Remote Sensing of Environment 227 (2019) 110–124

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data

Piotr Tompalski[a,*], Joanne C. White[b], Nicholas C. Coops[a], Michael A. Wulder[b]

[a] Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada
[b] Canadian Forest Service, Pacific Forestry Center, Natural Resources Canada, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada

ARTICLE INFO

Keywords:
Area-based approach
Lidar
Transferability
Models
Point density
Random forests
K-nearest neighbours
Ordinary least squares

ABSTRACT

Airborne laser scanning (ALS) is a reliable source of accurate information for forest stand inventory attributes including height, cover, basal area, and volume. The commonly applied area-based approach (ABA) allows the derivation of wall-to-wall geospatial coverages representing each of the modeled attributes at a grid-cell level, with spatial resolutions typically between 20 and 30 m. The ABA predictive models are developed using stratified inventory data from field plots, the requirement for which can increase the overall cost of the ALS-based inventory. Parsimonious use of ground plots is a key means to control variable costs in the operational implementation of the ABA. In this paper, we demonstrate how the prediction accuracy of Lorey's height (HL, m), quadratic mean diameter (QMD, cm), and gross volume (V, m^3) vary when existing ABA models are transferred to different areas or are applied to point cloud data with different characteristics than those on which the original model was developed. Specifically, we consider three scenarios of model transferability: (i) same point cloud characteristics, different areas; (ii) different point cloud characteristics, same areas; and (iii) different point cloud characteristics, different areas. We generated area-based models using three modeling approaches: linear regression (OLS), random forests (RF), and k-nearest neighbour (kNN) imputation. Results indicated that the prediction accuracy of area-based models varied by attribute and by modeling approach. We found that when the models were transferred their prediction accuracy decreased, with an average increase in relative bias up to 22.04%, and increase in relative RMSE up to 29.31%. Prediction accuracies for HL were higher than those of QMD or V when models were transferred, and had the lowest average increase in relative bias and relative RMSE of < 5% in the majority of cases. Likewise OLS models for HL had greater prediction accuracies when models were transferred compared to RF and kNN models, especially when the point cloud characteristics were similar. Conversely, we found that for QMD and V, RF models were found to be the most transferable in cases when models were applied to different areas with similar and different point cloud characteristics. While there is potential for cost savings by transferring models and reducing data acquisition costs, our results show the degree of transferability depends more on the attribute being modelled or the modeling approach applied, and less on the characteristics of the point cloud data.

1. Introduction

Globally the forest sector aims to satisfy an increasingly complex set of rules, standards, business practices, and public expectations including economic, environmental, and social policy goals. As a result, sustainable forest management activities need to generate the maximum possible value from each tree harvested (Woods et al., 2011). In order to maximize value, forest inventory information is required that spatially quantifies timber and other forest-related ecosystem services in a timely manner and with a high level of spatial detail (Kangas et al., 2018). Moreover, these detailed inventory data must then be integrated into new and existing models to support improved growth and yield modeling and forest planning (Hyyppä et al., 2008; Tompalski et al., 2016).

Forest inventories are undertaken at an array of spatial scales and are designed to support information needs that range from strategic to tactical to operational (Leckie and Gillis, 1995). Strategic inventories are often designed to cover large areas (i.e. nations, provinces), and are intended for long-term monitoring and planning for a variety of forest ecosystem goods and services (White et al., 2016). In contrast, operational forest inventories are designed to support short-term, localized decisions related to accessing and harvesting timber (Wulder et al.,

* Corresponding author.
E-mail address: piotr.tompalski@ubc.ca (P. Tompalski).

https://doi.org/10.1016/j.rse.2019.04.006
Received 20 September 2018; Received in revised form 1 April 2019; Accepted 7 April 2019

2008). Conventionally, strategic and tactical inventories are developed through the interpretation of aerial photography, which facilitates the delineation of forest stands and the assignment of forest attributes derived for these stands, including tree species composition, height, stocking, site quality, health status, and stand age (Woods et al., 2011). Supported by limited ground sampling and empirical yield table estimates, stand site productivity and growth is then inferred.

Airborne laser scanning (ALS) data can be used to estimate forest stand attributes at scales that are relevant for strategic and operational forest information needs (Tompalski et al., 2016). Due to reliability and ease of implementation, the area based approach (ABA) has been widely applied to process ALS point clouds into meaningful and relevant forest attribute information. Reliable estimates of height, basal area, stem volume, and aboveground biomass have all been demonstrated over a wide range of forest environments (Næsset et al., 2004; Hyyppä et al., 2008; Woods et al., 2011; Brosofske et al., 2014; Wulder et al., 2013; Bouvier et al., 2015) and are now considered operational (Wulder et al., 2013; Næsset, 2014).

Recently digital aerial photogrammetry (DAP) has emerged as an alternative data source to ALS for forest stand characteristics and has been successfully used to predict stand attributes with ABA (Bohlin et al., 2012; Puliti et al., 2016; White et al., 2015). DAP point clouds primarily represent the top portion of the forest canopy and therefore require an ALS-based ground surface to normalize the point elevations. The lower acquisition costs associated with DAP data have made these data a potential source of information for inventory updates (Goodbody et al., 2019).

Underpinning the ABA is the statistical relationship between predictor variables, derived from summaries of the vertical distribution of laser returns in the ALS point cloud, and co-located ground plot measurements of forest inventory attributes (Næsset, 2002). Area-based models can be developed using parametric (Næsset et al., 2004; Woods et al., 2011; Wulder et al., 2012) or non-parametric (Hudak et al., 2008; Penner et al., 2013) approaches. Ground plot data play an essential role in the ABA regardless of the approach (Næsset, 2002), serving to both calibrate and validate area-based models. Detailed measurements from accurately georeferenced ground plot data are compiled to provide the forest inventory response variables for modeling. Critically, these ground plot data must represent the full range of forest structural variability present in the area of interest. While ALS-derived statistical models are usually similar across forest types and locations, typically each model is calibrated and validated against local field data (Strunk et al., 2012; Sumnall et al., 2016). As with conventional forest inventories, ground plot costs represent a significant proportion of the total costs associated with an area-based approach (Eid et al., 2004), particularly in areas with constraints to forest access or over complex terrain.

To date, there have been few studies that have specifically investigated the spatial transferability of ABA models, with the aim of developing a model in one forest area and applying it in another forest area. Several studies have however explored the utility of universal models. For example, in an effort to increase the efficiency of estimating carbon density and reduce reliance on local ground-based calibration data, Asner et al. (2012) applied a single universal ALS model for carbon density across four tropical regions ($R^2 = 0.80$, RMSE = 27.6 Mg C ha^{-1}). Magnussen et al. (2012) proposed a generic area-based model for attributes related to tree size (e.g. biomass, volume, basal area, quadratic mean diameter, Lorey's height), and their proposed model included only two ALS predictors, the mean grid cell height, and the variance of heights within the grid cell. Magnussen et al. (2012) posited that the transferability of models is increased by excluding ALS density metrics as predictors, which are more sensitive to variations in ALS acquisition parameters and to forest structural conditions and species composition.

There have likewise been a number of studies that have investigated pooling the available ALS and inventory data to develop predictive models of forest attributes. Næsset and Gobakken (2008) developed above- and below-ground biomass models based on 1395 plots located in ten different areas in Norway. They reported that the differences between the areas were up to 32% for above-ground biomass, and up to 38% for below-ground biomass. In a study located in Sweden, Nilsson et al. (2016) combined 11,500 NFI sample plots and ALS data from 13 different acquisitions to predict stem volume and basal area. The relative RMSE of the linear regression models ranged between 17.2–22.0% and 13.9–18.2% for the two model stand attributes, respectively. Nilsson et al. (2016) concluded that the prediction accuracy was at least as good as the accuracy of the data typically used in forest management planning. In a similar study in Finland, Kotivuori et al. (2016) combined nine inventory projects to develop nationwide ALS-based regression models for dominant height, volume and biomass, demonstrating that both forest structure and ALS device had a considerable effect on the predictions, and that the local calibration helped to increase the accuracy of volume and biomass models. Sumnall et al. (2016) examined the transferability of a leaf area index model across two different ALS instruments, concluding that models incorporating absolute counts of ALS returns or intensity values were not transferable between different instruments. Deo et al. (2017) compared the accuracy of generic or pooled models versus site-specific models for estimating biomass using ALS and an ABA at four sites in Minnesota, Maine, Pennsylvania-New Jersey, and South Carolina. In their study, the authors tested both linear regression and random forests, concluding that the pooled model had accuracies that were comparable to the site-specific models, and that estimates made using linear regression were more accurate.

Fekety et al. (2015) demonstrated the temporal transferability of ALS-based imputation of forest inventory attributes by combining ground plot observations from 2003 and 2009. The authors reported R^2 values for the pooled models that were 0.87, 0.90, 0.89, and 0.87 for aboveground carbon, basal area, stand density, and total stem volume, respectively. Fekety et al. (2015) concluded that pooling of the ground reference data not only increased the number of reference observations, but also the accuracy of the estimations. Building upon this work, Fekety et al. (2018) analyzed the spatial transferability of basal area and stem density models by combining ALS and field measurements from ecologically similar forest stands. The authors used six separate forest units and developed random forest regression models using inventory and ALS-metrics from 5 units. The withheld forest unit was then used to assess model performance. The results showed that the transferred basal area models were more accurate than transferred stem density models, with relative RMSE between 32.3–50.1%, and 40.7–67.3%, respectively. Fekety et al. (2018) conclude that for new ALS acquisitions over forested areas without inventory plots, foresters may use predictive models developed in forest stands with similar ecological conditions. Worth mentioning is also the recent study by Karjalainen et al. (2018) who focused on transferability of tree-level models. Using kNN imputation authors predicted tree diameter at breast height, height, crown based height, stem volume, and sawlog volume. When transferred, the relative RMSE increased between 3.2% for height and 54.4% for sawlog volume.

In practical contexts, the appeal of a universal model cannot be understated. At stake are significant cost savings, which largely manifest from reduced ground sampling (Magnussen et al., 2012), as well as processing efficiencies for operational monitoring programs (Asner et al., 2012). While a universal model may be more challenging to define or adopt than a regionally-specific pooled or global model, the re-use of existing models in different areas or when new ALS data (or other sources of point cloud data) are acquired over the same area, is a current operational reality in many jurisdictions. Streamlining the area-based approach must therefore involve trade-offs between cost, efficiency, and the development of robust models that are relevant to local conditions. What is required therefore is an improved understanding of the factors that influence model transferability.

Specifically, we consider three scenarios of model transferability: (i) models are applied using data with the same point cloud characteristics, but in a different area than where the models were originally developed; (ii) models are applied to data with different point cloud characteristics, but in a similar area in which the model was originally developed; and (iii) models are applied using data with different point cloud characteristics and in different areas, relative to the data and location used to develop the original model. Three different modeling approaches are used to develop the area-based models according to these three aforementioned scenarios: linear regression, random forests, and k-nearest neighbour imputation. Our objective is therefore to assess the transferability of area-based models according to attribute, location, point cloud characteristics, modeling approach, and combinations thereof.

2. Methods

2.1. Study areas

Data used in our analyses were acquired over three different areas. The bulk of the analysis was performed on northern Vancouver Island (NVI), British Columbia, Canada, and represents an area of approximately 52,000 ha, spanning three geographically separate areas (Fig. 1). Area-based models were developed primarily at this study site and then transferred to the other two study areas, which included the Malcolm Knapp Research Forest (MKRF) near Maple Ridge, British Columbia, Canada, and the Sunshine Coast Community Forest (SCCF) near Sechelt, British Columbia, Canada (Fig. 1). MKRF and SCCF were approximately 5100 ha and 20,900 ha in size, respectively. All three areas represent typical topography and stand conditions for a highly productive, managed coastal temperate rainforest, with western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), and western red cedar (*Thuja plicata* Donn ex D. Don) as the most common tree species.

