NRC Research Press

ARTICLE

Comparison of forest age estimators using *k*-tree, fixed-radius, and variable-radius plot sampling

Brent D. Burch and Andrew J. Sánchez Meador

Abstract: Quantifying the age characteristics of a forest can provide valuable information about the forest's impact on the environment. For instance, the age of a forest can affect the ecosystem's carbon exchange, soil enzyme activity, and biodiversity. In this paper, we investigate the use of different sampling methods to estimate the age characteristics of three simulated ponderosa pine (*Pinus ponderosa* Dougl. ex P. Lawson & C. Lawson) forests having different spatial and age patterns. This includes estimating the mean tree age and the age-class distribution of the trees in the forest. The trees in the sample are selected using *k*-tree sampling, fixed-radius plot sampling, or variable-radius plot sampling, and we compare the properties of the resulting estimators via design-based and model-based approaches. Analyses of the different sampling methods applied to the three forests suggest that the estimator associated with *k*-tree sampling, with the addition of a few extra trees per plot, is feasible for forests having a spatially mosaic or random spatial pattern. The estimator associated with fixed-radius plot sampling performed well for the forest having a clustered spatial pattern.

Key words: design-based inference, model-based inference, nearest neighbor sampling, mean tree age, tree age distribution.

Résumé : L'évaluation des caractéristiques d'âge d'une forêt peut fournir des informations précieuses sur l'impact que peut avoir la forêt sur l'environnement. Par exemple, l'âge d'une forêt peut avoir un impact sur les échanges de carbone de l'écosystème, l'activité enzymatique du sol et la biodiversité. Dans cet article, nous étudions l'utilisation de différentes méthodes d'échantillonnage pour estimer les caractéristiques d'âge de trois forêts simulées de pin ponderosa (*Pinus ponderosa* Dougl. ex P. Lawson & C. Lawson) avec des configurations spatiales et des distributions d'âges différentes. Cela comprend l'estimation de l'âge moyen des arbres et l'estimation de la distribution des classes d'âge des arbres de la forêt. Les arbres de l'échantillon sont sélectionnés en utilisant l'échantillonnage de *k* arbres voisins (*k*-arbre), l'échantillonnage par placette à rayon fixe ou l'échantillonnage par placette à rayon variable. Nous comparons les propriétés des estimateurs qui en résultent par des approches fondées sur le plan d'expérience ou sur un modèle. L'analyse des différentes méthodes d'échantillonnage appliquées aux trois forêts indique que l'estimateur associé à l'échantillonnage *k*-arbre, avec l'ajout de quelques arbres supplémentaires par parcelle, est adéquat pour les forêts ayant une mosaïque spatiale ou une organisation spatiale aléatoire. L'estimateur associé à l'échantillonnage par placette à rayon fixe s'est bien comporté pour la forêt avec une organisation spatiale en grappes. [Traduit par la Rédaction]

Mots-clés : inférence fondée sur le plan d'expérience, inférence fondée sur un modèle, échantillonnage du plus proche voisin, âge moyen des arbres, distribution des classes d'âge des arbres.

1. Introduction

Forests are in a constant state of change, and tree demographics, particularly age, can be used to assess forest dynamics and to comprehend key forest processes and functions. For example, Vilén et al. (2012) stated that patterns of net carbon exchange are strongly affected by the age of a forest. Lucas-Borja et al. (2016) investigated the effects of stand age on microbiological soil properties and concluded that stand age influenced differences in soil attributes (e.g., pH, NH_4^+, etc.). Lie et al. (2009) found that large, old trees were conveyors of biodiversity in boreal forests. Rozas (2015) analyzed tree age and size and intertree competition in a dendroclimatic study involving individual-tree sensitivity to climate changes to provide insights into the effects of past management history.

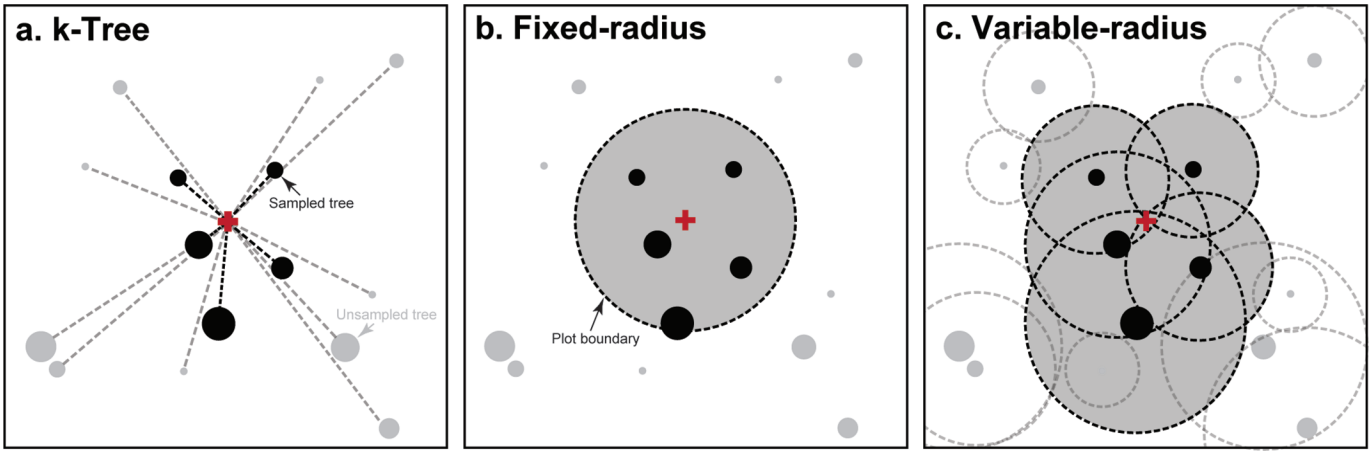
Determining the age characteristics of a forest, however, can be a challenging task and is especially difficult to do in an unbiased and yet highly efficient (i.e., cost- and time-effective) manner. While commonly measured tree attributes such as diameter at breast height (dbh; measured at 1.37 m above ground level), total height, leaf area index, crown shape, and location can be easily determined or estimated using field or remote sensing techniques, the methods most commonly used for determining tree age, and thus forest age, are quite limited. In some cases, the age of young trees can be estimated by counting whorls; however, the preferred field method of measuring tree age is by counting annual growth rings obtained using an increment borer (Avery and Burkhart 2002), and the most accurate method is dendrochronological crossdating of ring series (Stokes and Smiley 1996). In general, coring every tree in a forest is not feasible, and in practice, a sample of tree ages is obtained.

In the literature, numerous sampling methods have been used to study a variety of forest structural and compositional attributes, but few examine forest age via plot sampling and none explicitly compares sampling designs and quality of estimators. For example, Schreuder et al. (1987) compared three sampling

Received 20 March 2018. Accepted 2 May 2018.

B.D. Burch. Department of Mathematics and Statistics, Northern Arizona University, P.O. Box 5717, Flagstaff, AZ 86011, USA.
A.J. Sánchez Meador. School of Forestry, Northern Arizona University, P.O. Box 15018, Flagstaff, AZ 86011, USA.
Corresponding author: Andrew J. Sánchez Meador (email: Andrew.SanchezMeador@nau.edu).

