

**Enhancing the characterization of forest structure and resources for forest
inventory and monitoring using airborne and spaceborne single photon lidar**

by

Martin Queinnec

MSc, École Polytechnique Fédérale de Lausanne, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2022

© Martin Queinnec, 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Enhancing the characterization of forest structure and resources for forest inventory and monitoring using airborne and spaceborne single photon lidar

submitted by Martin Queinnec in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

in Forestry

Examining Committee:

Nicholas C. Coops, Professor, Forest Resources Management, The University of British Columbia

Supervisor

Joanne C. White, Research Scientist, Canadian Forest Service

Supervisory Committee Member

Verena C. Griess, Professor, Environmental Systems Science, Swiss Federal Institute of Technology in Zurich

Supervisory Committee Member

Naomi B. Schwartz, Assistant Professor, Geography, The University of British Columbia

Supervisory Committee Member

Sean M. Smukler, Associate Professor, Land and Food Systems, The University of British Columbia

University Examiner

Richard C. Hamelin, Professor, Forest Conservation Sciences, The University of British Columbia

University Examiner

Valerie A. Thomas, Professor, Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University

External Examiner

Abstract

Canada's forests cover about 347 million ha, of which 157 million ha are actively managed for timber procurement and 114 million ha are located in northern remote areas and mostly unmanaged. Information about forest resources is needed to inform sustainable forest management practices and monitor the evolution of forested ecosystems, particularly in the context of climate change. In the past decades, airborne acquisitions of light detection and ranging (lidar) data have been increasingly used to map forest structure. However, their acquisitions remain expensive and coverage is often limited to targeted managed forest units. Recent technological advances led to the development of single photon lidar (SPL) systems that enable high-altitude airborne and spaceborne lidar acquisitions with an unprecedented amount of detail and across large geographic extents. Two SPL systems are currently available: the airborne SPL100 instrument, commercially released in 2017, and the spaceborne Advanced Topographic Laser Altimeter System (ATLAS) onboard the Ice Cloud and Elevation Satellite-2 (ICESat-2) launched in September 2018. However, the ability of these two systems to accurately characterize forest structure and resources is not well defined. The primary objective of this dissertation is to examine the role that SPL100 and ICESat-2 can play to provide spatially-explicit, timely and accurate estimates of some key information needs in managed and unmanaged boreal forests of Canada. To that end, an enhanced inventory of key forest attributes driven by SPL100 data was implemented across an actively managed boreal forest located in Ontario, Canada. Strengths and limitations of ICESat-2 metrics to accurately characterize canopy height, cover and vertical vegetation structure were also highlighted by comparing them to SPL100 counterparts. Moving on to a broader monitoring perspective, a method to extrapolate ICESat-2 samples of canopy height estimates across over 19 million ha of unmanaged boreal forests was implemented. Time-series of this wall-to-wall product were also produced and used to examine the recovery of vegetation height after disturbance. The research presented in this dissertation helps better understanding the ability of SPL100 and ICESat-2 to characterize vegetation structure in both managed and unmanaged forests in order to inform policy-making and sustainable forest management practices.

Lay Summary

Spatially-explicit, accurate and timely information about forest resources is needed to inform policy-making and sustainable forest management practices. Remote sensing technologies such as light detection and ranging (lidar) are particularly well suited to characterize the three-dimensional structure of forests. However, the broad geographic extent of forests in Canada often limits the acquisition of lidar data to targeted forests that are actively managed. Recent technological advances have led to the development of single photon lidar (SPL) systems that can be operated across large extents from both airborne and spaceborne platforms. This dissertation examines the potential and limitations of an airborne and a spaceborne SPL to characterize forest structure and resources in both managed and unmanaged forests of Canada. Methods to assess and analyze the data acquired by these two systems, combine them with other data sources and demonstrate potential applications by producing novel products are presented.

Preface

The research objectives of this dissertation were developed through discussion with my supervisory committee. The work presented in Chapters 3, 4, 5 and 6 of this dissertation have been published in four peer-reviewed journals, with a second paper from Chapter 6 in preparation as listed below. In these publications, I implemented the methods discussed with co-authors and supervisory committee, performed formal analysis, prepared original manuscript drafts and implemented revisions. Co-authors provided critical feedback, comments and revisions to prepare the manuscript for submission and throughout the peer-review process.

Chapter 3

Queinnec, M., Coops, N. C., White, J. C., McCartney, G., & Sinclair, I. (2021). Developing a forest inventory approach using airborne single photon lidar data: from ground plot selection to forest attribute prediction. *Forestry: An International Journal of Forest Research*. <https://doi.org/10.1093/forestry/cpab051>

Chapter 4

Queinnec, M., Coops, N.C., White, J.C., Griess, V.C., Schwartz, N.B., McCartney, G (2022). In Press. Mapping dominant tree species groups by combining area-based and individual tree crown lidar metrics with Sentinel-2 data. *Canadian Journal of Remote Sensing*.

Chapter 5

Queinnec, M., White, J. C., & Coops, N. C. (2021). Comparing airborne and spaceborne photon-counting LiDAR canopy structural estimates across different boreal forest types. *Remote Sensing of Environment*, 262, 112510. <https://doi.org/10.1016/J.RSE.2021.112510>

Chapter 6

Queinnec, M., Coops, N.C. et. al. (2022) Mapping changes in canopy height over time using ICESat-2 and Landsat time-series products (Manuscript in preparation)

Queinnec, M., Tompalski, P., Bolton, D. K., & Coops, N. C. (2021). FOSTER—An R package for forest structure extrapolation. *PLOS ONE*, 16(1), e0244846. <https://doi.org/10.1371/JOURNAL.PONE.0244846>

In addition, I contributed to the following publications that are relevant to the scope of this dissertation:

Irwin, L., Coops, N. C., Queinnec, M., McCartney, G., & White, J. C. (2021). Single photon lidar signal attenuation under boreal forest conditions. *Remote Sensing Letters*, 12(10), 1049–1060.

<https://doi.org/10.1080/2150704X.2021.1962575>

Bolton, D.K., Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., Hermosilla, T., Queinnec, M., et al. 2020. “Optimizing Landsat Time Series Length for Regional Mapping of Lidar-Derived Forest Structure.” *Remote Sensing of Environment* 239: 111645.

<https://doi.org/10.1016/j.rse.2020.111645>.

Coops, N.C., Tompalski, P., Goodbody, T.R.H., Queinnec, M., Luther, J.E., Bolton, D.K., White, J.C., Wulder, M.A., Lier, O.R. van, and Hermosilla, T. 2021. “Modelling Lidar-Derived Estimates of Forest Attributes over Space and Time: A Review of Approaches and Future Trends.” *Remote Sensing of Environment* 260 (July): 112477. <https://doi.org/10.1016/J.RSE.2021.112477>.

Table of Contents

Abstract.....	iii
Lay Summary.....	iv
Preface.....	v
Table of Contents	vii
List of Tables	ix
List of Figures.....	xi
List of Abbreviations	xvi
Acknowledgments	xvii
Dedication.....	xviii
Chapter 1. Introduction and background.....	1
1.1. Information needs in managed and unmanaged forests of Canada	1
1.2. Remote sensing of forested ecosystems	3
1.3. Potential values and limitations of airborne and spaceborne single-photon lidar for forest management and monitoring	8
1.4. Research questions	9
1.5. Dissertation overview	10
Chapter 2. Study area and data.....	12
2.1. Study areas.....	12
2.2. Remotely sensed data.....	17
2.3. Field data.....	26
Chapter 3. Developing a forest inventory approach using airborne single photon lidar data: from ground plot selection to forest attributes selection.....	33
3.1. Introduction	33
3.2. Methods.....	36
3.3. Results.....	46
3.4. Discussion.....	53
Chapter 4. Mapping dominant tree species groups by combining area-based and individual tree crown lidar metrics with Sentinel-2 data.....	61
4.1. Introduction	61
4.2. Methods.....	63
4.3. Results.....	71

4.4. Discussion.....	80
Chapter 5. Comparing airborne and spaceborne photon-counting lidar canopy structural estimates across different boreal forest types.....	88
5.1. Introduction	88
5.2. Methods.....	90
5.3. Results.....	99
5.4. Discussion.....	108
Chapter 6. Mapping canopy height over time using ICESat-2 and Landsat time-series products .	115
6.1. Introduction	115
6.2. Methods.....	117
6.3. Results.....	125
6.4. Discussion.....	132
Chapter 7. Conclusions.....	136
7.1. Dissertation objectives.....	136
7.2. Research innovations and significance	139
7.3. Research limitations and future directions.....	141
7.4. Closing statement	143
Bibliography	145
Appendices.....	162

List of Tables

Table 2-1. Technical specifications of the SPL100 acquisition at the Romeo Malette Forest.....	18
Table 2-2. Non-exhaustive list of attributes recorded at the plot-level and individual tree level by field crews.	29
Table 2-3. Median, 25 th percentile and 75 th percentile of the distribution of measured tree DBH and height per species.....	32
Table 3-1. Description of the modeled forest attributes.....	40
Table 3-2. Description of species groups used to classify the field plots.	42
Table 3-3. Number of re-measured and new sample plots grouped by stand development stage and species group. Only conditions where at least 5 re-measured or new plots were observed are reported in the table. See Table 3-2 for the description of species groups.....	47
Table 3-4. Absolute and relative RMSE and bias for predictions of each forest inventory attribute of interest, stratified by forest type. See Table 3-2 for the description of species group.....	49
Table 4-1. Stratification of ground plots into 3 or 5 functional groups based on the species of the trees positioned in - or above - the dominant canopy layer. The number of trees in each plot was used to derive proportions.	64
Table 4-2. Description of the predictor variables derived from the ALS data and Sentinel-2 imagery.....	66
Table 4-3. Description of the predictor variables included in the different scenario runs. Refer to Table 4-1 for a description of the metrics	67
Table 4-4. Confusion matrix showing the results of the broader, 3-group classification model using all predictor sets. Predictions of the repeated folds are included in the matrix. SW = Softwood; MW = Mixedwood; HW = Hardwood.	73
Table 4-5. Confusion matrix for the five species groups classification model using all predictor sets. Predictions of the repeated folds are included in the matrix. JP = Jack Pine Dominated; BS = Black Spruce Dominated; MC = Mixed Conifer; MW = Mixedwood; HW = Harwood.....	74
Table 5-1. Characteristics of the analysis units (AUs). The standard deviation (SD) is indicated in parentheses.....	95
Table 6-1. Description of the predictor variables included in the k-NN imputation models. Spectral indices, change metrics and land cover metrics were derived for each year between 1984 and 2021.....	119
Table 6-2. Filters applied to the gridded <i>h_canopy</i> values at the RMF and NW Boreal study area in order to form a pool of reference observations acquired in optimal conditions.....	123
Table 6-3. Comparison of ICESat-2 and SPL100 estimates of canopy height (<i>h_canopy</i> and <i>p95</i> , respectively) in reference observations grouped by forest type (VLCE at the year of ICESat-2 data	

acquisition). pRMSD and pMD are calculated based on the mean values of SPL100 p95. Observations for which change was detected after 2018 were removed to account for the time difference between the SPL data acquisition (2018) and the ICESat-2 data acquisition (2019 – 2021).	127
Table 6-4. Summary of the evolution of canopy height after fire disturbance expressed as a percentage of pre-fire canopy height. The median values calculated across all sampled pixels having been disturbed between 1984 and 2021 are reported.	131

List of Figures

Figure 1-1. Classification of Canada’s forest management status according to Stinson et al. (2019). Pixels with crown closure < 10 % and tree height < 5 m, according to 2011 Canada-wide forest attribute maps derived by (Beaudoin et al., 2018), are masked out.	2
Figure 1-2. Overview of the core chapters of this dissertation.	11
Figure 2-1. Ecozones totally or partially covered by forests (A) and overview of the Romeo Malette Forest and NW Boreal study areas in Ontario (B).	13
Figure 2-2. Map of the Romeo Malette Forest study area overlaid with the network of main roads, waterbodies and watercourses and ground plots established as part of the enhanced forest inventory carried out in Chapter 3. The footprint of a large wildfire that occurred in the summer of 2012 is also displayed.	14
Figure 2-3. Monthly average of daily mean temperature and cumulated precipitation recorded from 2008 to 2022 at the weather station of Timmins, Ontario (Station ID 47547). Data obtained from Environment and Climate Change Canada through the R package weathercan.	15
Figure 2-4. Land cover in 2018 at the NW Boreal study area according to the Virtual Land Cover Engine classification (Hermosilla et al., 2018)	16
Figure 2-5. Overview of SPL100 acquisitions schedule across the forest management units of Ontario as of summer 2021.	17
Figure 2-6. Conical scanning footprint of a 10-s acquisition of SPL100 data from an aircraft flown at an altitude of 3,800 m above ground and cruising speed of 350 km/h in the West-East direction.	19
Figure 2-7. Visualization of a 1 x 1 km tile of the SPL100 point cloud acquired at the Romeo Malette Forest. Points are colored by elevation above sea level from blue (low) to red (high).	20
Figure 2-8. Configuration of the 3 pairs of ATLAS strong and weak beams for an acquisition over the Romeo Malette Forest.	21
Figure 2-9. Overview of the ATL03 and ATL08 data products. The ATL03 product contains the coordinates of each downlinked photon. The ATL08 product contains the classification of each photon (ATL08a) and canopy and terrain height summary metrics along 20-m and 100-m segments (ATL08b). Note that not all photons of the ATL03 product are classified by the ATL08 classification algorithm, as indicated by the NA values in the field classed_pc_flag of ATL08a.	23
Figure 2-10. SPL100 (A) and ICESat-2 (B) classified point clouds acquired along an ICESat-2 100-m segment (strong beam). The elevation of the ICESat-2 photons was transformed to the Canadian Geodetic Vertical Datum of 2013 (CGVD2013) to match SPL100 photons elevation.	24

Figure 2-11. Captures of 360° photographs taken from plot center in stands with different maturity code, species composition and crown closure.	28
Figure 2-12. Distribution of species of the 13,389 tallied trees grouped by the position of their crowns relative to their surrounding trees.	30
Figure 2-13. Distribution of DBH (≥ 7.1 cm) and tree height per species. Black dots indicate the mean of the distribution, wide and thin lines indicate where 66% and 95% of the values around the median are located, respectively. Two trees (one trembling aspen and one in the “other” category) with DBH of 81.2 cm and 89.5 cm are not included in the DBH graph.	31
Figure 3-1 Variance explained by the first 5 principal components obtained from the PCA of the 20 lidar metrics.	38
Figure 3-2. Distribution of field-measured forest attributes per species group. Red dots indicate the mean of the distribution. See Table 3-2 for the description of species groups.	43
Figure 3-3. Distribution of the number of plots per strata for the 15 subsets of the 250 plots used to examine the influence of the number of plots on the accuracy of the models. The graph at the top represents the smallest sample, with 1 plot per strata while the graph at the bottom represents the full sample of 250 plots with a maximum of 15 plots per strata. Bars are colored in green if an additional plot was added to a stratum compared to the previous smaller plot subset.	45
Figure 3-4. Feature space formed by the first two principal components PC1 and PC2. The horizontal and vertical dashed lines indicate the 41 strata that were used to stratify the feature space into equal intervals. Convex hulls are indicated by the solid colored lines.	46
Figure 3-5. Wall-to-wall and inset maps of predicted Lorey’s height (LOR), merchantable volume (MVOL) and stem density (DENS) at the Romeo Malette Forest. Inset maps have the extent of the box labelled as A is the wall-to-wall map.	48
Figure 3-6. Scatter plots and accuracy metrics comparing the forest attributes values predicted by the RF models and to their observed values.	50
Figure 3-7. Summary of the SPL100 metrics importance rank combined from all the models. Metrics that were not at least once the 5th most important variable in any of the model are not reported. pxx = xx percentiles of returns height; cov_xx = percentage of first returns above xx m; cv = coefficient of variation (standard deviation / average of returns height).	51
Figure 3-8. Pearson correlation values between the forest attributes values predicted by the RF models and the SPL100 metrics used as predictor variables.	52
Figure 3-9. Influence of the number of plots on the accuracy of the random forest models of forest attributes. Each point represents an addition of 1 plot per stratum used in the structurally guided sampling as	

long as additional plots are available in each stratum of the full sample of 250 plots. The distribution of the number of plots per strata is provided in Figure 3-3.....	53
Figure 3-10. Relative RMSE (RMSE%) by species group for the area-based models of forest attributes generated in this study compared to the ones reported in White et al. (2021b) and Woods et al. (2011). Lorey's height (LOR) modelled in this study and White et al. (2021b) is compared to average height (AVGHT) in Woods et al. (2011). Merchantable stem volume (MVOL) and stem density (DENS) in this study were not calculated with the same method as the one used in Woods et al. (2011) and White et al. (2021b) and RMSE% for these attributes are therefore not reported in this figure.	54
Figure 4-1. Influence of the local maxima window diameter on the F-score of the individual tree detection algorithm by species group. The window diameter for which the highest median F-score was observed is highlighted in green.	72
Figure 4-2. Comparison of the overall accuracy of the random forest classification model of the ground plots into softwood, mixedwood, and hardwood species groups (3 Species Groups) or jack pine dominated, black spruce dominated, mixed conifer, mixedwood, and hardwood species groups (5 Species Groups).	73
Figure 4-3. Chord diagram visualizing the confusion matrix of the random forest species classification model into (A) softwood, mixedwood, and hardwood species group; or (B) jack pine dominated, black spruce dominated, mixed conifer, mixedwood, and hardwood species groups for the scenario run that included all predictor variables. Predictions from all cross-validation folds were aggregated and are included in this diagram. The predicted species groups are shown in the bottom half and the reference species groups are shown in the upper half. The predicted species groups are shown in the bottom half and the reference species groups are shown in the upper half. Links illustrate the number of observations that were correctly classified (link between same predicted and reference species groups) and the number of classification errors (link originating from a predicted species group going to a different reference species group).....	75
Figure 4-4. Conditional random forest variable importance of the 3-group (A) and 5-group (B) species groups classification model including all the predictor variables. Raw importance values were scaled between 0 (least important variable) and 100 (most important variable).	76
Figure 4-5. Pairwise comparison of significant differences in predictor variable values across species groups. The blue and red color gradients indicate if the pairwise difference between the median values is negative or positive, respectively. JP = Jack Pine Dominated; BS = Black Spruce Dominated; MC = Mixed Conifers; MW = Mixedwood; HW = Hardwood.	77

Figure 4-6. Chord diagram showing the distribution of the most frequently predicted species groups in the stands of the photo-interpreted inventory. The most frequently predicted species groups in each stand are shown in the bottom half and the stand-level species groups derived from photo-interpreted species composition are shown in the upper half.	78
Figure 4-7. Wall-to-wall predictions of species groups overlaid with forest stands labeled by their inventory (photo-interpreted) species groups. Predictions were only made in pixels where $zp90 > 5$ m, $pzabove2 > 50\%$ and in productive forest stands. JP = Jack Pine Dominated; BS = Black Spruce Dominated; MC = Mixed Conifers; MW = Mixedwood; HW = Hardwood.	79
Figure 4-8. Comparison of species groups predictions and photo-interpreted stands with aerial photographs acquired at the same time as the SPL100 data. The aerial photograph is displayed as a false color composite (Red channel = NIR; Green channel = Red; Blue channel = Green).....	80
Figure 5-1. Flow diagram illustrating the main steps of the methodology applied in this study.	92
Figure 5-2. Illustration of the analysis units (A) and their analysis at the analysis unit level (B) and photon level (C). Small dots in B and C represent the normalized SPL100 point cloud. Large dots in B represent the normalized ATL03 photons located along the ICESat-2 ground track.	96
Figure 5-3. Transect of coincident normalized SPL100 and ICESat-2 point clouds. The canopy profile of SPL100 and ICESat-2 are displayed with the solid black and blue lines respectively. This figure illustrates the retrieved ICESat-2 canopy profile from the highest photons within a 3 m window in the along-profile direction. VF = Vegetation Fill; SPL = SPL100.	99
Figure 5-4. Violin plots and boxplots show the distribution of plot level field data by forest group, determined from k-means clustering. Groups 1, 2 and 3 are composed of 133, 83 and 33 plots respectively. All the attributes, except the rumple index and the stand maturity class, were used for the k-means clustering.	101
Figure 5-5. Normalized SPL100 profiles along analysis units illustrating the typical structure of the three forest groups. Each profile has a width of 11 m in the across-track direction.	102
Figure 5-6. (A) Variation of Pearson correlation coefficient and RMSD per forest group and SPL height metric. (B) Scatterplots showing the relationship between ATL03 top of canopy photons height and SPL100 90 th height percentile for all groups combined and per group. Plots are colored by point density on a blue (low density) to yellow (high density) color scale. The Pearson correlation coefficient is labelled as r . SPL = SPL100.	103
Figure 5-7. Scatterplots showing the relationship between ICESat-2 and SPL100 measure of canopy cover (A) as the percentage of returns above 2 m (CC_{2m}) and (B) as a measure of vegetation fill at maximum return height (VF_{max}). Plots are colored by point density on a blue (low density) to yellow (high density) color scale. The Pearson correlation coefficient is labelled as r . SPL = SPL100.	105

Figure 5-8. Scatterplots showing the relationship between SPL100 and ATL03 25 th , 50 th , 75 th and 95 th photons height percentiles at the AU level. SPL = SPL100.	106
Figure 5-9. Scatterplots showing the relationship between SPL100 and ATL03 standard deviation (SD) and coefficient of variation (CV) at the AU level. SPL = SPL100.....	107
Figure 6-1. Distribution of gridded h_canopy values of the reference observations at the Romeo Malette Forest (9,979 observations) and at the NW Boreal study area (10,000 observations).....	126
Figure 6-2. Scatter plot (A) and distribution (B) of p95 (SPL) and h_canopy (ICESat-2) estimates of canopy height in reference observations where no change was detected after 2018. In A, the scatterplot is visualized using 1-m hexagonal bins colored (shaded from light to dark blue) based on the number of observations contained within them. In B, the point indicates the median of observations and the thick and thin lines indicate intervals that contain 66% and 95% of the distribution around the median, respectively.	127
Figure 6-3. Map of ICESat-2 h_canopy estimates imputed across the study area located in the NW portion of the Boreal Shield Ecozone in Ontario, Canada in 2018.	128
Figure 6-4. Influence of the number of nearest neighbors (k) considered in the k-NN imputation on the Pearson correlation (r), relative RMSD (pRMSD) and relative MD (pMD) for different validation methods. The imputation was performed 10 times for each number of nearest neighbor and validation method. Points and lines represent the median and 75% intervals around the median of the distributions across the 10 repetitions, respectively.....	129
Figure 6-5. Scatterplots illustrating the results of the h_canopy k-NN imputation accuracy assessment at the Romeo Malette Forest and NW Boreal site. Points are aggregated into 1-m width hexagonal bins colored from light to dark blue based on point density.	130
Figure 6-6. Percentage of pre-fire canopy height before (years -3 to 0) and after (years 1 to 30) a fire event grouped by change magnitude (dNBR) values. Each point represents the median value calculated across all the sampled pixels having been disturbed by fire between 1984 and 2021.	131

List of Abbreviations

ABA	Area-Based Approach	NBR	Normalized Burn Ratio
AGB	Above Ground Biomass	NDVI	Normalized Difference Vegetation Index
ALS	Airborne Laser Scanning	NFI	National Forest Inventory
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer	NIR	Near-infrared
ATL03	ATLAS / ICESat-2 L2A Global Geolocated Photon Data	NISAR	NASA-ISRO Synthetic Aperture Radar
ATL08	ATLAS / ICESat-2 L3A Land and Vegetation Height	NN	Nearest Neighbour
ATLAS	Advanced Topographic Laser Altimeter	NOAA	National Oceanic and Atmospheric Administration
AU	Analysis Unit	NTEMS	National Terrestrial Ecosystem Monitoring System
BA	Basal Area	OA	Overall Accuracy
BAP	Best Available Pixel	OOB	Out-Of-Bag
CHM	Canopy Height Model	PCA	Principal Component Analysis
CV	Coefficient of Variation	QMDBH	Quadratic Mean DBH
DAP	Digital Aerial Photogrammetry	R ²	Coefficient of Determination
DBH	Diameter at Breast Height	RF	Random Forest
DEM	Digital Elevation Model	RMF	Romeo Malette Forest
DENS	Stem density	RMSD	Root Mean Square Difference
EFI	Enhanced Forest Inventory	RMSE	Root Mean Square Error
EVI	Enhanced Vegetation Index	SAD	Signal Attenuation Depth
FN	False Negative	SAR	Synthetic Aperture Radar
FP	False Positive	SGS	Structurally-guided sampling
GEDI	Global Ecosystem Dynamics Investigation	SPL	Single Photon Lidar
GLAS	Geoscience Laser Altimeter	SWIR	ShortWave InfraRed
GPS	Global Positioning System	TCA	Tasseled Cap Angle
ICESat	Ice, Cloud and land Elevation Satellite	TCB	Tasseled Cap Brightness
ITC	Individual Tree Crown	TCD	Tasseled Cap Distance
ITD	Individual Tree Detection	TCG	Tasseled Cap Greenness
LOR	Lorey's height	TCW	Tasseled Cap Wetness
MAD	Mean Absolute Bias	TP	True Positive
MD	Mean difference	TSRI	Topographic Solar Radiation Index
MODIS	MODerate Resolution Imaging Spectrometer	USGS	United States Geological Survey
MVOL	Merchantable stem Volume	VF	Vegetation Fill
NASA	National Aeronautics and Space Administration	VLCE	Virtual Land Cover Engine
		WVOL	Whole stem Volume

Acknowledgments

First, I would like to thank my supervisor Dr. Nicholas C. Coops for his continuous guidance throughout the completion of this research. I have been particularly impressed by the amount of attention and support that Nicholas provides to his students and by his determination to foster a social and connected working environment. I am extremely grateful to Nicholas for giving me the opportunity to work and learn at his side and for letting me get involved in teaching. I also extend my gratitude to the members of my committee, Dr. Joanne C. White, Dr. Naomi B. Schwartz and Dr. Verena C. Griess. Their guidance and feedback at various stages of my degree have been instrumental. Special thanks also go to Grant McCartney, Ian Sinclair, Geordie Robere-McGugan, Sumac Geomatics' field crews and the Ministry of Natural Resources and Forestry of Ontario for providing me with data and help without which this research would not have been possible. I am also grateful to the *Forestry Futures Trust Ontario* and the *Natural Sciences and Engineering Research Council of Canada* for funding this research.

Thank you to all the members of the Integrated Remote Sensing Studio for their friendship and support. Working side-by-side, sharing ideas and progress and above all sharing memorable times was essential to keep this research moving along.

Thank you to my parents, Isabelle and Benoît, my siblings Léo and Marie, and my partner Laura for your endless love and support.

Finally, I would like to acknowledge that the work carried out in this dissertation has implications in terms of land and forest management across territories of the Flying Post First Nation, Matachewan First Nation, Mattagami First Nation, Taykwa Tagamou First Nation, Chapleau Cree First Nation, Brunswick House First Nation, Wahgoshig First Nation, and the Métis Nation of Ontario. I would also like to acknowledge that the UBC Vancouver campus where I was based when conducting this research is located on the traditional, ancestral and unceded territory of the xʷməθkʷəy̓əm (Musqueam) people.

Dedication

To my family

Chapter 1. Introduction and background

1.1. Information needs in managed and unmanaged forests of Canada

Covering approximately 4.0 billion hectares (ha), or 31 % of the Earth land area (FAO, 2016), forests and their ecosystems are involved in the global terrestrial carbon cycle, drive hydrological processes and provide freshwater supply, contribute to nutrient cycling and host a large amount of biodiversity (Waring and Running, 2007). Canada's forests represent 9 % of global forest area (FAO, 2016), covering about 347 million ha (Mha) of land (Natural Resources Canada, 2018; Stinson et al., 2019), spanned across 650 Mha (Wulder et al., 2008b) of forested ecosystems (approximately 65 % of Canada's total land area). Most of Canada's forests are publicly owned (326 Mha; 94 %) and about 157 Mha (45.2 %) are actively managed for timber harvesting under land tenures, 30 Mha (8.5 %) are legally protected, and 20 Mha (6 %) are located in Federal or Indian reserves, restricted or treaty/settlement land (Stinson et al. 2019; Figure 1-1). About 114 Mha (33 %) of Canada's forest do not have any commercial forest tenure or other status and can therefore be considered unmanaged (Stinson et al. 2019; Figure 1-1).

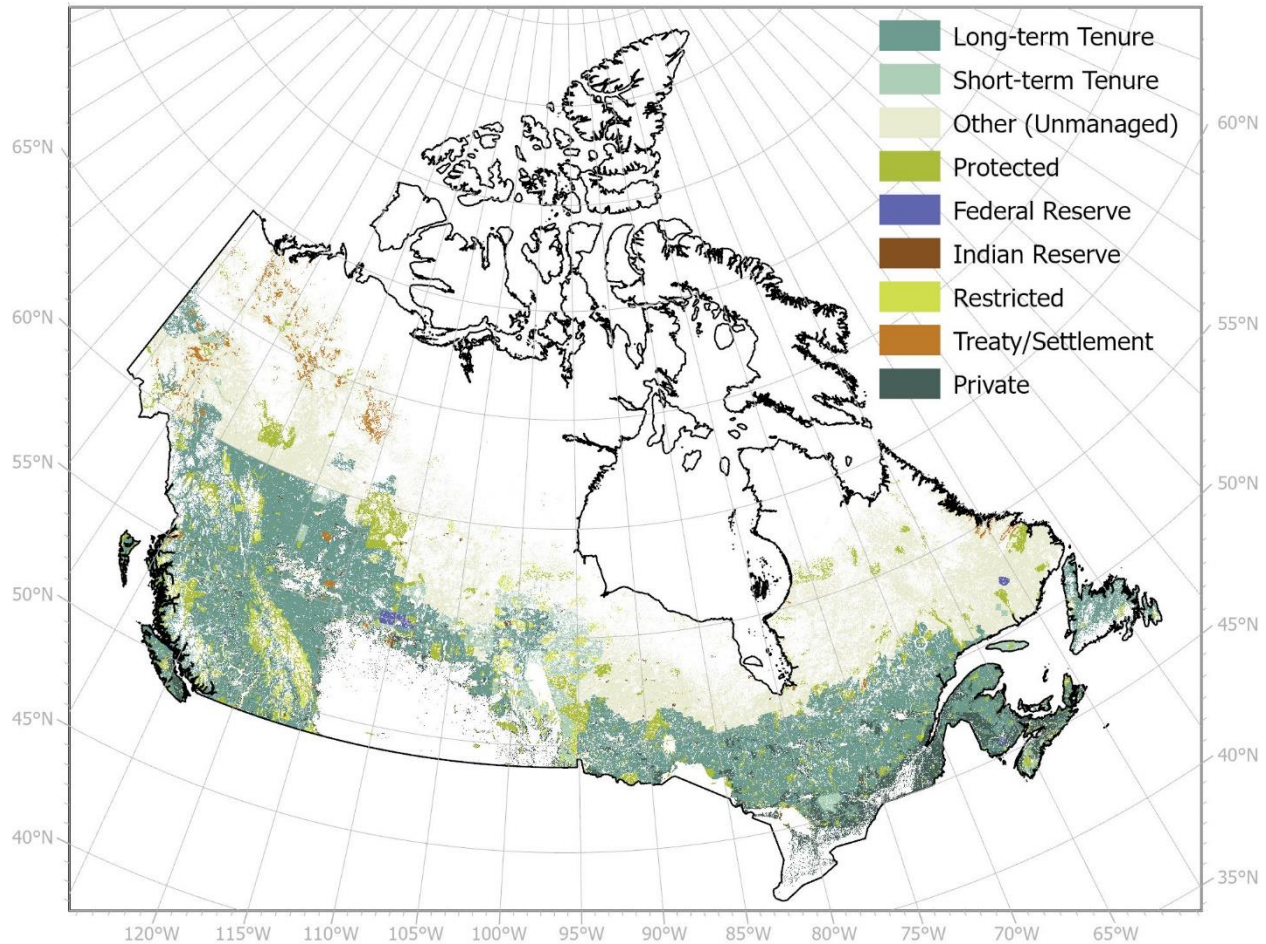


Figure 1-1. Classification of Canada's forest management status according to Stinson et al. (2019). Pixels with crown closure < 10 % and tree height < 5 m, according to 2011 Canada-wide forest attribute maps derived by (Beaudoin et al., 2018), are masked out.

Increase in CO₂ concentration, rising temperatures and increase in the magnitude and rate of disturbances such a wildfire, drought, insects and disease outbreaks, and land-use changes are shifting forested ecosystems dynamics globally (McDowell et al., 2020; Seidl et al., 2017). In areas where forests are actively managed, silvicultural practices need to account for these changing conditions in order to ensure a sustainable management of forest resources (Achim et al., 2022). Forest management practitioners have a particularly crucial role to play in ensuring sustainable management of forest resources in order to mitigate the impacts of climate change by enhancing carbon storage and sequestration and ensuring the adaptation of forested ecosystems to future climatic conditions (D'Amato et al., 2011). Spatially-explicit, timely and accurate inventory of forest resources are therefore needed to ultimately drive policy making as well as

management and operations planning. Typical information needs for forest management include stand-level¹ characterization of attributes such as canopy height, wood volume and biomass, species composition and age, to name a few. Finer-scale information, up to the individual tree level, can also be required to guide silvicultural operations such as thinning and selective logging for example. Beside being used for tactical planning, forest inventories also guide the development of forest management plans that must be approved by provincial and territorial governments before any harvesting can occur.

Beyond the precise inventory of managed forest resources, it is also important to account for forest resources at a broader scale that includes unmanaged forested areas. The boreal region, in which the majority of Canada's unmanaged forests are located, is expected to be particularly vulnerable to the effects of climate change resulting in an increase in the frequency and severity of biotic and abiotic disturbances (Gauthier et al., 2015). In this context, information needs in these regions are related to the monitoring of forest cover, biodiversity, forest productivity and carbon storage dynamics (DeFries, 2013). However, remote northern unmanaged boreal forests are difficult to access and therefore not as well characterized as managed forests because of the poorer accuracy and lack of data available (Andrew et al., 2012). For example, unmanaged forests are currently not taken into account in Canada's carbon budget reporting because of the large uncertainties associated with current carbon fluxes estimates (Kurz et al., 2013).

1.2. Remote sensing of forested ecosystems

Regardless of forest management level and information needs, spatially-explicit information about forest resources is required. Forest inventory for resources management, such as national forest inventories (NFIs) were traditionally compiled from ground sample plots data and manual interpretation of aerial imagery to delineate stands with homogeneous forest attributes (White et al., 2016). In the last decades, the increasing availability of airborne and spaceborne remote sensing data has provided the means to enhance the

¹ A stand can be defined as a community of trees possessing sufficient uniformity in composition, age, arrangement, or condition to be distinguishable from the forest or other growth on adjoining areas, thus forming a silvicultural or management entity (NRCan Forestry Glossary). The size of a stand typically ranges from one hectare to hundreds of hectares.

information available for forest resources assessment in both managed and unmanaged forests (Banskota et al., 2014; White et al., 2016; Wulder et al., 2008a). Below, an overview of two particular remote sensing data sources is provided: light detection and ranging (lidar) and optical imagery.

1.2.1. Light detection and ranging

Lidar is an active remote sensing technology that measures ranges (i.e. distances) based on the time required for a laser to travel towards a target, be reflected, and travel back to its source (Lim et al., 2003; Wehr and Lohr, 1999). A full lidar system consists of a laser transmitter unit that generate a laser pulse, a receiver unit that detects reflected laser pulses and records the time elapsed since their transmission with high precision, a Global Positioning System (GPS) receiver that records the position of the system and an inertial measurement unit that measures its orientation.

A lidar instrument can be operated from a fixed or mobile position on the ground (terrestrial laser scanning; mobile laser scanning) or mounted on airborne platforms such as airplanes, helicopters or remotely piloted aircraft systems in which case it is commonly referred to as airborne laser scanning, or ALS. Lidar instruments can also be operated from spaceborne platforms although the number of spaceborne lidar missions designed for Earth surface mapping has been limited. The Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat; Schutz et al., 2005), the Advanced Topographic Laser Altimeter (ATLAS) onboard ICESat-2 (Neumann et al., 2019) and the Global Ecosystem Dynamics Investigation (GEDI; Dubayah et al. 2020) mounted on the international space station are the three spaceborne lidar instruments that have been (GLAS) or are currently (ATLAS, GEDI) collecting data from space. Unlike ALS, current spaceborne lidar instruments collect samples of data along orbital tracks instead of scanning and providing wall-to-wall measurements.

The laser transmitter is based on a diode-pumped solid-state laser that generally emits a laser in a specific narrow band of the electromagnetic spectrum. A laser pulse emitted by the transmitter unit results in a circular horizontal sampling area (i.e., footprint) that expands with distance due to the laser beam divergence. Small-footprint lidar systems, with a typical diameter of a few decimeters on the ground, are

therefore more suitable to provide a detailed characterization of the surface three-dimensional structure than large-footprint lidar with a diameter of several meters (Lim et al., 2003). Depending on the footprint size and the vertical structure of the target, a single laser pulse can be reflected by multiple surfaces and result in multiple detection events. Full-waveform lidar systems are able to provide a continuous measure of the backscattered signal (i.e. amplitude over time) while discrete return lidar systems can generally return up to five returns per laser pulse (Lim et al., 2003). Full-waveform lidar systems generally have a larger footprint than discrete return lidar systems, with a few exceptions, and are generally used for research purposes, with limited applications in the operational forestry context (Hyypä et al., 2008; White et al., 2016; Wulder et al., 2008a).

