

RESEARCH ARTICLE

Assessing scale-dependent effects on Forest biomass productivity based on machine learning

Jingyuan He¹  | Chunyu Fan¹  | Yan Geng¹  | Chunyu Zhang¹  | Xiuhai Zhao¹  | Klaus von Gadow^{1,2,3} 

¹Research Center of Forest Management Engineering of State Forestry Administration, Beijing Forestry University, Beijing, China

²Faculty of Forestry and Forest Ecology, Georg-August-University, Göttingen, Germany

³Department of Forest and Wood Science, University of Stellenbosch, Matieland, South Africa

Correspondence

Xiuhai Zhao, Beijing Forestry University, Tsinghua East Road 35, Haidian District, Beijing 100083, China.
Email: zhaohx@bjfu.edu.cn

Funding information

Key Project of National Key Research and Development Plan, Grant/Award Number: 2017YFC0504104; National Natural Science Foundation of China, Grant/Award Number: 31971650; National Science Foundation of China, Grant/Award Number: 32171521

Abstract

Estimating forest above-ground biomass (AGB) productivity constitutes one of the most fundamental topics in forest ecological research. Based on a 30-ha permanent field plot in Northeastern China, we modeled AGB productivity as output, and topography, species diversity, stand structure, and a stand density variable as input across a series of area scales using the *Random Forest (RF)* algorithm. As the grain size increased from 10 to 200m, we found that the relative importance of explanatory variables that drove the variation of biomass productivity varied a lot, and the model accuracy was gradually improved. The minimum sampling area for biomass productivity modeling in this region was 140 × 140 m. Our study shows that the relationship of topography, species diversity, stand structure, and stand density variables with biomass productivity modeled using the RF algorithm changes when moving from scales typical of forest surveys (10 m) to larger scales (200m) within a controlled methodology. These results should be of considerable interest to scientists concerned with forest assessment.

KEYWORDS

above-ground biomass, productivity, random Forest algorithm, random spatial sampling, scale dependence

TAXONOMY CLASSIFICATION

Applied ecology

1 | INTRODUCTION

Terrestrial ecosystems play an important role in regulating the global and local climates (Gadow et al., 2021; Hwan & Chun, 2011; Zhou et al., 2013), and carbon cycles (Kuribayashi et al., 2017; Zhao et al., 2012), and maintaining biodiversity (Kitayama et al., 2018; Ren et al., 2017). Among these terrestrial ecosystems, the forest

ecosystem is the largest C reservoir, which comprises more than 80% and 40% of the global terrestrial C pools above-ground and below-ground, respectively (Dixon et al., 1994; Luo et al., 2020; Pan et al., 2011). Tree trunks and branches contain a massive ratio of these C, which is called above-ground biomass (AGB) (Fahey et al., 2010; Fotis et al., 2018). The relationships between biodiversity and AGB or productivity have drawn ample attention in ecology and

This work was supported by the Program of National Natural Science Foundation of China [grant numbers 31971650]; National Science Foundation of China [grant numbers 32171521]; and the Key Project of National Key Research and Development Plan [grant number 2017YFC0504104].

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

conservation biology (Cadotte, 2015; Cardinale et al., 2012; Isbell et al., 2011; Liang et al., 2016; Qiao et al., 2021). Numerous debates have been initiated with respect to the impacts of species diversity on productivity or AGB, with positive, negative, hump-shaped or U-shaped relationships (Mittelbach, 2010; Poorter et al., 2017; Ruiz-Benito et al., 2014; Whittaker, 2010; Zhang et al., 2015). Most studies have attended to scrutinize biomass instead of productivity (Holdaway et al., 2016). However, strictly speaking, the concepts of them are different in forest ecosystems (Chisholm et al., 2013; Schmid et al., 2009), though they are sometimes correlated positively (Stegen et al., 2011). Thus, biomass and productivity should be analyzed in different ways (Chisholm et al., 2013). The ambiguous results above declare our limited cognition on the diversity and productivity. Estimating forest biomass per unit area is important and challenging, and key to estimating forest carbon stock. Net biomass change per unit of area and time is an important index to measure forest productivity and effects of management (Hao et al., 2018; Luo et al., 2019).

Stand structure factors, such as stem density, tree size variation, and stand structural diversity, are critical components in characterizing forest productivity (Ali et al., 2016, 2019a; Morin et al., 2011; Rodriguez-Hernandez et al., 2021; Sullivan et al., 2017; Yachi & Loreau, 2007). It has been demonstrated that various layered stand structures imply multiple canopy, which contributes to more capture of light and other resources (Yachi & Loreau, 2007). Environmental conditions (i.e., topography heterogeneity in this study) in elevation, aspect and slope impact nutrient, plant traits, water availability, and biodiversity patterns directly and indirectly, which shapes them decisive predictors to project productivity (de Castilho et al., 2006; Liu et al., 2014; Zhang et al., 2012, 2016). Tree biomass has strong links to stand density (Dahlhausen et al., 2017; Mejstřík et al., 2022; Wertz et al., 2020; Xue et al., 2012). Węgiel et al. (2018) revealed that higher stand density might lead to higher total biomass production and carbon stock. The associations across diversified abiotic and biotic determinants of productivity remain debated (Ali et al., 2019a,b; Rodriguez-Hernandez et al., 2021). Moreover, the character of scale in unraveling the relationships is less well-known (but see Rodriguez-Hernandez et al., 2021). Community often shows large spatial variability due to environmental, community process, and disturbing factors (Chave et al., 2003; Mascaró et al., 2011). Generally, greater plot sizes reflect the characteristics of the community more accurately, but this ideal cannot be achieved because of the constraints of available funding. Therefore, it is necessary to design a suitable plot size to meet both scientific requirements and cost savings (Bradford et al., 2010; Chave et al., 2003; Kral et al., 2010; Laumonier et al., 2010; Pyle et al., 2008; Wagner et al., 2010). It is important to determine the smallest area of sampling plot that can be representative of the characteristics of the whole community (e.g., species composition and structure) (Peng & Guo, 2016). For instance, the sample plot area in most studies is generally smaller than 1 ha and they tend to focus on a single grain size (Holdaway et al., 2016; Li et al., 2019; Sande et al., 2017; Xu et al., 2015; Yuan, Ali, et al., 2018), which might lead to different results when viewed at even smaller

or larger spatial extents. Can the survey results based on the quadrats of the above area truly reflect forest productivity? Is there a minimum area that can fully reflect the spatial variability of stand productivity? So far, there seems to be no reliable evaluation basis. Furthermore, several studies have used the species–area curve to determine the minimum plot area (Harte et al., 2009; Kallimanis et al., 2008; Tikkanen et al., 2009), but this cannot reflect the characteristics of population and community structure. Thus, it is imperative and imminent that appraising the impact of biotic and abiotic factors on productivity at different spatial scales.

Although forest growth is a complex nonlinear continuous process, big data and machine learning algorithm provide new possibilities for productivity modeling. The Random Forest (RF) algorithm has been used to deal with the relationships between explanatory and response variables without assuming specific mathematical equations and statistical assumptions (De'ath, 2007). It can effectively avoid over-fitting, evaluate the relative importance of explanatory variables, and is insensitive to collinearity among explanatory variables (Prasad et al., 2006) when facing too many data dimensions. Therefore, we used RF in our study, which can better help us decide on the appropriate sampling sizes.

China's forest ecosystems represent an important carbon sink, especially those in northeastern China. This study is based on a 30-ha (500×600m) monitoring field plot established in an old-growth broadleaved Korean pine forest in Jiaohe Forest Experimental Administration of Jilin Province. This large plot allows the study of the spatial scale effects on productivity modeling. We used RF algorithm to simulate biomass productivity. Topography factors, diversity factors, stand structure factors, and a density factor were selected as explanatory variables, while biomass productivity was selected as response variable. The following questions were addressed in this study: (1) What is the minimum plot area for predicting the biomass productivity with relatively high accuracy? By analyzing the prediction accuracy of the RF models built at different scales, we may suggest a minimum sampling area for productivity. (2) What are the similarities and differences in variables importance across different grain scales? This study is thus expected to provide new insights regarding the estimation of biomass productivity in a natural forest.

2 | MATERIALS AND METHODS

2.1 | Study area

The study site is located in the Jiaohe Management Bureau of the Forest Experimental Administration in Jilin Province, China, in a temperate continental monsoon climate with a mean annual temperature of 3.8°C and mean annual precipitation of 695.9mm. The average temperature of the hottest month is 21.7°C in July, and the average temperature of the coldest month is -18.6°C in January (Xu et al., 2019). The soil type is a dark brown forest soil, with an average depth of 45cm. A permanent observational field plot,

covering an area of 30-ha (500m×600m), was established in 2010 at 43°57.928′–43°58.214′N, 127°45.287′–127°45.790′E. The elevation ranges between 576 m and 784 m. The forest is a typical unmanaged coniferous and broadleaved-mixed forest, which has been left undisturbed for more than 50 years (Figure 1). The dominant species are *Pinus koraiensis*, *Tilia amurensis*, *Ulmus laciniata*, *Acer mono* and *Betula costata*. We completely surveyed the plot in two consecutive inventories: the first inventory was conducted in 2010 and the second in 2015. All of the woody plants with diameters at breast height (DBH) ≥ 1 cm were tagged and mapped, and their species were identified. The tree DBHs, heights, crown widths, and height-to-live crowns were measured in both censuses.