Fig. 1. Schematic diagrams of the three plots associated with the sampling methods examined in this study: (*a*) *k*-tree, (*b*) fixed-radius, and (*c*) variable-radius plot sampling designs. [Colour online.]



methods to estimate stand parameters such as the total number of trees and the average tree dbh in a simulation study of tropical forests. Lessard et al. (1994) compared the statistical properties of estimators (mean board foot volume per acre and basal area per acre) obtained from distance sampling with point and plot sampling. Furthermore, Binkley (2008) examined the age distribution of aspen in Rocky Mountain National Park using a sampling approach that averaged variable-radius plot measurements located along a plot perimeter to provide a single estimate associated with each random plot, and Yocom-Kent et al. (2015) used *k*-tree to estimate age attributes of mixed-conifer and aspen forests in the North Rim area of Grand Canyon National Park, yet neither study compared alternative plot designs or suggested alternative methods.

In this paper, we estimate the ages of trees in a forest and compare the properties of mean tree age and age-class distribution estimators using *k*-tree sampling, fixed-radius plot sampling, and variable-radius plot sampling. Details concerning the formulaic expressions of the estimators and a description of the underlying sampling units are discussed. Both design-based and model-based approaches are considered, and the bias and variance of the estimators are used to measure the quality of the estimation techniques. We examine the performance of these estimators in three simulated ponderosa pine (*Pinus ponderosa* Dougl. ex P. Lawson & C. Lawson) forests having various spatial and age patterns. In scenarios where the overall cost of sampling is fixed, optimal values of *k* are discussed.

2. Methods

2.1. Design-based sampling

Let N be the total number of plots in the forest. Furthermore, let τ_{ntrees} denote the total number of trees in the forest so that τ_{ntrees}/N is the number of trees per plot in the forest. For example, a square grid of 2500 plots each of size 20 m × 20 m that covers a 100 ha forest of 30 000 trees results in 12 trees per plot. τ_{ntrees}/N also represents the average number of trees that must be aged for each plot selected. Suppose that n plots are randomly selected from the N plots available. Note that the actual plots used depend on the sampling method employed and the aforementioned square plots merely serve as a frame of reference. See Fig. 1 for schematic diagrams of the three different plots considered in this paper. The objective is to compare the estimators of forest age characteristics using *k*-tree, fixed-radius plot, and variable-radius plot sampling where each method is based on roughly $n\tau_{ntrees}/N$ trees in the sample.

2.1.1. *k*-tree sampling

The *k*-nearest trees to the center of plot i are measured if plot i is included in the sample. Using this nearest neighbor sampling approach, the distance to the *k*th-nearest tree is used to form the estimator of mean tree age. Equivalently, let A_i denote the effective area of plot i relative to the area of the forest under study with a radius r_i equal to the distance from the center of plot i to the *k*th-nearest tree. Let Y_{ij} be the age of the jth-nearest tree to the center of plot i, where $j \le k$. Then the ratio estimator of the mean tree age is

$$\hat{\mu}_1 = \frac{c\frac{1}{n}\sum_{i=1}^{n}\frac{1}{A_i}\left(\sum_{j=1}^{k}Y_{ij}\right)}{c\frac{1}{n}\sum_{i=1}^{n}\frac{k}{A_i}} = \frac{\sum_{i=1}^{n}\frac{1}{r_i^2}\left(\sum_{j=1}^{k}Y_{ij}\right)}{k\sum_{i=1}^{n}\frac{1}{r_i^2}} \quad (1)$$

where n is the number of plots in the sample, c is a bias correction term, and the constant k is determined in a manner that allows for a fair comparison of the competing estimators of mean tree age. For example, consider $k = \tau_{ntrees}/N$, where τ_{ntrees}/N is rounded to the nearest integer.

Let a denote a specific age-class and $n$class denote the total number of age classes. Then estimators of the proportion of trees in age-class a using *k*-tree sampling are given by

$$\hat{p}_{1a} = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k}I_{ij}\right)}{nk}, \quad a = 1, \ldots, n\text{class} \quad (2)$$

where I_{ij} takes on the value 1 if the jth-nearest tree to the center of plot i is in age-class a and 0 otherwise. Note that $\sum_{a=1}^{n\text{class}}\hat{p}_{1a} = 1$. As an example, $n$class = 14 if the width of an age-class is 25 years, where 0 to 25 years represents the first age-class, and the age of the youngest tree is 10 years and the age of the oldest tree is 340 years.

2.1.2. Fixed-radius plot sampling

The sampling units employed are circular where, for example, each plot has a radius equal to $\sqrt{400/\pi} \approx 11.3$ m in reference to a 100 ha forest. These 2500 circular plots overlay the forest, and a simple random sample of n circular plots is obtained. If a specific plot is included in the sample, then age measurements are made

on each tree in the plot. Let Y_{ij} denote the age of tree j in plot i. The ratio estimator of the mean tree age is

$$\hat{\mu}_2 = \frac{N\left[\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i} Y_{ij}\right)\right]}{N\left(\frac{1}{n}\sum_{i=1}^{n} N_i\right)} = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i} Y_{ij}\right)}{\sum_{i=1}^{n} N_i} \quad (3)$$

where N_i is the number of trees in plot i. Unlike k-tree sampling, the number of trees per plot in fixed-radius plot sampling is not constant.

The estimator of the proportion of trees in age-class a, denoted by $\hat{p}_{2a}$, is calculated in a similar manner as $\hat{p}_{1a}$, where, in this case, I_{ij} takes on the value 1 if tree j in plot i is in the age-class a and 0 otherwise. It follows that

$$\hat{p}_{2a} = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i} I_{ij}\right)}{\sum_{i=1}^{n} N_i}, \quad a = 1, \ldots, n\text{class} \quad (4)$$

2.1.3. *Variable-radius plot sampling*

A single-point sample is used for each plot, and a tree is counted as being "in" or "out" depending on its size and distance from the point based on a calibrated wedge prism. Trees were selected with probability proportional to size so that the expected number of "in" trees per prism point is τ_{ntrees}/N. In practice, a tree having dbh is tallied as "in" if it is within dbh/$(2\sqrt{\text{BAF}})$ metres of the prism point, where BAF is the associated basal area factor. Let Y_{ij} denote the age of the jth "in" tree with respect to prism point i. Then the ratio estimator of the mean tree age is

$$\hat{\mu}_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i^*} \frac{Y_{ij}}{\pi_{ij}}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i^*} \frac{1}{\pi_{ij}}\right)} = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i^*} \frac{Y_{ij}}{\text{dbh}_{ij}^2}\right)}{\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i^*} \frac{1}{\text{dbh}_{ij}^2}\right)} \quad (5)$$

where n is the number of prism points (plots) in the sample, N_i^* is the number of "in" trees with respect to prism point i, π_{ij} is the probability that tree j is "in" with respect to prism point i, and dbh_{ij} is the dbh of tree j that is "in" with respect to prism point i.

Using this point sampling approach, the estimator of the proportion of trees in age-class a, denoted by $\hat{p}_{3a}$, is calculated in a similar manner as $\hat{p}_{1a}$ and $\hat{p}_{2a}$, where, in this case, I_{ij} takes on the value 1 if the jth "in" tree with respect to prism point i is in age-class a and 0 otherwise. Specifically,

$$\hat{p}_{3a} = \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i^*} I_{ij}\right)}{\sum_{i=1}^{n} N_i^*}, \quad a = 1, \ldots, n\text{class} \quad (6)$$

Forest edge effects are accounted for by employing mirage boundary corrections. See Gregoire (1982), Ducey et al. (2001), and Lynch (2012) for details. Specifically, when tree boles were near the forest boundary, these trees were replicated equidistance outside the boundary on a line perpendicular to the boundary. The resulting mirage tree has the same characteristics as the actual tree, and trees occurring in the corner plots of the forest region have three mirage trees. The mirage boundary corrections are made when computing the values of the estimators.