Reflected laser pulses can be detected by the detectors of the lidar unit using two approaches: analog detection or single-photon detection (Harding, 2018). To date, the majority of lidar units that have been used in forestry applications carry analog detectors and are often referred to as linear-mode lidar. Analog detectors are usually composed of silicon avalanche photodiodes which have optimum detection sensitivity for lasers emitting near-infrared wavelengths of 900 nm (semiconductor lasers) or 1064 nm (Nd:YAG lasers) (Harding, 2018). Silicon avalanche photodiodes detectors produce a voltage proportional to the amount of energy (i.e., the number of photons) contained in the reflected pulse. For discrete-return system, thresholds are applied on the output voltage to determine when a return occurs. This means that high energy pulses need to be transmitted in order to have enough reflected photons (typically thousands) to exceed minimum detection thresholds (Harding, 2018). Single photon detectors on the other hand do not convert the reflected pulse to an output voltage but record the arrival of individual photons within a reflected pulse. This requires the use of highly sensitive photomultiplier tubes with short recovery time (minimum required time after a photon detection for a photomultiplier tube to return to a state where another photon can be detected). Currently available photomultiplier tubes have an optimum detection efficiency in the visible part of the spectrum and require the use of Nd:YAG lasers ($\lambda = 1064$ nm) with doubled frequency that produce a visible green light at a wavelength of 532 nm (Harding, 2018). The particularly high sensitivity

of single-photon detectors enables the capacity to transmit laser pulses with a lower energy than typically required for lidar units with analog detectors. In practice, it means that airborne single photon lidar (SPL) units can be operated from higher altitudes and cruise speeds, covering more area, and therefore decreasing operational acquisition costs (Degnan, 2016). The design of airborne and spaceborne single photon lidar units started to be discussed in the late 90's (Degnan, 2002) and prototypes have been completed during the last two decades (Degnan, 2016). This thesis focuses on the SPL100 airborne SPL system (described in more detail in section 2.2.1), one of the two airborne SPL systems currently available commercially, and the spaceborne ATLAS system onboard ICESat-2 (described in more detail in section 2.2.2).

ALS has been increasingly used in the past decades for forest structure and resources estimation and is now fully integrated into enhanced forest inventories (EFIs) frameworks. Estimation of forest attributes such as tree height, basal area, volume and biomass from ALS data is generally performed either with an area-based approach (ABA) or an individual tree crown (ITC) or individual tree detection (ITD) approach. The ABA (Næsset, 2002; White et al., 2013) is based on the development of predictive models based on ALS summary metrics of vegetation three-dimensional structure, such as height percentiles, standard deviation of height and canopy cover, and ground plot measurements. Predictive models are then applied over the region covered by ALS to provide wall-to-wall forest attribute estimations. The ABA is now implemented operationally (White et al., 2013; Woods et al., 2011; Wulder et al., 2008a). The positional and attribute accuracy associated with EFI has been shown to be significantly greater than that achieved from conventional polygon-based interpretation of aerial imagery (Holmgren, 2004; Næsset, 2004; Woods et al., 2011) due to the increased spatial accuracy of lidar data, as well as the ability to include additional geospatial information into the predictions such as information on terrain and local context. Alternatively, the ITC approach (Leckie et al., 2003) aims at automatically delineating tree crowns from the ALS data and derive tree level attributes information. By operating at the individual tree level, i.e. the smallest forest management unit possible, ITC approaches can enable the estimation of species-specific attributes which can be challenging to obtain from an ABA (Breidenbach et al., 2010a; Vastaranta et al., 2012). ITC

approaches are not yet commonly used operationally due to challenges in delineating all individual tree crowns present in a stand and subsequent impacts on estimation accuracy (White et al., 2016).

1.2.2. Optical imagery

Before the uptake of lidar into the forestry sector, forest inventories were mostly supported by the collection of airborne optical imagery. Such data, nowadays sometimes complemented by high (1 – 10 m) or very-high (< 1 m) spatial resolution optical satellite imagery, can be manually or automatically interpreted to delineate forest stands and estimate forest inventory attributes (Hall, 2003; White et al., 2016). A particular strength of optical imagery over lidar in a forest inventory context is the capability to measure reflectance in different parts of the electromagnetic spectrum, which can be particularly useful to estimate some attributes such as tree species composition (Fassnacht et al., 2016; White et al., 2016). The fusion of lidar data and very-high resolution multispectral or hyperspectral data or the use of digital aerial photogrammetry (DAP) in order to combine structural and spectral information is currently an active area of research for numerous applications (Goodbody et al., 2019; Zhang and Lin, 2017).

Long-term optical satellite imagery observations programs such as the Landsat suite of satellite (1972 – Present) or the MODerate Resolution Imaging Spectrometer (MODIS; 1999 – Present) also play a crucial role to monitor the state of forests. The free availability of the data (Wulder et al., 2012), advances in computing power, improved cloud masking and atmospheric corrections (Masek et al., 2006; Zhu and Woodcock, 2012), and more generally the generation of standardized and analysis-ready products (Dwyer et al., 2018; Hermosilla et al., 2016; White et al., 2014) have facilitated their uptake. The medium to low spatial resolution of Landsat (30-m) and MODIS (500 – 1000 m) limits their application for detailed forest inventory purposes but make them particularly suitable to monitor a wide range of applications such as land-use and forest cover, disturbance dynamics, vegetation health and phenology, and gross primary production to name a few, at a regional, national and global scale (Banskota et al., 2014; Justice et al., 2002; Wulder et al., 2022).

1.3. Potential values and limitations of airborne and spaceborne single-photon lidar for forest management and monitoring

In Canada, ALS acquisitions are often limited to southern managed areas, hence limiting the uptake of this technology for northern unmanaged forests characterization. As discussed above, airborne SPL systems decrease operational costs by enabling the acquisition of data from higher altitude and cruise speeds which result in an enhanced coverage (Degnan, 2016). Provided that airborne SPL systems are able to accurately characterize vegetation structure, they could play an important role in the future for systematic and updated EFIs. At the time of its operational deployment in 2018, limited information was known about the performance of airborne SPL sensors. Published literature focused on airborne SPL sensors design (Clifton et al., 2015; Degnan, 2016, 2002; Degnan et al., 2007; Degnan and Field, 2014) or data quality and processing (Gluckman, 2016; Tang et al., 2016; Ullrich and Pfennigbauer, 2016). A few studies had started to evaluate airborne SPL data for forest structure (Li et al., 2016; Swatantran et al., 2016) or ground elevation retrieval (Stoker et al., 2016). Despite clear operational advantages, several of the aforementioned studies highlighted potential issues related to (1) the use of a laser with a wavelength of 532 nm, required by SPL detectors as discussed previously, and (2) the use of laser pulses with low energy. Reflectance of leaves is indeed much reduced in the visible part of the electromagnetic spectrum (e.g., $\lambda = 532$ nm) compared to the near-infrared (e.g., $\lambda = 1064$ nm). Moreover, detectors operating in the visible part of the electromagnetic spectrum are much more sensitive to background solar noise and SPL data acquired during daylight conditions therefore requires intensive, and potentially aggressive, noise filtering (Swatantran et al., 2016; Tang et al., 2016). Finally, the use of low energy laser pulses could lead to a reduced capability to “penetrate” the canopy and characterize the ground and vegetation located below the dominant canopy layer, which is a particular strength of conventional linear-mode lidar systems operating in the near-infrared.

Unlike GEDI or GLAS which are both full-waveform systems, ATLAS onboard the ICESat-2 satellite is based on single-photon detection and enables the acquisition of nearly-continuous discrete-return lidar data

from space at an unprecedented level of detail. The ICESat-2 mission, launched in September 2018, was primarily designed as a cryospheric mission; however, the retrieval of canopy height to estimate biomass at a global-scale was identified as one of the scientific objectives (Neumann et al., 2019). ICESat-2 provides nearly global coverage ($\pm 98^\circ$ N/S latitude) of lidar data acquired along the orbital tracks of the satellite and could therefore enhance the characterization of forests structure, particularly where ALS data is not available. At the start of this thesis, on-orbit ICESat-2 data were not publicly available (launch on September 15, 2018; first data products public release on May 28, 2019) and only studies evaluating simulated data were available (Glenn et al., 2016; Gwenzi and Lefsky, 2014; McGill et al., 2013; Montesano et al., 2015; Neuenschwander and Magruder, 2016; Nie et al., 2018; Popescu et al., 2018; Wang et al., 2016). Potential issues with terrain detection under dense canopies (Neuenschwander and Magruder, 2016) and canopy height retrieval in ecosystems with sparse vegetation cover such as the tundra-taiga ecotone (Montesano et al., 2015) had already been identified prior to ICESat-2 launch for similar reasons as discussed above for airborne SPL systems; i.e. limited penetration of laser pulses through dense canopies and high-sensitivity to background solar noise.

1.4. Research questions

The primary objective of this dissertation is to examine the role that airborne and spaceborne SPL can play to provide spatially-explicit, timely, and accurate estimates for some of the key information needs in managed (e.g., height, volume, species composition) and unmanaged (e.g., height, canopy cover, monitoring through time) boreal forests of Canada. To that end, the dissertation is organized around four core research questions:

Q1: What is the accuracy of key forest inventory attributes derived from SPL100 data?

Q2: How can SPL100 data be used to include dominant tree species composition estimation in enhanced forest inventories?

Q3: How do ICESat-2 and SPL100 estimates of canopy height, cover and height variability compare across a gradient of boreal forest structure?

Q4: How can widespread ICESat-2 samples of forest structure be used to monitor canopy height over time?

1.5. Dissertation overview

The remainder of this dissertation is organized around six chapters (Chapter 2 – 7). Figure 1-2 provides a conceptual overview of the core research Chapters 3 – 6.

Chapter 2 describes the main study areas that this thesis focuses on and provides an overview of the remote sensing and field datasets that were used.

Chapter 3 establishes the ability of the SPL100 to provide accurate estimates of forest inventory attributes commonly derived in enhanced forest inventories.

Chapter 4 focuses on the derivation of species-composition information in managed forest stands, which is currently often derived from photo-interpreted sources, with SPL100 and Sentinel-2 data.

Chapter 5 compares SPL100 and ICESat-2 lidar metrics of height, canopy cover and height variability across a gradient of boreal forest conditions.

Chapter 6 implements a method to obtain wall-to-wall estimates of canopy height by extrapolating ICESat-2 metrics with Landsat times-series and explores some ways to examine canopy height growth and recovery trends over time in unmanaged boreal forested areas.

Chapter 7 concludes the dissertation with a synthesis of the work and highlights innovations as well as limitations and future research opportunities.

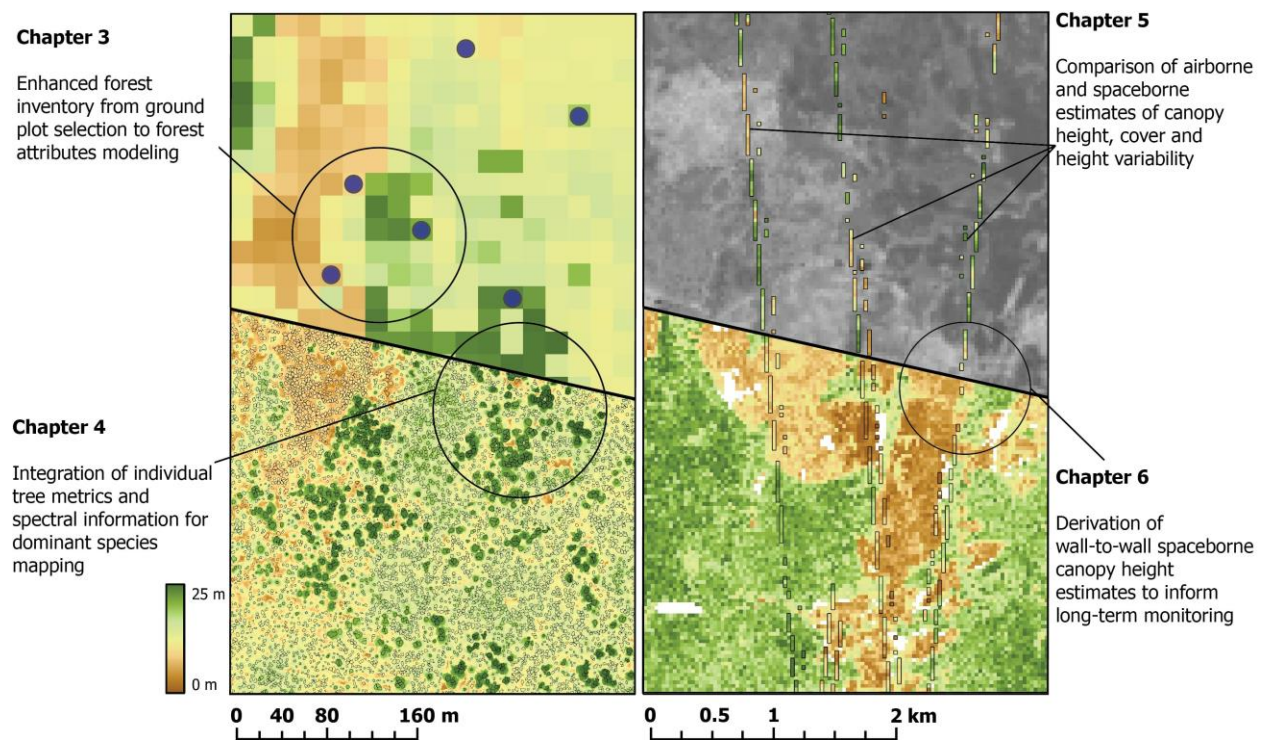


Figure 1-2. Overview of the core chapters of this dissertation.

Chapter 2. Study area and data

2.1. Study areas

This dissertation is positioned around two key focus areas located in the Canadian boreal zone and more specifically in the province of Ontario. The Canadian boreal zone intersects multiple ecozones² and spans an area of 550 million ha composed of forests, woodlands, wetlands and lakes (Brandt, 2009). The Boreal Shield is the ecozone containing the largest portion of the boreal zone and extends from Eastern Saskatchewan to Newfoundland (Figure 2-1A). According to the Ecological Stratification Working Group (1996), this ecozone is characterized by a continental climate with short warm summers and long cold winters. Forests are dominated by cold-tolerant coniferous species such as black spruce, white spruce and balsam fir. Broadleaf tree species such as white birch, trembling aspen and balsam poplar as well as other coniferous species such as jack pine and white pine are also common towards the southern latitudes of the ecozone. The portion of the Boreal Shield ecozone that intersects with the province of Ontario is the main focus of this dissertation. In particular, the Romeo Malette Forest, an actively managed forest, and an unmanaged portion of the Boreal Shield ecozone were studied (Figure 2-1B).

² The Ecological Stratification Working Group (1996) has divided Canada into 15 terrestrial ecological units, called ecozones, based on vegetation, climate, soil and geomorphology characteristics.

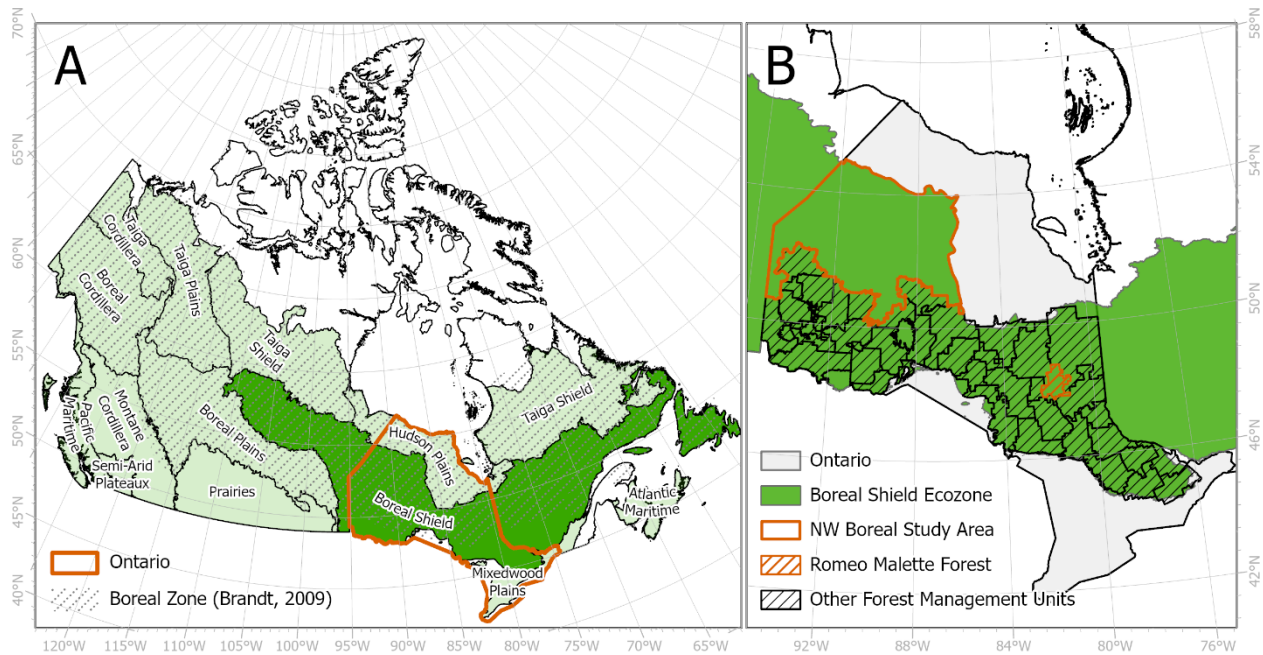


Figure 2-1. Ecozones totally or partially covered by forests (A) and overview of the Romeo Malette Forest and NW Boreal study areas in Ontario (B).

2.1.1. Romeo Malette Forest

The southern portion of the Boreal Shield of Ontario (~ 44°N – 51°N latitude) consists of 39 forest management units ranging in size from 300,000 to 3,600,000 ha (Figure 2-1B). The thesis focused particularly on the Romeo Malette Forest management unit which has an area of approximately 630,000 ha and is located in the vicinity of the city of Timmins, Ontario (Figure 2-2), where monthly average daily mean temperatures range from -16.1 °C in January to 17.8 °C in July and monthly cumulated precipitation (rain and snow combined) range from 34.4 mm in January and 88.2 mm in October (Figure 2-3). The terrain consists in flat to moderately rolling topography (305 – 380 m a.s.l) and poorly drained soils resulting in extensive wetlands, watercourses and lakes. The forest is mostly composed of black spruce (*Picea mariana*), jack pine (*Pinus banksiana*), poplar (*Populus* spp.), white birch (*Betula papyrifera*), white spruce (*Picea glauca*), cedar (*Thuja* spp.), larch (*Larix* spp.) and balsam fir (*Abies balsamea*).

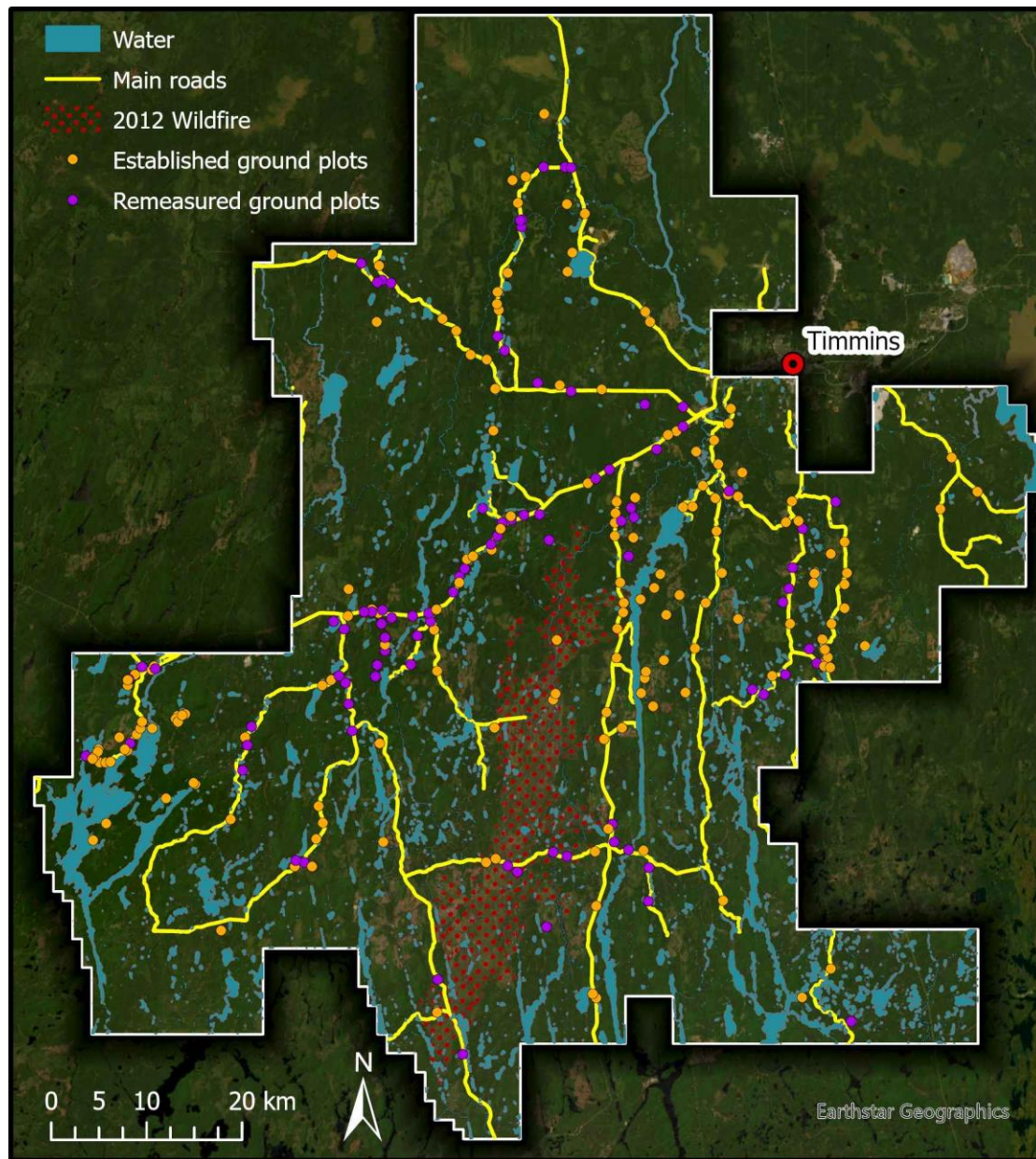


Figure 2-2. Map of the Romeo Malette Forest study area overlaid with the network of main roads, waterbodies and watercourses and ground plots established as part of the enhanced forest inventory carried out in Chapter 3. The footprint of a large wildfire that occurred in the summer of 2012 is also displayed.

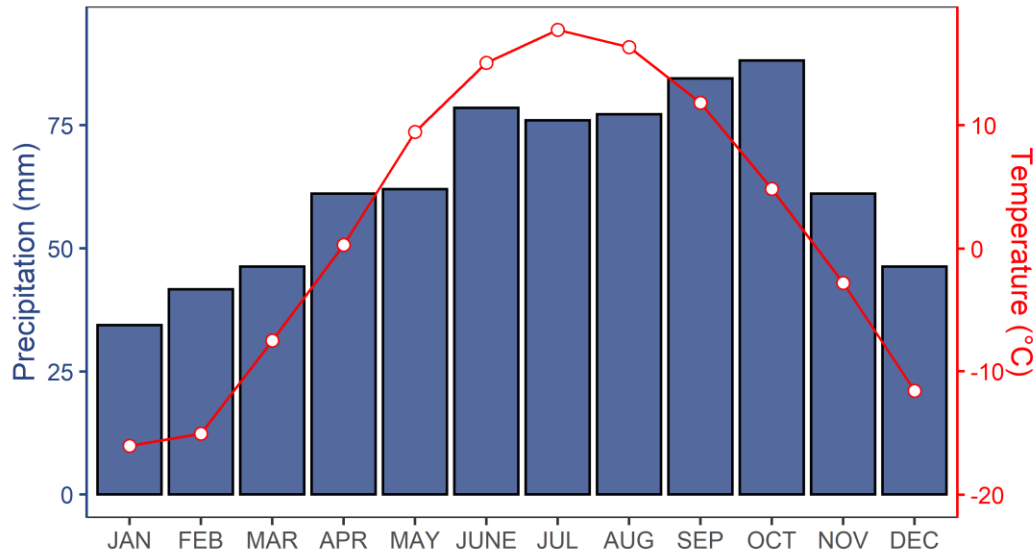


Figure 2-3. Monthly average of daily mean temperature and cumulated precipitation recorded from 2008 to 2022 at the weather station of Timmins, Ontario (Station ID 47547). Data obtained from Environment and Climate Change Canada through the R package [weathercan](#).

The Romeo Malette Forest is actively managed for timber production and fibre procurement. Management activities are guided by Ontario's Crown Forest Sustainability Act and related policies, and are based on a 10-year forest management plan, which includes a detailed description and mapping of logging and renewal operations and the development of planned road access. The current plan covers years 2019 to 2029 and is publicly available online at https://nrp.mnr.gov.on.ca/s/fmp-online?language=en_US. Annual work schedules identify the location of operations scheduled for implementation during each year, and provide information on the areas scheduled for harvest, road construction/maintenance, renewal, tending and protection operations. The Romeo Malette Forest has been certified under the National Boreal Standard of the Forest Stewardship Council since 2006.

2.1.2. North-Western Boreal Shield Ecozone study area

The north-western portion of the Boreal Shield ecozone of Ontario, referred herein as the NW Boreal study area, is an area of 19.6 Mha composed mainly of unmanaged forests, shrubs, wetlands and water. Disturbances detected from Landsat time-series between 1984 and 2018 (see section 2.2.3) were dominated by fire and non-stand replacing disturbances (e.g., insects or disease outbreaks), which represented 12% and 3% of the total area, respectively. Mixed wood, coniferous and wetland-treed land cover types represent 32%, 24% and 14% of the total area. A transition from boreal forest towards a landscape dominated by wetlands and arctic tundra (Hudson Plains ecozone) occurs at the northern and eastern boundary of this area (Figure 2-4).

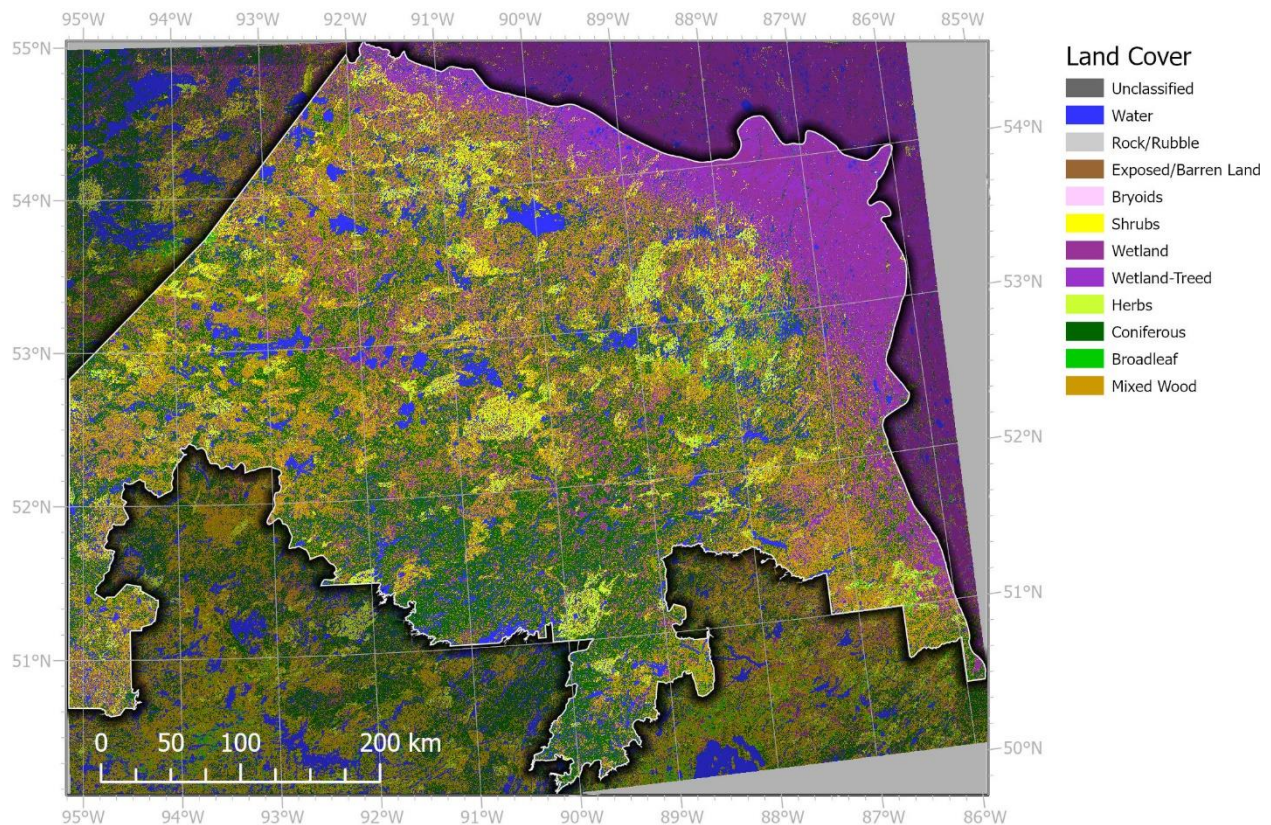


Figure 2-4. Land cover in 2018 at the NW Boreal study area according to the Virtual Land Cover Engine classification (Hermosilla et al., 2018)

2.2. Remotely sensed data

The primary source of remotely sensed data used throughout this dissertation are lidar data that have been acquired from airborne (Chapters 3, 4, 5) and spaceborne (Chapters 5, 6) platforms to characterize forest structure and model forest attributes. Spectral information obtained from Sentinel-2 imagery is also used in Chapter 4 for dominant tree species mapping. Finally, annual time-series (1984 – 2021) of Landsat products, in particular surface reflectance and disturbance mapping products, are used in Chapter 6.

2.2.1. Airborne SPL100 data

The Ontario Ministry of Natural Resources and Forestry launched a new EFI program in 2018 supported by the acquisition of airborne SPL100 (Hexagon / Leica Geosystems) data across the 39 forest management units of Ontario (Bilyk et al., 2021). The Romeo Malette Forest was one of the first scheduled acquisitions and took place between June and July 2018. Other forest management units were flown during the summers of 2019, 2020 and 2021. As of mid 2022, not all forest management units had been flown (Figure 2-5).

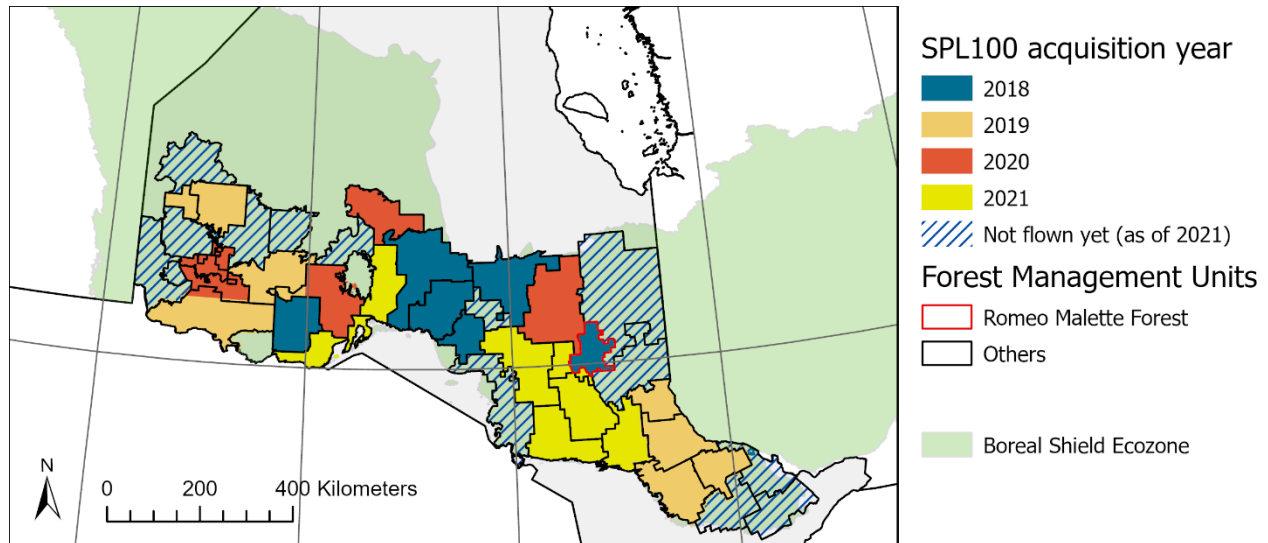


Figure 2-5. Overview of SPL100 acquisitions schedule across the forest management units of Ontario as of summer 2021.

The SPL100 transmits laser pulses in the visible part of the electromagnetic spectrum ($\lambda = 532 \text{ nm}$) that are split into an array of 10×10 low-energy beamlets by a diffractive optical element. The SPL100 laser beams have a divergence of 0.08 mrad , which results in footprints of approximately 30 cm diameter for

acquisitions at an altitude of about 4000 m above ground. The receiver unit is a 10 x 10 array of single-photon sensitive cells that each contain thousands of photon-sensitive channels. Each photon returning to a detector illuminates and therefore inactivates only a small number of the photo-sensitive channels, leaving the remaining portion available for subsequent photon detection (Degnan, 2016). The short recovery time (1.6 ns) of the detector cells after a photon is detected also allows for measurement of multiple returns for a single beamlet (Degnan, 2016).

Table 2-1. Technical specifications of the SPL100 acquisition at the Romeo Malette Forest.

Parameter	Romeo Malette Forest acquisition
Sensor	SPL100 (Leica / Hexagon Geosystems)
Acquisition date	June / July 2018 (leaf-on conditions)
Altitude	3,800 – 4,000 m above ground
Flying speed	350 km / h (180 knots)
Acquisition area	689,700 ha (6,897 tiles of 1 x 1 km)
Laser wavelength	532 nm
Laser pulse width	400 ps (10^{-12} s)
Laser beam divergence	0.08 mrad (footprint of ~ 30 cm diameter from 4,000 m altitude above ground)
Beam configuration	Each pulse split into a 10 x 10 array of beamlets by a diffractive optical element
Number of returns	Maximum 5 returns per beamlet
Pulse repetition frequency	60 kHz; resulting in an effective pulse repetition frequency of 6.0 MHz after splitting into the 10 x 10 array of beamlets
Scanning pattern	Conical
Field-of-view	30° (fixed)
Average return density	40.0 points / m ²

The Romeo Malette Forest SPL100 acquisition was carried out at an altitude of 3800 – 4000 m (~ 12,500 – 13,000 ft) above ground level with a cruise speed of 350 km/h (180 knots). The laser transmitter unit was operated at a pulse repetition rate of 60 kHz, which results in an effective pulse repetition rate of

6.0 MHz after diffraction into the 10 x 10 array of beamlets. The SPL100 uses a conical scanning pattern and was operated with a fixed field of view of 30° resulting in an average swath of 2000 m width (Figure 2-6). A total of 127 parallel flight lines with a parallel overlap of 50% were necessary to cover the entire Romeo Malette Forest. The lidar data were preprocessed by the data provider according to the processing chain presented in Gluckman (2016) which include geolocation, background solar noise removal, point cloud registration and classification of returns into ground, vegetation or water following LAS 1.4 specifications (Figure 2-10A). The horizontal coordinates of the point cloud were projected to the UTM Zone 17N. The Canadian Geodetic Vertical Datum 2013 (CGVD2013) was used to register points height. The data were delivered in a set of 6,897 tiles of 1 x 1 km (Figure 2-7).

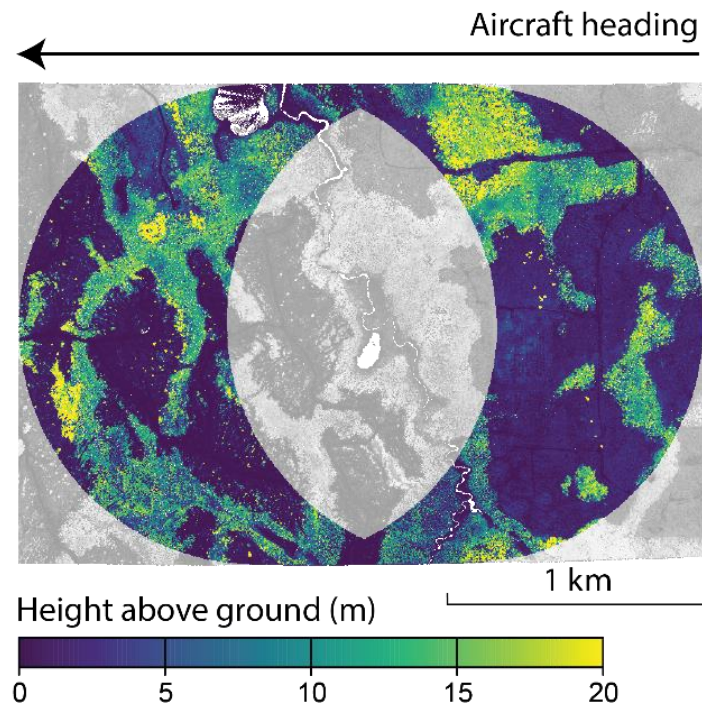


Figure 2-6. Conical scanning footprint of a 10-s acquisition of SPL100 data from an aircraft flown at an altitude of 3,800 m above ground and cruising speed of 350 km/h in the West-East direction.

