Estimating Forest Stock Volume Based on Airborne Lidar Data

Xiao Niu[1], Na Jiang[1], Ke Hou[1], Yuan Yin[1]

[1] Shandong Provincial Institute of Land Surveying and Mapping, China – xiaoniu710@126.com, Jiangna123321@163.com, 303475517@qq.com, 1078808122@qq.com

Keywords: Airborne Lidar data, Forest stock volume (FSV), Forest characteristic parameters, Random forest (RF), Model assessment.

Abstract

Forest stock volume (FSV) stands as an important indicator in evaluating the potential for carbon sequestration. It is crucial for forest resource management at local, regional, and national scales. In order to achieve an accurate estimation of FSV, this article takes Mengyin County, Shandong Province, China as the research area, builds a random forest (RF) model for four tree species based on airborne Lidar data, and forms a monitoring system of "individual tree - grid - county" granularities. The results demonstrated that all four models exhibited excellent generalization capabilities, with no signs of overfitting. In the test phase, the R^2 of the poplar and pine models exceeded 0.9, while the R^2 of the cypress model was 0.81, and the rRMSE was controlled within 20%, indicating that the fitting effect of the three tree species models was better; the accuracy of the robinia pseudoacacia model was relatively poor, with R^2 of 0.60 and rRMSE of 20.60%. This study provides a feasible method for estimating forest stock volume within the county, which provides strong technical support for forest resource management and planning, and helps promote sustainable forestry development.

1. Introduction

Forests are the mainstay of terrestrial ecosystems, their annual carbon sequestration accounts for about 2/3 of the entire terrestrial ecosystem (Post et al. 1982; He et al. 2022). They play an irreplaceable role in regulating global carbon balance, mitigating the greenhouse effect and tackling climate warming (Doelman et al. 2019). Accurately estimating the carbon sequestration potential of forests has important guiding significance for formulating action plans for addressing global climate change, increasing sequestration and reducing emissions under the carbon neutral target.

FSV is one of the most important indicators for assessing carbon sequestration potential (Hu et al. 2020). Traditional methods for calculating above-ground forest stock volume mainly rely on field survey data, which have high reliability (Liu et al. 2018). However, with these methods, it is cumbersome and difficult to implement dynamic monitoring of forest carbon stocks on a large regional scale. This kind of tasks is suitable for methods based on remote sensing data, which has the advantages of large scale, non-contact, multi-temporal, and high spatial resolution.

Traditional optical remote sensing techniques have certain limitations in monitoring forest resources. They mostly only provide texture and spectral information of the upper canopy of forests, and there exists a problem of easy saturation of spectral signals (Duncanson et al. 2010; Lu et al. 2012).The emergence of Lidar technology has effectively alleviated this issue. In particular, LiDAR data has a certain penetration ability for forests and can accurately describe the three-dimensional structure of the forest canopy (Nelson et al. 1984; Wilkes et al. 2018). Lidar data has been extensively utilized in research on the inversion of FSV. Yuan et al. (2021) combined airborne laser point cloud data with 800 ground sample plots to establish a stock volume model for these four coniferous forests using stepwise regression and partial least squares regression. The results showed that partial least squares regression was superior to stepwise regression. McRoberts et al. (2012) assessed the utility of lidar-based stratifications for mean growing stock volume per unit area. The results indicated that the stratifications based on nonlinear logistic regression model predictions of volume obtained from lidar data reduced variances of mean growing stock volume estimates. Liu et al. (2023) used a multiple linear to explore the relationship between forest stock volume and multi-source remote sensing features. This study explored an effective LiDAR sample collection scheme for estimating forest stock and can provide a reference for future LiDAR sample collection. However, the majority of these studies tend to rely on either multiple linear regression models or nonlinear regression models.

RF algorithm does not require statistical assumptions or predetermined model parameters. This algorithm can effectively handle nonlinear, interactive and collinear problems, while effectively avoiding overfitting (Sun et al. 2021). At present, RF has been applied in forest growth and forest carbon storage prediction (Mina et al. 2018; Jevšenak and Skudnik 2021; Tian et al. 2022). But, there are relatively few studies on the prediction of FSV using RF models for large-scale and multi-factor effects.

Mengyin County is located in southeast central Shandong Province, China. With a high forest coverage rate, complex terrain and large carbon sequestration potential, it is a representative and suitable region for forestry carbon sequestration monitoring researches using remote sensing technology. Therefore, this article takes Mengyin County as the research area, conducts an airborne Lidar data FSV inversion experiment based on the RF model, and evaluates the applicability and possibility of this technology in the inversion of forest stock volume in Shandong Province. It provides a reference for the application of new technologies in forest resource survey in the future.

