Canadian Journal of
Forest Research

Article

Stand validation of lidar forest inventory modeling for a managed southern pine forest

Jacob L. Strunk[a] and **Robert J. McGaughey**[b]

[a]USDA Forest Service Pacific Northwest Research Station, 3625 93rd Avenue SW, Olympia, WA 98512, USA; [b]USDA Forest Service Pacific Northwest Research Station, University of Washington, P.O. Box 352100, Seattle, WA 98195-2100, USA

Corresponding author: **Jacob L. Strunk** (email: jacob.strunk@usda.gov)

Abstract

We evaluated area-based approaches (ABAs) to light detection and ranging (lidar) predictions of plot- and stand-level forest attributes (tree count, height, basal area, volume, aboveground biomass, broadleaf/conifer, and diameter at breast height — "diameter"). ABA methods included post-stratification (PS), ordinary least squares (OLSs) regression, *k* nearest neighbors (*k*NN), and random forest (RF). This study was conducted on the Savannah River Site in South Carolina, USA. Plot- and stand-level predictions were validated against fixed-radius 0.04 ha (0.1 acre) plots in 49 ≈2.0 ha (5 acre) stands. Our findings demonstrate that lidar can be incorporated operationally into forest inventory systems to provide stand-level inferences for a wide range of forest attributes. Volume predictions for specific diameter classes, however, often fared poorly (root mean squared error (RMSE) > 100%) for the methods we explored, especially for larger (less common) diameter trees. Stand-level results were consistently better than pixel-level results (10–200+ percentage points). *k*NN and RF performed similarly and better than OLS and PS, but RF was the most robust to model configurations, while *k*NN has practical advantages such as simultaneous predictions of many attributes.

Key words: lidar, forest inventory, stand-level inference, area-based approach, sampling

1. Introduction

Forest inventories provide information needed by land managers to assess forest conditions with respect to objectives such as sustainable yield, preservation, and wildlife habitat. Under ideal conditions, forest managers would have access to both high spatial and high temporal resolution inventory data providing details at stand and substand levels. However, costs and operational complexities often result in relatively sparse samples and infrequent measurements. To reduce costs and improve performance, forest inventories have leveraged remote sensing data (especially airborne imagery) for over 100 years (e.g., Latham and McCarty 1972; Justice and Townshend 1982). More recently, airborne light detection and ranging (hereafter simply lidar) has been demonstrated to provide spatially explicit, high resolution, wall-to-wall information (Næsset 1997*a*; Magnussen and Boudewyn 1998). This study focuses on applications of lidar inventory using area-based analysis, or the quantification of forest attributes for areas including plots, pixels, stands, and tracts. Lidar individual-tree-based methods are not discussed.

Lidar provides a snapshot of terrain and vegetation structural characteristics at a resolution which is finer than tree scale. Pulse densities for modern lidar acquisitions in the United States often exceed 4 pulses/m^2 (≈ 8 pulse/m^2 for this study) which provides enough detail to produce accurate canopy and terrain surfaces, and exceeds the thresholds needed to support various methods to model vegetation attributes (Holmgren 2004; Strunk et al. 2012). Lidar provides direct measurements of forest structure information such as vegetation height, crown shape (Vauhkonen et al. 2009), height quantiles (Magnussen and Boudewyn 1998), and canopy cover (Korhonen et al. 2011). While lidar has proven effective to quantify forest attributes such as volume per hectare, aboveground biomass, and carbon per hectare, lidar can also be used to support a broad range of natural resource applications including geology, hydrology, habitat evaluation, and fire risk assessment (Hudak et al. 2009). The effectiveness of lidar in forest monitoring applications has been demonstrated around the world (Næsset 2007; Woods et al. 2011; Magnussen et al. 2018; Cosenza et al. 2022) including to replace traditional forest inventories (e.g., Næsset 2014; Maltamo and Packalen 2014).

A distinction is henceforth made between the terms "estimation" and "prediction". Estimation is the calculation of a statistic from a sample with the intention to characterize an attribute of the population, such as (commonly) its mean, total, or variance. Prediction, in contrast, uses a model and auxiliary variables to calculate a plausible value for an unknown or unmeasured observation (Weisberg 2005, see p. 34); for example, it would be prediction to obtain aboveground biomass for a pixel, plot, or stand from a fitted ordinary least square (OLS) model for corresponding values of auxiliary lidar variables. The distinction between estimation and prediction then, is that in estimation our sample represents a partial measurement of the population that is used to infer properties of the full population, whereas for prediction, we have no field-measured information about our target attribute for the unmeasured observation and instead depend on the fitted model paired with corresponding auxiliary values. Bradley Efron (2020, see p. 638) demonstrates this distinction for linear regression: model coefficients computed from sample data are "estimates" of population coefficients and individual values generated from the model with auxiliary values are "predictions" for new unmeasured observations.

Arguably, the dominant operational approach to forest inventory with lidar is the area-based approach (ABA). This is in contrast to individual tree detection (ITD) methods where individual trees are extracted from the lidar point cloud (e.g., Jeronimo et al. 2018), a step that is not required for ABA. In ABA, corresponding (areal) plot-level field and lidar measurements are related using models, which can then be used to infer forest inventory attributes from remotely sensed attributes (Means et al. 2000; Næsset 2002). The lidar metrics are typically statistics computed on the vertical distributions of return heights ($\text{height}_{\text{lidar}} = \text{elevation}_{\text{lidar}} - \text{elevation}_{\text{DTM}}$, where $\text{elevation}_{\text{DTM}}$ is derived from the lowest lidar elevations) for areas corresponding to field plots and for wall-to-wall grid cells (e.g., 30 m cells). Common height statistics include median, maximum, height quantiles, and metrics relative to height thresholds such as proportion of returns above 2 m. Models fit between area-level lidar metrics and field-measurement-derived metrics are then used with raster wall-to-wall lidar metrics to provide wall-to-wall predictions of forest inventory attributes. The raster predictions are commonly aggregated to convenient spatial units such as stands or watersheds.

A key advantage of an ABA lidar forest inventory is that it can provide a tract-wide inventory snapshot that can also be used for stand-level and finer-resolution inferences on a relatively short timeframe, such as 1 year, and nominally requires only tens to hundreds of plots. This small number of required plots is advantageous, since managing the measurement of plots and flow of data, especially through time, can be one of the most difficult parts of forest inventory implementation. The exact numbers of plots needed to fit reliable plot-level models will vary by forest attribute (Strunk et al. 2012) and model type (Cosenza et al. 2022), as well as other factors such as plot and sample design (Gobakken and Næsset 2008; Maltamo et al. 2011), and local forest characteristics (Cosenza et al. 2021). Research in this area has demonstrated that it is reasonable to expect actionable stand-level information from a lidar inventory, since lidar is able to explain a high proportion of variation in forest attributes strongly associated with forest height (e.g., Erdody and Moskal 2010; Woods et al. 2011), where demonstrating empirical stand-level performances is also one of the objectives of this study. In comparison, a stand-based inventory using plots in every stand may require thousands to tens of thousands of plots for an entire tract, and typically requires many years to implement (e.g., 5–20 years). An advantage of the stand sampling inventory over ABA lidar inventory is that any target stand-level precision may be achieved (on average) by simply adjusting the plot sampling intensity — although the amount of effort required to achieve higher precisions can be prohibitively expensive.

Small area estimation (SAE) is a hybrid approach that we mention but do not explore, that can leverage advantages of both direct estimates (based only on plots) and model predictions, such as those from ABA lidar inventory (Rao 2003). SAE methods typically rely on having at least some measurements in each of the small areas (e.g., stands). SAE estimators yield improved small area inferences through the weighted combination of predictions (synthetic estimates) with direct estimates for the small areas but can require fewer plots than direct estimation to achieve the equivalent precision. SAE methods present an exciting opportunity for improvement and is an active area of research (e.g., Breidenbach et al. 2018; Green et al. 2022), although SAE does require additional data in the form of stand-level measurements relative to a simple ABA lidar inventory, e.g., lidar ABA based on a few hundred plots on a grid. SAE is related to this study in that the term "stand prediction" used here (the mean of pixel predictions falling within a stand) would be equivalent to a "synthetic estimate" in SAE terminology.

Publications describing lidar forest inventory studies typically examine relatively few inventory attributes. Individual studies have typically focused on two to eight attributes with the most common being volume per ha, aboveground biomass per ha, stand height, number of trees per ha, and basal area per ha. In addition, models typically represent quantities such as total volume for all live trees but not for dead trees, and not for species nor size groupings. While lidar intensity values and height metrics have been used to distinguish broad species groups (Ørka et al. 2009), efforts to characterize inventory attributes for specific size or product classes tend to focus on distributions of diameters (where henceforth "diameter" is used for diameter outside bark at breast height, 1.37 m) and then use the predicted distribution to summarize inventory information for diameter classes (Packalén and Maltamo 2007; Strunk et al. 2017).

Tree species is an important component of most forest inventory and monitoring systems, yet the preponderance of lidar-based inventory research is dedicated to attributes such as volume per hectare and stand height. Studies have reported the utility of lidar to predict the proportions of various species, especially when differentiating between broad species groups rather than individual species. For example, Liang et al. (2007) used the difference between first- and last-return surfaces to differentiate broadleaf from coniferous species. Ørka et al. (2009) used leaf-on lidar data to discriminate between a multiple broadleaf species (birch; *Betula* spp.) and Norway spruce (*Picea abies* (L.) Karst.). They found that the difference between first and last returns was useful along with the intensity of first returns. Lin and Hyyppä (2016) used a variety of lidar-derived structure and spectral metrics to predict the species for individual trees segmented from the point cloud. These studies suggest that using lidar metrics to predict broad species groups (such as broadleaf or conifer) or individual species is possible.

The body of literature on ABA forest inventory with lidar typically focuses on plot-level performances for a single modeling method (e.g., commonly, OLSs regression), impeding our ability to contrast model types while holding everything else constant. Studies that compare multiple modeling methods at the plot level include Temesgen et al. (2003), who compared various nearest neighbor (NN) methods for accuracy when predicting tree lists (list of species and diameter for every tree), Ver Hoef and Temesgen (2013), who compared various NN methods with spatial linear models (SLMs) using simulated and observed forest measurements, Temesgen and Ver Hoef (2015), who compared SLM, gradient nearest neighbor (GNN) imputation, and random forest (RF) imputation using simulated and real forestry data, and Shin et al. (2016), who compared various NN methods, GNN, RF, OLS, SLM, and geographically weighted regression for predicting several inventory attributes. However, we are not aware of any studies that evaluate multiple modeling methods for multiple forest attributes at both plot and stand scales.

Raster data products (e.g., 30 m pixels) generated by ABA forest inventories are directly usable for forest management. However, managers typically aggregate pixels to make inferences at coarser scales such as stands, compartments, or management units. Performance metrics used to assess inventory results, in contrast, are typically derived at the plot level. While theoretical variance properties can be used to show that, on average, aggregation of predictions (e.g., to the stand) yields improved precision relative to pixels, with a few exceptions (Packalen et al. 2019; Kotivuori et al. 2021), the magnitude of the effect is not widely appreciated or documented empirically. In addition, both the manner of aggregation (e.g., compute the stand-level mean of pixels versus compute stand-level lidar attributes) and the size of pixels relative to coarser-scale units may affect the properties of the coarser-scale predictions. Finally, inferences about relative model performances at the plot scale may differ from inferences at the stand scale, e.g., the relative precision of OLS versus *k* nearest neighbor (*k*NN) for pixels or plots versus stands.

Better understanding of ABA performances for various modeling approaches, forest attributes, and aggregation units is needed for managers to understand the suitability of lidar-based inventory methods to support their inventory objectives. For example, it could be important for land managers to understand the precision of predicted diameter distributions when attempting to use lidar to assess the need for thinning, evaluate habitat suitability, or identify stands containing specific products. Independent stand-level validation can be expensive but is the best way to guarantee defensible and applicable measures of performance. Cross-validation approaches can be more cost-effective than independent validation, but there is less confidence that the results are free from confounding. White et al. (2014) used independent validation when they compared weight-scale measures of conifer (various species) merchantable volume to lidar-derived predictions of merchantable volume for 272 stands harvested between 2008 and 2010 in central Alberta, Canada and found that lidar-derived predictions overestimated scale volumes by only 0.6%. Peuhkurinen et al. (2007) used stand-level harvester data to demonstrate that ITD with lidar fared poorly for small trees (<20 cm diameter) but adequately for larger trees (≥20 cm diameter). Woods et al. (2011) used independent validation for predicted top height (average of the 100 stems per ha with the largest diameter), average height (average height of all trees 9.1 cm diameter and larger), basal area per hectare, gross total inside bark volume per hectare, gross merchantable inside bark volume per hectare, and aboveground biomass per hectare and found root mean squared error (RMSE) values were generally <20% of the mean values.