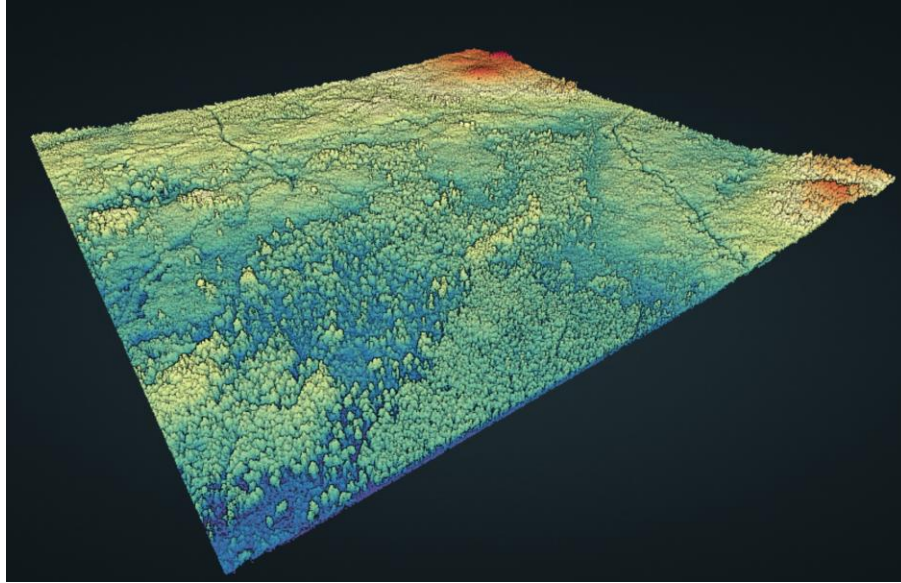


Figure 2-7. Visualization of a 1 x 1 km tile of the SPL100 point cloud acquired at the Romeo Malette Forest. Points are colored by elevation above sea level from blue (low) to red (high).

2.2.2. ICESat-2 data

The National Aeronautics and Space Administration (NASA) ICESat-2 mission, launched in September 2018, was primarily designed as a cryospheric mission; however, the retrieval of canopy height to estimate biomass at a global-scale was identified as one of the scientific objectives of the mission (Neumann et al., 2019). Unlike its predecessor GLAS (onboard ICESat), the ATLAS lidar unit onboard ICESat-2 is based on single-photon detection technology (Neumann et al., 2019). ATLAS transmits 3 pairs of laser beams at a wavelength of 532 nm. Each pair is composed of a strong energy beam and a weak energy beam (4:1 energy ratio) transmitted at a rate of 10 kHz (Neumann et al., 2019). The beam configuration results in 3 pairs of parallel profiles spaced by 3.3 km (in-between pairs spacing of 90 m) on the ground (Figure 2-8). The spacecraft orientation is shifted about twice a year to maximize sun illumination on solar panel arrays and result in a change of the relative position (left/right) of the strong and weak beams.

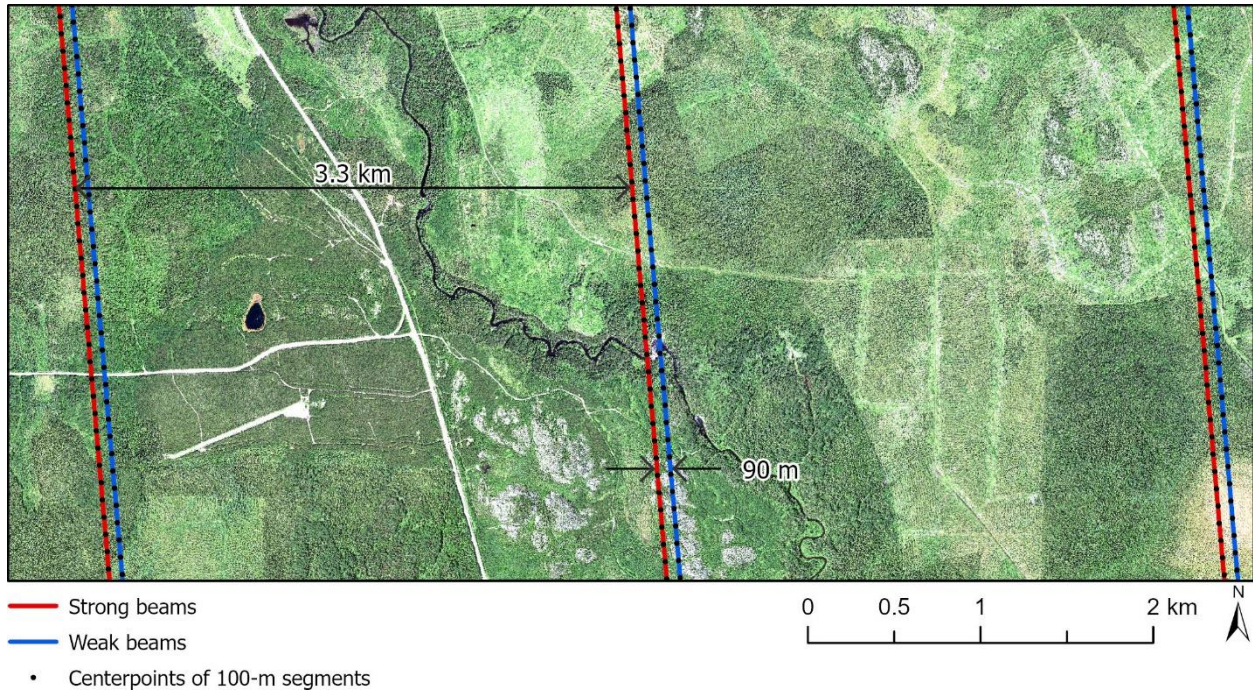


Figure 2-8. Configuration of the 3 pairs of ATLAS strong and weak beams for an acquisition over the Romeo Malette Forest.

ICESat-2 has an orbital repeat cycle of 91 days but the laser pointing direction is slightly shifted off-nadir at every repeat cycle over land to enhance the spatial distribution of measurements in the mid-latitudes (Neuenschwander and Pitts, 2019). Each ATLAS beam results in footprints of approximately 11 m diameter at the ground level spaced by 70 cm in the along-ground track direction (Magruder et al., 2020; Neuenschwander et al., 2020a). Each ATLAS pulse can result in multiple detection event. Prior to the ICESat-2 launch, it was anticipated that an average of 10 photons would be returned and detected by ATLAS for a single pulse of a strong energy beam over highly reflective surfaces such as snow and ice and under clear skies (Markus et al., 2017). Over vegetated lands where a larger amount of the incoming pulse energy is absorbed by vegetation and ground, 0 to 4 photons were expected to be returned and detected for a single strong energy pulse (Neuenschwander and Pitts, 2019). The horizontal geolocation accuracy of ATLAS pulses was expected to be well within the mission requirement of 6.5 m (Magruder et al., 2020; Neumann et al., 2019), with Luthcke et al. (2021) reporting beam-dependent geolocation errors ranging from 2.5 m for beam 6 to 4.4. m for beam 2.

The data collected by ICESat-2 is delivered in a set of Level-1, Level-2, and Level-3 products. The Level-2 Global Geolocated Photon Data product (ATL03; Neumann et al., 2020) provides geophysical-corrected latitude, longitude, and height above the WGS84 ellipsoid of each downlinked photon and is the source for all higher level products. The Level-3 Land and Vegetation Height data product (ATL08; Neuenschwander et al., 2020b), classifies the ATL03 photons as noise, ground, canopy, or top of canopy as shown in Figure 2-9 and Figure 2-10B (Neuenschwander and Pitts, 2019). The ATL08 data product also provides ground and canopy height summary metrics for 100-m segments (and 20-m segments since ATL08 version 005) in the along-track direction (Figure 2-9). ICESat-2 is publicly available and can be obtained from the NASA National Snow and Ice Data Center (<https://nsidc.org/data/icesat-2>).

ATL03

Photon index	20-m segment ID	Number of photons in segment	Latitude of photon	Longitude of photon	Height of photon above WGS84 ellipsoid
	<i>segment_id</i>	<i>segment_ph_cnt</i>	<i>lat_ph</i>	<i>lon_ph</i>	<i>h_ph</i>
1	269661	79	48.5623	-81.6138	351.40
2	269661	79	48.5623	-81.6138	306.10
3	269661	79	48.5623	-81.6138	256.92
...
80	269662	65	48.5625	-81.6138	312.50
81	269662	65	48.5625	-81.6138	266.23
82	269662	65	48.5625	-81.6138	182.79
...
145	269663	66	48.5627	-81.6138	305.15
146	269663	66	48.5627	-81.6138	206.67
147	269663	66	48.5627	-81.6138	267.06
...
211	269664	66	48.5629	-81.6139	321.44
212	269664	66	48.5629	-81.6139	265.29
213	269664	66	48.5629	-81.6139	251.06
...
277	269665	71	48.5630	-81.6139	300.50
278	269665	71	48.5630	-81.6139	279.84
279	269665	71	48.5630	-81.6139	258.62

ATL08a

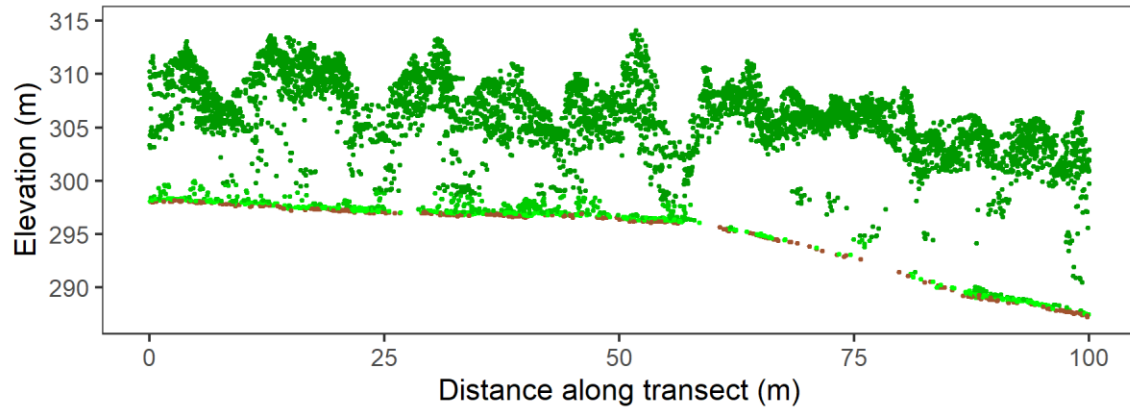
20-m segment ID	Index of photon in the segment	Photon classification
<i>ph_segment_id</i>	<i>classed_pc_idx</i>	<i>classed_pc_flag</i>
269661	1	NA
269661	2	ground
269661	3	NA
...
269662	1	NA
269662	2	ground
269662	3	NA
...
269663	1	top of canopy
269663	2	canopy
269663	3	NA
...
269664	1	NA
269664	2	NA
269664	3	NA
...
269665	1	canopy
269665	2	ground
269665	3	NA

ATL08b

Start segment ID	End segment ID	Canopy height	Canopy height for each 20-m segment	Other canopy and terrain metrics
<i>segment_id_beg</i>	<i>segment_id_end</i>	<i>h_canopy</i>	<i>h_canopy_20m</i>	...
269661	269665	19.43	19.96; 15.98; 17.8; 9.98; 8.95	...
269666	269670

Figure 2-9. Overview of the ATL03 and ATL08 data products. The ATL03 product contains the coordinates of each downlinked photon. The ATL08 product contains the classification of each photon (ATL08a) and canopy and terrain height summary metrics along 20-m and 100-m segments (ATL08b). Note that not all photons of the ATL03 product are classified by the ATL08 classification algorithm, as indicated by the NA values in the field *classed_pc_flag* of ATL08a.

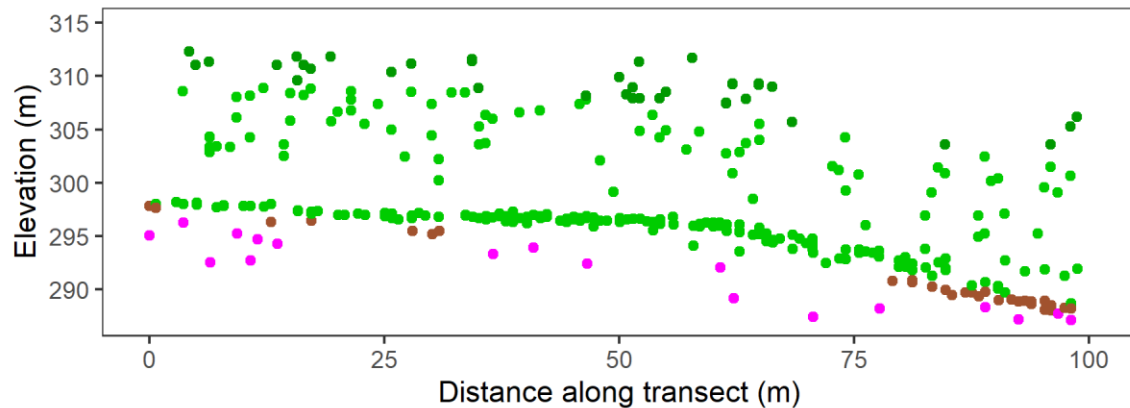
A SPL100



SPL100 Classification (LAS 1.4 standard)

• Ground • Low Vegetation • Medium Vegetation • High Vegetation

B ICESat-2



ATL08 classification

• Ground • Canopy • Top of canopy • Noise

Figure 2-10. SPL100 (A) and ICESat-2 (B) classified point clouds acquired along an ICESat-2 100-m segment (strong beam). The elevation of the ICESat-2 photons was transformed to the Canadian Geodetic Vertical Datum of 2013 (CGVD2013) to match SPL100 photons elevation.

2.2.3. Landsat time-series

The NASA/United States Geological Survey (USGS) Landsat satellites program has been continuously collecting multispectral imagery of the Earth from 1972 – Present. Landsat 1, 2 and 3 (1972 – 1983) carried

the Multispectral Scanner (MSS) sensor which collected images in four spectral bands at a spatial resolution of 60 m. Radiometric and spatial resolution improved with the Thematic Mapper (TM) onboard Landsat 4 and 5 (1982 – 2013) and Enhanced Thematic Mapper Plus (ETM+) onboard Landsat 7 (1999 – Present) which both collected images across seven spectral bands and at a 30-m spatial resolution. Landsat 8 (2013–Present) and Landsat 9 (2021–Present) carry the Operational Land Imager (OLI) / OLI-2 and the Thermal Infrared Sensor (TIRS) / TIRS-2 which collect images in eleven spectral bands and a spatial resolution of 30 m.

The free and open release of the Landsat archive to the public in 2008 (Woodcock et al., 2008) together with advances in data calibration, cloud masking and atmospheric corrections for the generation of surface reflectance products, and high-performance computing have enabled a rapid development of Landsat applications by the scientific community (Wulder et al., 2022, 2019). In Canada, the National Terrestrial Ecosystem Monitoring System (NTEMS) program supported the development of annual Landsat-based analysis-ready products at a 30-m spatial resolution across the country since 1984 (White et al., 2014). NTEMS products relied first on the generation of an annual time-series of surface reflectance images using a best-available-pixel (BAP) compositing method of Landsat images acquired ± 30 days of August 1 (White et al., 2014) where data gaps were in-filled using a protocol considering spatio-temporal trends (Hermosilla et al., 2015a). Methods were developed to detect and quantify spectral changes in the BAP time-series (Hermosilla et al., 2015a) and attribute them to a disturbance type such as fire, harvesting, non-stand replacing disturbances (e.g. insects or disease outbreaks) or road construction (Hermosilla et al., 2015b).

These standardized and analysis-ready products were used for numerous applications such as disturbance mapping (Hermosilla et al., 2016) post-disturbance forest recovery (White et al., 2017b) land cover mapping (Hermosilla et al., 2022, 2018) or carbon accounting (C. Boisvenue et al., 2016; Céline Boisvenue et al., 2016). In the Chapter 6 of this dissertation, 1984 – 2021 yearly time-series of NTEMS gap-free surface reflectance (Hermosilla et al., 2015a; White et al., 2014), land cover (Hermosilla et al., 2018; Wulder et al., 2018) and change characterization (Hermosilla et al., 2015b) were used.

2.2.4. Sentinel-2 data

The European Space Agency Copernicus Sentinel-2 mission (Drusch et al., 2012) is currently composed of a constellation of two satellites (Sentinel-2A and Sentinel-2B) collecting multispectral imagery of the Earth across 13 spectral bands and at a spatial resolution of 10, 20 or 60 m depending on the spectral bands with the MultiSpectral Instrument (MSI). Sentinel-2A and Sentinel-2B were launched in 2015 and 2017, respectively. The type of data collected by Sentinel-2 and Landsat are similar and harmonized products combining both have been produced (Claverie et al., 2018). In the chapter 4 of this dissertation, Sentinel-2 data were selected over Landsat because the 10-m or 20-m spatial resolution of Sentinel-2 spectral bands of interests (visible, near-infrared and shortwave infrared) were more suitable to be combined with the EFI data produced at a spatial resolution of 20 m.

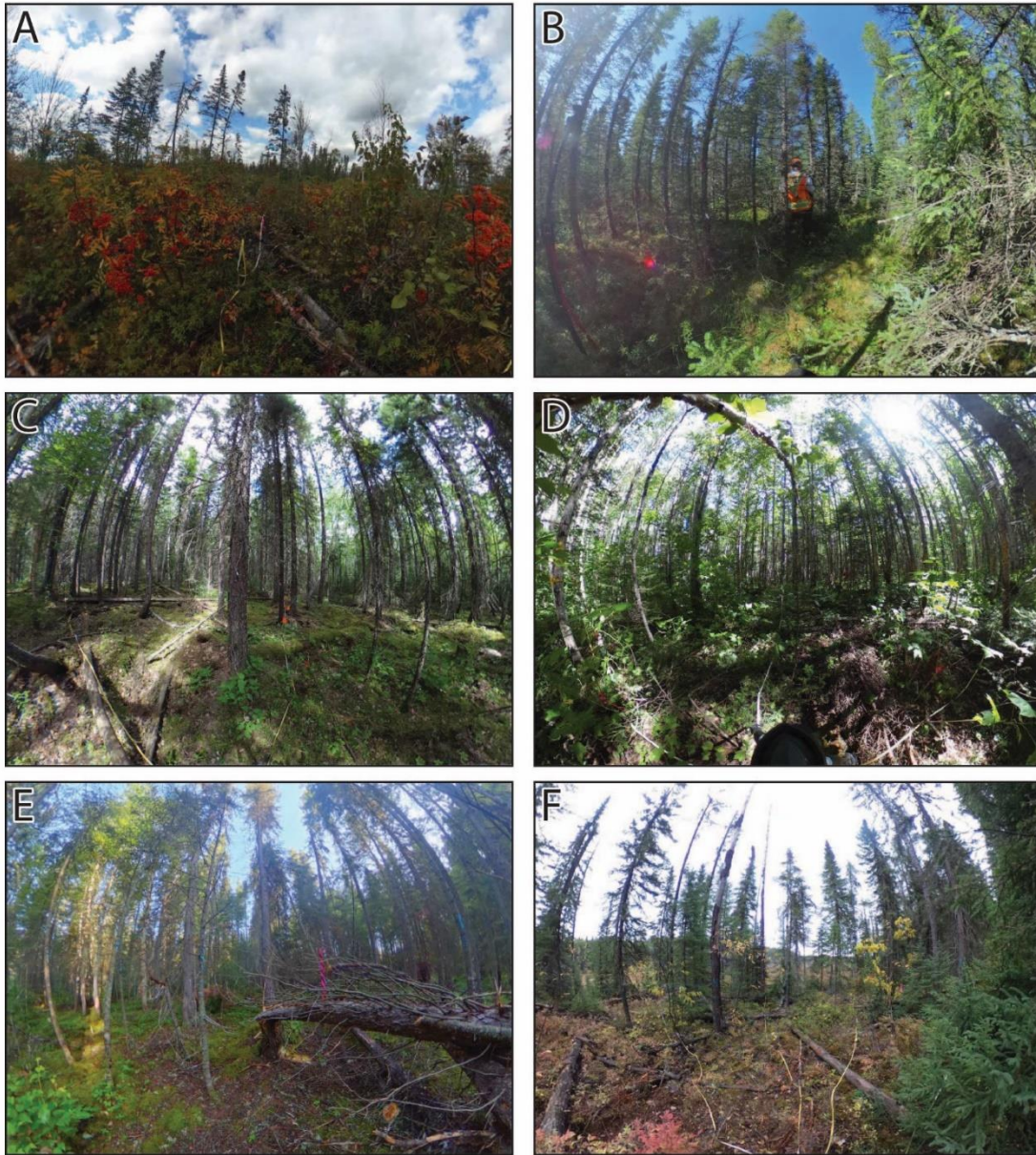
2.3. Field data

Field data were collected across fixed-radius circular ground plots ($r = 11.28$ m; 400 m²) at the Romeo Malette Forest between June and December 2019 to support the development of an EFI. The selection of ground plot locations was achieved using a stratified random sampling guided by metrics derived from the SPL100 acquisition that is described in Chapter 3. An overview of the main types of measurements collected at each plot is provided in Table 2-2, followed by an overall description of the sampled forest conditions.

Plot centre coordinates were obtained with the EOS Arrow 100 GPS receiver by acquiring a minimum of 600 points per plot centre and the use of differential post processing to achieve sub-metre position accuracy. The height, species, and diameter at breast height (DBH) of each tree with $DBH \geq 7.1$ cm located within the plot boundary were recorded at each plot. Additionally, 58 of the plots were stem mapped by recording the distance and azimuth from plot centre to each stem ($DBH \geq 7.1$ cm) located in the plot boundary using an azimuth and range finder instrument (TruPulse® 360). For leaning trees, the distance and azimuth between the stem and the vertical projection of the crown top on the ground were also measured. Trees with a crown located inside the plot boundary but with stem outside were also mapped. The cartesian coordinates

of tree tops were determined from the polar coordinates (distance and azimuth of tree tops) measured from the plot centre.

The plots covered a range of canopy structure, with 146 plots (58.4%) considered to have a complex structure with no discernable layers, 91 plots (36.4%) with a single canopy layer including 15 plots where veteran trees were observed above the dominant canopy layer, and 13 plots (5.2%) with two distinct canopy layers. A range of development stages were also observed, with 124 plots (49.6%) identified as stable, 58 (23.2%) as overmature, 38 (15.2%) as declining, 17 (6.8%) as regenerating and 13 (5.2%) as growth (see description of maturity classes in Table 2-2). Exemplar photos of ground plots covering a range of development stages are provided in Figure 2-11.



- A) Maturity Code: Regenerating; Dominant Species: Balsam Fir; Crown Closure: 20%
- B) Maturity Code: Growth; Dominant Species: Black Spruce; Crown Closure: 70%
- C) Maturity Code: Stable; Dominant Species: Black Spruce; Crown Closure: 70%
- D) Maturity Code: Stable; Dominant Species: Trembling Aspen; Crown Closure: 70%
- E) Maturity Code: Overmature; Dominant Species: Black Spruce; Crown Closure: 60%
- F) Maturity Code: Declining; Dominant Species: Jack Pine; Crown Closure: 45%

Figure 2-11. Captures of 360° photographs taken from plot center in stands with different maturity code, species composition and crown closure.

Table 2-2. Non-exhaustive list of attributes recorded at the plot-level and individual tree level by field crews.

Scale	Attribute	Description	Type of measurement	Number of plots
Plot-level	Plot Center	Coordinates of plot center	GPS (EOS Arrow 100)	250
	Canopy Structure	Vertical and horizontal stratification of tree crowns by species, height, age, and size.	Visual Assessment	250
	Maturity Class	Development stage of the dominant cohort. Assigned as either <i>regenerating</i> (few to no trees but capacity to grow), <i>growth</i> (young stand haven't achieved crown closure), <i>mature</i> (stand having achieved crown closure and with no more available growing space), <i>overmature</i> (small gaps in the canopy and mortality in the overstory) or <i>declining</i> (large gaps in the canopy and significant mortality in the overstory).	Visual Assessment	250
	Crown Closure	Proportion of the plot covered by tree crowns	Visual Assessment	250
Tree-level	Tree Status	Indicates if the tree is live or dead	Visual Assessment	250
	Tree Species	Species of the tree	Visual Assessment	250
	DBH	Diameter at breast height	Measured with DBH tape	250
	Tree Height	Height to the tree top	Measured with an electronic hypsometer	250
	Crown Class	Position of the tree in relation to surrounding trees (i.e., emergent, dominant, codominant, intermediate or suppressed)	Visual Assessment	250
	Tree position	Distance and azimuth of tree stems from plot center.	Range finder and azimuth instrument (TruPulse 360)	58
	Crown Width	Width of the tree crown represented as a circle or ellipsoid.	Width along the major and minor axis of the crown	58

Across all plots, a total of 13,389 individual trees were tallied. Species that occurred most frequently were black spruce (n = 2,968; 22.2 %), trembling aspen (n = 2,611; 19.5 %), balsam fir (n = 2,512; 18.8 %), jack pine (n = 2,233; 16.7 %) and paper birch (n = 1,925; 14.4 %). Some species such as trembling aspen mostly occurred in the dominant or codominant canopy layer while other species such as balsam fir were more often suppressed or located in intermediate canopy layers (Figure 2-12).

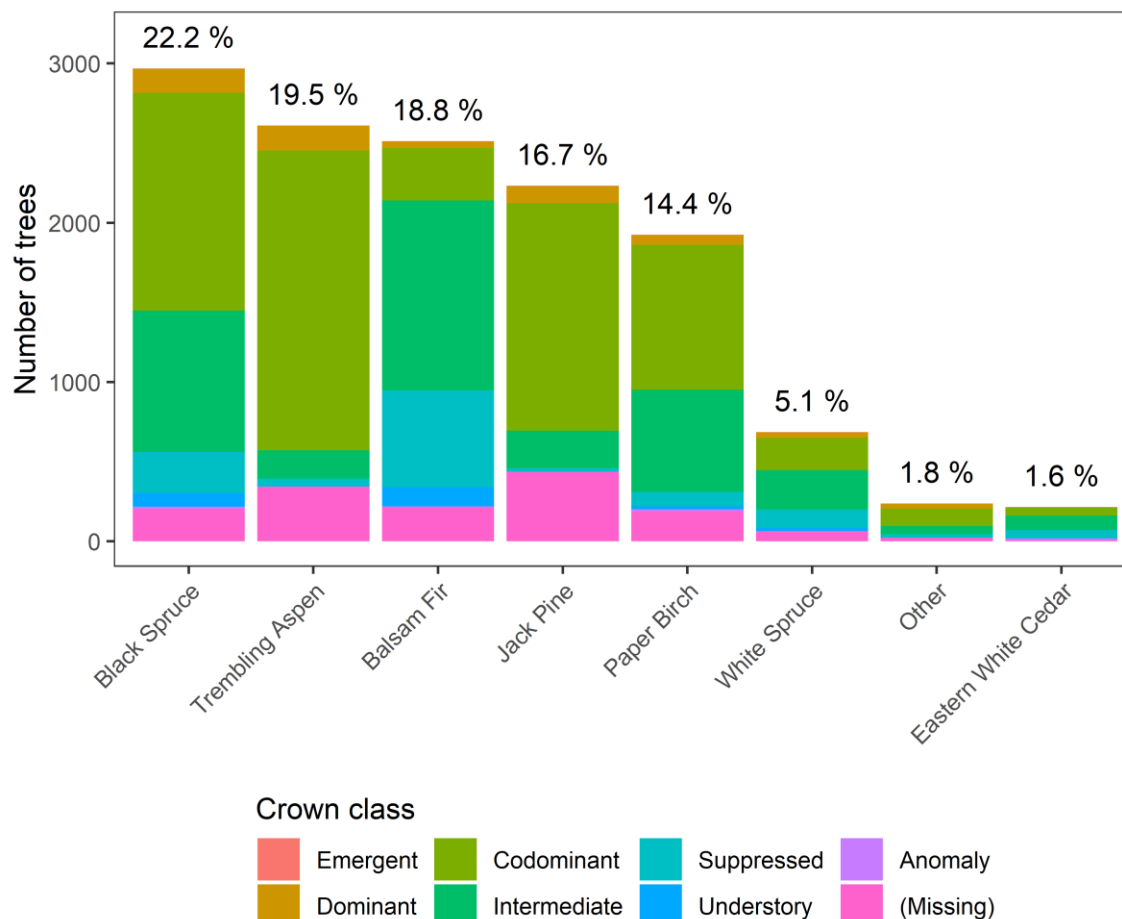


Figure 2-12. Distribution of species of the 13,389 tallied trees grouped by the position of their crowns relative to their surrounding trees.

Trembling aspen trees generally had the largest height and DBH with median values of 22.5 m and 23.8 cm, respectively (Figure 2-13; Table 2-3). Other species were more similar in terms of DBH and height, with median values ranging between 12.5 – 15.7 cm for DBH and 9.5 – 16 m for height.

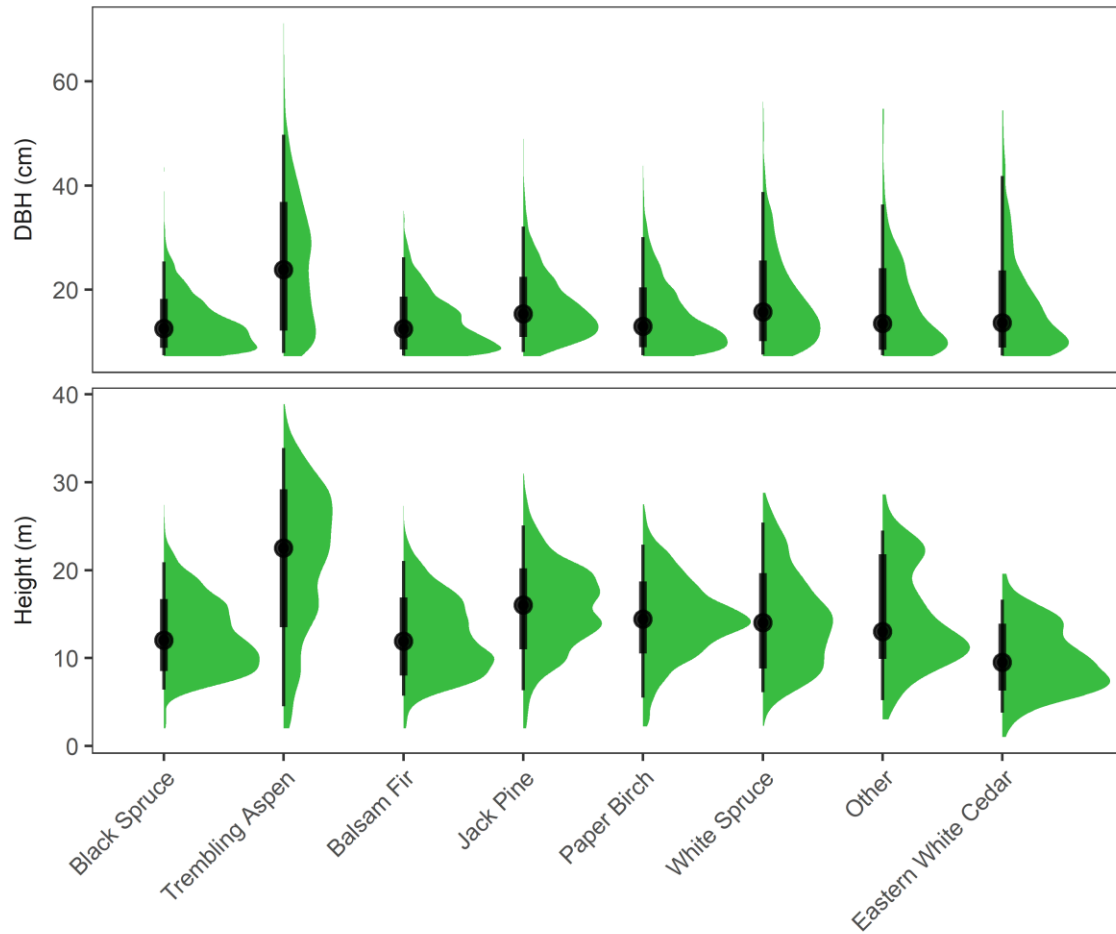


Figure 2-13. Distribution of DBH (≥ 7.1 cm) and tree height per species. Black dots indicate the mean of the distribution, wide and thin lines indicate where 66% and 95% of the values around the median are located, respectively. Two trees (one trembling aspen and one in the “other” category) with DBH of 81.2 cm and 89.5 cm are not included in the DBH graph.

Table 2-3. Median, 25th percentile and 75th percentile of the distribution of measured tree DBH and height per species.

Attribute	Species	Number of trees tallied	Median	25 th percentile	75 th percentile
DBH (cm)	Black Spruce	2,968	12.5	9.7	16.3
	Trembling Aspen	2,611	23.8	14.5	33.3
	Balsam Fir	2,512	12.4	9.3	16.7
	Jack Pine	2,233	15.3	12	20.1
	Paper Birch	1,925	12.9	9.8	18.2
	White Spruce	687	15.7	11.4	22
	Other	237	13.5	9.3	20.9
	Eastern White Cedar	216	13.6	9.7	20.2
	All combined	13,389	14.3	10.5	20.4
Height (m)	Black Spruce	2,968	12	9.3	15.6
	Trembling Aspen	2,611	22.5	15.9	27.5
	Balsam Fir	2,512	11.9	9	15.7
	Jack Pine	2,233	16	12.5	19.2
	Paper Birch	1,925	14.4	11.9	17.3
	White Spruce	687	14	10	17.9
	Other	237	13.1	10.5	18.7
	Eastern White Cedar	216	9.5	7	12.5
	All combined	13,389	14.3	10.4	18.7

Chapter 3. Developing a forest inventory approach using airborne single photon lidar data: from ground plot selection to forest attributes selection³

3.1. Introduction

Spatially explicit, timely and accurate information on the forest estate is critical in order for forest managers to meet societal expectations around the sustainable management of the resource (White et al., 2016). Knowledge of tree, stand, and site attributes is paramount and lack of such information can lead to suboptimal forest management decisions with negative impacts on the environment and along the entire forest value chain (Alam et al., 2014). Forest inventories can serve multiple resource management objectives for example provisioning of timber, assessment of habitat for biodiversity, as well as drive decision making around optimal silvicultural strategies (Wulder et al., 2008a). A forest inventory can also be used for future planning to support investment decisions that are necessary to enhance the competitiveness of the forest sector (Alam et al., 2014).

Today the forest industry has seen the widespread uptake of ALS data into forest inventory programs. Significant innovations in linear-mode lidar systems have enhanced the precision and amount of detail they can capture. However, linear-mode lidar systems are essentially limited in the number of pulses per second they are able to emit and the strength of the returned signal to exceed minimum detection thresholds, which places upper limits on airborne acquisition parameters (Harding, 2018; Nelson, 2013). While innovations that lead to the development of SPL systems, including the SPL100 instrument, were not driven by the information needs of forest inventory applications directly, the capacity for more efficient, large area data

³ The content of this chapter has been adapted from:

Queinnec, M., Coops, N. C., White, J. C., McCartney, G., & Sinclair, I. (2022). Developing a forest inventory approach using airborne single photon lidar data: from ground plot selection to forest attribute prediction. *Forestry: An International Journal of Forest Research*, 95(3), 347-362.

acquisitions with a consistent instrument and parameters offers significant benefits for operational inventory programs. However, concerns exist about the ability of the SPL100 low-energy laser pulses to penetrate through the canopy and detect understory vegetation and the ground surface due to their low energy and green wavelength. Brown et al. (2020), Yu et al. (2020), and White et al. (2021a) have reported that the SPL100 point cloud was dominated by first returns originating from the outer canopy layer. The larger beam density of the SPL100 therefore did not necessarily lead to a uniform increase in point cloud density in the upper and intermediate layers of the canopy, nor on the ground surface (Li et al., 2016; White et al., 2021a). Irwin et al. (2021) also observed a 31% reduction of returns reaching the mid-canopy understory vegetation with the SPL100 compared to a linear-mode lidar instrument mounted on a remotely piloted aircraft system in forest types with closed canopy. Brown et al. (2020) and Stoker et al. (2016) reported that the SPL100 had the poorer range precision and ground elevation accuracy when compared to linear-mode lidar instruments for both vegetated and non-vegetated targets. White et al. (2021a) examined the relative and absolute accuracies of one leaf-on and two leaf-off SPL100 datasets, acquired at different altitudes, to derive elevation information in forest environments. Results showed that the SPL100 datasets were markedly different when acquired under leaf-off vs leaf-on conditions and that under leaf-on conditions derivation of terrain information was poorer; however, accuracies were still within accepted error limits for elevation products. Results suggested that the density of the vegetation, as well as the vertical vegetation configuration impacted the derivation of the terrain surface (White et al., 2021a).

Recent research has used SPL100 data to derive area-based estimates of forest attributes in boreal and hemi-boreal forests of Sweden (Wästlund et al., 2018), Finland (Raty et al., 2021; Yu et al., 2020), and mixedwood forests of Ontario, Canada (White et al., 2021b). Wästlund et al. (2018) and Yu et al. (2020) both compared the SPL100 against the near-infrared channel of the Optech Titan linear-mode lidar system, acquired at 3800 m and 400 m above ground level, respectively. Both studies found similar accuracy between SPL100 and Optech Titan area-based predictions. For example, the relative root mean square error (RMSE) of area-based estimates of stem volume ranged between 16.81 – 21.23% for the SPL100 and 17.72

– 24.31% for the Optech Titan. Note that these two studies were undertaken in relatively small areas (1300 ha and 2000 ha) in managed boreal forests of northern Europe, which did not reflect large variations in forest productivity, disturbance, or landform types. White et al. (2021b) also recently produced SPL100 area-based estimates of forest attributes in a 15,000 ha temperate mixedwood forest in Ontario, Canada. When assessing the model estimation accuracy at the plot level, the relative error of the forest attribute estimates were generally larger than the ones reported by Wästlund et al. (2018) and Yu et al. (2020). However, White et al. (2021b) also found that the plot-level validation overestimated error and underestimated bias compared to independent validation data acquired at the forest stand level.