All three study areas are located within the Coastal Western Hemlock biogeoclimatic zone (CWH), which can be characterized by high annual precipitation (3000–5000 mm), mild winters (average temperature 0 °C to 2 °C), and cool summers (average temperature 18 °C to 20 °C) (Meidinger and Pojar, 1991).

The majority of the NVI study area is located within the CWH Very Wet Maritime subzone, Submontane Variant (CWH vm1), which has a wet, humid climate, and elevation range from sea level to approximately 600 m. Forests in this subzone are dominated by Western hemlock, Amabilis fir, and Western red cedar. A portion of NVI is also located within the CWH Very Wet Maritime subzone, Montane Variant (CWH vm2), which occurs on higher elevations than CWH vm1, and may contain stands with Yellow cedar (*Chamaecyparis nootkatensis*), and Mountain hemlock (*Tsuga mertensiana*) that do not occur in the adjacent CWH vm1. The climate in CWH vm2 is cooler with short summers and cool winters featuring substantial snowfall. The average age of stands, compiled according to standard provincial procedures (Ministry of Forest Lands and Natural Resource Operations, 2014), was 144 years (standard deviation = 127 years). Overall, the elevations ranged from 0 to 1398 m a.s.l. in the NVI, with an average slope of 43.7%.

The MKRF study area is located within three different subzones of CWH. Dry Maritime Coastal Western Hemlock Subzone occurs at low elevations and can be characterized with warm, relatively dry summers and moist, mild winters. Forests in this subzone are dominated by Western hemlock, Western red cedar, and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco). The higher elevations of MKRF are located with CWH vm1 and CHM vm2 subzones. On average stands in the MKRF were younger, with the average age of 65 years (standard deviation = 69 years). Elevations in MKRF ranged from 98 to 1025 m a.s.l., with an average slope of 40.5%.

SCCF is located within the CWH dm subzone, with higher elevations located with the CWH vm1 subzone (both subzones described above). The average age of stands in this study area was 119 years (standard deviation = 125 years). Elevations in SCCF ranged from sea level to 1493 m a.s.l., with an average slope of 35.5%.

2.2. Ground plot data

In each of the study areas, the same methods were used to acquire ground plot measurements. Plots were circular with a radius of 14 m (615 m^2) in NVI and 11.2 m (400 m^2) in MKRF and SCCF. Plots were

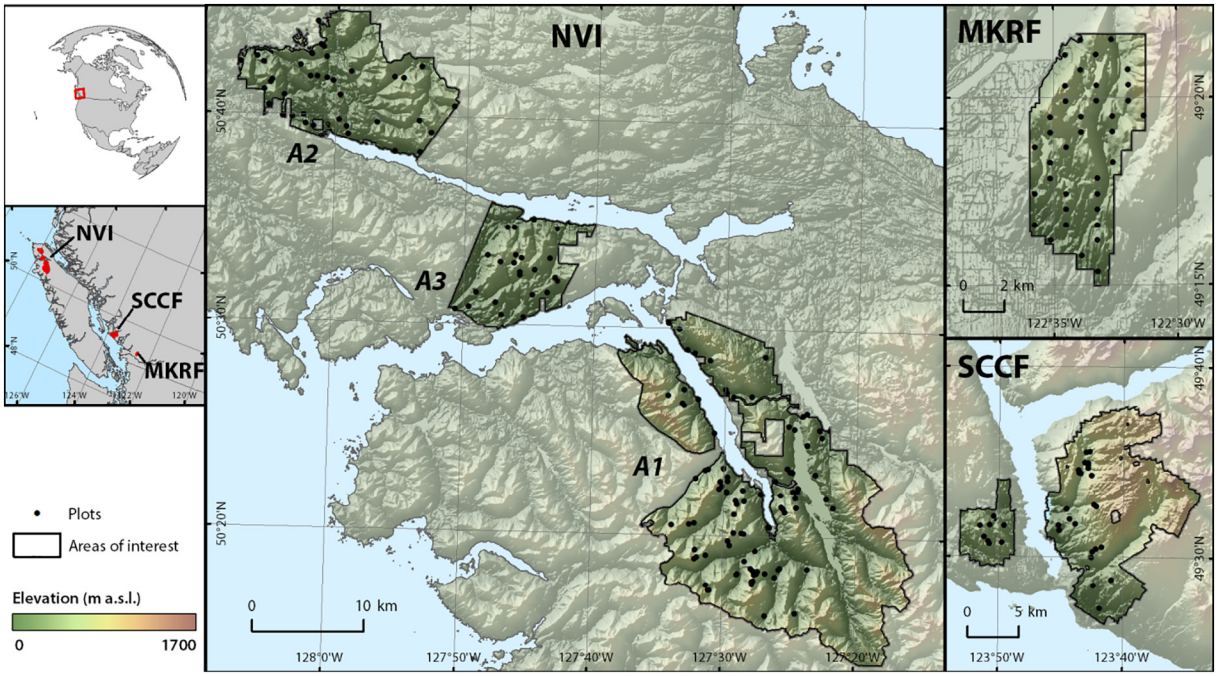

Fig. 1. Outline of the study areas used in the presented research. NVI – North Vancouver Island and sub-areas A1, A2, and A3; MKRF – Malcolm Knapp Research Forest; SCCF – Sunshine Coast Community Forest. The distance between NVI and SCCF was approximately 270 km, and between NVI and MKRF approximately 350 km.

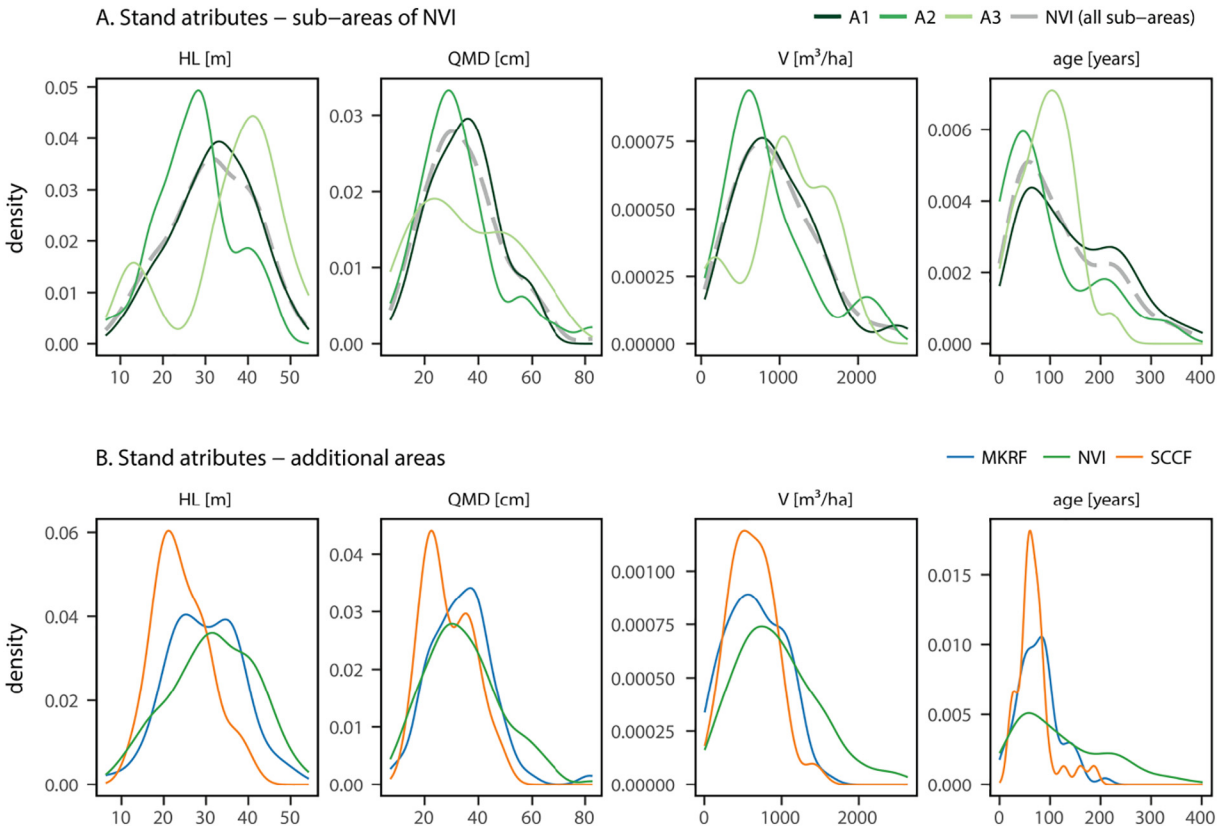

Fig. 2. Comparison of forest stand characteristics for different sub areas of NVI (A1, A2, and A3), and the additional areas used in the study (MKRF and SCCF).

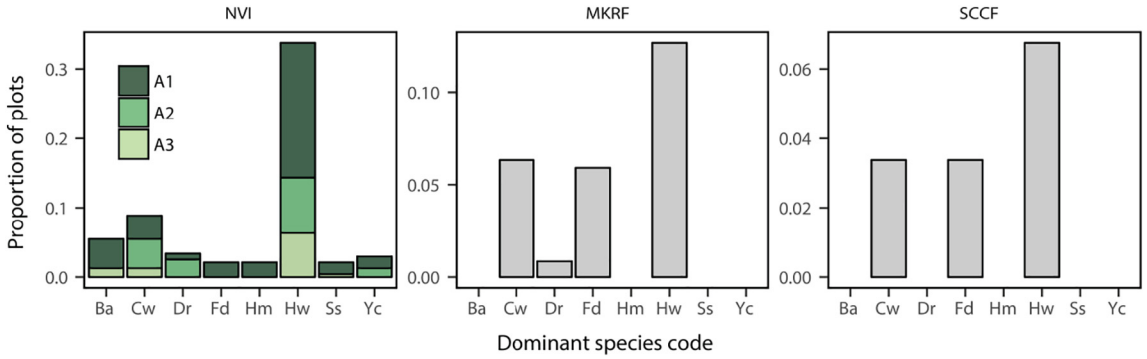

Fig. 3. Comparison of the species composition in for different sub areas of NVI and the additional area used in the study (MKRF and SCCF). Species codes: Ba – amabilis fir, Cw – western red cedar, Dr – red alder, Fd – Douglas-fir, Hm – mountain hemlock, Hw – western hemlock, Ss – sitka spruce, Yc – yellow cedar.

distributed following stratified random sampling with centers measured with differentially corrected GNSS measurements with sub-meter planimetric precision. Within each plot, all live standing trees with diameter at breast height (DBH) ≥12.0 cm were measured. Individual tree measurements included DBH (cm), tree height (m), and species. Individual tree-based estimates of stem height and diameter were used to compute estimates of Lorey's mean height (HL), quadratic mean diameter (QMD), and gross volume per hectare (V). The total number of plots was 144, 63, and 32, for NVI, MKRF, and SCCF, respectively. The A1, A2, and A3 sub-areas of NVI consisted of 84, 38, and 22 plots, respectively. A summary of ground plot characteristics for each of the sub-areas of NVI, as well as for MKRF and SCCF is presented in Figs. 2 and 3.