The goal of this study is to evaluate the utility of various lidar-based ABA forest inventory approaches at both plot and stand scales for a broad range of essential forest inventory attributes. To address our goal, we define three primary objectives with respect to the evaluation of forest inventory approaches:

1. Quantify lidar ABA forest inventory performance with RMSE for a diverse set of forest attributes including attributes by diameter class and conifer versus broadleaf species using independent validation data,
2. Compare RMSE for several plot and stand prediction approaches including post-stratification (PS), OLS regression, *k*NNs, and RF, and
3. Compare plot- and stand-level RMSEs.

2. Materials and methods

2.1. Study site

This study was conducted on the Savannah River Site, a National Environmental Research Park covering 80 267 ha (198 344 acres). The Savannah River Site is in the southeastern coastal area of the United States in west central South Carolina (Fig. 1). In partnership with the Department of Energy (DOE), the USDA Forest Service Savannah River manages nearly 73 653 ha (182 000 acres) of commercial forest and more than 4856 ha (12 000 acres) of non-forest land for a variety of natural resources. The remaining 8% of the area (6614 ha; 16 344 acres) is developed for DOE industrial activities and is considered non-forest.

The topography is gently sloping with most slopes <10% (5.7°). Elevations range from 19 m (62 ft) near the Savannah River along the southwestern border of the site to 130 m (426 ft) in the northwest corner of the site.

Forests at the Savannah River Site are about 69% conifer dominant and 31% broadleaf dominant. Dominant pine species include longleaf (*Pinus palustris* Mill.) and loblolly pine (*Pinus taeda* L.), and common broadleaf species include various oaks (*Quercus* spp.), yellow poplar (*Liriodendron tulipifera* L.), blackgum (*Nyssa sylvatica* Marsh.), sweetgum (*Liquidambar styraciflua* L.), red maple (*Acer rubrum* L.), hickories (*Carya* spp.), and holly (*Ilex* spp.). Bottomland broadleaf forests are found along streams and on the "islands" or "ridges" in swamps and major drainages. In these forests, typical broadleaf tree species include water oak (*Quercus nigra* L.), laurel oak (*Quercus laurifolia* Michx.), sweetgum, elms (*Ulmus alata* Michx. and *Ulmus americana* L.), red maple, and yellow poplar. Bald cypress

Fig. 1. The Savannah River Site is in South Carolina in the southeastern United States. Data are UTM NAD83 Zone 17 (feet), basemap was obtained with ESRI ArcMap 10.7.1 (credits: ESRI, TomTom, US Department of Commerce, Census Bureau).

(*Taxodium distichum* L.) and water tupelo (*Nyssa aquatic* L.) are common in the swamp forests found along the southwestern boundary of the site, adjacent to the Savannah River.

2.2. Field measurements

Our field sample was designed to support both model fitting and independent plot- and stand-level validation. Model fitting plots were primarily arranged on a regular grid, and validation plots were clustered in “validation stands”.

2.2.1. Plot locations

An initial grid was laid out using a 1063 × 1063 m grid overlaid on the Savannah River Site boundary with a random start. These grid plot locations were intersected with a non-forest mask composed of industrial areas, water bodies, major roads, and other areas known to be inaccessible due to environmental and regulatory precautions. Grid plot locations falling in the masked area (non-forest) were dropped from further consideration and new locations were drawn from a set of randomly located replacement locations (a regular grid with a supplementary set of completely random plot locations in “forest” conditions). All potential plot locations were evaluated using 2017 1-m-resolution orthophotos collected by the National Agriculture Imagery Program. The goal for this evaluation was to identify plot locations with visual evidence of recent harvest or other significant disturbance which would cause a condition mismatch between field measurement and lidar measurement. Any disturbed plots were also replaced. Our final sample design included 555 plots.

2.2.2. Validation stands

The validation data set was designed to support both plot- and stand-level validation and consisted of 50 new stands designed for the purpose of this study. One of the 49 validation stands was not measured due to the depth of standing water on the target stand location at the time of tree measurements. Validation stands were measured using a sample of nine circular 0.04 ha (0.1 acre) plots arranged in a 3 × 3 grid and separated by 48.2 m (158 feet) producing a total of 450 validation plots. Each cluster of nine validation plots, hereafter called validation stands, represents a 2.023 ha (5 acre) square area. Validation stands were distributed using a stratified random process based on vegetation types defined by predicted basal area per hectare, 95th percentile height, and first return canopy cover above 2 m (6.56 feet). Vegetation types were derived using data from a Savannah River Site lidar inventory project implemented in 2009. Height and cover were adjusted to account for harvest activity. For stands harvested since the 2009 lidar collection, both height and cover were set to 0%. Three strata were defined for each of the attributes. Vegetation type was "conifer" for areas with 80% or more conifer basal area per hectare, "broadleaf" for areas with <20% conifer basal area per hectare, and "mixed" for all other forested areas. The three height strata were 0.0–18.3 m (60 feet), 18.3–24.4 m (80 feet), and >24.4 m. First return cover above 2 m strata were 0%–40%, 40%–70%, and >70%. Validation stands were placed such that each validation stand fell entirely within a single condition (combination of broadleaf versus conifer, height, and cover strata) and such that validation stands did not overlap one another.

2.2.3. Field plot measurements

All 1005 target plots (555 single plots and 450 validation plots) were located and monumented prior to the collection of field measurements. Crews navigated to target plot locations using recreational grade Global Navigation Satellite System (GNSS) receivers and then plot location data were collected using Javad Triumph 2 GNSS receivers. GNSS data were collected for at least 15 min (1 s epochs) and post-processed using a base station approximately 37 km (23 miles) from the Savannah River Site study area.

Plot field measurements were collected by a contractor between May 2018 and February 2019 and recorded using an electronic data recorder. For validation stands, all plots were 0.04 ha (0.1 acre). For the sitewide 1063 m grid of plots, there were three possible (circular) plot sizes for large-tree measurements. A 0.16 ha (0.4 acre) plot, 0.08 ha (0.2 acre) plot, or 0.04 ha plot was used if the number of live dominant or codominant trees (based on light availability) in the initial 0.04 ha plot had <9, 9–18, or >19 trees, respectively.

For all live and dead trees with diameter larger than 7.62 cm (3 in.), the following information was collected: tree number, species, diameter, height, and crown class based on light availability. Small (2.54–7.62 cm diameter) live tree counts were taken by species for 2.54–5.08 cm (1–2 in.) and 5.08–7.62 cm (2–3 in.) diameter classes. Small-tree tallies were always done on a 0.04 ha (0.1 acre) plot regardless of the large-tree plot size.

Some of the grid and replacement plots could not be located due to lost plot markers or flooding, and some of the validation plots were harvested before the contractor could collect their measurements. As a result, field measurements were collected on 549 of the 555 grid (and random replacement) plots and 439 of 450 target stand validation plots. Replacements for plots not measured were not established because the plot measurement contractor was not equipped to install new plots and collect GNSS data.

Field measurements were summarized using R (R Core Team 2020). Volumes for individual trees were predicted using the National Volume Estimator Library produced by the US Forest Service (2020) and accessed as a dynamic link library in R. Bole (wood + bark) aboveground biomass for individual trees was predicted from equations assembled from various sources (see Wang 2014 for a list of biomass equations and sources). A summary of plot-level field attributes is provided in Table 1.

For each of the attributes shown in Table 1, statistics were computed for live trees in various subgroups of trees, including broad species categories (broadleaf, conifer, and all species) and 2.54 cm (1 in.) diameter classes.

2.3. Remote sensing data

Lidar data were acquired for the study site on 7–8 March 2018. The total area covered was approximately 96 000 ha (237 000 acre). The acquisition campaign was designed to produce a nominal pulse density of 8 pulses/m^2 (including 50% swath overlap) using a Leica ALS70-HP scanner. Acquisition parameters are shown in Table 2.

Lidar deliverables included classified point cloud in LAS version 1.2 format with classes 1, 2, 7, and 18 indicating unclassified, ground, low noise, and high noise, respectively; 1-m-resolution bare earth surface model; index files for LAS and bare earth tiles; and metadata.

Point data were processed using FUSION software (McGaughey 2018) to produce point height (from point elevations less ground elevations), canopy height, and topographic metrics for field plots and for 30 m × 30 m raster pixels using all and first returns. Point-cloud metrics were computed using returns above 2 m and cover metrics were based on the ratios of returns above a 2 m height threshold. Canopy height metrics for 30 m × 30 m pixels were computed from a 1 m × 1 m canopy height model. Additional sets of point-cloud metrics, used for the broadleaf/conifer classification, were computed using all and first returns above 2 m and first returns within 2 m of the canopy surface (called upper canopy returns hereafter) for 2 m × 2 m pixels. Example of lidar-derived metrics include height statistics such as mean, standard deviation, height percentiles (p05, p10, …, p90, p95, p99), various moments, height band ratios (e.g., 1–2 m), and the proportion of returns above 2 m

Table 1. Plot-level attributes computed from field measurements with units, abbreviations, and summary statistics.

Attribute	Unit	Abbrev.	Mean	Std. dev.	Min.	Max.
Basal area	m^2/ha (ft^2/acre)	ba	23.9 (104.4)	10.8 (47.0)	0.1 (0.5)	61.6 (268.4)
Total stem volume	m^3/ha (ft^3/acre)	vol	216.9 (3101.0)	118.6 (1695.4)	0.3 (3.7)	800.5 (11446.7)
Bole (wood + bark) aboveground dry biomass	tonnes/ha (short tons/acre)	bbm	112.3 (50.1)	60.6 (27.0)	0.2 (0.1)	410.6 (183.3)
Density	trees/ha (trees/acre)	den	625.5 (253.2)	431.1 (174.5)	6.2 (2.5)	3558.2 (1440.0)
Quadratic mean diameter	cm (inches)	qmd	24.1 (9.5)	6.6 (2.6)	7.9 (3.1)	48.1 (18.9)
Lorey's height	m (ft)	lor	20.4 (67.0)	4.9 (16.2)	4.0 (13.0)	31.9 (104.5)
Arithmetic mean height of tallest 40 trees per acre	m (ft)	t40	23.4 (76.9)	5.6 (18.3)	4.0 (13.0)	39.5 (129.8)
Arithmetic mean plot height	m (ft)	mht	15.7 (51.4)	3.8 (12.3)	4.0 (13.0)	29.0 (95.2)
Median (50th percentile) plot height	m (ft)	med	15.6 (51.0)	5.0 (16.3)	2.0 (6.5)	29.6 (97.0)
Canopy base height	m (ft)	cbh	9.9 (32.5)	3.4 (11.1)	0.6 (2.0)	19.9 (65.3)
Live crown length	m (ft)	lcl	6.4 (21.0)	1.2 (4.0)	3.3 (10.8)	11.5 (37.8)

Table 2. Lidar acquisition parameters.

Parameter	Value
Nominal pulse spacing (m)	0.7
Nominal single swath pulse density (pls/m^2)	4.4
Nominal flight height (AGL metres)	1500
Nominal flight speed (kts)	130
Nominal flight speed (kts)	40 (± 20)
Scan frequency (Hz)	47.3
Pulse rate of scanner (kHz)	339.2
Line spacing (m)	902
Beam divergence (Mrad)	0.22
Nominal swath width (m)	1500
Nominal swath overlap (%)	50
Scan pattern	Triangle

(cover2). Refer to the FUSION software documentation for a more complete description of metrics.

2.4. Construction of a foliage presence index

To facilitate modeling forest attributes for broadleaf and conifer species separately, lidar was used to create an index of foliar presence. This index is used alongside other lidar-derived attributes for the prediction of forest inventory attributes by species groups. Since the lidar was collected in the spring before deciduous trees had leaves, both structure and intensity attributes differed for most conifer and broadleaf species. In general, conifer species had foliage and broadleaf species did not. However, the separation between conifer and broadleaf species was partially confounded by the fact that the study site has both deciduous conifers (e.g., bald cypress, *Taxodium distichum* L.) and evergreen broadleaf species (e.g., laurel oak, *Quercus laurifolia* Michx. and holly, *Ilex* spp.). In addition, dead trees without foliage regardless of species have structure and intensity attributes similar to those of live broadleaf trees without foliage.

