ECOGRAPHY

Research article

How many trees are there in the North American boreal forest?

Kun Xu[1,2], **Jingye Li**[1], **Jian Zhang**[3], **Dingliang Xing**[3] **and Fangliang He**[1,3]

[1]Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada
[2]College of Life Sciences, Hubei Normal University, Huangshi, Hubei, China
[3]ECNU-Alberta Joint Lab for Biodiversity Study, Tiantong Forest Ecosystem National Observation and Research Station, School of Ecology and Environmental Sciences, East China Normal University, Shanghai, China

Correspondence: Fangliang He (fhe@ualberta.ca)

Ecography
2025: e07677
doi: 10.1002/ecog.07677
Subject Editor: Jason Pither
Editor-in-Chief:
Jens-Christian Svenning
Accepted 6 March 2025

Boreal forests, the largest terrestrial biome on Earth, are highly varied in local tree density. Despite previous attempts to estimate tree density in boreal forests, the accuracy of the estimation is unknown, leaving the question how many trees there are in boreal forests largely unanswered. Here, we compiled tree density data from 4367 plots in North American boreal forest and developed tree height-based generalized linear and machine learning models to address this question. We further produced the current boreal tree density map of North America, and projected tree density distribution in 2050 under the shared socioeconomic pathways (SSP) 126, 245 and 585 climate change scenarios. Our best-performed and cross-validated random forest model estimated a total of 277.2 (± 137.7 SD) billion trees in the North American boreal forest, 31.3% higher than the previously estimated 211.2 billion. Our projected tree density distributions in 2050 showed at least 11% increase in tree density in the region. This study improves our knowledge about boreal tree density and contributes to understanding the role of boreal forests in regulating forest ecosystem functions and informing adaptation and mitigation policy-making. The projected warming-induced increase in tree density suggests the potential of the North American boreal forest for carbon sequestration.

Keywords: estimating number of trees, forest inventory plots, North American boreal forest, random forest, SSP climate change scenarios, stand height

Introduction

Knowledge about tree density (number of trees per unit area; hereafter per hectare) at local, regional and global scales is critical to forest management (Kays and Harper 1974, Long 1985, Oliver and Larson 1996, Pretzsch 2009), biodiversity maintenance (Clark and Clark 1984, ter Steege et al. 2013, 2023), understanding ecosystem functioning (Tobner et al. 2014, Godlee et al. 2021), and informing climate change mitigation (Seppälä et al. 2009, Brunet-Navarro et al. 2016, Sterck et al. 2021, Woodall and Weiskittel 2021). Underlying these multifaceted roles of tree density is the process

of competition that drives stand dynamics and shakes up the development of stand structure, leading to quantitative relationships between tree density and stand volume or biomass (Mohler et al. 1978, Westoby 1984, Pretzsch 2009), and mean tree size (White and Harper 1970). Much of the foundation of growth and yield modelling and silviculture is derived from the theory and practice of managing stand density to minimize the effects of competition (Oliver and Larson 1996, Pretzsch 2009). Despite the wide importance of stand density, our knowledge about it and the factors responsible for its variation across landscapes remains limited. Tree density in a forest can be subject to as many factors as the number of trees themselves, including geography, topography, soils, nutrients, climate, stand structure, stand age and disturbances (Oliver and Larson 1996, Seidl et al. 2017, Madrigal-González et al. 2023). It is thus challenging to develop models to capture the variation in stand density across forests. Crowther et al. (2015) took up this challenge to model global tree density at the biome level (thereafter the biome models). This exceptional effort led to an estimate of 3.04 trillion trees on Earth and also made tree abundance estimation for each of the 14 biomes. These models were developed on the assumption that stand density is determined by stand topographic and vegetative characteristics and is regulated by climate and human development. While the biome models of Crowther et al. (2015) fill in a long-missing knowledge gap critical to management of forest resources and monitoring global carbon cycling, the quality of their estimation varies greatly across biomes, with particularly poor estimation for boreal and tundra biomes (Fig. 2 in Crowther et al. 2015).

Boreal forests are the largest terrestrial biome on Earth, storing approximately 11% of global terrestrial carbon (Gauthier et al. 2015), or 30% of global forest carbon (Pan et al. 2024). Boreal forests also play important roles in maintaining biodiversity, mitigating climate change impacts, and supporting livelihood and economy of north regions (Seppälä et al. 2009, Ma et al. 2012, Gauthier et al. 2015, Zhang et al. 2015). Given the importance of boreal forests, knowledge about tree density and its variation in distribution is necessary for sustainably managing boreal forests and informing policy-making (D'Amato et al. 2011, Brecka et al. 2018). However, accurate estimation of boreal tree density has proven challenging because boreal forests are one of the most varied forest ecosystems in terms of stand density, which can range from hundreds of trees ha^{-1} to 10 000s (Brandt 2009). Crowther et al. (2015) estimated 749.3 billion trees in boreal forests (accounting for 25% of global trees by their estimation), and 211.2 billion trees were estimated for North America's boreal forest, calculated from their tree density map. However, the accuracy of this estimation is unknown but is suspected to be low given that the boreal biome model is one of the two poorest performers of their 14 biome models. Various reasons likely contribute to the poor performance of their boreal biome model, perhaps most notably a lack of data and missing important explanatory variables. The 8688 ground plots used by Crowther et al. (2015) to build their boreal biome model only included 346 ground plots from the North American boreal forest that covers a vast 627-million-hectare area of land (the majority of their boreal plots are from Scandinavia). As another issue, although topographic, geographic and climatic conditions are believed to dictate the formation and distribution of global vegetation (Peel et al. 2007), they are insufficient in capturing the local variation of tree density (Aussenac 2000). An important process that regulates stand density is competition, e.g. for soil nutrients and light (Krajicek et al. 1961, Hart et al. 1989, Madrigal-González et al. 2023). Tree canopy height has been widely recognized as an aspect of key stand architecture that controls forest light condition (Krajicek et al. 1961, Hart et al. 1989, MacFarlane et al. 2000, Xu et al. 2019), and should be considered in modelling stand density. It is particularly relevant given that data on tree canopy height are now widely available from both ground and airborne data.

In this study, we aimed to estimate tree density for the boreal forest of North America. We compiled a large set of data consisting of 4367 ground plots from the region, and developed stand density models that include the effects of human development, land use, topography, vegetation, soil, soil-water balance, nitrogen deposition and climate as well as stand height. The results showed that our models performed well with high accuracy, and stand height was found to be the most important predictor for plot tree density. Our model estimated 277.2 billion boreal trees in North America, 31.3% higher than the previously estimated 211.2 billion.

Material and methods

Tree density data

We compiled ground plot data from three data sources, including permanent sample plots (PSPs), Canadian National Forest Inventory plots (NFI, the data source of Crowther et al.'s study for the North American boreal forest; Gillis et al. 2005) and Alberta Biodiversity Monitoring Institute plots (ABMI; Stadt et al. 2006). The PSPs consist of Canadian provincial and territorial forest inventory plots (Bonnor and Magnussen 1987) and the Cooperative Alaska Forest Inventory plots (CAFI; Malone et al. 2009) from the Alaskan boreal forest where no plot data were included in Crowther et al.'s biome model. These PSPs, NFI and ABMI plots together cover the North American boreal zone delineated by Brandt (2009), which were distributed in Alaska and nine Canadian provinces and territories (Fig. 1). Consistent with Crowther et al. (2015), we included live trees with diameter at breast height (DBH) $\geq$ 10.0 cm in this study. Plots were selected for this study if: 1) they were $\geq$ 100 m^2 in size with complete coordinates, 2) there were $\geq$ 5 trees and field-measured height records for $\geq$ 3 largest trees, 3) they were naturally regenerated without silviculture treatments (e.g. fertilizing or thinning) or records of disturbances (e.g. fire, pest, landslide, flood or other extreme weather events),