2.2 | Sample design and dataset

A random spatial sampling design was employed to investigate the relationship between forest biomass productivity and topographic, species diversity, stand structure, and a stand density variable at different spatial scales. For our design, different spatial scales

represented a series of square-shaped quadrats of different size, which increased from 10×10 m to 200×200 m in stepwise 5 m increase of the side length. Thus, a total of 39 spatial scales ranging from 10 to 200 m were studied: 10×10 m, 15×15 m, ..., 195×195 m, and 200×200 m. For each of the 39 quadrat sizes, 100 quadrats of the same size were randomly placed in the permanent observational field plot. We set up a buffer zone in the plot that extended from the boundary of the plot to half the side length of the quadrat. The center of the quadrat was extracted from the unbuffered zone to ensure that all the quadrats are in the plot. For each quadrat, tree species, DBH, and heights were extracted.

Four quadrat-level biotic and abiotic attributes were derived: (a) topographic variables (Topography), (b) species diversity variables (Diversity), (c) stand structure variables (Structure), and (d) stand density variable (Density). In our study, the quadrat biomass productivity (P) as output was modeled using the RF algorithm with the mentioned variables above as inputs. Thus, the basic model form was expressed as follows:

$$P = f(\text{Topography}, \text{Diversity}, \text{Structure}, \text{Density}) \quad (1)$$

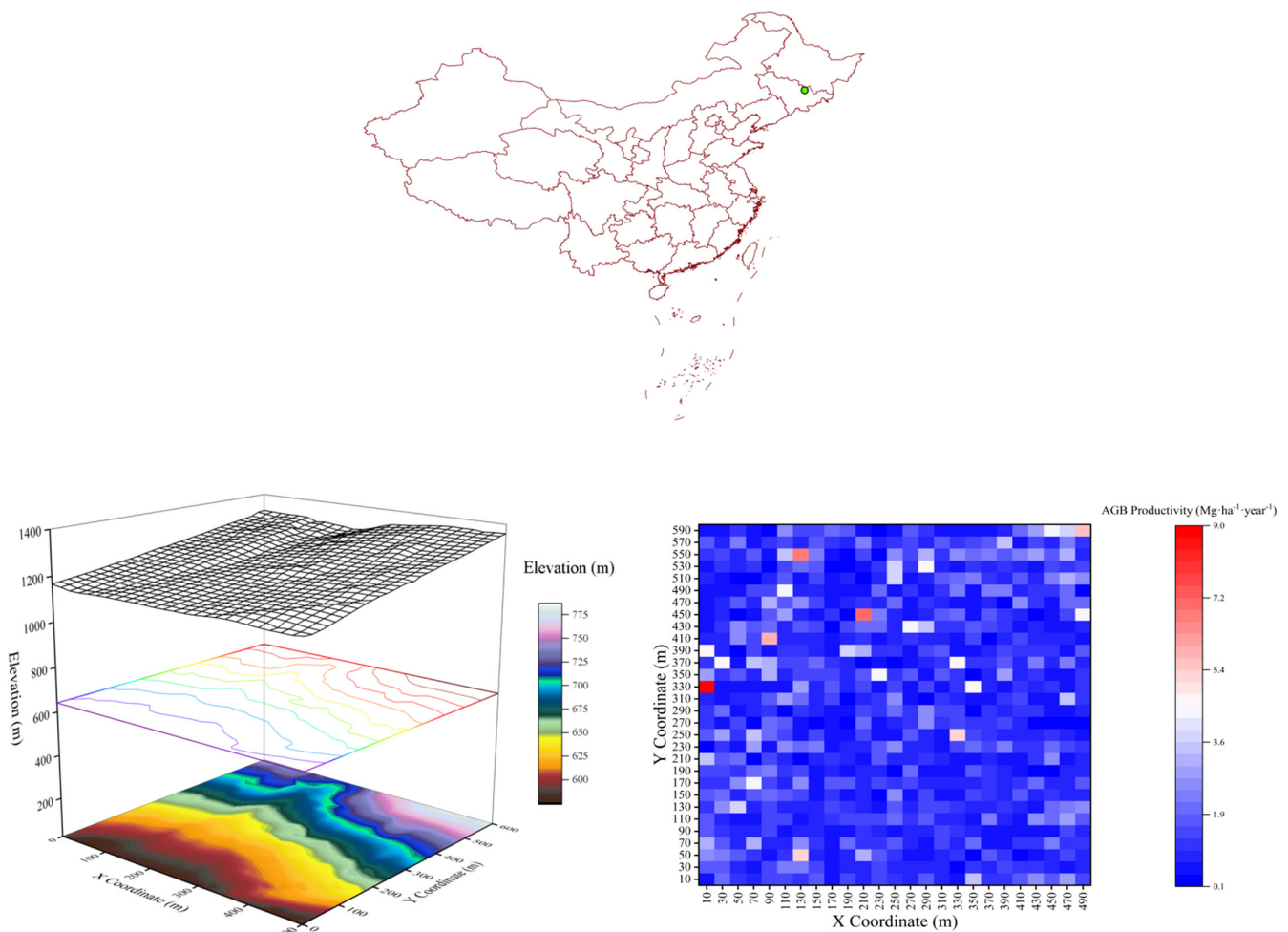


FIGURE 1 Top: Location of the study area in northeastern China. Bottom-left: Map depicting elevation patterns. The color from dark to light means the observed values are from low to high. Bottom-right: Map depicting forest biomass productivity patterns at the scale of 20 m. The color from blue to red means the observed values are from low to high

For each of the 39 scales, we utilized a completely randomized spatial sampling design, which was similar to Luo et al. (2019). Because of each quadrat standing for a virtual sample plot, the sampling design facilitated the study between biomass productivity and four biotic and abiotic attributes along the area scales. The 100 quadrats of the same size were randomly placed in the study area with replacement. This process was repeated 1000 times.

2.3 | Biomass productivity

In our study, the AGB of all individual trees were estimated using a set of region-specific allometric equations with DBH as an independent variable (Table A.2) (Yuan, Wang, et al., 2018). Each quadrat's total AGB was computed as the sum of the AGB of all the individuals within the quadrat. According to Prado-Junior et al. (2016) and Luo et al. (2019), forest productivity (P) of the quadrats could be measured as the periodic annual increment of total AGB per unit area and time. Only individuals with a DBH ≥ 5 cm were included since such trees were responsible for most of the AGB (Chiang et al., 2016; Hao et al., 2018).

2.4 | Topographic variables

The plot was subdivided into 750 cells of 20×20m, and the elevation of four corners of each cell was measured during plot establishment. Ordinary kriging was employed to construct a trend surface (Luo et al., 2019). Based on the trend surface, we extracted the elevation values of the center and four corners of each quadrat which we randomly placed in the permanent observational field plot. We further estimated slope and aspect values (Luo et al., 2019). When measuring the slope, any three of the four corners of the quadrat would form a plane, and the average angle between the plane and the horizontal plane was the slope of the quadrat. The average value of the angle between the plane composed of the corners and the vertical plane of the quadrat is the aspect value of the quadrat. We selected elevation (E), slope (SLO) and aspect (ASP), that is, CE , SLC as topographic variables, following Xiang et al. (2016):

$$CE_{ij} = \cos(ASP_{ij}) \times \ln(E_{ij}) \quad (2)$$

$$SLC_{ij} = \tan(SLO_{ij}) \times \cos(ASP_{ij}) \quad (3)$$

2.5 | Diversity variables

Seven indices were used to quantify the diversity of the species. Species richness (*Richness*) refers to the number of tree species that were present in each quadrat. We also studied Hill numbers (Hill, 1973) and species evenness as diversity measures. Hill number is defined as follows:

$$Dq_{ij} = \left(\sum_{ijk=1}^{Richness_{ij}} p_{ijk}^q \right)^{\frac{1}{1-q}} \quad (4)$$

where ijk means k th species of the ij th quadrat, p_{ijk} is the proportion of the k th species in the ij th quadrat in terms of stem number, q is the sensitivity of the measure to the relative frequencies.

