Forestry *An International Journal of Forest Research*

Institute of Chartered Foresters

Forestry 2023; **96**, 49–61, https://doi.org/10.1093/forestry/cpac032
Advance Access publication 2 September 2022

Circular or square plots in ALS-based forest inventories—does it matter?

Petteri Packalen[1,*], Jacob Strunk[2], Matti Maltamo[3] and Mari Myllymäki[1]

[1]*Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Latokartanonkaari 9, FI-00790 Helsinki, Finland*
[2]*USDA Forest Service, 3625 93rd Ave SW, Olympia, WA 98512, USA*
[3]*School of Forest Sciences, University of Eastern Finland, PO Box 111, 80101 Joensuu, Finland*

*Corresponding author E-mail: petteri.packalen@luke.fi

Received 6 May 2022; accepted 25 July 2022

In airborne laser scanning (ALS)-based forest inventories, there is commonly a discrepancy between the plot shape used for model fitting (typically circular) and the shape of population elements (typically square) where predictions are needed. Circular plots are easy to establish, locate and have the smallest number of edge trees on average. Therefore, a circle is the most common plot shape in both traditional and remote sensing-based forest inventories. In contrast, the shape of population elements used for remote sensing-based predictions is nearly always a square because it enables division of the target population into a grid of non-overlapping plots. In this study, we investigate shape effects for ALS-based forest inventories using circular and square plot shapes. This has not been examined earlier. Aboveground biomass was used as the response variable. The sampling design was created in a way that the probability of selection for any location inside a stem-mapped 30 m × 30 m plot was the same for the circular (radius 7.95 m) and square (side length 14.09 m) plot. This configuration enabled us to compare circular and square plots with the same areas and identical sampling probabilities for every tree in the population. Our primary finding is that for equal area square and circular plots, there is no evidence of systematic prediction error when a model fitted to one shape is used to predict for the other shape. Our secondary finding is that root mean square error (RMSE) value is slightly underestimated (1.2 per cent) when a model fitted to circular plots is used to predict for square plots. A small underestimation of RMSE due to plot shape effect has hardly practical significance in stand-level forest management inventories, but the plot shape effect may be problematic in large area forest surveys.

Introduction

A circular sample plot shape is commonly used in forest inventories. It has several advantages as it is easy to establish, can be readily used to determine whether a tree is inside the plot or not and contains less edge trees than other plots shapes with the same area (Kangas and Maltamo, 2006). In remote sensing applications, the advantage of a circular plot is that only the centre point needs to be located (Schreuder *et al*., 1993), which makes plot positioning rapid and straightforward to implement without, for example, the need to verify right-angled corners. In remote sensing inventories, the shape of the plots used to fit a model (typically circular) often differs from the shape of population elements used for prediction or estimation where the inventory area is typically divided into parts without gaps (i.e. tessellated) using a grid of square plots that cover the entire population. This approach is also used with an area-based approach (ABA) to forest inventories when utilizing airborne laser scanning (ALS) data (Næsset, 2002).

Several plot characteristics that affect the performance of ABA have been studied, such as the sample plot size (Gobakken and Næsset, 2008; Frazer *et al*., 2011), the accuracy of variable radius plots versus fixed area plots (Maltamo *et al*., 2007; Deo *et al*., 2016), the use of subplot configuration, such as the Forest Inventory and Analysis plot design of the USDA Forest Service (Andersen *et al*., 2011; Gopalakrishnan *et al*., 2015) and the use of different plot sizes in fitting and prediction (Zhao *et al*., 2009; Packalen *et al*., 2019). In general, the prediction error decreases with larger plot sizes due to spatial averaging of errors (Goetz and Dubayah, 2011; Fassnacht *et al*. 2018; Packalen *et al*., 2019). Many studies have shown that sample plots must be positioned accurately (Mauro *et al*., 2011; Fadili *et al*., 2019) and that terrain slope has a minimal effect on root mean square error (RMSE) values (Breidenbach *et al*., 2008; Ørka *et al*., 2018). Studies have also explored how many sample plots are needed in an ABA inventory (Strunk *et al*., 2012; Junttila *et al*., 2013), the effect of plot selection strategy (Ruotsalainen *et al*., 2019; Nilsson *et al*., 2017) and how best to use existing ALS data in plot selection (Hawbaker *et al.*, 2009; Gobakken *et al.*, 2013).

In remote sensing-based forest inventories, trees located near to the plot edges strongly contribute to prediction errors due to differences between what is measured by remote sensing and how we define a tree that falls within a sample plot (Packalen *et al.*, 2015; Knapp *et al.*, 2021). An established practice is to designate a tree as falling within a sample plot if the centre of the trunk is located inside the plot. However, this causes inconsistency in two ways: (1) a trunk falls inside a plot, but some portion of the crown observed by remote sensing falls outside the plot, or (2) a remotely sensed tree crown falls partially within the plot, yet the trunk is located outside the plot. The ratio of edge to interior trees, and hence edge errors, is affected by the plot size and shape, and by tree crown size. The smaller the plot compared with the tree crowns, the larger the proportion of trees found on the plot edge. Since a circular plot has the smallest perimeter area ratio of any two-dimensional plot shape, it will also have the fewest errors resulting from edge trees. For example, for plots with the same area, the length of the perimeter in a square plot is $2/\sqrt{\pi} \approx 1.13$ times longer than the perimeter of a circular plot.

In the prediction phase, square-shaped population elements have an advantage in that they can be placed next to each other to ensure that there are no gaps between the units. This arrangement is a straightforward and, seemingly optimal, way of tessellating the population. However, if a model is fitted with circular plots and the prediction is carried out with square-shaped population elements (plots), then there is clearly a mismatch in plot characteristics between the fitting and prediction processes. In this paper, we ask whether the shape mismatch for circular and square plots of the same area really matters and has practical implications.

Hence, our aim is to reveal possible issues that arise from using square or circular plots to relate ABA metrics and forest attributes, and then predict for the alternate plot shape. The corresponding specific research questions are:

1) Do the parameter estimates for models of the same form differ when fitted with circular and square plots of the same area?
2) How do RMSE and mean difference (MD) values vary when predicting for the same and different plot shapes?
3) Does an increase in the expected number of edge trees on square plots explain the larger RMSE values relative to circular plots?

We used empirical methods to gain insights into research questions by comparing prediction performances for different combinations of fitting and prediction shapes (square and circular) with the same plot area.

Materials

Study area

The study was implemented in a boreal forest area located in eastern Finland (62°31′N, 29°123′E). The size of the study area is about 40 000 ha. The main tree species in the region are Norway spruce (*Picea abies* [L.] Karst.) and Scots pine (*Pinus sylvestris* L.). Broadleaved tree species are in a minority. Most common broadleaved species are downy birch (*Betula pubescens* Ehrh), silver birch (*Betula pendula* Roth) and European aspen (*Populus tremula* L.). Nearly all stands contain trees of several tree species.

Table 1 Mean, standard deviation (SD), minimum and maximum AGB at the tree-level and the 30 m × 30 m plot-level

	Mean	SD	Min	Max
Tree AGB (kg)	101.7	133.5	3.2	1262.8
Plot AGB (Mg·ha^{-1})	110.6	45.4	38.3	254.1

Field data

We used field measurements from 110 square sample plots (30 m × 30 m) distributed over the study area. Plot locations were selected to ensure an appropriate representation of the main tree species and development classes in the area. Two or more plots were never placed in the same stand. Seedling and sapling stands (mean height < 7 m) were excluded for reasons not related to this study. Field measurements were conducted in summer 2017. A more detailed description of the sampling design can be found in Kotivuori *et al*. (2021).

We measured diameter at breast height, height, position and species for all trees with a diameter ≥ 5 cm. Tree positions were determined with the help of ALS data collected before field measurements. First, we detected clearly visible treetops from the ALS-based canopy height model (CHM) using the principle of local maxima. Most locations of CHM treetops were also used as tree locations. Locations of trees not detected or erroneously positioned according to CHM were determined using the protocol described by Korpela *et al*. (2007). This tree positioning procedure uses CHM-derived tree locations as reference locations, and missing tree locations were determined using the distances and azimuths to the nearest reference locations. Aboveground biomass (AGB) of the individual trees was predicted as a function of diameter at breast height and tree height by tree species using the models presented in Repola (2008) and Repola (2009). Plot level AGB per hectare (Mg·ha^{-1}) for a single plot was predicted as the sum of area-expanded tree biomass in the plot. A summary of AGB at tree- and plot-levels is provided in Table 1. The proportions of tree species with respect to AGB were 43 per cent, 37 per cent and 20 per cent for spruce, pine and deciduous trees, respectively.