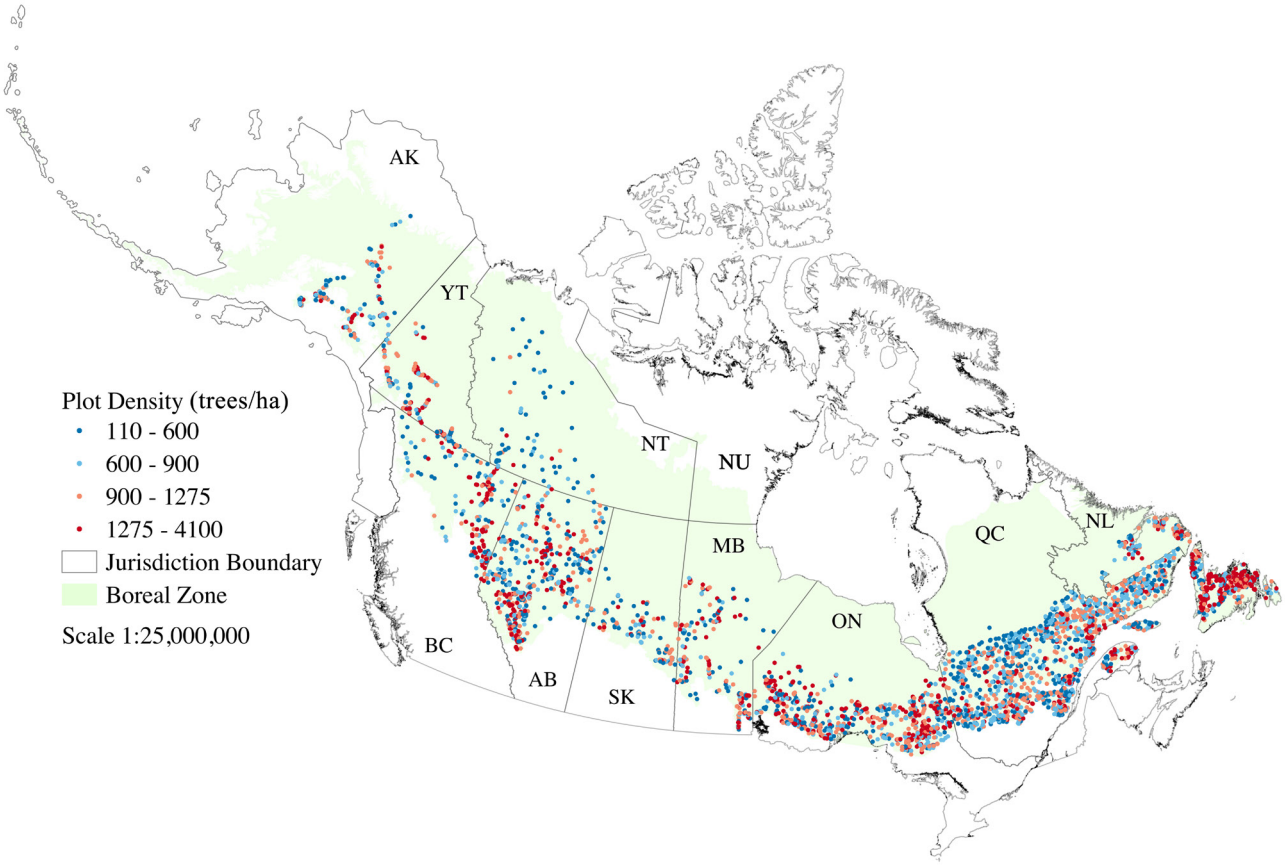

Figure 1. Distribution of 4367 plots in the North American boreal zone. The boreal zone is in light green color with latitude and longitude ranging from 46°36'36" to 69°33'00"N and from 52°38'24" to 178°58'48"W, respectively. The tree density varies from 110 to 4100 trees ha^{-1} (categorized into four blue-to-red colors from the lowest to the highest density). The jurisdiction abbreviations are: Alaska (AK), Yukon (YT), British Columbia (BC), Northwest Territories (NT), Alberta (AB), Saskatchewan (SK), Manitoba (MB), Ontario (ON), Quebec (QC) and Newfoundland and Labrador (NL). The boreal zone covers the southern part of Nunavut (NU) with no forest plots.

and 4) they were censused between 1999 and 2019. For plots with multiple censuses, the census closest to 2009 was used to minimize the temporal variation of plot because this census is comparable to the ground data used in Crowther et al. (2015). The final data included 4367 ground plots, of which 3829 were from PSPs, 346 from NFI and 192 from ABMI plots (Supporting information).

For each plot, the following data were compiled or calculated: 1) plot location (latitude, longitude and elevation) and distance to the nearest road (roads registered to the administrations including highway, secondary highway, local connector, and local road based on US Geological Survey National Transportation Dataset and roads under the categories of highway and road in Canada Road Network File), 2) field-measured stand height (mean height of the tallest three or more trees), and 3) the observed tree density (number of trees per ha; see the Supporting information for its frequency distribution). We also extracted plot canopy height data from the 2020 10-m global canopy height map (Lang et al. 2022) and from the 2005 1-km map (Simard et al. 2011), which were the two available sources of data for canopy height with a complete coverage for the North American boreal zone. In our study, we initially compiled 48 explanatory variables (see the Supporting information for the description and for the summary of these variables), including the same set of 20 variables of Crowther et al. (2015), i.e. one human development, six topographic, eight climatic, two vegetative and three second-order texture measures of vegetation index. The other 28 variables included one human development, one land use (i.e. land cover classes, the only categorical variable in this study), one topographic, one vegetative (i.e. the widely used normalized difference vegetation index, NDVI), one second-order texture measure of vegetation index, three soil, two soil-water balance, two anthropogenic nitrogen deposition, and 15 climatic variables plus stand height. Other than longitude, latitude, elevation, and field-measured stand height, explanatory variables were extracted from corresponding imaging data (Supporting information). The 23 climatic and two soil-water balance variables were 1970–2000 annual means, and the three vegetative indices and two nitrogen deposition variables were 2005–2014 monthly moving averages and 2005–2014 annual means, respectively. In cases where imaging data of the variables were not available at the plot location, we set 10 km (the lowest accuracy in spatial location of the 4367

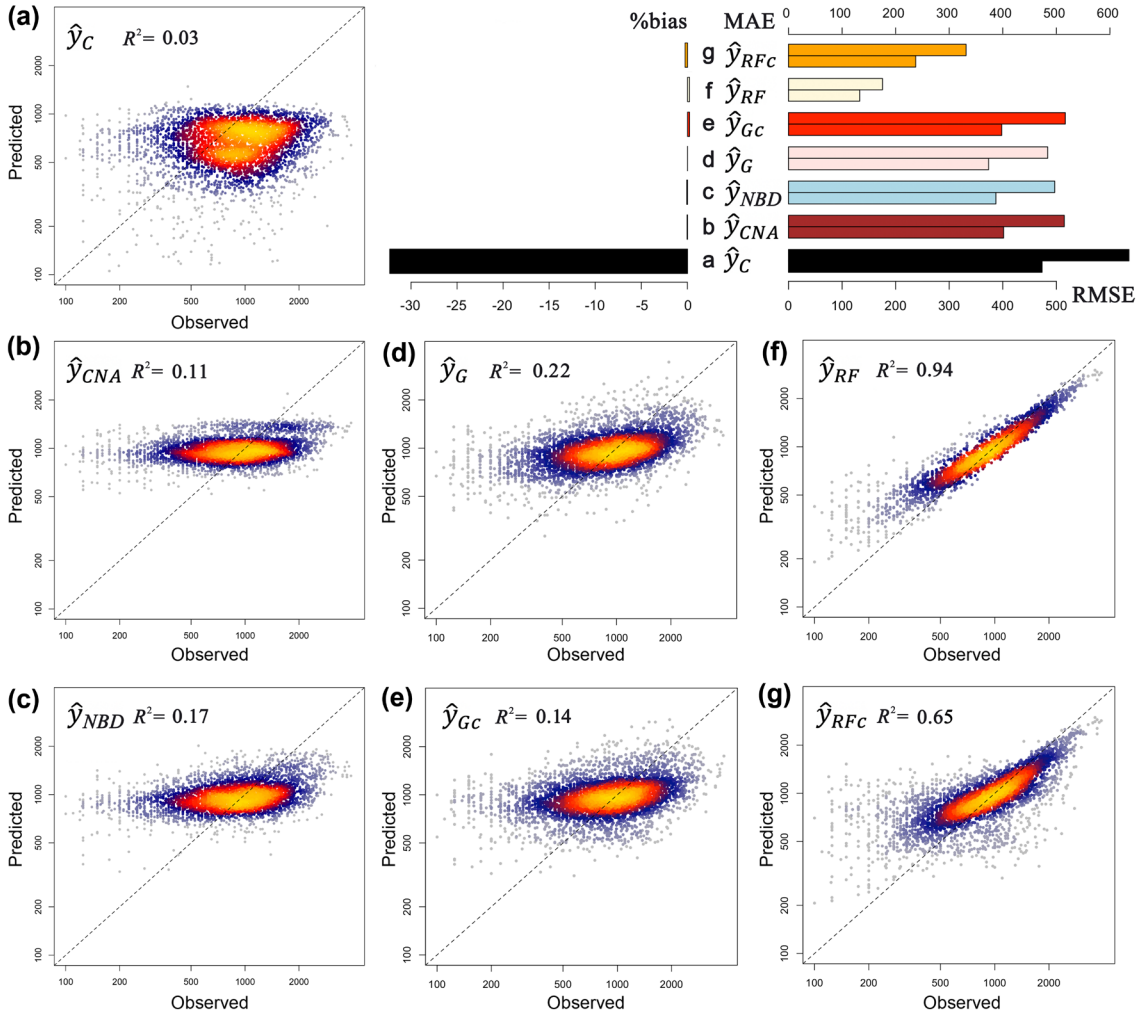

Figure 2. Heat scatter plots for the prediction of each of the seven models versus the observed tree density (y). The seven models are shown in each panel (see Table 1 for the notation of the models). The R^2 for each model is corrected for spatial autocorrelation between the observed density and each model prediction. The black dashed line is the 1:1 line, and the x- and y-axes are in log scale with the range 100–4100 100–4100trees ha^{-1}. The upper-right panels show the measures of the goodness-of-fit (%bias, MAE and RMSE) for each model in the bar plot grouped by different colors.

plots) as the maximum radius in searching the nearest location where such data were available.