2. Study area and Data used

2.1 Study area

The study is situated in the Yimeng Mountain Area in southeast part of Shandong Province, in Mengyin County, in the

coordinates 117°45′-118°15′E and 35 °27′-36°02′N (Mengyin County Bureau of Statistics, 2023). The total area of Mengyin County is 1,601.6 km², characterized by high-altitude terrain in the north and south and low-altitude terrain in the middle. The area has a warm temperate continental monsoon climate with four distinct seasons. The mean annual temperature is 14.4℃, and the mean annual precipitation is 813.8mm.

The main dominant tree species in Mengyin County are poplar, pine, cypress, and robinia pseudoacacia, accounting for more than 95% of the county's forest area. The location of the study area and the distribution of main dominant tree species can be found in Figure 1.
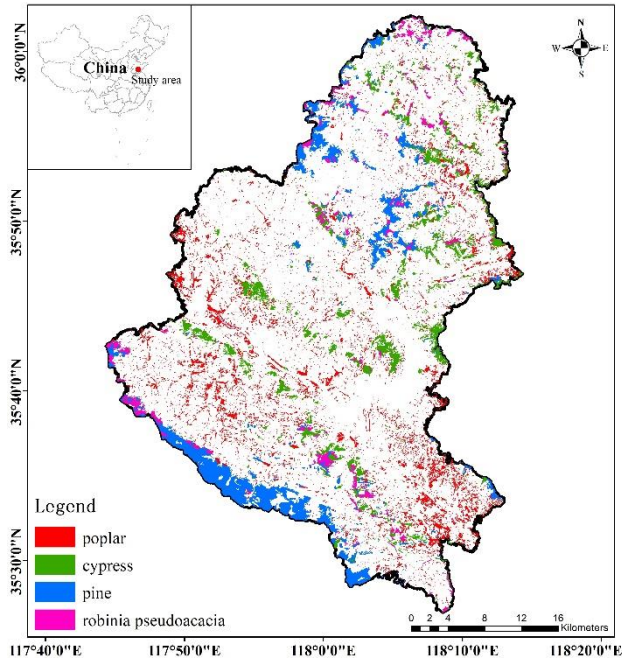

Figure 1. Location and the main dominant tree species in the study area.

2.2 Data Source

2.2.1 Lidar data

The airborne Lidar data mainly comes from the Shandong Provincial 14th Five-Year Basic Surveying and Mapping Planning project, with a point density of 1pts/m². The sensor system is CityMapper-2L and the laser model is Hyperion2. The data's currency spans from April to June 2023.

2.2.2 In situ sample plot data

The sample data was obtained by measuring the tree height, diameter at breast height (DBH) and location of the trees, taking into full consideration the information such as tree species, age group, and topography.

This study measured 71 plots with an area ranging from 1000 to 1500 square meters. The tree species in the plots are four dominant tree species, containing more than 8,800 trees.

2.2.3 Map of forest resources inventory data

Map of forest resources inventory data comes from forestry departments. The data is current as of 2021. Based on this data, the vegetation coverage and types of tree species can be derived.

3. Methods

3.1 calculation of sample plot FSV

In this research, a 20*20meter grid system is selected as the unified analytical units. The stock volume of each tree was calculated based on the "Timber Volume Table of Major Tree Species in Shandong Province". Based on the DBH of each tree, the corresponding volume values can be found in the table according to the DBH classification and different tree species. Then, the total volume of all trees contained within the standard grid range is calculated. Finally, 163 standard grid samples with FSV parameters were formed, including 36 poplars, 43 pines, 49 cypresses, and 35 robinia pseudoacacia trees.

3.2 Forest characteristic parameters extraction

After point cloud rough classification, noise removal and vegetation reclassification, original Lidar data are ultimately transformed to normalized point cloud data. Among them, the high-vegetation points after vegetation reclassification are considered as trees.
Based on normalized point cloud data, forest characteristic parameters are calculated according to a 20*20meter grid, mainly including height variables, density variables and vertical structure variables (Table 1). This article extracts totally 34 forest characteristic parameters.

