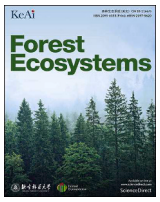
Forest Ecosystems 15 (2026) 100396

Contents lists available at ScienceDirect

Forest Ecosystems

journal homepage: www.keaipublishing.com/cn/journals/forest-ecosystems



# Prediction of individual tree seedling volume, biomass, and carbon content via allometric projection

James A. Westfall *

*U.S. Forest Service, Northern Research Station, 3460 Industrial Drive, York 17402, PA, USA*

ARTICLE INFO

*Keywords:*
Tree taper
Seedling height
Groundline diameter
Understory vegetation
Forest inventory
Tree regeneration

ABSTRACT

When making assessments of forest resources, there is nearly ubiquitous interest in quantifying current status and trends in tree biomass and carbon stocks. While important at various spatial scales, typical estimations pertinent to broad forest management and policy issues are conducted for large areas such as state, regional, and national perspectives. These assessments are usually accomplished using large-area forest inventory data collected by National Forest Inventory (NFI) programs. While NFI efforts commonly collect size data for individual trees, there is often limited information for tree seedlings, e.g., frequency by species. To fully describe the tree population across the entire range of sizes present, this study proposes methods to predict individual seedling groundline diameter and height using models developed from trees having a diameter at breast height (DBH) less than 7.62 cm. These attributes are subsequently used for the prediction of seedling stem volume, total aboveground biomass, and carbon content. The results suggest a smooth transition in tree attributes as size increases to where direct measurement of individual trees and prediction of their volume, biomass, and carbon are implemented as part of standard inventory protocols. Analyses including the full spectrum of tree sizes show that seedlings contribute roughly 0.6%–0.7% of the total tree volume/mass. This additional suite of information provides opportunities for more holistic assessments across the full spectrum of the tree resource or for specialized subdomains that include the seedling component.

## 1. Introduction

Individual tree volume, biomass, and carbon attributes often serve as the basis for forest resource assessments that are critical to forest management and policy decisions. A substantial amount of research exists, with significant emphasis on volumes for trees of merchantable size to facilitate analyses related to wood procurement, product output potential, and forest management objectives (Avery and Burkhart, 2002; Kershaw Jr et al., 2016). As assessments of forest sector contributions to biomass and carbon sequestration and storage became more prevalent, the importance of including smaller-sized trees was recognized (Nelson et al., 2014; Singh et al., 2011). However, many studies only include trees having attained a minimum height such that diameter at breast height (DBH) could be measured, as DBH was often used as the primary predictor variable in regression models (Chave et al., 2005; Jenkins et al., 2003). Consequently, very small trees (seedlings) are usually left out of the tree pool and are commonly subsumed into a broader category that includes various forms of understory vegetation (Muukkonen et al., 2006; Smith et al., 2006). These practices preclude understanding of tree structure and dynamics occurring across the full spectrum of tree sizes present in forest ecosystems.

Nonetheless, some studies exist where individual seedling attributes are specifically addressed (Annighöfer et al., 2016; Menéndez-Miguélez et al., 2022; Pati et al., 2022; Thippawan et al., 2023). While the strict definition of what tree size thresholds constituted classification of seedling status can vary among studies, typical predictor variables include groundline diameter and height, and models are often developed by species and/or species groups. A limitation for detailed studies of tree seedlings is that the requisite data collection is often time-consuming and costly. Thus, where data are lacking but there is a need for seedling-specific information, it may be advantageous to use indirect methods that build upon extensions of existing knowledge. For example, the United States National Forest Inventory (NFI) implemented by the Forest Inventory and Analysis (FIA) program predicts stem volume, aboveground biomass, and aboveground carbon for saplings having 2.54 cm $\leq$ DBH $<$ 12.70 cm (Burrill et al., 2024). However, stems

* Corresponding author.
*E-mail address:* james.westfall@usda.gov.
Peer review under the responsibility of Editorial Office of Forest Ecosystems.

https://doi.org/10.1016/j.fecs.2025.100396
Received 7 February 2025; Received in revised form 3 July 2025; Accepted 26 September 2025
2197-5620/Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

where DBH < 2.54 cm are classified as seedlings, are not subject to individual-unit measurements, and consequently have no attribute predictions per seedling. In pursuance of obtaining predictions for seedling characteristics with no observed individual seedling data available, the objectives of this study are to 1) develop a strategy to build upon allometric relationships observed in neighboring tree size data realms to predict groundline diameter and height for individual seedlings, 2) use the seedling size information to develop stem volume, aboveground biomass, and aboveground carbon predictions, and 3) assess the contribution of seedlings to these attribute stocks in the context of the full range of tree sizes.

## 2. Materials and methods

### 2.1. Study data sources

Two datasets were used in the analysis. The primary dataset arises from NFI data collected by the Northern Research Station, Forest Inventory and Analysis (NRS-FIA) program within the U.S. Forest Service. NRS-FIA collects a myriad of data within their nested plot design (Fig. 1). Relevant to this study are the delineation of any boundaries between forest and nonforest conditions occurring within the plot area. These data allow for the determination of the proportion of the plot area that is forested. Another notable feature is that trees are measured in the different plot sizes based on their DBH of 1.37 m. Trees having attained DBH ≥ 12.70 cm are measured within the subplot area (radius 7.32 m), whereas seedling (<2.54 cm DBH) and sapling (2.54 cm ≤ DBH <12.70 cm) sized trees are measured within the microplot area (radius 2.07 m). The macroplot area is not used in this research. The study focused on a subset of 1,031 inventory plots measured between 2015 and 2019 in the state of Minnesota (MN), USA, that were designated to receive detailed measurements of seedling information (regeneration indicator (RI)) (McWilliams et al., 2015). Under these protocols, the numbers of seedlings are recorded by species and total height (HT) class where the height classes are defined as (1) 0.05 m ≤ HT < 0.15 m, (2) 0.15 m ≤ HT < 0.30 m, (3) 0.30 m ≤ HT < 0.91 m, (4) 0.91 m ≤ HT < 1.52 m, (5) 1.52 m ≤ HT < 3.05 m, and (6) HT ≥ 3.05 m. Data collected for saplings that are relevant to this study include species, DBH, and HT. For both tree and sapling size trees, predictions of total stem volume, biomass, and carbon content, branch biomass, and total aboveground biomass and carbon (excluding foliage) based on Westfall et al. (2024) are already stored in the database. For analytical purposes, all species were assigned to either the hardwood (39 species) or softwood (9 species) class using FIA definitions (Burrill et al., 2024), and the sapling data were reduced to only those with DBH ≤7.62 cm. A summary of the FIA sapling data is provided in Table 1. Information regarding the seedling data will be provided later in this section.