ALS data

ALS data were collected using an Optech Titan instrument on 2–10 July 2016. Optech Titan is a multispectral airborne LiDAR system that provides measurements from three different wavelengths, although only the 1064 nm channel was used in this study. The data were collected using a fixed wing aircraft operating at 850 m above ground level. The half angle was 20°, which resulted in a 650 m nominal swath width on the ground. The lateral overlap between flight lines was 55 per cent, which means that each location was visible from at least two flight lines. Pulse repetition frequency was set to 250 kHz. This configuration resulted in an average pulse density of 14 pulses m^{-2}.

Terrain level was estimated by initially classifying points as ground and non-ground hits by the method described in

Axelsson (2000). A raster terrain model of 0.5 m resolution was then interpolated by Delaunay triangulation (Okabe *et al*. 1992). Echo heights were scaled to aboveground level by subtracting terrain model from the corresponding location. Hereafter, we refer to aboveground height as echo height. ALS data processing routines were implemented using in-house C/C++ programs.

Methods

Comparison of circular and square plots

Our analysis is based on simulations composed of randomly located square and circular plots placed entirely within the corner-modified 30 m × 30 m plots. Since we have tree positions for all trees in the 30 m × 30 m plots, we can identify the trees that fall within each simulated plot. The plot sampling design was created in such a way that the probability of selection for any location inside the corner-modified 30 m × 30 m plot is the same for circular and square plots, although the probability of selection varies within the corner-modified 30 m × 30 m plot. This is explained in detail in the section 'Probability of selection'. To obtain equal probabilities of selection, the radius of the circular plot was 7.95 m and the side length of the square plot was 14.09 m, yielding equal plot areas of 198.56 m^2. This configuration enables us to compare the circular and square plots with the same areas and the identical sampling probabilities for every tree in the population. The main steps in our comparison were as follows:

1. One circular plot location and one square plot location were randomly placed within each corner-modified 30 m × 30 m plot (Figure 1), such that a plot was entirely located inside the corner-modified 30 m × 30 m plot.
2. Plot AGB and ALS metrics were computed for square and circular plots using the known tree locations and plot boundaries for simulated plot locations.
3. Separate regression models were fitted to circular plots ($n = 110$) and square plots ($n = 110$). Plot AGB was the response variable and ALS metrics were predictor variables.
4. New circular ($n = 110$) and square plots ($n = 110$) were established as explained in step 1, and plot AGB and ALS metrics were computed as outlined in step 2.
5. Both models fitted in step 3 were used to predict plot AGB for the new circular and square plots outlined in step 4.
6. RMSE and MD values were computed for the predictions made in step 5.
7. Steps 1–6 were repeated 100 000 times.

We studied the differences in the fitted models and variables between the circular and square plots (from step 3). For each of the 100 000 replicates, we calculated the RMSE and MD values between the predicted (from step 5) and actual values (from step 4) when predicted for the same plot type and across plot types, and investigated the resulting distributions. Finally, we analysed the effect of edge trees by plot type. Analyses were implemented in the R environment (R Development Core Team, 2011).

Probability of selection

We denote a 30 m × 30 m plot by $W = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, where $x_{\min}$ and $y_{\min}$ define the lower left corner of the plot,

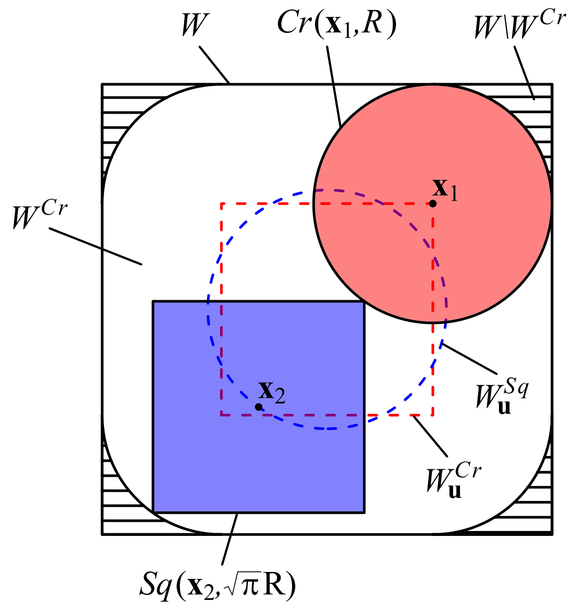

Figure 1 Region W of size 30 m × 30 m with horizontal black lines that designate areas ($W\backslash W^{Cr}$) outside the sample frame. The dashed red square designates the region $W_{\mathbf{u}}^{Cr}$ in which centre $\mathbf{u}$ of the circular sample plot is placed randomly, and the dashed blue circle designates the corresponding region $W_{\mathbf{u}}^{Sq}$ for the square sample plot. The blue square and red circle are example plot placements with their centre points $\mathbf{x_1}$ and $\mathbf{x_2}$ falling on the borders of $W_{\mathbf{u}}^{Cr}$ and $W_{\mathbf{u}}^{Sq}$, respectively, although plot centres may fall anywhere in the respective regions $W_{\mathbf{u}}^{Cr}$ and $W_{\mathbf{u}}^{Sq}$.

and we let $\mathbf{x}_{\text{center}}$ stand for the centre location of W. We use the notation $|\cdot|$ for the area of any region: for example, plot W has area $|W| = 900$ m^2. Furthermore, we use the notation $Cr(\mathbf{u}, R)$ for the circular plot with centre $\mathbf{u}$ and radius $R > 0$, and $Sq(\mathbf{u}, L)$ for the square plot with centre $\mathbf{u}$, side length L and sides parallel to the corresponding sides of W. The areas of the plots are $|Cr(\mathbf{u}, R)| = \pi R^2$ and $|Sq(\mathbf{u}, L)| = L^2$, respectively, for any $\mathbf{u}$. We chose $R = 7.95$ m and $L = \sqrt{\pi}R \approx 14.09$ m, such that the areas of the square and the circular plots were equal ($\sim$198.56 m^2).

We define W^{Cr} as the sub-region of W that can be sampled by a circular plot $Cr(\mathbf{u}, R)$ that falls entirely within W (Figure 1). Its complement, $W\backslash W^{Cr}$, i.e. the sub-region of W that cannot be sampled by the circular plot (Figure 1) has an area $|W\backslash W^{Cr}| = R^2(4-\pi)$. Consequently, W^{Cr} has an area $|W^{Cr}| = |W| - |W\backslash W^{Cr}| \approx 845.75$. We can show that when we restrict square plots $Sq(\mathbf{u}, \sqrt{\pi}R)$ from falling within the corners $W\backslash W^{Cr}$ (to match the extent of the circular plot), the probability to be included in a uniformly randomly placed circular or square sample plot is identical for any location $\mathbf{x}$ in W^{Cr}, although the probability is non-constant in W^{Cr}. In Figure 1, the regions $W_{\mathbf{u}}^{Cr}$ (red dashed square) and $W_{\mathbf{u}}^{Sq}$ (blue dashed circle) denote the regions where the centre $\mathbf{u}$ of the circular and square plots must lie. The region $W_{\mathbf{u}}^{Cr}$ is defined by

$$W_{\mathbf{u}}^{Cr} = \{\mathbf{u} : Cr(\mathbf{u}, R) \in W\}$$
$$= [x_{\min} + R, x_{\max} - R] \times [y_{\min} + R, y_{\max} - R]. \quad (1)$$