2.3. Point-cloud data

2.3.1. ALS point clouds

ALS point clouds for NVI were acquired in September of 2012 using an Optech ALTM3100EA scanning system operated at an altitude of approximately 700 m above ground level (Table 1). The average pulse density was 11.6 pulses/m^2 (Tompalski et al., 2015). ALS data for MKRF and SCCF were acquired using a Riegl LMS-Q1560 system. These datasets had a higher pulse density of 17.2 and 15.4 pulses/m^2, respectively and were acquired with a higher scan angle than for the NVI (58° vs 25°). The minimum flightline overlap for all three acquisitions was 60%.

2.3.2. NVI DAP

Digital imagery was acquired for the NVI study area using a Vexcel UltraCamX (Vexcel Imaging, 2018) camera in September 2012. The imagery was 4-band (RGB and NIR) with a 0.30 m ground sampling

Table 1
Airborne laser scanning (ALS) data acquisition parameters.

	NVI	MKRF	SCCF
Acquisition date	September 2012 (leaf off)	June 2016 (leaf on)	November 2015 (leaf off)
Sensor	Optech ALTM3100EA	Riegl LMS-Q1560	Riegl LMS-Q1560
Flying height	700 m AGL	1200 m AGL	1100 m AGL
Flight speed	240 km/h	210 km/h	260 km/h
Pulse repetition rate	70 kHz	533 kHz	533 kHz
Scan angle	25° (± 12.5°)	58° (± 29°)	58° (± 29°)
Beam divergence	0.3 mrad	0.25 mrad	0.25 mrad
Average pulse density [pulses/m^2]	11.6	17.2	15.4

distance (GSD) and was acquired along 6 flight lines, with a minimum 60% along-track and 20% across-track overlap. We used the semi-global matching (SGM) algorithm (Hirschmüller, 2008), as implemented in the Remote Sensing Software Package Graz (Joanneum Research Digital, 2018; RSG version 7.46.11) to generate dense image point clouds. Only along-track stereo pairs (n = 49) were used for image matching. The resulting point density of the image-based point cloud was 12.27 points/m^2 (White et al., 2015).

2.3.3. Point cloud data processing

ALS point clouds were processed following standard processing routines, which included tiling, ground classification, and height normalization. Processing was performed using a combination of point cloud processing tools: LAStools (Isenburg, 2017, version 180322), and the lidR library for R (R Core Team, 2013; Roussel and Auty, 2018, lidR v1.5.1). A tile size of 1000 × 1000 m with a 20 m buffer was used. Ground classification was based on adaptive TIN models implemented in LAStools (Axelsson, 2000). Points classified as ground were then used to normalize the ALS point cloud elevations relative to the ground surface. DAP point clouds were processed using similar routines. Data were tiled using a similar tile size and tile overlap. However, because the DAP data primarily characterize the top of the canopy, height normalization for each tile was performed by incorporating the ALS-based points classified as ground (White et al., 2015). Point clouds were then clipped to match the plots and metrics were then calculated for both the ALS and DAP point clouds. Metrics included measures of central tendency (mean, median, mode), measures of dispersion (variance, standard deviation, interquartile range), percentiles, proportions, and densities of point heights above ground. Metrics were calculated using all returns above 2 m.

2.4. Experimental design

To examine the transferability of ALS-derived models, we designed an experimental framework that included different modeling approaches and combinations of plot data to train and test the models. The modeling approaches included ordinary least squares regression (OLS), random forest (RF, implemented in randomForest package for R, version 4.6–14), and k-nearest neighbour imputation (kNN, implemented in yaimpute package for R, version 1.0–30). We used k = 5 and a random forest-based distance metric during kNN modeling and imputed the three response variables as part of the same model (Crookston and Finley, 2008). The experimental design included three scenarios. In the first scenario, we tested how prediction accuracies for HL, QMD, and V varied when area-based models were developed and applied on the different sub-areas of NVI (Fig. 1). Under this scenario, the point cloud characteristics were the same (the data are from the same ALS acquisition), but the areas were different. In this scenario we used different combinations of plot data available for subsets of NVI study area (i.e. A1, A2, and A3), to first develop a model and then apply it to each of the other three sub-areas. For example, in this scenario an OLS model predicting HL was developed using 84 reference plots in A1, and then was applied to areas A1, A2, and A3. This process was repeated for all sub-areas, resulting in nine training-testing combinations. By replicating the procedure for all modeling approaches and all stand attributes, there were in total 81 models developed and tested for this first scenario.

In the second scenario, we tested how differences in point cloud density influenced the prediction accuracies of HL, QMD, and V. In this scenario we altered the ALS point cloud data available for NVI and generated data with pulse densities ranging from 1 to 10 pulses m^{-2} (in 1 pulse m^{-2} increments). Point cloud thinning was performed by randomly selecting a proportion of lidar pulses (all echoes corresponding to a given pulse). For this scenario, we used only the A1 subset of NVI study area. We developed predictive models using each point cloud density variant, and applied it to every other density variant. This resulted in a total of 100 (10 × 10) combinations for each of our three modeling approaches for HL, QMD, and V (900 training-testing combinations in total).

In the third scenario we investigated how the prediction accuracy for HL, QMD, and V changes if ALS-derived models are applied to different areas with different point cloud data properties. In this scenario we included different types of point cloud data acquired over the same area (DAP data for NVI), as well as similar types of point cloud data acquired over different areas (ALS data MKRF and SCCF). The models were developed using all plot data available for NVI (for A1, A2, A3; n = 144 plots). The developed models for HL, QMD, and V were then applied to the DAP data for NVI, and the ALS data for MKRF and SCCF. This scenario resulted in a total of 27 model outcomes for evaluation (three modeling approaches used to predict three stand attributes, over three areas).

To facilitate the aforementioned transferability scenarios, we first developed global models for HL, QMD, and V using all available NVI ground plot data (n = 144). The accuracy of these global models was then used as a benchmark for assessing variation in prediction accuracy for each of the aforementioned modeling scenarios. OLS models were developed to be parsimonious, and for these models we carefully selected independent variables that resulted in the highest adjusted R^2 value, while keeping a single independent variable for the HL model, and a maximum of four variables for QMD and V, following the approach of Bouvier et al. (2015). We tested different combinations of variables as well as different variable transformations. We evaluated correlations between potential predictor variables and kept only predictors that had a correlation coefficient < 0.8. A log-log transformation was required to achieve linearity for models predicting QMD and V. To simplify the analysis, the same OLS global model forms were used within each scenario. In scenario 1, model parameters for each of the sub-areas were derived using subsets of plots located in A1, A2, or A3. In scenario 2, model parameters were derived for each point cloud density variant. In scenario 3 global models were used to derive predictions for the additional datasets (NVI DAP, MKRF, SCCF). Global models were similarly developed using RF and kNN, although model development was simplified for these non-parametric approaches, as the independent variables were selected automatically by each of the applied algorithms. As per the OLS models, we evaluated potential predictors for the RF and kNN approaches by examining the correlations between potential predictors and chose those predictors that had a correlation coefficient lower than 0.8.

2.5. Assessing model performance

Prediction accuracy was assessed by calculating the R^2 coefficient, bias (absolute and relative), and RMSE (absolute and relative). We also used a paired Wilcoxon signed rank test to evaluate the null hypothesis that the median difference between the observed and predicted values was zero at α = 0.05. In cases where the same data were used to train and test the models (i.e., global models), leave-one-out cross validation

was used for model assessment. In cases where models required data transformation (i.e. OLS), model statistics were calculated using the back-transformed variables with bias correction factors applied, as per Sprugel (1983).

The R^2 was calculated using the following formula:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (\hat{y} - y)^2}{\sum (\bar{y} - y)^2} \quad (1)$$

where SSE is sum of square error, SST is total sum of squares, y is the observed value, $\hat{y}$ is the predicted value, and $\bar{y}$ is the mean of observed values. This formula was also used to calculate the R^2 value for models developed with RF and kNN.

Relative and absolute bias and RMSE were calculated as follows:

$$bias = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \quad (2)$$

$$bias\% = \frac{bias}{\bar{y}} * 100 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (4)$$

$$RMSE\% = \frac{RMSE}{\bar{y}} * 100 \quad (5)$$

where: N is the number of plots, y_i is the observed value at plot i, $\hat{y}_i$ is the predicted value at plot i, and $\bar{y}$ is the mean of the observed variable over all plots.

The average difference in RMSE% and bias% between models applied locally and models transferred to different areas was calculated for each attribute, modeling approach, and scenario. For scenario 1 the change in accuracy (i.e. ΔRMSE%, Δbias%) was calculated as the mean difference between RMSE% or absolute bias% for models trained and tested on the same sub-area, and RMSE% or absolute bias% for models tested on different sub-areas. For scenario 2 the change in accuracy was calculated as the mean difference between RMSE% or absolute bias% for models trained and tested using the same point cloud density, and RMSE% or absolute bias% for models tested point clouds with different point cloud density. Finally, for scenario 3 the change in accuracy was calculated as the mean difference between RMSE% and absolute bias% for models tested using different point cloud datasets, and RMSE% and absolute bias% calculated for global models.

In addition, to provide more insights into the resulting accuracies of the transferred models, we compared relationships between the modeled variables and the metrics used as predictors. The successful application of an already existing predictive model developed in a different area, depends not only on the similarity of forest conditions, or the similarity of point cloud data. Rather, the success of model transferability is also influenced by the relationship of the dependent variable with the predictors that influence the accuracy of the model transfer. As a decrease in prediction accuracy of a transferred model may originate from a number of factors (i.e. ALS acquisition parameters, variation in ground plot size, forest conditions, etcetera), we performed a simple comparison of the relationship between five selected percentiles (10th, 50th, 75th, 90th, and 99th percentile of normalized point heights) and the first three principal components derived from all available metrics.

3. Results

3.1. Global area-based models: generation of baseline data

The evaluation of the global predictive models using leave-one-out-cross-validation is presented in Fig. 4, with model forms for the OLS modeling approach showed in Table 2. A power regression approach was used for modeling QMD and V. The proportion of explained variance was highest for models predicting HL, and was similar across modeling approaches; however, OLS based predictions of HL also had the lowest RMSE values (both absolute and relative). Predictions of QMD and V had R^2 values ranging from 0.13 to 0.59, with the lowest value reported for QMD modelled using kNN, and the highest achieved for V modelled using OLS. Models derived using the non-parametric approaches had a small bias%, which was largest for V modelled using kNN (−2.38%). The RMSE% was similar for each of the modelled stand attributes across modeling approaches. The largest RMSE% was for V modelled with RF and kNN, which were 37.88 and 48.95%, respectively. The results of the paired Wilcoxon signed-rank test indicated that we could not reject the null hypothesis that the median difference between the observed and predicted values was 0 for any of the modelled attributes and modeling approaches.

Predictors used in OLS models included 90th percentile of ALS heights (HL), 99th percentile, percent returns above 2 m, and kurtosis (QMD), and 75th percentile and percent returns above 2 m (V). Predictors with highest variable importance that were most frequently used in non-parametric approaches included standard deviation of point heights, 20th percentile of point heights, and proportion of returns in three height strata: 5–10 m, 20–30 m, and above 30 m.

