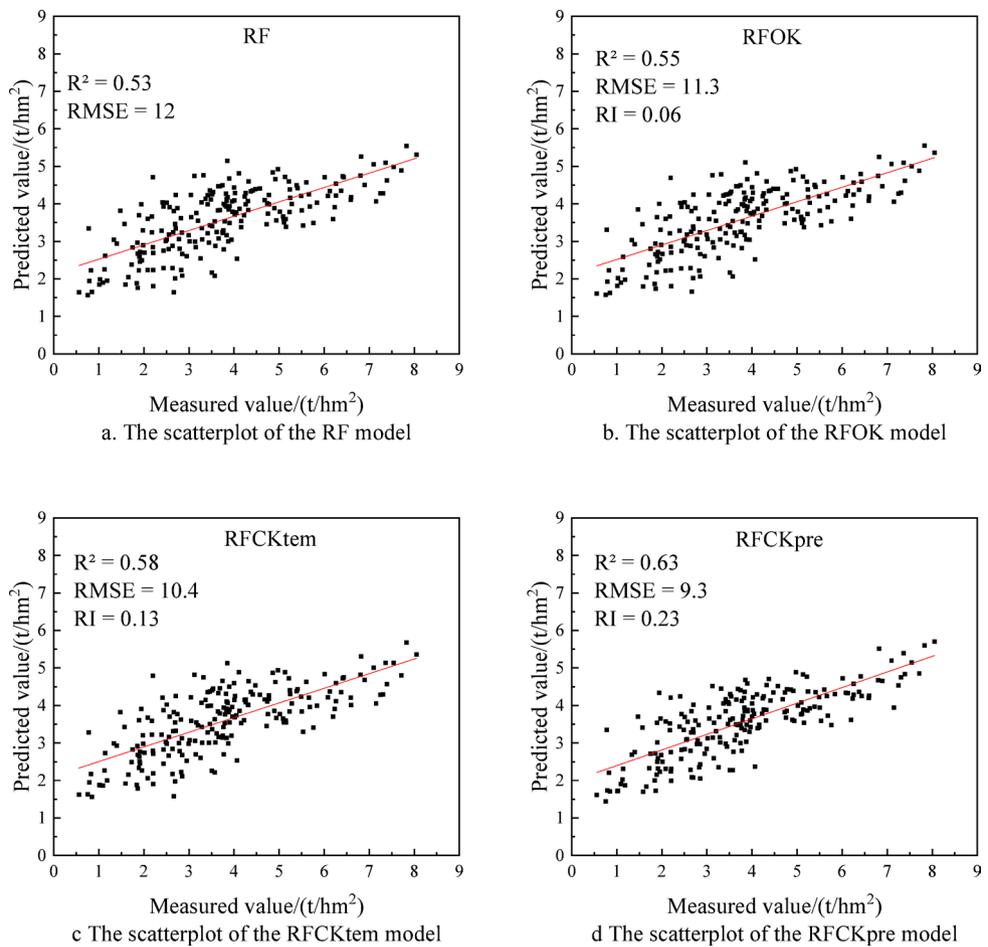


Fig. 4. Model fitting effect.

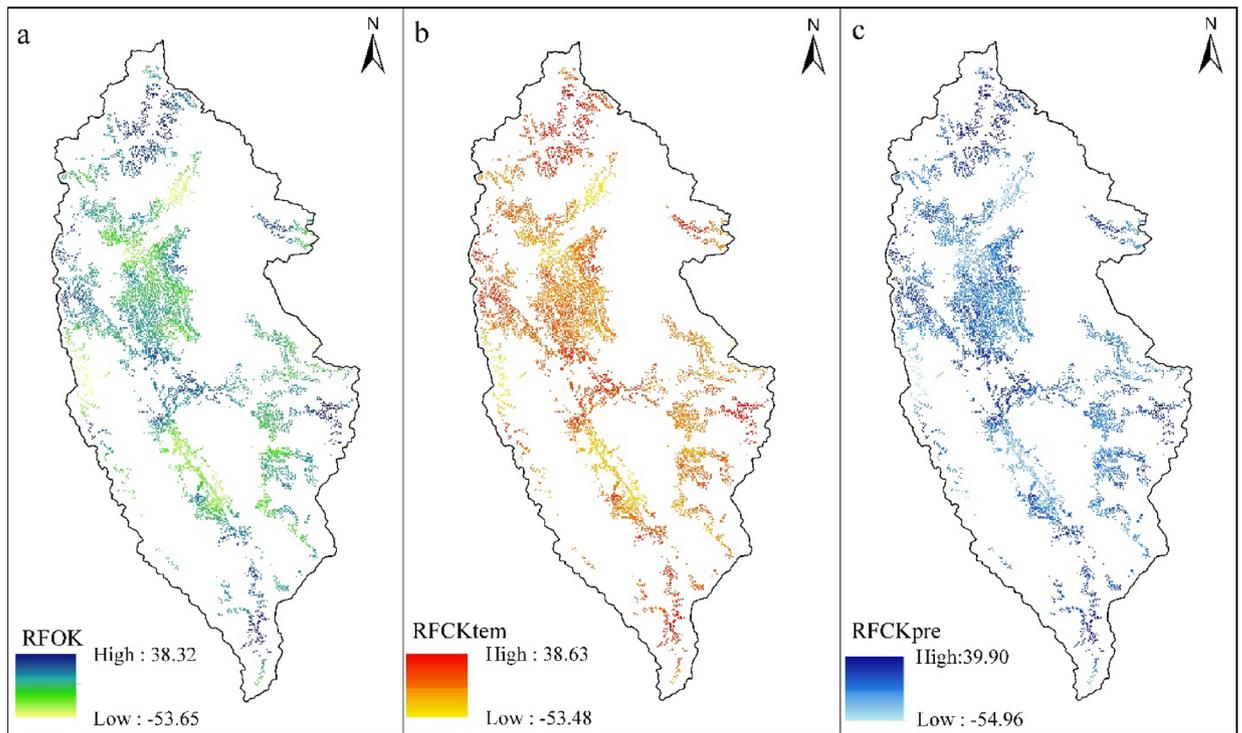


Fig. 5. Residual Kriging interpolation. The map was created by the author using ArcGIS 10.8 (<https://www.arcgis.com/>).