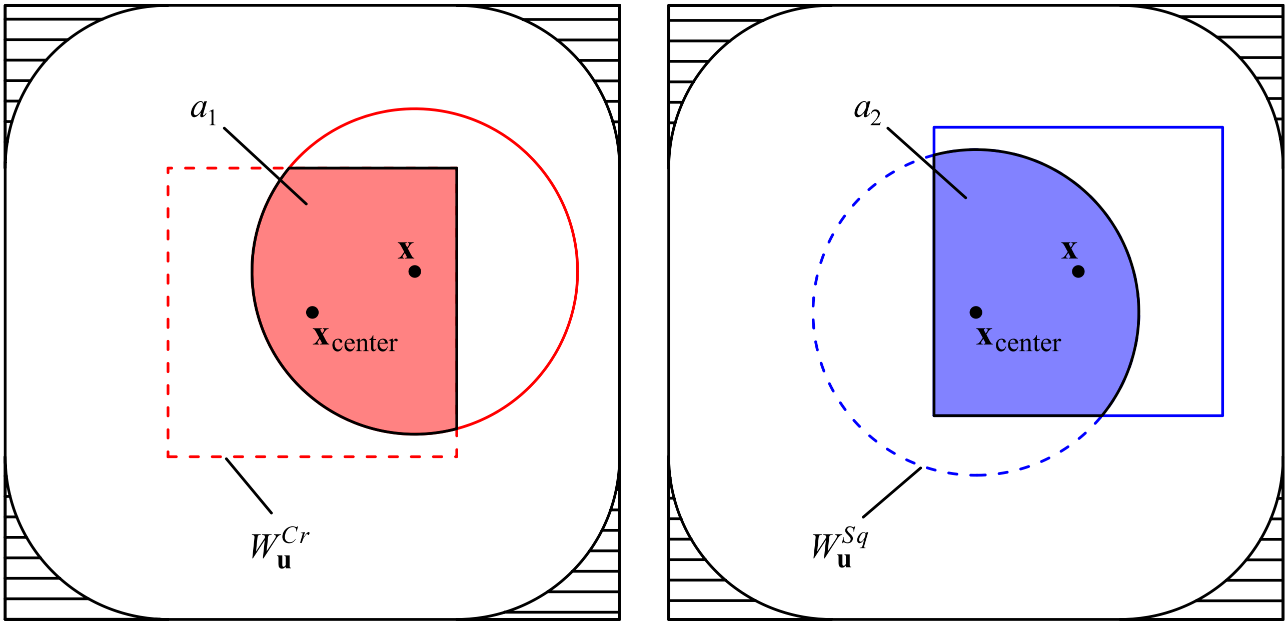

Figure 2 Illustration of the equality of the probabilities that a location $\mathbf{x} \in W^{Cr}$ is selected by a circular (on the left) or square plot (on the right) when $R = 7.95$ m and the centre location $\mathbf{u}$ of the plot is selected randomly in $W_{\mathbf{u}}^{Cr}$ (dashed red square) or $W_{\mathbf{u}}^{Sq}$ (dashed blue circle), respectively. The colored regions, a_1 and a_2, represent the regions where $\mathbf{u}$ needs to lie so that $\mathbf{x}$ is included in the sample plot. The probabilities of selection for $\mathbf{x}$ are simply $|a_1|/|W_{\mathbf{u}}^{Cr}|$ and $|a_2|/|W_{\mathbf{u}}^{Sq}|$ for the circular and square plots, respectively; because $|a_1| = |a_2|$ and $|W_{\mathbf{u}}^{Cr}| = |W_{\mathbf{u}}^{Sq}|$ (see text for details), the probabilities are also equal. The location $\mathbf{x}_{\text{center}}$ is the centre of W, i.e. it is also the centre of $W_{\mathbf{u}}^{Cr}$ and $W_{\mathbf{u}}^{Sq}$.

For our choice of R and L, $W_{\mathbf{u}}^{Sq}$ takes a circular form, as shown in Figure 1. The probability that any location $\mathbf{x}$ in W^{Cr} is included in a circular plot $Cr(\mathbf{u}, R)$ whose centre $\mathbf{u}$ is randomly placed in $W_{\mathbf{u}}^{Cr}$ is simply the proportion of the area in $W_{\mathbf{u}}^{Cr}$ from which it is possible to sample $\mathbf{x}$. This is the proportion of the area of the red square that is shaded red in Figure 2, or more formally, given that $\mathbf{u}$ is uniformly randomly sampled from $W_{\mathbf{u}}^{Cr}$,

$$\pi^{Cr}(\mathbf{x}) = \mathbf{P}(\mathbf{x} \in Cr(\mathbf{u}, R)) = \mathbf{P}\left(\mathbf{u} \in Cr(\mathbf{x}, R) \cap W_{\mathbf{u}}^{Cr}\right) = \frac{|Cr(\mathbf{x}, R) \cap W_{\mathbf{u}}^{Cr}|}{|W_{\mathbf{u}}^{Cr}|} \tag{2}$$

where $\cap$ denotes the intersection. Similarly, for a square plot, the probability of selection is the proportion of the area in $W_{\mathbf{u}}^{Sq}$ that also falls in the square centred on $\mathbf{x}$

$$\pi^{Sq}(\mathbf{x}) = \mathbf{P}\left(\mathbf{x} \in Sq(\mathbf{u}, \sqrt{\pi}R)\right) = \mathbf{P}\left(\mathbf{u} \in Sq(\mathbf{x}, \sqrt{\pi}R) \cap W_{\mathbf{u}}^{Sq}\right) = \frac{\left|Sq(\mathbf{x}, \sqrt{\pi}R) \cap W_{\mathbf{u}}^{Sq}\right|}{\left|W_{\mathbf{u}}^{Sq}\right|}. \tag{3}$$

In fact, for the 30 m $\times$ 30 m plot and $R = \frac{30}{2+\sqrt{\pi}} \approx 7.95$ m, $W_{\mathbf{u}}^{Cr} = Sq(\mathbf{x}_{\text{center}}, \sqrt{\pi}R)$, because for this R it holds that $\sqrt{\pi}R + 2R = 30$. Likewise, $W_{\mathbf{u}}^{Sq}$ equals $Cr(\mathbf{x}_{\text{center}}, R)$ for this R. Therefore,

$$\pi^{Cr}(\mathbf{x}) = \frac{|Cr(\mathbf{x}, R) \cap Sq(\mathbf{x}_{\text{center}}, \sqrt{\pi}R)|}{|Sq(\mathbf{x}_{\text{center}}, \sqrt{\pi}R)|} \text{ and}$$
$$\pi^{Sq}(\mathbf{x}) = \frac{|Sq(\mathbf{x}, \sqrt{\pi}R) \cap Cr(\mathbf{x}_{\text{center}}, R)|}{|Cr(\mathbf{x}_{\text{center}}, R)|}. \tag{4}$$

Since $|Sq(\mathbf{x}_{\text{center}}, \sqrt{\pi}R)| = |Cr(\mathbf{x}_{\text{center}}, R)|$ and $|Cr(\mathbf{x}, R) \cap Sq(\mathbf{x}_{\text{center}}, \sqrt{\pi}R)| = |Cr(\mathbf{x}_{\text{center}}, R) \cap Sq(\mathbf{x}, \sqrt{\pi}R)|$ it holds that $\pi^{Cr}(\mathbf{x}) = \pi^{Sq}(\mathbf{x})$ for all $\mathbf{x}$ in W^{Cr}, i.e. the probability of selection for any point $\mathbf{x}$ in W^{Cr} is identical for circular and square plots. The map of probabilities is shown in Figure A.1. However, these probabilities are not used in the modelling or in the RMSE or MD computations in this study.

ALS metrics

The Optech Titan ALS instrument can record up to four echoes per pulse. Here, we used a combined set of echoes from echo categories 'first of many' and 'only', which we call 'first echoes'. Metrics included traditional ABA height percentiles and cover variables (Næsset, 2002; Packalen *et al.*, 2019). Ultimately, we used two metrics in modelling: height percentile of 80 (h80) and vegetation cover >2 m (d2). Height percentile was computed as proposed by Hyndman and Fan (1996) using all 'first echoes'. Cover metric d2 was computed as the ratio of the number of 'first echoes' > 2 m divided by the total number of 'first echoes'. The

ALS metrics were computed in the same way in the square and circular plots.

Model

Plot-level AGB was modelled using ALS metrics. We chose AGB because it is a common attribute in different types of forest inventories and is strongly correlated with the carbon stock. We selected predictor variables and the model form such that the model was relatively simple but also fitted the data well. The model form was as follows

$$\sqrt{AGB} = \beta_0 + \beta_1 \mathrm{d2} + \beta_2 \mathrm{h80} + \varepsilon \tag{5}$$

where β_0, β_1 and β_2 are coefficients to be estimated from the data, d2 and h80 are the ALS metrics and ε is the residual error. The model was fitted with the least squares method (Chambers, 1992). Because of the square root transformation of the response variable, the back-transformed value was corrected for bias (Gregoire *et al*., 2008)

$$\hat{AGB} = \left(\hat{\sqrt{AGB}}\right)^2 + \hat{\sigma_\varepsilon}^2 \tag{6}$$

where $\hat{\sigma_\varepsilon}$ is the estimated residual standard deviation of the fitted model.