In 2018, the province of Ontario (107,520,421 ha including 56,094,000 ha of forests) has initiated a new EFI program based on SPL100 data acquisitions across Ontario's forest management units (Bilyk et al., 2021). This globally unique acquisition of SPL100 data represents one of the first widespread adoptions of the technology, and allows for analysis in terms of its use in prediction of forest attributes over much larger areas than has previously been attempted. In addition to the SPL100 data acquisition, calibration plots need to be established to generate forest attribute estimates following the ABA. Current forest resources inventory standards in Ontario were initiated in 2005 and are based on the interpretation of high-resolution digital airborne imagery (Bilyk et al., 2021). Systematic networks of calibration plots (1 plot per 8 km² or 5 km²) and permanent sample plots (1 plot per 400 km²) are used to link photo interpretation observations with field measurements (Bilyk et al., 2021; OMNR, 2013). The shift to an ALS-based EFI approach requires the adoption of ground plot sampling schemes guided by the structural information provided by the lidar data and compatible with the types of parametric or non-parametric models used to estimate the forest attributes (White et al., 2017a, 2013). Structurally guided sampling (SGS) approaches incorporating vegetation structure information (e.g. by stratification of ALS metrics) ensure that the full range of structural variability occurring across the inventoried area is considered, which is not necessarily the case with random or systematic sampling approaches (Gobakken et al., 2013; Grafström and Ringvall, 2013; Hawbaker et al., 2009; Junttila et al., 2013a; Maltamo et al., 2011; Papa et al., 2020). Acquiring

the SPL100 data prior to the establishment of plots allows for a highly structured statistical design for the plot network to ensure the plots cover the variation observed in the three-dimensional point cloud while minimizing the number of plots and thus the cost of measuring the forest in the field (Papa et al., 2020). Approaches to stratify three-dimensional point clouds information to allow plot placement and to ensure that the distribution of forest structures across the landscape are effectively sampled is therefore of key interest to forest managers as they aim to produce accurate forest inventory information more efficiently and at the lowest cost.

Previous research using the ABA with SPL100 data focused on the forest attribute accuracy and used sample plots established prior to the SPL100 acquisition, following either systematic sampling (Wästlund et al., 2018), systematic stratified sampling (Raty et al., 2021), or stratified random sampling based on linear-mode lidar metrics of canopy height and density (Yu et al., 2020, 2015), and principal component analysis of linear-mode lidar metrics (White et al., 2021b). Moreover, these studies were conducted in forest areas of small to moderate size (1300 – 160,000 ha). The objective of this chapter was to develop an ABA fully driven by the SPL100 data, from the sample plot selection to the generation of forest attributes, across a large boreal forest management unit (630,000 ha) located in Ontario, Canada. First, structural metrics derived from the SPL100 data were used to develop a SGS approach and ensure that field data was collected in the optimum number of plots without repetition across the forest estate. Then, an ABA was developed using the sample plots obtained from the SGS approach with the SPL100 data, to predict a number of forest inventory attributes including height, stem volume, biomass, diameter, and basal area. The accuracy of the models was assessed and compared to a previously developed EFI produced over the same area using low-density (< 1 point/m²) linear-mode lidar data (Woods et al., 2011).

3.2. Methods

3.2.1. Study area

This chapter focuses on the Romeo Malette Forest, which is described in section 2.1.1 of this dissertation.

3.2.2. Lidar metrics

The SPL100 data were processed using the LAStools software (Isenburg, 2017) and lidR R package (R Core Team, 2022; Roussel et al., 2020). The point clouds were first normalized to above ground height using the lasheight function of LAStools, which creates a triangular irregular network from ground returns to interpolate a ground surface. The tiles were processed with a 50 m buffer to avoid edge artifacts. A set of 20 lidar point cloud summary metrics were then generated at a 20 x 20 m resolution. Metrics included measures of central tendency (mean, median), dispersion (coefficient of variation, skewness, kurtosis) and percentiles of vegetation returns height above 1.3 m (5th, 10th, 20th, 30th, 40th, 60th, 70th, 80th, 90th, 95th and 99th percentiles). A measure of canopy cover was also derived by calculating the proportion of first returns above 2 m, 5 m, 10 m and 15 m thresholds.

3.2.3. Structurally guided sampling for plot placement

The availability of wall-to-wall SPL100 data over the entire Romeo Malette Forest allowed for an examination of the forest structural variation, as observed with the SPL100, across the entire management area prior to plot placement and establishment. An SGS approach is useful for both establishing a new network of plots within the spatial extent of lidar acquired coverages to support an ABA, as well as for modifying or augmenting existing plot networks to ensure that the plot locations adequately represent the full structural variation of the forest resource (Gobakken et al., 2013; Grafström and Ringvall, 2013; Hawbaker et al., 2009; Junttila et al., 2013a; Maltamo et al., 2011).

As discussed above, a large number of lidar point cloud metrics may be derived within the 20 x 20 m grid tessellation over an entire forest estate. Rather than developing a SGS approach using all of these lidar metrics, the well-established Principal Component Analysis (PCA) data reduction technique was employed, which summarizes a large number of correlated variables into a smaller number of uncorrelated variables referred to as principal components (Li et al., 2008; Stephens et al., 2012; White et al., 2013). Random sampling can then be performed within the feature space defined by the principal components to ensure a representative characterization of all forest structures occurring within the study area.

The PCA was performed with the rasterPCA function from the RStoolbox package (Leutner et al., 2019) over the 20 x 20 m cells of the 20 wall-to-wall lidar metrics. The first (PC1) and second (PC2) principal components captured 76.3% and 11.3% of the variance, respectively (Figure 3-1). Since the majority of the variance (87.6%) was captured in the first two principal components, the PC1 x PC2 feature space was used for stratification.

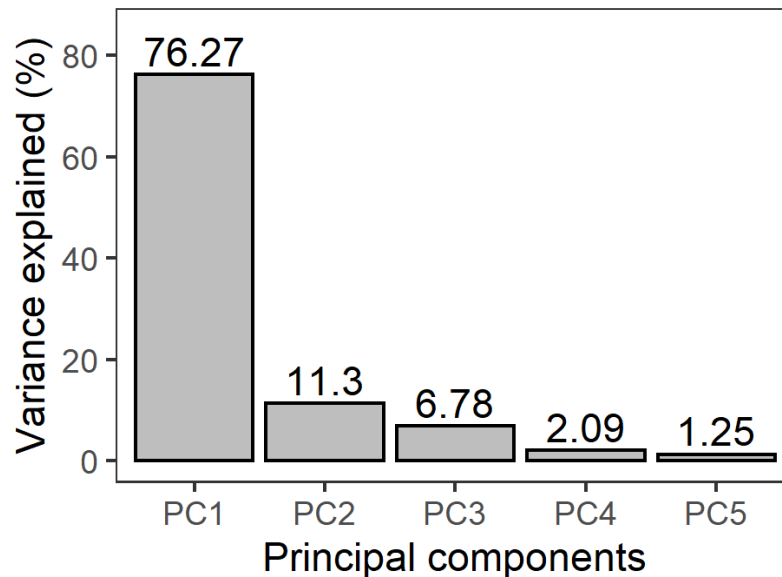


Figure 3-1 Variance explained by the first 5 principal components obtained from the PCA of the 20 lidar metrics.

The range of values of the PC1 and PC2 were divided into 10 and 5 equal intervals, respectively, forming a 10 x 5 matrix of strata. From the 50 available strata, 41 overlapped with the PC1 x PC2 feature space (Figure 3-4). Random sampling was then performed within each of the 41 strata to develop the plot network. An existing network of 183 plots was already established across four broad forest types (jack pine, black spruce, intolerant hardwoods and mixedwoods) in the Romeo Malette Forest from the previous EFI (Woods et al., 2011). If an existing plot occurred within a stratum it was flagged for re-measurement. To ensure representative plot placement across the entire forest estate, a minimum of 10 plots (either existing or established) were placed within each stratum with the exception of less frequent strata where a lower

number of plots was allowed. In order to avoid situations where a sampled cell within a given strata was surrounded by cells of another strata, the selection of cells that had surrounding cells (3 x 3 cells window) of the same stratum was also prioritized. This ensured that when locating these plots in the field using GPS, a small positional error would not result in sampling a boundary between two strata. Additionally, any sampled cell was required to be at least 200 meters away from an existing plot or previously sampled cell in the same stratum to avoid issues of spatial autocorrelation. With an objective of minimizing the costs and safety risks associated with the field data collection program, only the cells located between 30 m and 200 m of accessible roads and not restricted by land ownership were considered for plot establishment. Despite a potential impact of edge effects on forest structure and related attributes (Harper et al., 2005), this restriction on plot location had relatively minor implications in terms of the representativity of the sample of the range of forest structure occurring across the Romeo Malette Forest. The cells only located within the 30 – 200 m road buffer had a comparable convex hull in the PC1 and PC2 feature space than the wall-to-wall cells.

3.2.4. Forest inventory attributes

Based on the field data, a suite of plot-level forest attributes (basal area (BA), Lorey's height (LOR), quadratic mean diameter at breast height (QMDBH), stem density (DENS), whole stem volume (WVOL), merchantable stem volume (MVOL) and above-ground biomass (AGB)) were calculated or modelled from allometric equations as described in Table 3-1.

Table 3-1. Description of the modeled forest attributes

Forest attribute	Description	Calculation or modeling method	Unit
Basal area (BA)	Tree cross sectional area (approximated as a circle) at breast height (1.3 m)	$\frac{\pi}{4} \sum_i^n DBH_i^2 \times \frac{1}{A}$, where n is the number of stems and A the plot area in ha	m ² /ha
Lorey's height (LOR)	Average tree height weighted by basal area	$\frac{1}{n} \times \sum_i^n h_i \times BA_i$, where n is the number of stems and h is the tree height	m
Quadratic mean DBH (QMDBH)	Quadratic mean of DBH	$\sqrt{\frac{\sum_i^n DBH_i^2}{n}}$, where n is the number of stems	cm
Stem density (DENS)	Number of stems with DBH > 7.1 cm per ha	$\frac{n}{A}$, where n is the number of stems and A the plot area in ha	stems/ha
Whole stem volume (WVOL)	Total whole stem volume per hectare	Honer, (1983) and (Ung et al., 2013)	m ³ /ha
Merchantable stem volume (MVOL)	Total merchantable stem volume per hectare. Merchantable volume is defined as the stem volume between stump height (0.1 m) stem height at a bark diameter of 10 cm.	Honer, (1983) and (Ung et al., 2013)	m ³ /ha
Above-ground biomass (AGB)	Total tree biomass per hectare	Ter-Mikaelian & Korzukhin, (1997) and C.-H. Ung, Bernier, & Guo, (2008)	t C / ha

3.2.5. Random forest models and accuracy assessment

The plot-level forest attributes described in Table 3-1 were modeled from the SPL100 metrics using random forest (RF) regression models (Breiman, 2001) implemented in the R package randomForest (Liaw and Wiener, 2002). Only the lidar metrics having a pairwise correlation < 0.9 with the other metrics were used in the models; however, the 20th, 50th and 95th height percentiles were required to be included in the model in order to ensure that measures of the lowest, central, and highest parts of the distribution of return height were included. The accuracy of each RF model was assessed using a k-fold cross validation approach implemented with the R package caret (Kuhn et al., 2015). The training data were split into $k = 5$ folds of the same size ($n = 50$ observations), with each observation falling only into a single fold. At each iteration, one fold was held-out and a RF model was trained on the remaining 4 folds ($n = 200$ observations). After

5 iterations, the predictions of the 5 separated held-out folds were combined into a single set of predicted values and compared to their observed values to assess the model accuracy by calculating the coefficient of determination (R^2), root mean square error (RMSE) and bias.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}$$

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

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

where y_i is the observed value of the response variable, \hat{y}_i is the predicted value (from any of the 5 folds) and n is the total number of observations. The RMSE and bias were also calculated relative to the mean of all observed values (RMSE%; bias%). The 5-fold cross validation approach was also used to select an optimal value for the RF parameter *mtry* that defines the number of predictor variables randomly sampled at each node of the RF trees. At each fold, RF models were trained for *mtry* values ranging between 1 and the number of predictor variables. The *mtry* value that minimized the RMSE was retained in the final RF models and the number of trees for each model was set to 500.

The contribution of each predictor variable to the models was assessed using the RF measure of variable importance obtained from the randomForest package, calculated as the percentage of increase in mean square error when a predictor variable is permuted while others remain unchanged (Breiman, 2001). Variable importance was calculated globally, as an average variable importance across all trees in the RF model, and locally, where the importance of predictor variables is assessed for each observation. For an observation i , the local variable importance is calculated as the average importance across all the trees where i is in the out-of-bag sample. It is important to note that Strobl et al. (2007) have demonstrated that RF measures of variable importance can be severely biased in classification models when importance is assessed with the Gini index. If variable importance is assessed by permutation of the predictor variables and scaled by the standard deviation of the importance measures, which is the case herein, Strobl et al.

(2007) observed a less pronounced, yet existing bias. Variable importance measures obtained with the RF regression models developed and reported herein should therefore be interpreted with caution.

Although one global RF model was trained for each forest attribute, the RMSE, bias and RF variable importance were also examined by groups based on species composition. The ground plots were classified either as jack pine, black spruce, coniferous, deciduous, or mixedwoods following the criteria described in Table 3-2. This classification is similar to the one used by Woods et al. (2011) in the previous EFI performed at the Romeo Malette Forest. The distribution of forest attribute values for each species group is shown in Figure 3-2.

Table 3-2. Description of species groups used to classify the field plots.

Species group	Description	Number of plots
Jack pine	$\geq 50\%$ of jack pine and $\geq 70\%$ of other coniferous species	33
Black spruce	$\geq 50\%$ of black spruce and $\geq 70\%$ of other coniferous species	38
Coniferous	$\geq 70\%$ of coniferous species but $\leq 50\%$ of jack pine or black spruce	43
Deciduous	$\geq 70\%$ of deciduous species	43
Mixedwoods	Coniferous or deciduous species $\geq 30\%$ and $\leq 70\%$	93

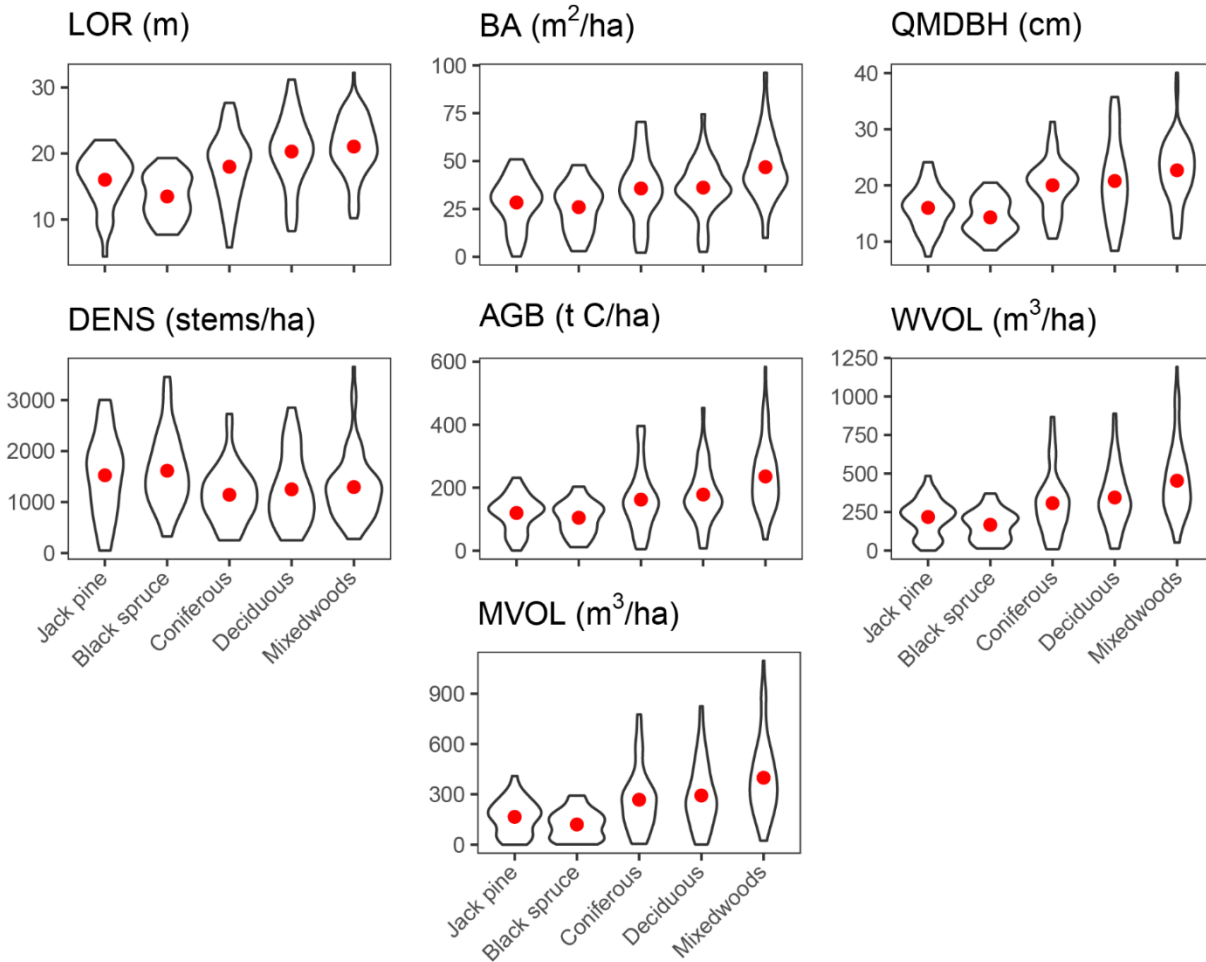


Figure 3-2. Distribution of field-measured forest attributes per species group. Red dots indicate the mean of the distribution. See Table 3-2 for the description of species groups.

3.2.6. Influence of sample size on the accuracy of the models

As described above, the models were developed using the 250 plots that were selected across the Romeo Malette Forest using SGS. The influence of decreasing the number of plots on the models accuracy while maintaining a sample of plots representative of the 41 unique strata used in the SGS was also examined. The scenario with the minimal number of plots consisted in randomly sampling one plot in each stratum from the 250 available plots. The number of plots was then gradually increased by randomly sampling additional plots in each stratum, until the original sample of 250 plots was reached. Increasing the number of plots was not always possible in strata characterizing uncommon forest types because of the lack of additional plots available in the full sample of 250 plots. This process resulted in 15 distinct sample subsets

with a sample size ranging between 41 and 250 plots. The distribution of the number of plots per strata as the sample size increased is illustrated in Figure 3-3. For each subset of the 250 plots, RF models were trained and their accuracy were assessed with the R^2 , RMSE% and bias% using 5-fold cross-validation as described previously.

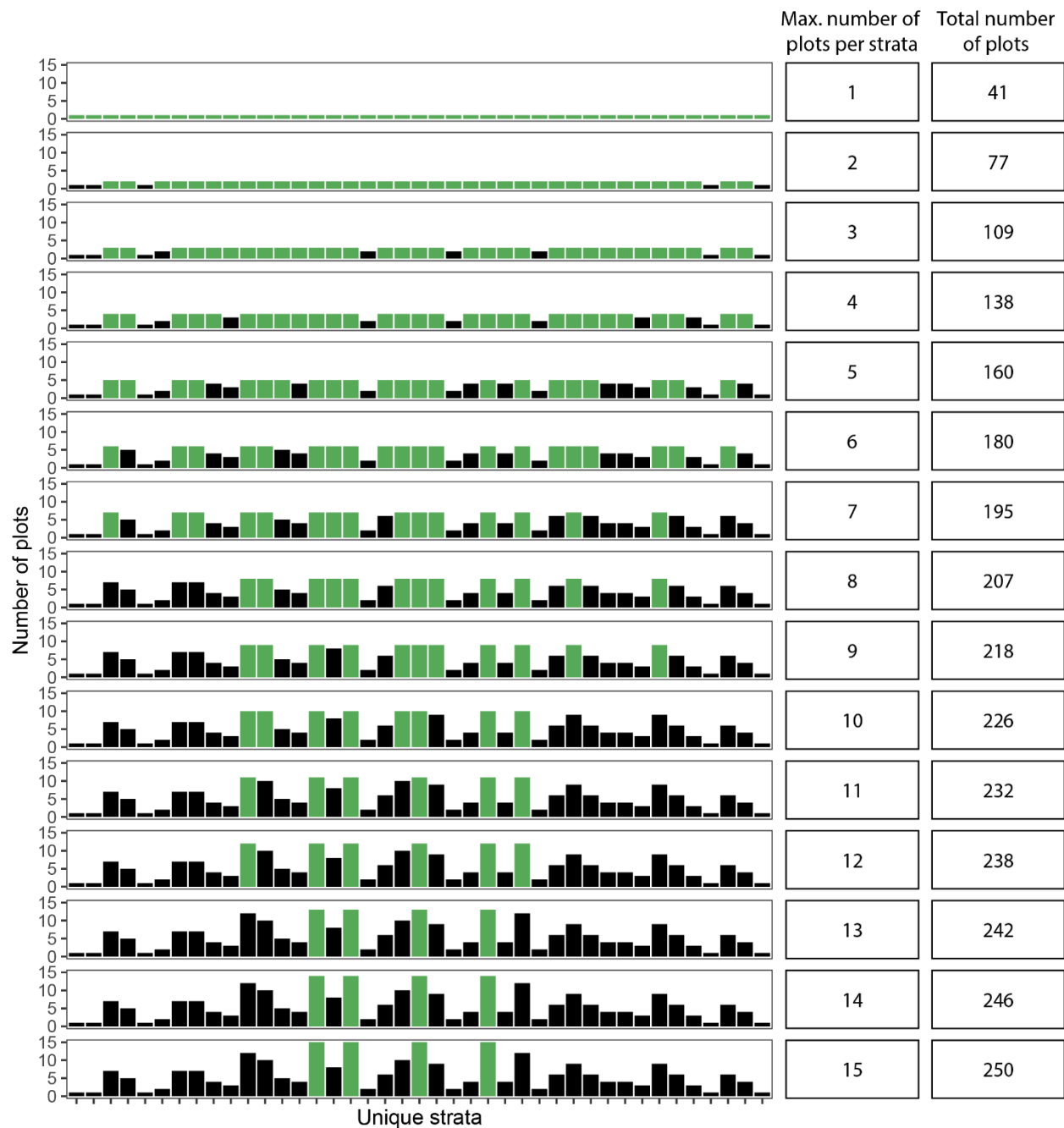


Figure 3-3. Distribution of the number of plots per strata for the 15 subsets of the 250 plots used to examine the influence of the number of plots on the accuracy of the models. The graph at the top represents the smallest sample, with 1 plot per strata while the graph at the bottom represents the full sample of 250 plots with a maximum of 15 plots per strata. Bars are colored in green if an additional plot was added to a stratum compared to the previous smaller plot subset.

3.3. Results

3.3.1. Structurally guided sampling

Figure 3-4 provides a visualization of the PC1 and PC2 feature space for the entire forested cells of the Romeo Malette Forest area overlaid with the re-measured existing and new sample plots. The figure shows that the re-measured existing plots covered a smaller area of the feature space compared to the new plots. The figure also shows that some types of forests have no existing sample plots associated with them and clearly demonstrates the benefit of the SGS approach for identifying which forest types are poorly sampled using the existing plot network.

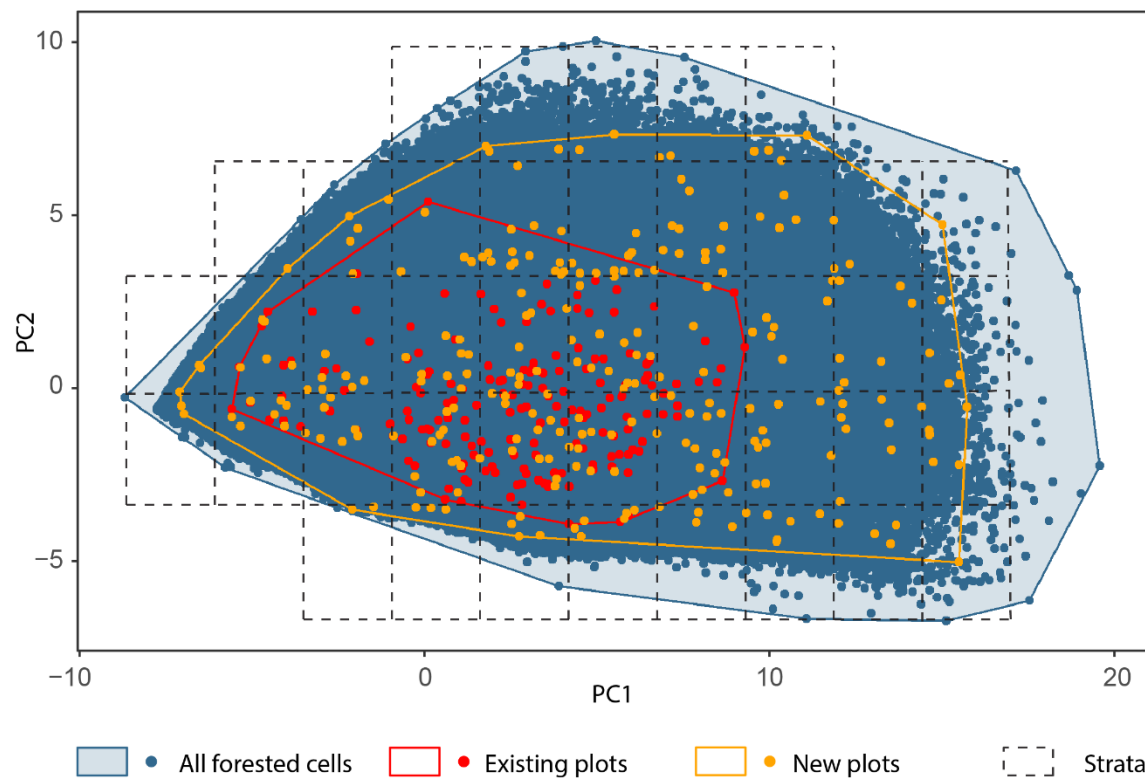


Figure 3-4. Feature space formed by the first two principal components PC1 and PC2. The horizontal and vertical dashed lines indicate the 41 strata that were used to stratify the feature space into equal intervals. Convex hulls are indicated by the solid colored lines.

The SGS process resulted in decreasing the number of plots in oversampled strata covered by the existing plot network and increasing the number of plots in poorly represented strata. Stands in the initiation and old-growth stage of their development were poorly represented by the existing plot network whereas the

SGS resulted in new plots established in these forest types (Table 3-3). The stem exclusion stage in jack pine, black spruce, coniferous and mixedwoods stands was relatively well sampled by the existing sample plots (47.2% of plots) but not at all in deciduous stands. The SGS resulted in 17.4% of the new plots representing deciduous stands in the stem exclusion stage (Table 3-3). Similarly, mixedwoods stands in the understory reinitiation stage were not sampled by the existing plots whereas 31 new plots (19.3%) were established with the SGS (Table 3-3).

Table 3-3. Number of re-measured and new sample plots grouped by stand development stage and species group. Only conditions where at least 5 re-measured or new plots were observed are reported in the table. See Table 3-2 for the description of species groups.

Stand development stage	Species group	Number (percentage) of re-measured existing sample plots (n = 89)	Number (percentage) of additional sample plots (n = 161)
Stand initiation	Jack pine	0	5 (3.1%)
	Coniferous	0	6 (3.7%)
Stem exclusion	Jack pine	16 (18.0%)	5 (3.1%)
	Black spruce	19 (21.3%)	5 (3.1%)
	Coniferous	7 (7.9%)	6 (3.7%)
	Deciduous	0	28 (17.4%)
	Mixedwoods	11 (12.4%)	26 (16.1%)
Understory reinitiation	Black spruce	6 (6.7%)	0
	Coniferous	5 (5.6%)	7 (4.3%)
	Mixedwoods	0	31 (19.3%)
Old-growth	Coniferous	5 (5.6%)	6 (3.7%)
	Deciduous	0	8 (5.0%)
	Mixedwoods	0	15 (9.3%)

3.3.2. Random forest models accuracy and variable importance

Wall-to-wall maps of select forest attributes estimated across the Romeo Malette Forest are presented in Figure 3-5.

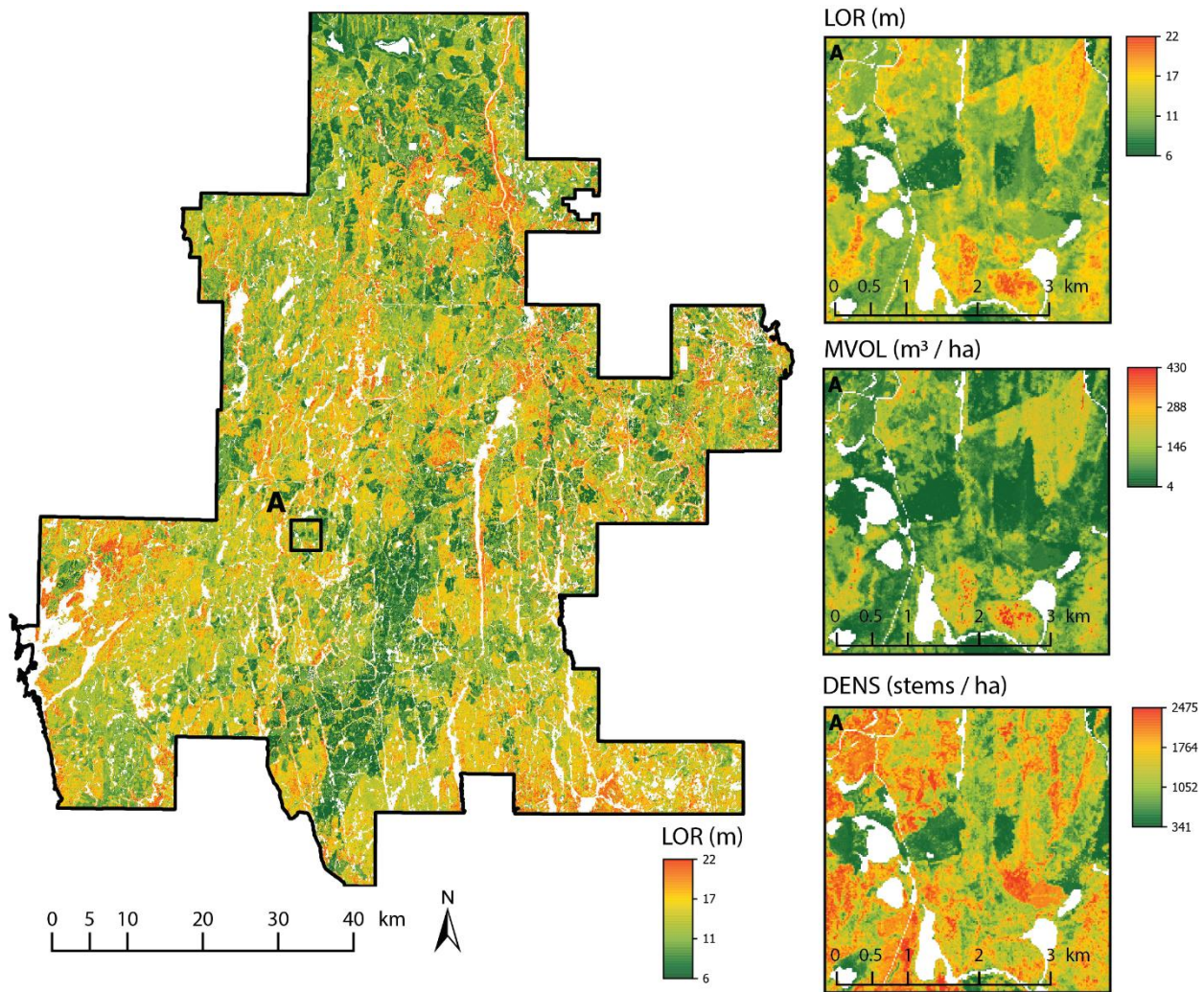


Figure 3-5. Wall-to-wall and inset maps of predicted Lorey's height (LOR), merchantable volume (MVOL) and stem density (DENS) at the Romeo Malette Forest. Inset maps have the extent of the box labelled as A is the wall-to-wall map.

The forest attribute predicted with the highest accuracy was LOR with $R^2 = 0.91$ and $RMSE\% = 8.5\%$. The attributes BA, QMDBH, WVOL, MVOL and AGB had R^2 values ranging between 0.79 and 0.9 and $RMSE\%$ ranging between 13.97% and 23.79%. DENS was predicted with markedly lower accuracy ($R^2 = 0.63$; $RMSE\% = 30.82\%$). Bias% was $< 1\%$ for all forest attributes. Scatterplots comparing the predicted and observed values of each response variable at the plot level are shown in Figure 3-6. The $RMSE\%$ was the lowest in black spruce-dominated plots for LOR (6.8%), BA (19.37%) and QMDBH (9.69%) but in plots with mixed conifers for MVOL (19.02%), WVOL (21.32%) and AGB (17.22%). DENS was predicted with the lowest $RMSE\%$ (24.77%) in mixedwoods plots. Jack pine-dominated plots

had the highest RMSE% compared to the other species groups for LOR (10.65%), BA (22.29%), QMDBH (17.21%), DENS (39.31%) and WVOL (24.28%). Predictions at the deciduous plots exhibited higher absolute bias% than for the other species groups for BA (9.01%), WVOL (6.80%), MVOL (7.10%) and AGB (7.99%). The accuracy of the models grouped by species composition is also reported in Table 3-4 .

Table 3-4. Absolute and relative RMSE and bias for predictions of each forest inventory attribute of interest, stratified by forest type. See Table 3-2 for the description of species group.

	RMSE; Bias (RMSE%; Bias%)				
	Jack pine (n = 33)	Black spruce (n = 38)	Coniferous (n = 43)	Deciduous (n = 43)	Mixedwoods (n = 93)
LOR (m)	1.71; -0.07 (10.65%;-0.42%)	0.92; -0.18 (6.8%;-1.35%)	1.68; 0.09 (9.36%; 0.5%)	1.92; -0.35 (9.47%;-1.72%)	1.52; 0.26 (7.22%;1.21%)
BA (m ² /ha)	6.33;-0.55 (22.29%-1.92%)	5.03; -1.34 (19.37%; 5.18%)	6.92; -1.36 (19.37%;-3.8%)	7.89; 3.26 (21.81%;9.01%)	8.48; -0.73 (18.12%;-1.55%)
QMDBH (cm)	2.75; 1.00 (17.21%;6.26%)	1.39; 0.19 (9.69%;1.32%)	2.69; -0.20 (13.45%;0.09%)	2.87; -0.06 (13.82%;-0.28%)	3.13;-0.36 (13.8%;-1.59%)
DENS (stems/ha)	599.63; -92.71 (39.31%; 6.08%)	452.84; -89.01 (28.05%;-5.51%)	343.53; 54.17 (30.05%;4.74%)	440.12; 79.68 (35.22%;6.38%)	320.68; 40.68 (24.77%;3.14%)
WVOL (m ³ /ha)	52.96; -10.10 (24.28%; 4.63%)	32.20; -4.21 (19.25%;-2.52%)	58.43; -9.5 (19.02%;-3.09%)	74.34; 23.42 (21.58%;6.80%)	91.8; -0.17 (20.25%;-0.04%)
MVOL (m ³ /ha)	38.67;-5.34 (23.42%;-3.23%)	28.04; -5.09 (23.39%;-4.24%)	57.01; -8.71 (21.32%;-3.26%)	69.28; 20.78 (23.68%;7.1%)	88.03; -0.27 (22.09%;-0.07%)
AGB (t C/ha)	23.52; -0.01 (19.62%; 0.01%)	21.11; -4.87 (20.14%;(4.65%)	27.94; 0.75 (17.22%;0.46%)	39.43; 14.25 (22.11%;7.99%)	54.75; -4.89 (23.22%;-2.07%)

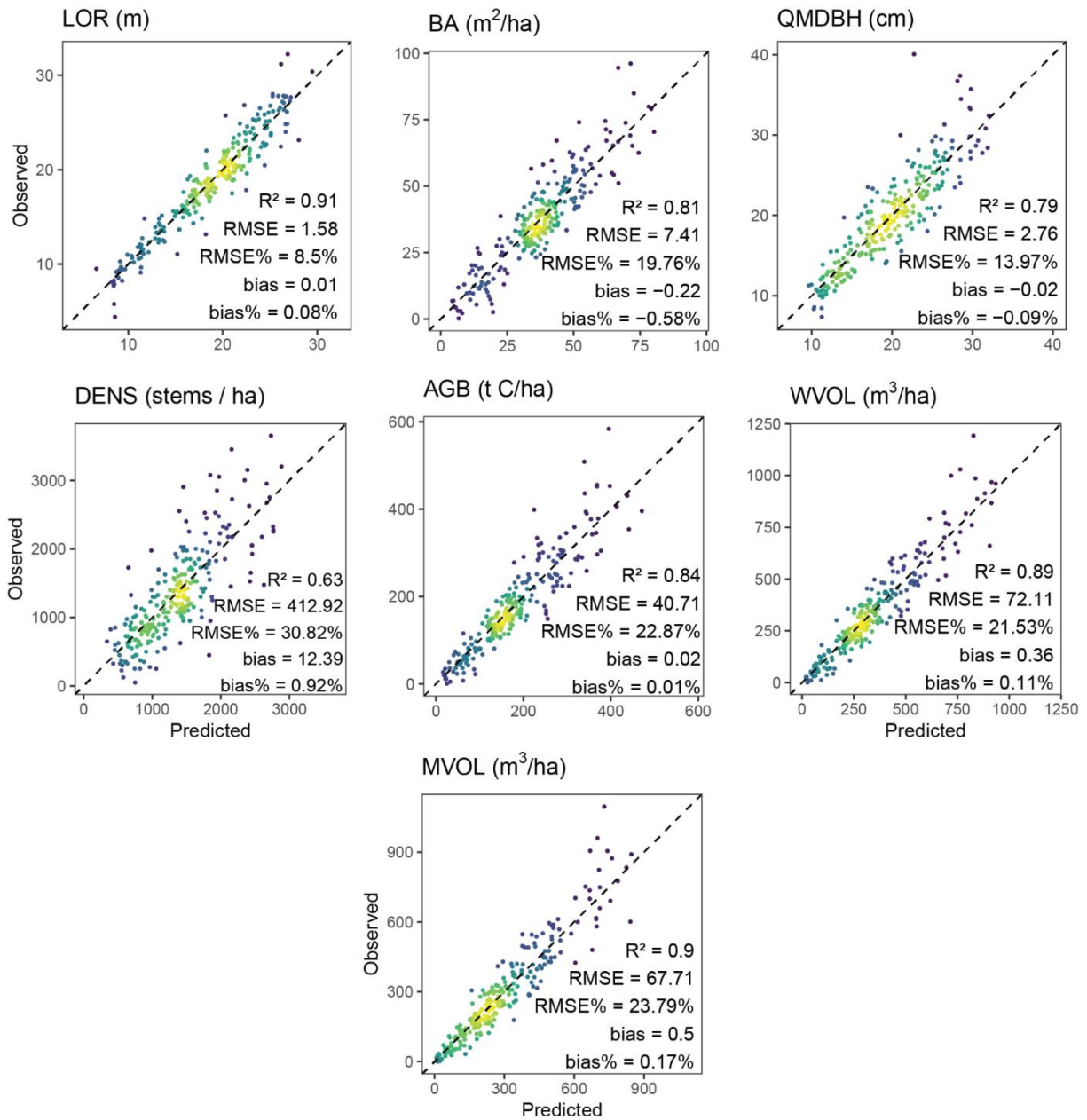


Figure 3-6. Scatter plots and accuracy metrics comparing the forest attributes values predicted by the RF models and to their observed values.