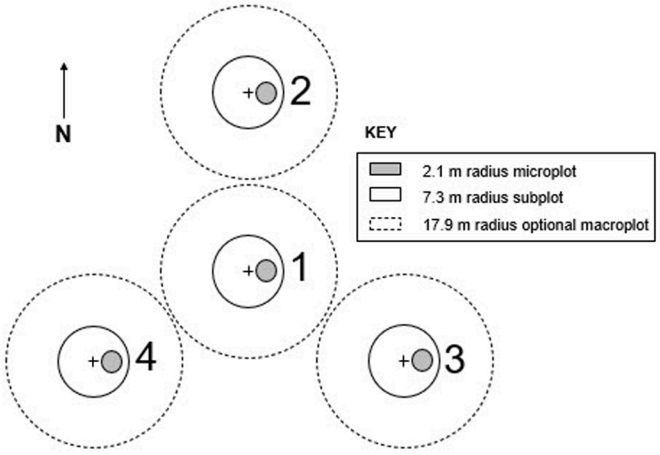



**Fig. 1.** The FIA plot design.

The second dataset used in this study was obtained from Radtke et al. (2023). Specifically, the tree taper (STEM) and tree summary (TREE) tables were used. These data included study location information, species, measurements of diameter: height (*d*: *h*) pairs along the tree stem, HT, and DBH. Due to the intent of developing models for small trees, data were only retained for trees having a DBH <7.62 cm. Also, the data were limited to only those species found in the FIA data for MN. Table 2 provides information regarding these data.

### 2.2. Attribute prediction

To facilitate understanding of the detailed methodological construct to be presented below, a general outline of procedural steps is initially provided:

1) Assign heights to individual seedlings ($HT_{seed}$) in the RI data based on the observed seedling height class and a rescaled sapling height distribution.
2) Fit a tree taper model for sapling-sized trees (Radtke et al., 2023) and use this model to predict sapling groundline diameter ($DGL_{sap}$) for the FIA data.
3) Use observed sapling height ($HT_{sap}$) and predicted $DGL_{sap}$ to develop a model that predicts $DGL_{sap}$ as a function of $HT_{sap}$.
4) Use the preceding model to predict the RI seedling groundline diameter ($DGL_{seed}$) as a function of $HT_{seed}$.
5) Use the sapling taper data to estimate a stem shape parameter *r* and use this parameter along with $DGL_{seed}$ and $HT_{seed}$ to estimate seedling stem volume ($V_{seed}$).
6) Convert $V_{seed}$ to seedling stem biomass ($BS_{seed}$) via specific gravity (SG) information.
7) Use predicted sapling branch biomass ($BR_{sap}$) to develop a model that predicts $BR_{sap}$ as a function of $DGL_{sap}$ and $HT_{sap}$.
8) Use the preceding model to predict seedling branch biomass ($BR_{seed}$) using $DGL_{seed}$ and $HT_{seed}$.
9) Sum $BS_{seed}$ and $BR_{seed}$ to obtain seedling total aboveground biomass ($AGB_{seed}$).
10) Convert $AGB_{seed}$ to seedling total aboveground carbon content $AGC_{seed}$ using carbon fraction (CF) information.

Given the goal to predict volume, biomass, and carbon content of individual seedlings, the primary analytical effort is focused on developing seedling height and diameter predictions. Lacking an empirical basis for the seedling height distribution, the distribution of sapling (2.54 cm ≤ DBH <12.70 cm) heights was examined. The distribution of sapling heights is conditional on the sapling having attained a minimum DBH of 2.54 cm. Based on a visual inspection of the data, a minimum sapling height of 4.88 m was established to create a prototype height distribution exhibiting the "reverse-J" characteristic when plotted against frequency. Also, these data were rescaled to mimic the height range of the seedling data, where the largest seedling height expected to be observed would be less than 4.88 m. This rescaled sapling height distribution was then summarized by calculating the proportion of observations in each 0.3 m height class (0.1–4.9 m). To closely match these proportions, a number of distributions were created by generating a random rescaled height for each sapling in the dataset over a range of values for $\lambda$ (Eq. 1):

$$HT_{ij}^{\lambda} = LC_j + \left(UC_j - LC_j\right)R(U)^{\lambda} \quad (1)$$

where $HT_{ij}^{\lambda}$ is the rescaled height (m) for sapling *i* in height class *j* for a specified value of $\lambda$, $R(U)$ is the random variate from a uniform $U \sim (0, 1)$ distribution, and $LC_j$ and $UC_j$ are the respective upper and lower thresholds for height class *j*. The means of $HT_{ij}^{\lambda}$ values in each 0.3 m

**Table 1**
Sample size ($n$) and distribution information for sapling DBH and HT by species class.

| Species class | $n$ | Diameter breast-height (DBH, cm) | | | | Total height (HT, m) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Min. | Mean | Max. | Std. dev. | Min. | Mean | Max. | Std. dev. |
| Hardwood | 6540 | 2.54 | 4.38 | 7.62 | 1.47 | 1.52 | 6.40 | 15.85 | 1.93 |
| Softwood | 2812 | 2.54 | 4.76 | 7.62 | 1.51 | 1.83 | 4.77 | 11.89 | 1.52 |

Note: 3 inch max were retained, only RI plots.

