



OPEN ACCESS

EDITED BY

Jiixin Jin,
Hohai University, China

REVIEWED BY

Shan Wang,
Beijing Forestry University, China
Aqib Gul,
Sher-e-Kashmir University of Agricultural
Sciences and Technology, India

*CORRESPONDENCE

Zhidong Zhang
✉ zhangzd@hebau.edu.cn

RECEIVED 10 August 2025

REVISED 14 November 2025

ACCEPTED 17 November 2025

PUBLISHED 02 December 2025

CITATION

Zhang J, Wang Z, Li M, Chen X, Pang Y and
Zhang Z (2025) Using a random forest model
to predict volume growth of larch, birch, and
their mixed forests in northern China.
Front. Plant Sci. 16:1682940.
doi: 10.3389/fpls.2025.1682940

COPYRIGHT

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

Using a random forest model to predict volume growth of larch, birch, and their mixed forests in northern China

Junfei Zhang, Ziyi Wang, Minghao Li, Xiaotong Chen,
Yue Pang and Zhidong Zhang*

College of Forestry, Hebei Agricultural University, Baoding, Hebei, China

Accurately quantifying forest volume and identifying its driving mechanisms are critical for achieving carbon neutrality objectives. Using data from the National Forest Inventory (NFI), plot-level measurements, and environmental variables from pure larch (LP), birch (BP), and mixed larch-birch (LB) forests in the mountainous region of northern Hebei, China, this study employed random forest (RF) algorithms to evaluate the relative importance and partial dependence of biotic and abiotic factors on stand volume growth. A total of 33 predictors related to climate, topography, and soil were analyzed, and model hyperparameters were optimized through grid search combined with blocked cross-validation to mitigate spatial autocorrelation. The RF models exhibited strong predictive performance, with the BP model achieving the highest R^2 (0.92). The minimum temperature of the coldest month (Bio12) was identified as the most influential predictor across all stand types, while stand age also exerted a substantial effect on growth dynamics. Young and middle-aged forests demonstrated higher productivity compared with near-mature and mature stands, suggesting that the latter require improved management interventions to sustain growth. The LB stands exhibited higher productivity than pure stands, likely due to species complementarity and interspecific facilitation. In LP, growth was primarily driven by the interaction between stand age and canopy density, whereas in BP, slope position was more decisive. The management of LB stands offers potential to maintain or enhance forest productivity. The findings emphasize the importance of adaptive forest management strategies that optimize forest structure and mitigate climate change impacts. These insights contribute to advancing carbon sequestration efforts and supporting the development of carbon neutrality policies by enhancing forest productivity and resilience to climate variability.

KEYWORDS

carbon neutrality, random forest algorithm, stand volume growth, relative importance, partial dependence

1 Introduction

Global climate change, primarily driven by anthropogenic activities, poses a profound threat to ecosystems and human well-being (Malhi et al., 2002; Solomon et al., 2009; Rehman et al., 2021). In response, the Chinese government has established the ambitious “dual carbon” goals, aiming to peak CO₂ emissions by 2030 and achieve carbon neutrality by 2060. Enhancing forest carbon sequestration has been recognized as a key national strategy for achieving these goals, as forests constitute the largest terrestrial carbon pool (Pan et al., 2011). Forest stocking volume, serving as a direct proxy for biomass and carbon storage, represents a critical metric for assessing carbon sink capacity (Aalde et al., 2006). The National Forestry and Grassland Administration has set explicit targets to increase national forest coverage to 24.1% and forest volume stock to 19 billion m³ by 2025 (National Forestry and Grassland Administration, 2021). Therefore, accurately quantifying forest volume and identifying its driving mechanisms are essential for developing scientifically informed forest management practices and supporting climate change mitigation efforts.

Forest volume growth is governed by the complex interplay among stand attributes, climatic conditions, soil properties, and topographic factors (Lei, 2019). Traditional statistical models have been widely used to examine these relationships; however, they often fail to adequately capture the inherent nonlinearity and intricate interactions among multiple driving factors (Tian et al., 2022). This limitation underscores the need for advanced analytical approaches capable of handling such ecological complexity. In recent years, machine learning (ML) techniques, particularly the Random Forest (RF) algorithm, have emerged as powerful tools in ecological research due to their high predictive accuracy and ability to model nonlinear patterns without requiring predefined functional relationships (Breiman, 2001). The RF approach has proven effective in various forest modeling applications, including the estimation of basal area increment (Jevšenak and Skudnik, 2021) and the identification of key biomass drivers (He et al., 2022), demonstrating strong potential for elucidating the combined effects of site and climatic factors on forest growth (Tian et al., 2022).

Larch (*Larix principis-rupprechtii*) and birch (*Betula platyphylla*) forests are ecologically and economically vital in northern China. Previous studies have primarily focused on individual aspects of their productivity, including climate-growth relationships (Wang et al., 2022a; Liu et al., 2024), the impact of stand age and site conditions on productivity (Zhao et al., 2015; Dincă et al., 2020), and the effects of soil quality (Wang et al., 1998; Zhang et al., 2023). Although some studies have reported facilitative effects in mixed stands (Zhang et al., 2022b), a comprehensive assessment that integrates multiple ecological drivers such as stand characteristics, climate, soil, and topography to accurately model volume growth across pure and mixed stands remains lacking. This knowledge gap hinders our ability to predict forest dynamics under changing environmental conditions and to optimize management strategies aimed at enhancing carbon sequestration.

To address these research gaps, this study seeks to answer the following scientific questions: (1) How do the key drivers of volume

growth and their relative importance differ among the three stand types? (2) Do mixed stands exhibit higher productivity than pure stands, and what factors contribute to this potential advantage? (3) What are the critical thresholds and nonlinear responses of volume growth to major environmental gradients? Accordingly, the specific objectives were to: (1) develop volume growth models for the three stand types using RF; (2) quantify the relative importance of predictors on volume growth; and (3) explore the response patterns of key variables influencing stand volume growth. The findings are expected to provide a scientific basis for precision forestry management and contribute to regional carbon neutrality initiatives.

2 Materials and methods

2.1 Data collection and processing

2.1.1 Study area and stand volume growth data

This study was conducted in Weichang County (42°02′–42°36′ N, 116°51′–117°39′ E), Chengde City, Hebei Province, located at the ecotone between the North Hebei Mountains and the Mongolian Plateau (Figure 1). The area encompasses two forest farms: Saihanba Mechanical Forest Farm and Mulan Weichang National Forest Farm. According to the Ninth National Forest Inventory, the primary forest types in the study area were larch plantation, covering 38,182 ha (51.1% of the total forest area), followed by birch secondary forest, which occupies 16,039 ha (21.4%). Larch stands exhibited the highest unit volume at 172 m³/ha, while birch stand averaged 124 m³/ha (Zhang et al., 2025). Stand volume data were derived from three sources. First, subcompartment inventory data (2016 and 2022) from both forest farms were used, encompassing 8,996 subcompartments dominated by larch or birch. The volume for each subcompartment was calculated as the sum of individual tree volumes, with volume growth derived from the difference between 2022 and 2016 measurements. Second, data were extracted from the eighth and ninth forest resource inventories of Hebei Province, which included 123 sample plots in larch and birch stands, typically set up as squares of 0.0667 ha. Third, field survey data (2016 and 2022) from 128 sample plots—set as rectangular plots ranging from 600 to 2,500 m²—were incorporated. Individual tree volume was calculated using two-variable standing volume models (Equations 1, 2) based on diameter at breast height (DBH) and tree height. Stand volume was calculated as the total volume of all trees per plot, while volume increment was computed as the difference between the two survey years. Notably, only alive trees that exist in both periods were counted. A stand was classified as pure larch or pure birch if the dominant species accounted for more than 80% of the total stand volume; otherwise, it was designated as mixed larch-birch stand. To ensure comparability, the volume increment was standardized to an annual per-hectare rate (m³·ha⁻¹·yr⁻¹), adjusted based on plot area and the duration of the growth period.

Equations 1, 2 were used to estimate the individual tree volume for larch and birch, respectively (Standardization Administration of

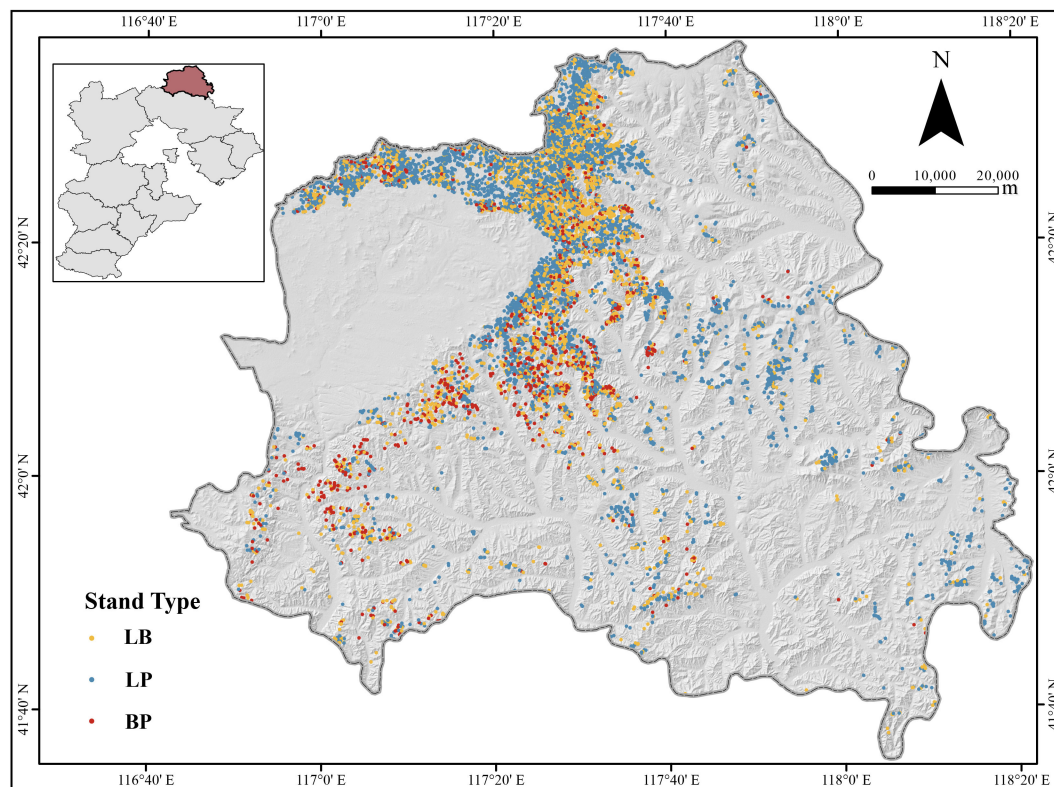


FIGURE 1

Geographical location of the study area and distribution of sample plots. LP, *Larix principis-rupprechtii* pure forest; BP, *Betula platyphylla* pure forest; LB, mixed *L. principis-rupprechtii*–*B. platyphylla* forest.

the People's Republic of China, 2024).

$$V = 0.00007267D^{1.80035} \times H^{0.99246} \quad (1)$$

$$V = 0.00008077D^{1.83864} \times H^{0.89108} \quad (2)$$

where: V is the standing tree volume (m^3); D is the diameter at breast height (DBH, cm); and H is the tree height (m). The single-tree volume equations were calibrated based on data collected from a broad region of northern China, with the detailed parameters of each equation presented in [Supplementary Table S2](#) of the [Supplementary Material](#).