2.2. Simulated forests having various spatial and age patterns

The availability of large, censused areas where tree spatial pattern and age composition are known is extremely limited, and nonexistent at the stand level (~50–100 ha). To explore the quality of the three sampling methods at stand levels, simulated, single-species forests were created using individual-tree segmentation algorithms allied to subsets of research-grade airborne light detection and ranging (LiDAR) acquisition data for 100 ha regions of typical forest spatial pattern and tree age composition.

In brief, potential 100 ha sites were examined for which discrete return airborne LiDAR data were acquired by Watershed Sciences Inc. (current name: Quantum Spatial Inc.) between 2012–2016 using a Leica ALS50 Phase II and ALS60 laser systems mounted on fixed-wing aircraft on National Forest System lands in northern Arizona. Aircraft were flown at altitudes between 900 and 2000 m above the ground following topography. Data were acquired using an opposing flight line side-lap of ≥50% and a sensor scan angle ±14° from nadir. Average point densities (points·m^{-2}) ranged from 9.4 to 14.9. The vendor post-processed LiDAR data using automated methods in proprietary software (TerraScan) coupled with manual methods to identify ground points for development of the digital terrain model (DTM) with vertical DTM accuracies of approximately 15 cm at a 95% confidence level. LiDAR point cloud data were normalized (i.e., converted from elevation to height above the ground) and individual-tree location and size information were obtained using the Li et al. (2012) individual-tree segmentation algorithm using the lidR package (version 1.2.0; Roussel and Auty 2017) in R (version 3.1.0; R Core Team 2017).

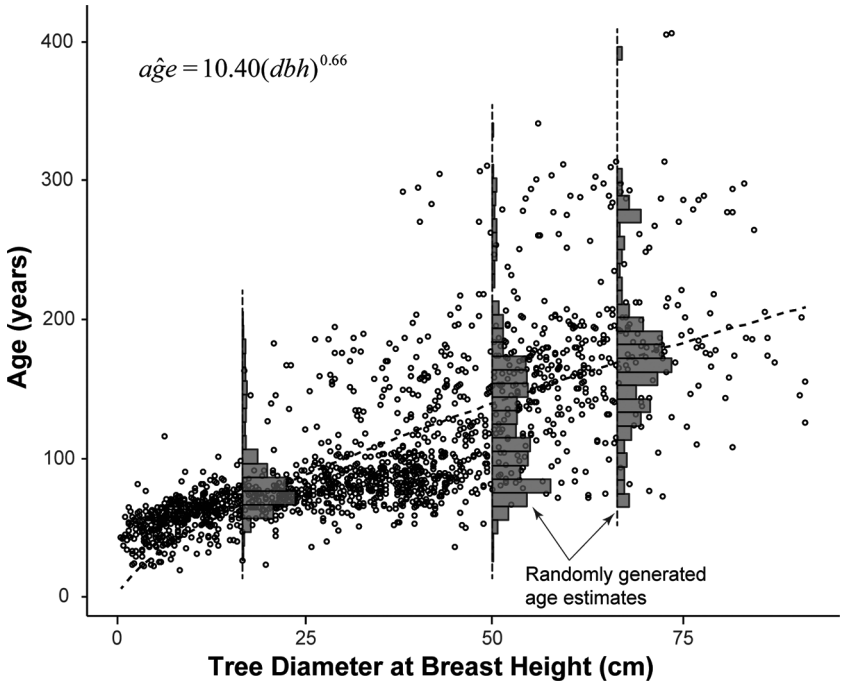
Because tree age data could not be obtained from the processed LiDAR-derived tree lists, we developed and used a nonlinear regression model of the form age = $a(\text{dbh})^b\varepsilon$, with parameters a and b and a multiplicative error term ε to obtain estimates of tree age. See Binkley (2008), Loewenstein et al. (2000), and Veblen (1986) for additional examples of models of age–size relationships for various tree species. The model was developed using ponderosa pine tree size and age data collected by Sánchez Meador et al. (2010) on the Coconino National Forest in the vicinity of Flagstaff, Arizona, and the fitted equation $\widehat{\text{age}} = 10.40(\text{dbh})^{0.66}$ was obtained for the 1589 trees in the study. Moreover, for a fixed value of dbh, we assumed that age was gamma-distributed with mean $10.4(\text{dbh})^{0.66}$ and standard deviation $\sqrt{10.4}(\text{dbh})^{0.66}$ and randomly generated age estimates from the appropriate distribution. Figure 2 displays the relationship between dbh and age for the 1589 trees, as well as sample histograms for randomly generated age distributions at fixed values of dbh.

The age–size model was then used to generate tree ages in three forests having varying spatial patterns. Specifically, age ~ Gamma(10.40, $(\text{dbh})^{0.66}$), where 10.40 is the rate parameter and $(\text{dbh})^{0.66}$ is the scale parameter of the gamma distribution for trees in spatially mosaic, random, and clustered forests. As dbh increases, the model indicates that the variability in age also increases. Figure 3 illustrates the resulting three simulated forest conditions used in the study, which are described briefly in the following sections.

2.2.1. *Two-aged mosaic forest*

The first simulated forest is generally composed of two age cohorts — remnant older trees (150+ years old) and younger trees originating after the Euro-American settlement (circa 1876; see Savage et al. 1996) — in a spatially mosaic pattern. The site is located on the Fort Valley Experimental Forest within the Coconino National Forest (35°15.94′N, 111°44.99′W). *Pinus ponderosa* is

Fig. 2. Modeled relationship between tree diameter at breast height (1.37 m above the ground) and age for 1589 ponderosa pine trees in northern Arizona (from Sánchez Meador et al. 2010). The histograms provide pictorial summaries of randomly generated age estimates for a given size tree.



the only tree species on the site, and the understory vegetation is predominantly perennial bunchgrasses. The site is centered on the 2.59 ha study site that Sánchez Meador et al. (2009) used to examine change in tree spatial pattern following selective harvest and regeneration dynamics.

2.2.2. All-aged random forest

The second simulated forest represents a functioning, intact ponderosa pine forest dominated by large, old trees with over 70% of the trees predating the year 1800 in origin and 20% predating 1700. The site is located within Grand Canyon National Park (36°16.36′N, 112°22.51′W) and is dominated by ponderosa pine with Gambel oak (*Quercus gambellii* Nutt.) and New Mexican locust (*Robinia neomexicana* Gray) trees occurring in drainages, and the understory is dominated by forbs and perennial grasses. The site is centered on the 315 ha study site that Fulé et al. (2003) used to examine change in fire regimes and suggested that they may represent "nearly natural" conditions.

2.2.3. Young-aged clustered forest

The third and last simulated forest is dominated by younger trees (<150 years old) in an aggregated or clustered spatial pattern, composed of many small groups (<10 distinct tree groups of approximately 20 trees per group) with many scattered, randomly arranged individuals throughout. The site is located on the North Kaibab Ranger District of the Kaibab National Forest Valley Experimental (36°31.85′N, 112°21.76′W) and consists of pure ponderosa pine stands with scattered groups of Gambel oak and aspen (*Populus tremuloides* Michx.) present but rare. The site is centered within the 2000 ha study site that Tuten et al. (2015) used to assess fine-scale forest spatial patterns and evaluate two common forest restoration approaches.

3. Results

3.1. Simulation outcomes using the design-based approach

For each simulated forest, a simple random sample without replacement of n plots (or units) is selected from the N = 2500 possible plots. Following the selection of plots, we compute $\hat{\mu}_1$ and $\hat{p}_{1a}$, or $\hat{\mu}_2$ and $\hat{p}_{2a}$, or $\hat{\mu}_3$ and $\hat{p}_{3a}$ using the methods described earlier. For comparison purposes, we considered a sample of size n = 25 plots (or units) for each sampling method so that 1% of the plots (and thus trees) are sampled. Although what constitutes a plot (or unit) differs for the three sampling approaches, the total number of trees sampled remained roughly constant.