**Table 2**
Summary of taper data tree DBH and HT characteristics ($n$ is the total number of trees and observations per tree is the number of within-tree taper measurements).

| Species class | $n$ | Diameter breast height (DBH, cm) | | | Total height (HT, m) | | | Observations per tree | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Min. | Mean | Max. | Min. | Mean | Max. | Min. | Mean | Max. |
| Hardwood | 633 | 1.78 | 4.62 | 7.37 | 2.74 | 6.96 | 12.95 | 3 | 6.50 | 20 |
| Softwood | 265 | 2.03 | 5.52 | 7.37 | 2.19 | 6.43 | 12.31 | 3 | 8.69 | 23 |

height class were compared to those from the original rescaled sapling heights, and the closest distribution was determined from a minimum mean-squared error criterion. The analysis was conducted separately for hardwood and softwood species classes, where the best outcomes occurred at $\lambda = 1.72$ and 2.26, respectively. Subsequently, each seedling in the RI dataset was assigned a height based on the observed height class and appropriate $\lambda$ value.

Having assigned a height to each seedling, an associated diameter is needed. However, many seedlings will not have attained a height (1.37 m) where DBH would be measured. Therefore, the diameter at ground-line (DGL) is more appropriate for these data. The first step in this process entails fitting taper models using the taper dataset presented in the preceding section. In this case, the model presented by Kozak et al. (1969) was chosen as the initial analysis with the small tree taper data suggesting difficulty in model fitting for more complex specifications, e.g., Max and Burkhart (1976). Due to the lack of independence of within-tree observations, a mixed effect model was specified and fitted in this form (Eq. (2)):

$$d^2 = \mathrm{DBH}^2\left[(b_1 + \theta)(x-1) + b_2\left(x^2 - 1\right)\right] + \in \tag{2}$$

where $d$ is the stem diameter (cm) at height $h$ (m), $x$ is $h$/HT, $b_1$ and $b_2$ are estimated parameters, $\theta$ is the random effect parameter, and $\in$ is the random residual error.

The taper model cannot be directly applied to seedlings at this point because 1) there is no DBH information for seedlings, and 2) some seedlings would not have attained a sufficient height for DBH measurement. As such, the next step is to predict $\mathrm{DGL_{sap}}$ in the FIA data in pursuance of ultimately developing a model to predict $\mathrm{DGL_{seed}}$ from seedling height ($\mathrm{HT_{seed}}$). Prediction of $\mathrm{DGL_{sap}}$ from Eq. 2 was accomplished using $x = 0$ in Eq. 3:

$$\mathrm{DGL_{sap}} = \sqrt{\mathrm{DBH}^2[b_1(x-1) + b_2(x^2-1)]} \tag{3}$$

At this point, observed $\mathrm{HT_{sap}}$ and predicted $\mathrm{DGL_{sap}}$ information are present, and the relationship between them was described by Eq. 4:

$$\mathrm{DGL_{sap}} = c_0\left(1 - \exp\left(c_1 \mathrm{HT_{sap}}\right)\right)^{c_2} + \in \tag{4}$$

where $c_0$, $c_1$, and $c_2$ are estimated parameters. Given now that DGL can be predicted from HT, model Eq. 4 was applied to the seedling data to obtain $\mathrm{DGL_{seed}}$:

$$\mathrm{DGL_{seed}} = c_0\left(1 - \exp\left(c_1 \mathrm{HT_{seed}}\right)\right)^{c_2} \tag{5}$$

Distribution summaries for the seedling data are now provided based on individual seedling predicted $\mathrm{DGL_{seed}}$ and $\mathrm{HT_{seed}}$ (Table 3).

Prediction of seedling stem volumes based on $\mathrm{DGL_{seed}}$ and $\mathrm{HT_{seed}}$ suggests a conic solution, and assuming a simple cone would be convenient. However, further refinement may be appropriate, e.g., to account for potential flare at the stem base. To better approximate the stem shape, the tree taper data were used to estimate the $r$ parameter in the following model (McTague and Weiskittel, 2021):

$$\left(\frac{d}{\mathrm{DBH}}\right)^2 = \left(\frac{\mathrm{HT} - h}{\mathrm{HT} - 1.37}\right)^r + \in \tag{6}$$

A value of $r \approx 2$ would indicate the conventional cone shape. Subsequently, the volume of the seedling stem is calculated from Eq. 7:

$$V_{\mathrm{seed}} = \frac{1}{r+1}\frac{\pi}{40000}\mathrm{DGL_{seed}}^2 \mathrm{HT_{seed}} \tag{7}$$

where $V_{\mathrm{seed}}$ is the seedling stem volume ($\mathrm{m}^3$) and $\pi \approx 3.1416$. At this point, the emphasis shifts to obtaining total aboveground biomass and

**Table 3**
Sample size ($n$) and distribution summaries for $\mathrm{DGL_{seed}}$ and $\mathrm{HT_{seed}}$ by height class and hardwood/softwood species class.

| | Height class | $n$ | $\mathrm{DGL_{Seed}}$ (cm) | | | $\mathrm{HT_{Seed}}$ (m) | | |
|---|---|---|---|---|---|---|---|---|
| | | | Min. | Mean | Max. | Min. | Mean | Max. |
| Hardwood | 1 | 10,592 | 0.008 | 0.018 | 0.040 | 0.050 | 0.087 | 0.150 |
| | 2 | 8795 | 0.044 | 0.068 | 0.113 | 0.160 | 0.211 | 0.300 |
| | 3 | 14,275 | 0.118 | 0.267 | 0.555 | 0.310 | 0.529 | 0.900 |
| | 4 | 4159 | 0.564 | 0.770 | 1.138 | 0.910 | 1.131 | 1.510 |
| | 5 | 4846 | 1.148 | 1.753 | 2.811 | 1.520 | 2.085 | 3.050 |
| | 6 | 1887 | 2.823 | 3.550 | 4.776 | 3.060 | 3.726 | 4.878 |
| Softwood | 1 | 1385 | 0.002 | 0.005 | 0.016 | 0.050 | 0.080 | 0.150 |
| | 2 | 1199 | 0.018 | 0.031 | 0.067 | 0.160 | 0.203 | 0.300 |
| | 3 | 2567 | 0.071 | 0.199 | 0.576 | 0.310 | 0.491 | 0.900 |
| | 4 | 1038 | 0.588 | 0.820 | 1.424 | 0.910 | 1.086 | 1.508 |
| | 5 | 958 | 1.442 | 2.207 | 3.970 | 1.520 | 1.984 | 3.048 |
| | 6 | 64 | 3.989 | 4.873 | 6.477 | 3.060 | 3.665 | 4.878 |