3.2. Scenario 1: same point cloud characteristics; different areas

The R^2, RMSE%, bias%, and p-values, derived for different combinations of data used during model development and data used during model application in scenario 1 are shown in Fig. 5. Similar to the global models, for the three modeled stand attributes, the highest R^2, lowest RMSE% and lowest bias%, were achieved for HL, for all of the modeling approaches. For QMD and V, results were consistent with the results reported for the global models, with several exceptions. Typically, when the developed models were applied to different sub-areas of NVI, lower R^2 and higher RMSE% values were observed, with increased bias. For HL, the smallest differences between combinations using the same and different sub-area for model training and testing (e.g. A1 model applied to area A2, A3 model applied to area A1) were observed when an OLS approach was used. For example, the largest difference between RMSE% values calculated for different combinations of HL OLS model was only 2.0% (10.14–8.14). For QMD and V, the smallest differences for training-testing combinations were observed for OLS and RF approaches, with the differences in RMSE% and R^2 sometimes exceeding 30% and 0.4, respectively. Application of the kNN models over different areas resulted in greater differences in R^2, RMSE%, and bias%. In general for QMD and V, both RF and kNN models trained and tested at the same sub-area often yielded more accurate results than OLS, however the RF models were markedly more accurate than kNN when transferred to different sub-areas. For example, when A1 model for QMD was applied to sub-area A2, the RMSE% value increased from 26.45 to 31.06% for OLS, from 27.63 to 29.97% for RF, and from 30.92 to 60.17 for kNN.

For scenario 1, the results of the paired Wilcoxon signed rank test showed that when models were trained and tested in the same area, there was no significant difference between median observed and predicted values. However, in many cases when models were transferred and applied to a different area from where they were trained, there were significant differences between observed and predicted variables for the transferred models. Note that while there was no significant difference for estimates of HL derived from an OLS approach, there were significant differences for both RF and KNN.

While the same training-testing combinations (i.e. A1 to A1) resulted in the lowest bias values, the combinations when a different sub-area was used for model training and model testing (i.e. A1 to A2, or A1 to A3) resulted in a marked increase in bias for every modeling approach (Fig. 4). Contrary to the R^2 and RMSE% values, which were consistent when the OLS modeling approach was used, for bias there

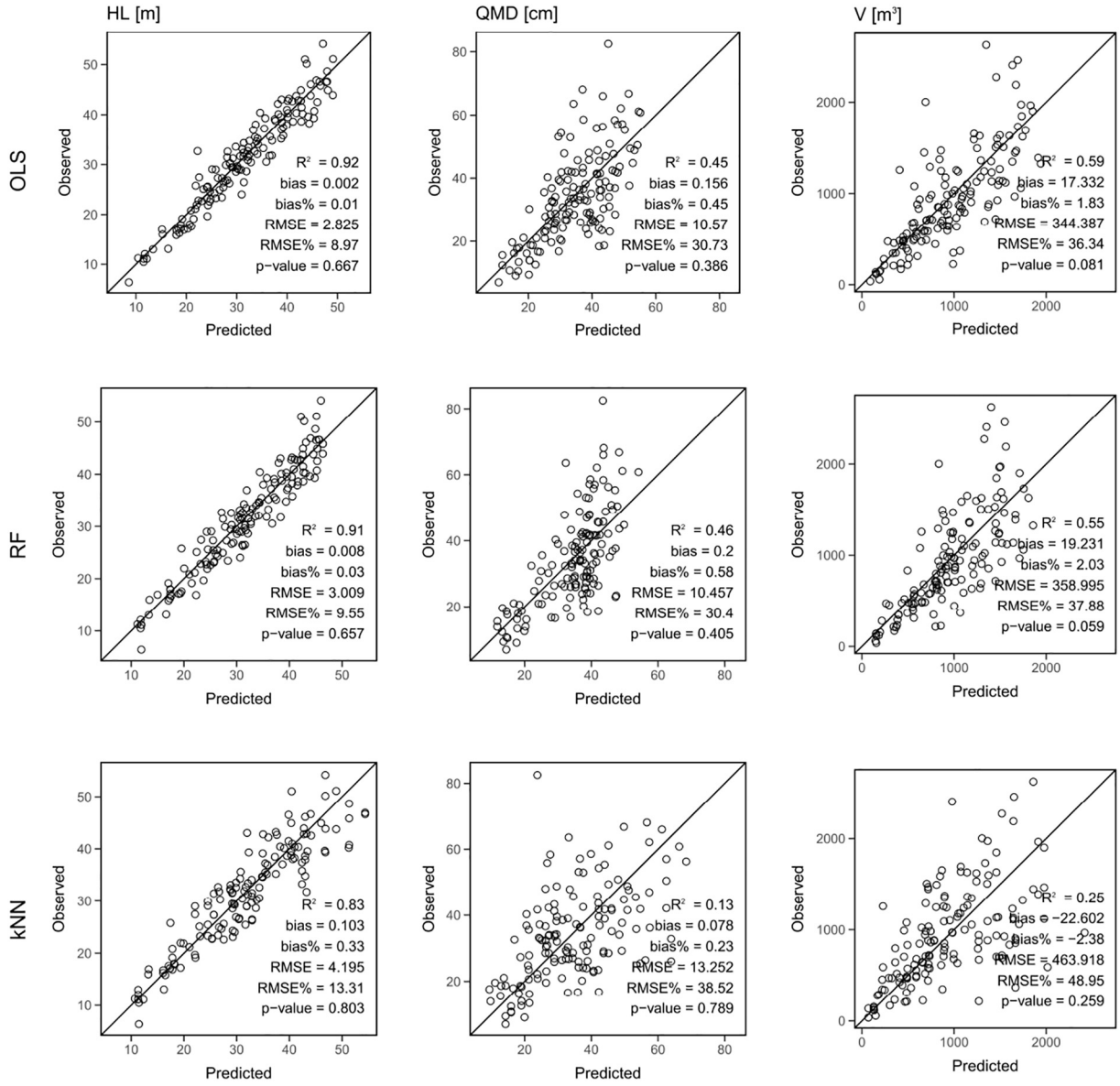

Fig. 4. Assessment of the global predictive models for HL, QMD, and V, developed using OLS, RF, and kNN modeling approaches and leave-one-out cross validation. Performance statistics derived using leave-one-out cross-validation.

was no clear advantage of OLS over the RF or kNN approaches. For example, the highest bias% values for V reported for OLS, RF, and kNN were similar (26.87, 15.4, 29.78%, respectively), and derived from the same training-testing combinations. In most cases, the large values of bias corresponded to training-testing combinations for which the *p*-values were lower than alpha (0.05).

3.3. Scenario 2: different point cloud characteristics; same area

The results for scenario 2, where the effect of different point cloud densities was analyzed, showed consistent performance statistics when the OLS modeling approach was used (Fig. 6) with a very minor change in R^2, RMSE%, and bias, when point clouds densities differed between training and testing datasets. When the non-parametric approaches were used, the differences in model outcomes were markedly greater, especially for QMD and V, although for HL, the values of R^2, RMSE%, and bias were more similar. For example, the distinct values of performance statistics on the diagonal of R^2 values for V modeled with kNN (Fig. 6) informs that the proportion of explained variance was markedly lower when the developed model was applied to a dataset

Table 2
Global predictive models developed with OLS. Predictors included 90th percentile of ALS heights (HL), 99th percentile, percent returns above 2 m, and kurtosis (QMD), and 75th percentile and percent returns above 2 m (V).

Dependent variable	Predictive model
Lorey's height (HL; m)	$2.34 + 0.92 * P90$
Quadratic mean diameter (QMD; cm)	$\exp(5.53 + 0.97 * P99 - 1.29 * perc\_return\_above\_2 + 0.25 * kurt)$
Gross total volume (V; m^3)	$\exp(5.47 + 1.58 * P75 - 0.89 * perc\_return\_above\_2)$

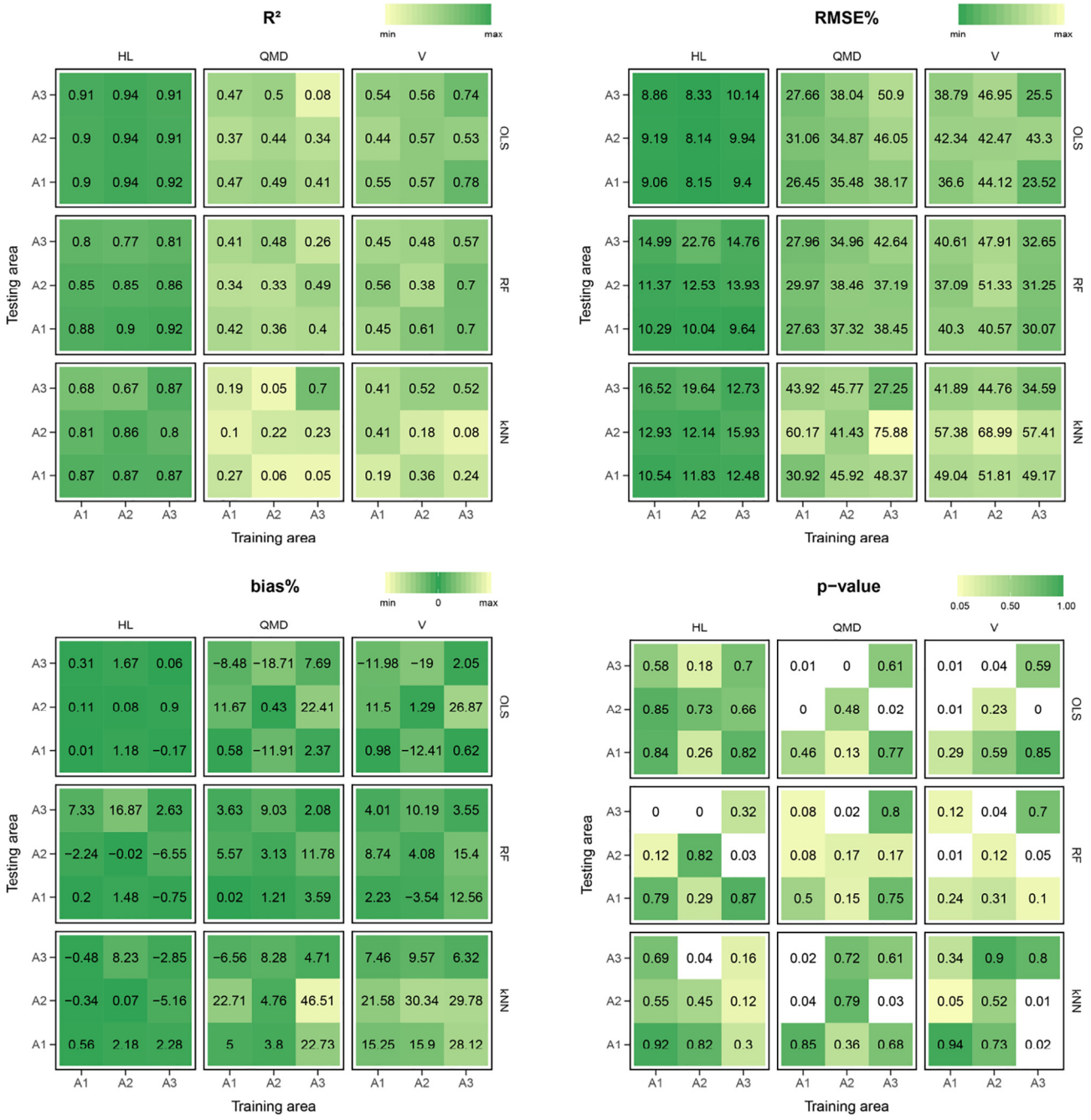

Fig. 5. Performance statistics calculated for models developed and applied on different sub-areas (A1, A2, and A3). *P*-values calculated with paired Wilcoxon test.

with different point density. In addition, it can also be observed that the difference in R^2 was greatest when the difference in point densities was largest, with the same effect for RMSE% but not for bias%. The results of the paired Wilcoxon signed-rank test indicated that the null hypothesis that the median difference between the compared values was 0 could not be rejected for any of the training-testing combinations.