Multicollinearity in the data of 4367 plots was dealt with by screening pairwise Pearson's correlations with $R^2 > 0.8$ (i.e. VIF > 5) among the 48 explanatory variables. None was collinear with stand height, while the Priestley–Taylor alpha coefficient (Priestley and Taylor 1972) and six bioclimatic variables (BIO; Hijmans et al. 2005), namely BIO6, 7, 10, 16, 18 and 19, were excluded due to collinearities with other variables. A final set of 41 explanatory variables was included for modelling (Supporting information). This set of explanatory variables included all explanatory variables used in Crowther et al. (2015).

Modelling tree density

Any model aiming to describe stand tree density ought to consider factors that regulate stand density. For this purpose, we followed Crowther et al. (2015) to include human development, topography, vegetation and climate as predictors, and we further compiled additional ecologically meaningful variables in these categories. These included land use, soil, soil-water balance, and nitrogen deposition as abiotic factors and stand height as a proxy for competition because height as a measure of stand architecture plays a key role in controlling stand light condition (MacFarlane et al. 2000). Our exploratory data analysis confirmed that stand height had a linear and quadratic relationship with stand density (Supporting information). Stand height was also one of the few stand architecture variables available in both ground measurement and airborne data (e.g. the Global Canopy Height map; Lang et al. 2022).

In this study, we used generalized linear models (GLMs) as in Crowther et al. (2015) and a decision-tree-based machine learning approach to estimate tree density using data of the 4367 plots. As shown in the Supporting information, plot

Table 1. Comparison of the performance of tree density models. The first five models are the generalized linear models, including the original boreal biome model in Crowther et al. (2015) ($\hat{y}_C$) fitted to their 346 plots data from the boreal forest of North America, their model refitting to our 4376 plots data ($\hat{y}_{\text{CNA}}$), negative binomial regression ($\hat{y}_{\text{NBD}}$), and Gamma regression ($\hat{y}_G$). The last two are the random forest regression ($\hat{y}_{\text{RF}}$) models. The measures of model goodness-of-fit include percent bias (%bias), Pearson's correlation corrected for spatial autocorrelation (R^2), mean absolute error (MAE), root mean squared error (RMSE), and the percentage of hot and cold spots matching the observed pattern (%hotspot). The units of MAE and RMSE are trees ha^{-1}. Stand height was the mean height of the three tallest trees in each of the data plots while canopy height was extracted from the Global Canopy Height map (Lang et al. 2022).

Model	Including tree height	% bias	R^2	MAE	RMSE	Hotspot (%)
Crowther biome ($\hat{y}_C$)	No	−32.3	0.03	473.3	635.6	38.0
Crowther refitted ($\hat{y}_{\text{CNA}}$)	No	−0.1	0.11	401.5	514.8	43.2
Negative binomial ($\hat{y}_{\text{NBD}}$)	Stand height	−0.1	0.17	387.2	497.0	48.6
Gamma ($\hat{y}_G$)	Stand height	2.3×10^{-8}	0.22	370.9	483.1	52.0
Gamma canopy ($\hat{y}_{Gc}$)	Canopy height	0.6	0.14	393.3	511.4	44.0
Random forest ($\hat{y}_{\text{RF}}$)	Stand height	0.5	0.94	134.2	179.4	74.1
Random forest canopy ($\hat{y}_{\text{RF}c}$)	Canopy height	−0.8	0.65	237.4	330.3	66.9

stand density (y) follows a Gamma distribution (χ^2 test with p-value = 0.36). We thus proposed Gamma regression to model tree density using an inverse link function $\frac{1}{\mu} = b_0 + b_1 x_1 + \cdots + b_p x_p$, where μ is the expected density and x's are the predictors. (Note the identity and log link functions were also used but the inverse link function turned out to have the lowest AIC.) All the numeric explanatory variables were standardized using $\frac{x-\mu}{\sigma}$, where μ and σ were mean and SD of the variable, respectively. In the Gamma regression model, both the first and quadratic terms of stand height were included. This model is denoted as $\hat{y}_G$ in Table 1.

Besides the above Gamma regression model, we also estimated three of Crowther et al.'s (2015) boreal biome models. The first was the biome model of Crowther et al. (2015) fitted to their 346 plots. This model assumed tree density followed a negative binomial distribution (NBD) with a log link function, i.e. $\hat{y}_C$ in Table 1. This is the model that produced the global tree density map reported in the study by Crowther et al. Model prediction of tree density in the 4367 plots was then extracted from their map (Fig. 3 in Crowther et al. 2015) using the latitude and longitude of the plots. In addition to this original model, we built two more NBD regression models also using the log link function to our data of 4367 plots. The first one was the refit of Crowther et al.'s model (i.e. using their set of 20 explanatory variables) to our 4367 plots data, i.e. model $\hat{y}_{\text{CNA}}$ in Table 1. This model was to test how our expanded plot data would improve the fitting of Crowther et al.'s model. The second model was to fit the NBD log link model to the full data (i.e. the 41 explanatory variables of the 4367 plot data), resulting in model $\hat{y}_{\text{NBD}}$ in Table 1. All the models were estimated using the generalized linear regression function *glm* in R ver. 4.2.2 (https://www.r-project.org). We used an AIC-based stepwise procedure for model selection. To assess model adequacy, we also checked the spatial correlation in residuals using the *variogram* function from R package 'gstat' (Pebesma 2004) but detected no spatial autocorrelation.

In addition to the above GLM models, we further adopted the machine learning algorithm random forest (Breiman 2001) to estimate tree density, i.e. $\hat{y}_{\text{RF}}$ in Table 1. Random forest regression models are more flexible than GLMs for characterizing complex relationships (Cutler et al. 2012, Greewell 2017) and are commonly used for spatial data analysis and mapping (Miller and Franklin 2002, Zhang et al. 2014, Lu and Hardin 2021). We trained the random forest model for the complete set of 41 explanatory variables using a portion of the 4367 plots and the rest as test data (see next section). Random forest can also assess importance of each variable in model prediction (Liaw and Wiener 2002). A widely used index for assessing variable importance is %IncMSE, calculated as the mean percentage change in prediction accuracy scaled by SD if this variable is excluded. The higher the %IncMSE, the more important this variable is for model prediction. The value can be extracted using the R package 'randomForest' (Liaw and Wiener 2002) to assess the rank of importance of each variable. We set the number of trees to grow to 5000 and the number of variables randomly sampled at each split to 13 (by default, the number of variables is divided by 3). Partial dependence between stand height and tree density prediction by marginalizing the other variables was produced using R package 'pdp' (Greenwell 2017) to assess the relationship between stand height and tree density. Prediction error of the random forest model, i.e. the conditional mean square prediction error, is generated using R package 'forestError' (Lu and Hardin 2021).

Model validation

We assessed the performance of the seven models in estimating tree density. Except for Crowther et al.'s model whose estimated tree density for each of the 4367 plots was extracted from their map (Fig. 3 in Crowther et al. 2015), tree density for other models was estimated by substituting respective