As q tends to 1, Hill number is the exponential of the Shannon entropy (Jost, 2006). As $q = 2$, Hill number is equivalent of the Gini-Simpson diversity, the inverse of the Gini-Simpson index (Jost, 2006) as follows: In addition, we also calculated species evenness (Pielou, 1975) as diversity measures.

$$D1_{ij} = \exp \left(- \sum_{ijk=1}^{Richness_{ij}} p_{ijk} \ln p_{ijk} \right) \quad (5)$$

$$D2_{ij} = \left(\sum_{ijk=1}^{Richness_{ij}} p_{ijk}^2 \right)^{-1} \quad (6)$$

$$E_{1_{ij}} = \frac{D1_{ij}}{\ln S} \quad (7)$$

$$E_{2_{ij}} = \frac{D2_{ij}}{\ln S} \quad (8)$$

Additionally, we considered the abundance-based coverage estimator (ACE index) (Anne & Mark, 1993) and the $Chao1$ index (Anne, 1984) as diversity variables as well because ACE index can be used to estimate the number of species not yet observed in the community and $Chao1$ index can be used to estimate the total number of species and is sensitive to rare species. The greater the ACE value, the more is the real species in the community. The greater the $Chao1$ value is, the greater is the species richness.

$$ACE_{ij} = S_{abund_{ij}} + \frac{S_{rare_{ij}}}{C_{ace_{ij}}} + \frac{F_{1_{ij}}}{C_{ace_{ij}}} \gamma_{ace_{ij}}^2 \quad (9)$$

$$\gamma_{ace_{ij}}^2 = \max \left[\frac{S_{rare_{ij}}}{C_{ace_{ij}}} \frac{\sum_{ijk=1}^{10} i(i-1)F_{1_{ij}}}{N_{rare_{ij}}(N_{rare_{ij}}-1)} - 1, 0 \right] \quad (10)$$

$$N_{rare_{ij}} = \sum_{ijk=1}^{10} ijkF_{ijk} \quad (11)$$

$$C_{ace_{ij}} = 1 - \frac{F_{1_{ij}}}{N_{rare_{ij}}} \quad (12)$$

where $S_{abund_{ij}}$ is the number of abundant (abundance threshold $> n$) species in the ij th quadrat; $S_{rare_{ij}}$ is the number of rare (abundance threshold $\leq n$) species in the ij th quadrat; $F_{1_{ij}}$ refers to the number of species represented by one individual only in ij th quadrat; $\gamma_{ace_{ij}}^2$ represents the estimated coefficient of variation of rare species in the ij th quadrat. $n = 10$ is commonly used as the abundance threshold to classify abundant or rare species. At this point, $\gamma_{ace_{ij}}^2$, $N_{rare_{ij}}$, and $C_{ace_{ij}}$ can be calculated by the above formula, where F_{ijk} is the number of the ijk th species represented by individuals.

$$\text{Chao1}_{ij} = \text{Richness}_{ij} + \frac{F_{1ij}(F_{1ij} - 1)}{2(F_{2ij} + 1)} \quad (13)$$

where F_{2ij} refers to the number of species which contain at least two individuals in ij th quadrat.

2.6 | Stand structural variables

Our stand structural variables include the number of stems (N_{all}), the number of large trees (N_{60} , i.e., number of stems with DBH ≥ 60 cm), skewness of the log-normal distribution fitted to all individuals' DBH data (*skewness*), the shape parameter of the Weibull distribution fitted to the same data (*shape*), Shannon index of DBH ($DBHShannon$), Simpson index of DBH ($DBHSimpson$), gini index of DBH ($GiDBH$), coefficient of variation of DBH ($CVDBH$), Shannon index of tree height ($HShannon$), Simpson index of tree height ($HSimpson$), gini index of tree height (GiH) and coefficient of variation of tree height (CVH). The calculation method of a structural diversity index is as follows: Taking 2 cm as a diameter class width and 1 m as a tree height class width, the number of DBH classes or tree height classes, and the number of individuals in each class in each quadrat were calculated respectively. According to the calculation, formulas of Shannon index and Simpson index, $DBHShannon$, $DBHSimpson$, $HShannon$, and $HSimpson$ were calculated by substituting the number of species with the number of DBH classes or tree height classes respectively:

$$GiDBH_{ij} = \frac{\sum_{ijm=2}^{N_{ij}} (2k - N_{ij} - 1) BA_{ijm}}{\sum_{ijm=2}^{N_{ij}} (N_{ij} - 1) BA_{ijm}} \quad (14)$$

$$GiH_{ij} = \frac{\sum_{ijm=2}^{N_{ij}} (2k - N_{ij} - 1) H_{ijm}}{\sum_{ijm=2}^{N_{ij}} (N_{ij} - 1) H_{ijm}} \quad (15)$$

$$CVDBH_{ij} = 100\% \frac{\sqrt{\frac{1}{N_{ij}} (DBH_{ijm} - \mu_{ij})^2}}{\mu_{ij}} \quad (16)$$

$$CVH_{ij} = 100\% \frac{\sqrt{\frac{1}{N_{ij}} (H_{ijm} - \lambda_{ij})^2}}{\lambda_{ij}} \quad (17)$$

where DBH_{ijm} is the diameter at breast height of the m th individual in the ij th quadrat; μ_{ij} is the average DBH of all individuals in the ij th quadrat; H_{ijm} is the tree height of the m th individual in the ij th quadrat; λ_{ij} is the average tree height of all individuals in the ij th quadrat.

2.7 | Density variables

We calculated the Reineke Stand Density Index (SDI) for each quadrat.

$$SDI_{ij} = N_{ij} \times \left(\frac{20}{D_{gij}} \right)^\beta \quad (18)$$

where D_{gij} is the mean diameter in the ij th quadrat, and β is the allometric exponent that expresses the relation between tree size and number of trees.

Definition, units for the key forest attributes studied here are summarized in Table A.3.

2.8 | "Random Forest" algorithm

The RF is a popular machine learning algorithm based on multiple decision trees (Leo, 2001). It can deal with both regression problems and classification problems. In the process of splitting each child node, some variables from all candidates for the splitting variables are selected randomly, and then the optimal ones are determined (Wang & Wang, 2021). For the initial dataset, the training data are chosen randomly to build the model, the data not included are defined as "out-of-bag" (OOB) (Catani et al., 2013). The RF error is approximated by the OOB error during the training process (Naghibi et al., 2017).

A total of 22 explanatory variables including 2 topographic variables, 7 species diversity variables, 12 stand structural variables, and 1 stand density variable were selected in this study to participate in the modeling of biomass productivity, resulting in a very large data dimension. Therefore, to avoid the "dimensionality curse" and to reduce the time required for RF algorithm modeling, it is essential to screen the predictors before modeling (Chandrashekar & Sahin, 2014; Zarshenas & Suzuki, 2016). A novel feature selection method, the Boruta algorithm (Kursa & Rudnicki, 2010), was adopted in this study. The goal is to screen all sets of features associated with the response variable, rather than selecting for a specific model the set of features that minimizes the model cost function, and thus to screen for explanatory variables (Kursa & Rudnicki, 2010) more comprehensively and efficiently. The Boruta algorithm has been applied in geology (Pourghasemi et al., 2020), hydrology (Amiri et al., 2019), and ecology (Arjasakusuma et al., 2020; Dobrowolska & Bolibok, 2019; Poona et al., 2016) studies.

Hyper-parameter tuning is a tedious but crucial task for machine learning algorithms (Ou et al., 2019), which aims to improve model fitting and the reduction of prediction errors. In the RF model, the hyper-parameter "mtry" specifies the number of variables in the node for a binary tree, with a default value of one-third of the number of dataset variables (Leo, 2001). The "tuneRF" function from the "randomForest" package in the R software was employed to determine the optimal "mtry," at which point the corresponding OOB error is the smallest. The hyper-parameter "ntree," with a default value of 500, controls the number of decision trees (Leo, 2001). However, "ntree" values ranged from positive integers, and it is not feasible to train models with corresponding all "ntree" values and evaluate them. The common practice to solve this problem is to select a certain value range and step size for the hyper-parameters and evaluate the model trained by the candidate hyper-parameters to

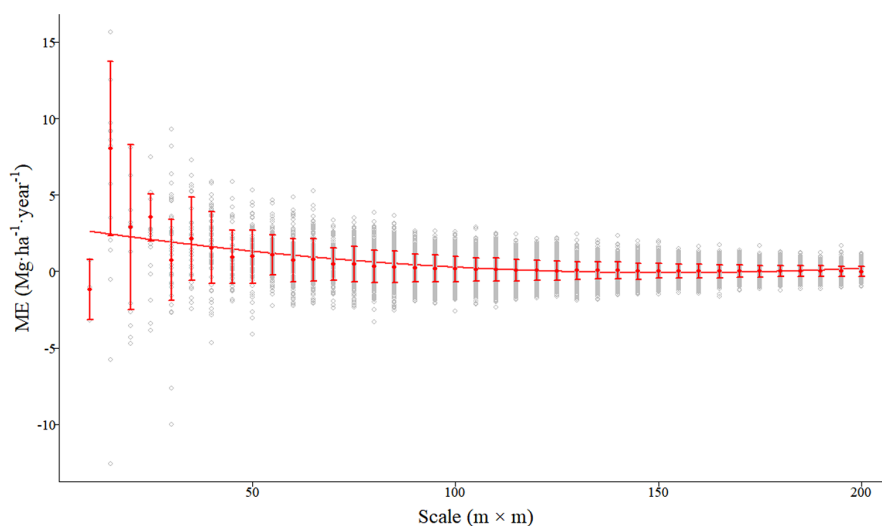
screen the optimal hyper-parameters. This compromise method of determining hyper-parameters balances computational costs and model performance, making the process of machine learning feasible (Ou et al., 2019).