In the two-aged mosaic forest, there are 21 285/2 500 = 8.51 trees per plot, so roughly 8.51(25) ≈ 213 trees are measured using the fixed-radius plot and variable-radius plot methods. Using k-tree sampling with k = 9, 9(25) = 225 trees are measured. In the all-aged random forest, there are 17 834/2 500 = 7.13 trees per plot, so roughly 7.13(25) ≈ 178 trees are measured using the fixed-radius plot and variable-radius plot methods. Using k-tree sampling with k = 7, 7(25) = 175 trees are measured. In the young-aged clustered forest, there are 56 782/2 500 = 22.71 trees per plot, so roughly 22.71(25) ≈ 568 trees are measured using the fixed-radius plot and variable-radius plot methods. Using k-tree sampling with k = 23, 23(25) = 575 trees are measured.

For each simulated forest, the properties of the ratio estimators of mean tree age, as well as the estimators of the age-class distribution, were computed by repeating the sampling process 10 000 times. Because ratio estimators are not unbiased, both the bias and standard deviation of the estimator are taken into consideration when assessing the quality of the estimator. This is accomplished by computing the square root of the mean square error (MSE) of an estimator, which is defined as

$$\sqrt{\text{MSE}(\hat{\mu})} = \sqrt{\text{Bias}^2(\hat{\mu}) + \text{Var}(\hat{\mu})} \quad (7)$$

For a ratio estimator, Bias^2 is often small relative to its variance. For unbiased estimators, $\sqrt{\text{MSE}(\hat{\mu})} = \sqrt{\text{Var}(\hat{\mu})}$, which simplifies to the standard deviation of the estimator.

3.1.1. Two-aged mosaic forest

Table 1 provides a summary of the bias, standard deviation, and $\sqrt{\text{MSE}(\hat{\mu})}$ of the ratio estimators of mean tree age. While k = 9 for this forest, additional values of k are included to

Fig. 3. Simulated forest conditions in three case study forest types: (*a*) two-aged mosaic, (*b*) all-aged random, and (*c*) young-aged clustered. Each forest represents varying spatial patterns, and the age and size distributions are shown with summary statistics (tree density, basal area, mean tree diameter at breast height (dbh), mean tree age, and Clark and Evans' (1954) *R*).

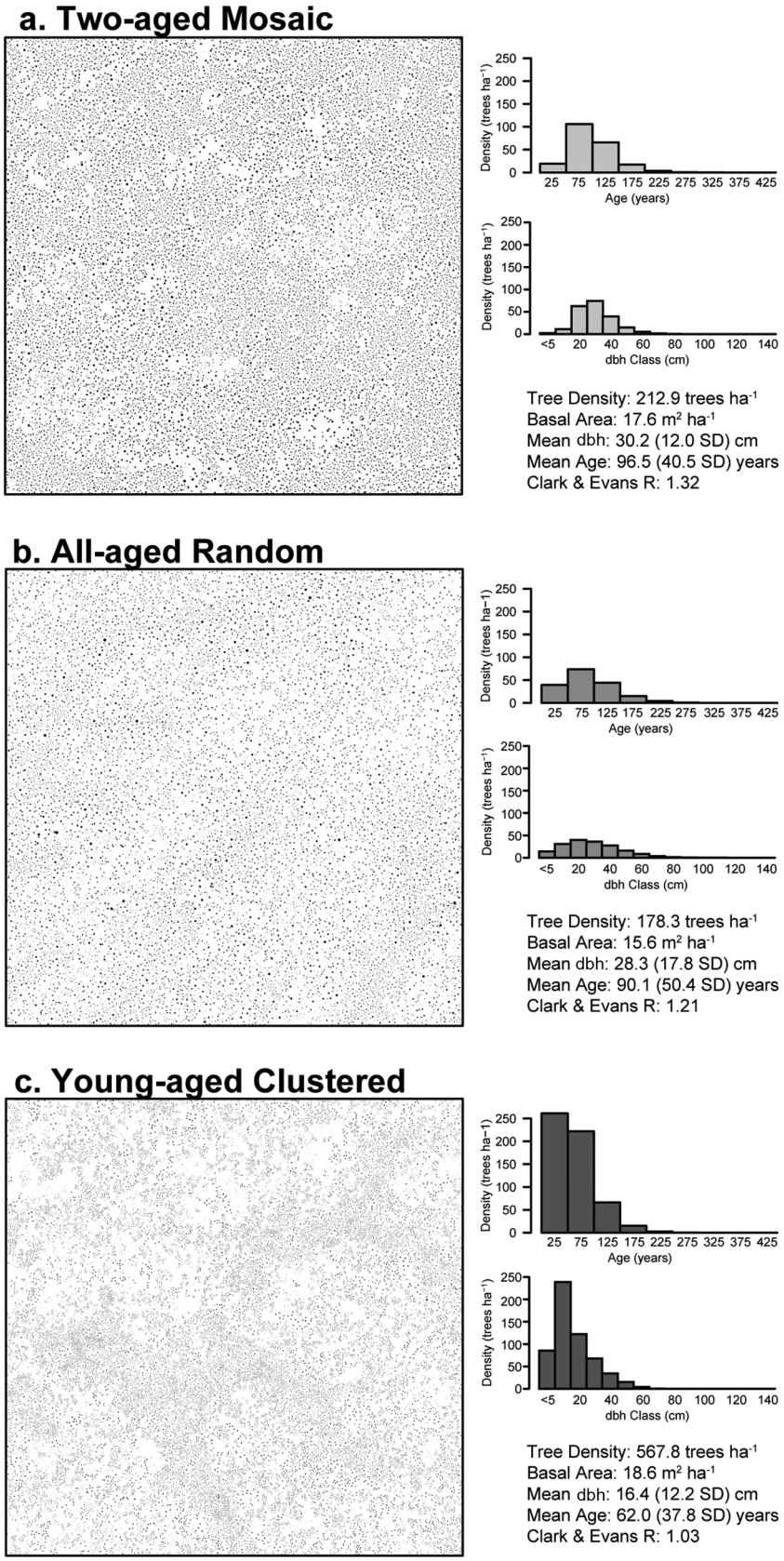


Table 1. Properties of the *k*-tree ($\hat{\mu}_1$; using various values of *k*), fixed-radius ($\hat{\mu}_2$), and variable-radius ($\hat{\mu}_3$) sampling estimators.

Estimator	Sampling estimator	No. of trees per plot	Bias	Standard deviation	$\sqrt{\text{MSE}(\hat{\mu})}$
Two-aged mosaic forest					
$\hat{\mu}_1$	*k*-tree	9	−0.476	3.205	3.240
$\hat{\mu}_1$	*k*-tree	10	−0.455	3.052	3.085
$\hat{\mu}_1$	*k*-tree	11	−0.473	2.976	3.014
$\hat{\mu}_1$	*k*-tree	12	−0.397	2.876	2.903
$\hat{\mu}_2$	Fixed-radius	8.5*	0.177	3.190	3.195
$\hat{\mu}_3$	Variable-radius	8.5*	0.830	5.037	5.105
All-aged random forest					
$\hat{\mu}_1$	*k*-tree	7	−1.573	4.620	4.880
$\hat{\mu}_1$	*k*-tree	8	−1.309	4.316	4.510
$\hat{\mu}_1$	*k*-tree	9	−1.056	4.119	4.252
$\hat{\mu}_1$	*k*-tree	10	−1.027	3.957	4.088
$\hat{\mu}_2$	Fixed-radius	7.1*	−0.207	4.400	4.405
$\hat{\mu}_3$	Variable-radius	7.7*	6.010	11.333	12.828
Young-aged clustered forest					
$\hat{\mu}_1$	*k*-tree	23	−0.744	4.406	4.469
$\hat{\mu}_1$	*k*-tree	31	−0.589	4.121	4.163
$\hat{\mu}_1$	*k*-tree	32	−0.553	4.077	4.115
$\hat{\mu}_1$	*k*-tree	33	−0.542	4.042	4.079
$\hat{\mu}_2$	Fixed-radius	22.7*	0.405	4.113	4.133
$\hat{\mu}_3$	Variable-radius	22.7*	1.285	7.231	7.344

Note: Bias, standard deviation, and $\sqrt{\text{MSE}(\hat{\mu})}$ are measured in years. Results for the fixed-radius plot sampling estimator $\hat{\mu}_2$ and the variable-radius plot estimator $\hat{\mu}_3$ are included for comparison purposes. Calculations are based on 10 000 repeated samples of 25 plots from the three case study forest types.