3.4. Scenario 3: different point cloud characteristics; different areas

The results of the third scenario are shown in Fig. 7 and Table 3. Results indicate that when the global models derived for NVI based on ALS were applied using NVI DAP metrics as predictor variables, the predictions of HL contained only a small bias with RMSE% between 14.39 and 22.45% for each of the modeling approaches. Predictions of QMD and V were not as similar across the modeling approaches and were more biased and had higher RMSE than HL models. The most accurate predictions of QMD were derived using the RF model, with RMSE% of 35.50%, lowest among the modeling approaches. Predictions of V using OLS and RF were more accurate than predictions derived with kNN and had similar RMSE% (43.05 and 40.25%, respectively) and R^2 values (0.43 and 0.49, respectively). In addition, the application of the kNN approach can result in this phenomenon where values are replicated from the nearest neighbour, visible as vertical stripes on scatterplot. The only case for which we could reject the null hypothesis that the median difference between the compared values was 0 was the prediction of QMD using OLS.

When the models were applied to the MKRF and SCCF datasets, prediction accuracies were similar for both of these areas. Results were consistent for HL, with similar accuracies reported for all modeling approaches and for both MKRF and SCCF datasets. The R^2 values were

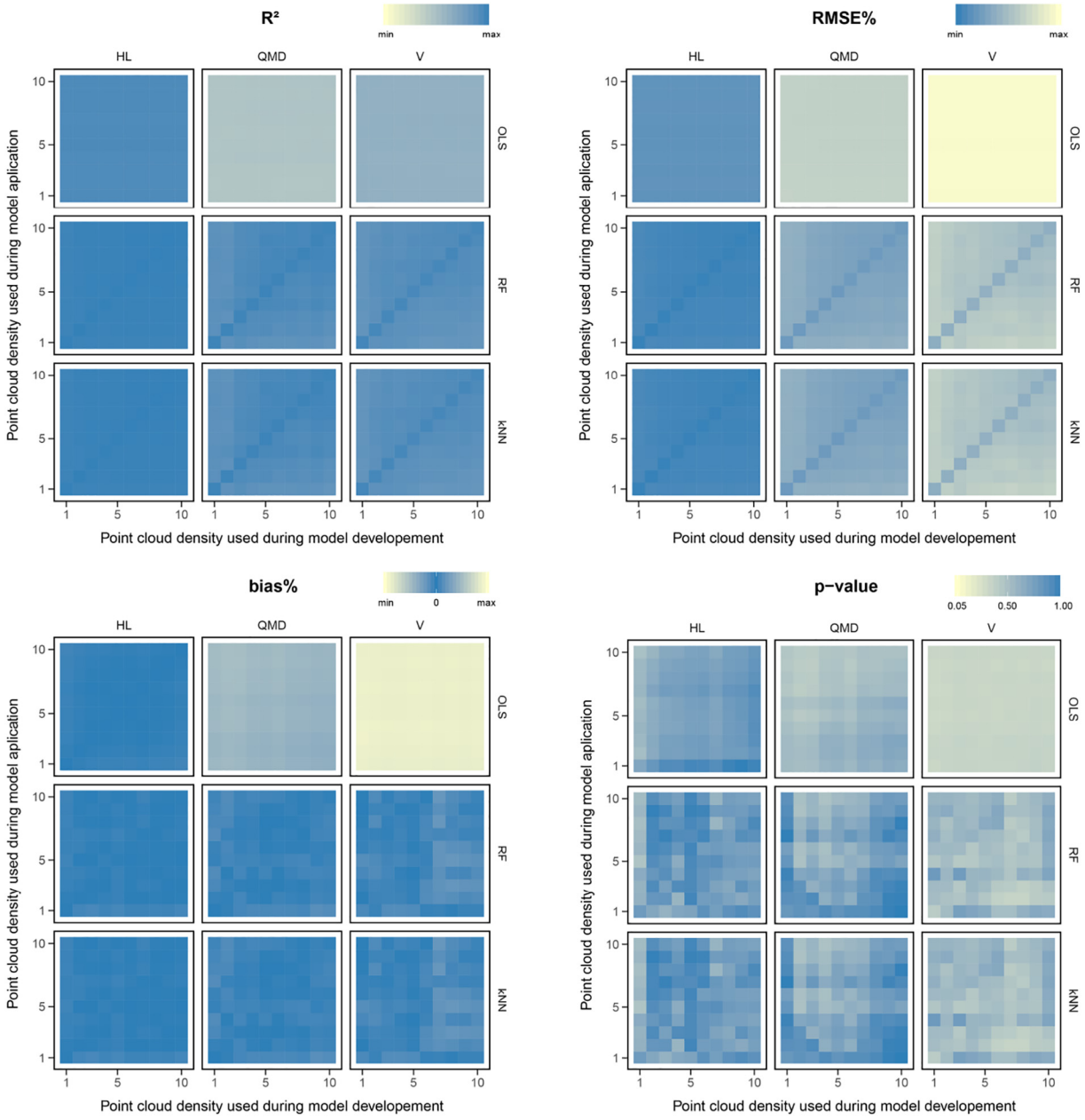

Fig. 6. Model performance statistics derived for models derived and applied on datasets with different point cloud density. Statistics calculated for A1 subset area only.

higher than 0.8 for all but one case (HL modeled with kNN in SCCF; $R^2 = 0.68$). The relative RMSE ranged between 10.47 and 14.99%, with lowest values of HL predicted for MKRF using RF. We observed a positive bias between 4.49% and 4.99% for HL predictions at MKRF, and a negative bias between −1.56% and −2.41% for SCCF. In both cases, the smallest bias was reported for predictions derived with kNN models.

Predictions of QMD and V were less consistent, with larger discrepancies between the performance statistics. For example, the RMSE % values reported for QMD predicted for MKRF were very high when the OLS approach was used (109.65%), while the values for RF and kNN were much lower (23.14% and 28.25%, respectively). For QMD the highest prediction accuracy was achieved with RF models for both MKRF and SCCF, with R^2 values of 0.58 and 0.36 respectively. Predictions of V were inaccurate for MKRF, with very low R^2 values (maximum of 0.18), extreme bias and RMSE. These predictions were unique as the least accurate across all variables, modeling approaches and datasets. For SCCF predictions of V were on par with what we observed for HL and QMD, with highest accuracy achieved using RF. The RF-based predictions resulted in R^2 of 0.82, bias% of 2.34%, and relative RMSE of 20.09%.

For the majority of our study areas, the relationship between the five selected height percentiles and the stand attributes (HL, QMD, and V) was similar (Fig. 8). Likewise, there was a similar relationship between the first three principal components of the ALS metrics and the stand attributes (HL, QMD, and V; Fig. 9). For example, the relationship of HL with the 90th percentile of normalized point heights (Elev.P90) was very similar for every dataset used. Also the relationship between the principal components derived from all point cloud metrics (especially PCA1) and HL was similar for all study area datasets, with a large overlap of data points among the four study areas. On the other hand, there was a clear difference in the relationship of height percentiles and V in MKRF and the other three datasets. For example, based on the linear trend summarizing the relationship between the 75th percentile (Elev.P75) and V, the value of 40 m (Elev.P75) corresponds to

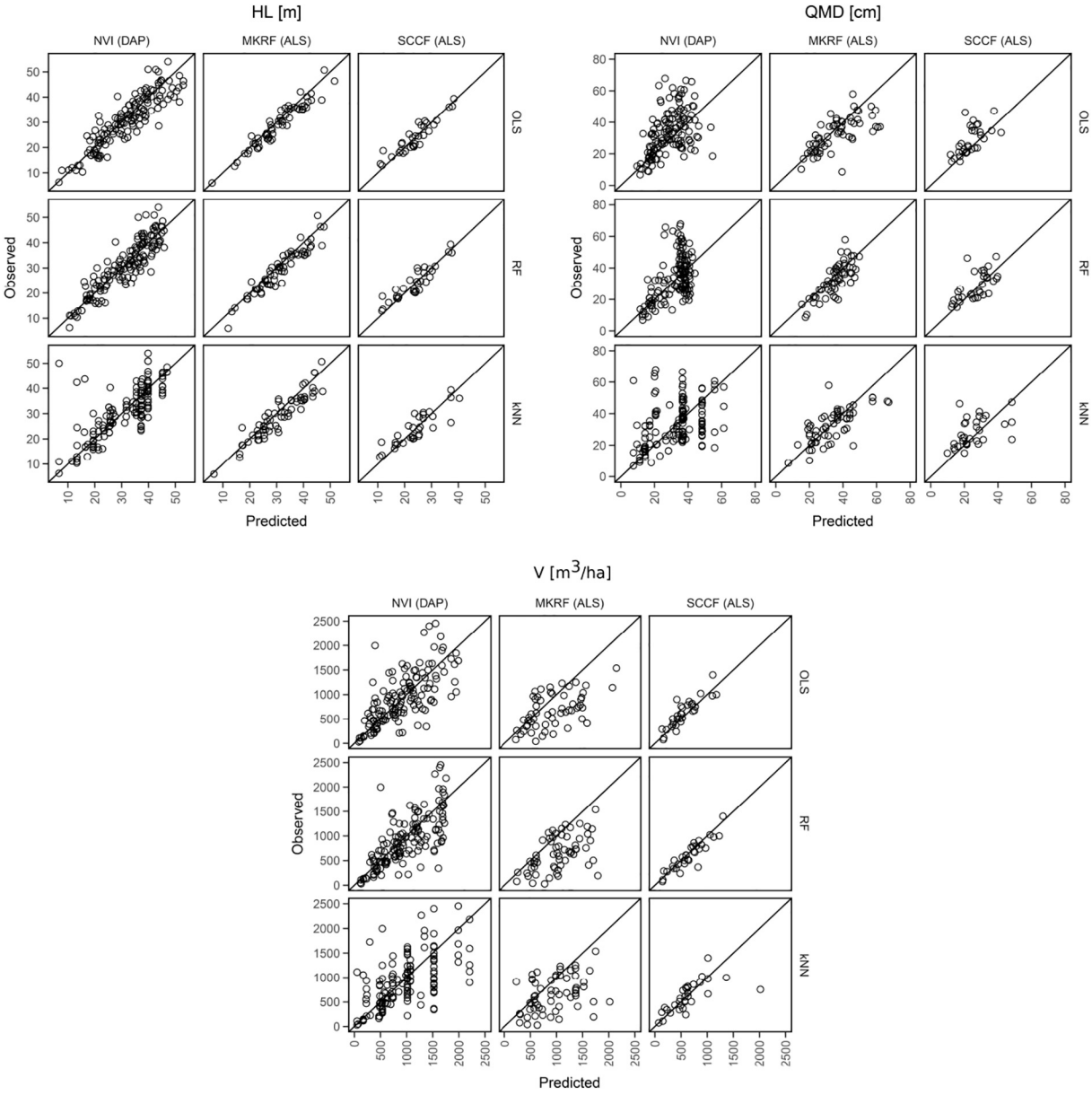

Fig. 7. Scatterplots of the three stand attributes (HL, QMD, and V) predicted using DAP data for NVI, ALS data for MKRF, and ALS data for SCCF, using transferred models originally developed using ALS data for NVI.

approximately 1500 m^3/ha for the ALS and DAP datasets in NVI, and the SCCF dataset. However, the same value of 40 m (Elev.P75) corresponds to much lower V in MKRF – approximately 1000 m^3/ha. This dissimilarity in relationship of stand attributes and point cloud metrics for MKRF is also somewhat reflected in the comparison of principal components and stand attributes. The overlap between data points originating from different datasets indicates discrepancies in the relationship that stand attributes and metrics have in MKRF and SCCF.