Therefore, in this study, the range of “ntree” values was set at [0, 7000]. The function “randomForest” was used to calculate the model error corresponding to each “ntree” value under the optimal “mtry” values. The function “plot” was used to plot the model error versus the number of decision trees, showing that the error decreased gradually as the decision tree increased, and finally leveled off, at which point the number of decision trees, that is, the optimal value of “ntree,” was taken. Stationarity in this study was defined as follows: Within the value range of “ntree,” the step length was set to 500. The difference in a corresponding error of the *i*th candidate value and the (*i* + 1)th candidate value was calculated. If the difference was less than 0.1, the “stationarity” was reached. The *i*th candidate value was determined as the value of the optimal “ntree.” If the difference was greater than 0.1 across the range of “ntree” values, the threshold was upregulated by 0.05, until the optimal “ntree” value was taken. We conducted model selection and estimation of hyper-parameters using the package “randomForest” of R version 4.0.4 (R Core Team, 2021).

We first randomly selected 70% of the data for training and 30% for testing. Common criteria for assessing the goodness-of-fit of model predictions have historically been the mean error (ME), the mean absolute error (MAE), the mean relative error (MRE), the root-mean-squared error (RMSE), the relative RMSE (rRMSE), the coefficient of determination (R^2). Lower absolute ME and absolute MAE values indicate higher goodness-of-fit. Lower MRE, RMSE, and rRMSE values indicate higher goodness-of-fit. Higher R^2 values indicate higher goodness-of-fit.

$$ME = \frac{\sum_{p=1}^n (obs_p - est_p)}{n} \quad (19)$$

$$MAE = \frac{\sum_{p=1}^n |obs_p - est_p|}{n} \quad (20)$$



$$MRE = \frac{\sum_{p=1}^n (obs_p - est_p)}{\sum_{p=1}^n obs_p} \quad (21)$$

$$RMSE = \sqrt{\frac{\sum_{p=1}^n (obs_p - est_p)^2}{n}} \quad (22)$$

$$rRMSE = \frac{\sqrt{\frac{\sum_{p=1}^n (obs_p - est_p)^2}{n}}}{\frac{\sum_{p=1}^n obs_p}{n}} \times 100 \quad (23)$$

$$R^2 = 1 - \frac{\sum_{p=1}^n (obs_p - est_p)}{\sum_{p=1}^n \left(obs_p - \frac{\sum_{p=1}^n obs_p}{n} \right)} \quad (24)$$

where obs_p and est_p are the p_{th} observation and estimate, respectively, and n is the number of observations.

3 | RESULTS

3.1 | Evaluating the RF simulation accuracy at different area scales

For the test data, ME and MRE showed the same pattern with smaller fluctuating ranges and achieved maximum of median absolute values at the scale of 15 m ($9.2157 \text{ Mg}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$ and 0.2932, respectively), and minimum of median absolute values at the scale of 185 m ($0.0019 \text{ Mg}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$ and 0.0001, respectively; Figure 2; Figure A.1) with increasing quadrat area. MAE had the highest median value ($28.4703 \text{ Mg}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$) at 10 m and had the lowest median value at 200 m ($1.0645 \text{ Mg}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$; Figure 3). RMSE continually decreased at all scales and resulted in the highest and lowest median values ($29.7232 \text{ Mg}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$ and $1.3658 \text{ Mg}\cdot\text{ha}^{-1}\cdot\text{year}^{-1}$) at 10 and 200 m, respectively (Figure A.2). R^2 had an increasing trend with increasing quadrat area scales, reaching the highest median values at the scale of 200 m (0.79; Figure 4). The rRMSE decreased

FIGURE 2 Boxplot of the mean error (ME) changes modeled by the RF algorithm at each quadrat area scale. The solid line represents the mean trend line values, whereas dots with horizontal bars represent mean the ME for each quadrat size value and its standard deviation (SD)

FIGURE 3 Boxplot of the mean absolute error (MAE) changes modeled by the RF algorithm at each quadrat area scale. The solid line represents the mean trend line values, whereas dots with horizontal bars represent mean the MAE for each quadrat size value and its standard deviation (SD)

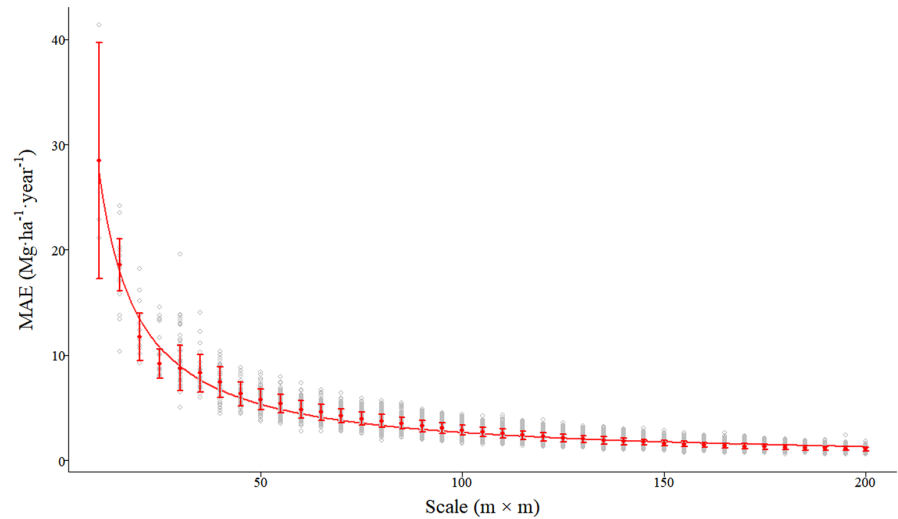
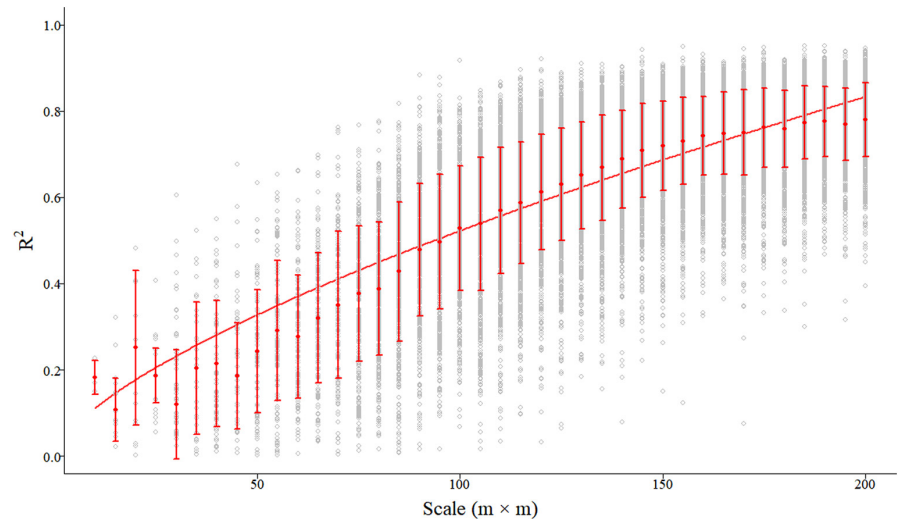


FIGURE 4 Boxplot of the coefficient of determination (R^2) changes modeled by the RF algorithm at each quadrat area scale. The solid line represents the mean trend line values, whereas dots with horizontal bars represent mean the R^2 for each quadrat size value and its standard deviation (SD)



with increasing scale, and the lowest median value (9.49%) were at the 200m plot size (Figure A.3). Thus, the optimal scale for estimating forest biomass productivity in this region may be chosen based estimated accuracy.