2.1.2 Bioclimatic factors

Bioclimatic data were obtained using ClimateAP software (v3.10) (<http://ClimateAP.net>). ClimateAP provides location-specific monthly and seasonal climate variables across the Asia-Pacific region for the period 1901–2100 (Wang et al., 2017b). For each plot's specific growth period, a corresponding time series of annual climate data was obtained. The annual values of each climatic variable were then averaged over the growth period, resulting in a dataset comprising 15 time-averaged bioclimatic variables for each sample plot (Table 1).

2.1.3 Topographic factors

Topographic variables—including elevation, slope, aspect, and slope position—were derived from a 30-m resolution Digital Elevation Model (DEM) obtained from the National Earth System Science Data Center (<http://www.geodata.cn>). Elevation, slope, and aspect were extracted from the DEM using ArcGIS 10.8, while slope position was recorded for each sample plot during field surveys. All DEM-based layers were projected onto the CGCS2000/Gauss–Krüger Zone 20 coordinate system to ensure spatial consistency. Raster alignment was subsequently checked to confirm that all layers shared the same extent, grid origin, and pixel alignment.

2.1.4 Soil factors

Soil data were sourced from the National Earth System Science Data Center (<http://www.geodata.cn>). Twelve soil variables were selected for analysis (Table 1). The original soil data, with a spatial resolution of 900 m, were resampled to 30 m using bilinear interpolation in ArcGIS 10.8 to ensure consistency with other variables. All soil layers were projected onto the CGCS2000/Gauss–Krüger Zone 20 coordinate system, and raster alignment was checked to ensure consistent extent, grid origin, and pixel alignment across all spatial layers.

TABLE 1 Candidate variables for stand volume growth modeling.

Type	Variable	Mean	SD	Min.	Max.
Stand attributes	ΔV ($\text{m}^3\cdot\text{ha}^{-1}\cdot\text{yr}^{-1}$)	10.2	6.1	1.7	27.9
	Stand age (yr)	39.2	10.2	10.0	90.0
	Canopy density	0.7	0.2	0.1	0.9
Topographic factors	Altitude (m)	1530.0	185.5	668.0	2034.0
	Slope ($^{\circ}$)	11.8	7.2	0.0	40.7
	Slope direction	–	–	–	–
	Slope position	–	–	–	–
Soil factors	AK ($\text{mg}\cdot\text{kg}^{-1}$)	173.5	56.0	43.1	308.3
	AN ($\text{mg}\cdot\text{kg}^{-1}$)	157.2	62.8	13.2	501.0
	AP ($\text{mg}\cdot\text{kg}^{-1}$)	10.6	5.1	1.9	16.3
	TK ($\text{mg}\cdot\text{kg}^{-1}$)	1789.0	375.5	1178.0	2485.0
	TN ($\text{mg}\cdot\text{kg}^{-1}$)	255.5	100.8	10.9	498.6
	TP ($\text{mg}\cdot\text{kg}^{-1}$)	71.9	15.7	30.8	112.4
	BD ($\text{g}\cdot\text{cm}^{-3}$)	1.3	0.1	0.8	1.4
	CL (%)	16.2	1.3	5.1	20.9
	GRAV (%)	15.8	7.4	0.0	42.0
	SA (%)	41.2	3.3	29.8	82.8
	SOM (%)	4.7	2.1	0.2	10.7
	pH	6.7	0.4	6.0	8.5
Climatic factors	Bio1 ($^{\circ}\text{C}$)	2.3	1.2	0.2	8.5
	Bio2 ($^{\circ}\text{C}$)	18.9	0.9	16.8	23.5
	Bio3 ($^{\circ}\text{C}$)	-16.9	1.3	-19.4	-9.4
	Bio4 ($^{\circ}\text{C}$)	35.7	0.5	30.7	37.8
	Bio5 (mm)	516.3	25.2	382.5	749.2
	Bio6 (days)	151.6	12.1	128.0	203.2
	Bio7 ($^{\circ}\text{C}$)	17.0	1.0	14.7	22.3
	Bio8 ($^{\circ}\text{C}$)	-14.3	1.3	-16.7	-6.9
	Bio9 (mm)	328.3	17.4	222.5	494.2
	Bio10 (mm)	17.4	5.3	6.3	26.5
	Bio11 ($^{\circ}\text{C}$)	25.1	0.8	23.3	29.0
	Bio12 ($^{\circ}\text{C}$)	-20.4	5.0	-26.0	-10.3
	Bio13 ($^{\circ}\text{C}$)	13.9	0.5	11.6	15.2
	Bio14 (mm)	158.3	9.1	97.5	235.5
	Bio15 (mm)	3.0	2.3	0.9	40.8

(1) “–” represents categorical variables; SD, standard deviation; Min, minimum; Max, maximum. (2) ΔV , average annual volume growth of forest stand; Stand age, average age of dominant tree species; Canopy density, the ratio of the tree canopy cover to the total forest area; AK, available K; AN, alkali-hydrolysable N; AP, available P; TK, total K; TN, total N; TP, total P; BD, bulk density; CL, percentage of clay; GRAV, gravel; SA, percentage of sand; SOM, soil organic matter; pH, soil pH value in the topsoil; Bio1, annual mean temperature; Bio2, mean warmest month temperature; Bio3, mean coldest month temperature; Bio4, monthly temperature difference (temperature difference between Bio2 and Bio3); Bio5, annual precipitation; Bio6, the number of frost-free days; Bio7, mean temperature of summer; Bio8, mean temperature of winter; Bio9, summer precipitation; Bio10, winter precipitation; Bio11, max temperature of warmest month; Bio12, min temperature of coldest month; Bio13, mean diurnal range [mean of monthly (max temp. – min temp.)]; Bio14, wettest-month precipitation; Bio15, driest-month precipitation.

2.2 Forest type classification

Landsat 8 OLI Level-2 surface reflectance imagery was obtained from the Geospatial Data Cloud (<http://www.gscloud.cn>) at a spatial resolution of 30 m. As July was identified as the period during which chlorophyll content peaked in the canopies of both larch and birch (Asner and Martin, 2008), the acquisition date was set to July 24, 2022. The near-infrared band (Band 5, 0.85–0.88 μm) and the shortwave infrared band (Band 7, 2.11–2.29 μm), which are sensitive to spectral differences between coniferous and broadleaf species, were used as key inputs for supervised classification (Liu et al., 2018; Soleimannejad et al., 2019).

Supervised classification of forest types was performed using the Random Forest (RF) algorithm in ArcMap 10.8, based on satellite imagery and plot data (see [Supplementary Figure S1](#) for classification results). Training and validation samples were derived from field survey data. A stratified random sampling strategy was employed to select these samples, effectively mitigating spatial dependence by ensuring geographic separation between the training and validation datasets. A confusion matrix was then constructed to evaluate classification accuracy. Four accuracy metrics, including Producer's Accuracy (PA), User's Accuracy (UA), Overall Accuracy, and the Kappa coefficient (K) (Olofsson et al., 2014; Morales-Barquero et al., 2019; Xie et al., 2019; Olorunfemi et al., 2020), were calculated. The classification yielded an Overall Accuracy of 91.11% and a K of 86.67%, with detailed results provided in [Supplementary Table S1](#) of the [Supplementary Material](#).

2.3 Random forest algorithm

RF is an ensemble learning method that combines multiple models to enhance overall predictive performance. An RF model comprises several decision trees and operates based on three core steps. First, n samples (where n equals the size of the original dataset) are randomly drawn with replacement from the original dataset to form a training subset. Second, a regression tree is fitted to each subset. At each node split, a random subset of features is selected, and the optimal split is determined based on variance reduction. This recursive process continues until terminal nodes contain either samples from a single class or no more than five observations. Finally, predictions from all individual trees are averaged to produce the final output (Breiman, 2001; Ou et al., 2019).

Forest ecosystems are characterized by multi-source, high-dimensional data involving topography, climate, and stand structure. The powerful nonlinear modeling capability of RF allows for the complex coupling among these factors. In previous studies on forest productivity, RF models have consistently demonstrated superior performance compared to alternative approaches (Hamidi et al., 2021; Liu et al., 2022; Zhang et al., 2022a). However, the internal decision-making logic of RF is difficult to interpret intuitively, and the relationship between input features and model output remains opaque. To enhance

interpretability, relative importance analysis and Partial Dependence Plots (PDPs) were employed (Breiman, 2001; Friedman, 2001). This combined approach has been widely adopted in ecological modeling (Jevšenak and Skudnik, 2021; Wang et al., 2021; He et al., 2022; Tian et al., 2022; Wang et al., 2022b; Zhang et al., 2022a).