3.5. Overall summary of average change in accuracy for transferred models

The average differences in bias% and RMSE% for all three scenarios are summarized in Table 4. The summary confirms that when the models were transferred, bias% and RMSE% increased for almost all variables and modeling approaches; however, the decrease in accuracy varied by attribute, modeling approach, and scenario. Across all scenarios, the smallest changes in bias% and RMSE% were observed for HL modeled with OLS. For example, the average difference in bias% for HL modeled with OLS under Scenario 1 was only 0.67%. While the differences for non-parametric model estimates of HL were higher, they exceeded 5% in only one case (scenario 1, HL modeled using RF). For QMD and V, the differences were much higher and in some cases exceeded 20%. For example, under Scenario 3 the average increase in bias % and RMSE% for V modeled with OLS was 22.04% and 23.54%, respectively, while the average increase in bias% and RMSE% for QMD modeled with OLS was 20.20% and 29.31%, respectively. For these two cases the decrease in accuracy was highest across all scenarios and modeling approaches. Contrary to HL modeled with OLS, the lowest differences in bias% and RMSE% were observed for QMD and V models developed with RF, which in many cases were markedly lower than differences in bias% and RMSE% for OLS and kNN models. For example, in scenario 1 the average increase in bias% for QMD modeled with RF was 4.49%, while for QMD modeled with OLS and kNN the average increase was 9.69% and 12.80%, respectively. Overall, the smallest changes in accuracies were observed in Scenario 2, where the average change in bias% was never larger than 0.5% and the difference

Table 3
Prediction accuracy of the three stand attributes (HL, QMD, and V) predicted using DAP data for NVI, ALS data for MKRF, and ALS data for SCCF, using transferred models originally developed using ALS data for NVI.

Stand attribute	Modeling approach	Area and data type	R^2	bias	bias%	RMSE	RMSE%	*p*-Value
HL	OLS	NVI (DAP)	0.79	0.55	1.74	4.69	14.90	0.1773
		MKRF (ALS)	0.87	1.36	4.60	3.32	11.23	0.0003
		SCCF (ALS)	0.84	−0.54	−2.24	2.59	10.65	0.3896
	RF	NVI (DAP)	0.80	−0.26	−0.81	4.53	14.39	0.6329
		MKRF (ALS)	0.90	1.48	4.99	3.10	10.47	0.0000
		SCCF (ALS)	0.80	−0.59	−2.41	2.91	11.95	0.3694
	kNN	NVI (DAP)	0.51	−0.63	−2.00	7.07	22.45	0.8434
		MKRF (ALS)	0.85	1.33	4.49	3.57	12.07	0.0036
		SCCF (ALS)	0.68	−0.38	−1.56	3.64	14.99	0.4319
QMD	OLS	NVI (DAP)	0.20	−5.70	−16.56	13.92	40.46	0.0000
		MKRF (ALS)	0.05	9.19	27.30	36.92	109.65	0.0259
		SCCF (ALS)	0.40	−5.18	−18.09	8.58	30.00	0.0001
	RF	NVI (DAP)	0.29	−2.37	−6.88	12.21	35.50	0.1548
		MKRF (ALS)	0.58	1.53	4.54	7.79	23.14	0.0112
		SCCF (ALS)	0.36	−2.07	−7.24	7.34	25.65	0.1386
	kNN	NVI (DAP)	0.15	−1.81	−5.26	15.43	44.87	0.3721
		MKRF (ALS)	0.35	−0.13	−0.38	9.51	28.25	0.9209
		SCCF (ALS)	0.23	−3.44	−12.04	10.21	35.70	0.0253
V	OLS	NVI (DAP)	0.43	−48.15	−5.08	408.00	43.05	0.2350
		MKRF (ALS)	0.03	343.03	52.51	718.89	110.04	0.0000
		SCCF (ALS)	0.77	−86.37	−14.02	163.51	26.54	0.0014
	RF	NVI (DAP)	0.49	−15.83	−1.67	381.68	40.27	0.9610
		MKRF (ALS)	0.03	315.59	48.31	479.23	73.35	0.0000
		SCCF (ALS)	0.82	14.40	2.34	123.76	20.09	0.8320
	kNN	NVI (DAP)	0.19	1.75	0.18	481.85	50.84	0.7055
		MKRF (ALS)	0.18	268.23	41.06	510.54	78.15	0.0000
		SCCF (ALS)	0.06	4.24	0.69	279.51	45.36	0.2386

in RMSE% exceeded 5% in only two cases.

4. Discussion

In this research, we demonstrated the effects of transferring existing predictive models of three selected forest stand attributes to areas with similar or different point cloud data characteristics, and similar or different forest stand structures. We used three different modeling techniques that included parametric (OLS) and non-parametric approaches (RF, kNN) and used them to estimate HL, QMD, and V. We partitioned the analysis into three scenarios that allowed for the examination of different questions related to model transferability. Our results indicate that the accuracy of existing predictive models applied to areas of new point cloud data acquisitions depends on the modeled attribute, modeling approach, similarity of forest conditions, and similarity in the relationship between predicted forest stand attribute and

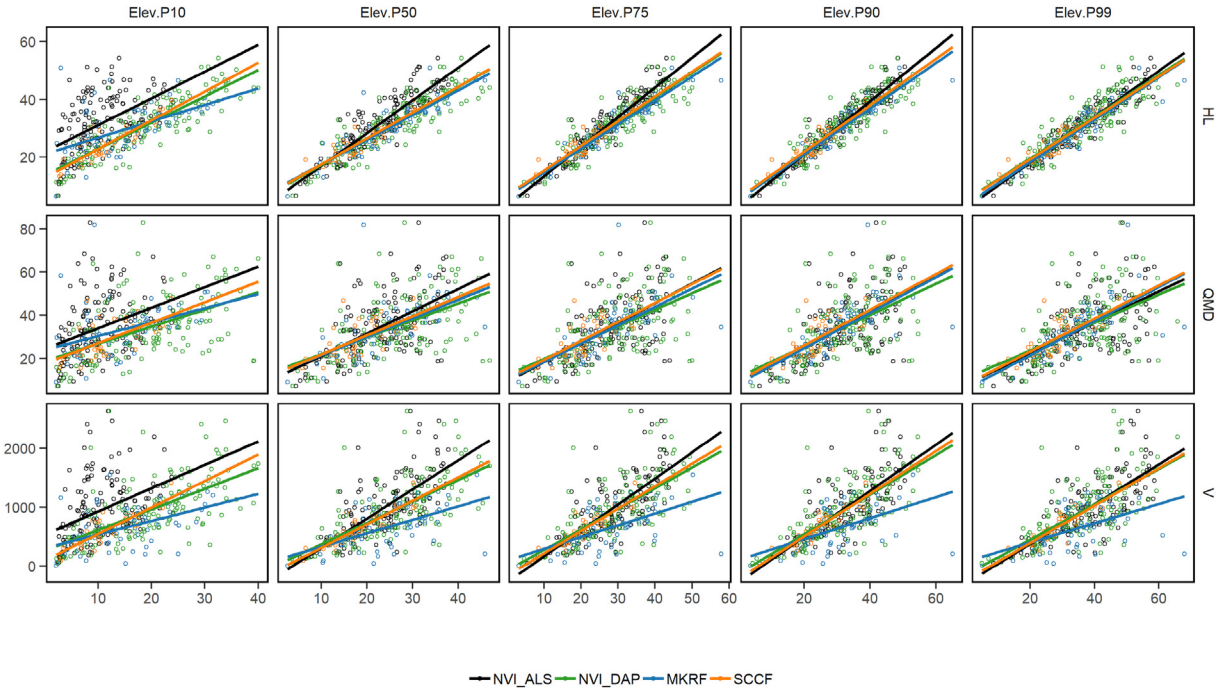

Fig. 8. Relationships between the three stand attributes (HL, QMD, V) and five selected point cloud metrics derived from the four available point cloud datasets (colors). Elev.PXX – XXth percentile of normalized point heights, where XX = {10, 50, 75, 90, 99}.

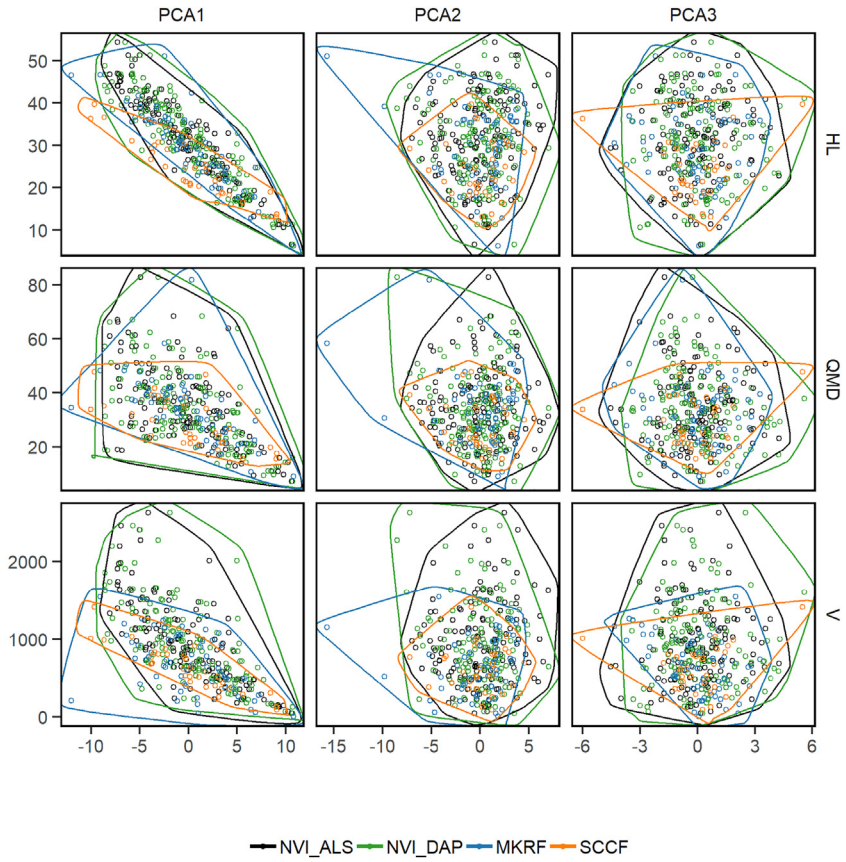

Fig. 9. Relationships between the three stand attributes (HL, QMD, V) and the first three principal components generated from all available point cloud metrics (PCA1, PCA2, PCA3). Comparison presented for all four available point cloud datasets (colors).

point cloud metrics. From the three selected forest stand attributes considered, HL was the most robust to transferability. When the forest conditions and ALS data acquisition parameters were similar (scenario 1), the application of transferred models resulted in greater prediction accuracies compared to scenario 3, when the forest conditions and data acquisition were different, however the accuracy in this context depended strongly on modeling approach. Results showed that the differences in accuracies between predictions of HL derived for training and testing areas under scenario 1 were lowest for OLS models. For predictions of QMD and V, differences between training and testing areas were comparable between OLS and RF, although often a much larger relative bias was introduced when OLS models were transferred. When compared to the accuracy of global models, RF models of QMD and V were most stable, with lowest decrease of accuracy.