AGB on the plot edge

An edge tree is a tree whose crown intersects the plot border. Let us consider the area of the border region for a tree with a circular crown with diameter u and radius $u/2$ (Figure 3). The length of the perimeter of a circular plot is $2\pi R$, and the length of the perimeter of a square plot is $4L = 4\sqrt{\pi}R$, where R is the radius of a circular plot and L is the side length of a square plot. The crown of the tree intersects the border of a sample plot when its centre point lies within distance $u/2$ from the border line. The ratio of the border areas of the square and circle plots is then:

$$\frac{(L+u)^2 - 4\left(\frac{u}{2}\right)^2\left(1-\frac{\pi}{4}\right) - (L-u)^2}{\pi\left(R+\frac{u}{2}\right)^2 - \pi\left(R-\frac{u}{2}\right)^2} = \frac{2}{\sqrt{\pi}} - \frac{1-\frac{\pi}{4}}{2\pi R}u \tag{7}$$

Assuming equal areas of the square and circular plots, equation (7) shows that the ratio of the border areas is not constant but depends on both the size of the sample plot and tree crown radius. The reason for this is that in the square plot the corner of the outer boundary of the border area is rounded by the factor that depends on the diameter of tree crown (Figure 3, left). The ratio goes to $\frac{2}{\sqrt{\pi}}$ when $R \to \infty$ for a fixed u, so its value increases as the plot size increases. With typical values of real forest data (e.g. $R = 7.95$ m, $u = 2-7$m in this study), the ratio falls between 1.10 and 1.12.

Equation (7) shows that a square plot has a larger border region than a circular plot of the same area. Therefore, we examined empirically the relationship between AGB of the edge trees and the RMSE values associated with plot-level predictions. We used edge AGB as a measure of 'edge tree noise' because AGB was the response variable of the models. Because tree crown radius was not measured in the field, we used the following protocol to determine the trees that intersect the plot border:

1. Tree crowns were identified by individual tree detection on ALS data as described in Kotivuori *et al*. (2021): a total of 6087 tree crowns were delineated.
2. The area of the segmented tree crown (*a*) was converted to crown radius (*crad*) assuming a circular crown shape (*crad* = $\sqrt{a/\pi}$).
3. The height of the highest ALS echo within a segmented tree crown was used to estimate tree height (*ht*) of the segment.
4. The no-intercept model $\boldsymbol{crad} = \beta_1 \boldsymbol{ht}$ was fitted using the least squares method (Chambers, 1992), which yielded the estimate β_1 = 0.1162 (*ht* and *crad* in meters).
5. Fitted model from step 4 was used to predict crown radius for each tree from field measured tree heights.

This approach provided a crown radius for each tree; including trees not identified by individual tree detection. Finally, tree crowns that intersected the plot border based on their predicted crown radii and field measured tree locations were identified and used to predict the sum of AGB for trees on the plot edge. The contribution of errors from trees that are outside of the 30 m × 30 m plots was not considered.

Results

Models

The distributions of the model coefficient estimates and the predictor variable values for the circular and square plots resembled each other (Figure 4). Parameter estimate $\hat{\beta}_0$ had, on average, a slightly larger value for the square plots, whereas parameter estimate $\hat{\beta}_1$, which is the coefficient of predictor variable d2, had a slightly larger value for the circular plots. Parameter estimate $\hat{\beta}_2$, which is the coefficient of predictor variable h80, had a slightly wider distribution in the square than circular plots. Distributions of the predictor variable values (d2 and h80) were almost indistinguishable.

The distributions of the estimates of the residual standard deviation ($\hat{\sigma_\varepsilon}$) showed a difference between the circular and square plots. Models fitted with the square plots had larger residual errors than the models fitted with the circular plots. The residual standard deviation was used to correct for bias due to square root transformation of the response variable.

The distributions of plot-level AGB means were visually indistinguishable between the circular and square samples (Figure 5, left panel). However, the distributions of standard deviations differed (Figure 5, right panel), with an approximate 0.3 per cent difference between the mean standard deviation of the square plots (51.83) and the mean standard deviation of the circular plots (51.67).

Performances of circular and square plots

Density plots showed that the RMSE values were, on average, smaller when predicting for circular plots (Figure 6, left column).

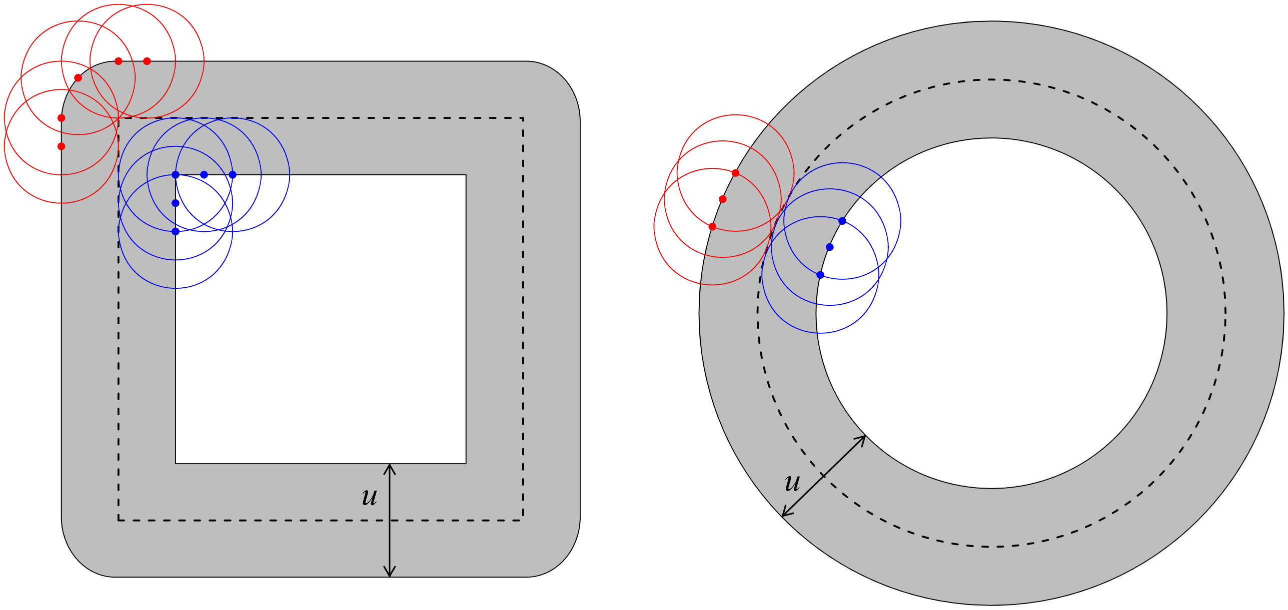

Figure 3 Border areas of square and circular sample plots assuming a circular tree crown with diameter ***u***. Dashed black line is the plot border and the grey area denotes the border area. Blue circles depict trees inside the plot whose crowns touch the plot border. Red circles depict trees outside the plot whose crowns touch the plot border.

The trend was similar whether the model was fitted with circular or square plots. This means that predictions for circular plots exhibited smaller RMSE values, even when the models were fitted with square plots. The magnitude of this difference (the difference of mean RMSE between the plot shape used in prediction) was ∼0.32 $Mg{\cdot}ha^{-1}$, or 1.2 per cent in relative terms (Table 2). Models fitted with circular plots performed slightly better than models fitted with square plots in both circular (26.192 vs. 26.205) and square (26.512 vs. 26.523) plots. The narrow 95 per cent confidence intervals of the mean RMSE values indicated that the number of repetitions (100 000) was sufficient in the simulations (Table 2).