For regression task, the relative importance of each predictor variable was determined using the Percent Increase in Mean Squared Error (%IncMSE) metric (Genuer et al., 2012; Kuhn, 2008). The resulting importance values were normalized as percentages of the total predictor importance for comparative analysis (Tian et al., 2022). Based on these rankings, PDPs were generated for the top eight predictors influencing volume growth, using the Locally Weighted Regression (LOESS) smoothing method (Cleveland and Devlin, 1988; Jacoby, 2000), which effectively reduces high-frequency noise while preserving critical ecological thresholds (Shi et al., 2019; Wang, 2024). To further improve model interpretability, the interaction strength among variables was quantified using Friedman's H-statistic (Friedman and Popescu, 2008), and the marginal effects of the top eight predictors were visualized through Accumulated Local Effects (ALE) plots, which account for potential feature correlations (Apley and Zhu, 2020). Additionally, SHapley Additive exPlanations (SHAP) values were calculated to assess the contribution of each predictor to individual model outputs (Lundberg et al., 2020). All interpretability analyses were performed using the "iml" package (Molnar et al., 2018), and the RF model was implemented using the "randomForest" package (Genuer et al., 2012) in R version 4.3.2.

2.4 Model development

A total of 33 predictor variables ([Table 1](#)) were selected as independent variables to model stand volume growth using the RF algorithm. During RF model construction, three key hyperparameters required tuning: the number of regression trees ($ntree$), the number of predictor variables randomly selected at each tree node for splitting ($mtry$) (Tian et al., 2022), and the minimum node size ($min.node.size$). Generally, the overall error rate decreases with increasing $ntree$ and stabilizes once $ntree$ exceeds 500. To ensure model reliability while maintaining computational efficiency, $ntree$ was set to 1000 in this study ([Supplementary Figure S2](#)) (Zhang et al., 2016). Although $mtry$ was often set to one-third of the total number of predictors in previous studies (Breiman, 2001), the optimal $mtry$ value was typically variable. Therefore, a comprehensive grid search was performed over the full range of $mtry$ values (1–33) and five $min.node.size$ values (3, 5, 10, 15, and 20). Hyperparameter tuning was carried out using the "caret" package (Kuhn, 2008) with the "ranger" method in R. Model selection followed the one-standard-error (one-SE) criterion, which selects the most parsimonious model whose cross-validation error falls within one standard error of the minimum, thereby improving generalizability and model stability.

To account for differences in tree species composition, all samples were categorized into three forest types: pure larch (LP),

pure birch (BP), and mixed larch–birch (LB) forests. For each forest type, spatial clustering was first applied to partition the samples into 10 spatially coherent clusters. These clusters were then used to divide the data into training and testing sets at an 8:2 ratio, ensuring that the residuals of the testing set exhibited no significant spatial autocorrelation. Independent RF models were then developed for each forest type using their respective training datasets.

2.5 Model evaluation

Traditional random or stratified k -fold cross-validation can produce overly optimistic performance estimates when applied to spatial data, as nearby samples often exhibit spatial autocorrelation. To address this issue, a blocked cross-validation approach that explicitly accounts for spatial dependence was adopted (Roberts et al., 2017). Specifically, the training samples for each forest type were divided into 10 spatial clusters based on geographic coordinates. A 10-fold blocked cross-validation was then performed, in which each cluster was used once as a validation set while the remaining nine clusters served as the training data (Supplementary Figures S3–S5). This procedure effectively reduces spatial leakage and provides a more realistic assessment of model predictive performance.

Model evaluation metrics included the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) (Equations 3–5). The optimal hyperparameter combination was determined based on the one-SE criterion described above. Finally, the predictive accuracy of each optimized RF model was independently validated using the reserved test set to ensure unbiased assessment of model performance.

Cross-validated coefficient of determination (R_{CV}^2):

$$R_{CV}^2 = \frac{1}{k} \sum_{j=1}^k (R_{CV}^2)_j = \frac{1}{k} \sum_{j=1}^k \left(1 - \frac{\sum_{i=1}^{n_j} (\Delta V_{ij} - \Delta \hat{V}_{ij})^2}{\sum_{i=1}^{n_j} (\Delta V_{ij} - \Delta \bar{V}_{ij})^2} \right) \quad (3)$$

Cross-validated root mean squared error (RMSE_{cv}):

$$\begin{aligned} \text{RMSE}_{cv} &= \frac{1}{k} \sum_{j=1}^k (\text{RMSE}_j) \\ &= \frac{1}{k} \sum_{j=1}^k \left(\sqrt{\frac{j}{n_j} \sum_{i=1}^{n_j} (\Delta V_{ij} - \Delta \hat{V}_{ij})^2} \right) \end{aligned} \quad (4)$$

Cross-validated mean absolute error (MAE_{cv}):

$$\text{MAE}_{cv} = \frac{1}{k} \sum_{j=1}^k (\text{MAE}_j) = \frac{1}{k} \sum_{j=1}^k \left(\frac{1}{n_j} \sum_{i=1}^{n_j} |\Delta \hat{V}_{ij} - \Delta V_{ij}| \right) \quad (5)$$

where, k is the number of folds (here, $k = 10$); ΔV_{ij} is the observed volume growth for the i -th sample in the j -th fold; $\Delta \hat{V}_{ij}$ is the corresponding predicted value; $\Delta \bar{V}_{ij}$ is the mean of the observed value in fold j ; and n_j is the number of samples in the j -th fold. The metrics R_j^2 , RMSE_j and MAE_j represent the evaluation results for the j -th fold. R_{CV}^2 reflects the model's goodness-of-fit, with values closer to 1 indicating stronger performance. RMSE_{cv} and MAE_{cv}

quantify the prediction error, with lower values signifying better predictive accuracy.

2.6 Spatial autocorrelation analysis

To characterize the spatial distribution and clustering patterns of stand volume growth across different forest types, a hotspot analysis (*Getis-Ord G_i^**) was conducted for pure larch (LP), birch (BP), and mixed larch–birch (LB) forest plots. The *Getis-Ord G_i^** statistic identifies local spatial autocorrelation by quantifying the degree to which high or low values are spatially clustered in the study area (Getis and Ord, 1992). The hotspot analysis was performed in ArcGIS 10.8. Stand volume growth was subsequently categorized using the equal-interval classification method to visualize spatial clustering patterns for each forest type.

3 Results

3.1 Spatial distribution pattern of stand volume growth

The volume growth of the three forest types (pure larch, pure birch, and mixed larch–birch forest) was categorized into three levels: low, medium, and high, using the equal interval classification method (Figure 2). The spatial distribution of volume growth exhibited a similar pattern across all forest types, with low-growth areas dominating the landscape and accounting for more than 50% of the total areas. Hotspot analysis using the *Getis-Ord G_i^** statistic revealed that high-growth areas were clustered in the northwestern part of study area, while low-growth areas were concentrated in the southeastern regions (Supplementary Figure S6). These findings highlight distinct spatial patterns of stand volume growth, with pronounced high-value clusters in the northwest and low-value clusters in the southeast.

3.2 Performance of RF models for stand volume growth

The 10-fold cross-validation procedure yielded three optimized RF models for stand volume growth corresponding to the three forest types. The optimal hyperparameter values were as follows: *mtry* = 5 and *min.node.size* = 15 for LP, *mtry* = 8 and *min.node.size* = 5 for BP, and *mtry* = 8 and *min.node.size* = 10 for LB. Among the models, the BP model achieved the highest predictive performance, with $R^2 = 0.92$, $\text{RMSE} = 1.89$, and $\text{MAE} = 1.24$. The LB model followed, with $R^2 = 0.88$, $\text{RMSE} = 2.68$, and $\text{MAE} = 1.89$, while the LP model showed the lowest performance, with $R^2 = 0.81$, $\text{RMSE} = 3.28$, and $\text{MAE} = 2.31$.

Model performance was further evaluated using an independent test set (Figure 3). The LB model achieved the highest predictive accuracy, with an R^2 of 0.54, RMSE of 2.73, and MAE of 1.91. The BP model followed, with an R^2 of 0.50, RMSE of 2.05, and MAE of 1.64, while the LP model showed the lowest performance, with an R^2 of 0.43, RMSE of 3.53, and MAE of 2.49. Additionally, residual analyses of the

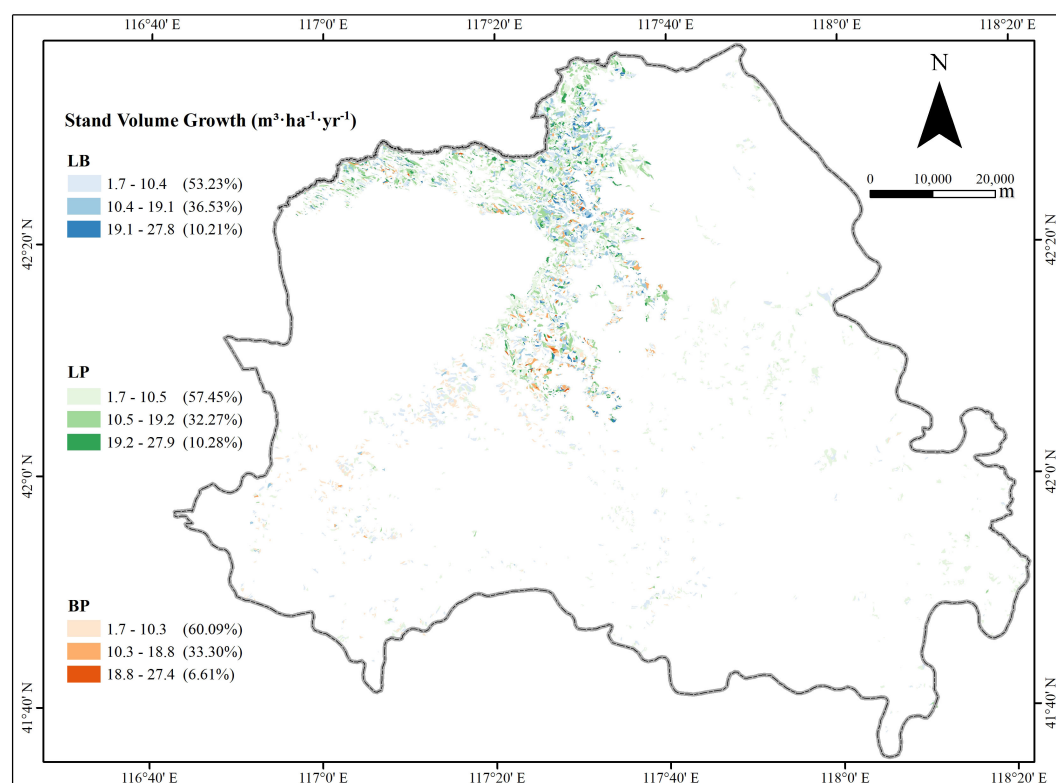


FIGURE 2
The spatial distribution of the volume growth for the three stand types.

test set revealed no significant spatial autocorrelation across all three models (Supplementary Figure S7), indicating that the predictions were not influenced by spatial dependencies and confirming the models' robustness and generalizability.