We found that there was only a small decrease in prediction accuracy when point clouds of different pulse densities were used, with OLS models being the most consistent and least sensitive to changes in predictors. This result shows that in the same forest environment, existing models can be used to predict forest attributes of interest using new data acquisitions, even if they differ in point density. This is in keeping with the results reported in the literature, which indicate that area-based prediction accuracy does not increase with increasing pulse density (Jakubowski et al., 2013). We acknowledge the fact that data density is not the only attribute that changes with different ALS data acquisition parameters, and that to fully understand the effects of differences in these parameters on model transferability, separate ALS acquisitions covering the same area, but with different parameters are required for robust benchmarking and disentangling of interactive effects of these parameters, particularly for scan angle effects (Montaghi, 2013; Roussel et al., 2018). However, in many cases the acquisition parameters (e.g. sensor, acquisition height, speed, scan angle, number of returns per pulse) are very similar for new data collections, especially

Table 4
Average differences in bias% and RMSE% for the transferred models. Positive values (increase in bias% or RMSE%) indicate a decrease in prediction accuracy.

Attribute	Method	Scenario 1		Scenario 2		Scenario 3	
		Δbias%	ΔRMSE%	Δbias%	ΔRMSE%	Δbias%	ΔRMSE%
HL	OLS	0.67	−0.14	0.12	0.00	2.85	3.29
	RF	6.32	1.51	0.10	1.25	3.55	3.75
	kNN	1.17	2.94	0.10	1.25	2.87	3.20
QMD	OLS	9.69	−1.33	−0.01	0.02	20.20	29.31
	RF	4.49	0.75	0.20	3.85	8.78	−2.24
	kNN	12.80	23.55	0.20	3.85	10.00	−4.27
V	OLS	12.29	4.98	−0.01	0.00	22.04	23.54
	RF	5.03	−2.69	0.17	6.43	19.58	9.52
	kNN	8.88	0.23	0.17	6.43	16.92	7.01

for datasets collected over large areas, or when the data is collected by the same vendor. Under such assumptions, the non-parametric modeling approaches were more sensitive than OLS for changes in point cloud densities, and the differences in prediction accuracies were more marked for QMD and V, compared to HL. The consistency in prediction accuracy of the OLS models was especially reflected in bias, which had similar values for the transferred models and the models trained and tested using the same pulse density, and was much lower than the bias introduced when the RF and kNN models were transferred. Interestingly, for both QMD and V, the R^2 values were higher and RMSE values were lower for both non-parametric approaches when compared to OLS, for all transferred models in scenario 2.

Although Roussel et al. (2017) demonstrated that pulse density affects calculations of point cloud metrics using mean point height as an example in a northern hardwood forest, our results showed that differences in metrics calculated for point clouds with different pulse densities have very little effect on the prediction accuracy of transferred OLS models. Our findings are similar to those reported by Yoga et al. (2017), who found no effect of density on the predictions using nonlinear generalized least squares regression. The influence of differences in pulse density between point clouds used during model development and model application were more pronounced for RF and kNN (Fig. 6). We theorize that the larger decrease in accuracy of the non-parametric models may be caused by the fact the non-parametric models cannot extrapolate beyond the range of data used to train them, as well as the different effects that pulse density has on different ALS metrics. Because the OLS models were developed using a maximum of three independent variables, the cumulative effect of changes in metric values on model outcome was less noticeable than for RF or kNN, which incorporated more predictors.

We found that models developed using ALS data can be applied using DAP-based predictors over the same area, and that although the results were less accurate than when ALS-based metrics were used, the achieved accuracy could be considered acceptable. The fact that the results were comparable for both parametric and non-parametric approaches, contrary to the results in Scenarios 1 and 2, demonstrated the agreement between ALS- and DAP-based metrics. This agreement, using the same data, was previously demonstrated by White et al. (2015) who found that the correlation between ALS- and DAP-based metrics increased with the increase of canopy cover, and was higher for upper canopy metrics. Because the upper canopy metrics are typically the most important predictor variables for both ALS and DAP-based models (Penner et al., 2013; Stepper et al., 2017), the high correlation between the metrics based on ALS and DAP data resulted in similar predictions. The prediction accuracies of the ALS-derived models applied to DAP-based metrics were lower than the accuracy of DAP-derived models presented in White et al. (2015). The differences were however markedly small for HL, with RMSE% of 14.0 reported by White et al. (2015) for models developed with DAP data, compared to RMSE% between 14.39 and 22.45% for the transferred models. The differences were also small for V – the bias% and RMSE% for the DAP-derived models reported by White et al. (2015) were −1.21 and 37.68%, respectively, while the most accurate predictions resulting from the model transfer had bias% of −1.67% and RMSE% of 40.27%. The most accurate predictions for the transferred models were derived with models developed using RF, which was also the modeling approach used by White et al. (2015). However, the reported set of the most important predictors used in White et al. (2015) was different from the predictors used in this study.

Transferring models developed using NVI data to areas with similar forest types (MKRF, SCCF) resulted in similar prediction accuracies in both areas, especially for HL. However, contrary to Scenarios 1 and 2, we could not identify a modeling approach that would consistently provide the greatest accuracies for the three stand attributes when the models were transferred to different areas or used with DAP data instead of ALS point clouds. Overall, in a majority of cases, transferred RF models provided slightly more accurate results, however, by comparison, OLS models had only slightly lower accuracies.

Predictions of V in MKRF were markedly less accurate compared to those in SCCF, even though the point cloud and stand attributes were similar in these two areas. By analyzing the relationship between the point cloud metrics and the modeled attribute, we found that the relationship is different for MKRF compared to other study areas (Fig. 8). Specifically, for similar values of height percentiles, volume had markedly lower values. This different relationship is reflected in the overestimated predictions of V in MKRF. We find it interesting that the difference in attribute-metrics relationship occurred only for V, and only in MKRF. The MKRF has similar dominant species and age distributions relative to the other areas.

ALS data for MKRF and SCCF were acquired with higher maximum scan angle than in NVI. The influence of scan angle on height metrics has shown to be minimal for acquisition with a maximum scan angle of 20° and that narrower scan angle (15 vs 30°) leads to slightly higher accuracies of volume estimates (Holmgren et al., 2003; Keränen et al., 2016). Although existing studies recommend that the maximum scan angle does not exceed 20° (Gatziolis and Andersen, 2008; White et al., 2013), the degree of overlap between the flight lines, which was ≥60% for all acquisitions used herein, must also be considered. Keränen et al. (2016) found that when the models were applied to datasets with different acquisition parameters, the bias increased by up to 10% for volume, although the RMSE% remained similar. They conclude that models should only be used for datasets collected under similar conditions. It is therefore possible that in our case the decrease of accuracy for models transferred from NVI to MKRF and SCCF was caused by the differences in the maximum scan angle used during ALS data acquisition. However, this effect is difficult to characterize because the same maximum scan angle was used for MKRF and SCCF ALS datasets. Therefore any marked effect of scan angle on the prediction accuracy of the transferred model should exist in both of the datasets, which is not the case. However, Keränen et al. (2016) found that height distributions also influenced the accuracy of transferred models. On average, heights in the MKRF are 5.29 m higher than those at SCCF, and 1.90 m lower than those at NVI. Further research would be required to quantify the effects of data acquisition parameters on the transferred models.

In addition, the accuracy of the models transferred from NVI to MKRF or SCCF may be influenced by the different plot sizes used in the study areas (625 m^2 in NVI vs 400 m^2 in MKRF and SCCF). However, given the relatively high point density of the ALS data (> 10 pulses/m^2 for all datasets), the plots used in all three areas are large enough to capture a sufficient number of returns that will result in a reliable characterization of canopy height distributions (Hudak et al., 2012). As demonstrated in this study (scenario 2), the effect of the number of pulses on accuracy of transferred models and thus on the calculated metrics is small. Hudak et al. (2016) demonstrated the consistency in ALS height and density metrics despite an 8-fold difference in plot size, while Packalen et al. (2019) quantified a maximum bias of 1.50% and a maximum change in relative RMSE of 0.97% when there was a 4-fold difference in plot size. We therefore likewise assumed that although there is a small difference in plot size, this difference likely had only a minimal influence on the model predictions.

Interest in applying existing models to predict forest inventory attributes is of high value not only from the perspective of extending the area of ALS data acquisitions, but also for updating forest inventories when new point cloud data is acquired over the same area. As demonstrated by Fekety et al. (2015) such temporal transferability is not only possible, but also allows for the development of “pooled models” that are based on plot and point cloud data acquired at different times (Kotivuori et al., 2016; Nilsson et al., 2016). Such models are therefore developed using an increased number of samples, which could potentially improve model results, particularly for non-parametric approaches to area-based modeling, which we have shown to be more sensitive to the range of calibration data used for modeling (Jin et al.,

2018). The potential of transferring already developed models temporally (Fekety et al., 2015) and spatially (Foody et al., 2003), as demonstrated in this study, enables investments in model development (including ALS and ground plot data acquisition costs) to be further leveraged, and provides opportunities to reduce costs for the operational implementation of area-based approaches through a reduction in the need for ground plot data.

5. Conclusions

The degree to which area-based models can be transferred to new areas or different point cloud data, and thereby reduce the requirement for extensive ground data collection, is of operational relevance for forest industry stakeholders, forest managers, and landowners. Our results demonstrate that the transferability of area-based models depends on the attribute, and the modeling approach applied, and to a much lesser extent, the point cloud characteristics. First, the stand conditions in the areas of new acquisitions should be similar to the forest stands where the models were developed. All of our study areas represent mature, coastal temperate forests dominated by western hemlock; however, these are natural forests that represent more complex stand conditions than the more uniform and intensively managed boreal forests that have been the primary focus of the majority of transferability investigations to date. Second, HL was the attribute that was most robust to transferability and QMD the least, while results for volume were often middling. Third, OLS regression was the most successful approach for model transferability for HL, maintaining prediction accuracies when models from one location were applied elsewhere. RF provided better transferability performance for QMD and V when models were transferred to different areas with either similar or different point cloud characteristics. The degree to which models can be transferred successfully depends on trade-offs between tolerance for decreases in estimation accuracy and the cost of acquiring additional field calibration plots. Variability in model outcomes by attribute and modeling approach, as demonstrated herein, highlight the challenges to making universal recommendations that will ensure successful model transferability. While specific details on the required number of plots and calibration methodology remain to be determined, from an operational perspective, it is reasonable to expect that a small subset of local calibration plots would need to be installed to improve the prediction accuracy of transferred models and optimize the value of new ALS acquisitions for forest attribute modeling.

Acknowledgements

This research was supported by the Canadian Wood Fibre Centre (CWFC) of the Canadian Forest Service, Natural Resources Canada, with data for northern Vancouver Island generously supplied by Western Forest Products Inc. and BC Timber Sales. Support was also provided by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to Nicholas Coops. Part of this research was funded through an NSERC CRD grant to Coops and Sunshine Coast Community Forest. We thank David Lesser for access to the ALS data at SCCF and Anna Yuill for leading the field data collection at SCCF. We thank MKRF staff for providing technical information and ALS data for the study. We thank the editors and three anonymous reviewers for their time and insightful comments that allowed us to improve our manuscript. Open Access supported by the Government of Canada.

References

Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J.S., van Breugel, M., 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia 168 (4), 1147–1160. https://doi.org/10.1007/s00442-011-2165-z.

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

Bohlin, J., Wallerman, J., Fransson, J.E.S., 2012. Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. Scand. J. For. Res. 27 (7), 692–699. https://doi.org/10.1080/02827581.2012.686625.

Bouvier, M., Durrieu, S., Fournier, R.a., Renaud, J.-P., 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sens. Environ. 156, 322–334. https://doi.org/10.1016/j.rse.2014.10.004.

Brosofske, K.D., Froese, R.E., Falkowski, M.J., Banskota, A., 2014. A review of methods for mapping and prediction of inventory attributes for operational forest management. For. Sci. 60 (2). https://doi.org/10.5849/forsci.12-134.

Crookston, N.L., Finley, A.O., 2008. yaImpute: an R package for κNN imputation. J. Stat. Softw. 23 (10), 1–16 (doi:doi).