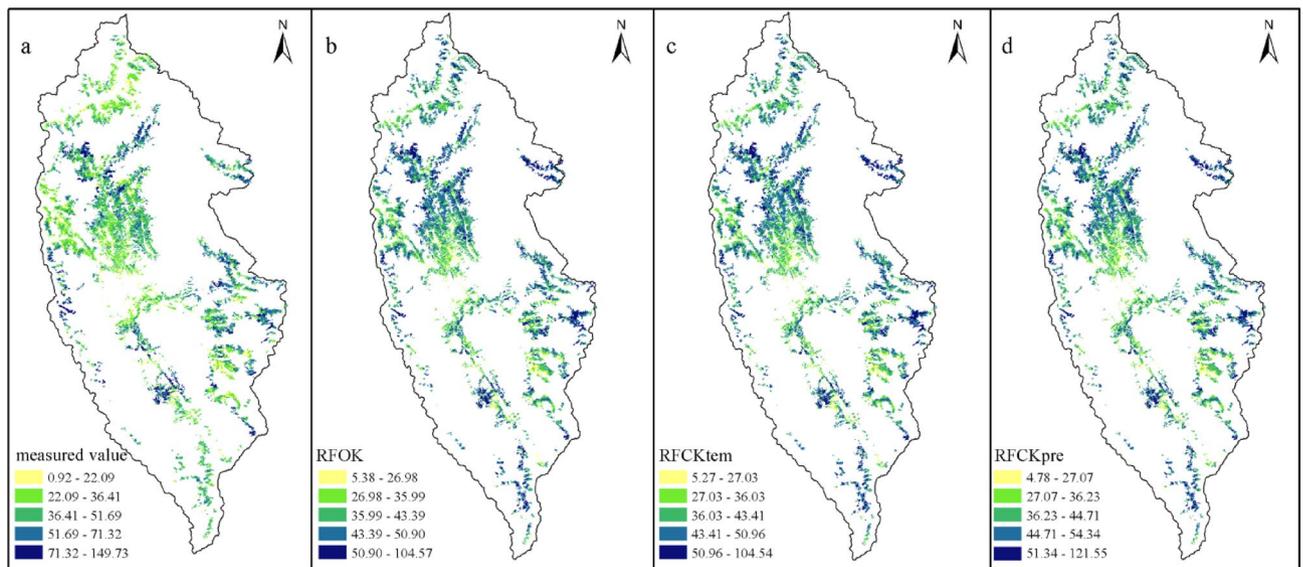


Fig. 6. Map of forest ACS. The map was created by the author using ArcGIS 10.8 (<https://www.arcgis.com/>).

follows: the residual range value of the RFOK model is from -53.65 to 38.32 t/hm^2 , the RFCK_{tem} model is from -53.48 to 38.63 t/hm^2 , and the RFCK_{pre} model is from -54.96 to 39.9 t/hm^2 .

Mapping of ACS

Figure 6 presents maps that detail the distribution of ACS in *Pinus densata* based on Forest Management Inventory data. Figure 6a shows the measured values of *Pinus densata* from the Forest Management Inventory data. Figure 6b shows the results of the RFOK model. Figure 6c,d illustrate the results of the RFCK model, with annual mean temperature and precipitation as covariates, respectively.

Figure 6 shows that incorporating Kriging interpolation into ACS estimation is feasible, with results closely aligning with the measured ACS values. The range for the RFOK model is 5.38 to 104.57 t/hm², for the RFCK_{tem} model is 5.27 to 104.54 t/hm², and for the RFCK_{pre} model is 4.78 to 121.55 t/hm².

After correcting the ACS residuals using kriging interpolation, the resulting forest ACS map accurately reflects both lower and higher AGB values, effectively alleviating issues of overestimating low values and underestimating high values, thus improving the accuracy of forest ACS estimation. Among them, the RFCK_{pre} model exhibits a result range that is closest to the measured values.

Discussion

Influence of sampling method selection on mapping accuracy

The choice of sampling method could significantly impact the accuracy of ACS thematic maps. Scholars found that stratified sampling offers higher accuracy with smaller sample sizes compared to systematic sampling methods^{40,41}. This underscores the critical importance of selecting appropriate sampling strategies for ensuring accurate mapping of ACS.

Research has indicated that an optimal age group structure supports sustainable forest management and ongoing utilization⁴². Large-scale afforestation is impractical because of space constraints, which advocates instead for near-natural regeneration principles to enhance forest quality through age group adjustments⁴³.

In this study, we employed age groups as the primary stratification criterion for sampling *Pinus densata* in Shangri-La. The stratified sampling approach focused solely on age groups without incorporating other stratification factors such as origin, geomorphology, or dominant tree species. Future studies could enhance sampling precision by incorporating additional stratification criteria into the sampling framework.

Impact of temperature and precipitation on model accuracy

Most studies on carbon storage in Shangri-La focus primarily on the influence of terrain factors^{44,45}, with limited consideration given to the impact of climate factors. Additionally, existing research suggests that changes in topographical factors are relatively small over longer time scales⁴⁶.

The spatial distribution and dynamics of ACS are significantly influenced by climatic factors, particularly temperature and precipitation⁴⁷. Climate change has a significant impact on forest ecosystems. Therefore, this study considers mean annual precipitation and temperature as covariates to explore the accurate estimation of ACS in *Pinus densata* forests. The results show that this approach provides relatively accurate estimates. Furthermore, the study indicates that including elevation and slope as covariates improves the accuracy of ACS estimation, although the precision is not optimal. This result is consistent with the findings of Luo Kai's study⁴⁸, further demonstrating that mean annual temperature and precipitation in the region play a positive role in enhancing the accuracy of forest ACS estimation.

Determination of the RFRK model for mapping

In this study, the RF model was constructed using 15 selected optimal feature sets through the RF-RFE method, achieving an RMSE of 3.948 t/hm², an R² of 0.871, and a P of 90.4%. These metrics indicate a strong model fit, high predictive accuracy, and good performance²⁴. The RF model is known for its ability to handle high-dimensional data, manage outliers effectively, and mitigate overfitting issues⁴⁹. However, the RF model may overlook the effects of spatial autocorrelation. To address this limitation, Kriging interpolation was employed, enhancing overall estimation accuracy by incorporating spatial autocorrelation patterns⁵⁰.

Combining the RF model with the Kriging method achieves higher mapping accuracy than using the RF model alone. Specifically, the RFCK model outperforms the RFOK and MLRRK models in terms of precision. Moreover, compared with the RFOK, the RFCK demonstrated superior accuracy. This outcome highlights the mitigation of the overestimation of low values and underestimation of high values in ACS estimation to some extent. This approach underscores the effectiveness of integrating machine learning models, such as RF, with spatial interpolation techniques, such as Kriging, to enhance predictive accuracy in ACS mapping studies. This study used the RFRK model to achieve regional ACS mapping with a relatively small dataset, whereas other non-parametric models may require larger datasets for accurate ACS estimation.

Furthermore, the RFRK method can correct model errors and improve estimation accuracy. Several researchers have employed random forest kriging/co-kriging methods along with ICESat-2/ATLAS data or Landsat data for AGB mapping and biomass estimation^{32,34,36}. Their findings consistently show that the method can effectively mitigate the underestimation of high values and the overestimation of low values, highlighting its significant potential in ACS estimation and mapping. This is consistent with the outcomes observed in the present study. In the future, additional models could be considered in combination with Kriging for ACS estimation.