*On average.

determine when the *k*-tree sampling estimator is favored over both of the other estimators. For instance, if $k = 10$, then $\sqrt{\text{MSE}(\hat{\mu}_1)} < \sqrt{\text{MSE}(\hat{\mu}_2)}$. Also from Table 1, it is evident that the estimator based on variable-radius plot sampling exhibited a greater degree of bias and variability than the other two estimators. Both the numerator and denominator of $\hat{\mu}_3$ depend heavily on dbh, and because the variation in dbh was relatively large, the estimators in the numerator and denominator of $\hat{\mu}_3$ exhibited large variances. This, in turn, resulted in a comparably large value of $\sqrt{\text{MSE}(\hat{\mu}_3)}$. Although $\hat{\mu}_3$ is often a viable estimator of tree characteristics that depend on dbh^2 such as basal area and volume, it is not recommended for tree age studies where selection bias of larger trees may result in an unintended bias in mean forest age and corresponding age distribution estimates.

Figure 4 displays the estimators of the proportion of trees in each age class (i.e., age distributions) using the three sampling methods. Included in the displays are standard deviation bars for each proportion, and the solid squares indicate the true values (from the simulated forest) of the proportion of trees in each age class. Plots of $\hat{p}_{1a}$, $\hat{p}_{2a}$, and $\hat{p}_{3a}$ suggest that the *k*-tree and fixed-radius plot sampling approaches resulted in estimators that approximately match the true age-class distribution. However, as expected, the variable-radius plot approach drastically underestimated the proportion of young trees and severely overestimated the proportion of old trees. In other words, the plot of the mean values of $\hat{p}_{3a}$ for each age class, based on the 10 000 simulations, was shifted to the right and thus did not coincide with the true age-class distribution. The standard errors of the three estimators of the age-class distribution appeared to be similar to one another for each age class.

3.1.2. All-aged random forest

Table 1 provides a summary of the bias, standard deviation, and $\sqrt{\text{MSE}(\hat{\mu})}$ of the ratio estimators of mean tree age. While $k = 7$ for this forest for all three sampling approaches to acquire about the same number of trees, Table 1 indicates that $\sqrt{\text{MSE}(\hat{\mu}_1)} < \sqrt{\text{MSE}(\hat{\mu}_2)}$ when $k = 9$. As in the previous forest, $\hat{\mu}_3$ is not recommended when it comes to estimating mean tree age. Figure 4 displays plots of $\hat{p}_{1a}$, $\hat{p}_{2a}$, and $\hat{p}_{3a}$, where the *k*-tree and fixed-radius plot estimators of age-class distribution approximately match the true age-class distribution, but $\hat{p}_{3a}$ underestimates the proportion of young trees and severely overestimates the proportion of old trees.

3.1.3. Young-aged clustered forest

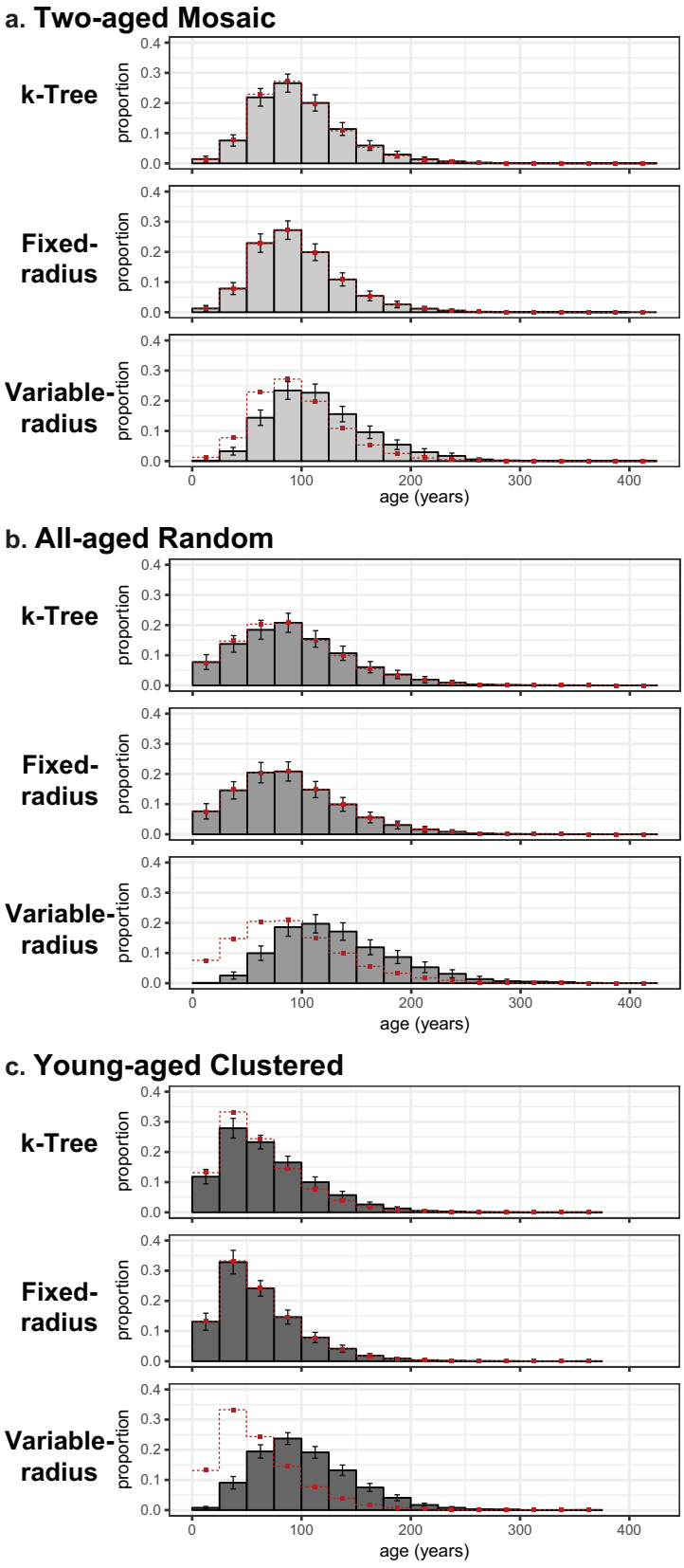
Table 1 provides a summary of the bias, standard deviation, and $\sqrt{\text{MSE}(\hat{\mu})}$ of the ratio estimators of mean tree age. While $k = 23$ for this forest so that all three sampling methods have about the same number of sampled trees, Table 1 indicates that $\sqrt{\text{MSE}(\hat{\mu}_1)} < \sqrt{\text{MSE}(\hat{\mu}_2)}$ occurs only when $k \geq 32$. That is, at least nine extra trees per plot must be obtained for the *k*-tree sampling estimator to be competitive with the fixed-radius plot estimator of mean tree age.