Deo, R.K., Russell, M.B., Domke, G.M., Andersen, H.E., Cohen, W.B., Woodall, C.W., 2017. Evaluating site-specific and generic spatial models of aboveground forest biomass based on Landsat Time-Series and LiDAR strip samples in the Eastern USA. Remote Sens. 9 (6). https://doi.org/10.3390/rs9060598.

Eid, T., Gobakken, T., Næsset, E., 2004. Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scand. J. For. Res. 19 (6), 512–523. https://doi.org/10.1080/02827580410019463.

Fekety, P.A., Falkowski, M.J., Hudak, A.T., 2015. Temporal transferability of LiDAR-based imputation of forest inventory attributes. Can. J. For. Res. 45 (4), 422–435. https://doi.org/10.1139/cjfr-2014-0405.

Fekety, P.A., Falkowski, M.J., Hudak, A.T., Jain, T.B., Evans, J.S., 2018. Transferability of Lidar-derived basal area and stem density models within a Northern Idaho Ecoregion. Can. J. Remote. Sens. 44 (2), 131–143. https://doi.org/10.1080/07038992.2018.1461557.

Foody, G., Boydb, D.S., Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sens. Environ. 85 (4), 463–474. https://doi.org/10.1016/S0034-4257(03)00039-7.

Gatziolis, D., Andersen, H.E., 2008. A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest. Forestry Sciences. United States Department of Agriculture, Forest Service.

Goodbody, T.R.H., Coops, N.C., White, J.C., 2019. Digital aerial photogrammetry for updating area-based forest inventories: a review of opportunities, challenges, and future directions. Curr. For. Rep. https://doi.org/10.1007/s40725-019-00087-2.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30 (2), 328–341. https://doi.org/10.1109/TPAMI.2007.1166.

Holmgren, J., Nilsson, M., Olsson, H., 2003. Simulating the effects of Lidar scanning angle for estimation of mean tree height and canopy closure. Management 29 (5), 623–632.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., Falkowski, M.J., 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sens. Environ. 112 (5), 2232–2245. Retrieved from. http://linkinghub.elsevier.com/retrieve/pii/S0034425708000138.

Hudak, A.T., Strand, E.K., Vierling, L.A., Byrne, J.C., Eitel, J.U.H., Martinuzzi, S., Falkowski, M.J., 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. Remote Sens. Environ. 123, 25–40. https://doi.org/10.1016/j.rse.2012.02.023.

Hudak, A.T., Bright, B.C., Pokswinski, S.M., Loudermilk, E.L., O'Brien, J.J., Hornsby, B.S., Klauberg, C., Silva, C.A., 2016. Mapping Forest Structure and Composition from Low-Density LiDAR for Informed Forest, Fuel, and Fire Management at Eglin Air Force Base, Florida, USA. Can. J. Remote. Sens. 42 (5), 411–427. https://doi.org/10.1080/07038992.2016.1217482.

Hyyppä, J., Hyyppä, H., Leckies, F., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. Int. J. Remote Sens. 29 (5), 1339–1366.

Isenburg, M., 2017. LAStools. Retrieved from. http://lastools.org.

Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between Lidar pulse density and forest measurement accuracy. Remote Sens. Environ. 130, 245–253. https://doi.org/10.1016/j.rse.2012.11.024.

Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., Guo, Q., 2018. The transferability of random forest in canopy height estimation from multi-source remote sensing data. Remote Sens. 10 (8), 1183. https://doi.org/10.3390/rs10081183.

Joanneum Research Digital, 2018. RSG-Photo. Retrieved from. https://www.remotesensing.at/remote-sensing-software/.

Kangas, A., Gobakken, T., Puliti, S., Hauglin, M., Næsset, E., 2018. Value of airborne laser scanning and digital aerial photogrammetry data in forest decision making. Silva Fenn. 52 (1), 1–19. https://doi.org/10.14214/sf.9923.

Karjalainen, T.M., Korhonen, L., Packalen, P., Maltamo, M., 2018. The transferability of airborne laser scanning based tree level models between different inventory areas. Can. J. For. Res. https://doi.org/10.1139/cjfr-2018-0128.

Keränen, J., Maltamo, M., Packalen, P., 2016. Effect of flying altitude, scanning angle and scanning mode on the accuracy of ALS based forest inventory. Int. J. Appl. Earth Obs. Geoinf. 52, 349–360. https://doi.org/10.1016/j.jag.2016.07.005.

Kotivuori, E., Korhonen, L., Packalen, P., 2016. Nationwide airborne laser scanning based models for volume, biomass and dominant height in Finland. Silva Fenn. 50 (4), 1–28. https://doi.org/10.14214/sf.1567.

Leckie, D.G., Gillis, M.D., 1995. Forest inventory in Canada with emphasis on map production. For. Chron. 71 (1), 74–88. https://doi.org/10.5558/tfc71074-1.

Magnussen, S., Næsset, E., Gobakken, T., Frazer, G., 2012. A fine-scale model for area-based predictions of tree-size-related attributes derived from LiDAR canopy heights. Scand. J. For. Res. 27 (3), 312–322. https://doi.org/10.1080/02827581.2011.624116.

Meidinger, D.V., Pojar, J., 1991. Ecosystems of British Columbia. B.C. (Victoria, BC).

Ministry of Forest Lands and Natural Resource Operations, 2014. Vegetation Resources

Inventory, Photo Interpretation Procedures. Version 3.0.
Montaghi, A., 2013. Effect of scanning angle on vegetation metrics derived from a nationwide airborne laser scanning acquisition. Can. J. Remote. Sens. 39 (Suppl. 1), S152–S173. https://doi.org/10.5589/m13-052.
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), 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.
Næsset, E., 2014. Area-based inventory in Norway – from innovation to an operational reality. In: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), For Appl Airborne Laser Scanning Concepts Case Stud. Springer, Dordrecht, pp. 215–240. https://doi.org/10.1007/ 978-94-017-8663-8_11.
Næsset, E., Gobakken, T., 2008. Estimation of above-and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sens. Environ. 112, 3079 – 3090. Retrieved from. http://linkinghub.elsevier.com/retrieve/pii/S0034425708000849.
Næsset, E., Gobakken, T., Holmgren, J., Hyyppa, H., Hyyppa, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, A., Soderman, U., 2004. Laser scanning of forest resources: the Nordic experience. Scand. J. For. Res. 19 (6), 482–499. https://doi.org/10.1080/02827580410019553.
Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., Olsson, H., 2016. 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.
Packalen, P., Strunk, J., Packalen, T., Maltamo, M., Mehtätalo, L., 2019. Resolution dependence in an area-based approach to forest inventory with airborne laser scanning. Remote Sens. Environ. 224 (December 2018), 192–201. https://doi.org/10.1016/j.rse.2019.01.022.
Penner, M., Pitt, D.G., Woods, M.E., 2013. Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. Can. J. Remote. Sens. 39 (5), 426–443. https://doi.org/10.5589/m13-049.
Puliti, S., Gobakken, T., Ørka, H.O., Næsset, E., Puliti, S., Gobakken, T., Ørka, H.O., Næsset, E., Puliti, S., Gobakken, T., Ørka, H.O., Næsset, E., 2016. Assessing 3D Point Clouds From Aerial Photographs for Species-specific Forest Inventories Inventories. 7581https://doi.org/10.1080/02827581.2016.1186727. (July).
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria Retrieved from. http://www.r-project.org/.
Roussel, J.-R., Auty, D., 2018. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. Retrieved from. https://github.com/Jean-Romain/lidR.
Roussel, J.R., Caspersen, J., Béland, M., Thomas, S., Achim, A., 2017. Removing bias from LiDAR-based estimates of canopy height: accounting for the effects of pulse density and footprint size. Remote Sens. Environ. 198, 1–16. https://doi.org/10.1016/j.rse.2017.05.032.
Roussel, J.R., Béland, M., Caspersen, J., Achim, A., 2018. A mathematical framework to describe the effect of beam incidence angle on metrics derived from airborne LiDAR: the case of forest canopies approaching turbid medium behaviour. Remote Sens. Environ. 209 (December 2017), 824–834. https://doi.org/10.1016/j.rse.2017.12.006.
Sprugel, D., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64, 209–210. https://doi.org/10.2307/1937343.
Stepper, C., Straub, C., Immitzer, M., Pretzsch, H., 2017. Using canopy heights from digital aerial photogrammetry to enable spatial transfer of forest attribute models: a case study in central Europe. Scand. J. For. Res. 32 (8), 748–761. https://doi.org/10.1080/02827581.2016.1261935.
Strunk, J., Temesgen, H., Andersen, H.E., Flewelling, J.P., 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 (5), 644–654. https://doi.org/10.5589/m12-052.
Sumnall, M., Peduzzi, A., Fox, T.R., Wynne, R.H., Thomas, V.A., Cook, B., 2016. Assessing the transferability of statistical predictive models for leaf area index between two airborne discrete return LiDAR sensor designs within multiple intensely managed Loblolly pine forest locations in the south-eastern USA. Remote Sens. Environ. 176, 308–319. https://doi.org/10.1016/j.rse.2016.02.012.
Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., 2015. Augmenting site index estimation with airborne laser scanning data. For. Sci. 61 (5(28)), 861–873. https://doi.org/10.5849/forsci.14-175.
Tompalski, P., Coops, N., White, J., Wulder, M., 2016. Enhancing forest growth and yield predictions with airborne laser scanning data: increasing spatial detail and optimizing yield curve selection through template matching. Forests 7 (12), 255. https://doi.org/10.3390/f7110255.
Vexcel Imaging, 2018. UltraCam-X Technical Specifications. Retrieved from. https://www.sfsaviation.ch/files/177/SFSUCX.pdf.
White, J.C., Wulder, M.A., Varhola, A., Vastaranta, M., Coops, N.C., Cook, B.D., Pitt, D., Woods, M., 2013. A Best Practices Guide for Generating Forest Inventory Attributes From Airborne Laser Scanning Data Using an Area-based Approach.
White, J., Stepper, C., Tompalski, P., Coops, N., Wulder, M., 2015. Comparing ALS and image-based point cloud metrics and modelled forest inventory attributes in a complex coastal forest environment. Forests 6 (10), 3704–3732. https://doi.org/10.3390/f6103704.
White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote sensing technologies for enhancing forest inventories: a review. Can. J. Remote. Sens. 42 (5), 619–641. https://doi.org/10.1080/07038992.2016.1207484.
Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., Treitz, P., 2011. Operational implementation of a LiDAR inventory in Boreal Ontario. For. Chron. 87 (4), 512–528. https://doi.org/10.5558/tfc2011-050.
Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., et al., 2012. Lidar sampling for large-area forest characterization: a review. Remote Sens. Environ. Elsevier B.V. 121, 196–209. https://doi.org/10.1016/j.rse.2012.02.001.
Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., White, J.C., 2008. The role of LiDAR in sustainable forest management. For. Chron. 84, 807–826. https://doi.org/10.5558/tfc84807-6.
Wulder, M.A., Coops, N.C., Hudak, A.T., Morsdorf, F., Nelson, R., Newnham, G., Vastaranta, M., 2013. Status and prospects for LiDAR remote sensing of forested ecosystems. Can. J. Remote. Sens. 39 (sup1), S1–S5. https://doi.org/10.5589/m13-051.
Yoga, S., Bégin, J., St-Onge, B., Riopel, M., 2017. Modeling the effect of the spatial pattern of airborne lidar returns on the prediction and the uncertainty of timber merchantable volume. Remote Sens. 9 (8), 1–13. https://doi.org/10.3390/rs9080808.