In this study, the RF model achieved an R²=0.871 and a P of 90.4%. The RFCK_{pre} exponential model demonstrated the highest fitting accuracy, with R²=0.63. In a similar study⁵¹ conducted in Shangri-La City, the RF model achieved a P of 88.3%, with the optimal model being the spherical model, which exhibited a fitting accuracy of R²=0.65. These results are highly consistent with those of the present study, indicating that the method for estimating the ACS of *Pinus densata* is feasible. However, since only a single data source was used in this study, the improvement in accuracy was somewhat limited. Future work will explore the use of multi-source remote sensing data for estimating the ACS of *Pinus densata*.

Conclusions

This study is based on Landsat 8 OLI data and Forest Management Inventory Data, incorporating elevation, slope, precipitation, and temperature as covariates. The residual kriging interpolation was performed using an exponential model, and the accuracies of the MLRRK and RFRK models were compared. The results indicate

that (1) Residual interpolation incorporating annual precipitation and temperature as covariates produces more favorable results than interpolation using elevation and slope; (2) the accuracy of the RFRK model outperforms that of the MLRRK model. These findings provide valuable insights into the impact of climatic factors on ACS in Shangri-La. Furthermore, future studies will consider incorporating additional climatic factors to further improve estimation accuracy.

Methods

Study area

Shangri-La is situated in northwestern Yunnan Province (99° 20' ~ 100° 19' E, 26° 52' ~ 28° 52' N), covering an area of 11,613 km²⁵². Due to its location in a high-altitude, low-latitude zone, the climate varies with increasing altitude. The region experiences distinct wet and dry seasons, characterized by wet summers and autumns and dry winters and springs. The average annual temperature is 5.4 °C, with annual rainfall ranging from 268 to 945 mm. The frost-free period spans from 129 to 197 days. According to the "Vegetation of Yunnan Province" classification standard, Shangri-La hosts ten vegetation types. Dominant forest types include *Pinus densata*, *Quercus aquifolioides*, and *Abies ferreana*^{53,54}. The target species for this study is the *Pinus densata*, which includes young, middle-aged, near-mature, mature, and over-mature stands forest, including natural and artificial forests distributed within the study area. The location and climate information map of the study area is shown in Fig. 7.

Collection and processing of remote sensing images

The data utilized in this study contains Landsat 8 OLI images obtained from the Geospatial Data Cloud (<https://www.gscloud.cn/>). Three scenes with minimal cloud cover from 2016 were selected, as shown in Table 4, and preprocessing was carried out using ENVI 5.6 (<https://envi.geoscene.cn/>) and ArcGIS 10.8 (<https://www.arcgis.com/>) software. Radiometric calibration and atmospheric correction were applied to mitigate the effects of terrain and aerosols on the surface reflectance⁵⁵. Images composites were generated to fill cloud gaps and reduce the data volume required for periodic forest cover monitoring⁵⁶.

Collection and processing of forest management inventory data

The field data employed in this study is the Forest Management Inventory. The dataset employed in this study is tree stand-based, focusing on the general characteristics of forest stands. This dataset includes key attributes

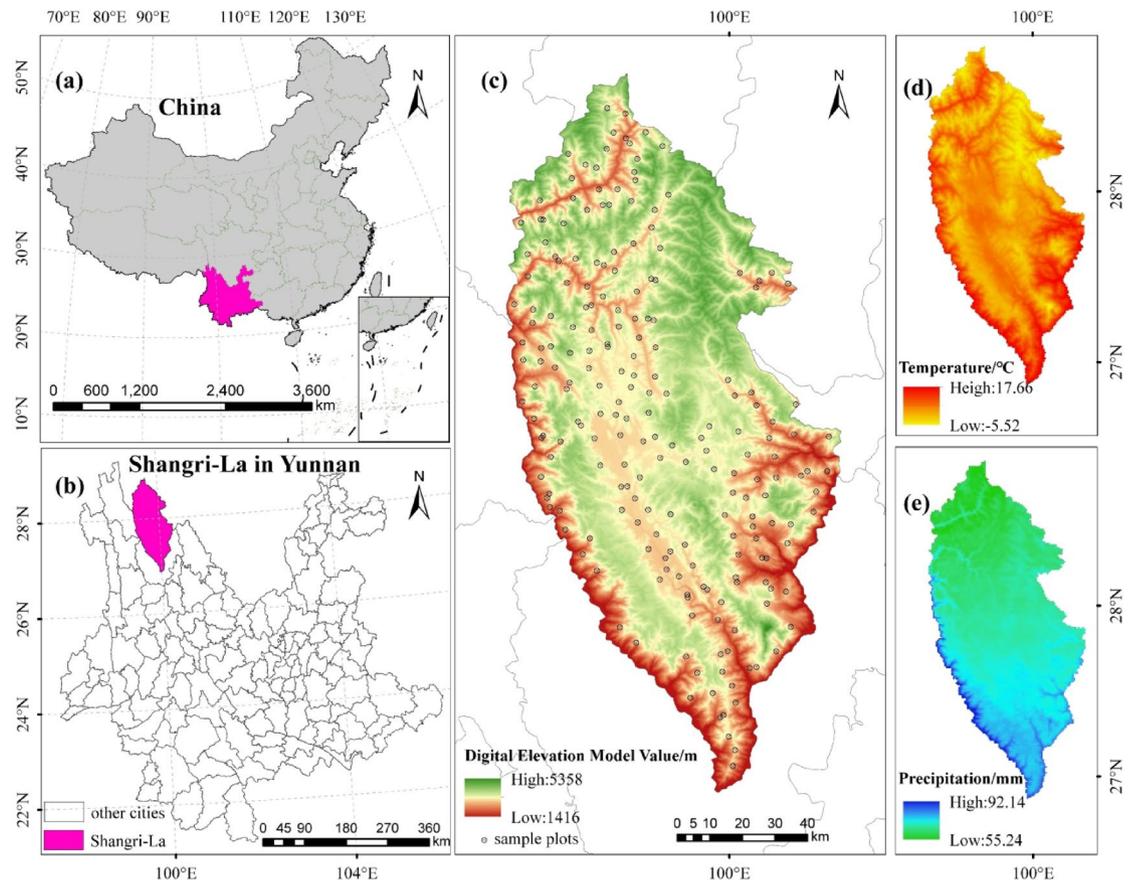


Fig. 7. Study area (a) Location of Yunnan; (b) Shangri-La in Yunnan; (c) Sample location in Shangri-La; (d) Distribution of Temperature; (e) Distribution of Precipitation. The map was created by the author using ArcGIS 10.8 (<https://www.arcgis.com/>).