carbon attributes for individual seedlings. The known species of each seedling was used to apply species-specific SG information in the conversion of stem volume to stem biomass. The SG values were based on those documented by the FIA program (Burrill et al., 2024):

$$BS_{seed} = V_{seed} \times SG \times 1000 \quad (8)$$

where $BS_{seed}$ is the seedling stem biomass (kg). However, seedling branches should also be included for estimation of aboveground biomass, and the analysis returns to the FIA sapling data where branch biomass predictions are present. As the models used to predict sapling branch biomass rely on DBH as a predictor (Westfall et al., 2024), there is no way to directly make predictions for seedlings using existing models. However, using the sapling $HT_{sap}$ and $DGL_{sap}$ information, a model that could be applied to seedlings may take the form of Eq. 9:

$$BR_{sap} = d_1 DGL_{sap}^{d_2} HT_{sap}^{d_3} + \in \quad (9)$$

where $BR_{sap}$ is the sapling branch biomass (kg) and $d_1$–$d_3$ are estimated parameters. Subsequently, predictions of $BR_{seed}$ (kg) are obtained from Eq. 10:

$$BR_{seed} = d_1 DGL_{seed}^{d_2} HT_{seed}^{d_3} \quad (10)$$

The total aboveground biomass for seedlings $AGB_{seed}$ (kg) arises from the sum of the stem and branch biomass:

$$AGB_{seed} = BS_{seed} + BR_{seed} \quad (11)$$

Finally, the $AGB_{seed}$ serves as the conduit for the calculation of seedling aboveground carbon $AGC_{seed}$ (kg) using CF values that define the carbon proportion of woody biomass. Although it is somewhat common to use 0.5 as a generic carbon fraction, the values used here were the species-specific fractions used by FIA (Burrill et al., 2024):

$$AGC_{seed} = AGB_{seed} \times CF \quad (12)$$

For the regression models described above, i.e., those with an $\in$ term, goodness-of-fit was assessed via $R$-squared ($R^2$), root mean squared error (RMSE), and mean absolute error (MAE) statistics. Technically, the $R^2$ values reported are pseudo-$R^2$ due to the nonlinear model forms. Eqs. 13–15 are given based on a generic dependent variable $y$:

$$R^2 = 1 - \frac{\sum^n (y_i - \hat{y}_i)^2}{\sum^n (y_i - \bar{y})^2} \quad (13)$$

$$RMSE = \sqrt{\frac{\sum^n (y_i - \hat{y}_i)^2}{n}} \quad (14)$$

$$MAE = \frac{\sum^n |y_i - \hat{y}_i|}{n} \quad (15)$$

where $y_i$ is the observation from tree $i$, $\hat{y}_i$ is the predicted value for tree $i$, $\bar{y}$ is the mean of the observed values, and $n$ is the number of observations.

### 2.3. Population estimation

Population estimates for volume, biomass, and carbon attributes were calculated using the full suite of data from the 1031 RI plots in MN. Calculations were performed according to the post-stratified estimation procedures described in Westfall et al. (2022). In the referenced documentation, details are provided on aggregating the individual stem observations to the plot level, including handling of stems occurring in different nested plots sizes and application of adjustment factors to account for plots that were only partially measured. For this study, the described ratio estimators were used to obtain estimates on a per hectare of forestland basis. Of primary interest were estimates pertaining to $V_{seed}$, $AGB_{seed}$, and $AGC_{seed}$. However, corresponding estimates that include both sapling and tree data were also calculated to provide context for the seedling contribution to the overall tree species population.

## 3. Results

There were a number of models employed to predict $V_{seed}$, $AGB_{seed}$, and $AGC_{seed}$ for individual seedlings. An important initial modeling effort was fitting the taper model Eq. (2) that allowed for the prediction of $DGL_{sap}$, where all estimated parameters were statistically significant at the 95 % confidence level and $R^2$ values exceeded 0.80 for both hardwood and softwood species classifications (Table 4).

Subsequent modeling of $DGL_{sap}$ as a function of $HT_{sap}$ (Eq. 4) to enable the ability for prediction of $DGL_{seed}$ using $HT_{seed}$ showed strong allometric relationships were present with $R^2$ values exceeding 0.95 and statistically significant estimated parameters at the 95% confidence level (Table 5).

Upon obtaining $DGL_{seed}$ and $HT_{seed}$, it was necessary to approximate the stem form via the $r$ parameter from model Eq. 6. The results showed the form to be intermediate between conic and neiloidal for hardwood species ($r = 2.524$) and between parabolic and conic form for softwood species ($r = 1.815$). The corresponding forms are shown in Fig. 2.

The preceding models provided the ability to estimate $V_{seed}$. However, for the attributes $AGB_{seed}$ and $AGC_{seed}$, it was desirable to also include the contribution of $BR_{seed}$. Again, the sapling data was used to describe the relationship between branch biomass and sapling size as characterized by $DGL_{sap}$ and $HT_{sap}$ in Eq. 9. In keeping with previous results, the model fit for hardwood species was weaker than for softwood species, although $R^2$ exceeded 0.80, and the estimated parameters were statistically significant based on the 95% confidence criteria (Table 6).

Population estimates pertaining to $V_{seed}$, $AGB_{seed}$, and $AGC_{seed}$ suggest that forests in MN, USA. contain nearly 1.1 $m^3 \cdot ha^{-1}$ of stem volume, about 0.5 $Mg \cdot ha^{-1}$ of biomass, and approximately 0.2 $Mg \cdot ha^{-1}$ of carbon. In the context of all tree stems, seedlings generally comprised between 0.6% and 0.7% of the population total (Table 7).

## 4. Discussion

A key to successful implementation of the methods was the tree taper data (Radkte et al., 2023) that provided the ability to fit the taper model Eq. 2. Subsequently, prediction of $DGL_{sap}$ for the sapling-sized trees in the FIA data formed the basis of the allometric projection, assuming relationships between $DGL_{sap}$ and $HT_{sap}$ were representative of $DGL_{seed}$ and $HT_{seed}$ for seedlings.