This may be viewed is an inordinate number of additional trees per plot so that the *k*-tree sampling estimator may not be a viable estimator in this case. Figure 4 containing plots of $\hat{p}_{1a}$, $\hat{p}_{2a}$, and $\hat{p}_{3a}$ confirms this result as well. Whereas $\hat{p}_{3a}$ clearly underestimates the proportion of young trees and severely overestimates the proportion of old trees, $\hat{p}_{1a}$ does so in a less severe manner. Only the fixed-radius plot estimator of the age-class distribution appears to be acceptable in this particular application.

3.2. Sampling outcomes using the model-based approach

Up to this point in the study, randomness was associated with a design-based approach in which a simple random sample of *n* plots were drawn without replacement from a total of *N* plots. In this section, we present properties of the estimators resulting from *k*-tree sampling and fixed-radius plot sampling under the assumptions that N_i, r_i, and Y_{ij} were random variables. Variable-radius plot estimators of mean tree age and age-class distribution were dismissed from further analyses as they exhibited undesirable properties.

Fig. 4. Values of the estimates of the proportion of trees in each age class for simulated forest conditions in three case study forest types ((*a*) two-aged mosaic, (*b*) all-aged random, and (*c*) young-aged clustered) using the three sampling methods. For each forest, the top panel is for *k*-tree sampling, the center panel is for fixed-radius plot sampling, and the bottom panel is for variable-radius plot sampling. Dotted lines and solid squares indicate the true values of the proportion of ponderosa pine trees in each age class. [Colour online.]



Thompson (1956), Eberhardt (1967), and Lessard et al. (2002) modeled N_i and r_i as random quantities. Specifically, consider the case in which $N_i \overset{\text{iid}}{\sim} \text{Poisson}(\tau_{ntrees}/N)$. Equivalently, assume that $N_i \overset{\text{iid}}{\sim} \text{Poisson}(10^6\,\lambda/N)$, where λ is the number of trees per square metre. The aforementioned authors used the Poisson distribution to model the locations of trees in a forest as having a random spatial pattern. Based on the Poisson distribution assumption, Thompson (1956) showed that $2\pi\lambda r_i^2 \overset{\text{iid}}{\sim} \text{Gamma}(k, 2)$. If one also considers the case in which Y_{ij} are independent and identically distributed random variables with mean and variance given by μ and σ^2, respectively, then it can be shown that

$$E(\hat{\mu}_1) \approx \mu \tag{8}$$

$$\text{Var}(\hat{\mu}_1) \approx \frac{k-1}{k(k-2)}\frac{\sigma^2}{n} \tag{9}$$

and

$$E(\hat{\mu}_2) \approx \mu \tag{10}$$

$$\text{Var}(\hat{\mu}_2) \approx \frac{N}{\tau_{ntrees}}\frac{\sigma^2}{n} \tag{11}$$

when the sample size is large. See the Supplementary information[1] for details.

Based on these assumptions about the random variables N_i, r_i, and Y_{ij}, one may conclude that the ratio estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ have negligible biases. This theoretical result is consistent with the simulation results in which $\sqrt{\text{Var}(\hat{\mu})} \approx \sqrt{\text{MSE}(\hat{\mu})}$. Furthermore, the variances of $\hat{\mu}_1$ and $\hat{\mu}_2$ can be compared with one another using the approximate results. For instance, if one selects $k = \tau_{ntrees}/N$ trees per plot, then $\text{Var}(\hat{\mu}_1) > \text{Var}(\hat{\mu}_2)$ and the fixed-radius plot sampling estimator of mean tree age will be more precise. Alternatively, if the investigator favored k-nearest neighbor sampling over the simple random sampling methods due to its ease of implementation, then a value of k greater than τ_{ntrees}/N should be considered. Specifically, if one selects k to be the smallest integer satisfying $k(k-2)/(k-1) > \tau_{ntrees}/N$, then the k-nearest trees distance sampling estimator will be more precise. This value of k can be expressed as

$$k = 1 + \text{ceiling}\left[\frac{\tau_{ntrees}}{2N} + \sqrt{1 + \left(\frac{\tau_{ntrees}}{2N}\right)^2}\right] \tag{12}$$

where ceiling(x) is the smallest integer greater than or equal to x. Of course, more trees must be measured in k-tree sampling as compared with fixed-radius plot sampling for the resulting estimators to have comparable variances. These results correspond to those reported by Lessard et al. (2002) when estimating tree density.

Eberhardt (1967) and Lessard et al. (2002) also considered the scenario in which the negative binomial distribution was used to model the locations of trees in a forest having a clustered or aggregated spatial pattern. Specifically, suppose that N_i has a negative binomial distribution with probability of success equal to $(\tau_{ntrees}/N)/(\tau_{ntrees}/N + h)$, where $h > 0$. The parameter h (which Eberhardt (1967) denotes as k) is called the constant of heterogeneity in which the trees become more clustered as h approaches 0 and the trees become more randomly spaced as h approaches ∞. Using this model, it can be shown that $\text{Var}(\hat{\mu}_1) \le \text{Var}(\hat{\mu}_2)$ when

$$k = 1 + \text{ceiling}\left\{\frac{\tau_{ntrees}}{2N}\frac{(h+1)}{h} + \sqrt{1 + \left[\frac{\tau_{ntrees}}{2N}\frac{(h+1)}{h}\right]^2}\right\} \tag{13}$$

An estimate of h, denoted by $\hat{h}$, can be obtained by solving

$$\hat{h} = \frac{(\tau_{ntrees}/N)^2}{\widehat{\text{Var}}(N_i) - \tau_{ntrees}/N} \tag{14}$$

Incorporating spatial autocorrelation of the Y_{ij}s in the model-based approach can also be accomplished. For simplicity, suppose that the covariance between Y_{ij}s in the same plot is constant and $\text{Cov}(Y_{ij}, Y_{i'j}) = 0$, $i \ne i'$, so that the covariance between Y_{ij}s in different plots is 0. In this manner, the model suggests that any pair of responses taken from the same plot would be equicorrelated and any pair of responses taken from different plots would be uncorrelated. Let ρ denote the intraplot correlation coefficient of the Y_{ij}s and suppose that $\rho > 0$; then it can be shown that $\text{Var}(\hat{\mu}_1) \le \text{Var}(\hat{\mu}_2)$ when

$$k = 1 + \text{ceiling}\left\{\frac{\tau_{ntrees}}{2N}\frac{(h+1)}{h} + \sqrt{1 + \left[\frac{\tau_{ntrees}}{2N}\frac{(h+1)}{h}\right]^2 + \frac{\tau_{ntrees}}{N}\frac{(h+1)}{h}\rho}\right\} \tag{15}$$

In the extreme case in which $\rho = 1$, the k-tree sampling estimator of μ outperforms the fixed-radius plot sampling estimator of μ when $k = 2 + \text{ceiling}[(\tau_{ntrees}/N)(h+1)/h]$. For the three forests under consideration, the last column of Table 2 displays the values of k required so that the k-tree sampling estimator is more precise than the fixed-area plot estimator of mean tree age based on the formula $k = 2 + \text{ceiling}[(\tau_{ntrees}/N)(h+1)/h]$. It is interesting to note that the theoretical results using the model-based approach are similar to those of the simulation results discussed earlier.