Variable type	Characteristic variable	description	formula
height variables	H_max	The maximum high-vegetation height	
	H_min	The minimum of high-vegetation height	
	H_mean	The average of high-vegetation height	
	H_stdv	The standard deviation of high-vegetation height	
	H_var	The variance of high-vegetation height	
	hp (hp25, hp50, hp75, hp95)	height percentage	Refer to Li et al. (2023)
	H_range	Range of height distribution	H_max-H_min
	H_cv	coefficient of variation	H_std/H_mean
	H_skew	skewness	$\dfrac{\sum_{i=1}^{n}(H_i - \text{H\_mean})^3}{(n-1)\text{H\_stdv}^3}$
	H_Kurt	kurtosis	$\dfrac{\sum_{i=1}^{n}(H_i - \text{H\_mean})^4}{(n-1)\text{H\_stdv}^4}$
	H_iqr	interquartile range	hp_75-hp_25

	H_crr	$(H\_mean-H\_min)/H\_range$
	Average_second_pow	$\sqrt[2]{\dfrac{\sum_{i=1}^{n}(H_i)^2}{n}}$
	Average_third_pow	$\sqrt[3]{\dfrac{\sum_{i=1}^{n}(H_i)^3}{n}}$
	dq (dp_25, dp_50, dp_75, dp_95)	density quantile	Refer to Li et al. (2023)
density variables	CC	canopy density	Total number of first return high-vegetation points/total number of returns points
	hvp	high vegetation proportion	
	LIR	Light Interception Rate	1- (total number of ground points/total number of points)
	LAI_total	Leaf area index	Total number of high-vegetation points/total number of first return points
vertical structure variables	LAI_mean,LAI_stdv,LAI_cv	Mean, standard deviation, and coefficient of variation of leaf area index. Note that the leaf area index in this row is different from the LAI_total in the previous row, and is obtained after slicing at a certain height.	Refer to Hopkinson et al. (2013)
	LAD_mean,LAD_stdv,LAD_cv	Mean, standard deviation, and coefficient of variation of leaf area density	Refer to Hopkinson et al. (2013)
	VFP_mean,VFP_std, VFP_cv	Mean, standard deviation, and coefficient of variation of the vertical foliage profile	Refer to Knapp et al. (2019)

Table 1. Forest characteristic parameters (Remark: where n is the high-vegetation numbers in each grid, H_i refers to the height of the i-th high vegetation point in the grid)

3.3 Model construction

The regression model is create using RF here in this paper, a popular machine learning method. The RF model is an ensemble learning method based on decision trees. It constructs a series of base learners through resampling, combines the prediction results of these base learners, and outputs the final prediction (Zhou 2016). It has the ability to solve both regression and classification problems. The prediction formula for the RF regression model can be expressed as follows:

$$\bar{g}(X) = \frac{1}{T}\sum_{i=1}^{t}\{g(x,\theta_t), t = 1,2,\cdots,T\} \qquad (1)$$

Where $g(x,\theta_t)$ represents the t-th decision tree model, θ_t is a random variable that follows an independent distribution, x is the independent variable, and T is the number of decision trees.

The main works of volume modelling are: (1) choosing characteristic variables. Pearson's correlation analysis is conducted between 30 forest characteristic variables and stock volume, and the parameters with higher correlation index ($\geqslant 0.6$) are selected as the characteristic variables of the random forest model. (2) creating training set and test set. The sample data are randomly divided into training set and testing set, 80% of them belong to the former and 20% belong to the latter. (3) the optimal combination of hyperparameters of the random forest model is determined by grid searching strategy. (4) building the models. Considering the heterogeneity of different tree species, this paper establishes models for the four dominant tree species respectively.

3.4 Model assessment method

Two statistical indicators, the coefficient of determination(R^2) and the relative root mean square error(rRMSE), were chosen to assess the performance of the RF model. The formula is as follows.

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

$$rRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}}{\bar{y}} \qquad (3)$$

Where y_i represents the measured FSV, $\bar{y}$ is the mean measured FSV, $\hat{y}_i$ is the predicted FSV, i is the same index, and n is the grid sample plots number.

4. Result and Discussion

4.1 Model assessment results

In this paper, the fitting quality of the models was tested by the R^2 and rRMSE. The results (Table 2) shown that the four models had good generalization ability and there's no overfitting. In the training phase, the pine model had the best performance with the highest R^2=0.97 and smallest rRMSE=10.26%. The poplar model was second with an R^2=0.95 and rRMSE=11.19%. This model was followed by the cypress model with an R^2=0.88 and rRMSE=13.81%. The robinia pseudoacacia model had the worst performance with the smallest R^2=0.61 and highest rRMSE=19.36%. In the test phase, it was clear that the pine

model also performed the best among the four models (R^2=0.94, rRMSE=18.10%); the next best was the poplar model, with an R^2=0.91 and rRMSE=15.25%; the cypress was third with an R^2=0.81 and rRMSE=18.30%; the robinia pseudoacacia model performed the worst, with the smallest R^2=0.60 and highest rRMSE=20.60%.