Analysis of the variable importance of the RF models shows that the 80th percentile (p80) and 95th percentile (p95) of height returns were among the 5 most important variables for each of the seven modeled forest attributes (Figure 3-7). Metrics representing canopy cover (cov) were also more important when height thresholds of 10 m or 15 m were used compared to 2 m. The coefficient of variation (cv) was the third most

important variable in two of the models, while skewness and kurtosis were never among the 5 most important predictor variables of the RF models.

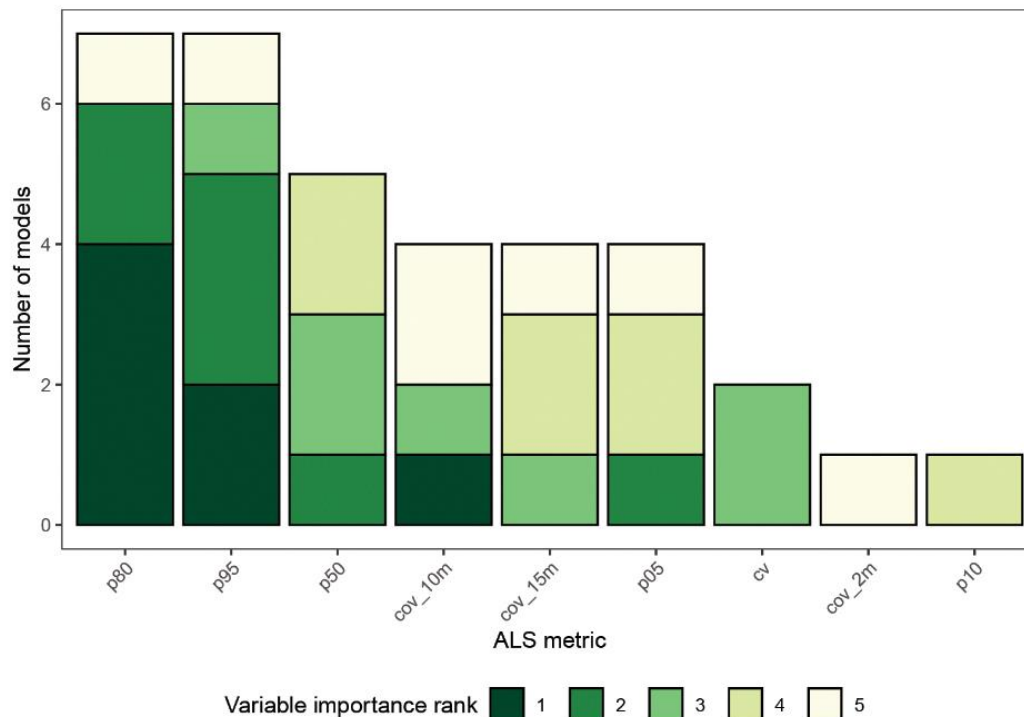


Figure 3-7. Summary of the SPL100 metrics importance rank combined from all the models. Metrics that were not at least once the 5th most important variable in any of the model are not reported. pxx = xx percentiles of returns height; cov_xx = percentage of first returns above xx m; cv = coefficient of variation (standard deviation / average of returns height).

The predictor variable importance analyzed by species group from the *randomForest* local importance is provided in Appendix A.

Figure 3-8 shows the correlation between predicted forest attributes and the SPL100 metrics included in the RF models. Higher height percentiles (p80, p95) had a $r > 0.82$ for all modelled forest attributes except DENS ($|r| = 0.24\text{--}0.29$). The linear relationship between DENS and height percentile metrics was stronger for the lower percentiles (p05, p10, p20, p30), with $r = 0.3\text{--}0.5$. The cv and kur were generally the most uncorrelated metrics with the modelled forest attributes ($|r| \leq 0.2$), except for QMDBH and DENS. In addition, metrics of canopy cover tended to have stronger correlation with modelled forest attributes when

calculated above height thresholds of 10 m and 15 m (cov_10m, cov_15m) compared to a height threshold of 2 m (cov_2m).

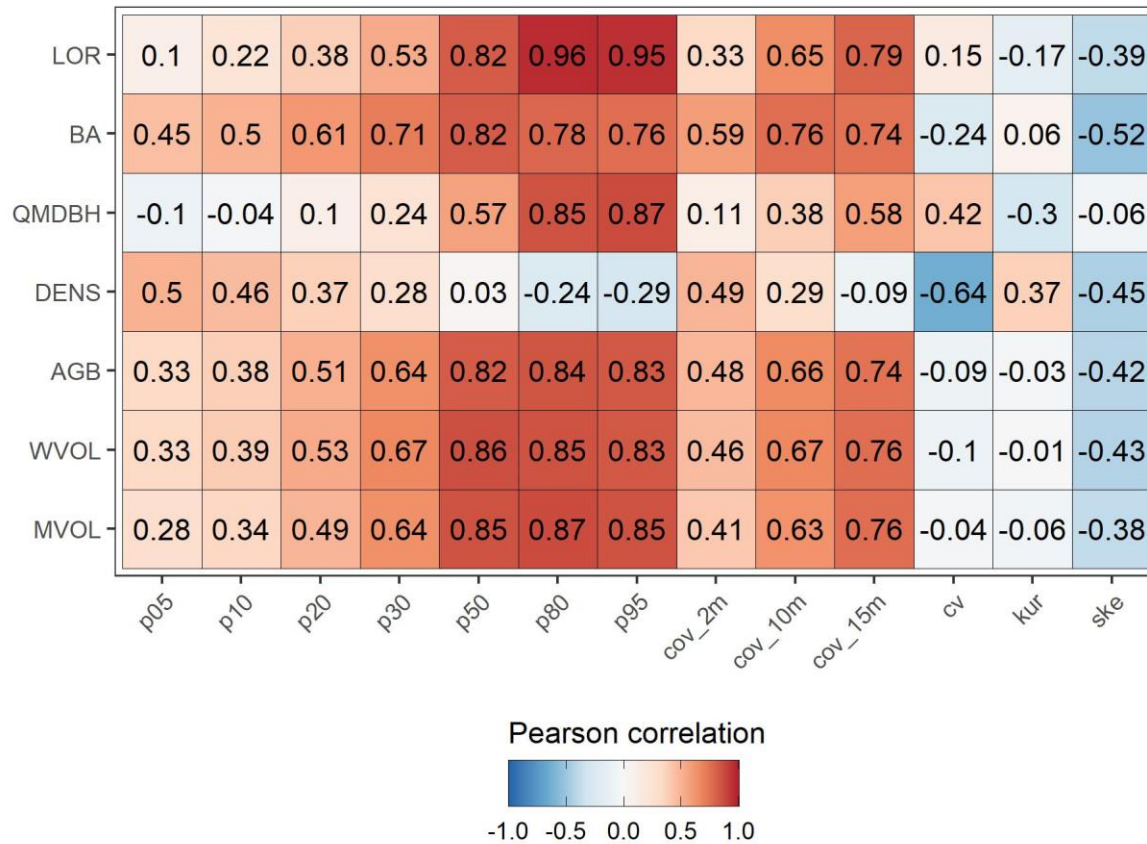


Figure 3-8. Pearson correlation values between the forest attributes values predicted by the RF models and the SPL100 metrics used as predictor variables.

3.3.3. Influence of sample size on models accuracy

Figure 3-9 shows the influence of the number plots on the R^2 , RMSE% and bias% of the RF models for each response variable. As the number of plots increased, the R^2 tended to increase, the RMSE% tended to decrease and the bias% tended to get closer to 0. DENS was the variable for which the accuracy improved the most as the number of plots increased, with a R^2 increase of 12%, RMSE% decrease of 20% and absolute bias% decrease of 42% between the smaller number of plots (1 plot per strata) and the full sample of 250

plots. For the other variables, the improvement in accuracy metrics was not as markedly significant. However, the stability of the model accuracy tended to improve as the number of plots increased, which indicate more robust models. Overall, these results indicate that the number of plots selected using the SGS could have been reduced up to about 160 plots (maximum of 5 plots per strata) without a markedly difference of accuracy compared to the models using the full sample of 250 plots.

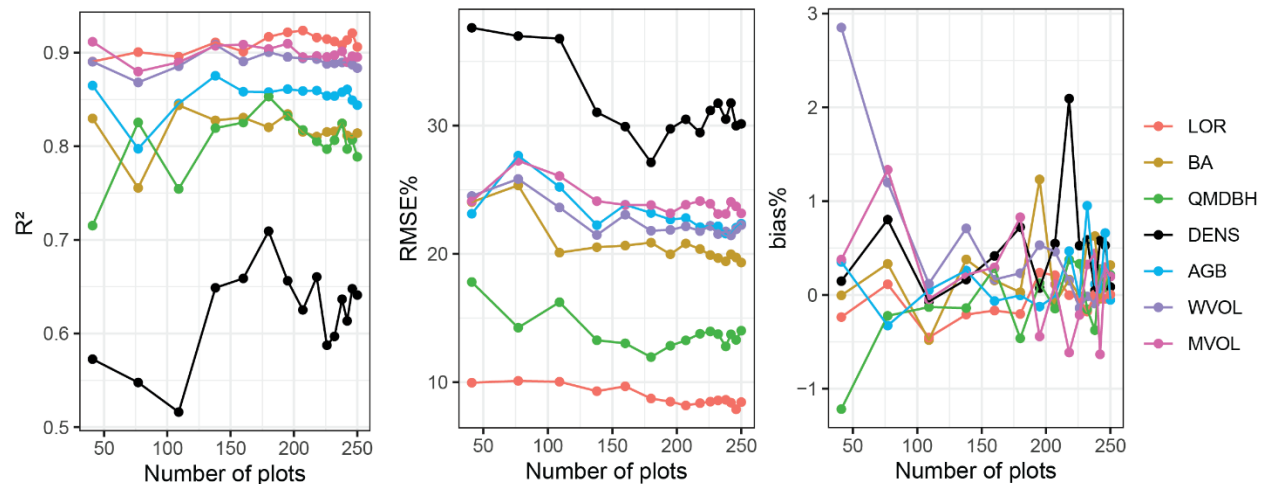


Figure 3-9. Influence of the number of plots on the accuracy of the random forest models of forest attributes. Each point represents an addition of 1 plot per stratum used in the structurally guided sampling as long as additional plots are available in each stratum of the full sample of 250 plots. The distribution of the number of plots per strata is provided in Figure 3-3.

3.4. Discussion

In this chapter, an EFI approach driven by a wall-to-wall SPL100 acquisition across a 630,000 ha forest estate was presented. A SGS approach was used to establish a network of ground plots that was representative of the full range of forest structural variability captured by the SPL100. A set of already established sample plots were integrated into the SGS, thereby reducing the number of plots in oversampled structural types and establishing new plots in undersampled structural types. A set of 250 sample plots were selected with the SGS and were used to generate forest attribute estimates using an ABA.

3.4.1. Accuracy of models compared to the previous EFI at the Romeo Malette Forest

The accuracy of the forest attribute models presented herein are very similar to the accuracy of existing area-based models developed at the Romeo Malette Forest using low density linear-mode lidar data (0.46

points/m²) acquired in 2004 and 2005 (Woods et al., 2011). The RMSE% of the previous area-based models reported in Woods et al. (2011) ranged from 5 to 36% for all forest types combined. Figure 3-10 indicates that the area-based models developed using the SPL100 data had a similar range in RMSE% (8–30%) and that the RMSE% for individual forest attribute models were of similar magnitude.

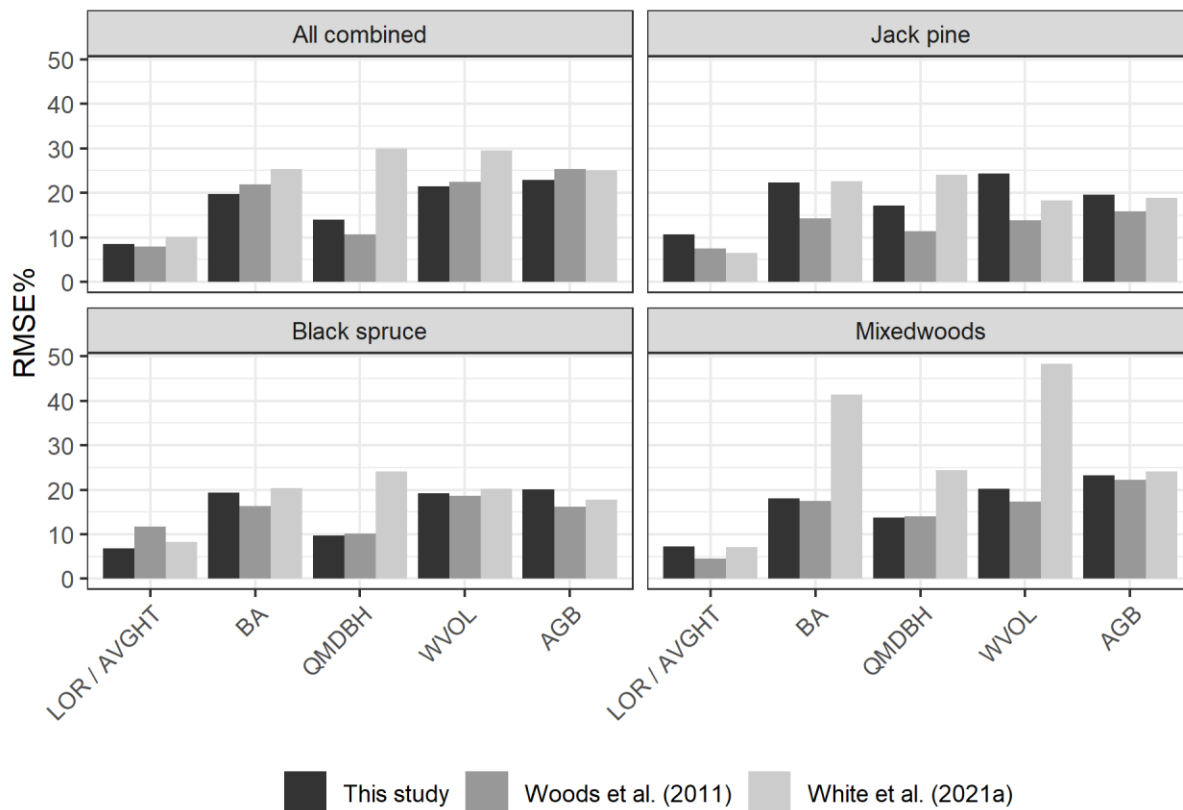


Figure 3-10. Relative RMSE (RMSE%) by species group for the area-based models of forest attributes generated in this study compared to the ones reported in White et al. (2021b) and Woods et al. (2011). Lorey's height (LOR) modelled in this study and White et al. (2021b) is compared to average height (AVGHT) in Woods et al. (2011). Merchantable stem volume (MVOL) and stem density (DENS) in this study were not calculated with the same method as the one used in Woods et al. (2011) and White et al. (2021b) and RMSE% for these attributes are therefore not reported in this figure.

As expected, the best performing models were for height-related attributes. For global models (all forest types), the previous inventory had an RMSE of 1.0 and 1.2 m for top and average height, respectively, whereas the RMSE for LOR estimated in this study was 1.6 m. In the case of WVOL the previous EFI had a RMSE of 52.9 m³/ha (RMSE% = 22.5%) whereas in this study WVOL had a RMSE of 72.1 m³/ha (RMSE% = 21.5%). In both studies, DENS was the poorest performing model with an RMSE of 489

stems/ha (RMSE% = 36.5%) in the previous study and 412 stems/ha (RMSE% = 30.8%) reported herein. Woods et al. (2011) also observed relatively small variation in RMSE% when comparing model predictions by species group. The post-hoc analysis of model predictions performed in this chapter likewise showed that similar RMSE% were observed across the species groups.

In the models developed by Woods et al. (2011), metrics of lidar returns height such as p90 and mean height were generally the most important variables for most forest types and forest attributes. In some models, lower height percentiles (p20, p30, p60, p70) and the cumulative proportion of returns at the 60th and 70th percentiles of return heights were also used. In this study, SPL100 metrics of percentile heights were also key variables for all forest attributes and species groups, particularly for LOR, QMDBH, and BA in coniferous stands. For the prediction of DENS, AGB, WVOL and MVOL however, the relative importance of the SPL100 metrics was more distributed between height-based, cover-based and height variance-based metrics. For DENS predictions, the lowest height percentiles p05 and p10 were more important in coniferous plots compared to deciduous and mixedwoods plots. This is likely due to the conical shape of these crowns with relatively more biomass lower in the stand due to the inherent shape of the trees (Coops et al., 2004). This contrasts with deciduous stands which have much more of the branch and leaf biomass higher in the canopy

3.4.2. Comparison to other area-based estimates with the SPL100

The accuracy of the forest attributes observed in this chapter contrast somewhat with those of White et al. (2021b) with SPL100 data in a 15,000 ha forest dominated by temperate mixedwoods (Petawawa Research Forest), located about 400 km south east of the Romeo Malette Forest. In their study, the authors also used RF to model forest attributes and assessed the model performance at the plot-level ($n = 269$) using the RF out-of-bag (OOB) observations (Breiman, 2001) and at the stand-level using 27 intensively sampled forest stands that were not used in model development. Model assessment was done overall and for 12 different forest types, which included plantations and uniform shelterwood systems and represented a more complex range of forest structures and management histories, including plantations, than the predominantly natural

boreal forest condition at the Romeo Malette Forest. The overall stand-level RMSE% of the models reported in White et al. (2021b) were consistently lower than the reported plot-level OOB RMSE ($RMSE_{stand}\% = 7.2\%–16.3\%$; $RMSE_{OOB}\% = 9.4\%–33.8\%$). All forest types combined, plot-level RMSE% reported herein and in White et al. (2021b) were relatively similar for LOR, BA, and AGB, with RMSE% of 10.2%, 25.4% and 25.1% reported by White et al. (2021b) compared to RMSE% of 8.5%, 19.8% and 22.9% in this study (Figure 3-10). Larger differences in plot-level RMSE% were observed for QMDBH and WVOL, with RMSE% of 29.9% and 29.6% in White et al. (2021b) compared to 14.0% and 21.5% herein. When comparing model accuracy by forest types, the most notable differences in plot-level RMSE% were observed in the mixedwoods plots for BA ($RMSE\% = 41.4\%$ in White et al. (2021b); $RMSE\% = 18.1\%$ herein) and WVOL ($RMSE\% = 48.3\%$ in White et al. (2021b); $RMSE\% = 21.5\%$ herein). Differences between the results presented herein and reported by White et al. (2021b) could be explained by notable differences in the number and complexity of the different forest types present in each of the study areas, particularly for mixedwood species. The range of observed values of the forest attributes, as well as differences in management interventions, study area extent, and ground sampling intensity could also explain these differences. At the Romeo Malette Forest, 250 sample plots for an area of 630,000 ha were measured, whereas White et al. (2021b) measured 269 sample plots for a 15,000 ha forest area. Other differences in outcomes may result from definitions and protocols used in the field measurement protocols and the compilation of ground plot data. For example, White et al. (2021b) considered all trees with $DBH \geq 2.5$ cm to calculate forest inventory attributes (except for merchantable attributes where $DBH \geq 9.1$ cm was used) whereas this study considered all trees with $DBH \geq 7.1$ cm.

Area-based forest attributes estimates were also produced with SPL100 data in boreal and hemi-boreal forests of southern Sweden (Wästlund et al., 2018) and southern Finland (Räty et al., 2021; Yu et al., 2020) mostly dominated by pine and spruce species. The findings of these studies were similar to the ones reported herein, with RMSE of 1.1–1.4 m (6.1%–6.7%) for height, 44.7–48.8 m³/ha (16.8%–21.2%) for WVOL and 20.9–29.6 t C / ha (15.8%–21.3%) for AGB. Räty et al. (2021) also integrated lidar intensity metrics in their

models and found that they did not improve the estimates of species-specific volume as much for SPL100 as they did for linear-mode lidar instruments.

Overall, the results presented in this chapter are consistent with results from other boreal forest contexts (Räty et al., 2021; Wästlund et al., 2018; White et al., 2021b; Yu et al., 2020) and indicate that the area-based inventory attribute predictions derived from SPL100 data are similar to the accuracy of attribute predictions generated using linear-mode lidar systems. As expected, the accuracy of the ABA predictions with the SPL100 does not improve significantly beyond what was obtained with a linear-mode lidar sensor, despite the increase in point density associated with the SPL100 for large areas (Lim et al., 2008; Treitz et al., 2012). However, applications that require high point cloud density, such as the extraction of individual tree attributes (Breidenbach et al., 2010a) or sub-canopy forest structure (Jarron et al., 2020; Joyce et al., 2019), could potentially be performed over large areas with widespread SPL100 acquisitions. Such applications with SPL100 data have yet to be examined in the literature. Moreover, with the exception of White et al. (2021b), the use of SPL100 to produce area-based forest inventory attributes predictions has focused primarily on boreal forest types in Canada and northern Europe. Additional insights could be gained from the application of EFI based on SPL100 data in broader ranges of forest types.

3.4.3. Benefits of a structurally-guided sampling approach

The development of EFI programs globally has demonstrated how key forest attributes can be predicted accurately across the landscape using co-located ground-based plot measurements and lidar metrics. However, the successful deployment of an EFI relies on the plot data accurately representing the structural variation that occurs within the forest management area of interest (Maltamo et al., 2011; White et al., 2013). Without representative sample plots, EFI predictions of forest attributes can be biased towards certain forest types and result in poorer prediction accuracy in the under-sampled forest types. For example, this chapter shows that the previously established plot network at the Romeo Malette Forest was biased towards productive coniferous stands and lacked representation of deciduous forest types, as well as stands in the initiation and old-growth stage of their development (Table 3-3). With the increasing availability of

wall-to-wall lidar data, forest managers have a wealth of information to guide inventory plot placement. Establishing plots and measuring tree inventory information is costly in terms of labor, travel, and safety. Minimizing the number of plots required to establish an accurate forest inventory is a key management objective (Papa et al., 2020). For example, Maltamo et al. (2011) demonstrated that a reduction of the number of ground plots did not necessarily markedly decrease the accuracy of forest attributes as long as the plot network was representative of the structural variability of the study area. The SGS approach presented here provides the opportunity to examine what stand structures exist and, through stratification, where plots should be located to capture that variation in order to develop an accurate EFI.

One key consideration associated with the SGS approach is that the selection of plots and their representative placement across the landscape is based only on the structural information provided by the lidar data acquisition. If there are other drivers that need to be considered in plot placement, such as species composition, these would not be used in the stratification for plot placement as lidar data is relatively insensitive to spectral changes associated with different species mixes. As a result, if species composition is a critical stratification layer to consider in the positioning of plots, then the method would need to be adapted to allow additional stratification layers to be added which represent for example species groupings.

The stratified random sampling approach adopted for this study, based on a stratification of the two principal components of a large set of ALS metrics, is only one of the many approaches that can be used in SGS. Other studies such as Hawbaker et al. (2009) and Maltamo et al. (2011), performed stratified random sampling using two ALS metrics for stratification. Rule-based stratification can also be achieved using thresholds on statistical metrics derived from the ALS point cloud such as L-moment ratios (Adnan et al., 2021; Valbuena et al., 2017). Sampling can also be performed on a large set of variables with methods such as conditional Latin hypercube sampling (Minasny and McBratney, 2006), which stratifies the multivariate distributions into equiprobable intervals, and balanced sampling (Grafström and Ringvall, 2013). From the range of approaches available, it is important to consider the suitability of the sampling scheme with the types of model approaches (e.g. regression or imputation; parametric or non-parametric) used in the ABA

(White et al., 2017a). For example, non-parametric models such as RF do not extrapolate beyond the range of observation values and thus require a representative plot sample. There is a need for a set of frameworks and tools targeted to forest managers that integrate multiple SGS methods and other factors such as land accessibility (e.g., roads, topography, water features) as well as existing sample plots to derive an optimum sample plot network. Costs associated with establishing and measuring sample plots are significant given the often large spatial extent of forest to sample and issues of accessibility in remote locations. As we see investment in lidar by companies and governments, there is also the need for a representative ground plot network with which to efficiently model forest inventory attributes of interest. Optimizing this network with respect to both the forest structures present and cost considerations is therefore critical to ensure robust information outcomes for forest managers and operational feasibility for implementation.

3.4.4. Benefits and challenges of SPL100 for enhanced forest inventory generation over large forested areas

One of the key advantages often discussed in relation to the SPL100 lidar data is the ability of the aircraft to fly higher, and faster, to acquire the commensurate (or greater) point density of conventional linear-mode lidar systems. This advantage results in efficient data acquisitions over larger areas with a single instrument and consistent acquisition parameters. The results presented here demonstrate that the development of an EFI using these SPL100 data produces modelled forest attributes that achieve similar accuracies to those derived using a linear-mode lidar system, indicating that the acquisition advantages afforded by the SPL100 data, and the associated differences in the laser and the instrumentation, do not adversely impact the quality of the derived forest inventory information. The SPL100 data were acquired at a nominal flight altitude of 3800 m above ground level and swath width of approximately 2000 m. As a comparison, the linear-mode lidar data used in the previous EFI was flown at 2740 m above ground level and had a swath width of approximately 1000 m. The capacity for large area acquisitions is important for operational implementation and has been well documented in other studies (Wästlund et al., 2018; Yu et al., 2020).

The costs of lidar acquisition are not driven exclusively by the coverage efficiency but also by the operational and logistical challenges associated with the data acquisition. While established vertical survey control networks that typically support municipal, provincial, hydroelectric and mining infrastructures can be leveraged for aerial lidar surveys, many remote portions of the boreal forest require additional ground control points to be established to support the specified point cloud vertical accuracy (Ministry of Natural Resources and Forestry, 2016). Establishing ground control is a key driver of cost, especially when road access is limited and ATV, boat, helicopter, or aircraft is required. Limited airport support is also an important consideration in the remote regions of the boreal forest. Moreover, high-level cloud and smoke from wildfires have resulted in a 30-50% reduction of suitable acquisition windows for the wall-to-wall SPL100 acquisition in Ontario, Canada (Bilyk et al., 2021). Adjusting acquisition altitude to fly under clouds and smoke could be a solution to avoid a reduction of suitable acquisition windows. However, differences in ALS acquisition parameters across multiple areas could have an effect on the characteristics of the ALS metrics derived from the point clouds. This is a particularly important consideration if multiple ALS acquisitions were combined into a single model or for the transferability of models across multiple study areas (Tompalski et al., 2019). Nighttime acquisitions could also be preferred due to the sensitivity of the single-photon detectors to noise coming from solar radiations if air traffic regulations and regional airport operating hours allow it.

Chapter 4. Mapping dominant tree species groups by combining area-based and individual tree crown lidar metrics with Sentinel-2 data⁴

4.1. Introduction

The previous chapter presented a framework to perform an EFI fully driven by the SPL100 data acquisition at the Romeo Malette Forest. The accuracy of the area-based estimates of Lorey's height, basal area, quadratic mean diameter, whole and merchantable volume, and above ground biomass were comparable to what would be expected with a conventional linear-mode lidar system. These attributes are strongly related to the structural information provided by the point cloud, however other attributes such as species composition are more challenging to derive accurately using summary metrics of lidar point cloud heights (Fassnacht et al., 2016). Information related to species composition is nevertheless needed for forest management, for example to assess species-specific growing stocks and yield (e.g. Maltamo and Packalen 2014; Tompalski et al. 2018) and monitor tree species richness (Gamfeldt et al., 2013).

Traditionally, assessment of species composition involved manual photo-interpretation of aerial imagery, from which stands of homogeneous forest type can be delineated, and field information obtained from plots sampled across the landscape (Leckie and Gillis, 1995). However, the accuracy of these photo-interpreted species compositions is rarely assessed (Tompalski et al., 2021), and when they are, have been found to have a relatively low level of accuracy (54–56% in Pinto et al. 2007 , ~36% in Thompson et al. 2007).

⁴ The content of this chapter has been adapted from:

Queinnec, M., Coops, N.C., White, J.C., Griess, V.C., Schwartz, N.B., McCartney, G (2022). In Press. Mapping dominant tree species groups by combining area-based and individual tree crown lidar metrics with Sentinel-2 data. *Canadian Journal of Remote Sensing*.

Moreover, the accuracy of photo-interpretation depends on the complexity of the forest environment, as well as the experience of the photo-interpreter, the latter of which are increasingly rare. For these reasons, there has been a growing interest into assessing species composition using automated classification techniques. However, to date, such methods have been unable to replicate the detailed species compositions associated with photo-interpreted forest inventories and scaling up species mapping approaches across large-areas remains a significant challenge (Fassnacht et al., 2016).

Species-related traits such as the pigments, water content and internal structure of leaves, branching, foliage density and crown shape influence the distribution and intensity of returns in ALS point clouds and the spectral response measured by optical sensors particularly in parts of the visible ($\lambda = 400 - 700$ nm), near-infrared ($\lambda = 700 - 1300$ nm) and shortwave infrared ($\lambda = 1300 - 2500$ nm) regions of the electromagnetic spectrum (Fassnacht et al., 2016). Research examining prediction of tree species with ALS data has largely focused on the individual tree level by using ITD and ITC segmentation algorithms from which tree-specific crown shape and density metrics can be derived (Holmgren and Persson, 2004; Michałowska and Rapiński, 2021). Nevertheless, structural information alone generally has lower potential than spectral information to classify tree species (Fassnacht et al., 2016). Spectral information is often integrated by combining ALS data and high resolution multispectral or hyperspectral imagery (Coops et al., 2004; Dalponte et al., 2014; Holmgren et al., 2008; Holmgren and Persson, 2004) or by using novel multispectral ALS sensors which emit laser pulses at different wavelengths (Axelsson et al., 2018; Budei et al., 2018; Kukkonen et al., 2019; Prieur et al., 2022). As tree-level approaches are prone to bias due to issues with segmentation (Kaartinen et al., 2012), approaches that combine ABA and ITC, known as semi-ITC, have also been explored (Breidenbach et al., 2010b). In areas where species-specific attribute estimates are required for the inventory (e.g. Finland), non-parametric supervised learning methods, such as k-nearest neighbours algorithms (k-NN) are commonly used for area-based estimates (e.g. Lindberg et al., 2013; Maltamo et al., 2021).

Area-based or stand-level predictions of species information have been achieved using spectral data from medium resolution multispectral satellite imagery such as Sentinel-2 or Landsat exclusively (Astola et al., 2019; Breidenbach et al., 2021; Grabska et al., 2019; Hemmerling et al., 2021; Persson et al., 2018) but the integration of these data with ALS in an area-based context has been less explored (Fassnacht et al., 2016). Therefore, the objectives of this chapter were: (1) to produce area-based predictions of dominant species groups using a combination of standard lidar height metrics, aggregated metrics derived from ABA and ITC methods, and Sentinel-2 imagery across the large extent of the Romeo Malette Forest; (2) to explore and assess the ability of these three sets of metrics to differentiate between species groups; and (3) to discuss implications of using this approach for large-area operational forest management at the stand level.

4.2. Methods

4.2.1. Study area

This chapter focuses on the Romeo Malette Forest, which is described in section 2.1.1 of this dissertation.

4.2.2. Dominant tree species groups

For each ground plot established at the Romeo Malette Forest (see section 2.3), the trees located within - or above - the dominant canopy layer were used to derive dominant species groups as described in Table 4-1. Two functional groupings were established. The first was more general and grouped plots according to the dominance of forest types: softwood, mixedwood and hardwood dominated groups. The second functional grouping distinguished the dominance of the two most common commercial tree species: black spruce and jack pine. Plots were assigned to five functional groups: black spruce-dominated, jack pine-dominated, mixed conifer, mixedwood, and hardwood. Similar species groupings were used in previous forest inventories at the Romeo Malette Forest (Queinnec et al., 2021a; Woods et al., 2011). The capacity of the developed approach using Sentinel-2 and ALS data to accurately predict each of these groupings was assessed.

Table 4-1. Stratification of ground plots into 3 or 5 functional groups based on the species of the trees positioned in - or above - the dominant canopy layer. The number of trees in each plot was used to derive proportions.

Species group	Number of plots	Number of stem-mapped plots	Description
<i>3 Functional Groups</i>			
Softwood	97	26	The proportion of softwood ¹ tree species is > 70%
Hardwood	114	25	The proportion of hardwood ² tree species is > 70%
Mixedwood	33	2	Plots not classified as Softwood or Hardwood. The proportion of softwood and hardwood tree species in these plots is between 30% and 70%
<i>5 Functional Groups</i>			
Black Spruce dominated	43	11	The proportion of softwood tree species is > 70% and > 50% of the trees are black spruce
Jack Pine dominated	40	10	The proportion of softwood tree species is > 70% and > 50% of the trees are jack pine
Mixed Conifer	12	4	The proportion of softwood tree species is > 70% but the proportion of black spruce or jack pine is not > 50%
Mixedwood	35	3	Plots not classified as other species groups. The proportion of softwood and hardwood trees in these plots is between 30% and 70%
Hardwood	114	25	The proportion of hardwood trees is > 70%

¹Softwood species include jack pine, black spruce, balsam fir (*Abies balsamea*), white spruce, american larch (*Larix laricina*) and eastern white cedar (*Thuja occidentalis*)

²Hardwood species include trembling aspen, paper birch and red maple (*Acer rubrum*)

4.2.3. Lidar data processing

A 25-cm spatial resolution canopy height model (CHM) was generated from the SPL100 normalized point cloud using the highest return within grid each cell. To densify the point cloud and reduce the number of pits, each return was considered as a disk of 30 cm diameter, which corresponds to the actual footprint of each beamlet on the ground (beam divergence of 0.08 mrad and altitude of 3800 m above ground). The CHM was then smoothed using a Gaussian filter ($\sigma = 1$) on a 5 x 5 pixel moving window in order to remove noise and pits.

Two sets of lidar summary metrics were calculated from the normalized SPL100 point cloud and derived CHM for each of the ground plots and across the entire study area at a resolution of 20 x 20 m matching the ground plot area of 400 m² (Table 4-2). The *lidR* package (Roussel et al., 2020) implemented in *R* (R Core Team, 2022) was used to generate the metrics. The first set of metrics included standard area-based metrics (as performed for the EFI presented in Chapter 3) calculated from the height distribution of the SPL100 returns, as well as the rumple index, which characterizes surface roughness of the canopy. The signal attenuation depth (SAD) of each return (Irwin et al., 2021; Véga et al., 2016), which indicates how far a return travels within the canopy, was also calculated. SAD is calculated for each return as the difference between the CHM and the return height and normalized to the CHM. $SAD = 0$ indicates a return on the outer canopy layer while $SAD = 1$ indicates a return on the ground. SAD was calculated only for returns where $CHM > 2$ m.

A second set of metrics was derived by summarizing the distribution of ITD and ITC metrics at the plot-level or 20 x 20 m pixel-level using the sum, median and standard deviation of values (Table 4-2). The ITD was performed using the local maximum filter algorithm implemented in *lidR* (Popescu and Wynne, 2004; Roussel et al., 2020). The algorithm uses a circular moving window to detect local maxima in the CHM and a size parameter controls the diameter of the window. The size parameter was varied between 0.5 m and 5 m at 0.5 m intervals to assess the influence of the window on the ITD performance window shapes. The individual tree tops detected by the ITD algorithm were then used as seed points in the marker-controlled watershed segmentation algorithm implemented in the *R* package *imager* to detect ITCs. The marker-controlled watershed algorithm was selected as it is parameter-free, not designed for particular crown shapes and computationally efficient, which are important considerations for large areas with varying species composition such as the Romeo Malette Forest. The algorithm inverts the CHM and sets the detected tree tops as local minima. A region is then propagated around each tree top by filling the inverted CHM. To avoid over-propagation, the region-growing algorithm was limited to the pixels of the CHM that

were not below 1 meter of the detected tree top height. That way, trees with a conical crown shape were expected to grow smaller regions than trees with a flatter and broader crown.

Table 4-2. Description of the predictor variables derived from the ALS data and Sentinel-2 imagery.