Species groups (conifer versus broadleaf) were modeled using RF (Breiman 2001), an ensemble classifier that uses bootstrap decision trees to generate a more robust classifier than a single decision tree model. Data for RF consisted of a subset of field plots representing "pure" conditions (100% basal area/ha of either conifer (25 plots) or broadleaf tree species (39 plots)). RF was used to classify pixels in a 2 m × 2 m raster covering the entire study site as either broadleaf or conifer.

Variables used in the final RF model for broadleaf versus conifer pixels were L-moment coefficient of variation of height (ht_lcv), ratio of the number of returns above 2 m to the total number of first returns (ht_ratio), 10th percentile of intensity values for upper canopy first returns (int_p10), and the median of intensity values for upper canopy first returns (int_med). The first two predictors (ht_lcv and ht_ratio) described the variability of "structure" and the penetration of laser pulses into the canopy. In the case of trees without foliage, we expected higher variability and more penetration through crowns to the ground resulting in lower values for ht_ratio. The latter two variables (int_p10 and int_med) provided information on the relative reflectance of material related to individual first returns near the upper canopy surface. In general, branches reflect less energy than green foliage (Kim et al. 2009). We expected somewhat lower reflectance values for cells containing trees with no foliage compared to cells containing trees with foliage.

The 2 m × 2 m raster was summarized to produce the proportion of broadleaf pixels within each pixel in a 30 m × 30 m raster. Figure 2 shows an example of the classification results (2 m × 2 m cells) next to a leaf-off orthophoto for the same location.

2.5. Forest modeling approaches

For this study we evaluated OLS regression, *k*NN, and RF modeling methods to predict forest attributes from remote sensing attributes at plot and stand (stand-level pixel aggregate) scales. The exploration of modeling approaches here is not intended or suited to identify the "best" modeling strategy, but to provide insights into types of models (or algorithms), scale-related issues, and various types of attribute predictions. The "best" modeling strategy may vary with location, data, and objectives.

 Can. J. For. Res. **53:** 71–89 (2023) | dx.doi.org/10.1139/cjfr-2022-0032

Fig. 2. Comparison of leaf-off orthophoto (15 cm resolution) and broadleaf and conifer species classification (2 m resolution).

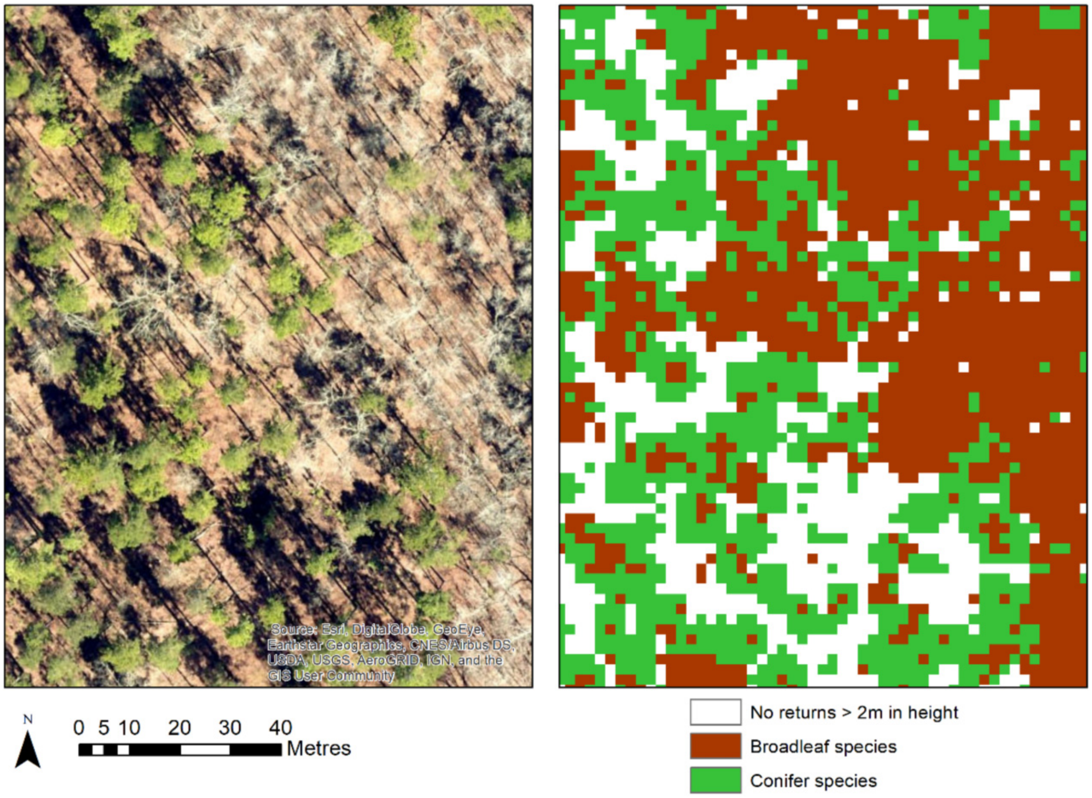


Table 3. Lidar-derived predictor groups.

Group	Variables
4 ×	p20, p95, profile.area, HW proportion
10 ×	4 ×, p40, p70, canopy relief ratio, coefficient of variation, skewness, cover 2 m
20 ×	10 ×, kurtosis, 1–2 m prop., 2–4 m prop., 4–8 m prop., 8–16 m prop., 16–32 m prop., 42–48 m prop., 64 m + prop., L1 mom., L2 mom.

Note: Abbreviations: mom., moment; prop., proportion.

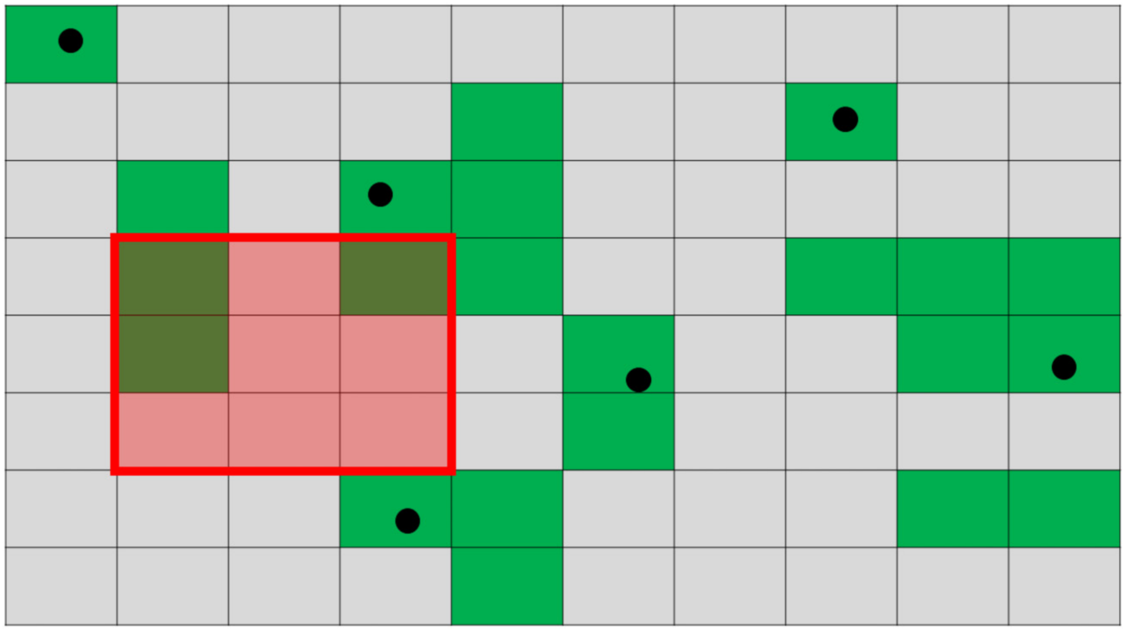
Models were fit with the R Statistical Language (R Core Team 2020). In the case of OLS, we compared three approaches to variable selection including (1) the same four lidar auxiliary variables selected by an "expert" for all forest attributes, (2) the same 20 lidar auxiliary variables for all forest attributes, which includes a wide range of height statistics, and (3) automated variable selection (AVS) for each forest attribute (the leaps and bounds algorithm by Furnival and Wilson 1974). For *k*NN, we looked at different numbers of neighbors ($k = 1$ and 5), 4 versus 20 lidar auxiliary variables, and different distance metrics available in the yaImpute package (Crookston and Finley 2008). The number of configurations reported for RF was more limited. We only demonstrated changes to the ntree RF parameter (500 versus 5000) and to the number of predictors (4 versus 20; Table 3) because changes to parameters had minimal impact on the results.

2.6. Post-stratification

In this study, the landscape was post-stratified with combinations of both one and two lidar-derived auxiliary variables (e.g., one bin could be p90 > 30 m and cover2 > 90% as in Fig. 3). Lidar-derived auxiliary variables were discretized into 100 target bins (strata) with five or six plots in each stratum. Lidar strata were then mapped to all pixels in the project area and intersected with field plot locations. Field plots falling within strata were used to estimate strata means — where "estimate" is the correct term in this instance because repeated random samples would generate strata mean distributions with means approximately equal to the true strata means. Pixel values in strata were then "predicted" as corresponding strata means — where "predict" is the appropriate term here because there is not an expectation that strata means are unbiased for the "true" forest attributes at pixel locations. To further illustrate this concept, we present a hypothetical forest in Fig. 3 where all of the green pixels fall within the same lidar stratum (because they fall within the same lidar metric bins) and are all assigned the same prediction: the mean of forest attributes computed for the plots falling within this stratum (represented by black dots in the figure).

The process of using post-strata to predict plot, pixel, and stand attributes is equivalent to modeling and prediction of

Fig. 3. Hypothetical forest composed of pixels (cells) with measured sample plot locations (black dots) for a hypothetical stratum (green defined as cover >90% and height >30 m), and other strata colored in gray. The red polygon represents a hypothetical stand covering an area composed of more than one stratum.



forest attributes based on a categorical auxiliary variable, where in this case the categorical variable is derived from discretized lidar metrics. Stand attributes were "predicted" using the means of plot predictions falling within stand boundaries (multiple lidar strata may fall within any stand, such as the red polygon in the hypothetical stand in Fig. 3) — again "predict" is appropriate here because there are no expectations about the distribution of possible estimates relative to the true field measurement derived stand-level forest attribute.

Post-stratification has a limitation related to the number of predictors that can be used for PS and the curse of dimensionality (Bellman 1966). The curse of dimensionality describes the idea that as the number of dimensions increases, the data become sparse. For example, for PS using a data set with 500 observations, two auxiliary variables with 10 bins each would yield up to 100 strata (~5 observations per stratum) and three auxiliary variables with 10 bins each would yield 1000 strata (<1 observation per stratum). In the context of this study, this means that the number of auxiliary variables (and hence explanatory information) used in PS is limited relative to the other approaches to prediction that are explored — the other approaches can use many more predictors. An approach we explore to overcome this limitation is to post-stratify on forest attributes predicted from lidar such as volume, biomass, and trees per hectare (e.g., Breidt and Opsomer 2008; Magnussen et al. 2015). In this case, OLS regression is first used to relate forest attributes to lidar. Predicted forest attributes were then used to create 100 bins with five to six plots each, and otherwise is identical to the PS process previously described. A preliminary investigation (not shown) explored numerous combinations of both lidar metrics and modeled forest attributes (univariate and bivariate), but we only report a few representative cases here.

An alternative not explored in this study is to create post-strata separately for each response variable. This would enable optimization of variable selection for each response, yielding improved performances but sacrificing the utility of having a single set of strata used for prediction of many attributes, including tree lists. Simultaneous predictions of a response vector is an advantage that PS shares with *k*NN over the other methods we explore. This is of considerable practical advantage, and a primary motivation to use PS. Since the simplicity of PS is lost when preparing separate strata for each response, we did not explore a multiple PS approach here.

2.7. Prediction performance

A common trend in lidar forest inventory studies is to report pixel- or plot-level performances. There is uncertainty, then, in how applicable the study results (e.g., plot-level RMSE or r^2) are to stand-level inferences, where stands are the dominant unit of area for operational forest inventory. Based on the properties of aggregation, we know that stand-level RMSE values are on average smaller than plot-level RMSE values, although the magnitude of the improvement from aggregation varies as a function of various factors including aggregation extent and spatial correlation — details which are not discussed here. This study instead focuses on empirical properties of plot- and stand-level prediction performances (RMSEs).