3.3 Relative importance of variables

The dominant factors influencing stand volume growth were generally consistent across the three forest types, with the minimum temperature of the coldest month (Bio12) identified as the most

influential predictor in all models (Figure 4). Specifically, Bio12 contributed 20.6%, 35.9%, and 28.3% to the total relative importance in the LP, BP, and LB models, respectively. Friedman's H-statistic further revealed that stand age exhibited the strongest overall interaction effects in the LP and BP models, with H values of approximately 0.2 in both cases (Figures 5, 6). In the LP model, the most pronounced interaction occurred between stand age and canopy density ($H = 0.16$), whereas in the BP model, the interaction between stand age and Bio12 was dominant ($H = 0.07$). In contrast, the LB model showed that Bio12 had the highest overall interaction strength ($H = 0.31$, Figure 7), with its strongest interaction observed with stand age ($H = 0.08$).

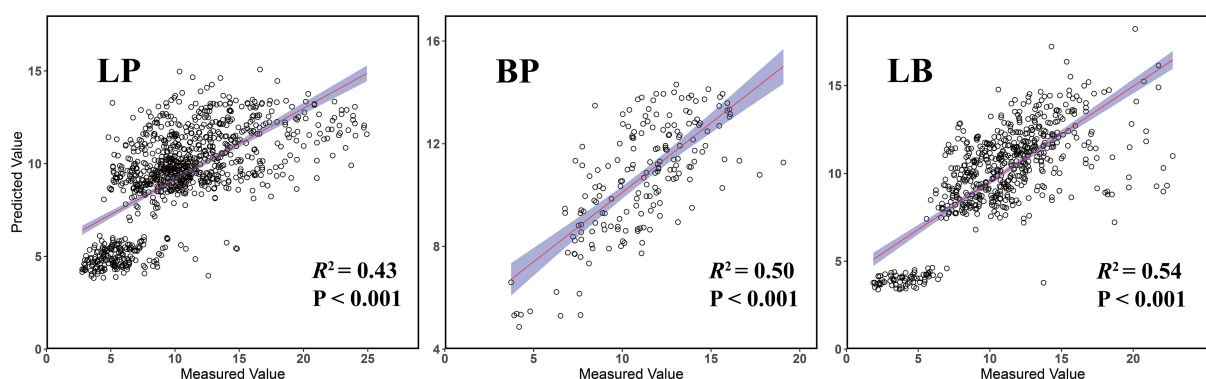
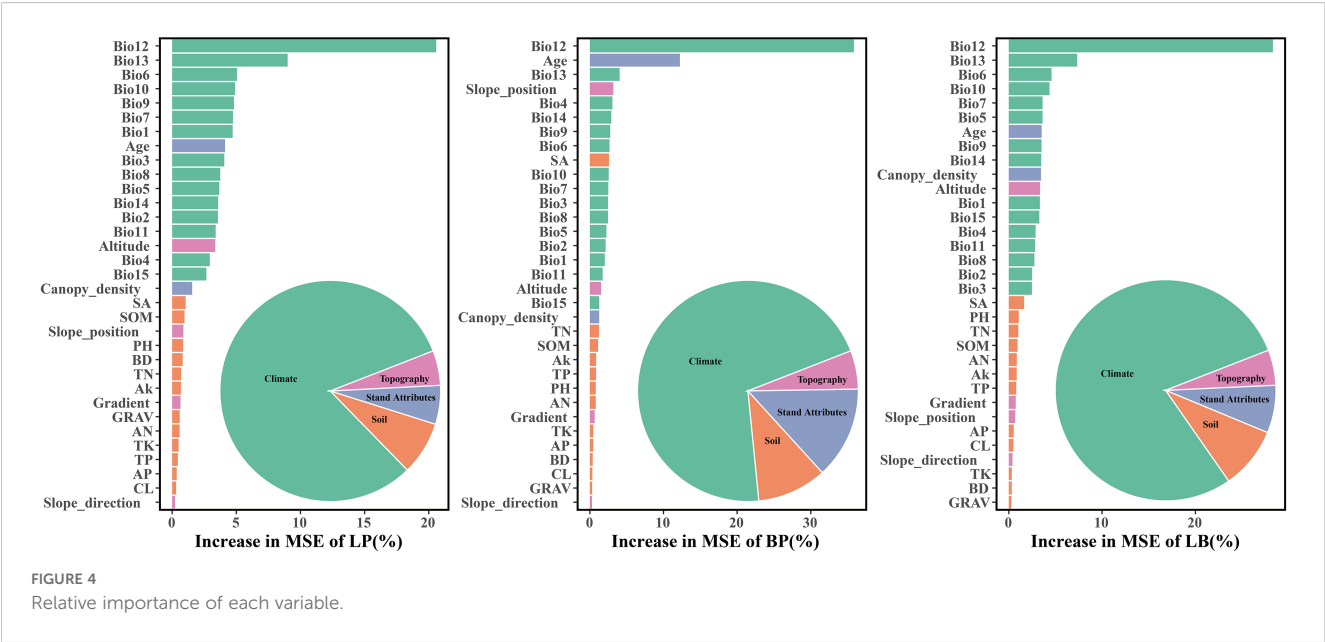


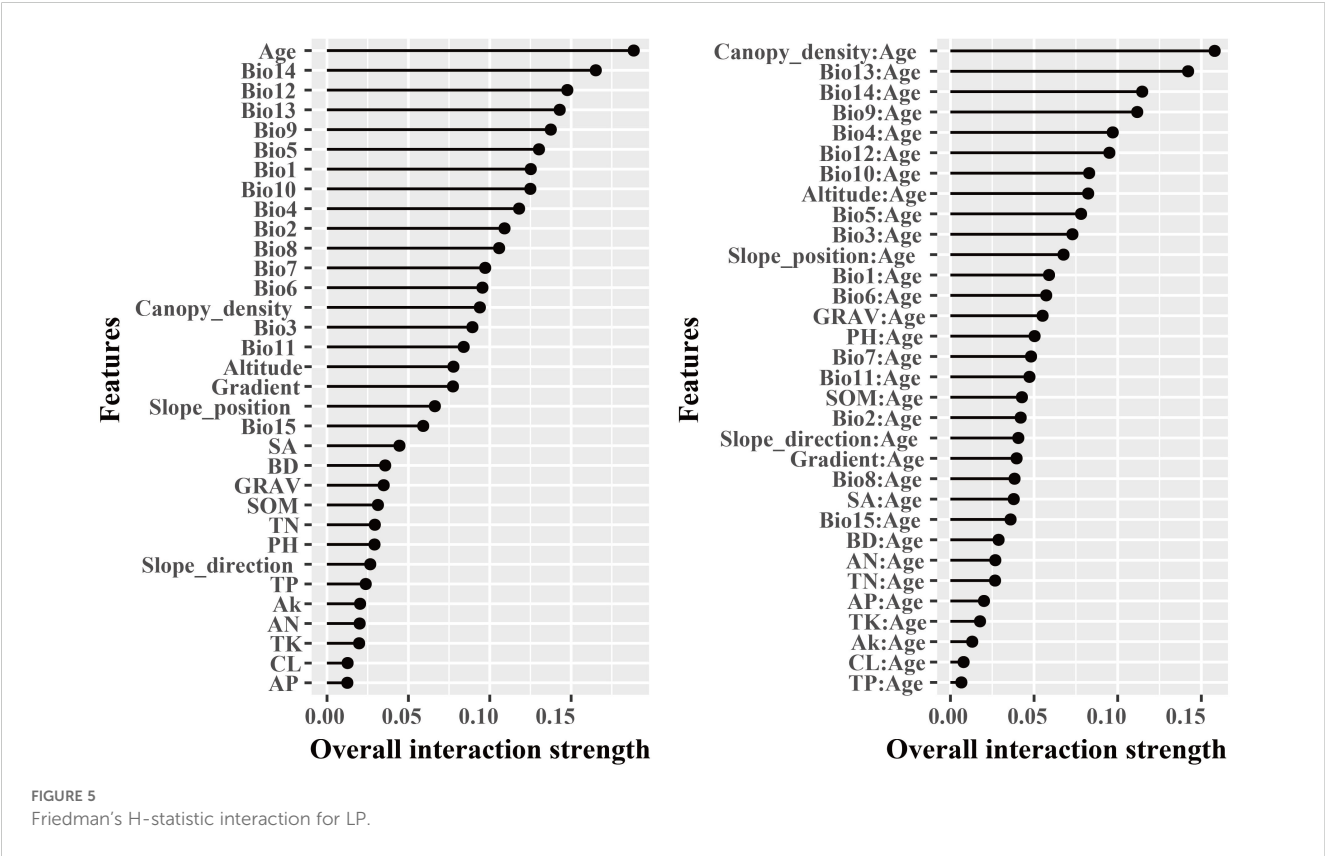
FIGURE 3
Validated prediction of random forest regression models of three forest types.