Investigation year	ID	Date	Path	Row	Sun Elevation	Sun azimuth	Cloud cover (%)
2016	LC81310412016325LGN00	2016-11-20	131	41	39.353	156.697	0.4
	LC81320402016348LGN00	2016-12-13	132	40	34.220	156.677	0.73
	LC81320412016348LGN00	2016-12-13	132	41	35.443	156.001	0.76

Table 4. Provides essential details about the remote sensing images used in this study.

Number of strata	Age groups	ACS in <i>Pinus densata</i> (t/hm ²)			
		Mean	Max	Min	Number of sample units
I	Young	11.34	13.56	7.98	6
II	Intermediate	25.44	84.37	9.88	36
III	Near-mature mature	43.5	92.25	6.69	168
Total	Over-mature	26.76	92.25	6.69	210

Table 5. Distribution of ACS in the sample.

such as tree height (H), diameter at breast height (DBH), stocking density, stand tree count, small group size, age classes, and topographic characteristics. These data were collected from various forest regions using standardized survey techniques, offering a comprehensive perspective for forest management and ecological research⁵⁷.

The forest biomass expansion factor (BEF) method was employed in this study to convert stand stock to biomass, using the formula shown in Eq. (1). Specifically, a BEF of 1.650 was applied for *Pinus densata*, with a specific volume density (SVD) of 0.413⁵⁸. Forest ACS is typically calculated by multiplying biomass by the carbon content (C_c) factor⁵⁹. Different forest types have varying carbon conversion coefficients, which are often difficult to obtain. Internationally, a carbon conversion coefficient of 0.5 is commonly used⁶⁰. Therefore, in this study, we also adopt a carbon conversion coefficient of 0.5 to calculate the carbon stock of the *Pinus densata* forests in Shangri-La City. As shown in Eq. (2).

$$B = V \times SVD \times BEF \quad (1)$$

$$C = B \times C_c \quad (2)$$

where⁶¹ B is forest aboveground biomass, V is stand stock, SVD is wood density, BEF is forest biomass conversion factor, C is *Pinus densata* ACS, C_c is carbon conversion factor.

Collection and processing of digital elevation model data

The Digital Elevation Model (DEM) data based on ASTER utilized in this study were sourced from the Geospatial Data Cloud (<https://www.gscloud.cn/>). This dataset was employed to extract elevation, slope, and aspect parameters essential for the research. The DEM images underwent necessary preprocessing, with a spatial resolution of 30 m.

Collection and processing of annual average temperature and precipitation data

The annual mean temperature and precipitation data used in this study were acquired from the National Tibetan Plateau Scientific Data Center (<https://data.tpdc.ac.cn/>). These datasets are comprehensive, providing month-by-month information at a 1-km resolution across China.

Stratified sampling

Existing studies have clearly demonstrated that forests ACS increases with the continuous growth of stand age. Accordingly, the present study adopts stand age as the sole criterion for implementing stratified sampling⁶². The age of *Pinus densata* stands was categorized into three strata based on its correlation with ACS values obtained by stratified sampling, enabling targeted data collection across different age groups, which are shown in Table 5. With a 95% confidence level, the sampling accuracy was set to 95%, and the sampling unit size was 30 × 30 m, which is consistent with the resolution of Landsat 8 OLI remote sensing images. A total of 210 sample units were selected, as shown in Eq. (3)⁶¹.

$$n_0 = \frac{t^2 \left(\sum W_i \sigma_i \right)^2}{E^2 \left(\sum W_i \bar{y}_i \right)^2} \quad (3)$$

where n_0 is the total number of sample units, W_i is the proportion of overall units in layer i, σ_i is the standard deviation of layer i, \bar{y}_i is the mean value of layer i, t is a reliability indicator, E is the relative error.

Remote sensing variable extraction

Previous studies²¹ have demonstrated a strong correlation between forest ACS and the optimal set of remote sensing features. In this study, a total of 527 optimal features were extracted, including 524 remote sensing

factors and 3 environmental factors. Correlation analysis using SPSS 27 (<https://www.ibm.com/cn-zh/products/spss-statistics>) software identified 271 optimal features that exhibited significant correlations with ACS, as detailed in Table 6.

Kriging

In this study, mean annual temperature and mean annual precipitation were used as covariates. Exponential, Spherical, and Gaussian⁶³ functions were employed to fit models for ordinary residual Kriging interpolation and residual co-Kriging interpolation. The nugget effect, R^2 , and the SSR were used as evaluation metrics. The nugget effect represents the ratio of the nugget value to the sill value, which describes the strength of spatial autocorrelation. The smaller the nugget effect, the stronger the spatial autocorrelation⁶⁴. SSR represents the sum of the squared differences between the measured values and the predicted values. A higher R^2 and lower SSR indicate better fitting performance, allowing for the selection of the best-fitting model³⁶.

$$y(h) (Exponential) = \begin{cases} 0, & h = 0 \\ C_0 + C \left(1 - e^{-\frac{h}{a}}\right) & h > 0 \end{cases} \quad (4)$$

$$y(h) (Spherical) = \begin{cases} 0, & 0 \\ C_0 + C \left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right), & 0 < h \leq a \\ C_0 + C & h > 0 \end{cases} \quad (5)$$

$$y(h) (Gaussian) = \begin{cases} 0, & h = 0 \\ C_0 + C \left(1 - e^{-\frac{h^2}{a^2}}\right) & h > 0 \end{cases} \quad (6)$$

Variable type	Variable name	Definition
Terrain factors	Elevation	Extraction with DEM data
	Slope	
	Aspect	
Vegetation indices	DVI	NIR-Red
	EVI	$2.5 \times (\text{NIR} - \text{Red}) / (\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1)$
	B4/B2	Red/Blue
	ND32	$(\text{Green} - \text{Blue}) / (\text{Green} + \text{Blue})$
	ND43	$(\text{Red} - \text{Green}) / (\text{Red} + \text{Green})$
	ND57	$(\text{NIR} - \text{SWIR2}) / (\text{NIR} + \text{SWIR2})$
	ND67	$(\text{SWIR1} - \text{SWIR2}) / (\text{SWIR1} + \text{SWIR2})$
	ND563	$(\text{NIR} + \text{SWIR1} - \text{Green}) / (\text{NIR} + \text{SWIR1} + \text{Green})$
	NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
	SAVI	$1.5 \times (\text{NIR} - \text{Red}) / 8 \times (\text{NIR} + \text{Red} + 0.5)$
	SR	NIR/Red
Raw single-band factors	B2	B2 (Blue)
	B3	B3 (Green)
	B4	B4 (Red)
	B5	B5 (NIR)
	B6	B6 (SWIR1)
	B7	B7 (SWIR2)
Information enhancement factors	Albedo MID(57) VIS234	Blue + Green + Red + NIR + SWIR1 + SWIR2
		NIR + SWIR2
		Blue + Green + Red
Textural features	CO	Correlation (CO)
	CN	Contrast (CN)
	DS	Dissimilarity (DS)
	EN	Entropy (EN)
	HO	Homogeneity (HO)
	VA	Variance (VA)
	ME	Mean (ME)
	SM	Second Moment (SM)

Table 6. Remote Sensing Variables. The windows for the textural features in the table are 3×3 , 5×5 , 7×7 , 9×9 , 11×11 , 13×13 , 15×15 , 17×17 , 19×19 .

where C_0 is the nugget constant, a is the range, C is the peak height, $C_0 + C$ is the sill value, and h is the distance between any two points.

$$SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{7}$$

where y_i represents the measured values, \hat{y}_i represents the predicted values, and n is the sample size.