In this study, we report performances for PS, OLS, *k*NN, and RF at both plot and stand levels using RMSE or a scaled equivalent, where the plot-level RMSE is

$$\text{(1)} \qquad \text{RMSE}_{\text{pl}} = \sqrt{\frac{\sum {e_i}^2}{n}}$$

where n is the number of sample plots, $e_i = y_i - \widehat{y}_i$ is the predicted residual for observation i, y_i is the observed value (e.g., volume) for plot i, and $\widehat{y}_i$ is the value of y_i predicted with a model.

We are also interested in the true (but unknown) stand-level error variation (σ_{st} or σ^2_{st}) computed from all the tessellations for each stand (e.g., a census of plots in a stand).

$$\text{(2)} \qquad \sigma_{\text{st}} = \sqrt{\sigma^2_{\text{st}}}$$
$$\sigma^2_{\text{st}} = \frac{\sum {e_j}^2}{m}$$

where $e_j = \bar{y}_j - \widehat{\bar{y}}_j$ is the true residual for stand j, m is the number of stands, $\bar{y}_j$ is the true mean for stand j, and $\widehat{\bar{y}}_j$ is the mean of predictions $\widehat{y}_{ij}$ for all plots in stand j.

However, we can only estimate the value of σ_{st} from our sample since we do not actually have all 50 possible 0.04 ha plots that could hypothetically be tessellated from a 2 ha stand with respect to area. A naïve (biased) estimator simply substitutes the sample-based estimate of stand-level error $\widehat{e}_j$ for the true (unknown) e_j:

$$\text{(3)} \qquad \text{RMSE}_{\text{st},9} = \sqrt{\text{MSE}_{\text{st},9}}$$
$$\text{MSE}_{\text{st},9} = \frac{\sum \widehat{e}^2_j}{m}$$
$$\widehat{e}_j = \bar{y}_{9,j} - \widehat{\bar{y}}_{9,j}$$

where $\widehat{e}_j$ is the predicted residual for stand j from nine plots, $\bar{y}_{9,j}$ is the response mean for stand j, and $\widehat{\bar{y}}_{9,j}$ is the mean prediction for stand j.

Unfortunately, $E\,[\text{MSE}_{\text{st},9}] \neq \sigma^2_{\text{st}}$ because the stand-level prediction residual from nine plots, $\widehat{e}_j$, has an additional source of variation $\left(\widehat{e}_j - e_j\right)$ or $\widehat{e}_j = e_j + \left(\widehat{e}_j - e_j\right)$, so

$$\text{(4)} \qquad V\left(\widehat{e}_j\right) \approx V\left(e_j\right) + V\left(\widehat{e}_j - e_j\right) \neq V\left(e_j\right)$$

The additional error term results in stand-level MSE estimates that are too large when computed for a sample of plots from each stand instead of the whole stand (nine plots in our case). We can estimate the contribution of this extra term $\left(\widehat{e}_j - e_j\right)$ to $\text{MSE}_{\text{st},9}$ using the squared standard error $\left(\text{SE}^2_j\right)$:

$$\text{(5)} \qquad E\left[\text{SE}^2_j\right] = V\left(\widehat{e}_j - e_j\right)$$
$$\text{SE}^2_j = \left(1 - \frac{n_j}{N_j}\right) \sum_{i=1}^{n_j} \frac{\widehat{e}^2_{i,j}}{n_j\left(n_j - 1\right)}$$

where n_j is the number of plots in stand j (9), and N_j is the number of possible plots in a stand (50), which motivates the stand-level variance and standard deviation estimators MSE_{st} and RMSE_{st}, respectively.

$$\text{(6)} \qquad \text{MSE}_{\text{st}} = \text{MSE}_{\text{st},9} - \sum_{j=1}^{m} \frac{\text{SE}^2_j}{m}$$
$$\text{RMSE}_{\text{st}} = \sqrt{\text{MSE}_{\text{st}}}$$

To facilitate comparisons, RMSE values are reported in per cent of the mean:

$$\text{(7)} \qquad \text{RMSE\%} = \frac{\text{RMSE}_y}{\text{mean}\,(y)} \times 100\%$$

where RMSE_y is the RMSE associated with y, and y here is some forest attribute.

3. Results

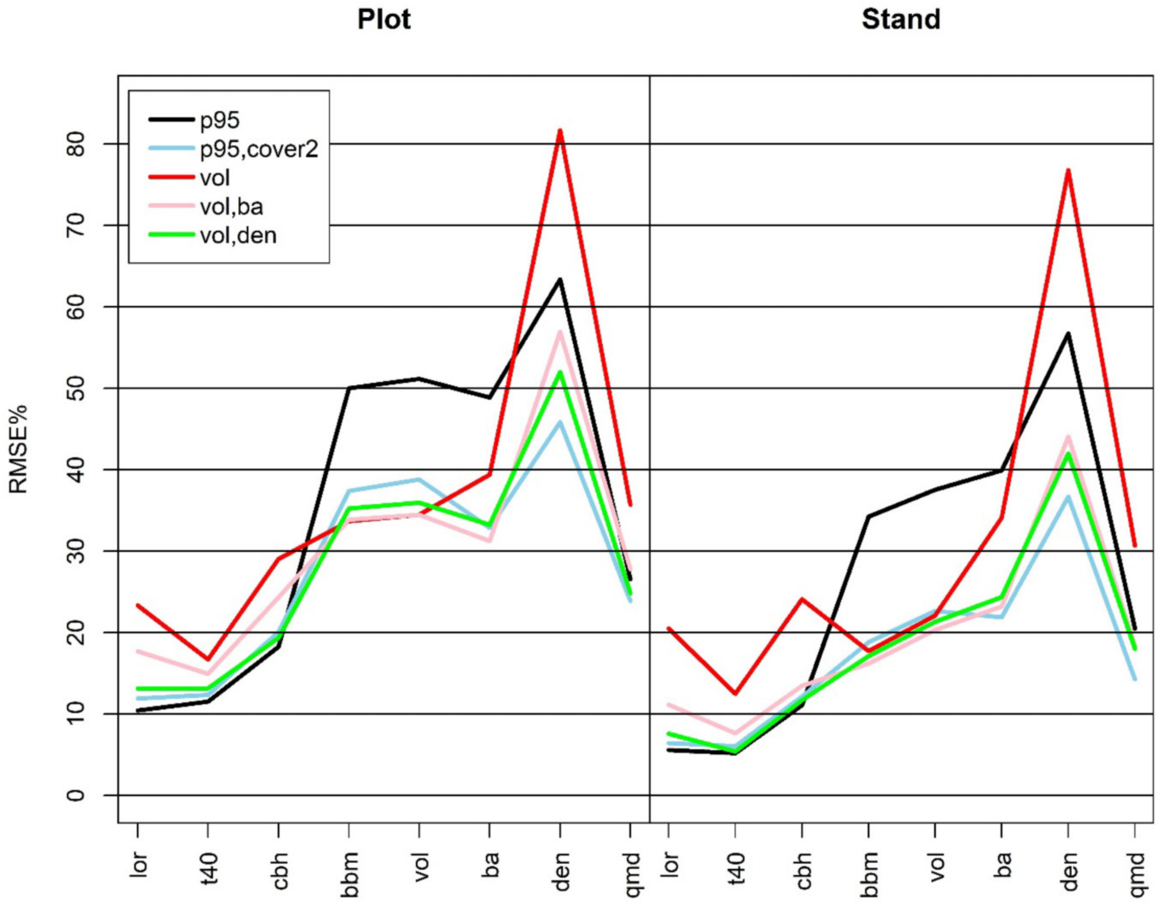
The following results describe performances of various prediction approaches at both plot and stand scales with respect to RMSE(%). The plot and stand inferences are compared to demonstrate the relative performances achieved at the two scales, but inferences about various prediction approaches are concentrated on the stand-level results. For example, the ranking of different prediction methods at plot and stand scales may differ slightly, although they tend to be similar. We focus on interpreting performances at the stand level, which is probably the most common scale of inference in operational forest inventories. Although not discussed in detail, the relative plot-level RMSE values are also valuable: plot-level results are the most useful in understanding prediction performances at the substand or pixel level, e.g., mapped predictions. Remote sensing provides an opportunity to make inferences at much finer scales than has historically been feasible from plot data alone (commonly stands), but there is typically a sacrifice in terms of prediction performance when working with finer spatial scales. The plot-level results demonstrate to what degree the RMSE increases when inferences are at a finer than stand scale.

3.1. Post-stratification performance

Stand-level performance (RMSE%) in Fig. 4 for PS was typically at least 20 percentage points better (lower RMSE% values) than at the plot level. At the stand level, PS with two auxiliary variables performed better than one auxiliary variable, with no clear winner between p95 and cover2, and predicted vol and ba. However, PS on p95 and cover2 generally performed as well or better than predicted vol and ba but was simpler and did not require modeling response attributes prior to PS. Post-stratification on modeled attributes did not warrant the additional effort of first fitting models.

While our choice of variables used in PS may seem arbitrary, we attempted to improve on these results with post-strata based on tens of combinations and weightings of both direct lidar metrics and modelled forest attributes (results not shown). Improvements were feasible for individual variables or groups of variables. However, our arbitrary initial selections of p95 and cover2 performed as well as any other choice of variables for balanced performance across a wide

Fig. 4. Plot- and stand-level post-stratification performances (RMSE% relative to the mean) for lidar-derived p95 and cover2 and predicted vol, ba, and den strata.



variety of forest attributes, and was simpler to implement than a modeled PS approach.

3.2. Ordinary least squares regression performance

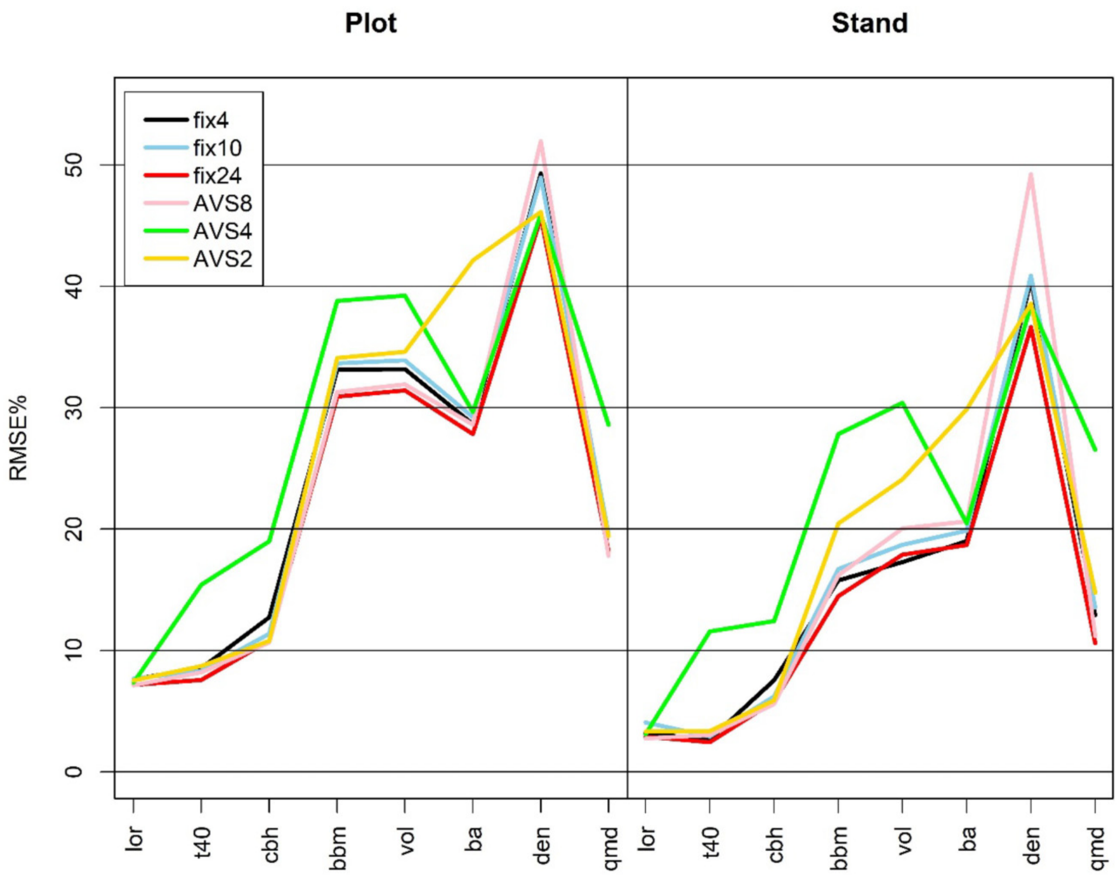
The best performances for OLS (Fig. 5) were similar, but slightly improved (≈ 2 + percentage points) over PS performances. As with PS, stand-level performances were better than plot-level performances, on the order of 5 percentage points for the best cases. Our primary observation for OLS was with respect to problems observed with AVS strategies.