Tree species	Training set accuracy		Test set accuracy	
	R^2	rRMSE	R^2	rRMSE
poplar	0.95	11.19%	0.91	15.25%
cypress	0.88	13.81%	0.81	18.30%
pine	0.97	10.26%	0.94	18.10%
robinia pseudoacacia	0.61	19.36%	0.60	20.60%

Table 2. Accuracy assessment of four dominant tree species models.

4.2 Map of the FSV estimation

Through the map of forest resources inventory data, the forest area of Mengyin County was extracted, with a total of 592,169 20*20m grids data. According to the experimental results, the stock volume in Mengyin County was mainly distributed in the southwest and northeast regions. The major types of landforms of the two regions were low mountains and hills. The Figure 2 shown that the FSV in Mengyin County ranged from 0.97m^3 to 10.21m^3. Among the tree species, poplar had the largest total stock volume, accounting for 44% of the total; pine was second, accounting for 40%; cypress accounted for 10%; and robinia pseudoacacia accounted for 6%.
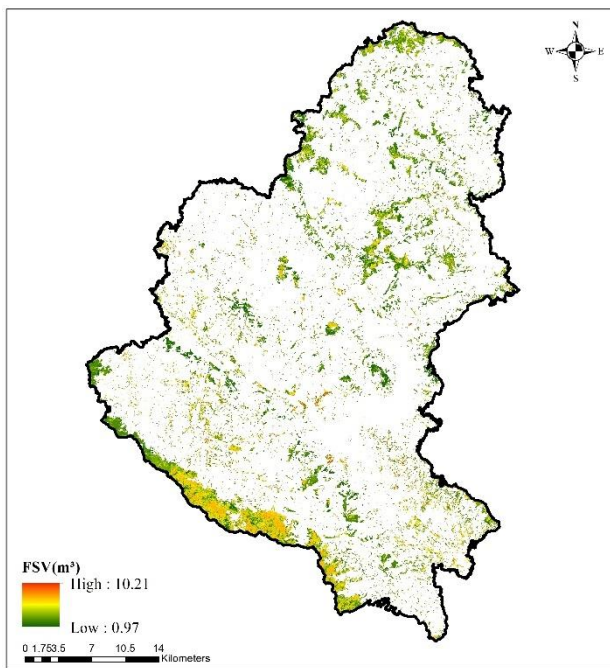

Figure 2. Map of the FSV estimation in Mengyin County

4.3 Spatial pattern analysis of FSV

After the entire FSV in Mengyin County was estimated, its spatial pattern was analyzed. The distribution of the stock volume, including its direction and range, was described using the standard deviation ellipse, and its spatial autocorrelation was described using Moran's I.

The standard deviation ellipse is a spatial statistical technique for measuring the distribution pattern of geographical elements (Liu et al. 2021). It can be used to analyze the spatial distribution characteristics of FSV in through parameters of standard deviation. The Figure 3 illustrated that the ellipse centered at 118.0086°E, 35.6976°N, with a major axis of 24.4km and a minor axis of 13.6km, can cover about 60% of FSV in Mengyin County. The direction of the principal axis of the standard deviation ellipse for FSV in Mengyin County was oriented at 14° north by east, suggesting a primary concentration of storage volume within the county towards the southwest-northeast axis.
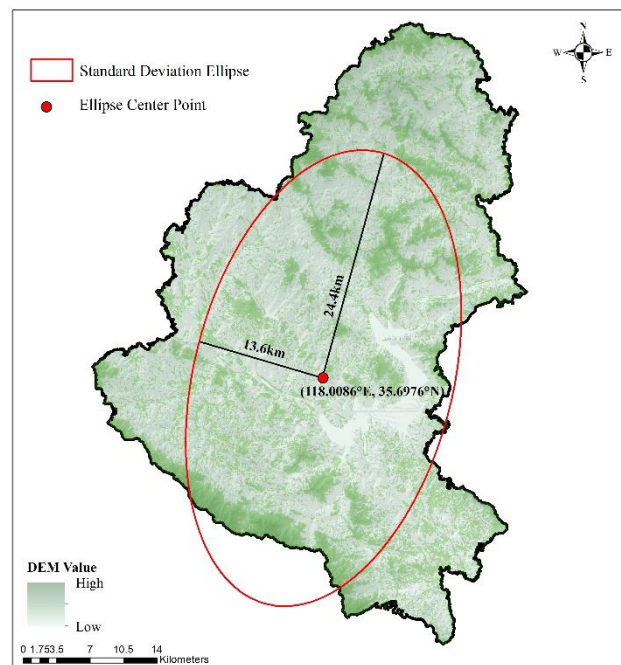