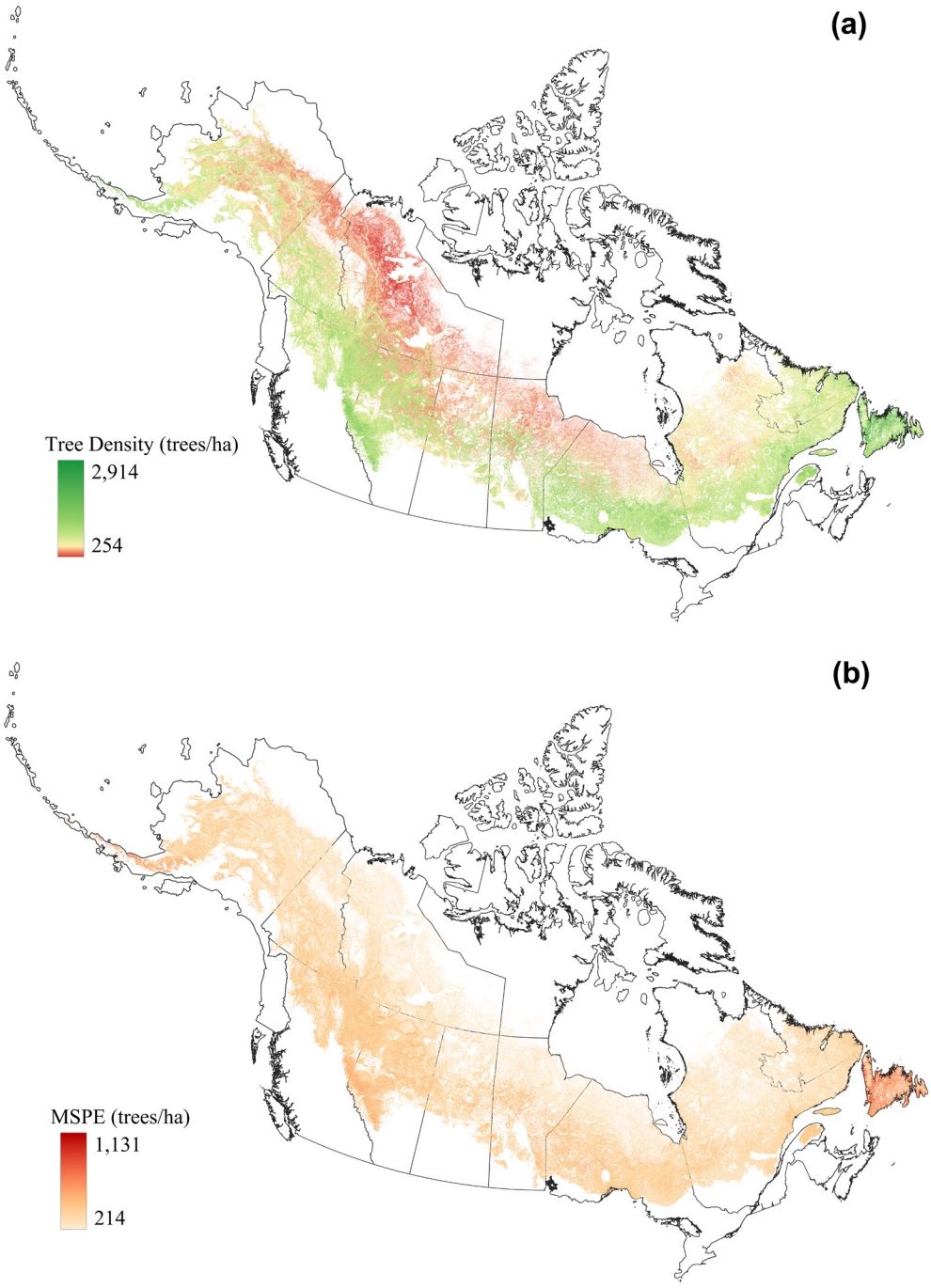

Figure 3. Maps of tree density distribution (a) and the mean square prediction error (MSPE) (b) for the boreal forest of North America. The distribution is produced from the prediction of the stand height-based random forest model ($\hat{y}_{RF}$; see Table 1). The scale is 1:30 000 000 and the map resolution is 1-km.

explanatory variables to each model. Two height data were respectively used: stand height (i.e. the mean height of the three tallest trees of each plot) and canopy height extracted from the global canopy height map. We used the 10-m map (Lang et al. 2022) instead of the 1-km mapped height (Simard et al. 2011) because the former was more accurate for predicting boreal canopy height (Yang and Kondoh 2020; also for the 4367 plots). Those models that were estimated using canopy height are denoted with subscript *c*, e.g. $\hat{y}_{Gc}$ and $\hat{y}_{RFc}$ in Table 1. Those without are models that either did not use height as a predictor at all (e.g. $\hat{y}_C$, $\hat{y}_{CNA}$) or used stand height (e.g. $\hat{y}_G$, $\hat{y}_{NBD}$, and $\hat{y}_{RF}$). It is important to note that no matter whether stand height or canopy height was used for parameterizing the models, canopy height must

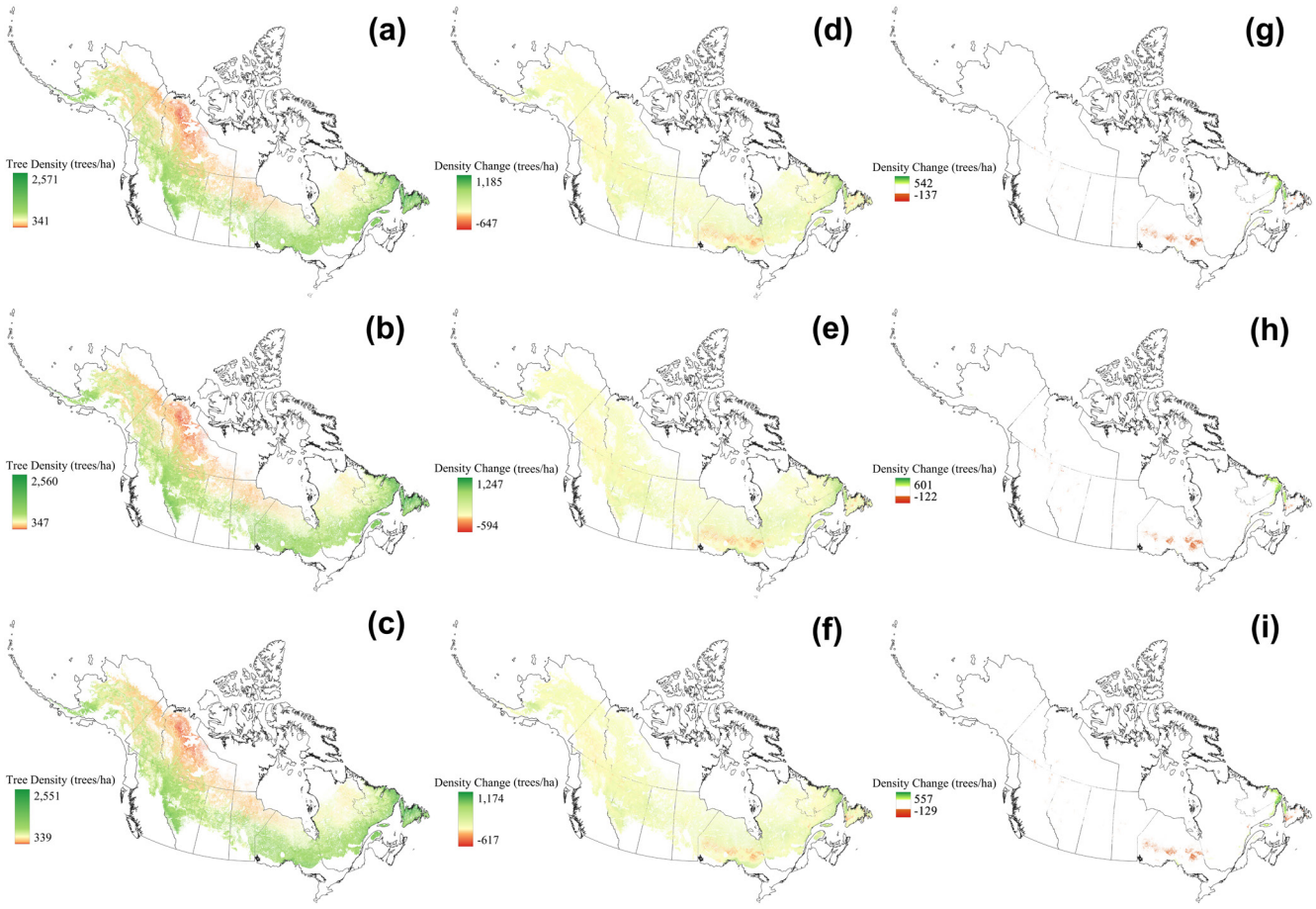

Figure 4. Projected tree density distribution of the North American boreal forest in 2050 under the shared socioeconomic pathways (SSP) 126 (a), 245 (b) and 585 (c) climate change scenarios. Prediction was made based on the random forest model by assuming only the bioclimatic variables (BIO1, 2, 3, 4, 5, 8, 9, 11, 12, 13, 14, 15 and 17) changed but other variables unchanged. The projected total number of trees in the region is 308.6, 313.2 and 313.5 billion under the SSP 126, 245 and 585 scenarios, respectively. The panels of the middle column (d–f) show the changes in boreal tree density (from −594 to 1247 trees ha^{-1}) that are the difference between projected density under the three climate change scenarios and the current estimate of the random forest regression model (i.e. Fig. 3a). The panels of third column (g–i) show those areas of highest decrease (1% quantile in density change in red) in tree density (southern Ontario) under the three climate change scenarios, and the areas of highest increase (99% quantile in density change in green) (northern Labrador). The 1% quantiles in density change for the SSP 126, 245 and 585 scenarios are −137, −122 and −129 trees ha^{-1}, and the 99% quantiles are 542, 601 and 557 trees ha^{-1}, respectively.

be used for model prediction because stand height data were not available for areas outside the 4367 plots.

These seven models ($\hat{y}_C$, $\hat{y}_{CNA}$, $\hat{y}_{NBD}$, $\hat{y}_G$, $\hat{y}_{Gc}$, $\hat{y}_{RFc}$ and $\hat{y}_{RF}$) in Table 1 were evaluated against the following metrices: 1) Pearson's correlation (R^2) between observed and estimated density corrected for spatial autocorrelation (Dutilleul 1993) using the *modified.ttest* function in R package 'SpatialPack' (Vallejos et al. 2020). We also presented the correlation in heat scatter plots using R package 'LSD' (Schwalb et al. 2020); 2) percent bias (%bias), calculated as $\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)}{\sum_{i=1}^{n} y_i}\times 100\%$ to assess estimation accuracy (positive %bias represents overestimation, negative %bias underestimation); 3) mean absolute error (MAE: $\frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{n}$); and (4) root mean square error (RMSE: $\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}}$). In these formulas, n is the number of plots, $\hat{y}_i$ estimated density (either $\hat{y}_C$, $\hat{y}_{CNA}$, $\hat{y}_{NGB}$, $\hat{y}_G$, or $\hat{y}_{RF}$) and y_i the observed tree density of the ith plot.