Our results suggest that for the independent validation data set, AVS was an unreliable variable selection method compared to fixed variable selection (fix). In some cases, AVS outperformed other methods, e.g., AVS with two variables (AVS2) for canopy base height (cbh), but AVS methods were generally unreliable, resulting in performances as much as 12 percentage points poorer than other methods in some cases. There was not a conclusive result with respect to more or fewer numbers of AVS predictors performing better. Each number of AVS predictors resulted in a dramatically higher prediction RMSE% for one or more variables, although AVS8 and AVS2 each only had one such instance (den and ba, respectively) while AVS4 fared poorly in five instances (t40, cbh, bbm, vol, and qmd).

Fixed (fix) sets of predictors were stable relative to AVS methods, with no instances of fixed sets of predictors performing 5 percentage points poorer than the best performing approach. A fixed set of 24 predictors (fix24) performed consistently better than other methods in nearly all cases, but we hesitate to recommend this approach in practice. In our case, there were 549 field sample plots for model training, and we did not evaluate outliers or collinearity among predictors. A fixed set of four variables (fix4) performed better than 10 variables, nearly as well as 24 variables, is suitable for use with a much smaller set of training observations, and can be selected to avoid collinearity issues.

These findings should serve as a cautionary note regarding validation and variable selection in the absence of independent validation data when using OLS. Inferences about model fit based only on training data may be incorrect or misleading compared to inferences based on independent validation data. For variable selection, our findings using an independent validation data set showed AVS to be unreliable and to generally not perform as well as a fixed set of predictors. As would be expected, when we evaluated results for a fixed set of predictors and AVS using only the training data set (not validated on independent data, results not shown), we found

Fig. 5. Plot and stand OLS performances (RMSE% relative to the mean) for models fit using fixed ("fix") sets of predictors where the number of predictors was 4, 10, and 24, and automated variable selection (AVS) with different numbers of selected predictors (8, 4, 2) from a total of 24.



that AVS outperformed the fixed set of predictors in all cases and did not exhibit the instability seen for AVS in Fig. 5.

3.3. *k* nearest neighbor performance

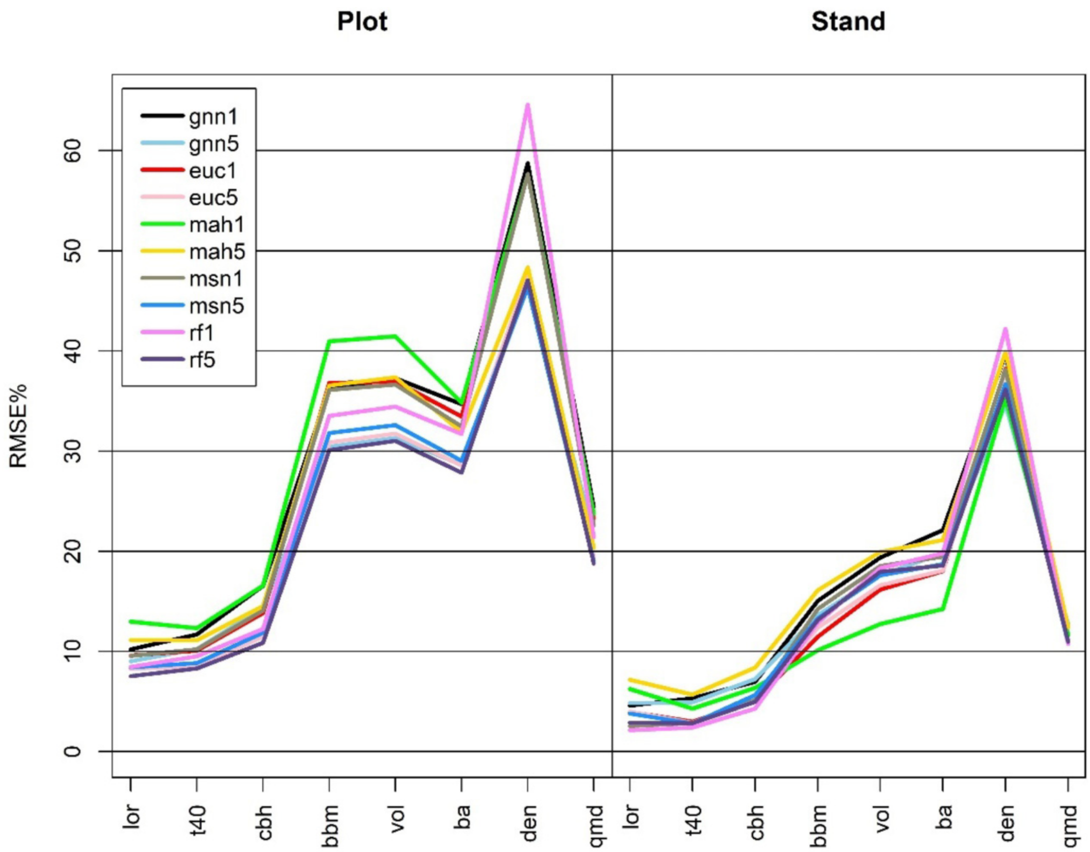
Plot- and stand-level *k*NN results (Fig. 6) were slightly better than OLS, and stand-level *k*NN results were better than plot-level *k*NN results. There was inconsistency, however, between plot- and stand-scale model inferences. At the plot-level, *k*NN based on RF classification with five neighbors consistently performed the best, but at the stand-level multiple distance metrics and both $k = 1$ and $k = 5$ nearest neighbors fared better, depending upon the forest attribute. Our primary finding for *k*NN is that *k*NN results were highly variable depending upon distance metric and number of neighbors; *k*NN RMSE values varied by up to nearly 20 percentage points at the plot level, and nearly 10 percentage points at the stand level. The best *k*NN configuration also varied by response, for example, "mah1" fared the best for ba prediction and "rf1" fared the best for lor prediction. While highly variable, at the stand level, *k*NN models demonstrated the best (lowest) absolute RMSE values among all the methods explored.

The findings presented for *k*NN may be poorly suited to guide users in model implementation. This is due to (1) the small number of validation stands (50 stands), (2) using a single auxiliary variable set with *k*NN, (3) the large number of comparisons made for *k*NN, and (4) the variability in "best" model performances. These findings are better suited to provide a general indication of the performance feasible with *k*NN, and perhaps suggest configurations to explore. Better results with *k*NN may also be feasible by evaluating various distance metrics and number of neighbors while tuning the training variable set (fixed versus AVS and different number of predictors). Variable selection was not an issue addressed in the analysis of *k*NN provided here due to the large number of combinations of factors already considered with *k*NN.

3.4. Random forest performance

The RF performances shown in Fig. 7 resemble those provided for OLS and *k*NN in that stand performances were better than plot performances, although RF fared slightly poorer than the best *k*NN models at the stand level. However, RF was the only approach that was highly robust to changes in the model parameters at both plot and stand levels. As with *k*NN, the results did not suggest an overall superior modeling configuration. For example, there are instances of both 4 and 24 predictors faring best. Variation due to the number of predictors was at most approximately 3.5 percentage points. There was no detectible difference between RF ntree parameters of

Fig. 6. Plot- and stand-level *k*NN performances (RMSE% relative to the mean) for $k = 1$ and 5, and various distance metrics. "gnn" or "gradient nearest neighbor" uses canonical correspondence analysis for distance and "msn" or "most similar neighbors" is based on canonical correlation analysis. "euc" and "mah" use Euclidean and Mahalanobis distances, respectively. "rf" distance is based on the number of shared nodes in a random forest model.



500 and 5000. This would suggest that the less time-intensive 500 ntree setting should be used. It may also be worth exploring even smaller values of ntree (faster still), as RF was the slowest approach to implement (RF is computer resource-intensive).

While we did not delve very deeply into modifying the RF parameters or variable selection (aside from three fixed numbers of predictors), our preliminary results for RF suggest that RF is incredibly robust with respect to model specification. The results for different values of the parameters "mtry" (not shown) and "ntree" had no discernible impact on inferences. Given the large number of parameters which can be adjusted in any of the modeling approaches, the fact that RF produced similar results in every case is a noteworthy characteristic. This is especially the case given that many lidar forest-inventory analysts will not have access to independent validation data. It appears that analysts can be confident in achieving decent results with RF for the structural forest attributes used here, even without the ability to validate model configurations against independent observations.

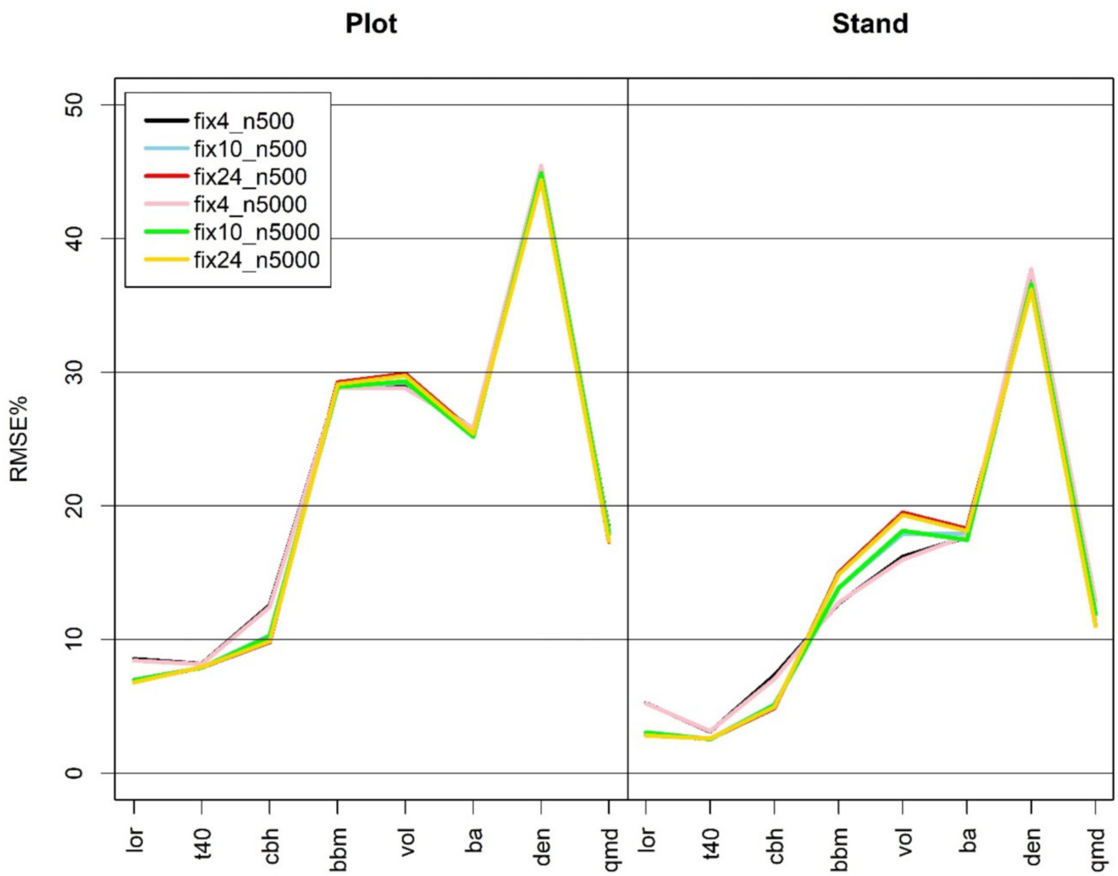
3.5. Species group predictions

In this section, we describe performances for forest attributes by species groups (Fig. 8). The results here are provided for configurations which performed best at the stand level in the previous sections. Aside from trees per hectare (den), predictive performances were better for all species combined than when broadleaf or conifer species attributes were predicted separately (Fig. 8). In the case of trees per hectare, conifer predictions were approximately the same (best case) as all species combined. The effects of species groups (all, conifer, broadleaf) on performances varied by scale (plot versus stand), forest attribute, and model type. Conifer performance demonstrated a similar pattern of performance to all species but the best cases had from 7 to 20 percentage points higher RMSE% values than all species at the plot level, and from 0 to 11 percentage points higher at the stand level (best cases). The best broadleaf cases, in contrast, ranged from 24 to 89 percentage points higher than all species RMSE% at the plot level, and 10 to 35 percentage points higher at the stand level (best cases). Broadleaf species regenerate naturally within our study area and generally exhibit more structural complexity than the conifer species that are typically planted and managed (especially pine) on the study site.