Residual values

The residual is the difference between the measured values and the predicted values. Both Fig. 8a,b illustrate that these residuals approximate a normal distribution. This conformity indicates that the residual values of ACS, as derived from both MLR and RF models, meet the necessary assumptions for kriging interpolation.

Establishment of MLRRK model

MLR analyzes correlations between multiple independent variables and a dependent variable to develop a predictive model. The MLR equation is expressed as follows⁶⁵:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n \tag{8}$$

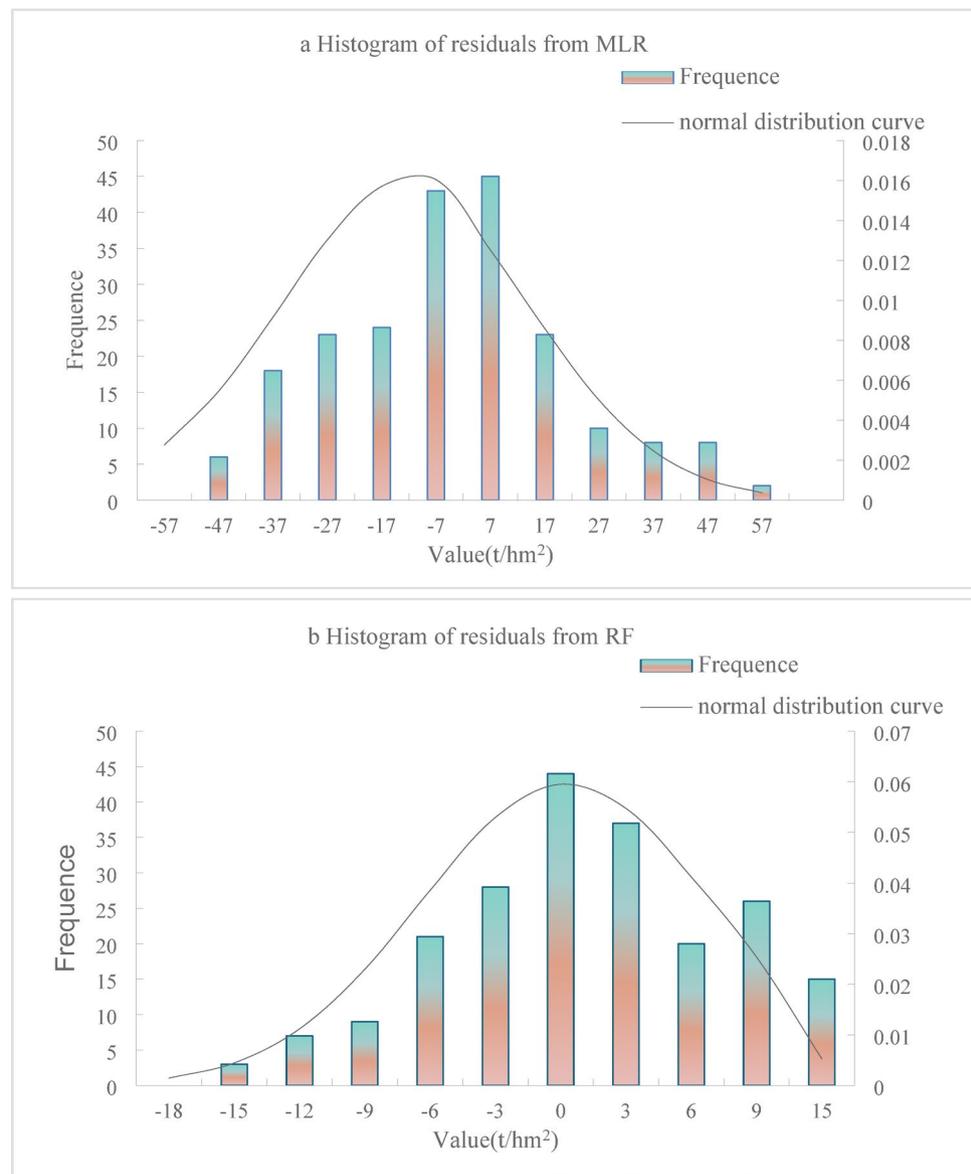


Fig. 8. Histogram of residual distribution.

Regression equation	R ²	F	Significance
C = 0.004*R03B6HO - 0.005*R05B4ME + 3.554	0.169	22.263	0.001

Table 7. Regression model and their associated parameters.

where Y represents the ACS at the sample site, α_0 represents constant, $\alpha_1, \alpha_2, \alpha_n$ are the regression coefficients, and X_1, X_2, X_n are the independent variables (Selected in this paper are R03B6HO, R05B4ME).

Given the stochastic nature of the data, the MLR model may not fully capture all information, which leaves residual information in the data. Therefore, this study developed MLR residual Kriging models aimed at correcting model errors and improving estimation accuracy.

Using SPSS 27 software, the characteristics of the 15 selected optimal features identified by Random Forest were analyzed using MLR equations. The entry and removal probabilities were set to 0.05 and 0.10, respectively. This process yielded the regression equations for the model, along with R², F-statistic, and significance values, as detailed in Table 7.

The steps for constructing the MLRRK model are the same as those for the RFRK model, as described in the next section.

Establishment of RFRK model

RF generates numerous random classification trees, which can select the final result based on the most frequently occurring tree^{66,67}. When constructing these trees, RF randomly samples n observations from the original dataset using the Bootstrap method, where n represents the number of observations for the dependent variable Y and i denotes its independent variables⁶⁸.

Random Forest, as noted by³⁵, is a non-spatial approach. However, it may struggle to account for spatial autocorrelation, leading to issues such as underestimating high values and overestimating low values of ACS. To address this, spatial prediction models are necessary. In this study, we employed both ordinary kriging and co-kriging for interpolation.