To compare the spatial aggregation of each of the seven estimates, we applied the optimized Getis-Ord GI* hotspot analysis (Getis and Ord 2010) in ArcGIS 10.3 (https://www.arcgis.com) for the observed density and each model

estimate. Hot spots were defined as those plots surrounded by high density plots with the confidence level 95%, and vice versa for cold spots (Nelson and Boots 2008). The percentage of matches in plot types (i.e. hot and cold spots; %hotspot) measured the degree of agreement between the observed tree density and the model-predicted density.

Ten-fold cross-validation was adopted to compare the prediction accuracy of Crowther's refitted ($\hat{y}_{\mathrm{CNA}}$), negative binomial ($\hat{y}_{\mathrm{NBD}}$), Gamma ($\hat{y}_G$), and random forest ($\hat{y}_{\mathrm{RF}}$) regression models. Given that the 4367 plots distribute widely in the study area, ten-fold cross-validation is preferred to examine the general applicability of the models for mapping tree density (Arlot and Celisse 2010). We randomly divided the plot data to 3930 and 437 plots as training and testing sets, respectively, and estimated each model using the training set. We then estimated plot tree density for the testing set based on each trained model. %bias, Pearson's R^2 corrected for spatial autocorrelation, MAE and RMSE of each prediction were calculated for the testing set. We iterated the cross-validation 1000 times, and conducted paired t test on each pair of the four predictions for each performance metric. We also adopted the ratio of 3:1 between training and testing data sets where 3275 plots were randomly taken to train the model and 1092 plots as testing data. Cross- validation allowed us to evaluate the performance of the models and select the best model for tree density estimation and mapping.

Mapping tree density and estimating tree abundance

The random forest regression model was selected as the best model. We used this model to estimate and map tree density of our study region. We also mapped the mean square prediction error of density estimates using the R package 'forestError' (Lu and Hardin 2021). The two maps were produced at the 1-km grid resolution (the same resolution as Crowther et al.'s map) for the North American boreal zone defined by Brandt (2009). Note that for non-forested areas where canopy height values did not exist or land cover types were not forest, a density value of 0 was assigned to them. We calculated the mean ($\bar{y}$), SD and coefficient of variation ($\mathrm{CV} = \frac{\mathrm{SD}}{\bar{y}} \times 100\%$) across the grid cells of the map. Tree density in Crowther et al. ($\hat{y}_C$) was also obtained from their map (Fig. 3 in Crowther et al. 2015). These would allow us to compare the distributions of boreal density estimated from our random forest model and Crowther et al.'s models. We estimated the total number of trees in the North American boreal forest across all forested grids for a given map as $\sum_{i=1}^{n} (\hat{y}_i \times \mathrm{s}_i)$, where s_i was the size (in ha) of the ith of all n grid cells. In addition, we calculated prediction error of the total number of trees as $\sum_{i=1}^{n} (\mathrm{MSPE}_i \times \mathrm{s}_i)$, where MSPE_i is the mean square prediction error of density estimate of the ith grid cell (Lu and Hardin 2021). We also calculated the number of trees per person for each jurisdiction. We did this for two populations. One is to divide the total number of boreal trees by population in the boreal zone only. This population is the sum of grid values from the Gridded Population of the World 2010 (GPWv4; https://sedac.ciesin.columbia.edu/gpw-v411-app). The other is to divide the number of trees by the entire population of the jurisdiction, using 2021 census data from the US Census Bureau 2020 Profile for Alaska and Statistics Canada Census of Population 2021 for Canadian jurisdictions. For example, in Ontario the estimated population in the boreal region is 0.18 million, while the population of the entire province in 2021 is 14.2 million.

Our last analysis was to use our best model to project boreal tree density in 2050 under three climate change scenarios assuming changing climate but else being equal, e.g. no change in stand height, land cover, topography, soil, nitrogen deposition and human development (though those factors are likely to change). The three climate change scenarios were SSPs 126, 245 and 585 representing the moderate, intermediate and severe climate change scenarios, respectively (Juckes et al. 2020). Under each SSP, the bioclimatic variables were averages of the 2050 projections by 23 general circulation models (GCMs; Supporting information; Flato et al. 2014) from the ensembled CMIP6 climate data (30-arc-second spatial resolution; https://wcrp-cmip.org/cmip6/). Maps of tree density distribution in 2050, their changes against the baseline (the current distribution estimated from our best model), and the 1 and 99% quantiles of the changes were produced under the three climate change scenarios. Based on these maps, we projected the total number of trees of the North American boreal forest in 2050.

Results

Tree density in North American boreal forest varies hugely from 110 to 4100110–4100 trees ha^{-1}, with the mean and SD of the observed density being 991.6 and 545.1, respectively (Fig. 1, Supporting information). Of the seven models that were used to model tree density (Table 1), the Gamma regression model ($\hat{y}_G$) is the best of the five generalized linear models (GLMs), and the stand height-based random forest regression model ($\hat{y}_{\mathrm{RF}}$) is the best of all models. Although refitting Crowther et al.'s model to the North American plot data ($\hat{y}_{\mathrm{CNA}}$ that excludes stand height and $\hat{y}_{\mathrm{NBD}}$ that includes height) improves model accuracy, their performance is still inferior to the Gamma regression model ($\hat{y}_G$) (Table 1). The random forest model outperforms all the GLMs with a higher R^2, lower MAE and RMSE, and a higher percentage match between the observed and predicted tree density hotspots (%hotspot) (Table 1, Supporting information). The superiority of $\hat{y}_{\mathrm{RF}}$ is also evident from the predicted versus observed heat scatter plots shown in Fig. 1 and the cross-validation tests (Supporting information).

The best selected Gamma regression model ($\hat{y}_G$) shows the linear and quadratic terms of stand height are the most important predictors of stand density, while human development and second-order texture measure of vegetation index

Table 2. The Gamma regression model with stand height and 16 other predictors. The best link function is the inverse link ($1/\mu = bx$), with its AIC 24.8 lower than the next best model. With this inverse link, a positive regression coefficient indicates a negative association between tree density and the predictor, and vice versa for a negative coefficient. Predictors are presented in descent order of their importance listed by the AIC-based stepwise selection. The estimated coefficients indicate effect size and are comparable because the *x*'s were all standardized except landcover which was a categorical variable. No spatial autocorrelation is detected from the semi-variogram of the residuals. The reference level for land cover is the type of temperate or sub-polar needleleaf forest.

Predictor	Estimated coefficient ($\times 10^{-5}$)	SE ($\times 10^{-5}$)	p-value
(Intercept)	90.30	1.07	< 0.001
Stand height	−10.34	1.11	< 0.001
Stand height^2	14.03	0.76	< 0.001
Nitrogen deposition as NO_y	14.37	1.52	< 0.001
BIO11 (mean temperature of coldest quarter)	−13.78	1.48	< 0.001
NDVI (normalized difference vegetation index)	−8.88	1.01	< 0.001
BIO5 (max temperature of warmest month)	−12.48	1.83	< 0.001
BIO9 (mean temperature of driest quarter)	7.50	1.29	< 0.001
Solar radiation	9.00	1.65	< 0.001
BIO13 (precipitation of wettest month)	9.88	1.84	< 0.001
Elevation	−6.93	1.32	< 0.001
Actual evapotranspiration	−1.13	2.29	< 0.001
Wind speed	−5.50	1.26	< 0.001
Aridity index	−10.49	2.73	< 0.001
Topsoil carbon content	−3.66	0.96	< 0.001
Topsoil pH	−3.44	1.05	0.001
Slope	2.85	9.16	0.002
Eastness	−2.28	0.72	0.002
Land cover: sub-polar taiga needleleaf forest	50.78	26.56	0.056
Land cover: temperate or sub-polar broadleaf deciduous forest	9.60	2.62	< 0.001
Land cover: mixed forest	6.81	1.79	< 0.001
Land cover: temperate or sub-polar shrubland	8.15	4.66	0.081

are not significant (Table 2). The other 16 significant predictors comprise land cover type, three topographic, NDVI, two soil, a soil-water balance, a nitrogen deposition and seven climatic variables. Similarly, the random forest regression model ($\hat{y}_{RF}$) also retains stand height as the most important predictor (with 177% IncMSE), followed by NDVI (88%) and the other variables (below 70%) (Supporting information).

The tree density distribution for North American boreal forest and the mean square prediction error estimated from the best selected random forest model are shown in Fig. 3. The distribution shows a northward trend of decrease in tree density. We also identified the areas where the tree density in the region is the highest or lowest, i.e. the hot and cold spots of the tree density, shown in the Supporting information.