The MD values were centred on zero for all combinations of model fitting and prediction (Figure 6, right column), regardless of whether a model was fitted with circular or square plots and predicted for circular or square plots. Due to the large number of simulations, the 95 per cent confidence intervals for mean MD were narrower than 0.015 $Mg{\cdot}ha^{-1}$, yet still covered zero in every case (Table 2). This indicates that there was no evidence of a systematic prediction error associated with modelling and prediction for the same or different plot shapes. Although some of the MD values from individual simulations were observed to be both practically (e.g. magnitude of MD > 5 per cent in 0.8 per cent of cases) and statistically (e.g. P-value <0.05 in 1.8 per cent of cases) significant, these proportions were almost identical across all combinations of plot shapes; there was no evidence that the plot shape (circle or square) used to model or predict AGB led to any systematic prediction error.

Effect of edge trees

The edge AGB and RMSE values were positively correlated in both circular and square plots with Pearson's r values of 0.061 and 0.050, respectively (Figure 7). In other words, RMSE values increased concomitantly with an increase in the amount of edge AGB. The slopes of the trendlines for the circular plots (solid red line) and square plots (solid blue line) were similar (Figure 7). The trendlines were also extrapolated across the range of other plot shapes (dashed red and blue lines) to illustrate that the fitted lines align quite well with the trendlines of the other plot shape. Although the relationship is overwhelmed by noise, the similarity in fitted relationships for circular and square plots suggests that the amount of edge tree noise may be one reason for greater RMSE values with the square plots.

Discussion

In this study, we delved into the effect of using different plot shapes in ABA forest inventories. We concentrated on the most common case, where circular plots are used in model fitting and the predictions are required on square population elements equivalent to square plots. Using actual tree and ALS data to explore this issue is a challenge as there is a need to control for confounding factors, such as distributions of field and remote sensing observations on the plots, and for the population. To effectively assess the shape effects, circular and square sample plots should represent the same population and other factors that affect the outcome should be eliminated. This study approached the issue through a sample design that guarantees identical sampling probabilities for circular and square plots for every tree in our target population. This was feasible with empirical simulations on large stem mapped sample plots (our population) and by sampling in a specific way.

The distributions of ABA metrics used as predictor variables were almost identical. This indicates that plot shape had little

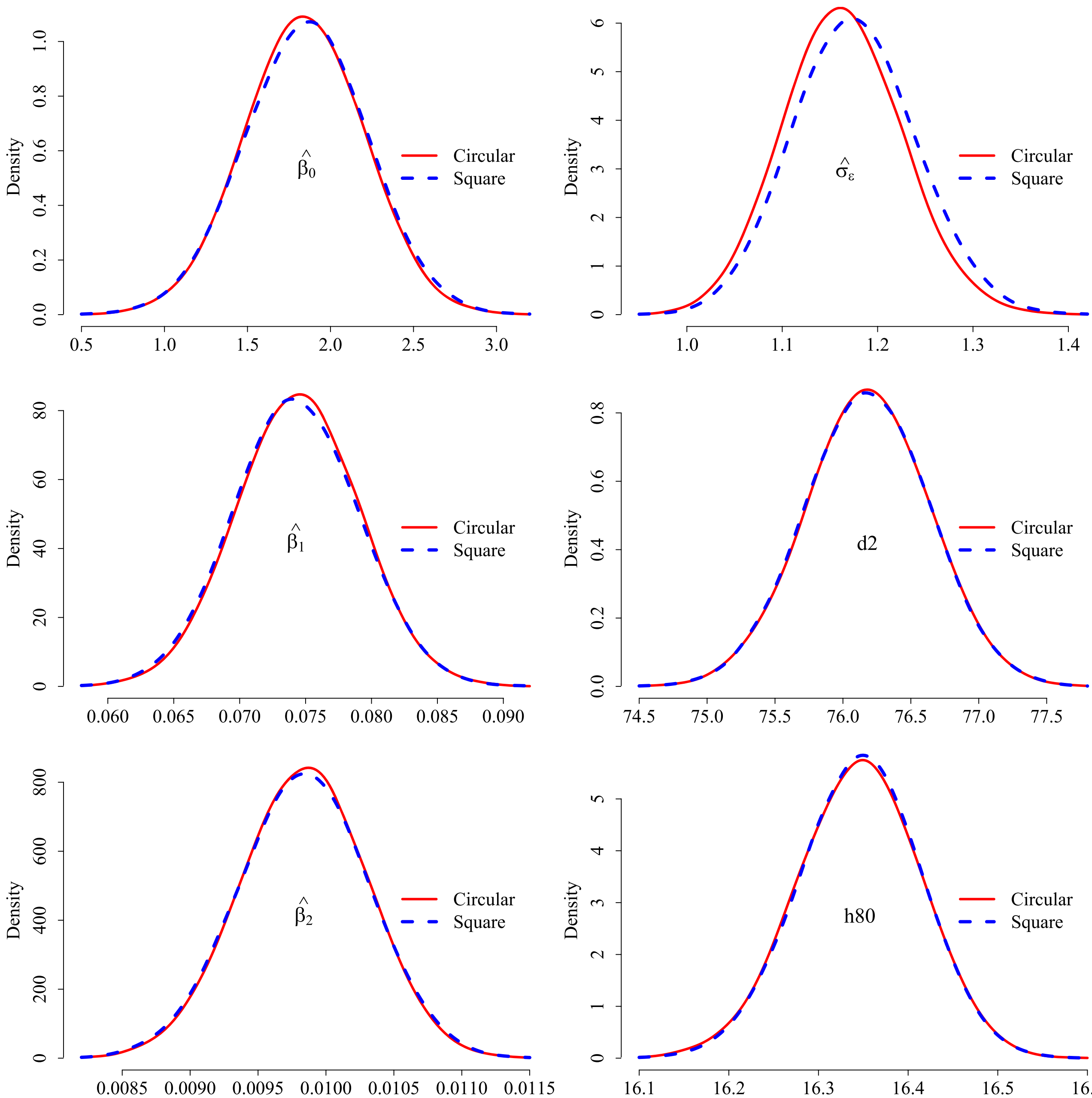

Figure 4 Density plots of the model coefficient estimates ($\hat{\beta_0}, \hat{\beta_1}, \hat{\beta_2}$), residual standard deviation for the square root of AGB ($\hat{\sigma_\varepsilon}$) and predictor variables (d2, h80) for circular and square plots.

effect on the values of ABA metrics. The distributions of the model coefficient estimates were also very similar between the circular and square plots, although the distributions of residual standard deviation differed notably. Residual standard deviation was used in bias correction terms due to the use of square root transformation of the response variable. Larger residual standard deviation values for square plots indicates that the residual variation is smaller for models fitted with circular plots than for square plots. One reason for the greater residual variation on square plots could be the inherently greater variation exhibited by forest attributes on square plots (Figure 4) as the maximum distance between trees is greater for these plots. Due to spatial autocorrelation, particularly in managed forests, trees close together tend to be more similar than those further apart. The greater residual variation also means that there is greater uncertainty in the model parameter estimates.

Another factor that may account for the smaller residual variation observed in the circular plots is the ratio of edge tree area to plot area, which is smallest for circular plots—a relationship that varies with the size of the tree relative to the plot. With the same plot areas, our calculated ratio was typically 1.10–1.12 times larger on square plots than circular plots. Therefore, due

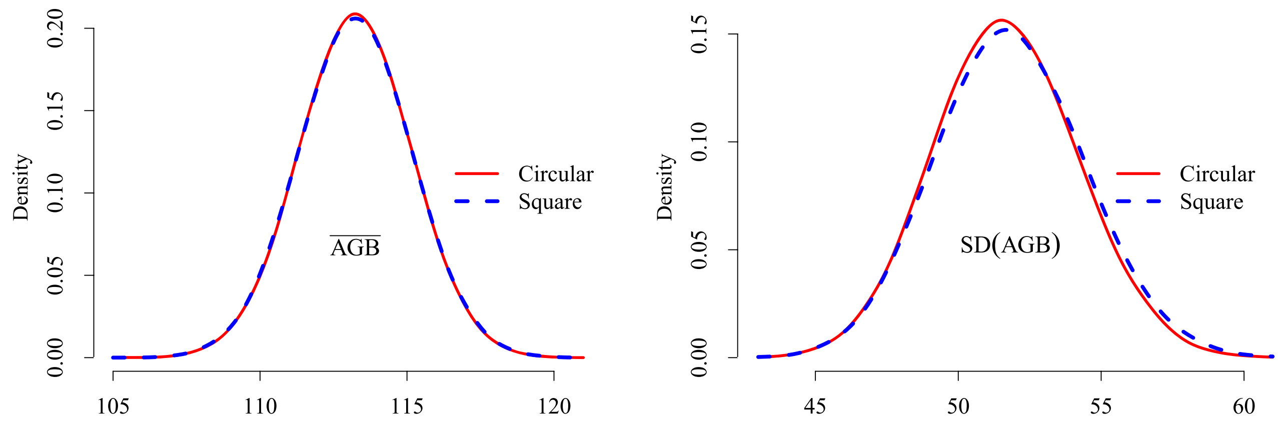

Figure 5 Density plots of the sample ($n = 110$) mean ($\overline{\text{AGB}}$) and standard deviation (SD(AGB)) of plot AGB (Mg·ha^{-1}) used in model fitting.