3.4 Partial dependence relationships of predictors

Among the three forest types, the LB stands exhibited the highest stand volume growth (ΔV). All three models showed a fluctuating decline in ΔV in response to Bio12 (Figure 8). The

distribution of Bio12 was bimodal, concentrated in both low and high value ranges (Figure 9). When Bio12 was in the lower range, ΔV increased, while it declined at higher Bio 12 values (Supplementary Figures S8-S10). The response of ΔV to stand age showed a unimodal pattern across all forest types, peaking at approximately 30 years and gradually declining thereafter. The



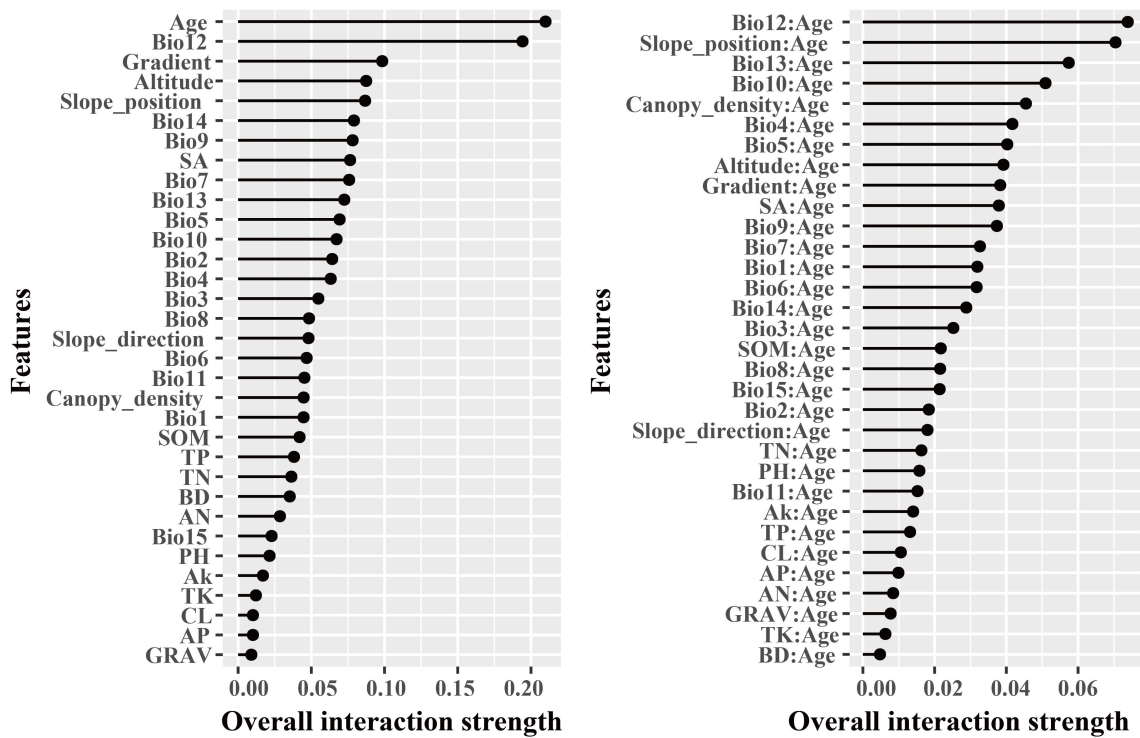


FIGURE 6
Friedman's H-statistic interaction for BP.

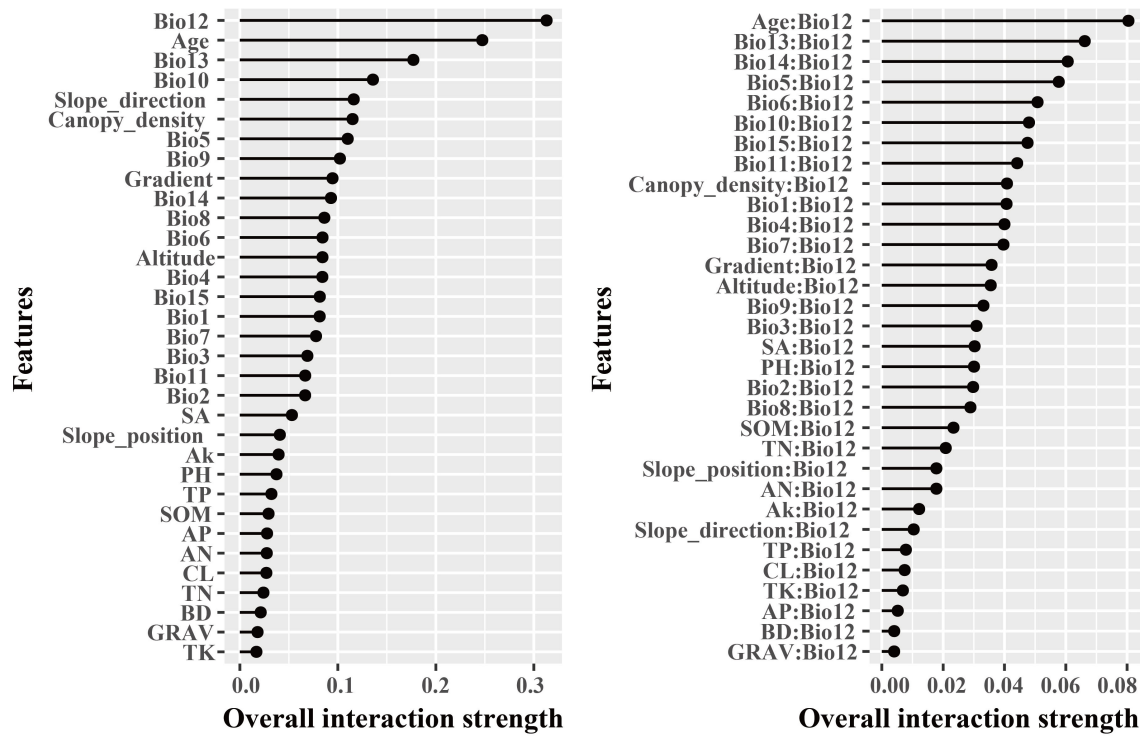
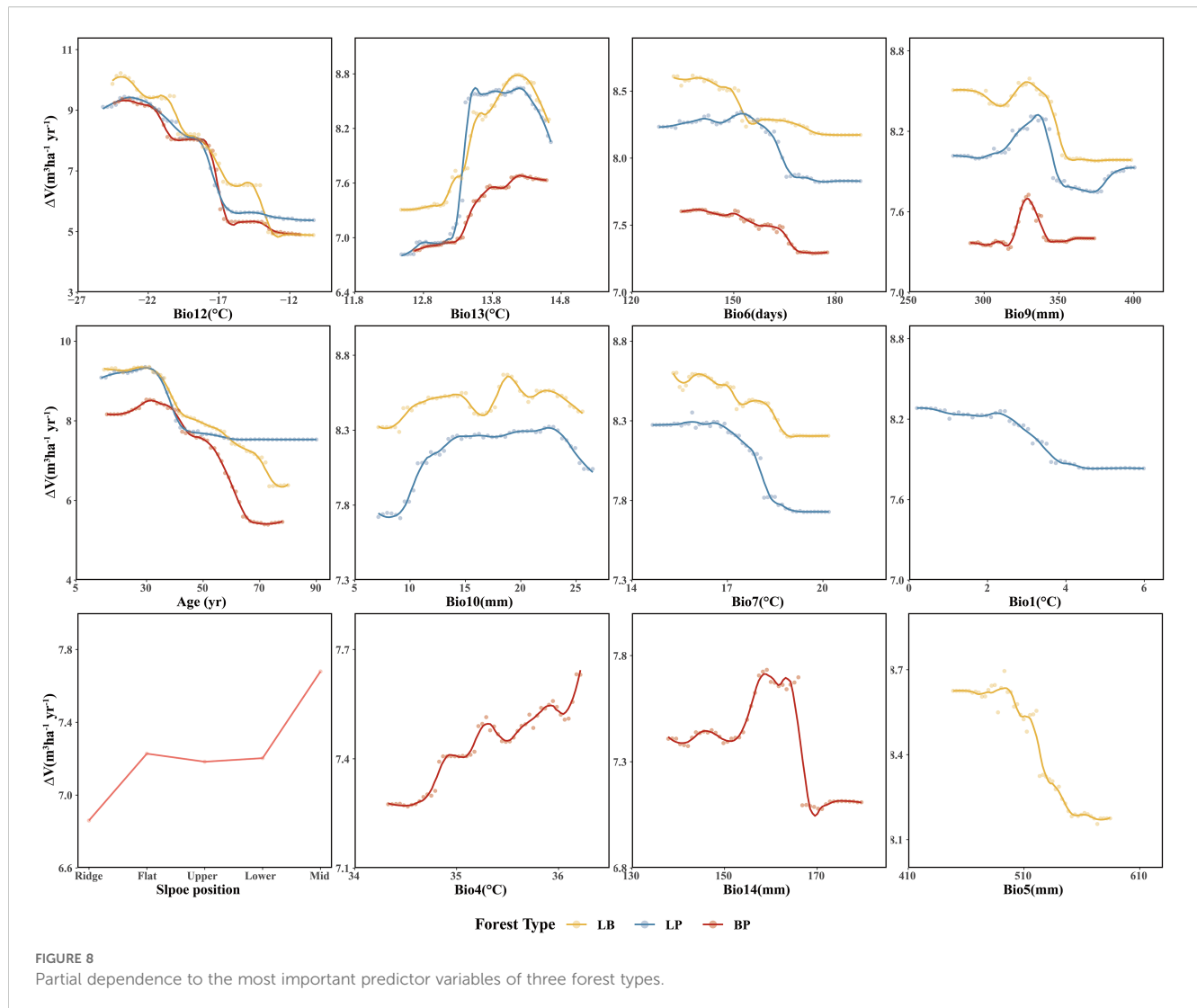


FIGURE 7
Friedman's H-statistic interaction for LB.