The best performances at both plot and stand scales were observed for *k*NN and RF, which had comparable performances. PS performed the worst of the four methods in most cases. In some cases, RF and *k*NN produced only modest improvements compared to OLS (e.g., plot-level BROADLEAF

 Can. J. For. Res. **53:** 71–89 (2023) | dx.doi.org/10.1139/cjfr-2022-0032

Fig. 7. Plot- and stand-level RF performances (RMSE % relative to the mean) for different fixed numbers of predictors (“fix”) and different values of ntree (“n”) parameters.



lor). However, in other cases, the differences were more dramatic. For broadleaf volume, RMSE% for OLS was more than 40% larger than for RF and *k*NN. In addition, the average reductions in performances associated with predicting species groups broadleaf and conifer were smallest for *k*NN and RF.

3.6. Diameter and species group predictions

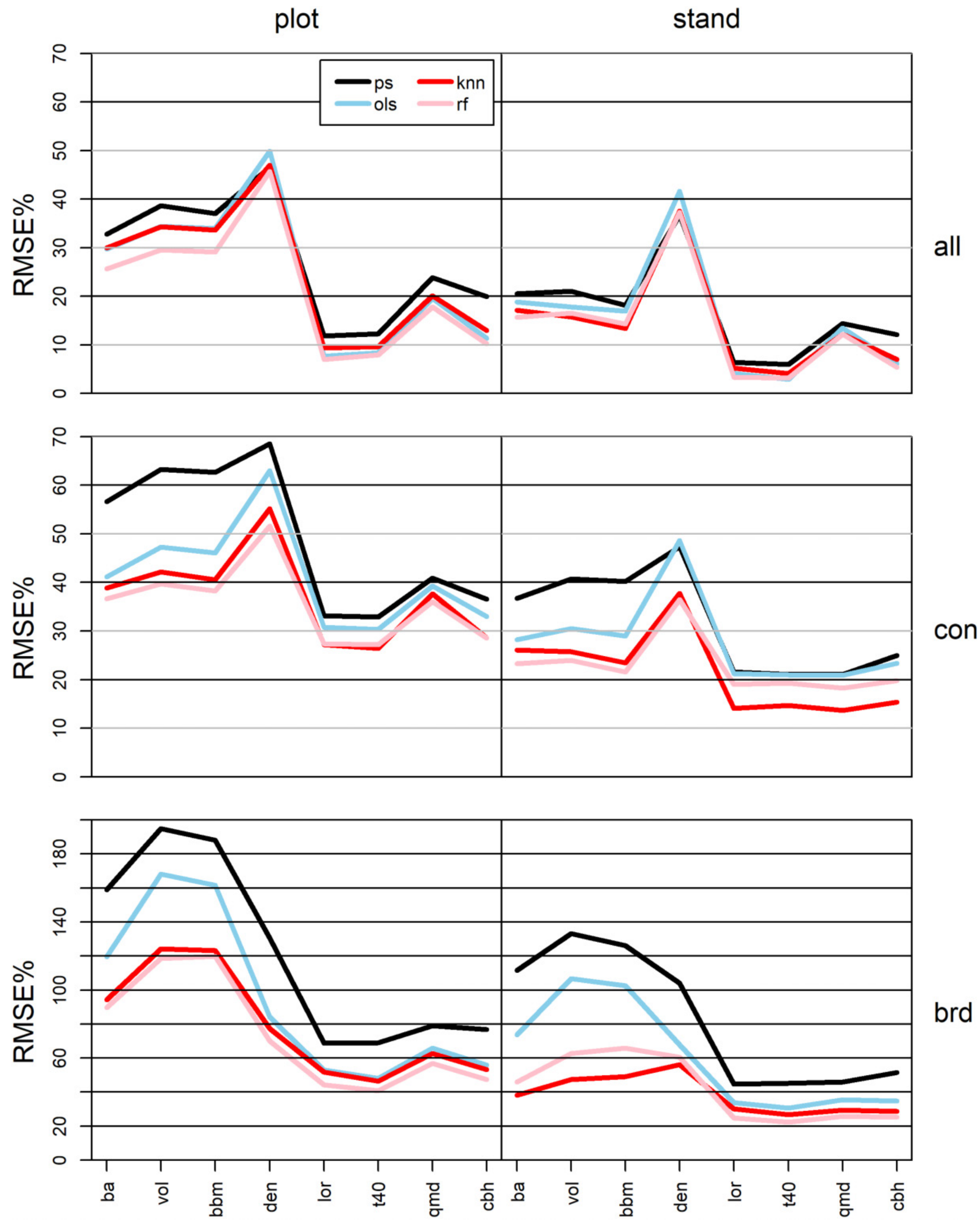
Figure 9 shows prediction performances (RMSE%) for volume by diameter class and species group. As in the previous section (Species group predictions), the results here are for model configurations that performed best within each model type. As in previous results, *k*NN and RF generally performed better than the other methods. Unlike the results in previous sections, PS performed as well or better than some other methods at plot and stand scales in some instances (instead of always performing the worst), although it never performed the best. In addition, unlike in previous results, the model types that performed best at the plot level did not always perform best at the stand level. The results also resemble previous sections in that conifer results are similar to all-species results, with broadleaf results faring poorest.

Unfortunately, the results in Fig. 9 suggest that for individual 5 cm (≈ 1 in.) diameter classes, the methods used here to predict volume by diameter class are highly unreliable regardless of how species were grouped. All-species combined predictions fared the best, but even for all-species the smallest RMSE% observed was at the stand level for 40 cm and was 37%, which would yield a 95% confidence interval of approximately (26%, 174%). In most other instances, RMSE% values exceed 50% (95% CIs would include zero), which means that volume by diameter predictions are not differentiable from 0.0 at the 0.05 α level. While this is concerning for inferences for individual diameter classes, in practice that is rarely how diameter data are used. Instead, they are typically used to look at the distribution of tree sizes across multiple diameter classes rather than from individual bins. These data may still be usable for broader inference, such as which range of diameters bins have greater proportions of volume relative to others. One of the reasons for the larger confidence intervals on 5 cm diameter bin predictions is that most plots did not have trees in every diameter bin, i.e., observations are 0.0 (zero inflated) for many or most bins. The methods used herein tend to perform poorly in this scenario.

4. Discussion

This study provides independent stand-level confirmation that lidar-derived predictors can be used to successfully model a wide range of forest attributes including forest attributes by species groups, but that the methods employed here proved less successful in prediction of volumes by diameter classes. Our findings are in agreement with

Fig. 8. Plot- and stand-level performances (RMSE% relative to the mean) for gross forest attributes for broadleaf and conifer (all), conifer (con), and broadleaf (brd). Note that due to differences in magnitude, the *y*-axis for brd differs from con and all.

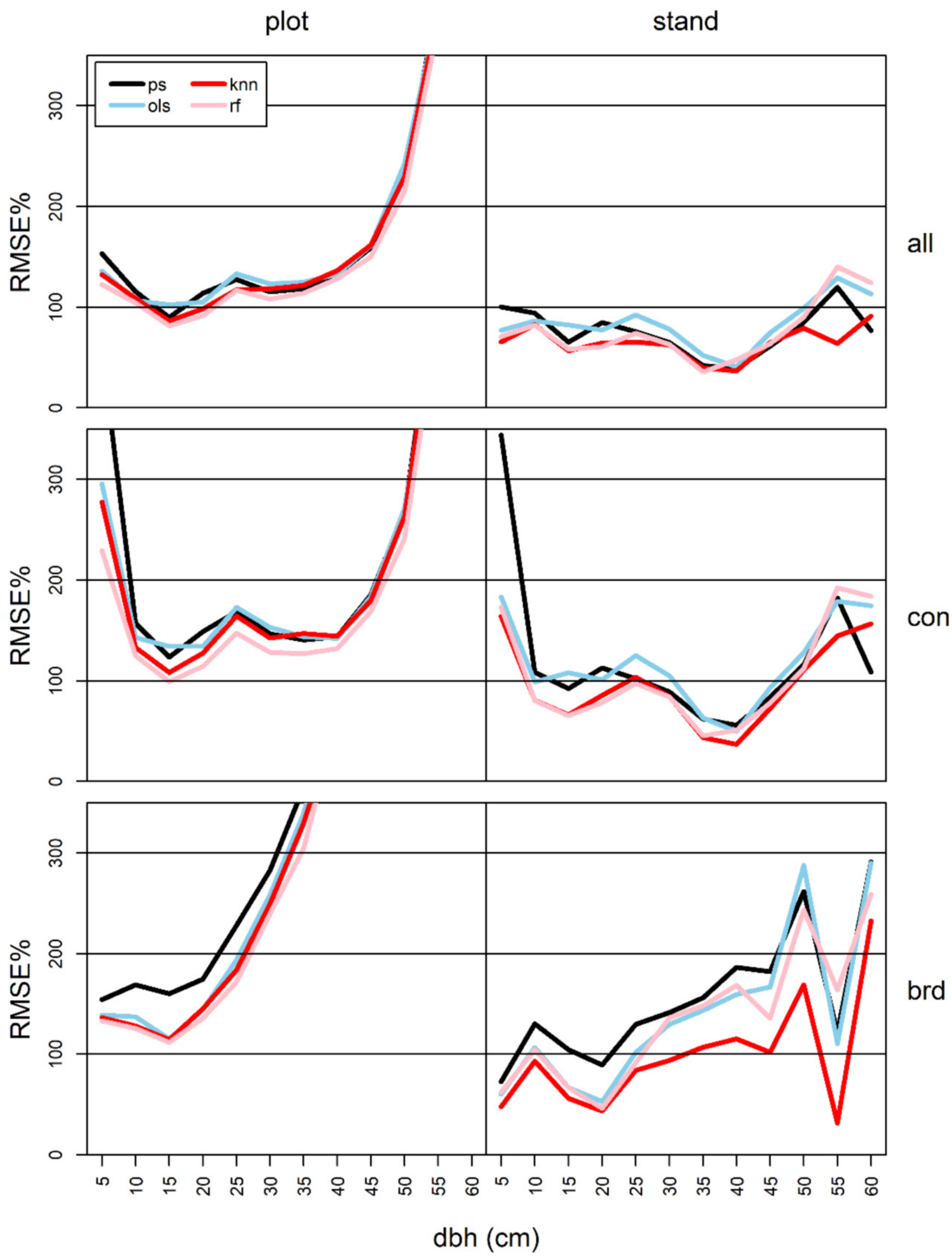


numerous other studies which demonstrate that lidar has good predictive performance for volume (Maclean and Krabill 1986; Næsset 1997*b*; Means et al. 2000; Cosenza et al. 2021), biomass (Nelson et al. 1988; Lefsky et al. 1999; Andersen and Breidenbach 2007), heights (Clark et al. 2004; Andersen et al. 2006; Breidenbach et al. 2008), trees per unit area (Strunk et al. 2012), and conifer versus broadleaf (Liang et al. 2007; Hamraz et al. 2019). Some studies which used other methods to evaluate diameters found optimistic results (Breidenbach et al. 2008; Räty et al. 2019; Strunk et al. 2017). The results presented here are distinct from other studies in that they demonstrate performances for a wide range of forest attributes and modeling approaches using independent validation at both plot and stand scales for a southeastern USA site composed of both pine plantations and mixed broadleaf species.

4.1. Resolution effects

In line with the theoretical properties of aggregation, our findings demonstrate that stand-level ABA prediction performances generally exceed those for individual plots or pixels. From an operational perspective, this finding is useful because stand-level inferences (or other spatial groupings) are a common spatial scale of inference for operational forest

Fig. 9. Plot- and stand-level performances (RMSE% relative to the mean) for volume by diameter class for broadleaf and conifer (mix), conifer (con), and broadleaf (brd) trees with diameters smaller than 62.5 cm. Diameter labels on *x*-axis represent mid-points of 5 cm diameter classes.



inventory systems. Our findings suggest that performances found at the pixel level can be used as conservative approximations for performances at the stand level (pixel-level performances underestimated stand-level performances). It is clearly not a perfect approximation, but it is generally considered better to underestimate performance than to overestimate performance. Our findings that stand-level performances are better than plot-level performances agree with those observed in a similar study using both lidar and photogrammetric point clouds (Puliti et al. 2017). Another related study compared multiple scales of data aggregation (Packalen et al. 2019, see p. 197) and found similarly that performance increased when the size of the area of inference was increased.