	Metric	Description
Area-based predictors	zq5, zq10, ..., zq90, zq95	5 th , 10 th , ... 90 th , 95 th percentiles of returns height > 1.3 m classified as vegetation
	zsd, zskew, zkurt	Standard deviation, skewness and kurtosis of returns height > 1.3 m classified as vegetation
	pzabove2, pzabovemean	Percentage of first returns > 2 m or > average returns height
	zpcum1, zpcum2, ... zpcum9	Cumulative percentage of returns in layers (Woods et al., 2008)
	zentropy	Shannon diversity index quantifying the evenness of the distribution of lidar returns height
	zrumple	Ratio of canopy outer surface area to ground surface area
	depth_q25, depth_q50, depth_q75	25 th , 50 th and 75 th percentile of signal attenuation depth of all returns
ITC predictors	n_itd	Number of detected tree tops
	cr_a_sum; cr_a_med; cr_a_sd	Sum, median or standard deviation of segmented crown area
	cr_vol_sum; cr_vol_med; cr_vol_sd	Sum, median or standard deviation of the product of segmented crown area and corresponding tree top height (Duncanson et al., 2015)
	cr_lor	Average of tree top height divided by corresponding segmented crown area (Duncanson et al., 2015)
Sentinel-2 predictors	s2_blue, s2_red, s2_green, s2_redge1, s2_redge2, s2_redge3, s2_nir, s2_siwr1, s2_swir2	Blue (band 2: 490 nm); Green (band 3: 560 nm); Red (band 4: 665 nm); Red-edge 1 (band 5: 705 nm); Red-edge 2 (band 6: 740 nm); Red-edge 3 (band 7: 783 nm); NIR (band 8A: 865 nm); SWIR1 (band 11: 1610 nm); SWIR2 (band 12: 2190 nm)

4.2.4. Sentinel-2 imagery

A total of 44 Sentinel-2A and Sentinel-2B level-1C (top-of-atmosphere reflectance) scenes acquired across the Romeo Malette Forest between June 1st and September 30th 2018 were obtained and processed to level-2A (bottom-of-atmosphere reflectance) using the *R* package *sen2r* (Ranghetti et al., 2020). This *R* package

streamlines the downloading and pre-processing of Sentinel-2 imagery using the atmospheric correction algorithm Sen2Cor (Main-Knorn et al., 2017). All pixels classified as missing values, shadows, clouds or cirrus by the level-2A scene classification map were masked out and scenes were projected and resampled to align with the 20 x 20 m SPL100 metrics grid using nearest neighbour interpolation. The Sentinel-2 bands used in this study had a spatial resolution of either 10 x 10 m or 20 x 20 m so downscaling to the resolution of the gridded metrics was not required. A cloud-free composite product was then produced by averaging the pixel values across all available cloud-free observations for each pixel. Spectral bands listed in Table 4-2 were selected as predictor variables for species group classification.

4.2.5. Classification of dominant tree species

Random forest models were used to classify the dominant species groups (Table 4-1) using the area-based, ITC and Sentinel-2 predictors (Table 4-2). The conditional random forest algorithm implemented in the function *cforest* of the *party* package in *R* (Hothorn et al., 2006; Strobl et al., 2008, 2007) was selected due to its ability to compute an unbiased measure of variable importance (see below). Predictor variables with a pairwise Pearson correlation coefficient > 0.9 were identified and from each pair of variables, only the one with the lowest average correlation among the other predictor variables was retained. Different scenario runs were also performed, as described in Table 4-3, to assess the influence of different sets of predictor variables on the accuracy of the species groups classification.

Table 4-3. Description of the predictor variables included in the different scenario runs. Refer to Table 4-1 for a description of the metrics

Scenario run	Abbreviation
Sentinel-2 only (s2_only)	s2_blue, s2_green, s2_red, s2_red_edge_2, s2_swir2
Lidar height only (z_only)	zq5, zq85, zsd, zkurt, zskew, zentropy, pzabove2, pzabovemean, zpcum1, zpcum2, zpcum8, zpcum9, sad_q25, sad_q50, sad_q75
Aggregated ITD and ITC metrics only (itd_crown)	n_itd, cr_lor, cr_a_sum, cr_a_med, cr_a_sd, cr_vol_sum, cr_vol_sd
All lidar metrics (all_lidar)	Metrics included in scenario runs 2 and 3
All predictor variables (all)	Metrics included in all the other scenario runs

4.2.6. Accuracy assessment of the individual tree detection and area-based species classification

The performance of the ITD was assessed using the ground plots that were stem-mapped (see section 2.3). Detected trees were automatically matched to their nearest field mapped trees based on the 3-dimensional distance between their tree tops. Only pairs of detected and mapped trees with a horizontal distance ≤ 5 m and height difference ≤ 3 m were considered as valid matches, or true positives (TP). Since the ITD algorithm only detects trees in the upper canopy layer, only field trees that were in or above the dominant canopy layer and that were not matched to any detected tree top were considered as false negatives (FN). All remaining detected tree tops not matched to any field tree were considered as false positives (FP). The accuracy of the ITD algorithm was assessed for each set of parameters by calculating the recall (Eq 1), precision (Eq 2) and F-score (Eq 3). Recall, or producer's accuracy, indicates the proportion of correctly detected trees out of all trees mapped on the field, whereas precision (user's accuracy) indicates the proportion of correctly detected trees out of all trees detected by the ITD algorithm. The F-score combines recall and precision measures by calculating their harmonic mean, which penalizes more low or unbalanced recall and precision values than the arithmetic or geometric means.

$$recall = \frac{TP}{TP+FN} \quad (\text{Eq 1})$$

$$precision = \frac{TP}{TP+FP} \quad (\text{Eq 2})$$

$$F\text{-score} = 2 \times \frac{recall \times precision}{recall + precision} \quad (\text{Eq 3})$$

The accuracy of the random forest classification models was assessed using repeated k-fold cross-validation (k = 10; repeated 5 times). Due to the imbalance of species groups in the observations, the training folds were randomly sampled so that all species groups had the same number of observations as the least frequent species group. The testing folds remained unchanged. The accuracy of the models was assessed using

overall accuracy (OA; Eq 4), recall (Eq 1), precision (Eq 2) and F-score (Eq 3) from the predictions of the repeated folds. The OA is a global measure of accuracy that indicates the proportion of observations that were correctly classified across all species groups whereas recall, precision and F-score are calculated for each species group.

$$OA = \frac{TP}{TP+FN+FP} \quad (\text{Eq 4})$$

The cross-validation was also used to optimize the parameter *mtry* that controls the number of predictor variables that are randomly selected to split the data in the nodes of each tree. The *mtry* value of the model with the lowest overall accuracy across the folds was used in the final model for both the 3-group and 5-group classification.

4.2.7. Variables driving the classification

The importance of the predictor variables to predict the species groups was assessed using the conditional variable importance measure from the conditional random forest model (Strobl et al., 2008). Variable importance in random forest is assessed by examining the effect of breaking the relationship between a predictor variable and the response, by permuting the predictor variable values, on the prediction accuracy (measured as OA). The difference between prediction OA before and after permutation of a predictor variable values is used as a measure of variable importance. Strobl et al. (2007) showed that when predictor variables are of different nature or vary in their scale of measurement, which is the case herein, the permutation scheme used in random forest tends to be biased toward certain variables. The conditional variable importance used herein is derived from a novel conditional permutation scheme, which enables to derive unbiased variable importance (Strobl et al., 2008). Conditional importance measures were linearly scaled between 0 (least important variable) and 100 (most important variable).

Random forest variable importance provides a measure of the overall contribution of each predictor variable to the species groups classification model but does not assess if some variables are more useful to discriminate between specific species groups than others. Therefore, a pairwise comparison of the predictor variable values between species groups was performed to gain insight into why some pairs of species groups were more or less separable. For each of the 15 most important predictor variables, the Kruskal-Wallis test, a non-parametric equivalent to one-way analysis of variance, was used to test the null hypothesis that the mean rank of the variable is equal across the species groups. When a significant difference ($p\text{-value} < 0.05$) was observed across species group, the post-hoc Dunn's test with Holm's p -value adjustment method was used to examine pairwise differences. That way, variables having more predictive power for some species groups compared to others could be highlighted.

4.2.8. Comparing predicted species group to existing photo-interpreted inventory

The estimates of species groups derived in this chapter were compared to stand-level species information available from the most up-to-date photo-interpreted forest resources inventory information available at the Romeo Malette Forest, which was initially established with manual photo-interpretation of multispectral aerial imagery acquired in 2005. About 20% of the stands initially established were updated using new aerial photographs or field-based surveys on a regular basis until 2014 to account for harvesting, natural depletion, silvicultural treatments, regeneration or changes in stand development stages. The objective of this comparison was to explore the classification results across a broader geographic extent than the cross-validation at the plot-level provided, and at the stand-level which is the management unit used for decision making. However, the photo-interpreted data were not considered as reference to assess the accuracy of the classification since their accuracy is poorly defined (Pinto et al., 2007; Thompson et al., 2007).

A total of 43,806 productive forest stands (median area of 5.6 ha) with SPL100 estimates of canopy height over 5 m ($zq90 > 5$ m) and over 50% canopy cover ($pzabove2 > 50\%$), to remove stands recently harvested or disturbed, were selected. In each stand, an estimate of overstory species composition as well as the percentage of crown cover each tree species occupies was available from the photo-interpreted inventory

and this information was used to classify each stand into the five functional species groups listed in Table 4-1. In each stand, the most frequently predicted species group was calculated and compared to the species group derived from the photo-interpreted inventory.

4.3. Results

4.3.1. Individual tree detection and crown segmentation

The accuracy of the ITD varied depending on the diameter of the local maxima filter window and species groups. The highest median F-score was observed for a circular window diameter of 1.5 m for black spruce dominated ($F_{\text{median}} = 0.82$) and mixed conifer ($F_{\text{median}} = 0.82$), 2 m for mixedwood ($F_{\text{median}} = 0.72$), 2.5 m for jack pine dominated ($F_{\text{median}} = 0.85$), and 3 m for hardwood ($F_{\text{median}} = 0.75$) species groups (Figure 4-1). All species groups combined, a highest median F-score of 0.72 was obtained for window diameters of 2 m, 2.5 m and 3 m and the window diameter of 2 m was retained to apply the ITD algorithm across the Romeo Malette Forest. Median values of recall, precision and F-score for each species group and window diameter are presented in Appendix B.

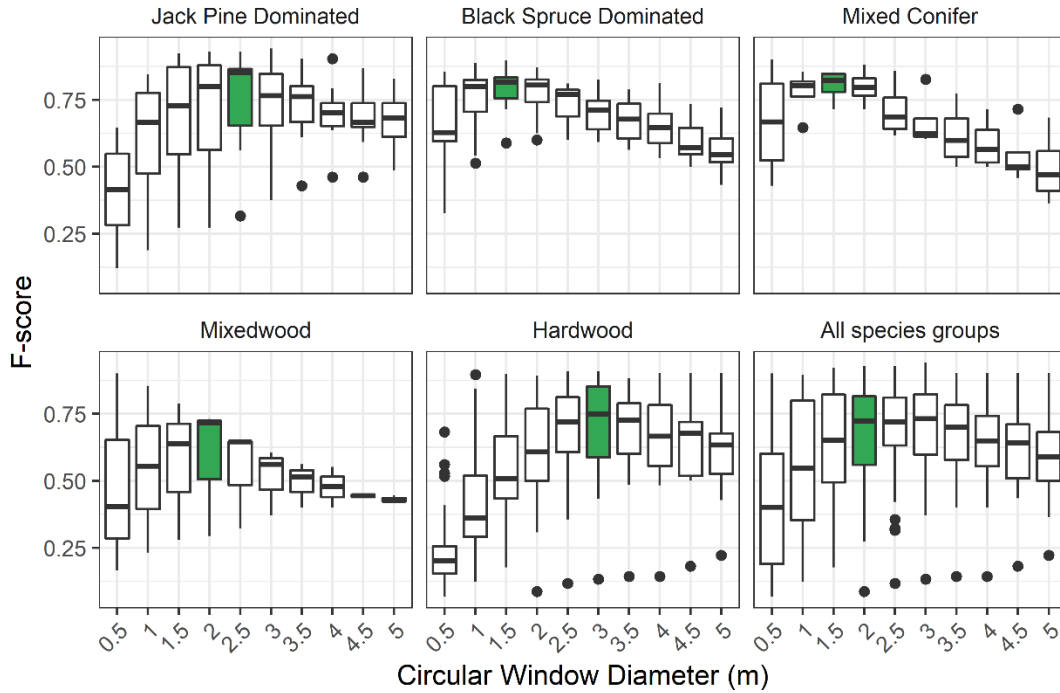


Figure 4-1. Influence of the local maxima window diameter on the F-score of the individual tree detection algorithm by species group. The window diameter for which the highest median F-score was observed is highlighted in green.

4.3.2. Classification of species groups

The classification of ground plots into the three functional species groups softwood, mixedwood and hardwood resulted in an overall accuracy ranging between 65.98% for the model including Sentinel-2 bands only, and 83.20% for the model including all predictor variables (Figure 4-2). When including all predictor variables, the softwood species groups had a precision and recall of 92.13% and 91.75% (F-score = 91.94%), and the hardwood species group had a precision and recall of 90.58% and 82.63% (F-score = 86.42%), respectively. The accuracy of the classification was poorer for the mixedwood species group with a precision of 45.62% and a recall of 60.00% (F-score = 51.83%). The confusion matrix is presented in Table 4-4 and illustrated in Figure 4-3A and recall, precision and F-score for each species group and model scenario are included in Appendix C and Appendix D.

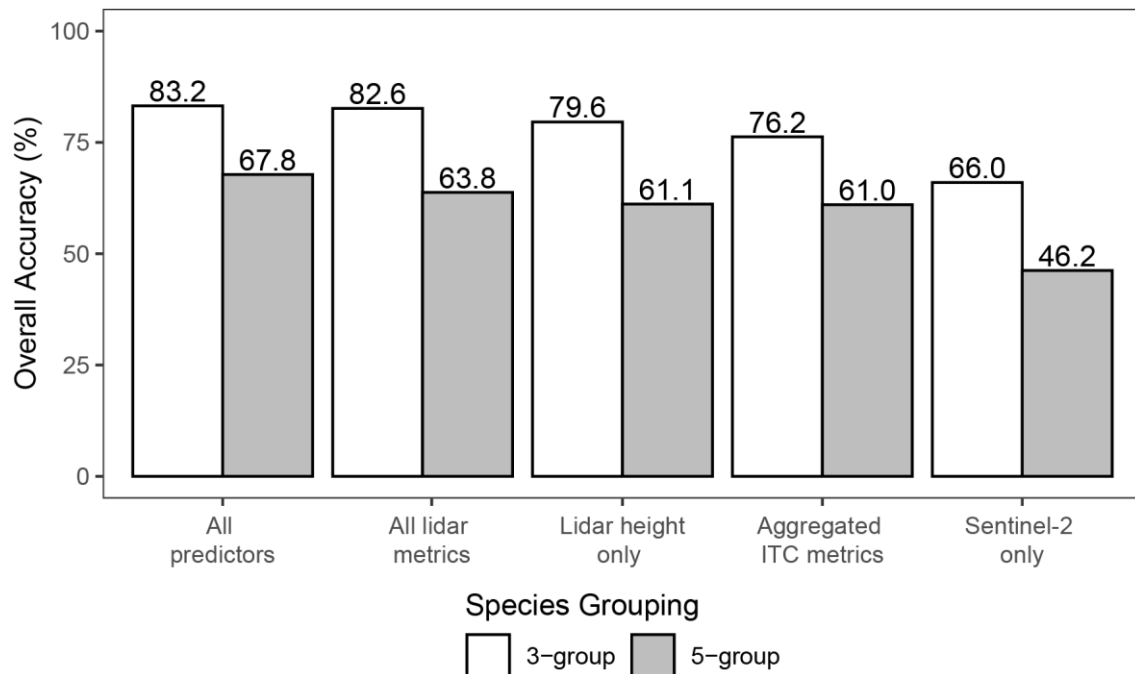


Figure 4-2. Comparison of the overall accuracy of the random forest classification model of the ground plots into softwood, mixedwood, and hardwood species groups (3 Species Groups) or jack pine dominated, black spruce dominated, mixed conifer, mixedwood, and hardwood species groups (5 Species Groups).

Table 4-4. Confusion matrix showing the results of the broader, 3-group classification model using all predictor sets. Predictions of the repeated folds are included in the matrix. SW = Softwood; MW = Mixedwood; HW = Hardwood.

		Reference			Total	Precision
		SW	MW	HW		
Predicted	SW	445	22	16	483	92.13%
	MW	35	99	83	217	45.62%
	HW	5	44	471	520	90.58%
	Total	485	165	570	1220	
Recall		91.75%	60.00%	82.63%		

When dividing the softwood species group into jack pine dominated, black spruce dominated and mixed conifer groups, the overall accuracy of the model ranged between 46.22% for the model including Sentinel-2 bands only and 67.89% for the model including all the predictor variables (Figure 4-2). The hardwood species group remained the most accurate with a precision of 87.34% and a recall of 87.19% (F-score = 87.27%). Jack pine and black spruce species groups had a similar precision of 62.50% and 61.76%,

respectively. Recall was, however, higher for black spruce (58.60%) than for jack pine (47.5%) dominated groups, which indicates that jack pine was misclassified as another species group more often than black spruce. Most classification errors in the jack pine and black spruce species groups occurred from confusion of one with the other or with the mixed conifers species group. Similar to the findings of the classification into softwood, mixedwood, and hardwood, low confusion was observed between the coniferous species and the mixedwood and hardwood species. The confusion matrix is presented in Table 4-5 and illustrated in Figure 4-3B and recall, precision and F-score for each species group and model scenario are included in Appendix C and Appendix D.

Table 4-5. Confusion matrix for the five species groups classification model using all predictor sets. Predictions of the repeated folds are included in the matrix. JP = Jack Pine Dominated; BS = Black Spruce Dominated; MC = Mixed Conifer; MW = Mixedwood; HW = Harwood.

		Reference					Total	Precision
		JP	BS	MC	MW	HW		
Predicted	JP	95	47	3	6	1	152	62.50%
	BS	58	126	8	12	0	204	61.76%
	MC	26	34	33	14	22	129	25.58%
	MW	18	8	14	76	50	166	45.78%
	HW	3	0	2	67	497	569	87.34%
	Total	200	215	60	175	570	1220	
	Recall	47.5%	58.60%	55.00%	43.43%	87.19%		

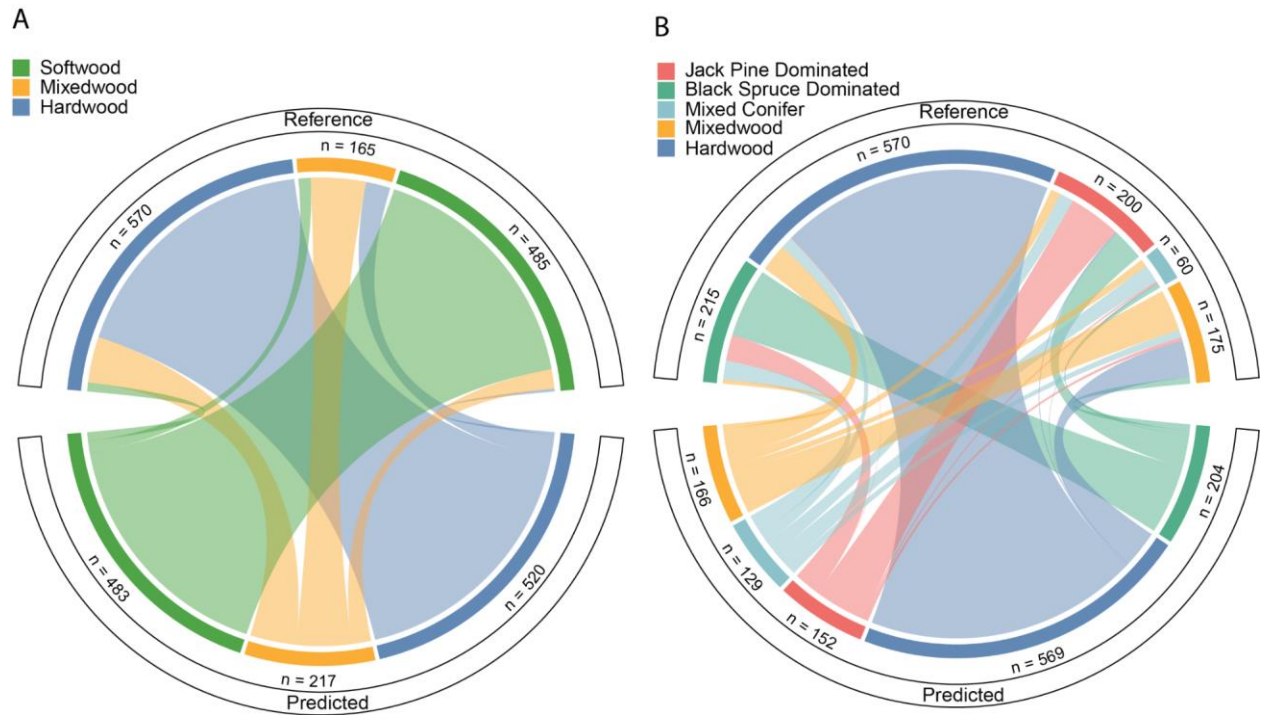


Figure 4-3. Chord diagram visualizing the confusion matrix of the random forest species classification model into (A) softwood, mixedwood, and hardwood species group; or (B) jack pine dominated, black spruce dominated, mixed conifer, mixedwood, and hardwood species groups for the scenario run that included all predictor variables. Predictions from all cross-validation folds were aggregated and are included in this diagram. The predicted species groups are shown in the bottom half and the reference species groups are shown in the upper half. The predicted species groups are shown in the bottom half and the reference species groups are shown in the upper half. Links illustrate the number of observations that were correctly classified (link between same predicted and reference species groups) and the number of classification errors (link originating from a predicted species group going to a different reference species group).

4.3.3. Variable importance for species groups classification

The variable importance of the 3-groups and 5-group species classification models including all predictor variables are presented in Figure 4-4. For the most detailed 5 species groups classification model, the second red edge band of Sentinel-2 (*s2_red_edge_2*; $\lambda = 740$ nm), the cumulative percentile of height returns *zpcum8* and the sum of segmented crown volumes (*cr_vol_sum*) showed markedly higher variable importance than other predictor variables for classifying species groups (Figure 4-4B). Six other variables, including four crown metrics, the 25th percentile of signal attenuation depth and *zpcum9*, had similarly scaled variable importance values ranging between 23% and 32%. The scaled variable importance values of the remaining variables were markedly lower (0 – 14%). The most important metrics for the 5-group species classification model also generally had high variable importance in the 3-group model (Figure

4-4A). Out of the 9 variables highlighted above that had markedly higher importance in the 5-group model (23 – 100%), all except *zpcum9* were also ranked in the 9 most important variables in the 3-group model.

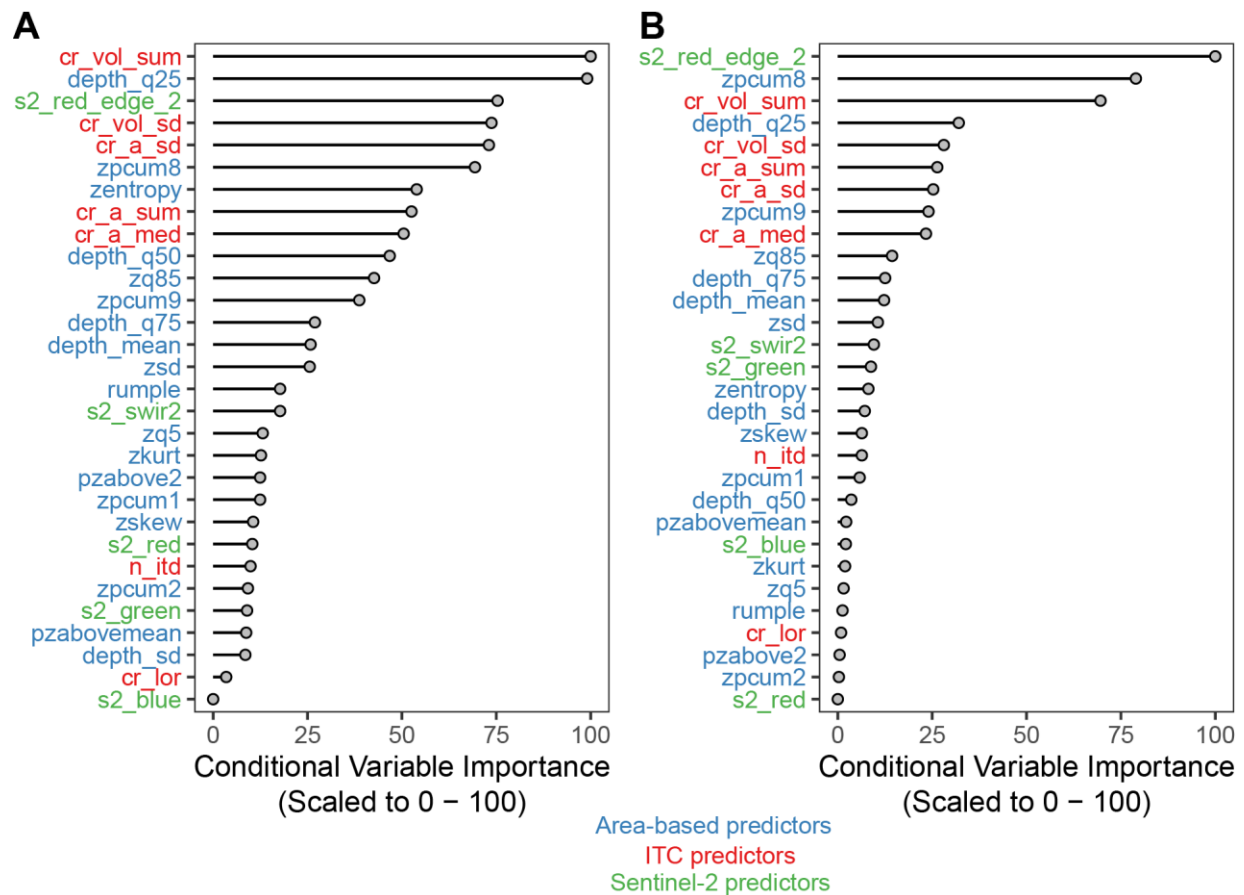


Figure 4-4. Conditional random forest variable importance of the 3-group (A) and 5-group (B) species groups classification model including all the predictor variables. Raw importance values were scaled between 0 (least important variable) and 100 (most important variable).

4.3.4. Comparison of predictor variables across species groups

Not all species groups were highly distinct across the predictor variables considered (Figure 4-5). Hardwood was generally the species group that was most distinct from others, especially when compared to jack pine and black spruce dominated groups. Importance of metrics related to crown area and volume were significantly different across all species groups except when comparing black spruce and jack pine dominated groups with the mixed conifer group as well as when comparing jack pine and black spruce dominated groups. The reflectance values of Sentinel-2 spectral bands were generally significantly different

when comparing jack pine and black spruce dominated groups with the other species groups, but not - or less - significantly, different when comparing mixedwood, mixed conifer, and hardwood groups. The SWIR2 band ($\lambda = 2190$ nm) reflectance value of Sentinel-2 was significantly different ($p < 0.001$) when comparing mixed conifer, mixedwood, and hardwood groups with the black spruce dominated group, but not significantly different when these groups were compared to the jack pine dominated group. The only predictor variables displaying significantly different values between jack pine and black spruce dominated groups were *zpcum9* ($p < 0.05$), *zpcum8* ($p < 0.01$) and the Sentinel-2 SWIR2 reflectance ($p < 0.05$).

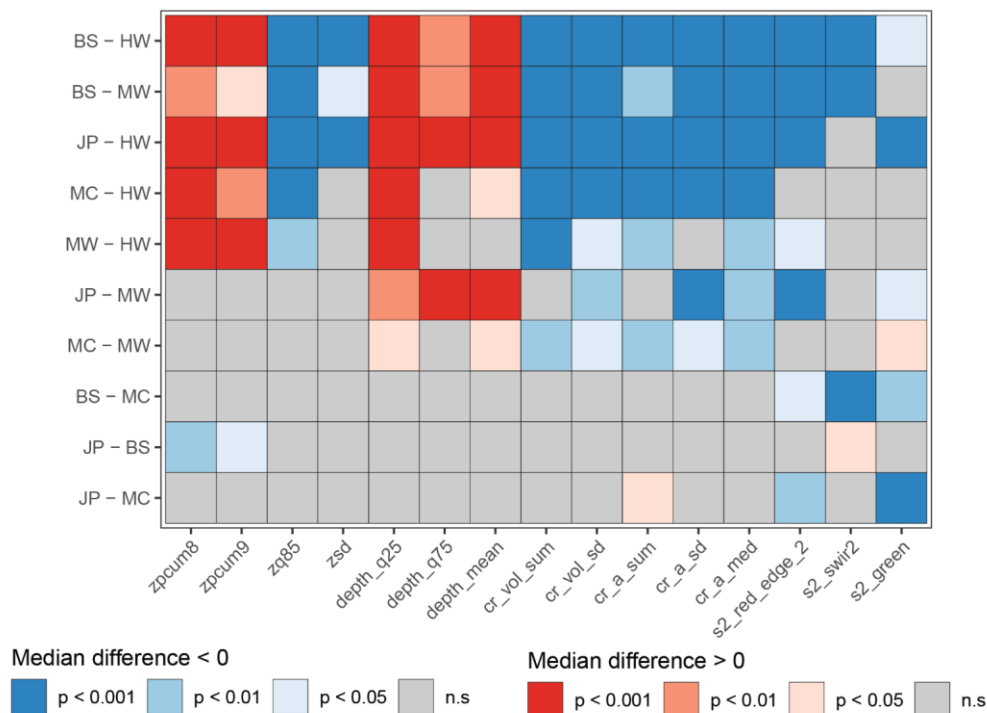


Figure 4-5. Pairwise comparison of significant differences in predictor variable values across species groups. The blue and red color gradients indicate if the pairwise difference between the median values is negative or positive, respectively. JP = Jack Pine Dominated; BS = Black Spruce Dominated; MC = Mixed Conifers; MW = Mixedwood; HW = Hardwood.

4.3.5. Comparing predicted species group to existing photo-interpreted inventory

Figure 4-6 shows the distribution of the most frequently predicted species group in the stands of the photo-interpreted forest resource inventory.

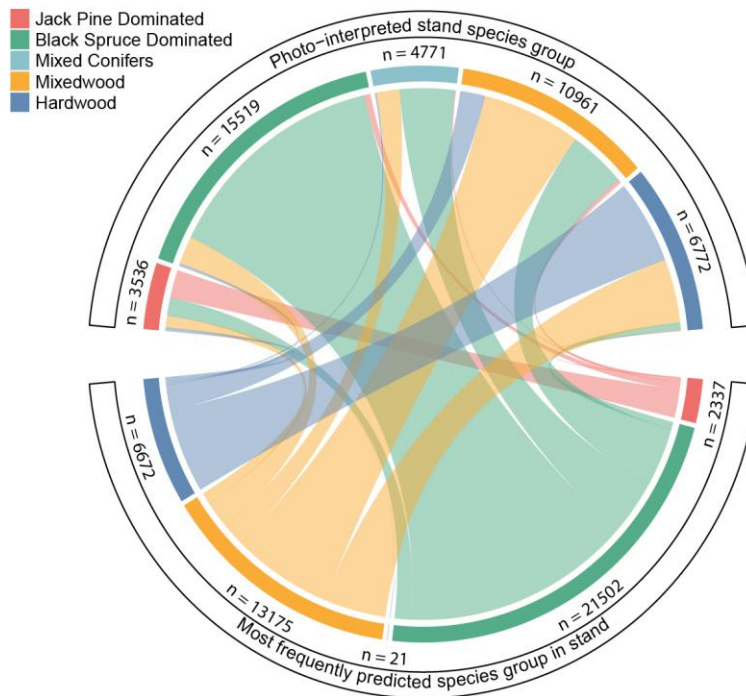


Figure 4-6. Chord diagram showing the distribution of the most frequently predicted species groups in the stands of the photo-interpreted inventory. The most frequently predicted species groups in each stand are shown in the bottom half and the stand-level species groups derived from photo-interpreted species composition are shown in the upper half.

The black spruce dominated species group was the most frequently predicted species group in 86.2% of the stands dominated by black spruce according to the photo-interpreted forest resource inventory. In stands classified as jack pine-dominated, jack pine was the most frequently predicted species group in 45.7% of the stands, and black spruce was the most frequently predicted species group in 29.7% of the stands. The most frequently predicted species groups in stands classified as hardwood in the photo-interpreted inventory were hardwood (53.8%) and mixedwood (40.7%). Overall, a negligible number of stands where mixed conifer was the most frequently predicted species group were observed. Black spruce dominated (68.1%) and mixedwoods (28.7%) were the most frequently predicted species groups in stands classified as mixed conifer by the photo-interpreted inventory. Figure 4-7A-D shows some examples of stands and their species group predictions. A comparison against aerial imagery is also provided in Figure 4-8.

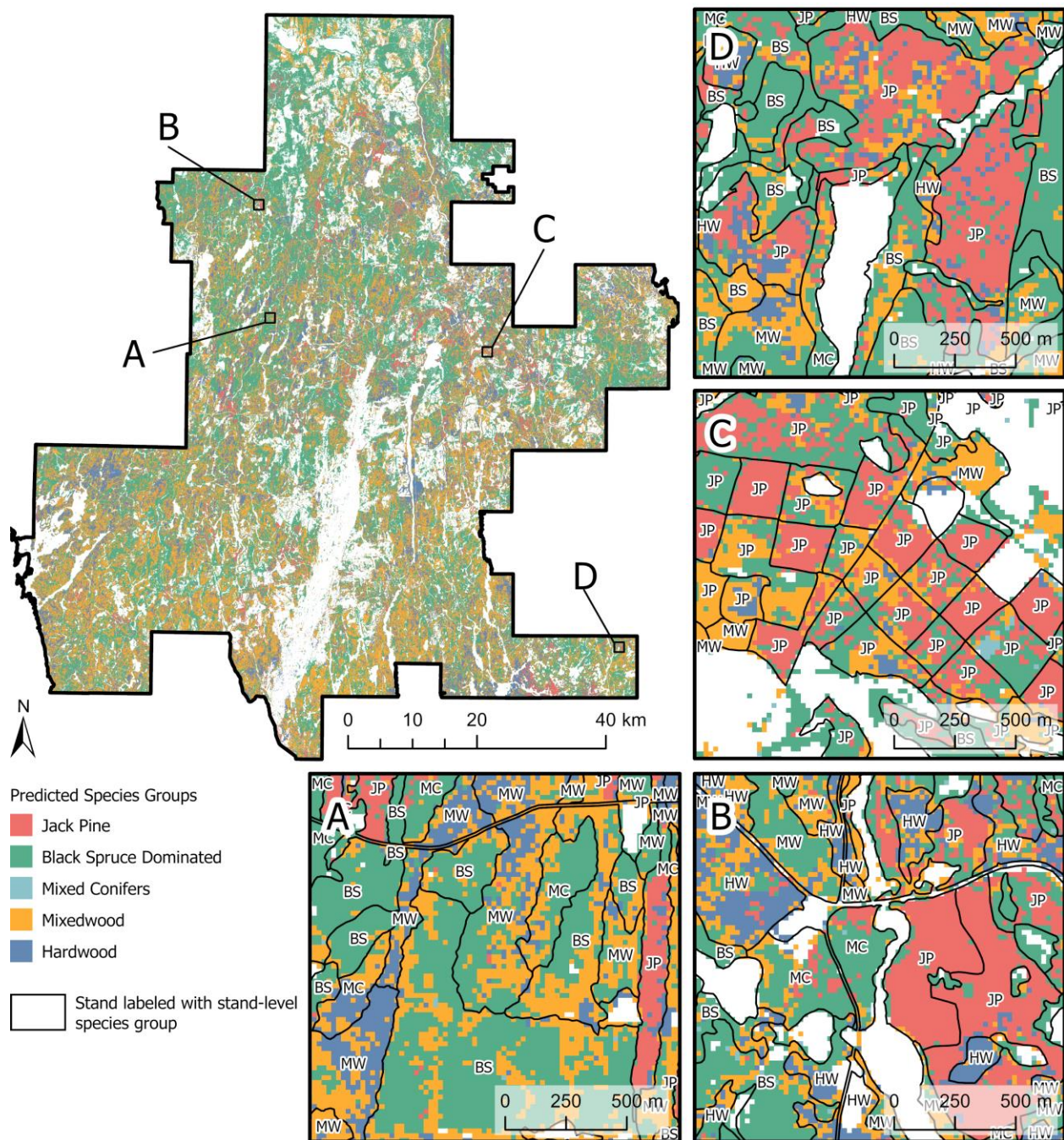


Figure 4-7. Wall-to-wall predictions of species groups overlaid with forest stands labeled by their inventory (photo-interpreted) species groups. Predictions were only made in pixels where $zp90 > 5$ m, $pzabove2 > 50\%$ and in productive forest stands. JP = Jack Pine Dominated; BS = Black Spruce Dominated; MC = Mixed Conifers; MW = Mixedwood; HW = Hardwood.