The steps are as follows:

- (1) RF Modeling: Random Forest (RF) modeling was initially employed to obtain the predicted value of ACS.
- (2) Residual Calculation: The residuals were calculated by comparing the measured values with the predicted ACS values, as shown in Eq. (6).

$$R(Xi) = ACS_{measured} - ACS_{predicted} \quad (9)$$

where R(Xi) is the residual value.

- (3) Normal Distribution Test: The residuals were tested for normal distribution, and the raw data of these residuals followed a normal distribution (Fig. 8b). The residuals of the RF model were subsequently interpolated using kriging.
- (4) Final Adjustment: The residuals were added to the predicted ACS values to adjust the estimation, as shown in Eq. (7). The adjusted results were then compared with the measured values to identify the optimal model.

$$C_i = ACS_{predicted} + R(Xi) \quad (10)$$

where C_i is the ACS value, and R(Xi) is the residual value.

- (5) Mapping ACS: ACS maps were generated for each model and compared with the measured values from the Forest Management Inventory data to identify the model that most closely approximates the true values.

Accuracy assessment

To evaluate model performance, 80% of the data was allocated for training, while the remaining 20% was used for validation. Several evaluation indexes were used in this study to evaluate the predictive accuracy of the ACS model and compare the predicted values with the measured values. These indexes include the coefficient of determination (R²), Relative Improvement (RI), relative Root Mean Square Error (rRMSE), prediction accuracy (P), and Root Mean Squared Error (RMSE). The RI is used to evaluate the improvement of the RFOK and RFCK models compared to the RF model; a higher RI value indicates a more significant improvement in the model. The P reflects the average predictive ability of the model, with higher values indicating better predictive performance. These metrics assess both the prediction accuracy and validation accuracy of the ACS model. The specific evaluation indexes and their formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

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

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

$$P = \frac{1}{n} \sum_{i=1}^n \left(1 - \left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right|\right) \times 100\% \quad (14)$$

$$RI = \frac{RF_{RMSE} - R_{FOK} / RF_{CK_{RMSE}}}{RF_{RMSE}} \quad (15)$$

Implementation

In this study, Landsat images were pre-processed and remote sensing factors were extracted using ArcGIS 10.8 (<https://www.arcgis.com/>) and ENVI 5.6 (<https://envi.geoscene.cn/>). SPSS 27 (<https://www.ibm.com/cn-zh/products/spss-statistics>) was used for Pearson analysis of remote sensing factors. Furthermore, in order to establish the RF model, Anaconda3 (<https://www.anaconda.com/>) was used to build a Python 3.7 environment.

Data availability

The Landsat OLI data and DEM data are available through <https://www.gscloud.cn/> (accessed on 6 November 2024), the meteorological data are available through <https://data.tpdc.ac.cn/home> (accessed on 6 November 2024). Forest Management Inventory data presented in this study are available on request from the corresponding author; the data are not publicly available due to the confidentiality of the dataset.

Received: 6 November 2024; Accepted: 13 May 2025

Published online: 20 May 2025

References

- Chang, J. & Huang, C. Three decades of spatiotemporal dynamics in forest biomass density in the Qinba Mountains. *Ecol. Inform.* **81**, 102566. <https://doi.org/10.1016/j.ecoinf.2024.102566> (2024).
- Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993. <https://doi.org/10.1126/science.1201609> (2011).
- Jiang, F. et al. Retrieving the forest aboveground biomass by combining the red edge bands of Sentinel-2 and GF-6. *Acta Ecol. Sin.* **41**, 8222–8236. <https://doi.org/10.5846/stxb202012173204> (2021).
- Wu, S. et al. Study on forest carbon storage and spatial distribution in the alpine gorge region of northwest Sichuan: Take Sichuan Luoxu nature reserve as an example. *Ecol. Environ. Sci.* **31**, 1735–1744. <https://doi.org/10.16258/j.cnki.1674-5906.2022.09.003> (2022).
- Zhang, Y., Li, X. & Wen, Y. Forest carbon sequestration potential in China under the background of carbon emission peak and carbon neutralization. *J. Beijing Forest. Univ.* **44**, 38–47. <https://doi.org/10.12171/j.1000-1522.20210143> (2022).
- Hu, Z. & Su, J. Forest vegetation carbon stock research review and perspectives. *Agric. Technol.* **42**, 58–62. <https://doi.org/10.19754/j.nyyjs.20220630014> (2022).
- Fararoda, R. et al. Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm. *Ecol. Inform.* **65**, 101392. <https://doi.org/10.1016/j.ecoinf.2021.101392> (2021).
- Chen, K., Zhang, H., Zhang, B. & He, Y. Spatial distribution of carbon storage in natural secondary forest based on geographically weighted regression expansion model. *Chin. J. Appl. Ecol.* **32**, 1175–1183. <https://doi.org/10.13287/j.1001-9332.202104.002> (2021).
- Avitabile, V. & Camia, A. An assessment of forest biomass maps in Europe using harmonized national statistics and inventory plots. *For. Ecol. Manage.* **409**, 489–498. <https://doi.org/10.1016/j.foreco.2017.11.047> (2018).
- Liu, Y., Gao, X., Fu, C., Yu, G. & Liu, Z. Estimation of carbon sequestration potential of forest biomass in China based on national forest resources inventory. *Acta Ecol. Sin.* **39**, 4002–4010. <https://doi.org/10.5846/stxb201805071016> (2019).
- Bui, Q. T. et al. Hybrid machine learning models for aboveground biomass estimations. *Ecol. Inform.* **79**, 102421. <https://doi.org/10.1016/j.ecoinf.2023.102421> (2024).
- Prakash, A. J., Behera, M. D., Ghosh, S. M., Das, A. & Mishra, D. A new synergistic approach for Sentinel-1 and PALSAR-2 in a machine learning framework to predict aboveground biomass of a dense mangrove forest. *Ecol. Inform.* **72**, 101900. <https://doi.org/10.1016/j.ecoinf.2022.101900> (2022).
- Du, X., Wang, J., Bai, Y., Du, Z. & Meng, J. Development of thematic map of forest carbon storage based on kriging interpolation method in Cili County, Hunan Province. *J. Northwest Forest. Univ.* **37**, 198–204. <https://doi.org/10.3969/j.issn.1001-7461> (2022).
- Guo, Q. et al. Combining GEDI and sentinel data to estimate forest canopy mean height and aboveground biomass. *Ecol. Inform.* **78**, 102348. <https://doi.org/10.1016/j.ecoinf.2023.102348> (2023).
- Reis, A. A. D. et al. Temporal stability of stratifications using different dendrometric variables and geostatistical interpolation. *Ciencia Florestal* **32**, 102–121. <https://doi.org/10.5902/1980509843274> (2022).
- Wu, H., Liu, L., Lu, C. & Wang, B. Sampling optimization of stand biomass survey based on regional characteristics analysis. *Forest Grassland Resour. Res.*, 57–63. <https://doi.org/10.13466/j.cnki.lyzygl.2023.02.008> (2023).
- Xu, M. et al. Integrating ward's clustering stratification and spatially correlated poisson disk sampling to enhance the accuracy of forest aboveground carbon stock estimation. *Forests* **15**, 2111 (2024).
- Cui, J., Chen, J. & Cui, W. Research about residuals in linear regression analysis. *J. Mudanjiang Univ.* **29**, 84–88. <https://doi.org/10.15907/j.cnki.23-1450.2020.10.018> (2020).
- Ahmad, A., Gilani, H. & Ahmad, S. R. Forest Aboveground Biomass Estimation and Mapping through High-Resolution Optical Satellite Images-A Literature Review. *Forests* **12**, <https://doi.org/10.3390/f12070914> (2021).
- Xu, Z., Li, Y., Li, M., Li, C. & Wang, L. Forest biomass retrieval based on Sentinel-1A and Landsat 8 image. *J. Central South Univ. Forest. Technol.* **40**, 147–155. <https://doi.org/10.14067/j.cnki.1673-923x.2020.11.018> (2020).
- Zhang, G. Spatial distribution characteristics of carbon storage of urban forests in shanghai based on remote sensing estimation. *Ecol. Environ. Sci.* **30**, 1777–1786. <https://doi.org/10.16258/j.cnki.1674-5906.2021.09.001> (2021).
- Wang, S., Shi, J., Yin, S. & Wu, M. In *Introduction to Linear Models* (ed Daqian Li) Ch. 121–127, pp. 121–127 (Science Press, 2004).
- 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* **8**, 128124–128139. <https://doi.org/10.1109/access.2020.3008686> (2020).
- Avina-Hernandez, J., Ramirez-Vargas, M., Roque-Sosa, F. & Martinez-Rincon, R. O. Predictive performance of random forest on the identification of mangrove species in arid environments. *Ecol. Inform.* **75**, 102040. <https://doi.org/10.1016/j.ecoinf.2023.102040> (2023).
- Cheng, F., Ou, G., Wang, M. & Liu, C. Remote sensing estimation of forest carbon stock based on machine learning algorithms. *Forests* **15**, 681. <https://doi.org/10.3390/f15040681> (2024).