Another important finding of this study was that plot-level inferences about relative performances among model types were broadly transferable to stand-level performances. At the plot level, RF and *k*NN generally performed the best, with OLS falling slightly behind, and PS last. The fact that PS performed generally poorer than *k*NN, RF, and OLS is unfortunate from an operational perspective, because PS enables simultaneous compatible and design-unbiased estimation. The findings here with respect to plot-level volume predictions with PS and OLS differ slightly from a related study (Strunk et al. 2019) where they found that PS performance slightly exceeded OLS performance for volume. In contrast, we observed that OLS exceeded PS in all cases, although for volume the best PS and OLS results were very similar (35% versus 32%, respectively). Another related study looked at PS methods with auxiliary lidar for simultaneous volume and forest area estimation (McRoberts et al. 2012) and achieved up to 3.2-fold relative efficiency for volume, which would be a similar but slightly lower efficiency than would be observed here (approximately 10% different). The difference is likely due to their use of smaller plots (250 m^2) than those used in this study ($\geq$ 405 m^2), where it has been shown that plot-level performance statistics improve with increased plot areas (Packalen et al. 2019).

4.2. Modeling

OLS is a common modeling tool across a wide range of disciplines, and AVS on the training data are a common variable selection approach. Our results suggest, however, that a fixed set of four variables performs better than AVS when evaluated on independent stand data. OLS performed better than PS on average but poorer than *k*NN and RF for both gross attributes and for volume by diameter class. It is worth noting, however, that OLS models may be fit with reasonable performance for as few as, e.g., 50 observations (Strunk et al. 2012; Cosenza et al. 2022), which is fewer than is typically considered good practice for the three other methods considered here. An exploration of other linear modeling approaches such as "elastics net" and related model types which shrink linear model coefficients towards zero for weak predictors, is another approach to handling many predictor variables which deserves further exploration (e.g., Pearse et al. 2018; Cosenza et al. 2022).

The best *k*NN configurations fared well at both plot and stand resolutions, either performing the best or equivalent to the best in nearly every instance. However, there was great variation among *k*NN results depending upon the distance metric and number of neighbors. In our experience here, and in related studies modeling forest attributes from lidar, plot-level performance varies dramatically by choice of predictors (Packalén et al. 2012), number of neighbors (Strunk et al. 2017), and distance metrics (Strunk et al. 2014; Shin et al. 2016). This is problematic from the perspective that the "best" configuration is uncertain, may vary by site and data set, and practitioners may not be prepared to compare and interpret results for various modeling approaches at plot and stand scales. An advantage of *k*NN, however, is that it is multivariate (fit a single model for all variables) and can be used simply to impute an entire tree list, along with predictions of the forest attributes used to fit the *k*NN model. Given that many applications of forest inventory and monitoring involve growth projections, and growth models generally require individual-tree data, there is a real practical advantage to *k*NN imputed tree lists.

From the perspectives of supporting robust model fits and reliable predictions, RF was the clear winner among modeling approaches. RF did not always perform the absolute best, and in the case of stand-level diameter predictions fell in the middle of the pack, but RF provided consistent competitive results regardless of how the models were fit, including different numbers of predictors and RF fitting parameters. This is consistent with the related study by Cosenza et al. (2022) which showed that results for RF volume models with auxiliary lidar were insensitive to the number of predictors and number of training observations when using 200 or more observations. RF is a bagging (or "**b**ootstrap **agg**regat**ing**") approach using Classification And Regression Trees (CART; Breiman 2001)—RF essentially uses the weighted average of bootstrap CART models for prediction. Breiman (1996) suggests that bagging aids with predictions in cases where the prediction function is inherently unstable, perhaps explaining why RF performed consistently well for our study; the high level of variation in results among other methods (PS, OLS, and *k*NN) suggests that those approaches were unstable. Breiman (1996) also demonstrates bagging methods for *k*NN and OLS which may have improved their reliability, but was not tested here.

4.3. Lidar inventory value

A motivation for using lidar to support forest inventory is that it can be used to create a wall-to-wall forest inventory update with high stand-level precision for many attributes (e.g., RMSE < 30%), for a relatively modest cost relative to a traditional inventory system. The material cost for this investigation (excluding validation data) was approximately US$ 190 000 including US$ 60 000 for lidar data and US$ 130 000 for the field measurements, or US$ 2.4/ha (US$ 0.94/acre). The resulting inventory provides actionable stand-level information for volume and other gross attributes.

Stand-level volume estimates in operational forest inventories are often derived from 10 to 30 plots per stand, which would be equivalent to 80 000–240 000 plots for the 8000 + stands at Savannah River. At US$ 40 per variable-radius field plot, field plots for a complete stand sample inventory would cost between US$ 3.2 and 9.6 million, or at least US$ 40/ha (US$ 16/acre). While the stand-level estimates from cruise data would typically be more accurate than lidar estimates for the same stands, a stand sampling inventory would cost at least 17 times more than the lidar inventory. Further work is also needed to understand the potential of lidar to contribute to predictions and estimates of components (species, diameter, products) using methods such as ITD. Direct comparisons with operational cruise data are problematic in our experience, however, as cruise specifications may cause wide divergence from lidar estimates, simply due to different measurement and sampling protocols (e.g., leave trees,

 Can. J. For. Res. **53:** 71–89 (2023) | dx.doi.org/10.1139/cjfr-2022-0032

cull, diameter thresholds, stand borders, measurement intensity, sampling intensity, and other differences).

Another approach to inventory which is relevant to this study is stratified random sampling in which, for example, approximately 30 plots are placed in each of an estimated 100 different strata based on the size and species composition inferred from aerial photography (or approximately 3000 plots). This approach also requires collection of and delineation of imagery for the site (approximately US$ 80 000) which brings the total to US$ 200 000, or very similar to the cost of a lidar-based inventory. The strata estimates can be assigned to stands, but they would not have stand-specific measures of forest structure (height, cover, and predicted volume); every stand in a stratum would have identical attributes. As with stand inventories, we are uncertain as to the relative performances of lidar- and stratum-based approaches.

4.4. Study limitations

A limitation of our assessment of PS is that we restricted the number of auxiliary variables for PS relative to other methods. For example, with OLS we considered from 4 to 24 auxiliary variables and both manual and AVS including an intensity derived metric which is indicative of broadleaf proportion. An approach to PS which facilitates more variables could potentially improve PS performance, especially since we did not consider broadleaf proportion as a covariate for PS. While more sophisticated approaches involving more diverse variables with PS are technically feasible, they are complicated by the fact that PS is more readily confounded by the "curse of dimensionality" (Berchtold et al. 1998; Indyk and Motwani 1998) than the other methods used. This is because the data are binned in PS, and there must be enough observations in each bin (stratum) to provide a reliable bin estimate. For example, if we require five observations per bin to produce a reliable bin estimate, and we bin two covariates into 10 bins each, this requires a minimum of 500 training observations. If we add a third covariate with 10 bins, this requires 5000 training observations.

Our use of PS with modeled forest attributes derived from more than two auxiliary variables did not demonstrate better performance than simple PS on height and cover. The modeled attributes were effective at explaining variation in the corresponding measured forest attribute (e.g., vol) but had less explanatory power than for other attributes (e.g., dens). More creative approaches to variable coordination such as principal components, canonical correlation analysis, or canonical correspondence analysis may help to incorporate additional explanatory information. Preparing separate post-strata for each response would yield better performance than was demonstrated here, but was not considered operationally viable, as one of the primary motivations to use PS is the simultaneous estimation of all attributes. Preparing separate post-strata would eliminate this advantage.

5. Conclusions

In this study, we demonstrate that stand-level RMSEs are generally smaller than plot- (or pixel-) level RMSEs from a lidar-augmented forest inventory — inventory consumers are often unsure whether stand-level performances are better or worse than the reported plot- and pixel-level performance statistics. Lidar was also demonstrated to have the capacity to predict some of the variation in forest attributes for broadleaf and conifer species groups. The area-based methods that we employed fared poorly in the prediction of individual diameter bins but could still be useful for larger-diameter bins (e.g., we speculated that perhaps 5 cm bins would be acceptable). Modeling performances were compared between PS, OLS, *k*NN. and RF, with RF arguably demonstrating the best overall performance, as it had nearly the best RMSE and was insensitive to the modeling configuration. Problematically, other modeling approaches were sensitive to changes in modeling configurations, which could be difficult to overcome without stand-level validation. While *k*NN was more sensitive than RF to parameterization, it performed similarly to RF in the best case, and is more convenient operationally than the other methods because it can be used to impute many variables simultaneously, including tree lists. Our study is a unique contribution to the understanding of forest inventory with lidar, providing insight into the scale of interest (plot or pixel versus stand), how lidar works across a suite of forest attributes including attributes for diameter classes, and contrasts the modeling performances of four modeling approaches. Our findings can help others to evaluate and implement lidar augmented inventory approaches.

Acknowledgements

We would like to thank Andy Horcher, Hans-Erik Andersen, Marin Palmer, and Tim Bryant for feedback on manuscript drafts and Jim Flewelling for feedback on RMSE computation. We would also like to thank Jason Jeffcoat and Jeffcoat Forestry for field measurement collection and Stefani Ryan, Graham Zimmerman, Mike Jonsson, and Paul Chisholm for plot installation and GNSS data collection for this project.

Article information

History dates

Received: 8 February 2022
Accepted: 5 October 2022
Version of record online: 5 January 2023

Copyright

Data availability

Data generated or analyzed during this study are available upon reasonable request from Jeff W. Atkins (jeffrey.atkins@usda.gov). Requests may require a high degree of scrutiny, due to the confidential nature of the US Department of Energy site.

Author information

Author ORCIDs

Jacob L. Strunk https://orcid.org/0000-0002-5114-4843

Author contributions

Conceptualization: JS, RM
Data curation: JS, RM
Formal analysis: JS
Funding acquisition: RM
Investigation: JS
Methodology: JS, RM
Project administration: JS, RM
Resources: RM
Software: RM
Supervision: JS, RM
Validation: JS
Visualization: JS, RM
Writing – original draft: JS, RM
Writing – review & editing: JS, RM

Competing interests

There are no competing interests for any of the authors or their institutions.

Funding information

Funding and support for this project was provided by the Department of Energy, Aiken, SC, to the Department of Agriculture, Forest Service, Savannah River under Interagency Agreement DE-AI09-EM0003622.

References

Andersen, H.-E., and Breidenbach, J. 2007. Statistical properties of mean stand biomass estimators in a LIDAR-based double sampling forest survey design. *In* Proceedings of the ISPRS Workshop Laser Scanning 2007 and SilviLaser 2007. ISPRS, Espoo, Finland. pp. 8–13.

Andersen, H.-E., Reutebuch, S.E., and McGaughey, R.J. 2006. A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Can. J. Remote Sens. **32**(5): 12. doi:10.5589/m06-030.

Bellman, R. 1966. Dynamic programming. Science, **153**(3731): 34–37. doi:10.1126/science.153.3731.34. PMID: 17730601.

Berchtold, S., Böhm, C., and Kriegal, H.-P. 1998. The pyramid-technique: towards breaking the curse of dimensionality. *In* Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, SIGMOD '98. 1–4 June 1998. Association for Computing Machinery, New York, USA. pp. 142–153.

Breidenbach, J., Gläser, C., and Schmidt, M. 2008. Estimation of bivariate diameter and height distributions using ALS. *In* Proceedings of SilviLaser 2008 Conference. 17–19 September 2008. Heriot-Watt University, Edinburgh, UK. pp. 366–372.

Breidenbach, J., Magnussen, S., Rahlf, J., and Astrup, R. 2018. Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data. Remote Sens. Environ. **212**: 199–211. doi:10.1016/j.rse.2018.04.028.

Breidt, F.J., and Opsomer, J.D. 2008. Endogenous post-stratification in surveys: classifying with a sample-fitted model. Ann. Stat. **36**(1): 403–427. doi:10.1214/009053607000000703.

Breiman, L. 1996. Bagging predictors. Mach. Learn. **24**(2): 123–140. doi:10.1007/BF00058655.

Breiman, L. 2001. Random forests. Mach. Learn. **45**(1): 5–32. doi:10.1023/A:1010933404324.