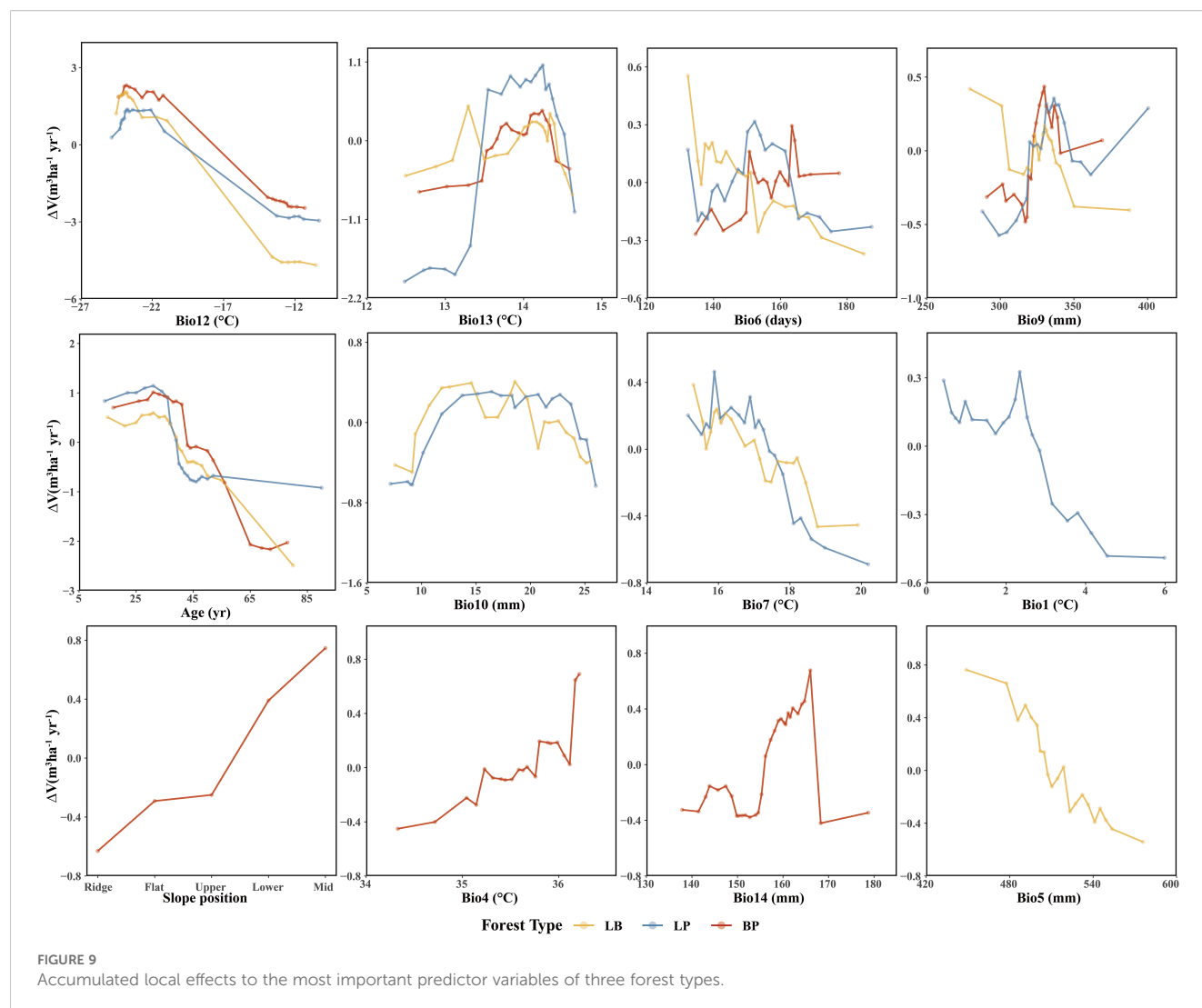
response of ΔV to Bio6 was also sensitive, with LB showing the highest ΔV in the high-value range of Bio6, while LP and BP exhibited the highest ΔV at intermediate Bio6 values. BP showed greater sensitivity to slope position, with higher ΔV observed at mid- and lower-slope positions.

For climatic factors, the response trends were similar across forest types, although the magnitude of variation differed. Among the precipitation-related factors, summer precipitation (Bio9), winter precipitation (Bio10), precipitation of the wettest month (Bio14), and annual precipitation (Bio5) showed similar nonlinear responses, characterized by an initial facilitation of ΔV followed by suppression at higher values. In contrast, temperature-related responses were more complex. The mean diurnal range (Bio13) exhibited an increase followed by a decrease in ΔV across all forest types. The response of LB and LP to the mean temperature of summer (Bio7) showed an overall declining trend, with higher ΔV at lower Bio7 values. Similarly, the response of LP to annual mean temperature (Bio1) resembled that of Bio7. In BP, the response to the monthly temperature difference (Bio4) showed a fluctuating increase, with higher ΔV observed at elevated Bio4 values.

4 Discussion

4.1 RF model evaluation

The 10-fold spatial cross-validation procedure demonstrated that all three random forest models (LP, BP, and LB) exhibited robust performance on the training set, confirming their effectiveness in modeling stand volume growth across different forest types. Among them, the BP model achieved the highest R^2 (0.92), followed by the LB ($R^2 = 0.88$) and LP ($R^2 = 0.81$) models. These results indicate that the RF models effectively captured the major determinants of forest growth, with particularly strong performance in the BP and LB forests. This superior performance likely reflects the inherent spatial structure and ecological complexity of these forest types. Specifically, BP forests, which are primarily natural in origin (Chen and Lou, 2019; Zhang et al., 2022a), and LB mixed forests, characterized by higher degree of complexity and environmental heterogeneity, provide richer variability for model learning and generalization. The LP model, derived mainly from homogeneous plantation stands, also



demonstrated good performance, consistent with previous studies that plantations tend to exhibit strong model fit due to uniform stand structure and controlled silvicultural practices (Tian et al., 2022). However, all models exhibited reduced performance on the independent test set, which may be attributed to the intrinsic ecological complexity of forest ecosystems. These complex factors such as interspecies competition, resource partitioning, and microenvironmental heterogeneity are difficult to capture fully using standard predictor variables, which may lead to decreased predictive accuracy when applied to novel data (Liu et al., 2022).

4.2 Differential performance of predictors in models

Variable importance was quantified using the Percent Increase in MSE (%IncMSE), which assesses the reduction in model accuracy when a given predictor is permuted or removed (Breiman, 2001). Across all forest types, climatic variables overwhelmingly dominated model performance, with the minimum temperature of the coldest

month (Bio12) emerging as the most influential factor in the LP, BP, and LB models, contributing 19.8%, 36.4%, and 28.6% to total relative importance, respectively (Figure 4). Stand age consistently ranked among the top eight predictors across models, reaffirming its pivotal role in regulating stand development and productivity. This finding is consistent with previous findings: Tian et al. (2022) identified stand age as the primary driver of *Larix* volume growth, while Zhang et al. (2022a) highlighted the importance of stand structure for *Betula* radial growth. Similar conclusions have been drawn in species-specific studies, where structural attributes and competition indices were found to be dominant predictors of forest productivity (Ou et al., 2019; Jevšenak and Skudnik, 2021). Beyond variable importance, Friedman's H-statistic provided additional insights into the strength of interactions among predictors (Figures 5–7). Interaction analyses revealed that stand age exhibited the highest overall interaction effects in both LP and BP models ($H \approx 0.2$), suggesting that developmental stage modulates the sensitivity of stands to environmental drivers. In the LP model, the interaction between stand age and canopy density ($H = 0.16$) indicated that stand structural attributes mediate the influence of climatic constraints on stand growth. In the BP model, the

interaction between stand age and Bio12 ($H = 0.07$) suggested that birch growth is highly sensitive to cold stress and that temperature effects are contingent on stand maturity. In contrast, the LB model showed that Bio12 itself had the highest overall interaction strength ($H = 0.31$), with its strongest interaction observed with stand age ($H = 0.08$), implying that temperature limitations and developmental dynamics jointly regulate mixed-stand growth patterns. Integrating the %IncMSE and H-statistic results indicates that the most influential predictors—particularly Bio12 and stand age—function not only as dominant main effects but also as key interaction hubs shaping forest growth responses. These findings support previous evidence that climatic factors and stand attributes jointly regulate productivity across temperate forest ecosystems (Tian et al., 2022; Zhang et al., 2022a; Li et al., 2022; Liu et al., 2025).

4.3 Key factors in the stand volume growth models of three forest types

This study revealed a strong dependence of volume growth in LP (pure larch) on stand age (Figure 8), exhibiting a staged growth pattern characteristic of typical allometric trajectory in plantation development (Zhao et al., 2011; Qi et al., 2016). In the young (≤ 20 years) and middle-aged (21–40 years) stages, elevated growth rates were primarily driven by larch's biological traits and favorable site conditions. During this period, larch accumulates trunk biomass most rapidly (Zhao et al., 2011), supported by sufficient sunlight and reduced belowground competition for nutrients (Rosenthal and Camm, 1996; Wang et al., 2017a). However, growth peaked around 30 years and subsequently declined, likely due to the effects of selective thinning. As stands entered the near-mature (41–50 years) and mature (> 50 years) stages, volume growth stabilized, possibly as a result of nutrient depletion in deeper soil layers, where root expansion becomes increasingly constrained by limited nutrient availability (Wang et al., 2019). This observed partial dependence on stand age aligns with the findings of Tian et al. (2022).

LP volume growth also exhibited a threshold response to Bio12 (minimum temperature of the coldest month), likely due to premature dormancy break. Trees, adapted to low temperatures, reduces metabolic activity and enters dormancy during winter. If winter temperatures are not sufficiently low, accumulated heat may trigger early budburst, increasing vulnerability to late frost events (Gömöry et al., 2015), which in turn depletes stored carbon and nutrient reserves due to the need for re-sprouting (Rubio-Cuadrado et al., 2021). Moreover, premature budburst prior to soil thaw may induce physiological drought. Similar threshold responses were observed for Bio1, and Bio7. In the study area, temperature emerged as the primary climatic driver of LP growth, consistent with national-scale findings (Tian et al., 2022). Bio1, Bio7, and Bio13 significantly influenced growth, while other precipitation-related factors (e.g., Bio9, Bio10) had relatively minor impacts.