3.2 | Importance of explanatory variables at different scales

The relative importance value of the explanatory variables changed with scale. The importance of structural predictors tends to increase with increasing scale (Figure 5). Of structural predictors, stand structural diversity had the strongest effect sizes in general at the scale of 10–50m and 110–115m, while stem density of all trees was the best factor at the scale of 55–105m, and at the scale of 120–155m, tree size inequality was the most dominant variable, and at the scale of 160–200m, stem density of big trees showed superiority (Figure 5). The importance of topographic factors at medium scale was better than that on small and large scale, and the performance of *CE* was stronger constantly than that of *SLC* (Figure 5). Of species diversity

variables, 1D and 2D were the excellent predictors across all grain sizes (Figure 5), though their importance values were lower than 5% (except for 30m) (Figure 6). *SDI* dominated increasingly with decreasing scales (Figure 6).

4 | DISCUSSION

This study uses repetitive measurements of a 30-ha broadleaved Korean pine forest in Jiaohe Forestry Experimental Administration Bureau of Jilin Province, China to estimate forest productivity for a five-year period. We analyzed differences in the contribution of explanatory variables at different area scales, and the variation in the goodness of model fit. Our results highlight the fundamental role of scale in determining the relationship among four factors (topography, diversity, structure, and density) and biomass productivity per unit area and time. Lin et al. (2013) and Xu et al. (2015) captured biomass spatial variances in 24-ha and 25-ha forests, respectively. Therefore, we believe our results are credible in such 30-ha forest. We will discuss the accuracy of the RF models at different scales,

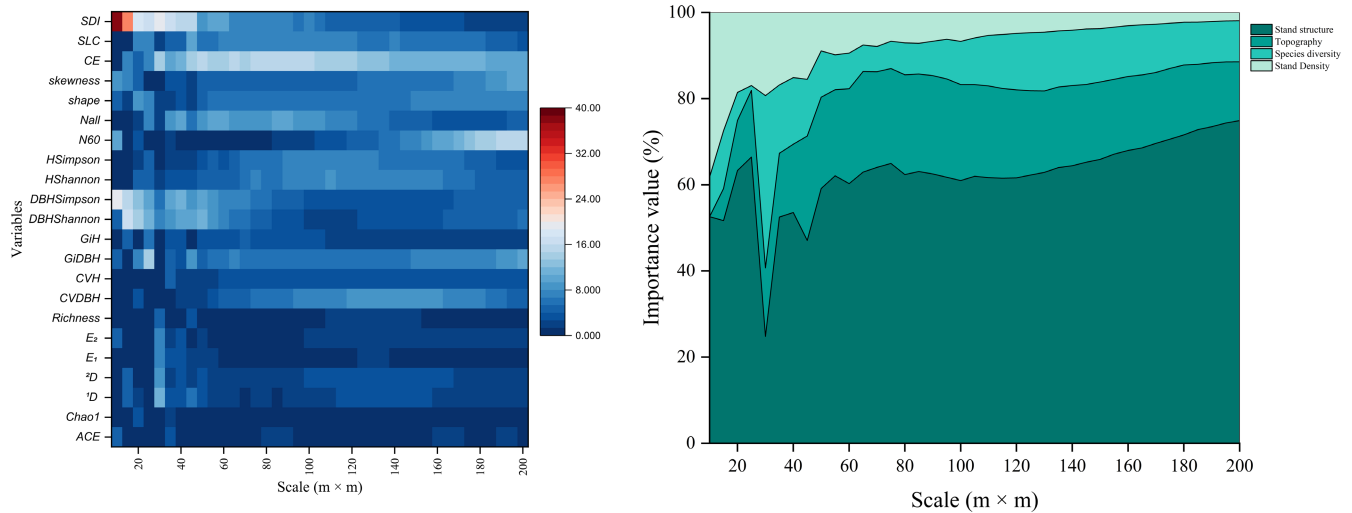
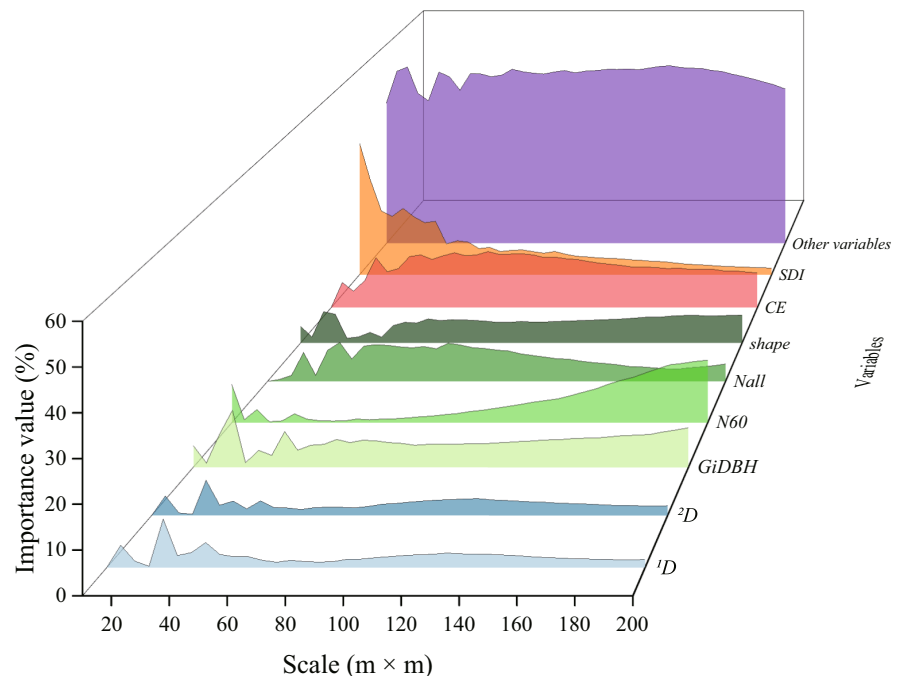


FIGURE 5 Left: Map depicting the relative importance values of variables patterns. The color from blue to red means the observed values are from low to high. Right: The relative importance value of the explanatory variables' categories at each scale

FIGURE 6 Map depicting the relative importance value patterns. The different color systems represent the different variables' categories



and the scale effect regarding the associations between the explanatory variables and forest productivity.

4.1 | Accuracy of the RF models at different scales

This study shows that the estimates of biomass productivity are affected by plot size. The accuracy of the RF model in estimating biomass productivity increases with increasing scale. Lin et al. (2013) and Keller et al. (2001) have shown that the variability of AGB decreased with increasing quadrat size. Similar results were reported by Sullivan et al. (2018) in tropical rainforests in South America and Asia, and by Chave et al. (2004) in a rainforest in Panama, plus by

Rodriguez-Hernandez et al. (2021) in a subtropical forest in China. Our study agrees with these findings with comparable results.

A small scale would provide less confident information and capture less characteristics in contrast to a big one (Rodriguez-Hernandez et al., 2021). This may be because too few samples may produce unreliable results (Leao et al., 2021), which also coincides with the statement that the bigger the number of sampling units, the greater the likelihood that new samples will lead to the same response (Brooks & Barcikowski, 2012). The spatial heterogeneity of sampling units can cause difficulties for community surveys, but large sampling areas are fundamental for reflecting community characteristics. Indeed, determining the optimum sampling area depends on the manpower and material resources consumed by the

field survey and the accuracy required by the investigators. Hetzer et al. (2020) showed that 1 ha is the effective area for mean biomass estimation with sufficient precision in South America (Hetzer et al., 2020). We consider a sampling area of at least 140×140 m to reach the requirements for estimates of biomass productivity in this region (Figure 4 and Figure A.3), as the turning point in R^2 and rRMSE indicates the minimal scale for effective sampling is the above scale, sampling at a median scale may be more cost effective ($R^2 > 0.7$, rRMSE < 20%). Using 3-PGmix model, Xie, Lei, and Shi (2020); Xie, Wang, and Lei (2020) explored the impacts of climate change on the biological rotation of *Larix olgensis* plantations for timber production and carbon storage in 492 sample plots of 0.0667 ha each in north-east China. Based on the results of this study, the results of these previous research may have some problems. Admittedly, a larger sampling area will lead to higher accuracy, but at the same time undoubtedly increase the manpower, material, financial resources consumed. Thus, sampling design is often faced with a dilemma (Peck & Zenner, 2021) that the proper sampling plot scale to assess forest productivity must take productivity variability and spatial distribution into consideration; however, it is difficult to predict such variables before an inventory is conducted. Hence, our results provide potential guidance for future sampling schemes.