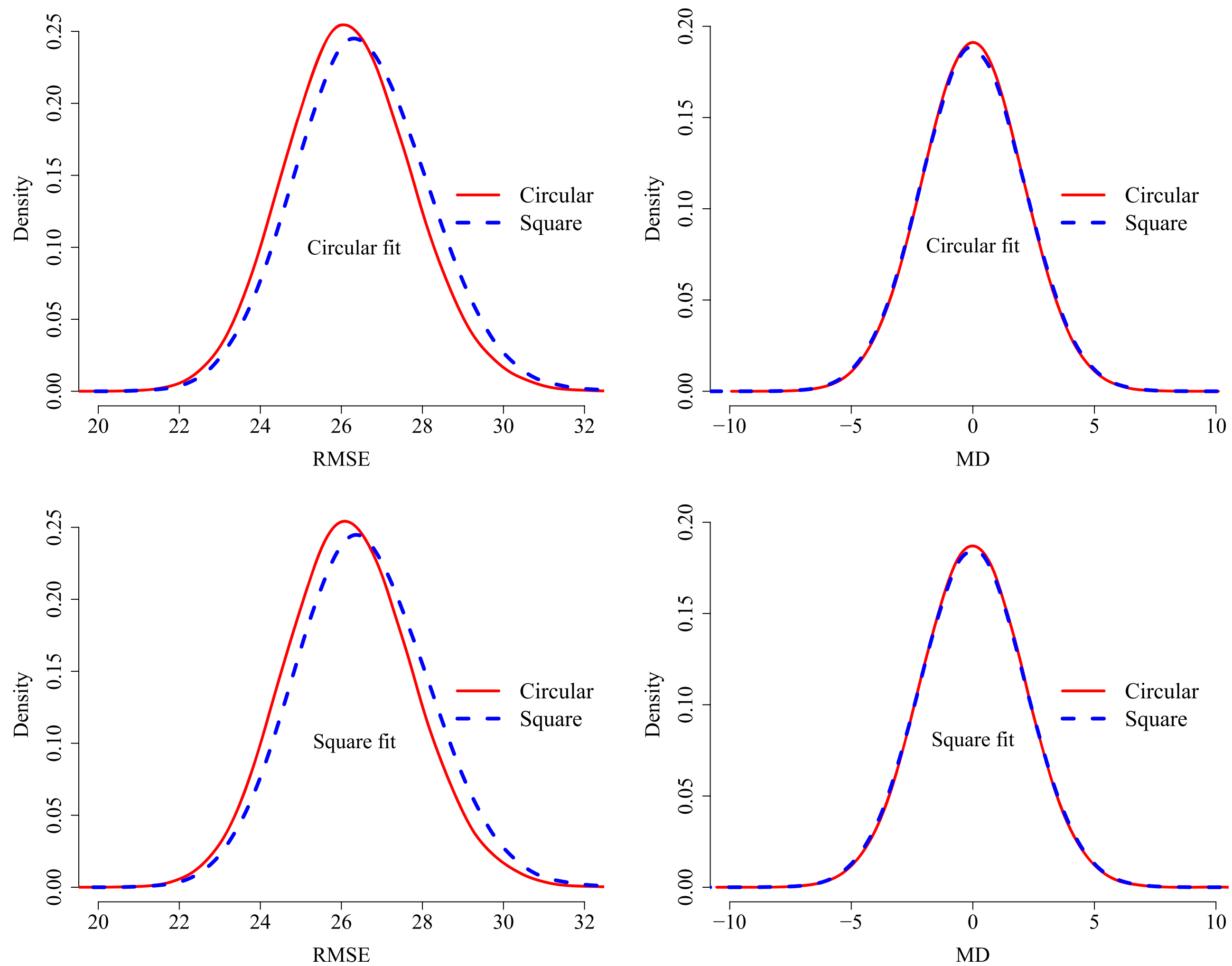

Figure 6 Density plots of RMSE (Mg·ha^{-1}) and MD (Mg·ha^{-1}) when predicted to circular (solid red line) and square (dashed blue line) plots. Models were fitted with another set of circular (first row) or square plots (second row).

to the greater errors caused by edge trees (omission and commission of edge tree crowns), it would be expected that errors would be greater in the square than circular plots. We tested this empirically by modelling the relationship between edge tree AGB and RMSE values, and found a relationship between edge tree AGB and RMSE, which was similar in circular and square plots.

Table 2 Mean values with 95 per cent confidence intervals of RMSE (Mg·ha^{-1}) and MD (Mg·ha^{-1}) when predicting for circular and square plots. Models were fitted with another set of circular or square plots

	Predict to circular plots RMSE	Predict to square plots RMSE	Predict to circular plots MD	Predict to square plots MD
Fit with circular plots				
Mean	26.192	26.512	0.002	0.000
Upper 95% limit	26.197	26.517	0.008	0.006
Lower 95% limit	26.188	26.508	−0.003	−0.006
Fit with square plots				
Mean	26.205	26.523	0.004	0.000
Upper 95% limit	26.209	26.527	0.009	0.006
Lower 95% limit	26.201	26.518	−0.003	−0.006