BP (pure birch) exhibited a similar ranking of predictor importance and response patterns to LP, reflecting convergent growth strategies under comparable environmental constraints (Gömöry et al., 2015; Oksanen, 2021). The stand volume growth

of BP showed high sensitivity to slope position, particularly with more vigorous growth observed on middle and lower slopes. This phenomenon is likely related to the impact of slope position on water availability, light exposure, and soil nutrient status. Middle and lower slopes tend to accumulate more surface runoff and groundwater, providing a more stable water supply. Additionally, these slopes, particularly those with favorable aspects, generally receive greater sunlight exposure, enhancing photosynthetic activity. Lower slopes also tend to have deeper and more fertile soils due to the accumulation of organic matter and nutrients, whereas upper slopes are typically characterized by faster drainage and lower soil fertility. Consequently, BP forests on middle and lower slopes benefit from more favorable hydrothermal and edaphic conditions, leading to enhanced stand volume growth.

In LB (larch-birch mixed forest), volume growth was consistently higher than that in pure stands (Figure 8), suggesting that moderate interspecific competition enhances mutual facilitation, in line with findings by Zhang et al. (2022b). The diurnal temperature range (Bio13) also exerted a significant effect: moderate increases enhanced energy accumulation and growth (Fu et al., 2016), whereas excessive daytime temperatures increased transpiration rates and water loss, leading to growth suppression (Sadok et al., 2021). The increased relative importance of Bio13 suggests that interspecific competition amplifies sensitivity to thermal variability, consistent with the findings of Forrester (Forrester, 2015). In both LP and BP pure forest types, the response to the frost-free period (Bio6) shows an adaptation to moderate climatic conditions. In these forests, stand growth is more vigorous under moderate frost-free periods, suggesting that such conditions provide an optimal balance between temperature, moisture availability, and physiological activity. This moderate duration likely supports sustained metabolic processes and long-term growth stability in these forest types. In contrast, LB mixed forests exhibited the highest growth under shorter frost-free periods, implying greater adaptability to cooler or more variable climatic conditions. This pattern may reflect reduced interspecific competition and enhanced resource-use efficiency in mixed stands, enabling them to maintain higher productivity even under shorter growing seasons.

4.4 Study limitations and recommendations

The RF models demonstrated strong predictive performance across all three stand types, owing to the algorithm's robust capability to interpret high-dimensional datasets. However, the representation of stand structure in this study was limited, relying solely on stand age and canopy density as structural predictors of volume growth. This limitation may have affected the model's generalization capacity by restricting its ability to capture the underlying structural complexity of forest ecosystems. While the lack of detailed structural variables likely contributed to this limitation, other sources of uncertainty—such as potential bias in volume equations and the exclusion of disturbance history—should also be considered when interpreting the model outputs. Moreover,

the exclusive reliance on the Random Forest (RF) model, without comparison to alternative approaches such as linear mixed-effects models (LMEM), generalized additive models (GAM), or gradient boosting methods (GBM), constitutes an limitation of this study. Although RF is well suited for analyzing complex and high-dimensional ecological data, future research should incorporate multiple modeling frameworks to benchmark performance and evaluate the generalizability of predictions across diverse forest types and climatic conditions. Such comparative analyses would provide a more comprehensive assessment of model suitability and strengthen confidence in the resulting predictions.

Future research should also incorporate a broader range of stand structural parameters, including interspecific competition indices, tree size distributions, and niche overlap metrics, to better represent ecosystem complexity and improve model performance. The high relative importance of Bio12 highlights the significant stress imposed by extreme climatic events and emphasizes the necessity of incorporating disturbance-related variables into future models' frameworks. Notably, mixed stands exhibited higher productivity than pure stands, indicating that species diversity enhances stand stability under climate variability. Based on these findings, forest management practices in the mountainous region of northern Hebei should prioritize optimizing stand structure and mitigating the effects of climate change on forest productivity. Recommended strategies include the underplanting of complementary tree species to improve the structural and functional diversity of pure stands, as well as the implementation of adaptive management practices in near-mature and mature forests to promote continued volume growth.

5 Conclusions

This study employed Random Forest (RF) models to predict the volume growth of larch (LP), birch (BP), and mixed larch–birch (LB) forests in northern China, using key climatic and environmental variables. The models demonstrated strong predictive performance, with the BP model achieving the highest explanatory power ($R^2 = 0.92$). Among the predictors, the minimum temperature of the coldest month (Bio12) emerged as the most influential factor across all forest types, underscoring the dominant role of temperature in regulating forest productivity. Stand age also exhibited a strong effect on volume growth and showed pronounced interactions with other predictors, particularly in the LP and BP models. The analysis further revealed distinct growth-regulating mechanisms among forest types: temperature-related factors and stand age were critical drivers for LP and BP stands, whereas the enhanced growth observed in LB mixed forests was largely attributed to species diversity and interspecific facilitation. These results suggest that mixed forests possess greater resilience to climatic variability and disturbance, leading to higher productivity than pure stands. From a management perspective, optimizing forest structure, promoting species diversity, and implementing adaptive management practices should be prioritized to enhance forest stability and productivity in the mountainous regions of northern Hebei. Future research should integrate additional structural descriptors, such as tree size distribution, interspecific

competition indices, and disturbance history, to refine model accuracy and deepen understanding of forest growth dynamics under changing climatic conditions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

JZ: Formal Analysis, Writing – original draft, Data curation, Visualization, Investigation, Conceptualization, Methodology, Software. ZW: Formal Analysis, Writing – original draft, Methodology, Validation. ML: Data curation, Writing – original draft, Formal Analysis. XC: Validation, Conceptualization, Writing – original draft. YP: Funding acquisition, Data curation, Supervision, Writing – review & editing. ZZ: Writing – review & editing, Supervision, Funding acquisition, Resources, Data curation.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was supported by the State Key Research and Development Program (2023YFD2200803); Major Science and Technology Support Program of Hebei Province (252L6802D; 252L6801D); Natural Science Foundation of Hebei Province (C2023204170); and Key Technological Innovation and Demonstration Projects of Forestry and Grassland Bureau of Hebei Province (2406101).

Acknowledgments

The authors thank everyone who helped with the field survey and the reviewers for their valuable comments.

Conflict of interest

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

Generative AI statement

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

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1682940/full#supplementary-material>

References

- Aalde, H., Gonzalez, P., Gytarsky, M., Krug, T., Kurz, W. A., Ogle, S., et al. (2006). "Agriculture, forestry and other land use," in *Guidelines for national Greenhouse Gas Inventories*. Eds. H. S. Eggleston, L. Buendia, K. Miwa, T. Ngara and K. Tanabe (IGES, Japan).
- Apley, D. W., and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B, Stat. Methodol.* 82, 1059–1086. doi: 10.1111/rssb.12377
- Asner, G. P., and Martin, R. E. (2008). Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote. Sens. Environ.* 112, 3958–3970. doi: 10.1016/j.rse.2008.07.003
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Chen, T., and Lou, A. (2019). Phylogeography and paleodistribution models of a widespread birch (*Betula platyphylla* Suk.) across East Asia: Multiple refugia, multidirectional expansion, and heterogeneous genetic pattern. *Ecol. Evol.* 9, 7792–7807. doi: 10.1002/eece3.5365
- Cleveland, W. S., and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* 83, 596–610. doi: 10.1080/01621459.1988.10478639
- Dincă, L., Murariu, G., Enescu, C. M., Achim, F., Georgescu, L., Murariu, A., et al. (2020). Productivity differences between southern and northern slopes of Southern Carpathians (Romania) for Norway spruce, silver fir, birch and black alder. *Not. Bot. Horti, Agrobo.* 48, 1070–1084. doi: 10.15835/nbha48211824
- Forrester, D. I. (2015). Transpiration and water-use efficiency in mixed-species forests versus monocultures: effects of tree size, stand density and season. *Tree Physiol.* 35, 289–304. doi: 10.1093/treephys/tpv011
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.1214/aos/1013203451
- Friedman, J. H., and Popescu, B. E. (2008). Predictive learning via rule ensembles. *Ann. Appl. Stat.* 2, 916–954. doi: 10.1214/07-AOAS148
- Fu, Y. H., Liu, Y., De Boeck, H. J., Menzel, A., Nijs, I., Peaucelle, M., et al. (2016). Three times greater weight of daytime than of night-time temperature on leaf unfolding phenology in temperate trees. *New Phytol.* 212, 590–597. doi: 10.1111/nph.14073
- Getis, A., and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geogr. Anal.* 24, 189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x
- Genuer, R., Poggi, J. M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recogn. Lett.* 31, 2225–2236. doi: 10.1016/j.patrec.2010.03.014
- Gömöry, D., Foffová, E., Longauer, R., and Krajmerová, D. (2015). Memory effects associated with early-growth environment in Norway spruce and European larch. *Eur. J. For. Res.* 134, 89–97. doi: 10.1007/s10342-014-0835-1
- Hamidi, S. K., Zenner, E. K., Bayat, M., and Fallah, A. (2021). Analysis of plot-level volume increment models developed from machine learning methods applied to an uneven-aged mixed forest. *Ann. For. Sci.* 78, 1–16. doi: 10.1007/s13595-020-01011-6
- He, X., Lei, X., Zeng, W., Feng, L., Zhou, C., and Wu, B. (2022). Quantifying the effects of stand and climate variables on biomass of larch plantations using random forests and national forest inventory data in north and northeast China. *Sustainability* 14, 5580. doi: 10.3390/su14095580
- Jacoby, W. G. (2000). Loess: a nonparametric, graphical tool for depicting relationships between variables. *Elect. Stud.* 19, 577–613. doi: 10.1016/S0261-3794(99)00028-1
- Jevšenak, J., and Skudnik, M. (2021). A random forest model for basal area increment predictions from national forest inventory data. *For. Ecol. Manage.* 479, 118601. doi: 10.1016/j.foreco.2020.118601
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Software* 28, 1–26. doi: 10.18637/jss.v028.i05
- Lei, X. (2019). Applications of machine learning algorithms in forest growth and yield prediction. *J. Beijing For. Univ.* 41, 23–36. doi: 10.12171/j.1000-1522.20190356
- Li, X., Pialuang, B., Kang, W., Ji, X., Zhang, H., Xue, Z., et al. (2022). Responses of radial growth to climate change over the past decades in secondary *Betula platyphylla* forests in the mountains of northwest Hebei, China. *Chin. J. Plant Ecol.* 46, 919–931. doi: 10.17521/cjpe.2021.0253
- Liu, D., An, Y., Li, Z., Wang, Z., Zhao, Y., and Wang, X. (2024). Differences and similarities in radial growth of *Betula* species to climate change. *J. For. Res.* 35, 40. doi: 10.1007/s11676-023-01690-7
- Liu, Y., Gong, W., Hu, X., and Gong, J. (2018). Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* 10, 946. doi: 10.3390/rs10060946
- Liu, J., He, J., Chai, L., Zhong, X., Jia, B., and Wang, X. (2022). A comparison of models of stand volume in spruce-fir mixed forest in northeast China. *Forests* 13, 1117. doi: 10.3390/f13071117
- Liu, Q., Shi, Q., Tong, X., Shi, M., Wang, Y., and Tan, N. (2025). The impact of hydrothermal conditions on the radial growth of *Larix gmelinii* var. *principis-rupprechtii* plantations. *Acta Ecol. Sin.* 10, 1–16. doi: 10.20103/j.stxb.202406281510
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 4765–4774. doi: 10.1038/s42256-019-0138-9
- Malhi, Y., Meir, P., and Brown, S. (2002). Forests, carbon and global climate. *Phil. Trans. R. Soc. Lond. A.* 360, 1567–1591. doi: 10.1098/rsta.2002.1020
- Molnar, C., Casalicchio, G., and Bischl, B. (2018). iml: An R package for interpretable machine learning. *J. Open Source Software* 3, 786. doi: 10.21105/joss.00786
- Morales-Barquero, L., Lyons, M. B., Phinn, S. R., and Roelfsema, C. M. (2019). Trends in remote sensing accuracy assessment approaches in the context of natural resources. *Remote Sens.* 11, 2305. doi: 10.3390/rs11192305
- National Forestry and Grassland Administration (2021). *The 14th five-year plan for forestry and grassland conservation and development* (China: Government Website of the National Forestry and Grassland Administration). Available online at: <http://202.99.63.178/c/www/lczy/12618.jhtml>.
- Oksanen, E. (2021). Birch as a model species for the acclimation and adaptation of northern forest ecosystem to changing environment. *Front. For. Glob. Change* 4. doi: 10.3389/ffgc.2021.682512
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote. Sens. Environ.* 148, 42–57. doi: 10.1016/j.rse.2014.02.015
- Olorunfemi, I. E., Fasimirin, J. T., Olufayo, A. A., and Komolafe, A. A. (2020). GIS and remote sensing-based analysis of the impacts of land use/land cover change (LULCC) on the environmental sustainability of Ekiti State, southwestern Nigeria. *Environ. Dev. Sustain.* 22, 661–692. doi: 10.1007/s10668-018-0214-z
- Ou, Q., Lei, X., and Shen, C. (2019). Individual tree diameter growth models of larch-spruce-fir mixed forests based on machine learning algorithms. *Forests* 10, 187. doi: 10.3390/f10020187
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., et al. (2011). A large and persistent carbon sink in the world's forests. *Science* 333, 988–993. doi: 10.1126/science.1201609
- Qi, G., Chen, H., Zhou, L., Wang, X., Zhou, W., Qi, L., et al. (2016). Carbon stock of larch plantations and its comparison with an old-growth forest in Northeast China. *Chin. Geogr. Sci.* 26, 10–21. doi: 10.1007/s11769-015-0772-z
- Rehman, A., Ma, H., Ahmad, M., Irfan, M., Traore, O., and Chandio, A. A. (2021). Towards environmental Sustainability: Devolving the influence of carbon dioxide emission to population growth, climate change, Forestry, livestock and crops production in Pakistan. *Ecol. Indic.* 125, 107460. doi: 10.1016/j.ecolind.2021.107460
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–929. doi: 10.1111/ecog.02881