Figure 3. The distribution of standard deviation ellipse of FSV

Global Moran's I of FSV in Mengyin County was calculated using ArcGIS's toolbox. The global Moran's I value stood at 0.77, with a p value below 0.01 and a Z score significantly exceeding 1.65, strongly suggesting that there was a notable autocorrelation present in the FSV of Mengyin County. In order to more intuitively analyze the spatial distribution pattern of the FSV, the local Moran's I was calculated in the experiment; the clusters were shown in Figure 4.

The high-high values were mainly distributed in the southern mountainous areas, forming distinct contiguous zones; the low-low values were distributed in the northern, central, and southwestern regions, forming distinct surface agglomerations locally. The distribution of low-high values and high-low values failed to exhibit any distinct clustering, instead, it was solely scattered around the peripheries of regions dominated by high-high values and low-low values.
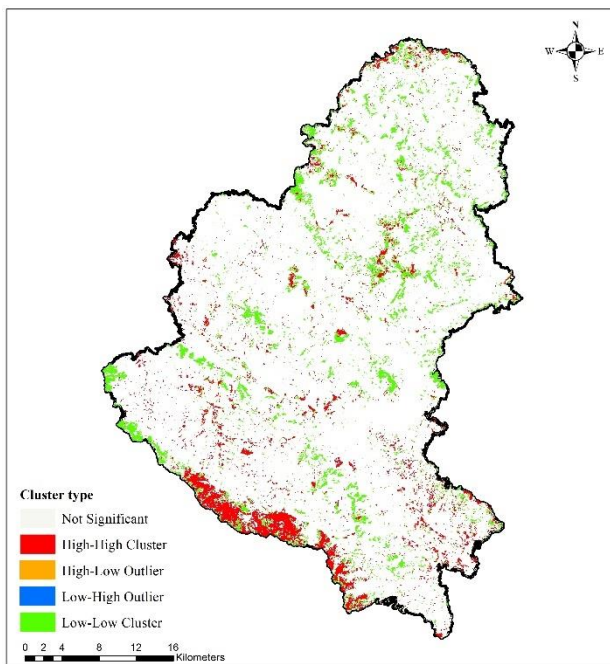
Figure 4. Cluster types of estimated FSV

4.3 Limitations and future work

The next step in this paper's research involves several crucial aspects. Firstly, during the field survey, the sample plots collected were mostly high-accumulation plots, and there were few samples with low accumulation. Therefore, the minimum estimated accumulation was relatively high. So we plan to expand the dataset by collecting more representative samples, especially for poplars and robinia pseudoacacia. This enhancement will ensure a more comprehensive and robust result for FSV. Secondly, while the current study focuses on evaluating the model's inherent accuracy, it lacks validation against third-party data. As a result, we intend to conduct a comparative analysis with external datasets to assess the model's reliability. This step is crucial for validating the model's performance in real-world scenarios. Additionally, we will explore potential improvements to the model's architecture and hyperparameters to further enhance its accuracy and efficiency. Overall, these works aim to strengthen the reliability and applicability of our model for estimating FSV in Mengyin County.

5. Conclusion

This study introduced a method for estimating FSV which jointly used low-density point cloud data and the random forest modelling algorithm. This method can greatly save manpower and material costs compared to traditional forestry estimation methods. Establishing models for individual tree species made the models more targeted and applicable, laying a solid foundation for calculating biomass and carbon storage in the next work.

References

Doelman, J., Stehfest, E., Vuuren, D., Tabeau, A., Hof, A., Braakhekke, M., Gernaat, D., Van den Berg, M., van Zeist, W.-J., Daioglou, V., Meijl, H.,Lucas, P., 2019. Afforestation for climate change mitigation: Potentials, risks and trade - offs. *Global Change Biol.,* 26(3), 1576-1591.

Duncanson, L.I., Niemann, K.O.,Wulder, M.A., 2010. Integration of GLAS and Landsat TM data for aboveground biomass estimation. *Can. J. Remote Sens.,* 36(2), 129-141.