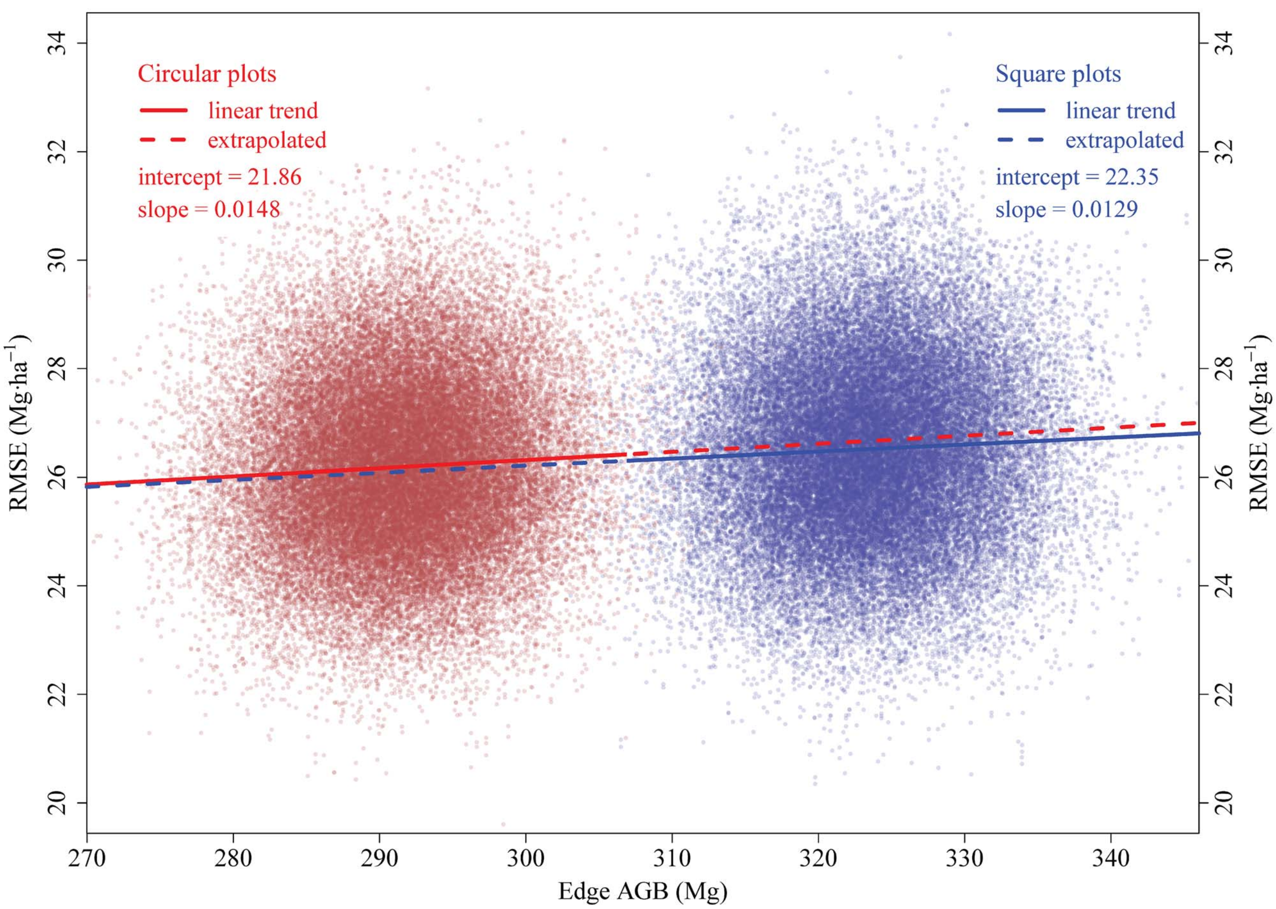

Figure 7 Relationship between edge AGB and RMSE values when models were fitted with circular plots and predicted to another sample of circular plots (red points) and square plots (blue points) over 100 000 repetitions. Edge AGB is the sum of AGB of edge trees in the sample (110 plots). Linear trends are shown separately for circular (red line) and square (blue line) plots.

This suggests that edge tree effects also contribute to the greater RMSE values found for the square plots.

We evaluated prediction performance with RMSE and MD using samples from circular and square plots across plot shapes. The results showed that the use of different plot shape did not cause systematic over- or under-prediction for a different plot shape. The prediction error (RMSE) was, however, greater for square plots regardless of the shape of the plots used in the model fitting. It is difficult to provide a definitive explanation as to why the RMSE values were less for circular plots. This would likely

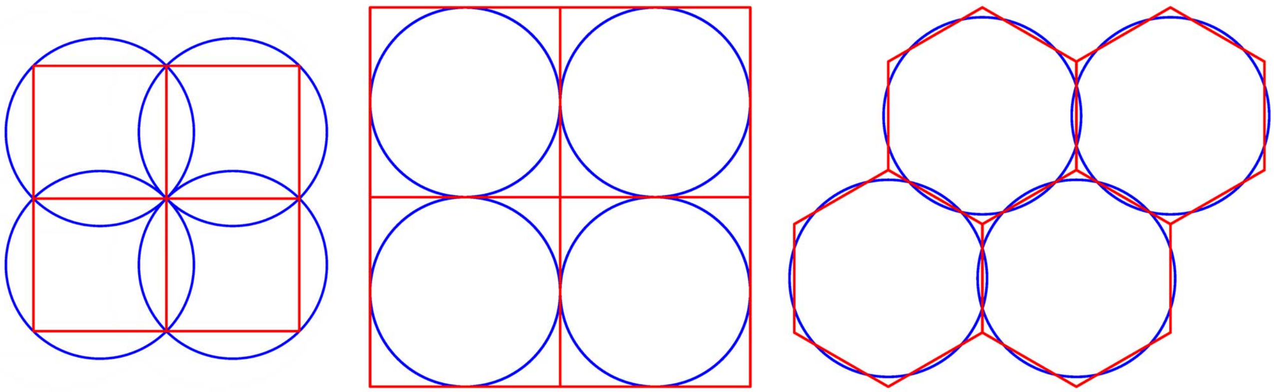

Figure 8 Schematics for three strategies to tessellate a population in an ABA with circular and hexagonal plots. In the left and centre figures, metrics are computed from the circular plots and assigned to grid cells. The areas of the circular plots (blue) are the same size as in the training plots but the area differs in the grid cells (red). In the left-hand figure, adjacent circles overlap. In the centre figure, circular plots do not overlap, but the ALS data in the corners of the square cells are not used at all. In the right-hand figure, there is a high degree of agreement (esp. footprint and edge length) between the equal-area circle and hexagonal footprints.

require that both tree and ALS data are simulated, which adds further complications with respect to the creation of a realistic population. We demonstrated two possible plot shape-related factors with measurable effects on performance, which include greater inherent variation of AGB on square plots and a greater proportion of edge tree AGB on square plots.

One clear finding in this study is that the RMSE values reported for circular plots underestimate the RMSE values associated with square plots, although the magnitude was so small (∼1.2 per cent) that it is not to be considered problematic in many cases. However, analysts should be aware of the issue when using different plot shapes in modelling and prediction. Conceptually, the simplest solution to eliminate the plot shape effect is to use either circular or square plots both in model fitting and prediction. However, this solution has more drawbacks than advantages. With respect to the plot shape measured in the field, it is much easier to establish and locate a circular plot than a square plot and it is likely not worth the added expense of migrating to square plots. Using circular plots to fit models and for prediction is arguably a more tractable approach, although there are also complications that arise from using circular grid cells, either from overlapping data in adjacent plots (Figure 8, left), or the presence of data in the corners of square cells that are not used at all (Figure 8, middle). A partial solution is the use of circular plots in model fitting and the use of hexagons when predicting (Packalén *et al*., 2011), which would reduce shape effects, since hexagonal and circular plots with the same area show greater agreement than square and circular plots (Figure 8, right). In contrast to circles, hexagons can also be placed in a way that they do not overlap but cover the population without any gaps occurring between plots. To date, hexagons have not been adopted to any significant degree into ABA forest inventories.

Large area forest surveys that aim at unbiased total and mean values may use ABA to improve the precision of estimators (Dettmann *et al*. 2022). Unbiasedness is a desirable property of a variance estimator in this context. If the variance estimate is based on the circular plots used in model fitting, but the square plots are used for the population tessellation, the variance will be underestimated. This is a generic issue not related to the estimator itself but originates from the different properties of the samples (circular plots) and population elements (square plots).

Conclusions

Our results show that there is no systematic prediction error when a model is fitted with circular plots and when the prediction is performed for square plots, and vice versa. This implies that there is no issue with mean or total estimates or estimators when using circular versus square plots to fit or predict in ABA, provided the plot areas are the same. This was expected as the distributions of the model coefficient estimates were almost identical regardless of the plot shape used in model fitting.

Although the model coefficients were identical, the RMSE values were greater for the square plots. We identified two possible reasons for this; (1) the variance of the response variable is greater on square plots, and (2) the number of edge trees is greater on square plots. This indicates that a measure of model performance (e.g. RMSE) for the population composed of square plots is an underestimate if it is based on circular plots. This configuration paints an overly positive picture with respect to the expected precision of the predictions, although the effect from the use of different plot shapes was small.

A small underestimation of RMSE due to plot shape effect has little practical significance in the stand-level forest management inventories that use ABA. In large area forest surveys that report confidence intervals of total or mean estimates, the effect of plot shape is at least a theoretical issue.

Data availability

The plot simulations used this article will be shared on reasonable request to the corresponding author.

Funding

Academy of Finland through projects (323484 to P.P., 295100 to M. Myllymäki, 327211 to M. Maltamo) and through Finnish Flagship Programme for the Forest-Human-Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) (337127 to M.M, 337655 to M.M.).

References

Andersen, H.-E., Strunk, J. and Temesgen, H. 2011 Using airborne light detection and ranging as a sampling tool for estimating Forest biomass resources in the upper Tanana Valley of interior Alaska. *West. J. Appl. For.* **26**, 157–164. https://doi.org/10.1093/wjaf/26.4.157.

Axelsson, P. 2000 DEM generation from laser scanner data using adaptive TIN models. *Int. Arch. Photogr. Remote Sens.* **B4**, 110–117.

Breidenbach, J., Koch, B., Kändler, G. and Kleusberg, A. 2008 Quantifying the influence of slope, aspect, crown shape and stem density on the estimation of tree height at plot level using lidar and InSAR data. *Int. J. Remote Sens.* **29**, 1511–1536. https://doi.org/10.1080/01431160701736364.

Chambers, J.M. 1992 Linear models. In *Statistical Models*. Chambers, T.J., Hastie (eds.). Wadsworth & Brooks/Cole, pp. 95–144.