26. Effendi, N. A. F. N. et al. Unlocking the potential of hyperspectral and LiDAR for above-ground biomass (AGB) and tree species classification in tropical forests. *Geocarto Int.* **37**, 8036–8061. <https://doi.org/10.1080/10106049.2021.1990419> (2022).
27. Hoover, C. M., Ducey, M. J., Colter, R. A. & Yamasaki, M. Evaluation of alternative approaches for landscape-scale biomass estimation in a mixed-species northern forest. *For. Ecol. Manage.* **409**, 552–563. <https://doi.org/10.1016/j.foreco.2017.11.040> (2018).
28. Zimbres, B. et al. Mapping the stock and spatial distribution of aboveground woody biomass in the native vegetation of the Brazilian Cerrado biome. *Forest Ecol. Manage.* **499**, 119615. <https://doi.org/10.1016/j.foreco.2021.119615> (2021).
29. An Thi Ngoc, D. et al. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. *Ecol. Inform.* **50**, 24–32. <https://doi.org/10.1016/j.ecoinf.2018.12.010> (2019).
30. Sekulic, A., Kilibarda, M., Heuvelink, G. B. M., Nikolic, M. & Bajat, B. Random forest spatial interpolation. *Remote Sens.* **12**, 1687. <https://doi.org/10.3390/rs12101687> (2020).
31. Zhou, Y., Xie, B. & Li, M. Mapping regional forest aboveground biomass from random forest Co-Kriging approach: A case study from north Guangdong. *J. Nanjing Forest. Univ.* **48**, 169–178. <https://doi.org/10.12302/j.issn.1000-2006.202202015> (2024).
32. Yu, J. et al. Estimation of forest canopy cover by combining ICESat-2/ATLAS data and geostatistical method/co-kriging. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, 1824–1838. <https://doi.org/10.1109/JSTARS.2023.3340429> (2023).
33. Aullo-Maestro, I. et al. Integration of field sampling and LiDAR data in forest inventories: Comparison of area-based approach and (lognormal) universal kriging. *Ann. Forest Sci.* **78**, 1–14. <https://doi.org/10.1007/s13595-021-01056-1> (2021).
34. Song, H., Xi, L., Shu, Q., Wei, Z. & Qiu, S. Estimate forest aboveground biomass of mountain by ICESat-2/ATLAS data interacting cokriging. *Forests* **14**, 13. <https://doi.org/10.3390/f14010013> (2023).
35. Fayad, I. et al. Regional scale rain-forest height mapping using regression-kriging of spaceborne and airborne LiDAR data: Application on French Guiana. *Remote Sens.* **8**, 240. <https://doi.org/10.3390/rs8030240> (2016).
36. Su, H., Shen, W., Wang, J., Ali, A. & Li, M. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *Forest Ecosyst.* **7**, 1–20. <https://doi.org/10.1186/s40663-020-00276-7> (2020).
37. Guo, P.-T. et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma* **237**, 49–59. <https://doi.org/10.1016/j.geoderma.2014.08.009> (2015).
38. Cao, J. et al. Estimation and uncertainty analysis of aboveground carbon storage of *Pinus densata* based on random forests and Monte Carlo. *For. Res.* **36**, 131–139 (2023).
39. Liao, Y., Zhang, J., Bao, R. & Xu, D. Estimating the dynamic changes of aboveground biomass of *Pinus densata* based on Landsat. *J. Southwest Forest. Univ.* **43**, 117–125. <https://doi.org/10.11929/j.swfu.202111058> (2023).
40. Tian, Z., Yan, Y. & Zhang, C. Comparative study on sampling design of forest biomass based on forest management inventory data. *Forest Invent. Plan.* **48**, 9–15 (2023).
41. Wang, R., Yang, Q., Ou, G. & Xu, H. Forest biomass investigation design using stratified sampling in Simao District. *Forest Grassland Resour. Res.*, 197–202. <https://doi.org/10.13466/j.cnki.lyzygl.2021.01.025> (2021).
42. Feng, D., Wang, S., Zhao, Y. & Feng, Z. Analysis of changes in china's forest resources and age group optimization. *J. Southwest Forest. Univ. (Nat. Sci.)* **43**, 122–131. <https://doi.org/10.11929/j.swfu.202206052> (2023).
43. Lu, Y. et al. From normal forest to close-to-nature forest: Multi-functional forestry and its practice at national, regional and forest management unit levels in Germany. *World Forest. Res.* **23**, 1–11. <https://doi.org/10.13348/j.cnki.sjlyyyj.2010.01.004> (2010).
44. Liao, Y., Zhang, J., Bao, R., Xu, D. & Han, D. Modelling the dynamics of carbon storages for *Pinus densata* using landsat images in Shangri-La considering topographic factors. *Remote Sens.* **14**, 6244. <https://doi.org/10.3390/rs14246244> (2022).
45. Zhang, J. & Xu, H. Establishment of remote sensing based model to estimate the aboveground biomass of *Pinus densata* for permanent sample plots from national forestry inventory. *J. Beijing Forest. Univ.* **42**, 1–11 (2020).
46. Qu, Y., Su, Z., Li, Z. & Lin, Y. Effects of topographic factors on the distribution patterns of ground plants with different growth forms in montane forests in North Guangdong, China. *Chin. J. Appl. Ecol.* **22**, 1107–1113. <https://doi.org/10.13287/j.1001-9332.2011.0184> (2011).
47. Peng, S., Zhao, C., Zheng, X., Xu, Z. & He, L. Spatial distribution characteristics of the biomass and carbon storage of Qinghai spruce (*Picea crassifolia*) forests in Qilian Mountains. *Chin. J. Appl. Ecol.* **22**, 1689–1694. <https://doi.org/10.13287/j.1001-9332.2011.0240> (2011).
48. Luo, K. et al. Developing a method to estimate above-ground carbon stock of forest tree species *Pinus densata* using remote sensing and climatic data. *Forests* **15**, 2023. <https://doi.org/10.3390/f15112023> (2024).
49. Bao, R., Zhang, J., Lu, C. & Chen, P. Estimating above-ground biomass of *Pinus densata* Mast. using best slope temporal segmentation and Landsat time series. *J. Appl. Remote Sens.* **15**, 024507–024507. <https://doi.org/10.1117/1.JRS.15.024507> (2021).
50. Hengl, T., Heuvelink, G. B. M. & Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **120**, 75–93. <https://doi.org/10.1016/j.geoderma.2003.08.018> (2004).
51. Hanyue, S., Lei, X., Qingtai, S., Zhiyue, W. & Shuang, Q. Estimate forest aboveground biomass of mountain by ICESat-2/ATLAS data interacting cokriging. *Forests* **14**, 13 (2022).
52. Chen, P., Zhang, J., Xu, D. & Xiong, D. Land cover change detection in Shangri-La based on Landsat time series. *J. Southwest Forest. Univ.* **42**, 171–176. <https://doi.org/10.11929/j.swfu.202002048> (2022).
53. Zhang, J. & Xu, H. Establishment of remote sensing based model to estimate the aboveground biomass of *Pinus densata* for permanent sample plots from national forestry inventory. *J. Beijing Forest. Univ.* **42**, 1–11 (2020).
54. Yue, C. *Forest biomass estimation in Shangri-La country based on remote sensing*, Beijing Forestry University, (2012).
55. Cao, J. et al. Estimation of ACS and uncertainty analysis for *Pinus densata* based on random forest and Monte Carlo methods. *For. Res.* **36**, 131–139 (2023).
56. Potapov, P., Turubanova, S. & Hansen, M. C. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sens. Environ.* **115**, 548–561 (2011).
57. Qinyu, Y., Rui, W. & Hui, X. Optimal design of second-order sampling for forest biomass in Shangri-La City based on the forest management inventory. *J. Southwest Forest. Univ. (Nat. Sci.)* **41**, 160–167 (2021).
58. Xu, H., Zhang, Z. & Ou, G. In *Estimation and distribution of forest biomass and carbon stocks in Yunnan province* (ed Guanglong Ou) Ch. 107,109, 195 (Yunnan Science and Technology Press, 2019).
59. Sun, H., Song, Y. & Sun, Z. Dynamic change analysis of the forest carbon storage in Heilongjiang province. *China Forest. Econ.*, 55–58. <https://doi.org/10.13691/j.cnki.cn23-1539/f.2020.06.015> (2020).
60. Liu, G., Fu, B. & Fang, J. Carbon dynamics of Chinese forests and its contribution to global carbon balance. *Acta Ecologica Sinica*, 733–740. <https://doi.org/10.3321/j.issn:1000-0933.2000.05.004> (2000).
61. Han, X. et al. Sampling estimation of *Pinus densata* carbon storage based on remote sensing feature variables. *J. Southwest Forest. Univ.* **43**, 117–124. <https://doi.org/10.11929/j.swfu.202208047> (2023).
62. Xia, Z. Calculation of biomass and carbon storage of ecological public welfare forest in Liaoning province based on stratified sampling method. *J. Green Sci. Technol.* **24**, 153–156. <https://doi.org/10.16663/j.cnki.lskj.2022.09.028> (2022).
63. Haijian, L. et al. Analysis of groundwater glow field distribution at a typical chemical plant plots in Guangdong-Hong Kong-Macau greater bay area based on kriging interpolation. *J. Huizhou Univ.* **42**, 100–106. <https://doi.org/10.16778/j.cnki.1671-5934.2022.06.016> (2022).
64. Zimmerman, D. L. & Zimmerman, M. B. A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics* **33**, 77–91. <https://doi.org/10.1080/00401706.1991.10484771> (1991).