Figure 5 displays the general relationship between h and k, when $\rho = 1$, for which the k-tree sampling estimator competes favorably with the fixed-area plot estimator for the three forest types under consideration. Note that k approaches $2 + \text{ceiling}(\tau_{ntrees}/N)$ as h approaches ∞ and k is unbounded as h approaches 0. Thus k-tree sampling results in a viable estimator when the value of h is along the horizontal asymptote associated with each forest type.

Table 2. Model-based sampling results where the value of k for which the k-tree sampling estimator is more precise than the fixed-area plot estimator of mean tree age for the three simulated forest conditions.

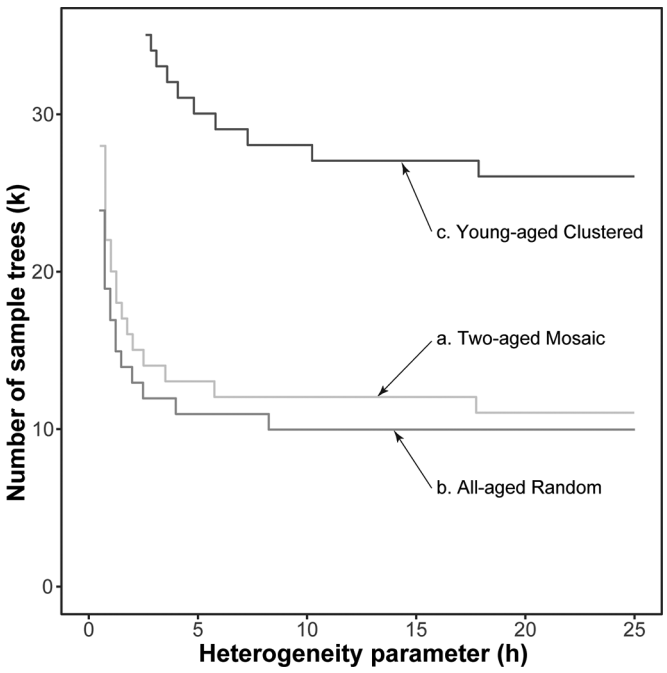
Forest	τ_{ntrees}/N	$\hat{h}$	$k = 2 + \text{ceiling}[(\tau_{ntrees}/N)(h+1)/h]$
Two-aged mosaic	8.5	269.6	11
All-aged random	7.1	16.8	10
Two-aged clustered	22.7	2.1	36

4. Discussion

The actual value of k employed in a particular study is under the purview of the investigator. For the simulated forests in this paper, we used $k = 7$, 9, and 23 trees per plot and focused on comparing three sampling methods rather than selecting an optimal value of k. In forest inventory applications, $k = 6$ appears to be a popular choice based on bias and root mean square error criteria. See Magnussen et al. (2008), Kleinn and Vilčko (2006), and Prodan (1968) for more details. When estimating the density of a popula-

[1]Supplementary information is available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfr-2018-0098.

Fig. 5. Relationship between h, the heterogeneity parameter of the spatial pattern of trees, and the value of k for which the k-tree sampling estimator compares favorably with the fixed-area plot estimator in the three case study forest types: (*a*) two-aged mosaic, (*b*) all-aged random, and (*c*) young-aged clustered.



tion of trees, Kronenfeld (2009) suggests that a good sampling strategy is to select 5 or more trees at each of between 20 and 40 plots. Recently, studies using k-tree sampling to estimate tree and stand age have become more prevalent in the literature and reported values of k range from $k = 10$ (Yocom-Kent et al. 2015) to $k = 30$ (Brown 2006; Brown and Wu 2005; Brown et al. 2008).

In many applications, the overall cost of the study is fixed and the investigator must decide on the values of n and k. A cost function for this scenario (see Thompson (2012), p. 178) is

$$C = c_0 + c_1 n + c_2 nk \tag{16}$$

where C is the overall cost, c_0 is a constant overhead cost, c_1 is the cost per plot, and c_2 is the cost per individual tree within a plot. The overhead cost includes planning and administrative activities that are not directly related to the sampling process. The cost per plot includes the cost of locating and traveling to the plot, as well as the cost of enumerating the trees in the plot. The cost per tree pertains to the work required to measure the age of the tree. For k-tree sampling, holding both the cost function and the heterogeneity parameter h constant, the variance of the estimator of mean tree age is related to k in the following manner

$$\mathrm{Var}(\hat{\mu}_1) \propto \frac{k-1}{k(k-2)}(c_1 + c_2 k)[1 + \rho(k-1)] \tag{17}$$

In other words, for a fixed overall cost of sampling from a forest, the precision of the k-tree sampling estimator of the mean tree age depends on the value of k and the intraplot correlation coefficient ρ. Furthermore, the optimal choice of k depends on the value of ρ. In general, the value of k that minimizes $\mathrm{Var}(\hat{\mu}_1)$ decreases as ρ increases. For example, in the scenario in which $c_1 = c_2 = 1$, if $\rho = 0.01$, then one would choose $k = 16$, whereas if $\rho = 1.0$, then one would select the smaller value $k = 4$.

Sampling costs are also important to consider when selecting a particular sampling method. For instance, if the respective costs c_0, c_1, and c_2 involved in k-tree sampling and fixed-radius plot sampling are equal and there is no prior information about the age and spatial configuration of the trees in the forest, then fixed-radius plot sampling is preferred because it yields an estimator that has greater precision. However, fixed-radius plot sampling may incur additional costs per plot because the number and location of all of the trees in the plot must be determined. In other words, if the plot-level costs of fixed-radius plot sampling are large compared with the plot-level costs of k-tree sampling, then k-tree sampling may be advisable even though more trees must be measured to attain the same precision.

5. Summary

In this study, we considered different sampling methods used to estimate the age characteristics of a forest. Trees in the forest were sampled using k-tree sampling, fixed-radius plot sampling, or variable-radius plot sampling. While tree attributes such as location and dbh can be easily obtained on the ground or estimated via remote sensing methods, obtaining quality tree age measurements requires one to core a tree. Auxiliary information such as the number of trees per plot, the distance that trees are from plot center, and the dbh of trees were used to form ratio estimators.

The properties of the estimators were examined using design-based and model-based approaches, both of which suggest that the estimator associated with the k-tree sampling method may be preferred if a few extra trees per plot can be acquired for forests having spatially mosaic or random spatial patterns. The estimator based on fixed-radius plot sampling outperformed the other two estimators for the forest having a clustered spatial pattern. The spatial pattern of the trees, as determined by the heterogeneity parameter h, was shown to influence the final choice of k in the k-tree sampling method. Furthermore, when holding sampling costs fixed, the intraplot correlation coefficient of trees ages ρ had an impact on the selection of k.

As expected, estimators of mean tree age and age-class distribution based on variable-radius plot sampling did not fare well for our simulated ponderosa pine forest applications. Specifically, $\hat{\mu}_3$ exhibited a noticeable bias and the standard error of $\hat{\mu}_3$ was large compared with the standard error of the other estimators. Furthermore, the variable-radius plot sampling strategy underestimated the proportion of younger trees and overestimated the proportion of older trees for our forest. The poor performance of the variable-radius plot sampling method may be rooted in the fact that the coefficient of variation of the ponderosa pine tree dbh was rather large. Variable-radius plot sampling estimators of tree age may perform better in applications where the observed coefficient of variation of dbh is small.

While the design-based and model-based methods corroborated one another in terms of the describing the properties of the estimators, further investigations should consider other forest types and structures. In the end, the decision to use the k-tree sampling method for a particular application may depend on a combination of its statistical properties, ease of implementation, and overall cost. This decision is an important one as estimating tree and forest age is instrumental to understanding forest structure and function and important when evaluating management actions.

Acknowledgements

The authors thank Dan Binkley for his contributions during research question development and extend a special thanks to three anonymous reviewers and Dr. David Auty, whose helpful comments greatly improved this article. Northern Arizona University is an equal opportunity provider.