In conjunction with the estimated stem shape parameter $r$ from model Eq. 6, $DGL_{seed}$ and $HT_{seed}$ were used for seedling stem volume prediction and subsequent conversion to biomass and carbon content. In the context of the entire tree population, the results show a small input (<1%) from seedlings to these attribute pools (Table 7). It should be noted that the seedling contribution will depend on forest conditions within the population being studied. For example, a larger seedling component may be expected in areas where relatively young stands predominate due to both the deficiency of large trees as well as the likelihood that a robust seedling population will be present due to understory light availability. In the specific context of the state of MN used in this analysis, there has been a trend towards older stand ages (Hillard et al., 2022), and thus the seedling component is apt to be small.

**Table 4**
Estimated model parameters ($b_1$ and $b_2$) and goodness-of-fit statistics ($R^2$, RMSE, and MAE) for model Eq. 2 by species class.

| Species class | $b_1$ | $b_2$ | $R^2$ | RMSE | MAE |
|---|---|---|---|---|---|
| Hardwood | −4.0845 | 2.1479 | 0.83 | 10.95 | 5.20 |
| Softwood | −2.5632 | 0.9346 | 0.92 | 5.74 | 3.16 |

**Table 5**
Estimated model parameters ($c_0$, $c_1$, and $c_2$) and goodness-of-fit statistics ($R^2$, RMSE, and MAE) for model Eq. 4 by species class.

| Species class | $c_0$ | $c_1$ | $c_2$ | $R^2$ | RMSE | MAE |
|---|---|---|---|---|---|---|
| Hardwood | 13.4479 | −0.1438 | 1.5114 | 0.96 | 1.29 | 0.99 |
| Softwood | 10.2285 | −0.3394 | 2.1555 | 0.97 | 1.11 | 0.87 |

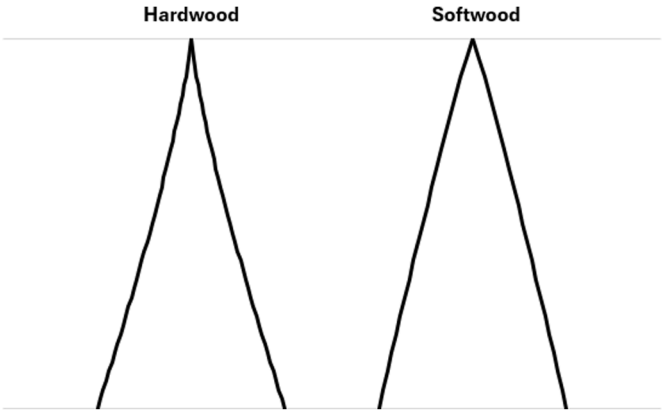



**Fig. 2.** Predicted seedling stem shape for hardwood ($r = 2.524$) and softwood ($r = 1.815$) species classes.

**Table 6**
Estimated model parameters ($d_1$, $d_2$, and $d_3$) and goodness-of-fit statistics ($R^2$, RMSE, and MAE) for model Eq. 9 by species class.

| Species class | $d_1$ | $d_2$ | $d_3$ | $R^2$ | RMSE | MAE |
|---|---|---|---|---|---|---|
| Hardwood | 3.0814 | −0.6454 | 0.0064 | 0.82 | 0.40 | 0.23 |
| Softwood | 2.7137 | −0.5242 | 0.0170 | 0.96 | 0.30 | 0.23 |

**Table 7**
Estimates of stem volume, aboveground biomass, and aboveground carbon for seedlings and for combined saplings and trees (DBH ≥2.54 cm) for the state of MN, USA.

| Attribute | Seedlings | | Saplings and Trees | | Seedling proportion |
|---|---|---|---|---|---|
| | Estimate | Std. err. | Estimate | Std. err. | |
| Stem volume ($m^3 \cdot ha^{-1}$) | 1.06 | 0.08 | 146.48 | 1.25 | 0.0072 |
| AGB ($Mg \cdot ha^{-1}$) | 0.48 | 0.04 | 82.73 | 0.69 | 0.0058 |
| AGC ($Mg \cdot ha^{-1}$) | 0.23 | 0.02 | 40.34 | 0.34 | 0.0057 |

Although various models were needed to assign $DGL_{seed}$ and $HT_{seed}$ to every seedling, an advantage of building on the FIA sapling data is that it helps ensure a smooth transition between the stem size classes. When there is no explicit linkage, incongruent values can occur when using different models depending on tree size class (Dixon, 2002; Woodall et al., 2011). Examination of the attribute values based on the methods presented showed consistent patterns across the range of seedling and sapling tree sizes (Fig. 3). This behavior is important, e.g., when seedling information is used as the basis for stand projections that implement growth and mortality assumptions to predict the future forest composition and structure (Eerikäinen et al., 2007; Weiskittel et al., 2011). The height class data collected under FIA RI protocols helps to better define seedling population characteristics, but having observed individual seedling diameter and height information along with volume/mass may encourage further study of seedling demographics and functions in forests at various developmental stages (Canham and Murphy, 2017; Chang-Yang et al., 2021). Such studies may lead to