Clark, M.L., Clark, D.B., and Roberts, D.A. 2004. Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sens. Environ. **91**(1): 68–89. doi:10.1016/j.rse.2004.02.008.

Cosenza, D.N., Korhonen, L., Maltamo, M., Packalen, P., Strunk, J.L., Næsset, E., et al. 2021. Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. Forestry, **94**(2): 311–323.

Cosenza, D.N., Packalen, P., Maltamo, M., Varvia, P., Räty, J., Soares, P., et al. 2022. Effects of numbers of observations and predictors for various model types on the performance of forest inventory with airborne laser scanning. Can. J. For. Res. **52**(3): 385–395. doi:10.1139/cjfr-2021-0192.

Crookston, N.L., and Finley, A.O. 2008. yaImpute: an R package for kNN imputation. Journal of Statistical Software, **23**: 1–16.

Efron, B. 2020. Prediction, estimation, and attribution. J. Am. Stat. Assoc. **115**(530): 636–655. doi:10.1080/01621459.2020.1762613.

Erdody, T.L., and Moskal, L.M. 2010. Fusion of LiDAR and imagery for estimating forest canopy fuels. Remote Sens. Environ. **114**(4): 725–737. doi:10.1016/j.rse.2009.11.002.

Furnival, G.M., and Wilson, R.W., Jr. 1974. Regressions by leaps and bounds. Technometrics, **16**(4): 499–511. doi:10.1080/00401706.1974.10489231.

Gobakken, T., and Næsset, E. 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. Can. J. For. Res. **38**(5): 1095–1109. doi:10.1139/X07-219.

Green, P.C., Hogg, D.W., Watson, B., and Burkhart, H.E. 2022. Small area estimation in diverse timber types using multiple sources of auxiliary data. J. For. **120**(6):646–659.

Hamraz, H., Jacobs, N.B., Contreras, M.A., and Clark, C.H. 2019. Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. ISPRS **158**: 219–230. doi:10.1016/j.isprsjprs.2019.10.011.

Holmgren, J. 2004. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. Scand. J. For. Res. **19**(6): 543–553. doi:10.1080/02827580410019472.

Hudak, A.T., Evans, J.S., and Stuart Smith, A.M. 2009. LiDAR utility for natural resource managers. Remote Sens. **1**(4): 934–951. doi:10.3390/rs1040934.

Indyk, P., and Motwani, R. 1998. Approximate nearest neighbors: towards removing the curse of dimensionality. *In* Proceedings of the 30th Annual ACM Symposium on Theory of Computing, STOC '98. 24–26 May 1998. Association for Computing Machinery, New York, USA. pp. 604–613.

Jeronimo, S.M., Churchill, D.J., Mcgaughey, R., and Franklin, J.F. 2018. Applying LiDAR individual tree detection to management of structurally diverse forest landscapes. J. For. **116**: 336–346.

Justice, C., and Townshend, J. 1982. A comparison of unsupervised classification procedures on landsat MSS data for an area of complex surface conditions in Basilicata, Southern Italy. Remote Sens. Environ. **12**(5): 407–420. doi:10.1016/0034-4257(82)90016-5.

Kim, Y., Yang, Z., Cohen, W.B., Pflugmacher, D., Lauver, C.L., and Vankat, J.L. 2009. Distinguishing between live and dead standing tree biomass on the north rim of Grand Canyon National Park, USA using small-footprint lidar data. Remote Sens. Environ. **113**(11): 2499–2510. doi:10.1016/j.rse.2009.07.010.

Korhonen, L., Korpela, I., Heiskanen, J., and Maltamo, M. 2011. Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. Remote Sens. Environ. **115**(4): 1065–1080. doi:10.1016/j.rse.2010.12.011.

Kotivuori, E., Maltamo, M., Korhonen, L., Strunk, J.L., and Packalen, P. 2021. Prediction error aggregation behaviour for remote sensing augmented forest inventory approaches. Forestry: Int. J. For. Res. **94**(4): 576–587. doi:10.1093/forestry/cpab007.

Latham, R.P., and McCarty, T.M. 1972. Recent developments in remote sensing for forestry. Journal of Forestry, **70**(7): 398–402. doi:10.1093/jof/70.7.398.

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., and Shugart, H.H. 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA. Remote Sens. Environ. **67**(1): 83–98. doi:10.1016/S0034-4257(98)00071-6.

Liang, X., Hyyppä, J., and Matikainen, L. 2007. Deciduous-coniferous tree classification using difference between first and last pulse laser signatures. IAPRS, **36**(3/W52): 253–257.

Lin, Y., and Hyyppä, J. 2016. A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. Int. J. Appl. Earth Obs. Geoinf. **46**: 45–55. doi:10.1016/j.jag.2015.11.010.

Maclean, G.A., and Krabill, W.B. 1986. Gross-merchantable timber volume estimation using an airborne LIDAR system. Can. J. Remote Sens. **12**: 7–18. doi:10.1080/07038992.1986.10855092.

Magnussen, S., and Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can. J. For. Res. **28**(7): 1016–1031. doi:10.1139/x98-078.

Magnussen, S., Andersen, H.-E., and Mundhenk, P. 2015. A second look at endogenous poststratification. For. Sci. **61**(4): 624–634.

Magnussen, S., Nord-Larsen, T., and Riis-Nielsen, T. 2018. Lidar supported estimators of wood volume and aboveground biomass from the Danish National Forest Inventory (2012–2016). Remote Sens. Environ. **211**: 146–153. doi:10.1016/j.rse.2018.04.015.

Maltamo, M., and Packalen, P. 2014. Species-specific management inventory in Finland. *In* Forestry applications of airborne laser scanning: concepts and case studies, managing forest ecosystems. *Edited by* M. Maltamo, E. Næsset and J. Vauhkonen. Springer, Dordrecht, the Netherlands. pp. 241–252.

Maltamo, M., Bollandsås, O.M., Næsset, E., Gobakken, T., and Packalén, P. 2011. Different plot selection strategies for field training data in ALS-assisted forest inventory. Forestry, **84**(1): 23–31.

McGaughey, R.J. 2018. FUSION/LDV: software for LIDAR data analysis and visualization. Version 3.8 [computer program]. USFS. Available from http://forsys.sefs.uw.edu/fusion/fusionlatest.html.

McRoberts, R.E., Gobakken, T., and Næsset, E. 2012. Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. Remote Sens. Environ. **125**: 157–166. doi:10.1016/j.rse.2012.07.002.

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., and Hendrix, C. 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogramm. Eng. Remote Sens. **66**(11): 1367–1371.

Næsset, E. 1997a. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS J. Photogramm. Remote Sens. **52**(2): 49–56. doi:10.1016/S0924-2716(97)83000-6.

Næsset, E. 1997b. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens. Environ. **61**(2): 246–253. doi:10.1016/S0034-4257(97)00041-2.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. **80**(1): 88. doi:10.1016/S0034-4257(01)00290-5.

Næsset, E. 2007. Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. Scand. J. For. Res. **22**(5): 433–442. doi:10.1080/02827580701672147.

Næsset, E. 2014. Area-based inventory in Norway—from innovation to an operational reality. *In* Forestry applications of airborne laser scanning. Springer, Dordrecht, Netherlands. pp. 215–240.

Nelson, R., Krabill, W., and Tonelli, J. 1988. Estimating forest biomass and volume using airborne laser data. Remote Sens. Environ. **24**(2): 247–267. doi:10.1016/0034-4257(88)90028-4.

Ørka, H.O., Næsset, E., and Bollandsås, O.M. 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sens. Environ. **113**(6): 1163–1174. doi:10.1016/j.rse.2009.02.002.

Packalén, P., and Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sens. Environ. **109**(3): 328–341. doi:10.1016/j.rse.2007.01.005.

Packalén, P., Temesgen, H., and Maltamo, M. 2012. Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. Can. J. Remote Sens. **38**(5): 557–569. doi:10.5589/m12-046.

Packalen, P., Strunk, J., Packalen, T., Maltamo, M., and Mehtätalo, L. 2019. Resolution dependence in an area-based approach to forest inventory with airborne laser scanning. Remote Sens. Environ. **224**: 192–201. doi:10.1016/j.rse.2019.01.022.

Pearse, G.D., Dash, J.P., Persson, H.J., and Watt, M.S. 2018. Comparison of high-density LiDAR and satellite photogrammetry for forest inventory. ISPRS J. Photogramm. Remote Sens. **142**: 257–267. doi:10.1016/j.isprsjprs.2018.06.006.

Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J., and Packalén, P. 2007. Preharvest measurement of marked stands using airborne laser scanning. For. Sci. **53**(6): 653–661.

Puliti, S., Gobakken, T., Ørka, H.O., and Næsset, E. 2017. Assessing 3D point clouds from aerial photographs for species-specific forest inventories. Scand. J. For. Res. **32**(1): 68–79. doi:10.1080/02827581.2016.1186727.

R Core Team. 2020. R: a language and environment for statistical computing. Available from https://www.r-project.org/ [accessed7 January 2020].

Rao, J.N.K. 2003. Small area estimation. Wiley-Interscience, Hoboken, New Jersey, USA.

Räty, J., Packalen, P., Kotivuori, E., and Maltamo, M. 2019. Fusing diameter distributions predicted by an area-based approach and individual-tree detection in coniferous-dominated forests. Can. J. For. Res. **50**(2): 113–125. doi:10.1139/cjfr-2019-0102.

Shin, J., Temesgen, H., Strunk, J.L., and Hilker, T. 2016. Comparing modeling methods for predicting forest attributes using LiDAR metrics and ground measurements. Can. J. Remote Sens. **42**(6): 739–765. doi:10.1080/07038992.2016.1252908.

Strunk, J., Temesgen, H., Andersen, H.-E., Flewelling, J.P., and Madsen, L. 2012. Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables. Can. J. Remote Sens. **38**(05): 644–654. doi:10.5589/m12-052.

Strunk, J.L., Temesgen, H., Andersen, H.-E., and Packalen, P. 2014. Prediction of forest attributes with field plots, landsat, and a sample of lidar strips. Photogramm. Eng. Remote Sens. **80**(2): 143–150. doi:10.14358/PERS.80.2.143-150.

Strunk, J.L., Gould, P.J., Packalen, P., Poudel, K.P., Andersen, H.-E., and Temesgen, H. 2017. An examination of diameter density prediction with k-NN and airborne lidar. Forests, **8**(11): 444. doi:10.3390/f8110444.

Strunk, J., Packalen, P., Gould, P., Gatziolis, D., Maki, C., Andersen, H.-E., and McGaughey, R.J. 2019. Large area forest yield estimation with pushbroom digital aerial photogrammetry. Forests, **10**(5): 397. doi:10.3390/f10050397.

Temesgen, H., and Ver Hoef, J.M. 2015. Evaluation of the spatial linear model, random forest and gradient nearest-neighbour methods for imputing potential productivity and biomass of the pacific northwest forests. For. Int. J. For. Res. **88**(1): 131–142.

Temesgen, B.H., LeMay, V.M., Froese, K.L., and Marshall, P.L. 2003. Imputing tree-lists from aerial attributes for complex stands of southeastern British Columbia. For. Ecol. Manag. **177**(1–3): 277–285. doi:10.1016/S0378-1127(02)00321-3.

US Forest Service (USFS). 2020. National Volume Estimator Library (NVEL). Avaliable from https://www.fs.usda.gov/forestmanagement/products/measurement/volume/nvel/index.php [accessed 23 September 2020].

Vauhkonen, J., Tokola, T., Packalen, P., and Maltamo, M. 2009. Identification of scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. For. Sci. **55**(1): 37–47.

Ver Hoef, J.M., and Temesgen, H. 2013. A comparison of the spatial linear model to nearest neighbor (k-NN) methods for forestry applications. PLoS ONE, **8**(3): e59129. doi:10.1371/journal.pone.0059129. PMID: 23527110.

Wang, Y. 2014. National Biomass Estimator Library.For. Manag. Serv. Cent. Available from https://www.fs.usda.gov/forestmanagement/products/measurement/biomass/index.php [accessed 3 January 2022].

Weisberg, S. 2005. Applied linear regression. John Wiley & Sons, Hoboken, New Jersey, USA.

White, J.C., Wulder, M.A., and Buckmaster, G. 2014. Validating estimates of merchantable volume from airborne laser scanning (ALS) data using weight scale data. For. Chron. **90**(03): 378–385. doi:10.5558/tfc2014-072.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., and Treitz, P. 2011. Operational implementation of a LiDAR inventory in Boreal Ontario. For. Chron. **87**(4): 512–528. doi:10.5558/tfc2011-050.