Based on the random forest regression model, we estimated a total number of 277.2 (± 137.7) billion trees in the boreal forest of North America, which is 31.3% higher than 211.2 billion trees estimated from Crowther et al.'s biome model (Table 3). The ratio of trees per person is 85 635 (277.2 billion trees divided by an estimated 3.24 million population in the boreal region based on GPW 2020), 203 times higher than the global average which is 422 trees per person (Crowther et al. 2015). The underestimation of 66.0 billion trees is equivalent to missing 20 370 trees per person in the North American boreal zone. However, this number of trees per person reduces to 7737 if the entire population of the jurisdictions is counted (see Table 4 for the estimated number of trees for each jurisdiction).

We projected that the total number of trees in the North American boreal forest in 2050 would increase by 31.4 (11.3%), 36.0 (13.0%) and 36.3 (13.1%) billion from the currently estimated 277.2 billion trees under the moderate (SSP 126), intermediate (SSP 245) and severe (SSP 585) climate change scenarios, respectively. Despite the projected overall increase in the number of boreal trees, the projected maps (Fig. 4a–c) show an overwhelming trend of decrease

Table 3. Tree density estimates of the random forest model ($\hat{y}_{RF}$) and a previous boreal biome model ($\hat{y}_C$), showing the estimated total number of trees in North American boreal forest. The mean, SD and coefficient of variation (CV) of tree density were calculated on the basis of the 1-km resolution density maps for the two models.

Model	Total number of trees (in billion)	Mean grid tree density (trees ha^{-1})	SD (trees ha^{-1})	CV (%)
Random forest ($\hat{y}_{RF}$)	277.2	726.9	261.4	36.0
Crowther et al.'s biome ($\hat{y}_C$)	211.2	553.8	262.5	47.4

Table 4. Mean tree density (trees/ha) per grid cell and the total number of boreal trees in each jurisdiction estimated by the random forest regression ($\hat{y}_{RF}$). The mean grid tree density (SD) and the total number of boreal trees (SE) are also presented. The fourth column shows the percentage of the trees in each jurisdiction over the total number of 277.2 billion boreal trees in North America. The last two columns are the number of trees per person in the boreal zone and in the entire jurisdiction. The jurisdiction abbreviations are: Alaska (AK), Yukon (YT), British Columbia (BC), Northwest Territories (NT), Alberta (AB), Saskatchewan (SK), Manitoba (MB), Nunavut (NU), Ontario (ON), Quebec (QC), and Newfoundland and Labrador (NL). Population in the boreal zone is extracted from the Gridded Population of the World 2020 data, while population per jurisdiction is obtained from the US Census Bureau 2020 Profile and Statistics Canada Census of Population 2021.

Jurisdiction	Mean tree density (SD) (trees ha^{-1})	Total no. of trees (SE) (in billion)	% of boreal trees	No. of trees per person in boreal zone	No. of trees per person in jurisdiction
AK	672.3 (179.0)	37.1 (19.4)	13.38	88 374	50 614
YT	705.7 (220.1)	21.2 (10.5)	7.66	662 934	526 944
BC	896.3 (200.5)	22.2 (10.5)	8.02	324 257	4439
NT	484.4 (166.1)	22.7 (13.7)	8.17	549 004	548 426
AB	820.7 (261.7)	30.5 (14.9)	11.00	21 073	7155
SK	675.5 (230.4)	13.3 (7.0)	4.80	104 609	11 744
MB	680.2 (295.4)	17.4 (9.3)	6.28	103 687	12 964
NU	455.1 (51.4)	0.3 (0.2)	0.11	31 212 213	8139
ON	871.6 (286.9)	39.4 (18.1)	14.22	220 369	2770
QC	764.7 (180.3)	49.4 (21.9)	17.83	176 388	5811
NL	957.2 (321.1)	23.6 (10.7)	8.53	51 202	46 225

in tree density in southern Ontario in 2050, while northern Labrador shows a considerable increase in tree density in 2050 (Fig. 4d–i).

Discussion

Knowledge about tree density at local, regional and global scales is important for forest management, understanding functioning of forest ecosystems, and formulating resource-based climate mitigation policies. However, such knowledge is often not available at the regional and global scales. Previous data have shown tree density varies hugely across forests (Oliver and Larson 1996, Gillis et al. 2005, Brandt 2009) and is elusive to estimate (Crowther et al. 2015). In this study, we compiled an extensive set of ground plot data and developed models with the aim to reduce the uncertainty in estimating tree density of the boreal forest of North America. The cross-validation of the models developed in this study showed the random forest model outperformed all other models as assessed by nearly all model validation criteria (Table 1, Supporting information).

Based on the best selected random forest model, we estimated a total number of 277.2 billion trees in the North American boreal forest, 66.0 billion (31.3%) more than the number estimated from a previous effort (Crowther et al. 2015) (Table 3). If this level of underestimation also occurred in Scandinavian and Siberian boreal forests, after a correction we would expect that globally there are 0.97 trillion boreal trees (1.313 × 0.74, where 0.74 trillion was the number of global boreal trees given by Crowther et al.). This underestimation of 0.23 trillion boreal trees was equivalent to 7.6% of their estimation of 3.04 trillion global trees.

The substantial improvement of our models in estimating tree density is due to two reasons. The first one is ecological, by which we incorporated stand height into our models. Tree height is a key stand architecture that controls stand light condition (MacFarlane et al. 2000) and drives tree competition (Hart et al. 1989, MacFarlane et al. 2000) and thinning mortality (Kays and Harper 1974, Mohler et al. 1978, Westoby 1984, Reyes-Hernández and Comeau 2014). Our models identified stand height to be the most important factor affecting tree density (Table 2, Supporting information). This revelation makes it easier to interpret our models mechanistically. The second reason is methodological, by which we adopted the Gamma regression and random forest methods for modelling. The random forest regression model performed remarkably well with $R^2 = 0.94$ (Table 1, Fig. 1).

Our Gamma regression model shows that the effect of stand height on tree density comes from linear and quadratic terms (Table 1; also detected in the Supporting information). These terms indicate that stand density initially increases with stand height but would reach the peak at an intermediate height, after which stand thinning processes start to act. Empirical evidence shows that when stands are short, canopy space is ample and competition for light is low (MacFarlane et al. 2000, Xu et al. 2019). With an increase in stand height, competition for canopy space, light and soil nutrients builds up, intensifying self-thinning (Reyes-Hernández and Comeau 2014). In addition to stand height, our model also found plot eastness, slope, elevation, NDVI, land cover type, actual evapotranspiration, two soil (topsoil carbon content and pH), nitrogen deposition as NO_y, and seven climatic variables (solar radiation, wind speed, aridity index, and BIO5, 9, 11 and 13) to be important. Crowther et al. (2015) did not consider stand height, NDVI, nor soil, despite that soils are considered to determine tree establishment (Paré and Bergeron 1996, Madrigal-González et al. 2023). Besides these differences, our model predicted that high maximum temperature of the warmest month (BIO5), low temperature of the driest quarter (BIO9), and high temperature of the coldest quarter (BIO11) were associated with high tree density. This is

expected as temperature is a major climatic factor related to tree growth, reproduction and mortality (Vayreda et al. 2012, Seidl et al. 2017, Lett and Dorrepaal 2018). Warmer winter (higher BIO11), cooler summer (lower BIO9), and warming (higher BIO5) stimulate boreal tree growth and reproduction (Vayreda et al. 2012, Lett and Dorrepaal 2018), while extreme cold (lower BIO5 and BIO11) would increase tree mortality (Seidl et al. 2017). Human development was not significant in our model, which could be due to the relatively low population density in the North American boreal zone (as mean and median of human development of the 4367 plots are 0.004 and 0, respectively; Supporting information). The relative rank of importance of these variables in our random forest regression model is in good agreement with our discussion on the Gamma regression model, as shown by the rank importance plot in the Supporting information. And the partial dependence of tree density on stand height (Supporting information) corroborates well the hump-shaped effect of stand height on tree density in the random forest model as in the Gamma regression model (Table 2).

The projected change in tree density in 2050 under three climate change scenarios shows an increase in tree density in the North American boreal zone compared to the current 277.2 billion trees (Fig. 4). However, the projected increases (Fig. 4d–f) are not evenly distributed across the region. Instead, tree density in southern Ontario, Newfoundland, northern Alberta, northeastern British Columbia and Yukon Territory is predicted to decrease in 2050, regardless of the climate change scenarios, while tree density is increased in the Canadian Prairies, northern Ontario, northern Quebec and Labrador, with Labrador predicted to have the highest increase (Fig. 4g–i). It is important to note, however, this impact analysis assumes 'all else being equal' but changes the bioclimatic variables under climate change scenarios. In reality, many other factors could also change over the next 30 years. For example, forest fires, land use change, and anthropogenic nitrogen deposition could greatly affect tree density in the region (Schlesinger 2009, Wells et al. 2020).