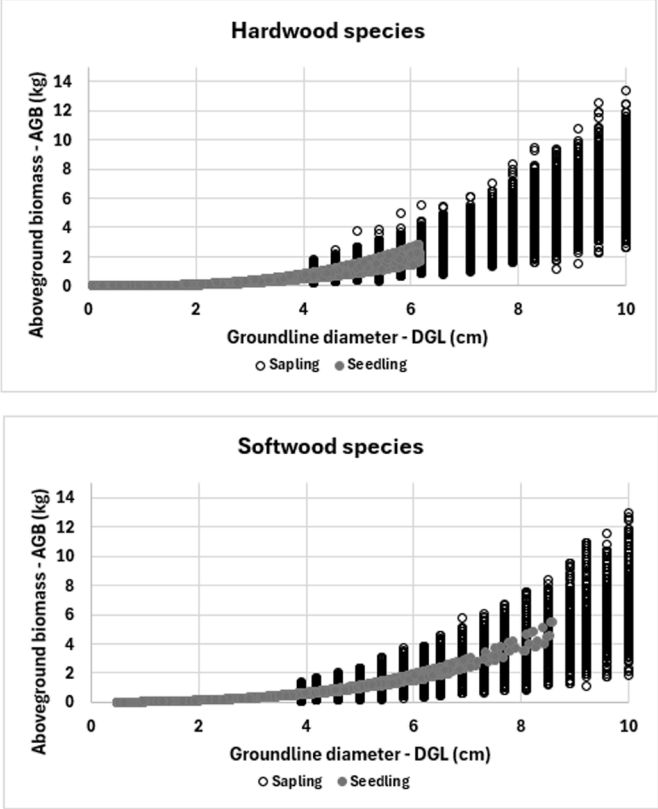



**Fig. 3.** Relationship between AGB (kg) and $DGL^2HT$ ($cm^3$) for seedlings and saplings ($DGL^2HT \leq 30{,}000$ $cm^3$) by species class.

further development or refinement of individual tree growth and mortality predictions across a continuum of tree sizes (Brose et al., 2008; Waller et al., 2024).

Perhaps the best comparison of results to an independent study in terms of similar species and forest types can be made with seedling biomass predictions reported by Annighöfer et al. (2016) for Western European forests. Based on the reported generic models for softwood and hardwood species, the $AGB_{seed}$ predictions were about 1.76 and 2.44 kg at $DGL^2_{seed}HT_{seed} = 10{,}000$ $cm^3$ for hardwood and softwood species, respectively. In this study, $AGB_{seed}$ predictions were roughly 1.32 and 1.80 kg at $DGL^2_{seed}HT_{seed} = 10{,}000$ $cm^3$ for hardwood and softwood species, respectively. This study predicts smaller values of $AGB_{seed}$ for both hardwood and softwood species in comparison to Annighöfer et al. (2016), which may be attributed to various factors such as dissimilar species composition and frequency and/or differing allometric relationships attributable to macro- and micro-environment factors. The agreement between studies may improve or deteriorate for other values of $DGL^2_{seed}HT_{seed}$ depending on the differences in model trajectories.

Other data sources may be consulted to make generalized comparisons with the results shown in this study. Smith et al. (2013) indicate that the Northern Lake States region (which includes the state of MN used in this study) had AGB of live trees (DBH ≥2.54 cm) of 52.9 $Mg \cdot ha^{-1}$, and understory vegetation AGB is 1.8 $Mg \cdot ha^{-1}$. However, more recently, Hoover and Smith (2021) show AGB of live trees as 43.9 $Mg \cdot ha^{-1}$ for the same region, which compares favorably with the result of 40.34 $Mg \cdot ha^{-1}$ reported in Table 7. Based on Smith et al. (2013), the ratio of understory vegetation C to C of live trees is 0.034, and extension of this result to the current analysis suggests understory vegetation C in MN to be approximately 1.37 $Mg \cdot ha^{-1}$. To approximate the seedling C component, Johnson et al. (2017) suggest that seedling C comprises about 1% of understory vegetation C based on a model that uses seedling

density as a predictor. Using the 1% approximation suggests seedling C in MN is 0.014 Mg·ha$^{-1}$. Yet, the estimate in this study was much larger (0.23 Mg·ha$^{-1}$). One likely contributor to the discrepancy is that Johnson et al. (2017) defined tree seedlings as stems having $HT_{seed}$ less than 1.37 m; whereas FIA defines seedlings as stems where DBH <2.54 cm. Thus, there are considerable size distribution differences between the two methods that are not accounted for by changes in per unit-area density alone. To assess the potential effect of definitional differences, seedling C for MN was recalculated only using seedlings having $HT_{seed}$ less than 1.37 m. This changed the seedling C estimate from 0.23 to 0.006 Mg·ha$^{-1}$, which is considerably closer to the 0.014 Mg·ha$^{-1}$ approximation and indicates a substantial effect of the FIA definition including seedlings of larger size.

In the preceding content, the RI plots having height class information were used to facilitate the prediction of $HT_{seed}$ and subsequently $DGL_{seed}$. Application of these methods to all NFI plots would be highly advantageous in terms of increasing sample sizes and widening spatial extents of application. One potential avenue may be to build upon the distribution modeling approach described in Westfall and McWilliams (2017), where rescaled sapling heights were used to develop seedling height distribution parameters estimated as functions of spatial location, stand conditions, and seedling density. $HT_{seed}$ values could then be generated from these predicted distributions. However, considerable research in this area is needed to understand the key drivers of seedling height distributions. Further, various factors may affect allometric relationships between $HT_{seed}$ and $DGL_{seed}$ (Zhang et al., 2020). Extension to other NFIs would depend on country-specific field work protocols and data availability (Tomppo et al., 2010).

The usage of predicted values in lieu of observed data with various interrelated models may raise concerns. Clearly, model error and related goodness-of-fit statistics cannot be interpreted in the same context when compared to using actual field data that likely exhibits larger variability. Readers should consider this circumstance when viewing the results for any given model. Also, there is potential for nontrivial error propagation as uncertainty due to model error, and the use of predicted values would need to be accounted for. While quantification of this uncertainty source in this specific context is warranted, in more typical situations, model error contributes relatively small amounts of additional variance to population estimates of forest resources. Particularly, models fit to large datasets with robust goodness-of-fit statistics produce little additional uncertainty compared to sampling error (McRoberts and Westfall, 2014; Ståhl et al., 2014). Prediction systems relying on various interrelated models are subject to uncertainty from each component model as well as compounding error among model predictions (Robinson and Ek, 2000), which describes the situation of this study. In terms of the conversions of volume to biomass and subsequently biomass to carbon, an analysis of this approach for trees 2.54 cm DBH and larger showed increases in overall error due to the models being about 1 % for estimates of aboveground tree biomass and carbon (Westfall et al., 2025).