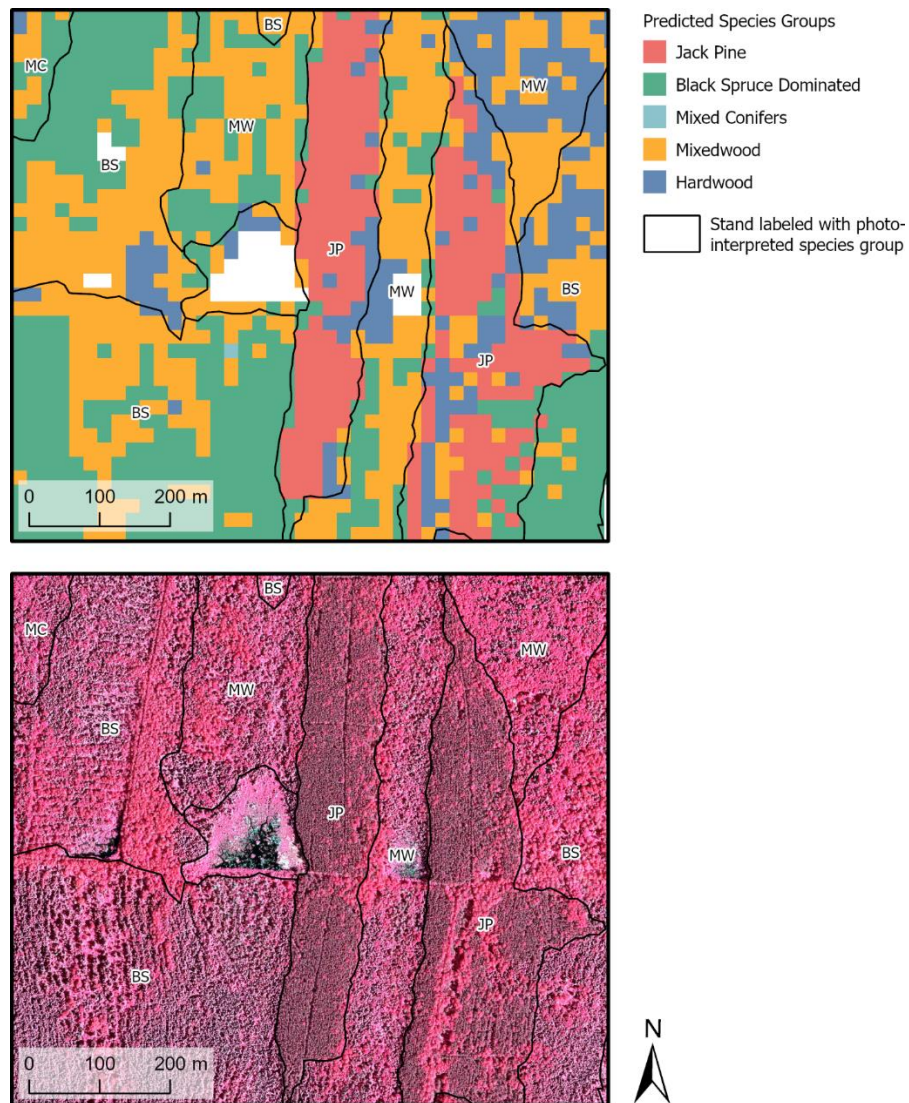


Figure 4-8. Comparison of species groups predictions and photo-interpreted stands with aerial photographs acquired at the same time as the SPL100 data. The aerial photograph is displayed as a false color composite (Red channel = NIR; Green channel = Red; Blue channel = Green).

4.4. Discussion

In this chapter, area-based and individual tree metrics obtained from ALS data were combined with Sentinel-2 data to predict dominant species groups across the Romeo Malette Forest. Below, the accuracy of the species group predictions, as well as the main variables that helped discriminate between species groups are discussed in the context of findings from previous research. Implications of using such approach

in an operational context, where information needs vary depending on the forest management approaches, are also considered.

4.4.1. Accuracy of species group predictions

Results indicated that the prediction of species groups was the most accurate (OA = 83.20% for 3-group classification; OA = 67.79% for 5-group classification) when models combined ALS and Sentinel-2 predictors relative to models that included only subsets of these predictors. Kukkonen et al. (2018) also observed that the accuracy of Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*) and broadleaved species volume improved when using both Sentinel-2 and ALS data (RMSE = 49.8 – 69.0%) compared to using ALS data only (RMSE = 53.9 – 77.1%). In a 14,000 ha temperate forest, Plakman et al. (2020) also found that the combination of Sentinel-2 data and crown structural metrics derived from ALS data yielded the most accurate classification of 8 species (OA = 78.5%) at the individual tree level. However, their results show that models based on Sentinel-2 alone performed better (OA = 73.7%) than models only based on crown structural metrics (OA = 58.4%), which was not the case here.

Accuracies comparable to the one obtained in this study with the best performing model combining ALS and Sentinel-2 have been observed in studies using Sentinel-2 data alone. For example, Persson et al., (2018) mapped 5 species in a 1500 ha boreal forest with an OA of 88.2% and Breidenbach et al. (2021) mapped Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*) and deciduous forests across Norway with an OA of 90.5%. In these two studies, models using multitemporal Sentinel-2 data, that can capture different phenological dynamics between species, outperformed models based on single-date Sentinel-2 acquisition. This could explain the fact that markedly lower accuracy was observed (OA = 65.98% for 3-group classification; OA = 46.23% for 5-group classification) in this chapter when using a single composite of Sentinel-2 data acquired between June and September. Sentinel-2 images acquired prior and after the growing season could have been integrated even though the presence of snow on the ground and canopy would have an impact on the measured reflectance and would therefore consider further considerations. Moreover, the consideration of mixed conifers and mixedwood species groups, which

comprise a diverse range of species and are not as clearly distinguishable compared to other forest types dominated by a single species, contributed to classification errors.

The accuracy of the models using ALS data only (OA = 76.2 – 82.6% for 3-group classification; OA = 61.0 – 63.8% for 5-group classification) were within the range of accuracy values that Fassnacht et al. (2016) compiled from 17 studies classifying species with ALS data (OA = ~ 55 – 90%). It is important to note however that these predictions ranged from the individual tree species to species groups, in various forest types and areas as well as with different ALS sensors and acquisition parameters. Only one study by Prieur et al. (2022) has examined tree species classification with the SPL100 to date. Their study area was also located in Ontario, Canada but in a more complex species-environment than the Romeo Malette Forest studied here. Prieur et al. (2022) obtained accuracies comparable to the ones observed in this chapter, with an OA of 82.9% if grouping predictions into softwood and hardwood species groups and 68.3% if grouping predictions into pine, spruce, other softwood hardwood and intolerant hardwood species groups.

4.4.2. Variables driving species groups classification

This work highlighted that the three sets of predictors provided valuable information to discriminate the modeled species groups. As discussed by Fassnacht et al. (2016), species influence the distribution of ALS returns within the canopy as a function of differences in crown shape or foliage density. Two metrics representing the cumulative distribution of returns in the upper layers of the canopy (*zpcum8*, *zpcum9*) had a markedly larger conditional variable importance than the other plot-level summary ALS metrics. The signal attenuation depth metrics, which are influenced by foliage density and canopy closure, were also important variables in the random forest classification model, especially to discriminate between coniferous species groups and mixedwood or hardwood species groups. This also confirms results from Irwin et al. (2021) who observed that the signal attenuation depth of the SPL100 was generally lower in closed-canopy deciduous forests, compared to coniferous forest types. It is also important to note here that point clouds of the SPL100 tend to be dominated by first returns in the upper canopy because of the low energy of the laser

beams and the filtering of noise returns that can decrease the amount of returns in the intermediate and lower layers of the canopy (Brown et al., 2020; Yu et al., 2020).

The red edge region of the electromagnetic spectrum is generally found to be of high importance in vegetation mapping because it results from a low reflectance in the red, due to chlorophyll absorption, and a high reflectance in the near-infrared (NIR), due to the internal structure of leaves (Horler et al., 1983). This high relevance is also reflected here, with the red edge 2 band of Sentinel-2 ($\lambda = 740$ nm) being the most important spectral band to discriminate species groups, a finding that is supported by other studies carried out in forests with both coniferous and deciduous trees (Astola et al., 2019; Persson et al., 2018). The red edge 2 band was also the most significantly different ($p\text{-value} < 0.001$) when comparing the jack pine and black spruce dominated species groups with the mixedwood or hardwood species groups. Among the blue ($\lambda = 490$ nm), green ($\lambda = 560$ nm), red ($\lambda = 665$ nm), red edge 2 ($\lambda = 740$ nm) and SWIR 2 ($\lambda = 2190$ nm) Sentinel-2 bands included in the models, the SWIR 2 band was the only one where significantly different reflectance values were observed between the jack pine and black spruce dominated groups. Sentinel-2 red edge 2 band ($\lambda = 740$ nm) and narrow-NIR band ($\lambda = 860$ nm), which was not included in the classification models due to its high correlation ($r > 0.99$) with the red edge 2 band, have also been shown to help discriminate Scots pine from Norway spruce (Kukkonen et al., 2018). These results are in line with previous research that has showed that the spectral reflectance of black spruce needles was lower than jack pine needles in the NIR ($\lambda = 700 - 1000$ nm) region of the electromagnetic spectrum (Middleton et al., 1997). In a more recent study also examining the shortwave infrared (SWIR) region of the electromagnetic spectrum, Hovi et al. (2017) observed that spectral differences between coniferous boreal tree species needles were larger in the SWIR ($\lambda = 1300 - 2500$ nm) region of the electromagnetic spectrum compared to the visible ($\lambda = 400 - 700$ nm) or NIR ($\lambda = 700 - 1300$ nm) regions. In addition to the needles, the site on which tree species occur can also influence the reflectance of each species groups. For example, black spruce often occurs on wetter sites than jack pine and the SWIR region of the electromagnetic spectrum is known to be sensitive to moisture (Zhu et al., 2014).

Individual tree crown features derived from ALS can be effective to separate species groups, in particular coniferous from broadleaf species (Fassnacht et al., 2016). On the other hand, Duncanson et al. (2015) found that the accuracy of area-based estimates of above ground biomass modeled from ITC metrics aggregated at the plot-level, as done in this research, did not outperform the accuracy of models based on plot-level height summary metrics. The ITC metrics included here marginally improved the accuracy of classification models compared to models with area-based metrics alone. ITC metrics were significantly different ($p\text{-value} < 0.05$) when comparing the hardwood species group to the conifer groups, but not among conifer groups. The metrics related to crown area or proxy of crown volume were more important if they were summed rather than averaged within a plot or a pixel. Summed metrics could possibly capture more information, e.g., in relation to canopy cover, than averaged metrics and explain why they were more important predictors. The fact that the marker-controlled watershed segmentation was limited to the top of each tree was important. Without this limit, delineated crowns tended to occupy the entire plot or pixel area and summed crown area metrics had limited variability across the species groups, especially when segmenting closed canopies. This approach was suitable to improve the explanatory power of the crown metrics but would not have been as effective if the objective was to isolate individual tree and their entire crowns as best as possible. More complex individual tree crown metrics, such as three-dimensional alpha shapes that represent the envelope of a tree crown (Korhonen et al., 2013), could have more explanatory power than the ones that were used herein to classify tree species. However, the generation of detailed individual tree metrics across large areas such as the Romeo Malette Forest is challenging due to the computational requirements and the necessity of tuning of individual tree segmentation algorithm parameters to different forest types.

The measure of intensity of linear-mode lidar instruments, which is related to the amount of energy in the reflected pulse, can be more useful than the structural information of the point cloud to discriminate tree species (Fassnacht et al., 2016; Shi et al., 2018). Equivalent intensity measures cannot be derived from lidar instruments based on single photon lidar detection since the returned signal is binary instead of being a

pulse of energy. A pseudo-intensity measure can be generated by aggregating the number of binary returns detected for each transmitted pulse but its properties and characteristics compared to the intensity of other ALS sensors is not well known (Hartzell et al., 2018). The SPL100 pseudo-intensity features were not included as predictor variables in this work, and other studies have shown that it provides limited amount of information for species-specific volume predictions or species classification. However, further research characterizing their properties across different surfaces and suitable standardization methods would be valuable.

4.4.3. Implication for operational forest management

Whether to use tree-level, area-based or stand-level forest inventory ultimately depends on the information needs. End users have come to expect detailed species composition estimates such as the ones provided by manually photo-interpreted inventories. However, in some contexts, knowledge of dominant species may be sufficient to support information needs, such as estimating timber volume (Tompalski et al., 2014). Area-based species groups predictions such as the ones derived herein can inform extensive forest management practices, which prevail in Canada where forests cover large extents, are mainly publicly owned and managed for multiple objectives (Wulder et al., 2007). In nations with a smaller forest landbase and a larger share of private ownership, intensive forest management practices are more common and require more detailed and precise species composition information than the ones derived herein. Changes in silvicultural approaches and practices, associated with changing climate and broader sustainable use of forest resources (Achim et al., 2022), could also require more precise inventory of forest resources across large scales.

The operational uptake of forest inventories based on individual tree attributes has partly been limited by the bias induced by over-segmentation or under-segmentation errors associated to the ITC delineation process (Breidenbach et al., 2010b; Vastaranta et al., 2012; White et al., 2016). Semi-ITC approaches, in which detected segments can be matched to one or multiple trees have been used to reduce this bias (Breidenbach et al., 2010b; Breidenbach and Astrup, 2014). Although the range of forest types on which the development and accuracy of ITC algorithms is being assessed is increasing (Kaartinen et al., 2012),

complex forest types in which a range of species and conditions coexist are limiting the application of ITC methods across large areas (Fassnacht et al., 2016; Zhen et al., 2016). Regardless, in the context of species mapping, there is no demonstrated gain in the accuracy of forest attribute prediction when using ITC or semi-ITC approaches compared to an ABA. For example, Ørka et al. (2013) used ALS and high-resolution multispectral and hyperspectral imagery to estimate the proportion of spruce, pine and deciduous tree species in a boreal forest of Norway with a $R^2 = 0.66 - 0.96$ when using an ITC or semi-ITC approach and a $R^2 = 0.83 - 0.89$ when using an ABA. The three approaches provided estimates with a relatively similar accuracy, but the ITC particularly outperformed the ABA to estimate the proportion of hardwood species that were not as frequent as pine and spruce in their study area. There are few studies however that have compared the relative performance of species predictions generated with area-based and ITC approaches and additional research quantifying differences in a range of forest environments would be informative.

The area-based predictions of species groups had more spatial detail than the stand-level estimates from the photo-interpreted inventory. The overall stand-level patterns of species distributions were also relatively well captured, except in stands of the mixed conifers species groups where misclassifications of black spruce or mixedwood dominated. It is important to consider that the photo-interpreted data should not be necessarily considered as reference data. Differences between the predicted species groups and the ones available in the forest inventory polygons could be explained by errors in the area-based predictions, photo-interpretation errors or time span between the photo-interpretation and the SPL100 acquisition. By using consistent data sources and methods, species predictions with remotely sensed data also avoid issues associated with subjective interpretation, particularly when many photo-interpreters are involved as is common when inventories span large forest management units.

The derivation of the aggregated crown metrics developed herein is more computationally intensive than the standard area-based summary metrics, which is an important consideration for operational uptake. However, the ITD and ITC algorithms used were applied on the raster-based CHM, which is faster than applying these algorithms directly on the point cloud. Moreover, no predictions were made at the individual

tree level since ITC metrics were aggregated, which is an important implication for processing time in large areas such as the Romeo Malette Forest. Area-based species information, together with area-based forest inventory attributes could be used to automatically delineate homogeneous forest stands similar to photo-interpreted stands as part of the forest inventory. In addition, the wall-to-wall species group predictions could be used to select which optimal set of ITC method and parameters should be used for a stand of interest.

A relatively marginal gain in species group classification OA was observed when combining ALS and Sentinel-2 predictor variables compared to using ALS metrics only, particularly for the 3-groups species classification (Figure 4-2). The Romeo Malette Forest is actively managed and some of the structural differences captured by the ALS summary metrics could be related to past silvicultural treatments targeting specific stands or species groups. In unmanaged forests, ALS metrics, in particular those related to height, could be less important and the gain in prediction accuracy obtained by combining ALS and Sentinel-2 data could be larger than the one observed herein. Sentinel-2 data are global, free and publicly available and their integration into the generation of EFI products is relatively straightforward. As discussed above, few studies have examined the integration of ALS and Sentinel-2 imagery for forest inventory products and further research, particularly in areas with more diverse species composition and different management regimes, would be interesting to better understand the benefits of integrating such data into forest inventory frameworks.

Chapter 5. Comparing airborne and spaceborne photon-counting lidar canopy structural estimates across different boreal forest types⁵

5.1. Introduction

The two previous chapters focused on the inventory of key forest attributes for forest management purposes. This chapter marks the transition from a management perspective towards a broader monitoring perspective by introducing ICESat-2 data and examining their ability to characterize different aspects of forest structure.

To date, the characterization and monitoring of forests across broad geographic extents, such as the 347 Mha of Canadian forests, has mostly been limited to utilizing data from satellite optical imagery. Optical imagery is well-suited to map attributes such as forest cover (Hansen et al., 2013), phenology (Bolton et al., 2020) and abiotic and biotic disturbances (Coops et al., 2018; Goodwin et al., 2008) but do not provide a direct measure of forests vertical structure as opposed to spaceborne lidar measurements. The GLAS instrument on board ICESat, launched in 2003 and decommissioned in 2010, used full waveform measurements along the laser line of sight to retrieve the vertical structure of the vegetation. GLAS offered the first opportunity to characterize canopy height (Lefsky et al., 2007; Sun et al., 2008) and biomass (Margolis et al., 2015) from spaceborne lidar measurements. Global maps of canopy height (Lefsky, 2010; Simard et al., 2011) and biomass (Hu et al., 2016) were also produced in combination with optical satellite imagery. In December 2018, GEDI (Dubayah et al., 2020), another full waveform instrument, was launched and installed on the international space station. GEDI was designed specifically for vegetation structure

⁵ The content of this chapter has been adapted from:
Queinnec, M., White, J. C., & Coops, N. C. (2021). Comparing airborne and spaceborne photon-counting lidar canopy structural estimates across different boreal forest types. *Remote Sensing of Environment*, 262, 112510.

characterization through canopy height retrievals to derive attributes such as canopy height (Potapov et al., 2021), structural diversity (Schneider et al., 2020), and biomass (Duncanson et al., 2020) within the latitudinal limits of the international space station orbit (i.e. $\pm 52^\circ$ latitude).

The ATLAS instrument onboard ICESat-2 is the first spaceborne lidar system based on SPL technology, which enable to reduce the laser power requirements and increase the sampling spatial resolution along the satellite track (Neumann et al., 2019). However, a range of limitations are associated with ATLAS measurements. The detection of ground or canopy is probabilistic by nature, as the likelihood of detecting a surface within the laser footprint depends on its reflective properties (Neuenschwander and Pitts, 2019). This is a particularly important consideration as it is anticipated that only 0 – 4 photons will be returned either from the ground or the canopy for each ATLAS footprint of 11 m diameter (Neuenschwander and Pitts, 2019). The photon-sensitive photomultiplier tubes of ATLAS are also detecting photons originating from solar radiation and atmospheric scattering and require advanced algorithms to separate photons resulting from true surface reflection versus those photons resulting from background noise—particularly for daylight acquisitions (Neuenschwander and Pitts, 2019; Popescu et al., 2018; Tang et al., 2016). As a result, limitations associated with ATLAS single-photon detection could result in a signal that is not necessarily representative of the vegetation structure.

Simulated ICESat-2 data products such as the global geolocated photon cloud (ATL03) and the land and vegetation height product (ATL08) have been used prior to launch to examine the filtering and classification of raw photons as well as the retrieval of canopy and terrain height (Gwenzi and Lefsky, 2014; Neuenschwander and Magruder, 2016; Popescu et al., 2018), and to derive estimates of biomass (Montesano et al., 2015; Narine et al., 2019b). (Narine et al., 2019b) An initial comparison of ICESat-2 land and vegetation height product (ATL08) canopy height estimates to reference ALS-derived canopy height estimates in Finland showed strong agreement ($R^2 = 0.99$, RMSE = 0.85 m; Neuenschwander and Magruder, 2019). Further ICESat-2 validation across southern Finland showed that canopy height was underestimated on average by 0.5 m – 0.6 m and that it was more accurate in forests with canopy cover

ranging between 40% and 85% (Neuenschwander et al., 2020a). However, there is still need to characterize the type of information that ICESat-2 data is providing regarding vegetation structure across a range of forest types (Neuenschwander and Pitts, 2019). For example, an estimate of canopy cover is planned to be delivered in ICESat-2 ATL08 data product (Neuenschwander and Pitts, 2019) but is not yet available. Moreover, to date, comparisons between ICESat-2 ATL08 and ALS data have used conventional linear-mode lidar data rather than airborne SPL data (Neuenschwander et al., 2020a).

The overall objective of this chapter is to compare coincident ICESat-2 geolocated and classified photon cloud data (ATL03 and ATL08 data products) and airborne SPL100 data acquired over the Romeo Malette Forest, where a range of boreal forest structural types are observed. The analysis aims at providing novel insights on the following questions:

- (1) How do individual ICESat-2 photons reflected from the top of canopy compare to coincident SPL100 point cloud height percentiles?
- (2) How do estimates of canopy cover and height variability derived from ICESat-2 and SPL100 data compare?
- (3) Are the relationships between estimates of canopy height, cover, and variability derived from ICESat-2 and SPL100 data consistent across a range of boreal forest structural types?

In addressing these questions, this chapter contributes to the characterization of ICESat-2 data for vegetation structure retrieval by comparing it to coincident SPL100 data acquired over a range of boreal forest types.

5.2. Methods

This study compared the SPL100 and ICESat-2 point clouds across different forest types observed at the Romeo Malette Forest, which is described in section 2.1.1 of this dissertation. The overall methodology can be broken down in three main steps (Figure 5-1). First, the field data collected at the ground plots selected following the SGS approach described in Chapter 1 (sections 2.3; 3.2.3) were used to identify and

delineate forest stands with relatively homogeneous forest composition and structure and assign those stands to three broad structural groups. The second step consisted in combining the ATL03 and ATL08 data products to normalize the ATL03 photons to heights above ground and classify them as either noise, ground, canopy and top of canopy (using the ATL08 product). In addition, the height normalized and classified ATL03 photons were used to identify continuous segments of ATL03 data that intersected the forest stands identified in the first step and were used as analysis units (AUs). Thirdly, the SPL100 point cloud was clipped to each AU extent and was then compared and analyzed against the ATL03 data (i) at the photon level, and (ii) at the AU level. All data processing and analysis were performed within the R environment (R Core Team, 2022).

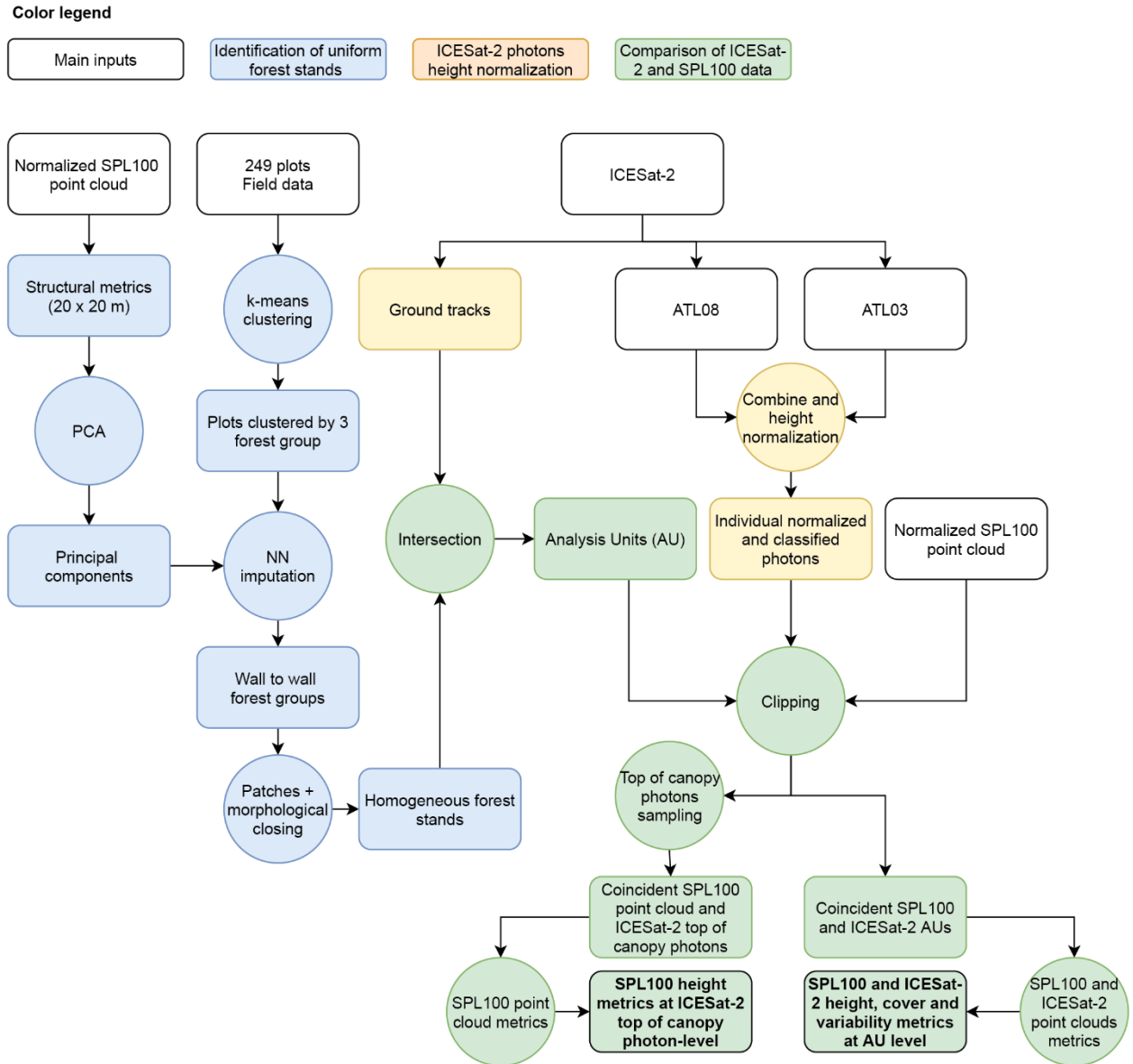


Figure 5-1. Flow diagram illustrating the main steps of the methodology applied in this study.

5.2.1. Delineation of homogeneous forest stands

The field measured 90th percentile and standard deviation of tree heights, stem density, percentage of coniferous tree species (as opposed to broadleaf species), mean diameter at breast height (DBH), crown closure, and tree species and height diversity, derived from the Shannon Diversity Index ($-\sum_i(p_i \times \log(p_i))$ where p_i is the proportion of each individual tree species i in a plot), were calculated at the plot level from the field inventory data. These attributes were then used to group the plots into three

groups using a k-means clustering algorithm to capture the main structural types and maturity classes (that is, stand establishment and regrowth, mature stands, and overmature and declining stands) observed at the Romeo Malette Forest. The three groups were composed of 133, 83, and 33 plots, respectively. The groups were then spatially extrapolated across the Romeo Malette Forest using a random forest-based nearest neighbor imputation (Crookston and Finley, 2008). The six first principal components of the SPL100 structural metrics, derived during the plot network establishment (section 2.4), were used as predictor variables. The rumple index (Kane et al., 2008), which characterizes the canopy surface roughness was also calculated for each plot and used as an indication of structural complexity of the three forest groups. The rumple index was calculated with the *lidR* package (Roussel et al., 2020), by triangulating the highest first returns falling within a 1 x 1 m grid, forming the surface of the canopy.

Contiguous patches with a minimum of 100 connected 20 x 20 m cells sharing the same structural group label were then identified and a morphological closing operation with a 3 x 3 moving window was then performed in order to remove small holes within the patches. Finally, the patches were converted to polygons to represent a set of distinct forest stands. In total, 4935 stands were identified with a median area of 13.6 ha. Due to active forest management practices and other types of disturbances such as wildfires, the structure of the forest changed between the SPL100 acquisition in summer 2018 and the analyzed ICESat-2 acquisition window (October 2018 – November 2020). In order to avoid comparing ICESat-2 and SPL100 vegetation structure estimates in areas that experienced past forest disturbances, differences in Normalized Difference Vegetation Index (dNDVI) and differences in Normalized Burn Ratio (dNBR) derived from two cloud-free Sentinel-2 Level-2A imagery, acquired on the 24th of June 2018 and the 11th of September 2020 were examined. Areas with abrupt changes in forest structure were identified using dNDVI and dNBR thresholds and subsequently masked out of the analysis.

5.2.2. ICESat-2 data processing and definition of analysis units

The ICESat-2 data used in this chapter consisted in the Level-2 Global Geolocated Photon Data product (ATL03) and the Level-3 Land and Vegetation Height data product (ATL08), which were also introduced

in section 2.2.2. The ATL03 product provides geophysical-corrected latitude, longitude, and height above the WGS84 ellipsoid of each downlinked photon (Neumann et al., 2020) and is the source for all higher level products. The ATL08 product classifies ATL03 photons as noise, ground, canopy, or top of canopy (Neuenschwander and Pitts, 2019). All available version 003 ATL03 (Neumann et al., 2020) and version 003 ATL08 (Neuenschwander et al., 2020b) granules covering the Romeo Malette Forest between October 2018 and November 2020 were obtained from the NASA National Snow and Ice Data Center.

Although the ATL08 algorithm classifies the ATL03 photons as noise, ground, canopy, or top of canopy, the coordinates of the photons are not stored within the ATL08 product. Therefore, the ATL03 and ATL08 products were combined to obtain an ICESat-2 point cloud composed of individual geolocated and classified photons (Figure 2-9; Figure 2-10B). Only photons collected by ATLAS strong beams (1 beam per pair), more suitable for the detection of ground and canopy (Neuenschwander et al., 2020a), were considered in the analysis. In order to avoid snow or atmospheric scattering to affect the results due to their influence on ICESat-2 ground or canopy sampling probability, only transects not covered by snow and with no atmospheric scattering were retained using the *segment_snowcover* and *msw_flag* parameters provided in the ATL08 data product. The *segment_snowcover* is derived from the daily NOAA Global Multi-sensor Snow/Ice Cover map (Palm et al., 2020) and indicates the presence of snow or ice on the ground. A *segment_snowcover* value of 1 indicates snow-free land. The presence of atmospheric scattering is indicated by the *msw_flag* in the ATL08 product, which has a value of 0 if not scattering is detected (Neuenschwander et al., 2020a). Additionally, the number of ground or canopy (including ATL08 canopy and top of canopy classes) photons per outgoing ATLAS pulse were calculated. This is defined as radiometry (Neuenschwander et al., 2020a) and gives an indication of the strength of the returned signal.

In order to obtain a height normalized ATL03 product, a continuous ground surface was linearly interpolated from the ATL03 returns classified as ground. Since data gaps can occur along the ground track, the interpolated ground surfaces were generated only for segments where the along-track distance between two ground photons was less than 10 m. The normalized ATL03 photon height was then calculated by

subtracting the interpolated absolute ground height from the absolute photon height above the WGS84 ellipsoid.

Segments were generated by the intersection of the homogeneous forest stands with the filtered ICESat-2 ground tracks (Figure 5-2). Only the segments with a length > 50 m and < 300 m were retained for further analysis. To generate the AUs, the centre lines of the segments were buffered by a diameter of 11 m to simulate the extent of ATLAS footprint. All normalized ATL03 photons within each of these AUs were extracted and the SPL100 point cloud was clipped to the extent of the AUs. Table 1 provides an overview of the AUs characteristics.

Table 5-1. Characteristics of the analysis units (AUs). The standard deviation (SD) is indicated in parentheses.

Forest group	Number of AUs	Mean AU length (SD) (m)	Mean AU ground radiometry (SD)	Mean AU canopy radiometry (SD)
All combined	846	117 (60)	1.03 (0.83)	1.10 (0.35)
1	154	111 (57)	1.00 (0.78)	1.11 (0.32)
2	296	113 (60)	0.88 (0.65)	1.20 (0.34)
3	396	121 (62)	1.16 (0.94)	1.03 (0.35)

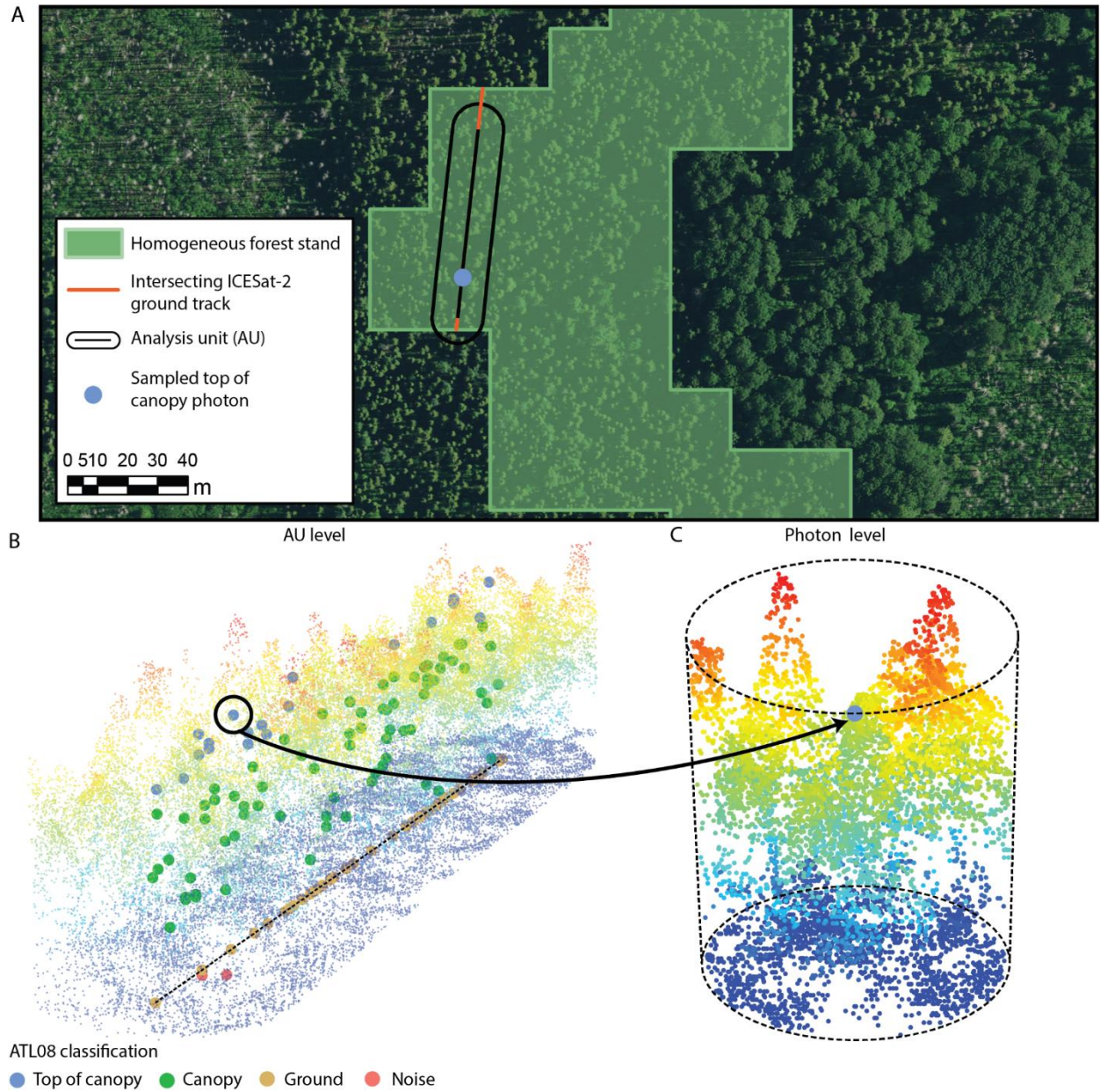


Figure 5-2. Illustration of the analysis units (A) and their analysis at the analysis unit level (B) and photon level (C). Small dots in B and C represent the normalized SPL100 point cloud. Large dots in B represent the normalized ATL03 photons located along the ICESat-2 ground track.

5.2.3. Comparison at the photon level: ATL03 top of canopy height and SPL100 height metrics

For each of the three forest groups, ATL03 photons classified as top of canopy were randomly sampled ($n = 300$ for each group) from the AUs. A minimum horizontal distance of 50 m between the sampled photons was required to avoid considering photons returned by the same tree and minimize the effect of spatial autocorrelation. The SPL100 point clouds matching the AUs' extent were then clipped to a circle of 11 m around the sampled top of canopy photons (Figure 5-2C). Height percentiles of all non-ground SPL100 returns within the circular footprint around the sampled ATL03 top of canopy photons were calculated. The normalized height of the sampled ATL03 top of canopy photons was then compared to the SPL100 height metrics and the Pearson correlation coefficient (r), root mean square difference (RMSD) and mean absolute difference ($MAD = 1/n \sum_i^n (|ATL_i - SPL_i|)$, where ATL and SPL are ICESat-2 and SPL100 height estimates and n the number of sampled photons) were calculated. RMSD and MAD were also calculated relative to the mean of SPL100 values (RMSD%, MAD%). Additionally, the difference between ICESat-2 top of canopy photons height and “true” canopy height was quantified by calculating the difference between SPL100 99th percentile of return heights, considered as a reference canopy height estimate, and the SPL100 height percentile that had the strongest agreement with ICESat-2 top of canopy photon height.

5.2.4. Comparison at the AU level: ATL03 and SPL100 estimates of canopy cover, height, and height variability

As canopy cover and height variability are evaluated from the distribution of photon heights within a spatial unit, the comparison of their SPL100 and ICESat-2 estimates was made at the AU level as opposed to the ATL03 top of canopy photon level. Two different approaches to derive lidar estimates of canopy cover (CC) were compared. Conventionally, ALS-derived CC is estimated by determining the percentage of first returns above a specified height threshold, commonly 2 m (White et al. 2013) which was the approach applied to calculate CC from the SPL100 data. This same approach was applied to the normalized ATL03 canopy and top of canopy photons within the AU by calculating the percentage of canopy and top of canopy

photons above 2 m. As an alternative approach, the vegetation fill index (VF), which was originally used in Wulder et al. (2009) as a 2-dimensional proxy of CC for airborne profiling lidar data was also calculated. VF represents the proportion of area under the canopy profile at a fixed height threshold along the AU and was calculated as follows:

$$VF_{th} = \frac{Area_{canopy}}{Area_{section}}$$

with

$$Area_{canopy} = \frac{1}{2} \sum_{i=1}^n [s_i \times (\min(h_{th}, h_{i-1}) + \min(h_{th}, h_i))]$$

and

$$Area_{section} = l \times h_{th}$$

where n is the number of photons from which the canopy profile is retrieved, s_i the horizontal distance between two consecutive photons from the canopy profile, h_{th} is a fixed height threshold, h_i is the height of the canopy profile photon i and l is the AU length.