References

Avery, T.E., and Burkhart, H.E. 2002. Forest measurements. 5th ed. Waveland Press, Inc., Long Grove, Ill.

Binkley, D. 2008. Age distribution of aspen in Rocky Mountain National Park, USA. For. Ecol. Manage. **255**: 797–802. doi:10.1016/j.foreco.2007.09.066.

Brown, P.M. 2006. Climate effects on fire regimes and tree recruitment in Black Hills ponderosa pine forests. Ecology, **87**: 2500–2510. doi:10.1890/0012-9658(2006)87[2500:CEOFRA]2.0.CO;2. PMID:17089659.

Brown, P.M., and Wu, R. 2005. Climate and disturbance forcing of episodic tree recruitment in a southwestern ponderosa pine landscape. Ecology, **86**: 3030–3038. doi:10.1890/05-0034.

Brown, P.M., Wienk, C.L., and Symstad, A.J. 2008. Fire and forest history at Mount Rushmore. Ecol. Appl. **18**: 1984–1999. doi:10.1890/07-1337.1. PMID:19263892.

Clark, P.J., and Evans, F.C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology, **35**: 445–453. doi:10.2307/1931034.

Ducey, M.J., Gove, J.H., Ståhl, G., and Ringvall, A. 2001. Clarification of the mirage method for boundary correction, with possible bias in plot and point sampling. For. Sci. **47**: 242–245.

Eberhardt, L.L. 1967. Some developments in 'distance sampling'. Biometrics, **23**: 207–216. doi:10.2307/2528156. PMID:6051587.

Fulé, P.Z., Heinlein, T.A., Covington, W.W., and Moore, M.M. 2003. Assessing fire regimes on Grand Canyon landscapes with fire-scar and fire-record data. Int. J. Wildland Fire, **12**: 129–145. doi:10.1071/WF02060.

Gregoire, T.G. 1982. The unbiasedness of the mirage correction procedure for boundary overlap. For. Sci. **28**: 504–508.

Kleinn, C., and Vilčko, F. 2006. A new empirical approach for estimation in *k*-tree sampling. For. Ecol. Manage. **237**: 522–533. doi:10.1016/j.foreco.2006.09.072.

Kronenfeld, B.J. 2009. A plotless density estimator based on the asymptotic limit of ordered distance estimation values. For. Sci. **55**: 283–292.

Lessard, V.C., Reed, D.D., and Monkevich, N. 1994. Comparing *N*-tree distance sampling with point and plot sampling in Northern Michigan forest types. North. J. Appl. For. **11**: 12–16.

Lessard, V.C., Drummer, T.D., and Reed, D.D. 2002. Precision of density estimates from fixed-radius plots compared to *N*-tree distance sampling. For. Sci. **48**: 1–6.

Li, W., Guo, Q., Jakubowski, M.K., and Kelly, M. 2012. A new method for segmenting individual trees from the lidar point cloud. Photogramm. Eng. Remote Sens. **78**: 75–84. doi:10.14358/PERS.78.1.75.

Lie, M.H., Arup, U., Grytnes, J.-A., and Ohlson, M. 2009. The importance of host tree age, size and growth rate as determinants of epiphytic lichen diversity in boreal spruce forests. Biodivers. Conserv. **18**: 3579–3596. doi:10.1007/s10531-009-9661-z.

Loewenstein, E.F., Johnson, P.S., and Garrett, H.E. 2000. Age and diameter structure of a managed uneven-aged oak forest. Can. J. For. Res. **30**(7): 1060–1070. doi:10.1139/x00-036.

Lucas-Borja, M.E., Hedo, J., Cerdá, A., Candel-Pérez, D., and Viñegla, B. 2016. Unravelling the importance of forest age stand and forest structure driving microbiological soil properties, enzymatic activities and soil nutrients content in Mediterranean Spanish black pine (*Pinus nigra* Ar. ssp. *salzmannii*) forest. Sci. Total Environ. **562**: 145–154. doi:10.1016/j.scitotenv.2016.03.160. PMID:27099995.

Lynch, T.B. 2012. A mirage boundary correction method for distance sampling. Can. J. For. Res. **42**(2): 272–278. doi:10.1139/x11-185.

Magnussen, S., Kleinn, C., and Picard, N. 2008. Two new density estimators for distance sampling. Eur. J. For. Res. **127**: 213–224. doi:10.1007/s10342-007-0197-z.

Prodan, M. 1968. Punktstichprobe für die forsteirichtung [A point sample for forest management planning]. Forst- Holzwirt, **23**: 225–226.

R Core Team. 2017. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Roussel, J.-R., and Auty, D. 2017. lidR: airborne LiDAR data manipulation and visualization for forestry applications [online]. R package Version 1.2.0. Available from https://github.com/Jean-Romain/lidR.

Rozas, V. 2015. Individual-based approach as a useful tool to disentangle the relative importance of tree age, size and inter-tree competition in dendroclimatic studies. iForest, **8**: 187–194. doi:10.3832/ifor1249-007.

Sánchez Meador, A.J., Moore, M.M., Bakker, J.D., and Parysow, P.F. 2009. 108 years of change in spatial pattern following selective harvest of a *Pinus ponderosa* stand in northern Arizona, USA. J. Veg. Sci. **20**: 79–90. doi:10.1046/j.1365-2893.1999.00142.x-i1.

Sánchez Meador, A.J., Parysow, P.F., and Moore, M.M. 2010. Historical stem-mapped permanent plots increase precision of reconstructed reference data in ponderosa pine forests of northern Arizona. Restor. Ecol. **18**: 224–234. doi:10.1111/j.1526-100X.2008.00442.x.

Savage, M., Brown, P.M., and Feddema, J. 1996. The role of climate in a pine forest regeneration pulse in the southwestern United States. Écoscience, **3**: 310–318. doi:10.1080/11956860.1996.11682348.

Schreuder, H.T., Banyard, S.G., and Brink, G.E. 1987. Comparison of three sampling methods in estimating stand parameters for a tropical forest. For. Ecol. Manage. **21**: 119–127. doi:10.1016/0378-1127(87)90076-4.

Stokes, M.A., and Smiley, T.L. 1996. An introduction to tree-ring dating. University of Chicago Press, Chicago, Ill.

Thompson, H.R. 1956. Distribution of distance to the *N*th neighbour in a population of randomly distributed individuals. Ecology, **37**: 391–394. doi:10.2307/1933159.

Thompson, S.K. 2012. Sampling. John Wiley & Sons, Hoboken, N.J. doi:10.1002/9781118162934.

Tuten, M.C., Sánchez Meador, A.J., and Fulé, P.Z. 2015. Ecological restoration and fine-scale forest structure regulation in southwestern ponderosa pine forests. For. Ecol. Manage. **348**: 57–67. doi:10.1016/j.foreco.2015.03.032.

Veblen, T.T. 1986. Age and size structure of subalpine forests in the Colorado Front Range. Bull. Torrey Bot. Club, **113**: 225–240. doi:10.2307/2996361.

Vilén, T., Gunia, K., Verkerk, P.J., Seidl, R., Schelhaas, M.-J., Lindner, M., and Bellassen, V. 2012. Reconstructed forest age structure in Europe 1950–2010. For. Ecol. Manage. **286**: 203–218. doi:10.1016/j.foreco.2012.08.048.

Yocom-Kent, L.L., Fulé, P.Z., Bunn, W.A., and Gdula, E.G. 2015. Historical high-severity fire patches in mixed-conifer forests. Can. J. For. Res. **45**(11): 1587–1596. doi:10.1139/cjfr-2015-0128.