Despite the effort made in this study, we identified that data availability and quality issues remain a major limitation to modelling tree density variation in boreal forests. First, the exact spatial coordinates of NFI and ABMI plots were not available to the public for the purpose of plot protection. Because of that, the spatial locations of NFI and ABMI plots are 10-km approximates to their published coordinates (Gillis et al. 2005, Stadt et al. 2006). This would inevitably increase uncertainty in density estimation regardless of which models are used. Second, there is considerable spatial variation in the distribution of ground plots, with fewer or no plots in northern and remote areas (Franklin et al. 2017). As a result, tree density estimates in the northern and remote areas are less precise than in southern regions. Further to that, the decreased tree density toward the high north shown in Fig. 3a could be caused by the 10 cm DBH cutoff used in this study and in Crowther et al. (2015), which excluded dwarf birch and spruce trees in the high latitudes (Huang et al. 2013). Third, our study excluded plots that showed obvious anthropogenic and natural disturbances, in order to be comparable with Crowther et al. (2015). Those plot data are meant to estimate tree density of naturally regenerated, undisturbed forests. However, in reality forests in the region are subjected to various disturbances. It is unknown how the estimates derived from our study and that from Crowther et al. (2015) truly represent the density of the real forest landscapes. We did a test by comparing the tree density estimated from our full 4367 plots against those 2740 plots at least 500 m away from the nearest roads. The result shows a little difference in tree density estimated from the two sets of plots (Supporting information), giving us a high confidence in our estimated tree density. In any case, five Canadian jurisdictions did not at all include plots that were affected by pest, pathogen and fire disturbances (Zhang et al. 2015, Xu et al. 2022). Therefore, data on disturbances are not available. To avoid potential bias, we decided not to include disturbed plots in our study as in Crowther et al. (2015). Fourth, despite the proven utility of canopy height for estimating density, the estimation accuracy is subjected to the accuracy and resolution of canopy height data. We noticed that the correlation between LiDAR-based canopy height data and ground-measured stand height was just moderate (the R^2 is 0.45 between the two heights for the 4367 plots) and the LiDAR-based canopy height measurement could be biased for boreal forests (Yang and Kondoh 2020). Compared with the 2019 10-m global canopy height map (Lang et al. 2022) that was used for predicting and mapping tree density in this study, we noticed that the 2005 1-km global canopy height data of Simard et al. (2011) fell closer into the plot measurement period of 1999–2019 but its correlation with stand height was even lower ($R^2 = 0.38$). Furthermore, the GLAS sensor was not well suitable for estimating vegetation height due to sparse LiDAR pulses (Yang and Kondoh 2020). Both maps yielded high prediction errors (up to ± 15 m), which were much larger than the mean decadal increase in canopy height of 2.5 m reported by Gamache and Payette (2004). Based on these two canopy height data, we found the predicted tree densities were in a good agreement (Supporting information). We thus adopted the more precise 10-m canopy height data for prediction and mapping in this study. It is nevertheless important to take note of the warning of Véga and St-Onge (2008) that LiDAR models alone are insufficient to predict plot canopy height. High uncertainty in the 10-m global canopy height map is present in Alaska and Yukon because the GEDI data are not available above 51.6°N though additional airborne LiDAR data are used to predict canopy height in these regions (Lang et al. 2022), which may result in higher uncertainty in our density prediction at the high latitudes. Unfortunately, due to the lack of resources and logistic support, the availability and quality of both ground and airborne data in boreal forests are behind those in temperate and tropical forests (Liang and Gamarra 2020). Finally, we would like to point out that the plots of Alaska used in this study are different from the widely available forest inventory and analysis (FIA) plots maintained by the US Forest Service. The FIA plots do not include trees with DBH < 12.7 cm which is

not aligned with the cutoff DBH of 10.0 cm used in this study and Crowther et al. (2015). However, it is worthwhile to explore how we may take advantage of the FIA plots for estimating tree density in North America.

In conclusion, our models improve the estimation of boreal tree density in North America and shed light on tree density of the global boreal biome. The maps of boreal tree density provide baseline data for modelling forest carbon stock and forest productivity, and for estimating forest biodiversity and competition-driven dynamics of boreal forests. We revealed the necessity to include stand height, NDVI, soil, nitrogen deposition, and other ecologically meaningful predictors in predicting tree density. The Government of Canada has made a commitment to planting two billion trees over 2020–2030 as a nature-based climate solution (https://www.canada.ca/en/campaign/2-billion-trees.html). This laudable goal, however, only accounts for 0.83% of our estimated total number of 240.0 billion boreal trees in Canada (Table 4). If the temperate biome were considered, this percentage would be even lower, speaking to the mitigation challenge through tree planting in the region. Considering the rapidly changing forests in the boreal (Seppälä et al. 2009, D'Amato et al. 2011, Gauthier et al. 2015), there is an urgent need for advanced and accurate data and models for informing adaptation and mitigation planning and policy-making.

Acknowledgements – We thank Tim Crowther for critical comments on an early draft that greatly improved the study.
Funding – Alberta Land Institute (Fangliang He), NSERC-Alliance (Fangliang He), CSC doctoral scholarship for Kun Xu, the Oversea Doctoral Study Program of Guangzhou Elite Project to Jingye Li.

Author contributions

Kun Xu: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (equal); Writing – original draft (lead); Writing – review and editing (equal). **Jingye Li**: Investigation (equal); Writing – review and editing (equal). **Jian Zhang**: Data curation (equal); Investigation (equal); Writing – review and editing (equal). **Dingliang Xing**: Investigation (equal); Methodology (equal); Writing – review and editing (equal). **Fangliang He**: Conceptualization (lead); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Supervision (lead); Writing – review and editing (lead).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ecog.07677.

Data availability statement

Data are available from the Figshare Digital Repository: https://doi.org/10.6084/m9.figshare.26539999.v3 (Xu et al. 2025).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Arlot, S. and Celisse, A. 2010. A survey of cross-validation procedures for model selection. – Stat. Surv. 4: 40–79.
Aussenac, G. 2000. Interactions between forest stands and microclimate: ecophysiological aspects and consequences for silviculture. – Ann. For. Sci. 57: 287–301.
Bonnor, G. M. and Magnussen, S. 1987. Forest inventories in Canada: a framework for change. – For. Chron. 63: 193–198.
Brandt, J. P. 2009. The extent of the North American boreal zone. – Environ. Rev. 17: 101–161.
Brecka, A. F., Shahi, C. and Chen, H. Y. 2018. Climate change impacts on boreal forest timber supply. – For. Policy Econ. 92: 11–21.
Breiman, L. 2001. Random forests. – Mach. Learn. 45: 5–32.
Brunet-Navarro, P., Sterck, F. J., Vayreda, J., Martinez-Vilalta, J. and Mohren, G. M. J. 2016. Self-thinning in four pine species: an evaluation of potential climate impacts. – Ann. For. Sci. 73: 1025–1034.
Clark, D. A. and Clark, D. B. 1984. Spacing dynamics of a tropical rain forest tree: evaluation of the Janzen–Connell model. – Am. Nat. 124: 769–788.
Crowther, T. W. et al. 2015. Mapping tree density at a global scale. – Nature 525: 201–205.
Cutler, A., Cutler, D. R. and Stevens, J. R. 2012. Random forests. – In: Zhang, C. and Ma, Y.-Q. (eds), Ensemble machine learning: methods and applications. Springer, pp. 157–175.
D'Amato, A. W., Bradford, J. B., Fraver, S. and Palik, B. J. 2011. Forest management for mitigation and adaptation to climate change: insights from long-term silviculture experiments. – For. Ecol. Manage. 262: 803–816.
Dutilleul, P. 1993. Modifying the *t* test for assessing the correlation between two spatial processes. – Biometrics 49: 305–314.
Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Rummukainen, M. 2014. Evaluation of climate models. – In: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M. B., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M. (eds), Climate change 2013: the physical science basis. Cambridge Univ. Press, pp. 741–866.
Franklin, S. E., Ahmed, O. S. and Williams, G. 2017. Northern conifer forest species classification using multispectral data acquired from an unmanned aerial vehicle. – Photogramm. Eng. Remote Sens. 83: 501–507.
Gamache, I. and Payette, S. 2004. Height growth response of tree line black spruce to recent climate warming across the forest-tundra of eastern Canada. – J. Ecol. 92: 835–845.
Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z. and Schepaschenko, D. G. 2015. Boreal forest health and global change. – Science 349: 819–822.
Getis, A. and Ord, J. K. 2010. The analysis of spatial association by use of distance statistics. – In: Anselin, L. and Rey, S. J. (eds), Perspectives on spatial data analysis. Springer, pp. 127–145.
Gillis, M. D., Omule, A. and Brierley, T. 2005. Monitoring Canada's forests: the national forest inventory. – For. Chron. 81: 214–221.