He, G., Zhang, Z., Zhu, Q., Wang, W., Peng, W.,Cai, Y., 2022. Estimating Carbon Sequestration Potential of Forest and Its Influencing Factors at Fine Spatial-Scales: A Case Study of Lushan City in Southern China. *Int. J. Environ. Res. Public Health,* 19, 9184.

Hopkinson, C., Lovell, J., Chasmer, L., Jupp, D., Kljun, N.,Gorsel, E.v., 2013. Integrating terrestrial and airborne lidar to calibrate a 3D canopy model of effective leaf area index. *Remote Sens. Environ.,* 136, 301-314.

Hu, Y., Xu, X., Wu, F., Sun, Z., Xia, H., Meng, Q., Huang, W., Zhou, H., Gao, J., Li, W., Peng, D.,Xiao, X., 2020. Estimating Forest Stock Volume in Hunan Province, China, by Integrating In Situ Plot Data, Sentinel-2 Images, and Linear and Machine Learning Regression Models. *Remote Sens,* 12(1), 186.

Jevšenak, J.,Skudnik, M., 2021. A random forest model for basal area increment predictions from national forest inventory data. *For. Ecol. Manage.,* 479, 118601.

Knapp, N., Fischer, R., Cazcarra-Bes, V.,Huth, A., 2019. Structure Metrics to Generalize Biomass Estimation from Lidar across Forest Types from Different Continents. *Remote Sens. Environ.,* 237

Li, C., Li, Z.,Dai, H., 2023: *Airborne Laser Radar Forest Resource Survey and Monitoring*. Beijing: Science Press

Liu, H., Shao, M.,Ji, Y., 2021. The spatial pattern and distribution dynamic evolution of carbon emissions in China: Empirical study based on county carbon emission data. *Scientia Geographica Sinica,* 41(11), 1917-1924.

Liu, H., Zhang, Z.,Cao, L., 2018. Estimating forest stand characteristics in a coastal plain forest plantation based on vertical structure profile parameters derived from ALS data. *National Remote Sensing Bulletin,* 22(5), 872-888.

Liu, J., Quan, Y., Wang, B., Shi, J., Ming, L.,Li, M., 2023. Estimation of Forest Stock Volume Combining Airborne LiDAR Sampling Approaches with Multi-Sensor Imagery. *Forests,* 14(12), 2453.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G.,Saah, D., 2012. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research,* 2012, 436537.

McRoberts, R.E., Gobakken, T.,Næsset, E., 2012. Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. *Remote Sens. Environ.,* 125, 157-166.

Mina, M., Huber, M.O., Forrester, D.I., Thürig, E.,Rohner, B., 2018. Multiple factors modulate tree growth complementarity in Central European mixed forests. *J. Ecol.,* 106(3), 1106-1119.

Nelson, R., Krabill, W.,MacLean, G., 1984. Determining forest canopy characteristics using airborne laser data. *Remote Sens. Environ.,* 15(3), 201-212.

Post, W.M., Emanuel, W.R., Zinke, P.J.,Stangenberger, A.G., 1982. Soil carbon pools and world life zones. *Nature,* 298, 156-159.

Mengyin County Bureau of Statistics., 2023: *Mengyin Statistical Yearbook*. Mengyin Statistics Press

Sun, Z., Gao, J., Wu, F., Gao, X., Hu, Y.,Gao, J., 2021. Estimating Forest Stock Volume via Small-Footprint LiDAR Point Cloud Data and Random Forest Algorithm. *Sci. Silvae Sin.,* 57(8), 68-81.

Tian, H., Zhu, J., He, X., Chen, X., Jian, Z., Li, C.,Guo, X., 2022. Projected Biomass Carbon Stock of Arbor Forest of Three Provinces in Northeastern China Based on Random Forest Model. *Sci. Silvae Sin.,* 58(4), 40-50.

Wilkes, P., Disney, M., Vicari, M.B., Calders, K.,Burt, A., 2018. Estimating urban above ground biomass with multi-scale LiDAR. *Carbon Balance Manage.,* 13(1), 10.

Yuan, Y.-N., Peng, D.-L., Wang, W.,Zeng, W., 2021. Estimating standing stocks of the typical conifer stands in Northeast China based on airborne lidar data. *Chin. J. Appl. Ecol.,* 32(3), 836-844.
Zhou, Z., 2016: *Machine learning*. Beijing: Tsinghua University Press