A limitation of this study is the presumption that the taper data properly reflect the allometry present in the FIA sapling data. In this analysis, the taper data were constrained to the species found in the FIA data, but there is no assurance that, e.g., the range of stand conditions was fully represented. A more rigorous approach would be to use taper data collected at the same locations as the FIA data. A second limitation of the study is the assumption that rescaling of sapling characteristics adequately portrays allometric relationships exhibited by seedlings. The results demonstrating the smooth transition between seedling and sapling tree size classes (Fig. 3) suggest this assumption is well-accommodated. However, further examination is warranted via actual measurement of individual seedling attributes $DGL_{seed}$, $HT_{seed}$, $V_{seed}$, $AGB_{seed}$, and $AGC_{seed}$. Collection of these data would also facilitate the lacking validation analyses for the models developed in this study. Further refinement of seedling: sapling relationships could also include quantification of competition, rotten and missing defects, and various forms of tree damage that may affect the characteristics of differently sized individuals. A challenge to accomplishment at a large spatial scale would not only be the cost to obtain samples and conduct destructive analyses, but also to cover a potentially wide range of forest conditions such as species composition, stand age, and various environmental gradients.

## 5. Conclusions

Due to the lack of individual seedling data over large areas, information from sapling-size trees (2.54 cm ≤ DBH <12.70 cm) was used to approximate allometric relationships between $HT_{seed}$ and $DGL_{seed}$, and subsequently those values served as predictors of seedling volume, biomass, and carbon attributes. Fundamentally, the methods relied on a series of models that transform observed relationships onto seedling size trees (DBH <2.54 cm) such that attribute values for individuals could be assigned. Although the volume, biomass, and carbon of seedlings are a relatively small component of all trees, quantification of these attributes allows for a more holistic view of their ecological importance beyond typical species-level number of stems assessments. Availability of this type of information from continuously remeasured permanent sample plots across a wide range of forest conditions may promote new investigative avenues for understanding seedling-specific roles in stand structure and development. A preferable alternative, albeit somewhat costly, would be to take individual stem measurements on seedling sized trees as part of standard inventory protocols.

## Declaration of competing 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.

## Acknowledgements

Credit is due to the associate editor and anonymous reviewers for their time commitment and expert commentary that considerably improved the manuscript.

## References

Annighöfer, P., Ameztegui, A., Ammer, C., Balandier, P., Bartsch, N., Bolte, A., Coll, L., Collet, C., Ewald, J., Frischbier, N., Gebereyesus, T., Haase, J., Hamm, T., Hirschfelder, B., Huth, F., Kändler, G., Kahl, A., Kawaletz, H., Kuehne, C., Lacointe, A., Lin, N., Löf, M., Malagoli, P., Marquier, A., Müller, S., Promberger, S., Provendier, D., Röhle, H., Sathornkich, J., Schall, P., Scherer-Lorenzen, M., Schröder, J., Seele, C., Weidig, J., Wirth, C., Wolf, H., Wollmerstädt, J., Mund, M., 2016. Species-specific and generic biomass equations for seedlings and saplings of European tree species. Eur. J. For. Res. 135, 313–329.

Avery, T.E., Burkhart, H.E., 2002. Forest Measurements, fifth ed. McGraw-Hill, New York.

Brose, P.H., Gottschalk, K.W., Horsley, S.B., Knopp, P.D., Kochenderfer, J.N., McGuinness, B.J., Miller, G.W., Ristau, T.E., Stoleson, S.H., Stout, S.L., 2008. Prescribing Regeneration Treatments for mixed-oak Forests in the Mid-atlantic Region. U.S. Department of Agriculture, Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-GTR-33.

Burrill, E.A., DiTommaso, A.M., Turner, J.A., Pugh, S.A., Christensen, G., Perry, C.J., Lepine, L.C., Walker, D.M., Conkling, B.L., 2024. The Forest Inventory and Analysis Database, FIADB user guides, volume database description (version 9.2). Nationwide Forest Inventory (NFI). U.S. Department of Agriculture, Forest Service. https://research.fs.usda.gov/understory/forest-inventory-and-analysis-database-user-guide-nfi. (Accessed 7 February 2025).

Canham, C.D., Murphy, L., 2017. The demography of tree species response to climate: sapling and canopy tree survival. Ecosphere 8 (2), e01701.

Chang-Yang, C.H., Needham, J., Lu, C.L., Hsieh, C.F., Sun, I.F., McMahon, S.M., 2021. Closing the life cycle of forest trees: the difficult dynamics of seedling-to-sapling transitions in a subtropical rainforest. J. Ecol. 109 (7), 2705–2716.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.P., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145, 87–99.

Dixon, G.E., 2002. Essential FVS: a User's Guide to the Forest Vegetation. U.S. Department of Agriculture, Forest Service, Forest Management Service Center (revised 16 February 2024).