Deo, R.K., Froese, R.E., Falkowski, M.J. and Hudak, A.T. 2016 Optimizing variable radius plot size and LiDAR resolution to model standing volume in conifer forests. *Can. J. Remote. Sens.* **42**, 428–442. https://doi.org/10.1080/07038992.2016.1220826.

Dettmann, G.T., Radtk, P.J., Coulston, J.W., Green, P.C., Wilson, B.T., Moisen, G. *et al.* 2022 Review and synthesis of estimation strategies to meet small area needs in Forest inventory. *Front. For. Global Change* **5**, Article 813569, 1–16. https://doi.org/10.3389/ffgc.2022.813569.

Fadili, M., Renaud, J.P., Bock, J. and Vega, C. 2019 RegisTree: a registration algorithm to enhance forest inventory plot georeferencing. *Ann. For. Sci.* **76**, 30. https://doi.org/10.1007/s13595-019-0814-2.

Fassnacht, F.E., Latifi, H. and Hartig, F. 2018 Using synthetic data to evaluate the benefits of large field plots for forest biomass estimation with LiDAR. *Remote Sens. Environ.* **213**, 115–128. https://doi.org/10.1016/j.rse.2018.05.007.

Frazer, G.W., Magnussen, S., Wulder, M.A. and Niemann, K.O. 2011 Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR derived estimates of forest stand biomass. *Remote Sens. Environ.* **115**, 636–649. https://doi.org/10.1016/j.rse.2010.10.008.

Gobakken, T., Korhonen, L. and Næsset, E. 2013 Laser-assisted selection of field plots for an area-based forest inventory. *Silva Fennica* **47**, 943.

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**, 1095–1109. https://doi.org/10.1139/X07-219.

Goetz, S. and Dubayah, R. 2011 Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Manage.* **2**, 231–244. https://doi.org/10.4155/cmt.11.18.

Gopalakrishnan, R., Thomas, V.A., Coulston, J.W. and Wynne, R.H. 2015 Prediction of Canopy Heights over a large region using heterogeneous Lidar datasets: efficacy and challenges. *Remote Sens.* **7**, 11036–11060. https://doi.org/10.3390/rs70911036.

Gregoire, T.G., Lin, Q.F., Boudreau, J. and Nelson, R. 2008 Regression estimation following the square-root transformation of the response. *For. Sci.* **54**, 597–606.

Hawbaker, T.J., Keuler, N.S., Lesak, A.A., Gobakken, T., Contrucci, K. and Radeloff, V.C. 2009 Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res. Atmos.* **114**, G00E04.

Hyndman, R.J. and Fan, Y. 1996 Sample quantiles in statistical packages. *Am. Stat.* **50**, 361–365.

Junttila, V., Finley, A.O., Bradford, J.B. and Kauranne, T. 2013 Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory. *For. Ecol. Manag.* **292**, 75–85. https://doi.org/10.1016/j.foreco.2012.12.019.

Kangas, A. and Maltamo, M. (Eds.) 2006 *Forest Inventory: Methodology and Applications. Series: Managing Forest Ecosystems* vol. **10**. Dordrecht. https://doi.org/10.1007/1-4020-4381-3.

Knapp, N., Huth, A. and Fischer, R. 2021 Tree crowns cause border effects in area-based biomass estimations from remote sensing. *Remote Sens.* **13**, 1592. https://doi.org/10.3390/rs13081592.

Korpela, I., Tuomola, T. and Välimäki, E. 2007 Mapping forest plots: an efficient method combining photogrammetry and field triangulation. *Silva Fennica* **41**, 457–469.

Kotivuori, E., Maltamo, M., Korhonen, L., Strunk, J.L. and Packalen, P. 2021 Prediction error aggregation behavior for remote sensing augmented forest inventory approaches. **94**, 576–587. https://doi.org/10.1093/forestry/cpab007.

Maltamo, M., Korhonen, K.T., Packalén, P., Mehtätalo, L. and Suvanto, A. 2007 Testing the usability of truncated angle count sample plots as ground truth in airborne laser scanning based frest inventories. *Forestry* **80**, 73–81. https://doi.org/10.1093/forestry/cpl045.

Maltamo, M., Packalen, P. and Kangas, A. 2021 From comprehensive field inventories to remotely sensed wall-to-wall stand attribute data—a brief history of management inventories in the Nordic countries. *Can. J. For. Res.* **51**, 257–266. https://doi.org/10.1139/cjfr-2020-0322.

Mauro, F., Valbuena, R., Manzanera, J.A. and García-Abril, A. 2011 Influence of global navigation satellite system errors in positioning inventory plots for tree-height distribution studies. *Can. J. For. Res.* **41**, 11–23. https://doi.org/10.1139/X10-164.

Næsset, E. 2002 Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **80**, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J. *et al.* 2017 A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the national forest inventory. *Remote Sens. Environ.* **194**, 447–454. https://doi.org/10.1016/j.rse.2016.10.022.

Okabe, A., Boots, B. and Sugihara, K. 1992 *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley.

Ørka, H.O., Bollandsås, O.M., Hansen, E.H., Næsset, E. and Gobakken, T. 2018 Effects of terrain slope and aspect on the error of ALS-based predictions of forest attributes. *Forestry* **91**, 225–237. https://doi.org/10.1093/forestry/cpx058.

Packalén, P., Heinonen, T., Pukkala, T., Vauhkonen, J. and Maltamo, M. 2011 Dynamic treatment units in eucalyptus plantation. *For. Sci.* **57**, 416–426.

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. https://doi.org/10.1016/j.rse.2019.01.022.

Packalen, P., Strunk, J., Pitkänen, J., Temesgen, H. and Maltamo, M. 2015 Edge-tree correction for predicting forest inventory attributes using area-based approach with airborne laser scanning. *IEEE J. Select. Topics Appl. Earth Observ.Remote Sens.* **8**, 1274–1280. https://doi.org/10.1109/JSTARS.2015.2402693.

R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Austria3-900051-07-0. http://www.R-project.org.

Repola, J. 2008 Biomass equations for birch in Finland. *Silva Fennica* **42**, 605–624.

Repola, J. 2009 Biomass equations for scots pine and Norway spruce in Finland. *Silva Fennica* **43**, 625–647.

Ruotsalainen, R., Pukkala, T., Kangas, A., Vauhkonen, J., Tuominen, S. and Packalen, P. 2019 The effects of sample plot selection strategy and the number of sample plots on inoptimality losses in ALS-based forest management planning. *Can. J. For. Res.* **49**, 1135–1146. https://doi.org/10.1139/cjfr-2018-0345.

Schreuder, H.T., Gregoire, T.G. and Wood, G.B. 1993 *Sampling Methods for Multiresource Forest Inventory*. Wiley.

Strunk, J., Temesgen, H., Andersen, H.-E., Flewelling, J.P. and Madsen, L. 2012 Effects of pulse density and sample size on a model-assisted approach to estimate forest inventory variables. *Can. J. For. Res.* **38**, 644–654.

Zhao, K., Popescu, S.C. and Nelson, R.F. 2009 Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.* **113**, 182–196.

Appendix

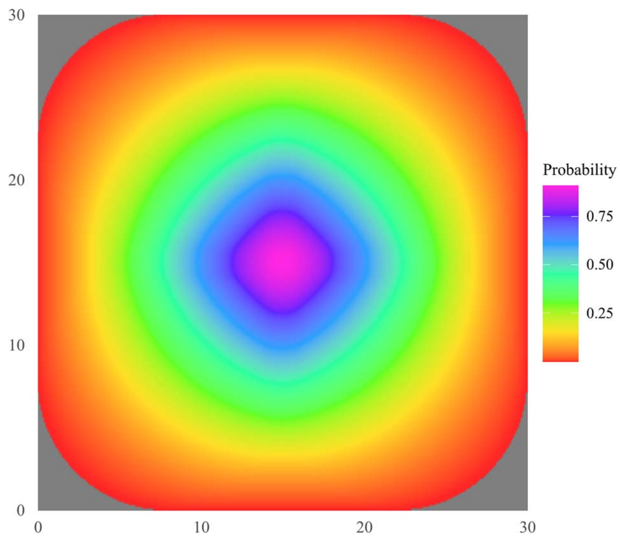

Figure A.1 Probability that locations within a 30 m × 30 m plot are included within the sample. The probability is equal in a circular plot with a radius of 7.95 m and a square plot with a side length of 14.09 m. Probability is zero in the grey area.