4.2 | Scale dependence of the association between explanatory variables and AGB productivity

Our results indicated the most critical structure predictors were different at different scales. At the scale of 10–50 m and 110–155 m, DBH diversity and tree size inequality had significant links to productivity not only because multilayered forest structure can capture light and other resources easier (Ali et al., 2016; Yachi & Loreau, 2007), but also because tree size inequality plays a vital role in interfering the indirect impacts on biodiversity and abiotic factors on forest AGB productivity (Rodríguez-Hernández et al., 2021). At the scale of 55–105 m and 160–200 m, stem density were the best factors, which is in line with Rodríguez-Hernández et al. (2021). The importance of large trees increased with the increasing scale, which is different from the results of Rodríguez-Hernández et al. (2021). This can be explained that we studied the net biomass change, not the biomass. It is demonstrated that large trees have limited contribution to annual biomass production in an old-growth forest (Ligot et al., 2018). Similar to Rodríguez-Hernández et al. (2021), CE had the greatest importance values at a large scale due to the various spatial distribution of topographic conditions. CE is the combination of aspect and elevation, which reflects complex moisture-fertility effects (Xiang et al., 2016). The cosine of aspect gets its maximum value at north and it gets its minimum value at south. Furthermore, the natural logarithm of elevation increases monotonically. High importance values of CE demonstrated north aspects at high elevations or south aspects at low elevation had the biggest influence on biomass productivity at a large area scale. Species diversity were less important variable to project productivity across all scales in line with Sullivan

et al. (2017), Fotis et al. (2018), Hao et al. (2018), Ali et al. (2019a), and Rodríguez-Hernández et al. (2021), which might be related to competitive exclusion (Ali et al., 2016; Grace et al., 2016). However, our study only pay attention to taxonomic diversity, which is just one of the facets of diversity. Since we established virtual sample plots, the other two diversity facets (i.e., functional diversity and phylogenetic diversity) are difficult to incorporate in our study. In that case, we would need to sample all the trees in the 30-ha field plot, which is obviously difficult to achieve. But we still hope that in the future we can achieve such work, as this will make our study more comprehensive. There exists no single correct scale which forest ecosystems AGB productivity should be studied under the premise of suitable accuracy. That means readers can select the variables, which play important role in forest AGB productivity based on their intentions from Figure 5.

5 | CONCLUSIONS

Our study shows that the estimation accuracy of AGB productivity is affected by scale. As the area scale increases, a greater number of factors contribute to the accuracy of productivity estimates. This important result could only be achieved based on a large fully enumerated and remeasured field plot. The results are essential for setting minimum sampling areas required for estimating AGB productivity in broadleaved Korean pine forests in Northeastern China, and possibly elsewhere. Further research is required to explore the optimum balance between sampling area, and assessment cost.

AUTHOR CONTRIBUTIONS

Jingyuan He: Formal analysis (lead); methodology (lead); writing – original draft (lead). **Chunyu Fan:** Writing – original draft (supporting). **Yan Geng:** Formal analysis (supporting); funding acquisition (equal); writing – original draft (supporting). **Chunyu Zhang:** Funding acquisition (equal). **Xiuhai Zhao:** Conceptualization (lead); data curation (lead); methodology (supporting); supervision (lead). **Klaus von Gadov:** Writing – original draft (supporting).

ACKNOWLEDGMENTS

This work was funded by the Program of National Natural Science Foundation of China [grant numbers 31971650]; National Science Foundation of China [grant numbers 32171521]; and the Key Project of National Key Research and Development Plan [grant number 2017YFC0504104].

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study can be accessed on Figshare: <https://doi.org/10.6084/m9.figshare.20151827>.

ORCID

Jingyuan He  <https://orcid.org/0000-0002-3116-7554>

Chunyu Fan  <https://orcid.org/0000-0002-3360-2919>

Yan Geng  <https://orcid.org/0000-0001-9312-2602>

Chunyu Zhang  <https://orcid.org/0000-0003-3091-5060>

Xiuhai Zhao  <https://orcid.org/0000-0003-0879-4063>

Klaus von Gadow  <https://orcid.org/0000-0003-3641-0397>

REFERENCES

- Ali, A., Lin, S. L., He, J. K., Kong, F. M., Yu, J. H., & Jiang, H. S. (2019b). Elucidating space, climate, edaphic, and biodiversity effects on aboveground biomass in tropical forests. *Land Degradation & Development*, 30, 918–927.
- Ali, A., Lin, S.-L., He, J.-K., Kong, F.-M., Yu, J.-H., & Jiang, H.-S. (2019a). Climate and soils determine aboveground biomass indirectly via species diversity and stand structural complexity in tropical forests. *Forest Ecology and Management*, 432, 823–831.
- Ali, A., Yan, E.-R., Chen, H. Y. H., Chang, S. X., Zhao, Y.-T., Yang, X.-D., & Xu, M. S. (2016). Stand structural diversity rather than species diversity enhances aboveground carbon storage in secondary subtropical forests in eastern China. *Biogeosciences*, 13, 4627–4635.
- Amiri, M., Pourghasemi, H. R., Ghanbarian, G. A., & Afzali, S. F. (2019). Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma*, 340, 55–69.
- Anne, C. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11, 265–270.
- Anne, C. & Mark, C., K., Yang (1993). Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika*, 80, 193–201.
- Arjasakusuma, S., Swahyu Kusuma, S., & Phinn, S. (2020). Evaluating variable selection and machine learning algorithms for estimating Forest Heights by combining lidar and hyperspectral data. *ISPRS International Journal of Geo-Information*, 9, 507.
- Bradford, J. B., Weishampel, P., Smith, M.-L., Kolka, R., Birdsey, R. A., Ollinger, S. V., & Ryan, M. G. (2010). Carbon pools and fluxes in small temperate forest landscapes: Variability and implications for sampling design. *Forest Ecology and Management*, 259, 1245–1254.
- Brooks, G. P., & Barcikowski, R. S. (2012). The PEAR method for sample sizes in multiple linear regression. *Multiple Linear Regression Viewpoints*, 38(2), 1–16.
- Cadotte, M. W. (2015). Phylogenetic diversity and productivity: Gauging interpretations from experiments that do not manipulate phylogenetic diversity. *Functional Ecology*, 29, 1603–1606.
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67.
- Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13, 2815–2831.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40, 16–28.
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 409–420.
- Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B., & Hubbell, S. P. (2003). Spatial and temporal variation of biomass in a tropical forest: Results from a large census plot in Panama. *Journal of Ecology*, 91, 240–252.
- Chiang, J.-M., Spasojevic, M. J., Muller-Landau, H. C., Sun, I. F., Lin, Y., Su, S.-H., Chen, Z. S., Chen, C. T., Swenson, N. G., & McEwan, R. W. (2016). Functional composition drives ecosystem function through multiple mechanisms in a broadleaved subtropical forest. *Oecologia*, 182, 829–840.
- Chisholm, R. A., Muller-Landau, H. C., Abdul Rahman, K., Bebbler, D. P., Bin, Y., Bohlman, S. A., Bourg, N. A., Brinks, J., Bunyavechewin, S., Butt, N., Cao, H., Cao, M., Cárdenas, D., Chang, L. W., Chiang, J. M., Chuyong, G., Condit, R., Dattaraja, H. S., Davies, S., ... Zimmerman, J. K. (2013). Scale-dependent relationships between tree species richness and ecosystem function in forests. *Journal of Ecology*, 101, 1214–1224.
- Dahlhausen, J., Uhl, E., Heym, M., Biber, P., Ventura, M., Panzacchi, P., Tonon, G., Horváth, T., & Pretzsch, H. (2017). Stand density sensitive biomass functions for young oak trees at four different European sites. *Trees*, 31, 1811–1826.
- de Castilho, C. V., Magnusson, W. E., de Araújo, R. N. O., Luizão, R. C. C., Luizão, F. J., Lima, A. P., & Higuchi, N. (2006). Variation in aboveground tree live biomass in a central Amazonian Forest: Effects of soil and topography. *Forest Ecology and Management*, 234, 85–96.
- De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, 88, 243–251.
- Dixon, R. K., Solomon, A. M., Brown, S., Houghton, R. A., Trexler, M. C., & Wisniewski, J. (1994). Carbon pools and flux of global forest ecosystems. *Science (New York, N.Y.)*, 263, 185–190.
- Dobrowolska, D., & Bolibok, L. (2019). Is climate the key factor limiting the natural regeneration of silver fir beyond the northeastern border of its distribution range? *Forest Ecology and Management*, 439, 105–121.
- Fahey, T. J., Woodbury, P. B., Battles, J. J., Goodale, C. L., Hamburg, S. P., Ollinger, S. V., & Woodall, C. W. (2010). Forest carbon storage ecology, anagement, and polic. *Frontiers in Ecology and the Environment*, 8, 245–252.
- Fotis, A. T., Murphy, S. J., Ricart, R. D., Krishnadas, M., Whitacre, J., Wenzel, J. W., Queenborough, S. A., & Comita, L. S. (2018). Aboveground biomass is driven by mass-ratio effects and stand structural attributes in a temperate deciduous forest. *Journal of Ecology*, 106, 561–570.
- Gadow, K.V., González, J., Zhang, C., Pukkala, T., & Zhao, X. (2021). Sustaining Forest ecosystems.
- Grace, J. B., Anderson, T. M., Seabloom, E. W., Borer, E. T., Adler, P. B., Harpole, W. S., Hautier, Y., Hillebrand, H., Lind, E. M., Pärtel, M., Bakker, J. D., Buckley, Y. M., Crawley, M. J., Damschen, E. I., Davies, K. F., Fay, P. A., Firn, J., Gruner, D. S., Hector, A., ... Smith, M. D. (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, 529, 390–393.
- Hao, M., Zhang, C., Zhao, X., & von Gadow, K. (2018). Functional and phylogenetic diversity determine woody productivity in a temperate forest. *Ecology and Evolution*, 8, 2395–2406.
- Harte, J., Smith, A. B., & Storch, D. (2009). Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology Letters*, 12, 789–797.
- Hetzer, J., Huth, A., Wiegand, T., Dobner, H. J., & Fischer, R. (2020). An analysis of forest biomass sampling strategies across scales. *Biogeosciences*, 17, 1673–1683.
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54, 427–432.
- Holdaway, R. J., Easdale, T. A., Carswell, F. E., Richardson, S. J., Peltzer, D. A., Mason, N. W. H., Brandon, A. M., & Coomes, D. A. (2016). Nationally representative plot network reveals contrasting drivers of net biomass change in secondary and old-growth forests. *Ecosystems*, 20, 944–959.
- Hwan, S. J., & Chun, J.-H. (2011). Forest monitoring in times of climate change from the Asian view. *Forest Science and Technology*, 7, 47–52.
- Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W. S., Reich, P. B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J.,