Godlee, J. L., Ryan, C. M., Bauman, D., Bowers, S. J.., Carreiras, J. M. B., Chisingui, A. V., Cromsigt, J. P. G. M., Druce, D. J., Finckh, M. and Gonçalves, F. M. 2021. Structural diversity and tree density drives variation in the biodiversity–ecosystem function relationship of woodlands and savannas. – New Phytol. 232: 579–594.
Greenwell, B. M. 2017. pdp: an R package for constructing partial dependence plots. – R J. 9: 421–436.
Hart, T. B., Hart, J. A. and Murphy, P. G. 1989. Monodominant and species-rich forests of the humid tropics: causes for their co-occurrence. – Am. Nat. 133: 613–633.
Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.
Huang, J. G., Stadt, K. J., Dawson, A. and Comeau, P. G. 2013. Modelling growth–competition relationships in trembling aspen and white spruce mixed boreal forests of western Canada. – PLoS One 8: e77607.
Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J. Y., Rixen, M. and Sénési, S. 2020. The CMIP6 data request (DREQ, version 01.00. 31). – Geosci. Model Dev. 13: 201–224.
Kays, S. and Harper, J. L. 1974. The regulation of plant and tiller density in a grass sward. – J. Ecol. 62: 97–105.
Krajicek, J. E., Brinkman, K. A. and Gingrich, S. F. 1961. Crown competition – a measure of density. – For. Sci. 7: 35–42.
Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R. and Wegner, J. D. 2022. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. – Remote Sens. Environ. 268: 112760.
Lett, S. and Dorrepaal, E. 2018. Global drivers of tree seedling establishment at alpine treelines in a changing climate. – Funct. Ecol. 32: 1666–1680.
Liang, J. and Gamarra, J. G. 2020. The importance of sharing global forest data in a world of crises. – Sci. Data 7: 1.
Liaw, A. and Wiener, M. 2002. Classification and regression by randomForest. – R News 2: 18–22.
Long, J. N. 1985. A practical approach to density management. – For. Chron. 61: 23–27.
Lu, B. and Hardin, J. 2021. A unified framework for random forest prediction error estimation. – J. Mach. Learn. Res. 22: 386–426.
Ma, Z., Peng, C., Zhu, Q., Chen, H., Yu, G.-R., Li, W.-Z., Zhou, X.-L., Wang, W.-F. and Zhang, W.-H. 2012. Regional drought-induced reduction in the biomass carbon sink of Canada's boreal forests. – Proc. Natl Acad. Sci. USA 109: 2423–2427.
MacFarlane, D. W., Green, E. J. and Burkhart, H. E. 2000. Population density influences assessment and application of site index. – Can. J. For. Res. 30: 1472–1475.
Madrigal-González, J. et al. 2023. Global patterns of tree density are contingent upon local determinants in the world's natural forests. – Commun. Biol. 6: 47.
Malone, T., Liang, J. and Packee, E. C. 2009. Cooperative Alaska Forest inventory, general technical report PNW-GTR-785. – USDA Forest Service, Pacific Northwest Research Station.
Miller, J. and Franklin, J. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. – Ecol. Modell. 157: 227–247.
Mohler, C. L., Marks, P. L. and Sprugel, D. G. 1978. Stand structure and allometry of trees during self-thinning of pure stands. – J. Ecol. 66: 599–614.
Nelson, T. A. and Boots, B. 2008. Detecting spatial hot spots in landscape ecology. – Ecography 31: 556–566.
Oliver, C. D. and Larson, B. C. 1996. Forest stand dynamics, Update edn. – FES other publications.
Pan, Y., Birdsey, R. A., Phillips, O. L., Houghton, R. A., Fang, J., Kauppi, P. E., Keith, H., Kurz, W. A., Ito, A., Lewis, S. L., Nabuurs, G. J., Shvidenko, A., Hashimoto, S., Lerink, B., Schepaschenko, D., Castanho, A. and Murdiyarso, D. 2024. The enduring world forest carbon sink. – Nature 631: 563–569.
Paré, D. and Bergeron, Y. 1996. Effect of colonizing tree species on soil nutrient availability in a clay soil of the boreal mixedwood. – Can. J. For. Res. 26: 1022–1031.
Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30: 683–691.
Peel, M. C., Finlayson, B. L. and McMahon, T. A. 2007. Updated world map of the Köppen–Geiger climate classification. – Hydrol. Earth Syst. Sci. 11: 1633–1644.
Pretzsch, H. 2009. Forest dynamics, growth and yield. – Springer.
Priestley, C. H. B. and Taylor, R. J. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. – Mon. Weather Rev. 100: 81–92.
Reyes-Hernández, V. and Comeau, P. G. 2014. Survival probability of white spruce and trembling aspen in boreal pure and mixed stands experiencing self-thinning. – For. Ecol. Manage. 323: 105–113.
Schlesinger, W. H. 2009. On the fate of anthropogenic nitrogen. – Proc. Natl Acad. Sci. USA 106: 203–208.
Schwalb, B., Tresch, A., Torkler, P., Duemcke, S., Demel, C., Ripley, B. and Venables, B. 2020. LSD: lots of superior depictions. Ver. 4.1-0. – https://cran.r-project.org/web/packages/LSD/LSD.pdf.
Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M. J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T. A. and Reyer, C. P. O. 2017. Forest disturbances under climate change. – Nat. Clim. Change 7: 395–402.
Seppälä, R., Alexander, B. and Katila, P. 2009. Adaptation of forests and people to climate change: a global assessment report (IUFRO World series volume 22). – Int. Union For. Res. Org. (IUFRO).
Simard, M., Pinto, N., Fisher, J. B. and Baccini, A. 2011. Mapping forest canopy height globally with spaceborne lidar. – J. Geophys. Res. 116: G04021.
Stadt, J. J., Schieck, J. and Stelfox, H. A. 2006. Alberta biodiversity monitoring program – monitoring effectiveness of sustainable forest management planning. – Environ. Monit. Assess. 121: 33–46.
Sterck, F., Vos, M., Hannula, S. E., de Goede, S., de Vries, W., den Ouden, J., Nabuurs, G. J., van der Putten, W. and Veen, C. 2021. Optimizing stand density for climate-smart forestry: a way forward towards resilient forests with enhanced carbon storage under extreme climate events. – Soil Biol. Biochem. 162: 108396.
ter Steege, H. et al. 2013. Hyperdominance in the Amazonian tree flora. – Science 342: 1243092.
ter Steege, H. et al. 2023. Mapping density, diversity and species-richness of the Amazon tree flora. – Commun. Biol. 6: 1130.
Tobner, C. M., Paquette, A., Reich, P. B., Gravel, D. and Messier, C. 2014. Advancing biodiversity–ecosystem functioning science using high-density tree-based experiments over functional diversity gradients. – Oecologia 174: 609–621.
Vallejos, R., Osorio, F. and Bevilacqua, M. 2020. Spatial relationships between two georeferenced variables: with applications in R. – Springer.
Vayreda, J., Martinez-Vilalta, J., Gracia, M. and Retana, J. 2012. Recent climate changes interact with stand structure and management to determine changes in tree carbon stocks in Spanish forests. – Global Change Biol. 18: 1028–1041.

Véga, C. and Stonge, B. 2008. Height growth reconstruction of a boreal forest canopy over a period of 58 years using a combination of photogrammetric and lidar models. – Remote Sens. Environ. 112: 1784–1794.
Wells, J. V., Dawson, N., Culver, N., Reid, F. A. and Morgan Siegers, S. M. 2020. The state of conservation in North America's boreal forest: issues and opportunities. – Front. For. Global Change 3: 90.
Westoby, M. 1984. The self-thinning rule. – Adv. Ecol. Res. 14: 167–225.
White, J. and Harper, J. L. 1970. Correlated changes in plant size and number in plant populations. – J. Ecol. 58: 467–485.
Woodall, C. W. and Weiskittel, A. R. 2021. Relative density of United States forests has shifted to higher levels over last two decades with important implications for future dynamics. – Sci. Rep. 11: 18848.
Xu, K., Huang, S. and He, F. 2022. Modeling fire hazards for the maintenance of long-term forest inventory plots in Alberta, Canada. – For. Ecol. Manage. 513: 120206.
Xu, K., Li, J., Zhang, J., Xing, D. and He, F. 2025. Data from: How many trees are there in the North American boreal forest? – Figshare Digital Repository, https://doi.org/10.6084/m9.figshare.26539999.v3.
Xu, Y., Li, C., Sun, Z., Jiang, L. and Fang, J. 2019. Tree height explains stand volume of closed-canopy stands: evidence from forest inventory data of China. – For. Ecol. Manage. 438: 51–56.
Yang, W. and Kondoh, A. 2020. Evaluation of the Simard et al. 2011 global canopy height map in boreal forests. – Remote Sens. 12: 1114.
Zhang, J., Huang, S., Hogg, E. H., Lieffers, V., Qin, Y. and He, F. 2014. Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. – Biogeosciences 11: 2793–2808.
Zhang, J., Huang, S. M. and He, F. 2015. Half-century evidence from western Canada shows forest dynamics is primarily driven by competition followed by climate. – Proc. Natl Acad. Sci. USA 112: 4009–4014.