65. Zhang, W., Jing, T. & Yan, S. Studies on prediction models of *Dendrolimus superans* occurrence area based on machine learning. *J. Beijing Forest. Univ.* **39**, 85–93. <https://doi.org/10.13332/j.1000-1522.20160205> (2017).
66. Lopez-Senespleda, E., Calama, R. & Ruiz-Peinado, R. Estimating forest floor carbon stocks in woodland formations in Spain. *Sci. Total Environ.* **788**, 147743. <https://doi.org/10.1016/j.scitotenv.2021.147734> (2021).
67. Dar, A. A. & Parthasarathy, N. Patterns and drivers of tree carbon stocks in Kashmir Himalayan forests: Implications for climate change mitigation. *Ecol. Process.* **11**, 58. <https://doi.org/10.1186/s13717-022-00402-z> (2022).
68. Ju, T. et al. Estimation of forest above-ground biomass in Guangxi, China, by integrating forest age and stack learning. *Land Degrad. Dev.* **34**, 4079–4093. <https://doi.org/10.1002/ldr.4740> (2023).

Acknowledgements

This research was funded by the National Natural Science Foundation of China (No. 32260390), “Young Top Talents” special project of the high-level talent training support program of Yunnan province, China, in 2020 (No. YNWR-QNBJ-2020-164), the Forestry Innovation Programs of Southwest Forestry University (Grant No: LXXK-2023Z06).

Author contributions

M.P. collected the data, conducted the data analysis, and wrote the draft of the paper. M.X. data analysis and part of the graphs. B.Q. revised the paper. C.T. software and revised the paper. C.C. provided suggestions and guidance. J.Z. supervised and coordinated the research project, project administration, and revised the paper.

Declarations

Competing interest

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025