- Weigelt, A., Wilsey, B. J., Zavaleta, E. S., & Loreau, M. (2011). High plant diversity is needed to maintain ecosystem services. *Nature*, 477, 199–202.
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113, 363–375.
- Kallimanis, A. S., Halley, J. M., Vokou, D., & Sgardelis, S. P. (2008). The scale of analysis determines the spatial pattern of woody species diversity in the Mediterranean environment. *Plant Ecology*, 196, 143–151.
- Keller, M., Palace, M., & Hurr, G. (2001). Biomass estimation in the Tapajos National Forest, Brazil: examination of sampling and allometric uncertainties. *Forest Ecology and Management*, 154, 371–382.
- Kitayama, K., Fujiki, S., Aoyagi, R., Imai, N., Sugau, J., Titin, J., Nilus, R., Lagan, P., Sawada, Y., Ong, R., & Kugan, F. (2018). Biodiversity observation for land and ecosystem health (BOLEH): A robust method to evaluate the management impacts on the bundle of carbon and biodiversity ecosystem Services in Tropical Production Forests. *Sustainability*, 10, 4224.
- Kral, K., Janik, D., Vrska, T., Adam, D., Hort, L., Unar, P., & Samonil, P. (2010). Local variability of stand structural features in beech dominated natural forests of Central Europe: Implications for sampling. *Forest Ecology and Management*, 260, 2196–2203.
- Kuribayashi, M., Noh, N.-J., Saitoh, T. M., Ito, A., Wakazuki, Y., & Muraoka, H. (2017). Current and future carbon budget at Takayama site, Japan, evaluated by a regional climate model and a process-based terrestrial ecosystem model. *International Journal of Biometeorology*, 61, 989–1001.
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36, 1–13.
- Laumonier, Y., Edin, A., Kanninen, M., & Munandar, A. W. (2010). Landscape-scale variation in the structure and biomass of the hill dipterocarp forest of Sumatra: Implications for carbon stock assessments. *Forest Ecology and Management*, 259, 505–513.
- Leao, F. M., Mendes Nascimento, R. G., Emmert, F., Alves Santos, G. G., Moraes Caldeira, N. A., & Miranda, I. S. (2021). How many trees are necessary to fit an accurate volume model for the Amazon forest? A site-dependent analysis. *Forest Ecology and Management*, 480, 118652.
- Leo, B. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Li, Y., Bao, W., Bongers, F., Chen, B., Chen, G., Guo, K., Jiang, M., Lai, J., Lin, D., Liu, C., Liu, X., Liu, Y., Mi, X., Tian, X., Wang, X., Xu, W., Yan, J., Yang, B., Zheng, Y., & Ma, K. (2019). Drivers of tree carbon storage in subtropical forests. *Sci Total Environ*, 654, 684–693.
- Liang, J., Crowther, T. W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E. D., McGuire, A. D., Bozzato, F., Pretzsch, H., de-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C. B., Glick, H. B., Hengeveld, G. M., Nabuurs, G. J., Pfautsch, S., ... Reich, P. B. (2016). Positive biodiversity-productivity relationship predominant in global forests. *Science*, 354, aaf8957.
- Ligot, G., Gourlet-Fleury, S., Ouédraogo, D., Morin, X., Bauwens, S., Baya, F., Brostaux, Y., Doucet, J. L., & Fayolle, A. (2018). The limited contribution of large trees to annual biomass production in an old-growth tropical forest. *Ecological Applications*, 28(5), 1273–1281.
- Lin, D., Lai, J., Mi, X., Ren, H., & Ma, K. (2013). Spatial variation in community structure of a subtropical evergreen broad-leaved forest: Implications for sampling design. *Chinese Science Bulletin*, 58, 1181–1186.
- Liu, J., Yunhong, T., & Slik, J. W. F. (2014). Topography related habitat associations of tree species traits, composition and diversity in a Chinese tropical forest. *Forest Ecology and Management*, 330, 75–81.
- Luo, W., Kim, H. S., Zhao, X., Ryu, D., Jung, I., Cho, H., Harris, N., Ghosh, S., Zhang, C., & Liang, J. (2020). New forest biomass carbon stock estimates in Northeast Asia based on multisource data. *Global Change Biology*, 26, 7045–7066.
- Luo, W., Liang, J., Gatti, R. C., Zhao, X., & Zhang, C. (2019). Parameterization of biodiversity-productivity relationship and its scale dependency using georeferenced tree-level data. *Journal of Ecology*, 107, 1106–1119.
- Mascaro, J., Asner, G. P., Muller-Landau, H. C., van Breugel, M., Hall, J., & Dahlin, K. (2011). Controls over aboveground forest carbon density on Barro Colorado Island, Panama. *Biogeosciences*, 8, 1615–1629.
- Mejstřík, M., Šrámek, M., & Matula, R. (2022). The effects of stand density, standards and species composition on biomass production in traditional coppices. *Forest Ecology and Management*, 504, 119860.
- Mittelbach, G. G. (2010). Understanding species richness-productivity relationships: The importance of meta-analyses. *Ecology*, 91, 2540–2544.
- Morin, X., Fahse, L., Scherer-Lorenzen, M., & Bugmann, H. (2011). Tree species richness promotes productivity in temperate forests through strong complementarity between species. *Ecology Letters*, 14, 1211–1219.
- Naghbi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of support vector machine, random Forest, and genetic algorithm optimized random Forest models in groundwater potential mapping. *Water Resources Management*, 31, 2761–2775.
- Ou, Q., Lei, X., & Shen, C. (2019). Individual tree diameter growth models of larch-spruce-fir mixed forests based on machine learning algorithms. *Forests*, 10, 187.
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A large and persistent carbon sink in the World's forests. *Science*, 333, 988–993.
- Peck, J., & Zenner, E. (2021). Structure area curves in eastern hardwoods: Implications for minimum plot sizes to capture spatially explicit structure indices. *Annals of Forest Science*, 78, 1–11.
- Peng, M., & Guo, Q. (2016). Minimum area of the community spatial structure of broadleaf-Korean pine forest in Shengshan Mountain, northeastern China. *Journal of Beijing Forestry University*, 38, 21–27.
- Pielou, E. (1975). *Ecological diversity*. John Wiley & Sons.
- Poona, N. K., van Niekerk, A., Nadel, R. L., & Ismail, R. (2016). Random Forest (RF) wrappers for waveband selection and classification of hyperspectral data. *Applied Spectroscopy*, 70, 322–333.
- Poorter, L., van der Sande, M. T., Arets, E. J. M. M., Ascarrunz, N., Enquist, B. J., Finegan, B., Licona, J. C., Martínez-Ramos, M., Mazzei, L., Meave, J. A., Muñoz, R., Nyctch, C. J., de Oliveira, A. A., Pérez-García, E. A., Prado-Junior, J., Rodríguez-Velázquez, J., Ruschel, A. R., Salgado-Negret, B., Schiavini, I., ... Peña-Claros, M. (2017). Biodiversity and climate determine the functioning of neotropical forests. *Global Ecology and Biogeography*, 26, 1423–1434.
- Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., & Cerda, A. (2020). Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Scientific Reports*, 10, 3203.
- Prado-Junior, J. A., Schiavini, I., Vale, V. S., Arantes, C. S., van der Sande, M. T., Lohbeck, M., & Poorter, L. (2016). Conservative species drive biomass productivity in tropical dry forests. *Journal of Ecology*, 104, 817–827.
- Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9, 181–199.
- Pyle, E. H., Santoni, G. W., Nascimento, H. E. M., Hutyrá, L. R., Vieira, S., Curran, D. J., van Haren, J., Saleska, S. R., Chow, V. Y., Carmago, P. B., Laurance, W. F., & Wofsy, S. C. (2008). Dynamics of carbon, biomass, and structure in two Amazonian forests. *Journal of Geophysical Research-Biogeosciences*, 113, G00B08.
- Qiao, X., Zhang, N., Zhang, C., Zhang, Z., Zhao, X., & Gadow, K. (2021). Unravelling biodiversity-productivity relationships across a large temperate forest region. *Functional Ecology*, 35, 2808–2820.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