Eerikäinen, K., Miina, J., Valkonen, S., 2007. Models for the regeneration establishment and the development of established seedlings in uneven-aged, Norway spruce dominated forest stands of southern Finland. For. Ecol. Manag. 242 (2–3), 444–461.
Hillard, S.C., Morin, R.S., Westfall, J.A., Butler, B.J., Crocker, S.J., Nelson, M.D., Walters, B.F., Luppold, W.G., Riemann, R.I., Woodall, C.W., Albright, T.A., Hemmer, B.J., Garner, J.D., 2022. Minnesota Forests 2018: Summary Report. U.S. Department of Agriculture, Forest Service, Northern Research Station.
Hoover, C.M., Smith, J.E., 2021. Current aboveground live tree carbon stocks and annual net change in forests of conterminous United States. Carb. Bal. Manag. 16 (1), 17.
Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49 (1), 12–35.
Johnson, K.D., Domke, G.M., Russell, M.B., Walters, B., Hom, J., Peduzzi, A., Birdsey, R., Dolan, K., Huang, W., 2017. Estimating aboveground live understory vegetation carbon in the United States. Environ. Res. Let. 12 (12), 125010.
Kershaw Jr, J.A., Ducey, M.J., Beers, T.W., Husch, B., 2016. Forest Mensuration. John Wiley & Sons, New York.
Kozak, A., Munro, D.D., Smith, J.H.G., 1969. Taper functions and their application in forest inventory. For. Chron. 45 (4), 278–283.
Max, T.A., Burkhart, H.E., 1976. Segmented polynomial regression applied to taper equations. For. Sci. 22 (3), 283–289.
McRoberts, R.E., Westfall, J.A., 2014. Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. 60 (1), 34–42.
McTague, J.P., Weiskittel, A., 2021. Evolution, history, and use of stem taper equations: a review of their development, application, and implementation. Can. J. For. Res. 51 (2), 210–235.
McWilliams, W.H., Westfall, J.A., Brose, P.H., Dey, D.C., Hatfield, M., Johnson, K., Laustsen, K.M., Lehman, S.L., Morin, R.S., Nelson, M.D., Ristau, T.E., Royo, A.A., Stout, S.L., Willard, T., Woodall, C.W., 2015. A regeneration indicator for Forest Inventory and Analysis. History, Sampling, Estimation, Analytics, and Potential Use in the Midwest and Northeast United States. U.S. Department of Agriculture, Forest Service, Northern Research Station.
Menéndez-Miguélez, M., Calama, R., Del Río, M., Madrigal, G., López-Senespleda, E., Pardos, M., Ruiz-Peinado, R., 2022. Species-specific and generalized biomass models for estimating carbon stocks of young reforestations. Biomass Bioenergy 161, 106453.
Muukkonen, P., Mäkipää, R., Laiho, R., Minkkinen, K., Vasander, H., Finér, L., 2006. Relationship between biomass and percentage cover in understorey vegetation of boreal coniferous forests. Silva Fenn. 40 (2), 231–245.
Nelson, A.S., Weiskittel, A.R., Wagner, R.G., Saunders, M.R., 2014. Development and evaluation of aboveground small tree biomass models for naturally regenerated and planted species in eastern Maine, USA. Biomass Bioenergy 68, 215–227.
Pati, P.K., Kaushik, P., Khan, M.L., Khare, P.K., 2022. Allometric equations for biomass and carbon stock estimation of small diameter woody species from tropical dry deciduous forests: support to REDD+. Trees For. People 9, 100289.
Radtke, P., Walker, D., Frank, J., Weiskittel, A., MacFarlane, D., Affleck, D., Temesgen, H., Poudel, K., Zhao, D., Auty, D., Sánchez Meador, A., Shaw, J., Gray, A., Westfall, J., Coulston, J., 2023. LegacyTreeData v2. Virginia Polytechnic Institute and State University. https://doi.org/10.7294/22582432.
Robinson, A.P., Ek, A.R., 2000. The consequences of hierarchy for modeling in forest ecosystems. Can. J. For. Res. 30 (12), 1837–1846.
Singh, V., Tewari, A., Kushwaha, S.P., Dadhwal, V.K., 2011. Formulating allometric equations for estimating biomass and carbon stock in small diameter trees. For. Ecol. Manag. 261 (11), 1945–1949.
Smith, J.E., Heath, L.S., Skog, K.E., Birdsey, R.A., 2006. Methods for Calculating Forest Ecosystem and Harvested Carbon with Standard Estimates for Forest Types of the United States. U.S. Department of Agriculture, Forest Service, Northeastern Research Station.
Smith, J.E., Heath, L.S., Hoover, C.M., 2013. Carbon factors and models for forest carbon estimates for the 2005–2011 national greenhouse gas inventories of the United States. For. Ecol. Manag. 307, 7–19.
Ståhl, G., Heikkinen, J., Petersson, H., Repola, J., Holm, S., 2014. Sample-based estimation of greenhouse gas emissions from forests—a new approach to account for both sampling and model errors. For. Sci. 60 (1), 3–13.
Thippawan, S., Chowtiwuttakorn, K., Pongpattananurak, N., Kraichak, E., 2023. Allometric models to estimate the biomass of tree seedlings from dry evergreen forest in Thailand. Forests 14 (4), 725.
Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., Gabler, K., Schadauer, K., Vidal, C., Lanz, A., Ståhl, G., Cienciala, E., 2010. National forest inventories: pathways for common reporting. Eur. Sci. Found. 1, 541–553.
Waller, D.M., Riege, D.A., Alverson, W.S., 2024. The regeneration ratio: combining seedling growth and mortality data to predict regeneration success. For. Ecol. Manag. 556, 121737.
Weiskittel, A.R., Hann, D.W., Kershaw Jr, J.A., Vanclay, J.K., 2011. Forest Growth and Yield Modeling. John Wiley & Sons, New York.
Westfall, J.A., McWilliams, W.H., 2017. Predicting tree-seedling height distributions using subcontinental-scale forest inventory data. For. Ecol. Manag. 400, 332–338.
Westfall, J.A., Coulston, J.W., Patterson, P.L., Scott, C.T., 2022. Foundational documentation. In: Westfall, J.A., Coulston, J.W., Moisen, G.G., Andersen, H.-E. (Eds.), Comps. Sampling and Estimation Documentation for the Enhanced Forest Inventory and Analysis Program: 2022. U.S. Department of Agriculture, Forest Service, Northern Research Station, pp. 4–21.
Westfall, J.A., Coulston, J.W., Gray, A.N., Shaw, J.D., Radtke, P.J., Walker, D.M., Weiskittel, A.R., MacFarlane, D.W., Affleck, D.L.R., Zhao, D., Temesgen, H., Poudel, K.P., Frank, J.M., Prisley, S.P., Wang, Y., Sánchez Meador, A.J., Auty, D., Domke, G.M., 2024. A national-scale Tree Volume, Biomass, and Carbon Modeling System for the United States. U.S. Department of Agriculture, Forest Service.
Westfall, J.A., Radtke, P.J., Walker, D.M., Coulston, J.W., 2025. Model error propagation in a compatible tree volume, biomass, and carbon prediction system. Carb. Bal. Manag. 20 (1), 1–10.
Woodall, C.W., Heath, L.S., Domke, G.M., Nichols, M.C., 2011. Methods and Equations for Estimating Aboveground Volume, Biomass, and Carbon for Trees in the US Forest Inventory, 2010. U.S. Department of Agriculture, Forest Service, Northern Research Station.
Zhang, W.P., Zhao, L., Larjavaara, M., Morris, E.C., Sterck, F.J., Wang, G.X., 2020. Height-diameter allometric relationships for seedlings and trees across China. Acta Oecol. 108, 103621.