- Rosenthal, S., and Camm, E. (1996). Effects of air temperature, photoperiod and leaf age on foliar senescence of western larch (*Larix occidentalis* Nutt.) in environmentally controlled chambers. *Plant Cell Environ.* 19, 1057–1065. doi: 10.1111/j.1365-3040.1996.tb00212.x
- Rubio-Cuadrado, Á., Gómez, C., Rodríguez-Calcerrada, J., Perea, R., Gordaliza, G. G., Camarero, J. J., et al. (2021). Differential response of oak and beech to late frost damage: an integrated analysis from organ to forest. *Agr. For. Meteorol.* 297, 108243. doi: 10.1016/j.agrformet.2020.108243
- Sadok, W., Lopez, J. R., and Smith, K. P. (2021). Transpiration increases under high-temperature stress: Potential mechanisms, trade-offs and prospects for crop resilience in a warming world. *Plant Cell Environ.* 44, 2102–2116. doi: 10.1111/pce.13970
- Shi, C., Shen, M., Wu, X., Cheng, X., Li, X., Fan, T., et al. (2019). Growth response of alpine treeline forests to a warmer and drier climate on the southeastern Tibetan Plateau. *Agr. For. Meteorol.* 264, 73–79. doi: 10.1016/j.agrformet.2018.10.002
- Soleimannejad, L., Ullah, S., Abedi, R., Dees, M., and Koch, B. (2019). Evaluating the potential of sentinel-2, landsat-8, and iris satellite images in tree species classification of hyrcanian forest of Iran using random forest. *J. Sustain. For.* 38, 615–628. doi: 10.1080/10549811.2019.1598443
- Solomon, S., Plattner, G.-K., Knutti, R., and Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1704–1709. doi: 10.1073/pnas.0812721106
- Standardization Administration of the People's Republic of China (2024). *Tree biomass models and related parameters to carbon accounting for major tree species*. Beijing: Standards Press of China.
- Tian, H., Zhu, J., He, X., Chen, X., Jian, Z., Li, C., et al. (2022). Using machine learning algorithms to estimate stand volume growth of *Larix* and *Quercus* forests based on national-scale Forest Inventory data in China. *For. Ecosyst.* 9, 100037. doi: 10.1016/j.fecs.2022.100037
- Wang, Y. (2024). The effect of climate change on forest fire danger and severity in the Canadian boreal forests for the period 1976–2100. *J. Geophys. Res. Atmos.* 129, e2023JD039118. doi: 10.1029/2023JD039118
- Wang, H., Chen, D., and Sun, X. (2019). Nutrient allocation to different compartments of age-sequence larch plantations in China. *Forests* 10, 759. doi: 10.3390/f10090759
- Wang, J. R., Hawkins, C., and Letchford, T. (1998). Photosynthesis, water and nitrogen use efficiencies of four paper birch (*Betula papyrifera*) populations grown under different soil moisture and nutrient regimes. *For. Ecol. Manage.* 112, 233–244. doi: 10.1016/S0378-1127(98)00407-1
- Wang, J., Li, S., Guo, Y., Yang, Q., Ren, R., and Han, Y. (2022a). Responses of *Larix principis-rupprechtii* radial growth to climatic factors at different elevations on Guancen mountain, North-Central China. *Forests* 13, 99. doi: 10.3390/f13010099
- Wang, H., Wang, W., and Chang, S. X. (2017a). Sampling method and tree-age affect soil organic C and N contents in larch plantations. *Forests* 8, 28. doi: 10.3390/f8010028
- Wang, T., Wang, G., Innes, J., Seely, B., and Chen, B. (2017b). ClimateAP: An application for dynamic local downscaling of historical and future climate data in Asia Pacific. *Front. Agr. Sci. Eng.* 4, 448–458. doi: 10.15302/J-FASE-2017172
- Wang, Z., Zhang, X., Chhin, S., Zhang, J., and Duan, A. (2021). Disentangling the effects of stand and climatic variables on forest productivity of Chinese fir plantations in subtropical China using a random forest algorithm. *Agr. For. Meteorol.* 304, 108412. doi: 10.1016/j.agrformet.2021.108412
- Wang, Z., Zhang, X., Zhang, J., and Chhin, S. (2022b). Effects of stand factors on tree growth of Chinese fir in the subtropics of China depends on climate conditions from predictions of a deep learning algorithm: A long-term spacing trial. *For. Ecol. Manage.* 520, 120363. doi: 10.1016/j.foreco.2022.120363
- Xie, Z., Chen, Y., Lu, D., Li, G., and Chen, E. (2019). Classification of land cover, forest, and tree species classes with ZiYuan-3 multispectral and stereo data. *Remote Sens.* 11, 164. doi: 10.3390/rs11020164
- Zhang, H., Feng, Z., Wang, S., and Ji, W. (2022a). Disentangling the factors that contribute to the growth of *Betula* spp. and *Cunninghamia lanceolata* in China based on machine learning algorithms. *Sustainability* 14, 8346. doi: 10.3390/su14148346
- Zhang, J., Liu, Q., Wang, D., and Zhang, Z. (2023). Soil microbial community, soil quality, and productivity along a chronosequence of *Larix principis-rupprechtii* forests. *Plants* 12, 2913. doi: 10.3390/plants12162913
- Zhang, R., Pang, Y., Li, Z., and Bao, Y. (2016). Canopy closure estimation in a temperate forest using airborne LiDAR and LANDSAT ETM+ data. *Chin. J. Plant Ecol.* 40, 102. doi: 10.17521/cjpe.2014.0366
- Zhang, Y., Wang, R., Liu, C., Liu, Q., Li, M., and Zhang, Z. (2025). Biodiversity and soil jointly drive ecosystem multifunctionality in larch forests. *Forests* 16, 745. doi: 10.3390/f16050745
- Zhang, J., Zhao, J., Cheng, R., Ge, Z., and Zhang, Z. (2022b). Effects of neighborhood competition and stand structure on the productivity of pure and mixed *Larix principis-rupprechtii* forests. *Forests* 13, 1318. doi: 10.3390/f13081318
- Zhao, Q., Liu, X., and Zeng, D. (2011). Aboveground biomass and nutrient allocation in an age-sequence of *Larix olgensis* plantations. *J. For. Res.* 22, 71–76. doi: 10.1007/s11676-011-0128-1
- Zhao, K., Wang, L., Wang, L., Jia, Z., and Ma, L. (2015). Stock volume and productivity of *Larix principis-rupprechtii* in northern and northwestern China. *J. Beijing For. Univ.* 37, 24–31. doi: 10.13332/j.cnki.jbfu.2015.02.011