- Ren, Y., Lu, Y., Fu, B., & Zhang, K. (2017). Biodiversity and ecosystem functional enhancement by forest restoration: A meta-analysis in China. *Land Degradation & Development*, 28, 2062–2073.
- Rodriguez-Hernandez, D. I., Deane, D. C., Wang, W., Chen, Y., Li, B., Luo, W., & Chu, C. (2021). Direct effects of selection on aboveground biomass contrast with indirect structure-mediated effects of complementarity in a subtropical forest. *Oecologia*, 196, 249–261.
- Ruiz-Benito, P., Gómez-Aparicio, L., Paquette, A., Messier, C., Kattge, J., & Zavala, M. A. (2014). Diversity increases carbon storage and tree productivity in Spanish forests. *Global Ecology and Biogeography*, 23, 311–322.
- Sande, M. T., Peña-Claros, M., Ascarrunz, N., Arets, E. J. M. M., Licona, J. C., Toledo, M., & Poorter, L. (2017). Abiotic and biotic drivers of biomass change in a neotropical forest. *Journal of Ecology*, 105, 1223–1234.
- Schmid, B., Balvanera, P., Cardinale, B. J., Godbold, J., Pfisterer, A. B., Raffaelli, D., Solan, M., & Srivastava, D. (2009). *Consequences of species loss for ecosystem functioning: Meta-analyses of data from biodiversity experiments*. In *Biodiversity, ecosystem functioning, and human wellbeing* (pp. 14–29). Oxford University Press.
- Stegen, J. C., Swenson, N. G., Enquist, B. J., White, E. P., Phillips, O. L., Jørgensen, P. M., Weiser, M. D., Monteagudo Mendoza, A., & Núñez Vargas, P. (2011). Variation in above-ground forest biomass across broad climatic gradients. *Global Ecology and Biogeography*, 20, 744–754.
- Sullivan, M. J., Talbot, J., Lewis, S. L., Phillips, O. L., Qie, L., Begne, S. K., Chave, J., Cuni-Sanchez, A., Hubau, W., Lopez-Gonzalez, G., Miles, L., Monteagudo-Mendoza, A., Sonké, B., Sunderland, T., Steege, H. T., White, L. J. T., Affum-Baffoe, K., Aiba, S.-I., & Zemagho, L. (2017). Diversity and carbon storage across the tropical forest biome. *Scientific Reports*, 7, 39102.
- Sullivan, M. J. P., Lewis, S. L., Hubau, W., Qie, L., Baker, T. R., Banin, L. F., Chave, J., Cuni-Sanchez, A., Feldpausch, T. R., Lopez-Gonzalez, G., Arets, E., Ashton, P., Bastin, J. F., Berry, N. J., Bogaert, J., Boot, R., Brearley, F. Q., Brienen, R., Burslem, D. F. R. P., ... Phillips, O. L. (2018). Field methods for sampling tree height for tropical forest biomass estimation. *Methods in Ecology and Evolution*, 9, 1179–1189.
- Tikkanen, O.-P., Punttila, P., & Heikkilä, R. (2009). Species-area relationships of red-listed species in old boreal forests: A large-scale data analysis. *Diversity and Distributions*, 15, 852–862.
- Wagner, F., Rutishauser, E., Blanc, L., & Hérault, B. (2010). Effects of plot size and census interval on descriptors of Forest structure and dynamics. *Biotropica*, 42, 664–671.
- Wang, H., & Wang, G. (2021). Improving random forest algorithm by lasso method. *Journal of Statistical Computation and Simulation*, 91, 353–367.
- Węgiel, A., Bembenek, M., Łacka, A., & Mederski, P. S. (2018). Relationship between stand density and value of timber assortments: A case study for scots pine stands in North-Western Poland. *New Zealand Journal of Forestry Science*, 48, 1–9.
- Wertz, B., Bembenek, M., Karaszewski, Z., Ochał, W., Skorupski, M., Strzeliński, P., Węgiel, A., & Mederski, P. S. (2020). Impact of stand density and tree social status on aboveground biomass allocation of scots pine *Pinus sylvestris* L. *Forests*, 11, 765.
- Whittaker, R. J. (2010). Meta-analyses and mega-mistakes: Calling time on meta-analysis of the species richness-productivity relationship. *Ecology*, 91, 2522–2533.
- Xiang, W., Lei, X., & Zhang, X. (2016). Modelling tree recruitment in relation to climate and competition in semi-natural *Larix-Picea-Abies* forests in Northeast China. *Forest Ecology and Management*, 382, 100–109.
- Xie, Y., Lei, X., & Shi, J. (2020). Impacts of climate change on biological rotation of *Larix olgensis* plantations for timber production and carbon storage in Northeast China using the 3-PG(mix) model. *Ecological Modelling*, 435, 109267.
- Xie, Y., Wang, H., & Lei, X. (2020). Simulation of climate change and thinning effects on productivity of *Larix olgensis* plantations in Northeast China using 3-PG(mix) model. *Journal of Environmental Management*, 261, 110249.
- Xu, W., Luo, W., Zhang, C., Zhao, X., von Gadow, K., & Zhang, Z. (2019). Biodiversity-ecosystem functioning relationships of overstorey versus understorey trees in an old-growth temperate forest. *Annals of Forest Science*, 76, 1–13.
- Xu, Y., Franklin, S. B., Wang, Q., Shi, Z., Luo, Y., Lu, Z., Zhang, J., Qiao, X., & Jiang, M. (2015). Topographic and biotic factors determine forest biomass spatial distribution in a subtropical mountain moist forest. *Forest Ecology and Management*, 357, 95–103.
- Xue, L., Jacobs, D. F., Zeng, S., Yang, Z., Guo, S., & Liu, B. (2012). Relationship between above-ground biomass allocation and stand density index in *Populus x euramericana* stands. *Forestry*, 85, 611–619.
- Yachi, S., & Loreau, M. (2007). Does complementary resource use enhance ecosystem functioning? A model of light competition in plant communities. *Ecology Letters*, 10, 54–62.
- Yuan, Z., Ali, A., Wang, S., Gazol, A., Freckleton, R., Wang, X., Lin, F., Ye, J., Zhou, L., Hao, Z., & Loreau, M. (2018). Abiotic and biotic determinants of coarse woody productivity in temperate mixed forests. *Sci Total Environ*, 630, 422–431.
- Yuan, Z., Wang, S., Ali, A., Gazol, A., Ruiz-Benito, P., Wang, X., Lin, F., Ye, J., Hao, Z., & Loreau, M. (2018). Aboveground carbon storage is driven by functional trait composition and stand structural attributes rather than biodiversity in temperate mixed forests recovering from disturbances. *Annals of Forest Science*, 75, 1–13.
- Zarshenas, A., & Suzuki, K. (2016). Binary coordinate ascent: An efficient optimization technique for feature subset selection for machine learning. *Knowledge-Based Systems*, 110, 191–201.
- Zhang, C., Zhao, Y., Zhao, X., & von Gadow, K. (2012). Species-habitat associations in a northern temperate forest in China. *Silva Fennica*, 46, 501–519.
- Zhang, Y., Chen, H. Y. H., & Coomes, D. (2015). Individual size inequality links forest diversity and above-ground biomass. *Journal of Ecology*, 103, 1245–1252.
- Zhang, Y., Chen, H. Y. H., Taylor, A. R., & Ostertag, R. (2016). Positive species diversity and above-ground biomass relationships are ubiquitous across forest strata despite interference from overstorey trees. *Functional Ecology*, 31, 419–426.
- Zhao, J., Yan, X., & Jia, G. (2012). Simulating net carbon budget of forest ecosystems and its response to climate change in northeastern China using improved FORCCHN. *Chinese Geographical Science*, 22, 29–41.
- Zhou, X., Fu, Y., Zhou, L., Li, B., & Luo, Y. (2013). An imperative need for global change research in tropical forests. *Tree Physiology*, 33, 903–912.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: He, J., Fan, C., Geng, Y., Zhang, C., Zhao, X., & Gadow, K. v. (2022). Assessing scale-dependent effects on Forest biomass productivity based on machine learning. *Ecology and Evolution*, 12, e9110. <https://doi.org/10.1002/ece3.9110>