Since the objective was to characterize VF under the top of canopy profile, h_{th} was set to the maximum height of SPL100 or ATL03 photons (excluding the ATL03 photons labeled as noise) within the AU (Figure 5-3). The canopy profiles of ICESat-2 and SPL100 were retrieved by locating the highest return classified as ground, or vegetation (canopy or top of canopy for ICESat-2) within a moving window along the AUs. The small point spacing of the SPL100 point cloud along the AU enabled the use of a relatively narrow window width (0.5 m) to retrieve the canopy profile (see Figure 5-3). However, since the spacing between ATL03 photons is larger and gaps between consecutive photons can be present, a variable along-track window width of 1 to 11 m was considered to identify ATL03 highest returns and retrieve the canopy profile.

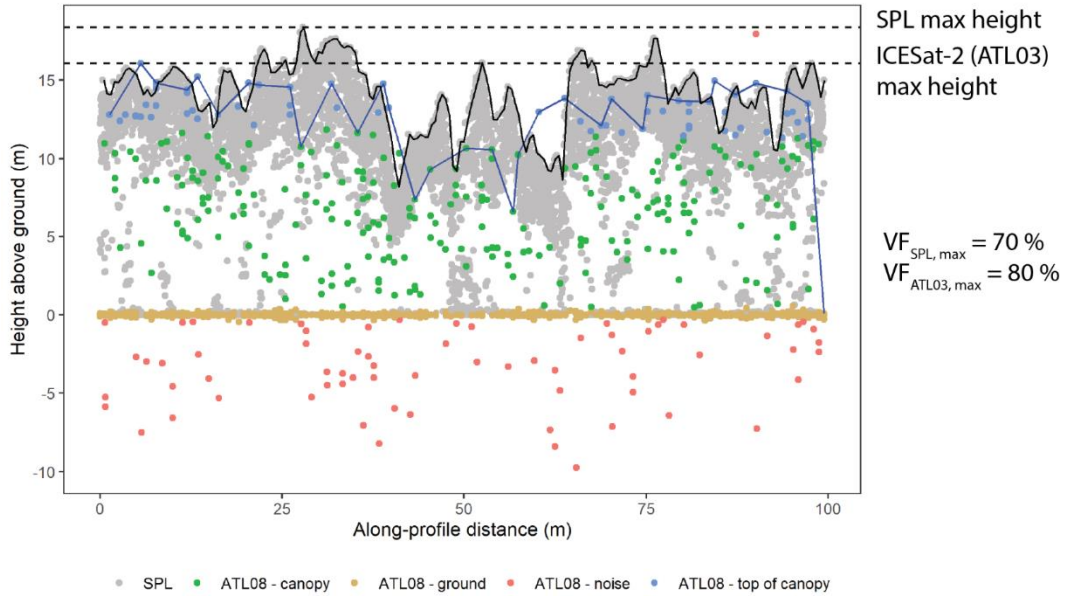


Figure 5-3. Transect of coincident normalized SPL100 and ICESat-2 point clouds. The canopy profile of SPL100 and ICESat-2 are displayed with the solid black and blue lines respectively. This figure illustrates the retrieved ICESat-2 canopy profile from the highest photons within a 3 m window in the along-profile direction. VF = Vegetation Fill; SPL = SPL100.

Lastly, the height variability of ICESat-2 and SPL100 profiles was assessed by calculating the standard deviation and coefficient of variation (standard deviation standardized by the mean) of returns height as well as the 25th, 50th, 75th and 95th percentiles of canopy and top of canopy photons height.

As for the aforementioned photon level canopy height comparison, the relationship between ICESat-2 and SPL100 canopy cover and height variability was assessed by calculating the Pearson correlation coefficient (r), RMSD, MAD, as well as their relative values. For the ICESat-2 measure of VF, the along-track window width that maximized r between ICESat-2 and SPL100 VF was retained.

5.3. Results

5.3.1. Attributes of forest groups

The three forest groups classified by the k-means algorithm were characterized by relatively distinct attributes (Figure 5-4). Overall, the three groups matched the stand development stages that are expected to

be found in this managed boreal forest as discussed previously (Oliver and Larson, 1996). Plots representing group 1 were mostly composed of mature (38%), overmature (35%), and declining (26%) field-assessed maturity classes, with larger tree height (18 – 38 m) and DBH (14 – 35 cm) than in the two other groups and largest variability in tree height. Plots in group 1 were also generally more diverse in terms of tree species, with a median species diversity index of 0.30 compared to 0.16 and 0.14 for groups 2 and 3 respectively and had a larger proportion of broadleaf species. Plots identified as being mature represented 83% of the plots assigned to group 2 which had taller trees ranging between 8 and 25 m and the largest stem density (450 – 3650 stems/ha) among the three groups. Group 3 conversely showed a higher proportion of stands experiencing regeneration (46%) or growth (18%), with a relatively low tree height (4 – 27 m), stem density (50 – 1300 stems/ha), DBH (7 – 23 cm), and crown closure (0 – 65%). SPL100 profiles of exemplars from each forest group are provided in Figure 5-5 for illustration.

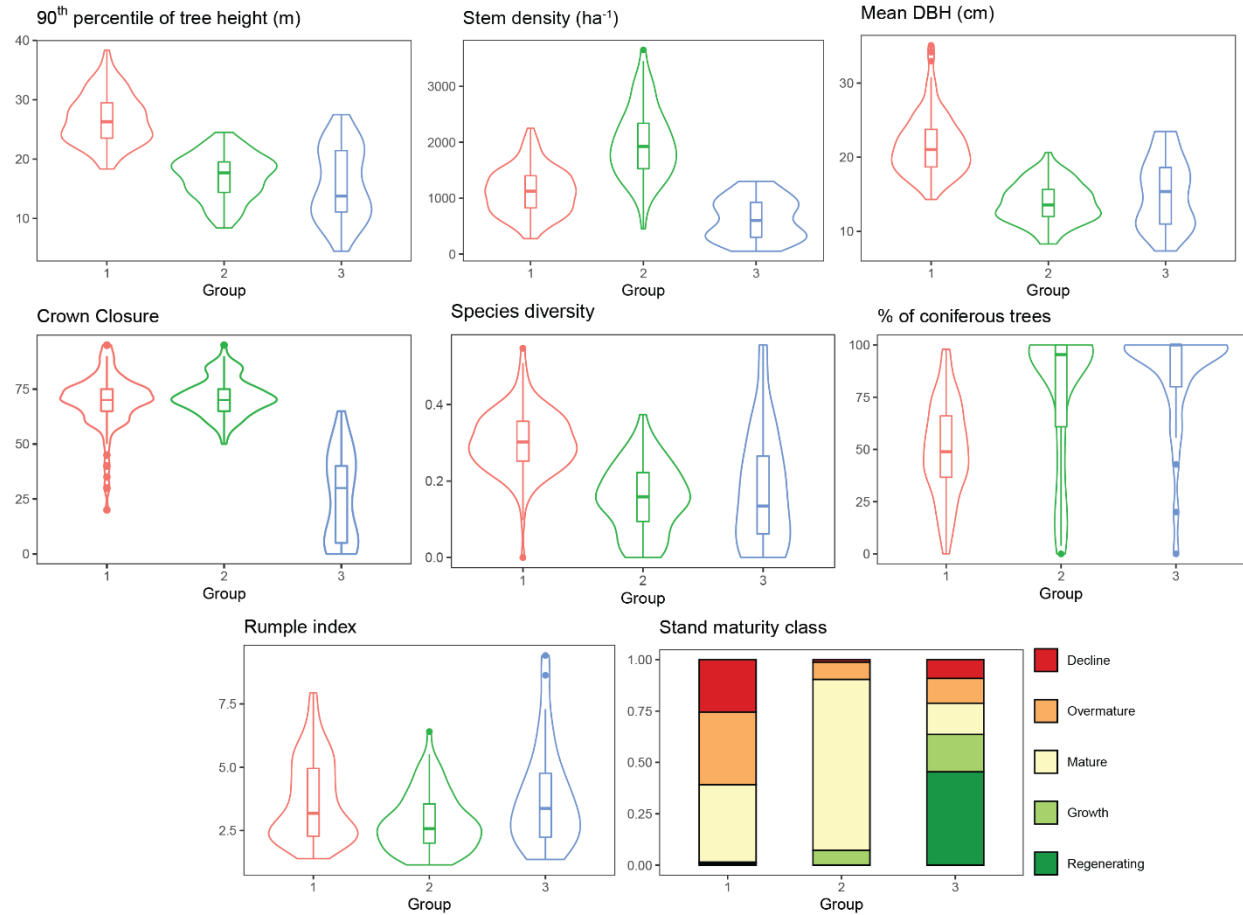


Figure 5-4. Violin plots and boxplots show the distribution of plot level field data by forest group, determined from k-means clustering. Groups 1, 2 and 3 are composed of 133, 83 and 33 plots respectively. All the attributes, except the rumple index and the stand maturity class, were used for the k-means clustering.

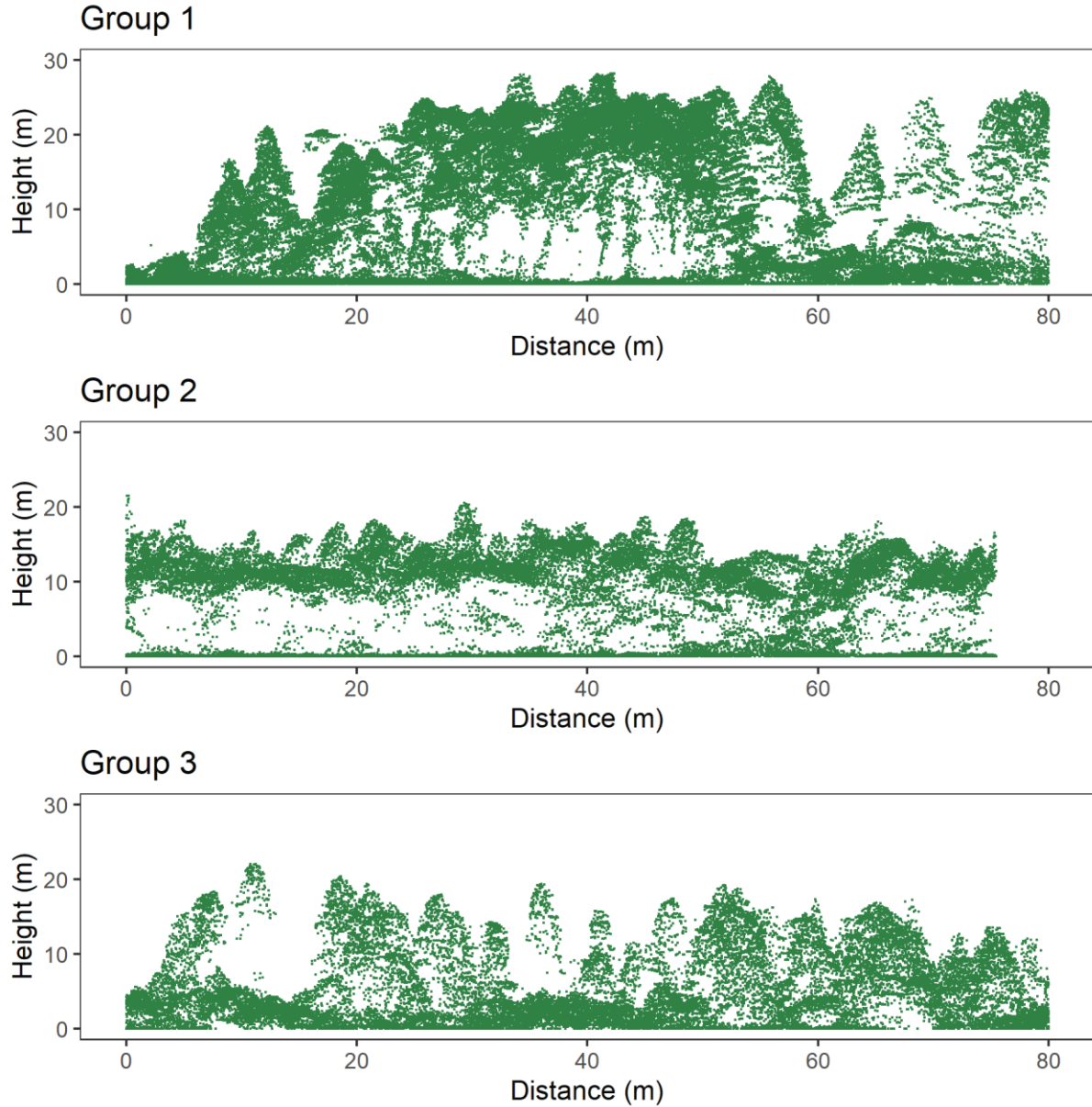


Figure 5-5. Normalized SPL100 profiles along analysis units illustrating the typical structure of the three forest groups. Each profile has a width of 11 m in the across-track direction.

5.3.2. Comparison of ATL03 top of canopy photon height to SPL100 height percentiles

Figure 5-6A shows the correlation and RMSD% between ATL03 individual top of canopy photons height and SPL100 height percentiles. All groups achieved their maximum correlation around the 90th and 95th SPL100 height percentiles. However, the relationship was notably weaker for group 1 ($r = 0.64$) than for

group 2 ($r = 0.73$) and group 3 ($r = 0.84$). Minimum RMSD% of 17.6%, 20.0%, and 20.2% were observed at the 90th SPL100 height percentile for groups 1, 2, and 3, respectively. RMSD% was markedly larger for group 3 (>50%) at the lower height percentiles. Figure 5-6B shows the scatterplots between ATL03 top of canopy photon heights and the SPL100 90th height percentile, for which the highest correlation and the lowest RMSD were observed when considering all groups combined.

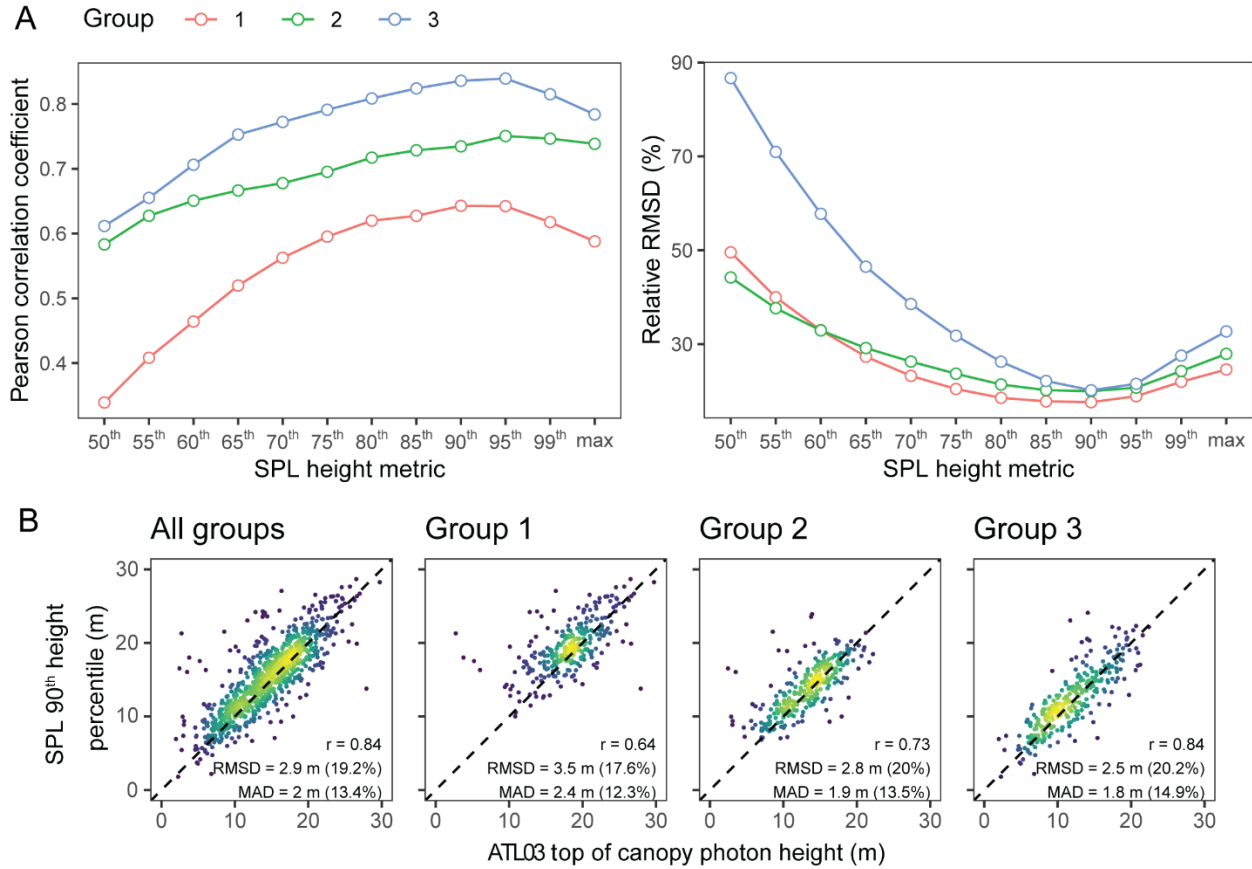


Figure 5-6. (A) Variation of Pearson correlation coefficient and RMSD per forest group and SPL height metric. (B) Scatterplots showing the relationship between ATL03 top of canopy photons height and SPL100 90th height percentile for all groups combined and per group. Plots are colored by point density on a blue (low density) to yellow (high density) color scale. The Pearson correlation coefficient is labelled as r . SPL = SPL100.

The average difference between the SPL100 99th percentile of returns height, considered as the best estimate of canopy height within the footprint from which the ATL03 top of canopy photon was returned, and the SPL100 90th percentile was of 2.3 m, 2.0 m and 2.7 m for groups 1, 2 and 3 respectively. Relative to the

mean of SPL100 99th percentile in each group, this corresponded to an average underestimation of canopy height by ATL03 top of canopy photons of 10.3%, 12.7% and 18.2% for groups 1, 2 and 3.

5.3.3. Comparison of ATL03 and SPL100 measures of canopy cover at the AU level

The comparison of CC resulted in a correlation coefficient of 0.51, a RMSD% of 23.8%, and a MAD% of 17.9% for all groups combined (Figure 5-7A). Most SPL100 measures of CC were contained between 0.75 and 0.95 for groups 1 and 2, which are characterized by mature and old forests with large trees and high crown closure, while the ICESat-2 equivalents exhibited more variability in cover estimates. This resulted in poor correlation coefficients of 0.26 and 0.07 and RMSD% of 21.8% and 21.6% for groups 1 and 2 respectively. SPL100 measures of CC in group 3, dominated by stands not having reached the crown closure stage, had a larger range and exhibited stronger correlation with ICESat-2 measure of CC ($r = 0.61$) but higher RMSD% (26.8%).

Highest correlation and lowest RMSD% between SPL100 and ICESat-2 measures of VF were achieved for an along-track window width of 6 m from the simulated width ranges of 1 m to 11 m ($r = 0.53 - 0.57$; RMSD% = 19.7% - 44.5%). The following results are therefore reported for ICESat-2 VF measures obtained with an along-track window of 6 m. In comparison to CC, similar trends were observed when comparing SPL100 and ICESat-2 VF measures across the groups but with highest correlation ($r = 0.57$ all groups combined; $r = 0.32 - 0.57$ across the groups) and lower RMSD% (19.7% all groups combined; 17.4% - 22.8% across the groups). Visual assessment of the scatterplots of CC (Figure 5-7A) and VF (Figure 5-7B) show a relatively better overall relationship between SPL100 and ICESat-2 measures of VF compared to CC, with more points aligned along the 1:1 line especially for values ranging between 0.5 and 0.9. Regardless of the method used to estimate canopy cover, Figure 5-7 indicates two notable trends. First, the SPL100 estimated a relatively limited range of canopy cover in groups 1 and 2 (60–90%). Second, the ICESat-2 has a tendency to underestimate cover relative to the SPL100, particularly for groups 1 and 2.

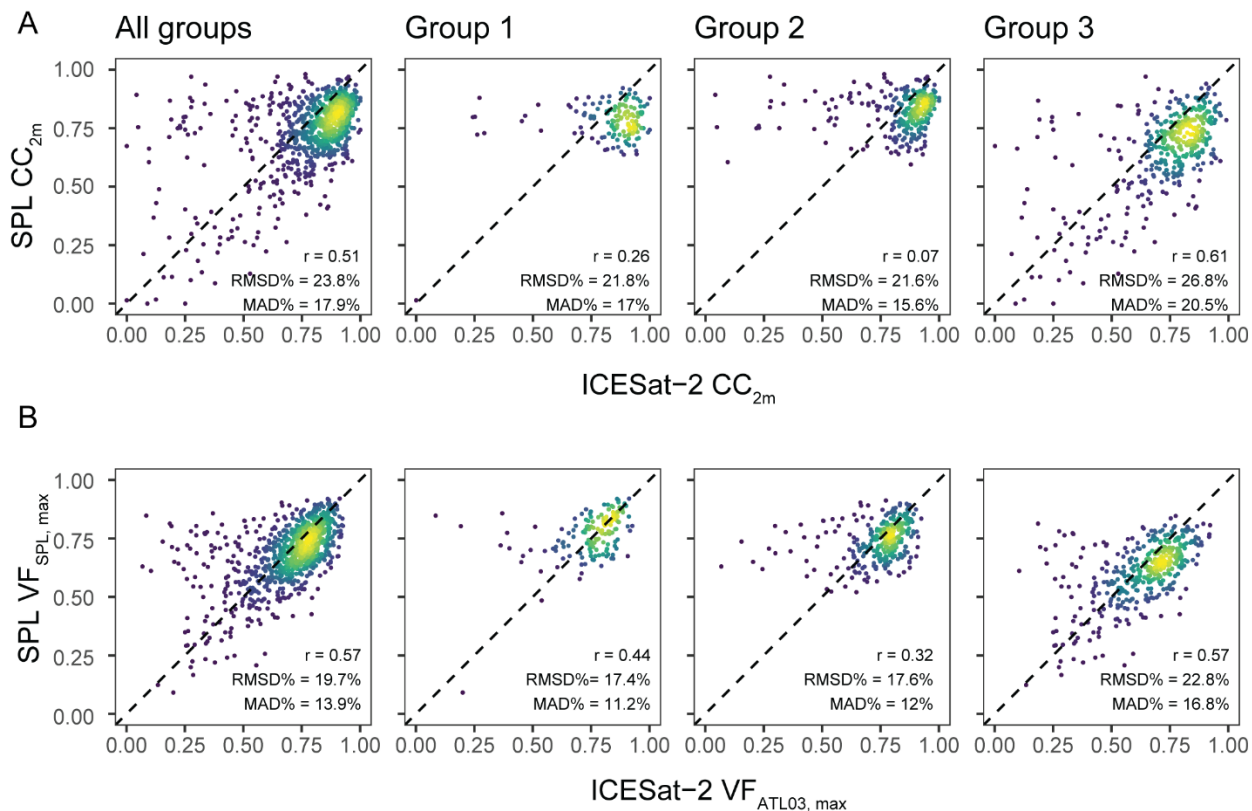


Figure 5-7. Scatterplots showing the relationship between ICESat-2 and SPL100 measure of canopy cover (A) as the percentage of returns above 2 m (CC_{2m}) and (B) as a measure of vegetation fill at maximum return height (VF_{max}). Plots are colored by point density on a blue (low density) to yellow (high density) color scale. The Pearson correlation coefficient is labelled as r . SPL = SPL100.

5.3.4. Comparison of ATL03 and SPL100 height and height variability at the AU level

For the AUs, the strength of the relationship between SPL100 and ATL03 height percentiles increased with increasing height percentiles. Overall, at the 25th percentile, correlation was 0.47 and RMSD% was 77.8% (MAD% = 53.6%), whereas at the 95th percentile, correlation was 0.74 and RMSD% was 23.7% (MAD% = 12.5%) (Figure 5-8). Among groups, RMSD% and MAD% were consistently lower for group 2 compared to groups 1 and 3. The strength of the relationship between the 25th and 50th height percentiles for groups 1 and 2 is in part affected by a number of AUs with ICESat-2 values close to 0 while SPL100 covers a larger range of values. Similar to results comparing height metrics at the photon level, ICESat-2 consistently underestimated canopy heights relative to the SPL100 data.

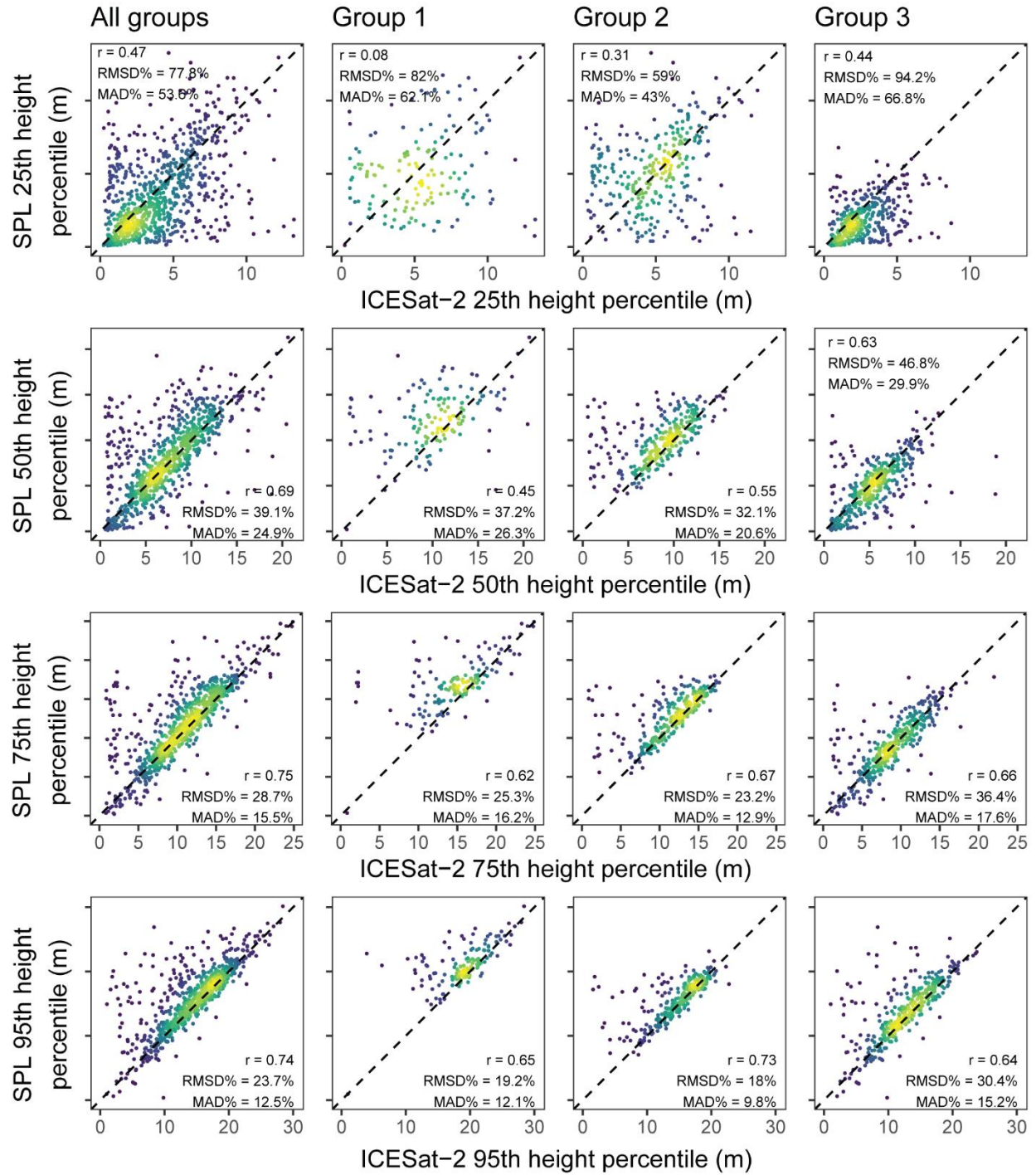


Figure 5-8. Scatterplots showing the relationship between SPL100 and ATL03 25th, 50th, 75th and 95th photons height percentiles at the AU level. SPL = SPL100.

Trends observed for the comparison of SPL100 and ATL03 photon height standard deviation (Figure 5-9) were very similar to those of the 95th height percentiles (Figure 5-8). Generally, ICESat-2 underestimated the variability in height relative to the SPL100 data. RMSD% and MAD% were similar for groups 1 and 3 and lower for group 2. The relationship between the two measures of coefficient of variation were notably weaker with larger differences between the two data sources ($r = 0.09 - 0.51$; RMSD% = 31.6% - 39.9%; MAD% = 22.7% - 26.3%). This suggests that the ICESat-2 and SPL100 have similar absolute variability in canopy heights (SD), but differ in the relative dispersion of canopy heights (CV). SPL100 generally has greater relative dispersion of canopy height compared to the ICESat-2, particularly for group 3.

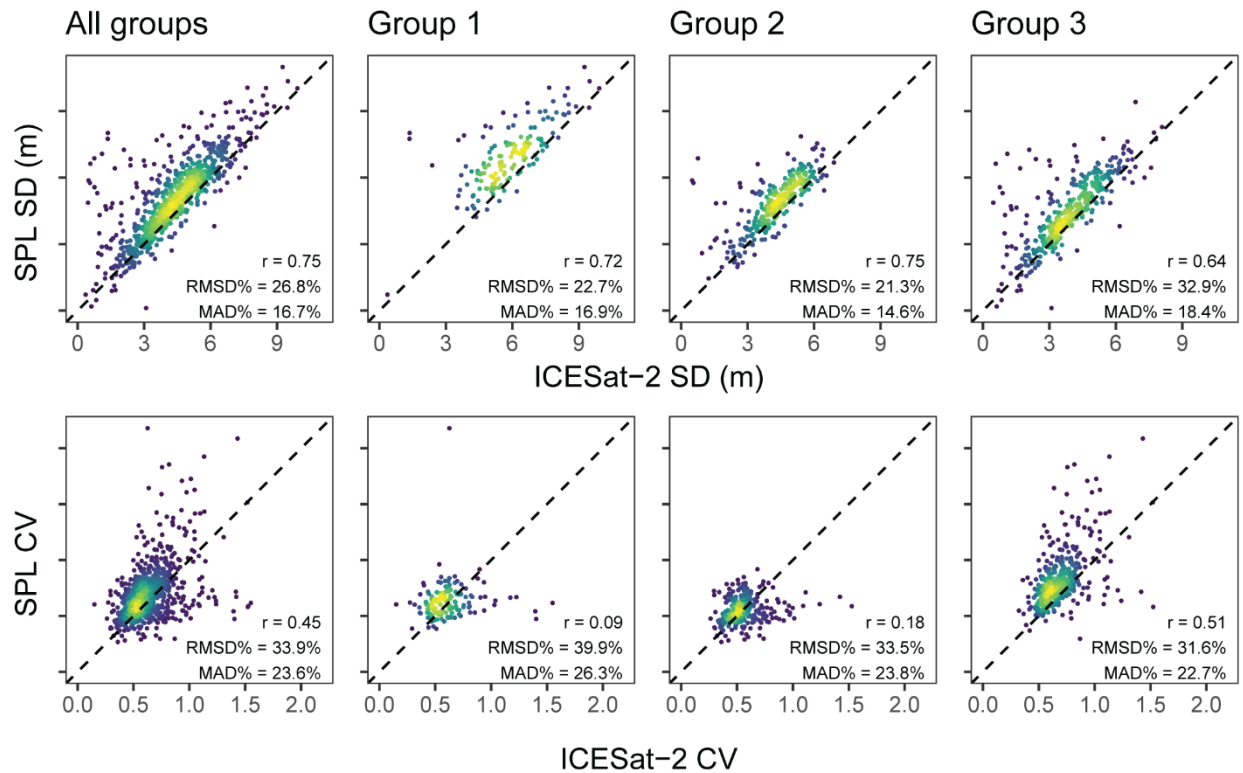


Figure 5-9. Scatterplots showing the relationship between SPL100 and ATL03 standard deviation (SD) and coefficient of variation (CV) at the AU level. SPL = SPL100.

5.4. Discussion

To date, given the recent launch of ICESat-2, most studies have used simulated ATL03 and ATL08 data products to assess the performance of ICESat-2 at characterizing vegetation structure and aboveground biomass (Gwenzi and Lefsky, 2014; Montesano et al., 2015; Narine et al., 2019b). The comparison of actual ATL08 data to ALS data has been limited to the retrieval of terrain and canopy height at the ATL08 product level (default 100 m segment; A. Neuenschwander et al., 2020; Neuenschwander and Magruder, 2019). In addition to canopy height, this study examines the retrieval of lidar estimates of canopy cover and vegetation height distribution from acquired ATL03 and ATL08 data at both the individual top of canopy photon level and along AUs of variable length intersecting homogenous forest stands that were defined using an extensive ground plot network. Moreover, ICESat-2 data were compared against airborne SPL data, inherently based on the same measurement principle, i.e., single-photon sampling, as the ATLAS instrument onboard ICESat-2.

5.4.1. Comparison of ATL03 top of canopy photon height to SPL100 height percentiles

Overall, the height of ICESat-2 ATL03 photons identified by the ATL08 algorithm as top of canopy, related most strongly to 90th height percentiles of coincident SPL100 footprints ($r = 0.84$, $MAD\% = 13.4\%$; Figure 5-6). Similar level of agreement between the ICESat-2 canopy heights and the SPL100 (RMSD = 2.9 m) was reported by Neuenschwander and Magruder (2019) (RMSE = 3.2 m) for canopy height retrievals in the boreal forests of southern Finland at the ATL08 100 m segment level. Results also correspond well to Popescu et al. (2018) who reported an average canopy height RMSE of 3.2 m using simulated datasets in a range of forested ecosystems and noise levels. Of note, photon-level comparisons were performed in this study, which are likely more influenced by geolocation errors than segment-level comparisons.

Moreover, results indicate that there is variability in the level of agreement between ICESat-2 top of canopy photons height and SPL100 height percentiles by forest structural groupings. Strongest correlation was observed for the stands with intermediate crown closure levels (group 3; $r = 0.84$) compared to stands with consistently higher crown closure (groups 1 and 2; $r = 0.64 - 0.73$). This is consistent with findings from

Neuenschwander et al. (2020a) who observed best agreement between ATL08 and ALS canopy height for canopy cover ranging between 40% and 85%. Lidar is known to under-sample the canopy and therefore underestimate the true canopy height (Næsset, 1997; Nelson et al., 1988). Dense lidar point clouds, such as the SPL100 data used herein (40 points/m²), mitigate this under-sampling issue but the larger footprint and lower energy of ICESat-2 ATLAS beams decrease the likelihood of having photons detecting tree tops (Neuenschwander and Magruder, 2016). As discussed by Montesano et al. (2015), this effect could be further emphasized in sparser forests, such as in northern Taiga/Tundra ecosystems. In denser forest types with high canopy cover, the retrieval of terrain, which is essential for accurate tree height estimates is more limited (Neuenschwander et al., 2020). The correlation between SPL100 height percentiles and ICESat-2 top of canopy photons height was markedly lower for stands in group 1 compared to group 2 even though they exhibited relatively similar high crown closure distribution (Figure 5). This could be explained by a highest proportion of stands in the overmature or declining stages of their development and with a highest proportion of broadleaf species in group 1 (Figure 5-4). Despite having the strongest correlation with the 90th percentile of SPL100 returns among all the groups, group 3 was also the group where the largest average underestimation of canopy height was observed (2.7 m). This could be explained by the fact that the upper canopy layer is less homogeneous in these sparser stands that have not reached maturity compared to stands in groups 1 and 2.

5.4.2. Comparison of ATL03 and SPL100 canopy cover at the AU level

ICESat-2 derived estimates of canopy cover have not been extensively considered in the literature. There is currently no estimate of canopy cover provided in the ATL08 data product. Due to the low number of photons returned per outgoing ATLAS pulse, Neuenschwander and Pitts, (2019) expect canopy cover, defined as the ratio of number of canopy photons to total number of terrain and canopy photons, to overestimate ALS equivalent measures. In this study, canopy cover estimates (CC and VF) were calculated from canopy returns only and ICESat-2 underestimated canopy cover relative to SPL100, particularly in the stands with higher stem density and crown closure (groups 1 and 2), regardless of the method used to

estimate cover (CC or VF). It is important to note here that SPL100 measures of canopy cover had low variability for groups 1 and 2 where high crown closure was also observed in the field (Figure 5-4). Low SPL100 CC and VF (<50%) were only observed in the sparser stands of group 3 and results indicate that ICESat-2 is able to provide relatively consistent estimates of canopy cover even in sparser forest environments.

The comparison of ICESat-2 and SPL100 estimates of canopy cover in this study highlight the challenges of finding an ICESat-2 equivalent for canopy cover as measured with airborne SPL100. The profiling nature of ICESat-2 data and the comparison to SPL100 provided herein indicate that VF is a more consistent representation of canopy cover than the conventional use of percentage of laser returns above a specified height threshold. If the lower ability of single-photon lidar instruments to characterize sub-canopy vegetation in certain forest conditions is problematic (either as a function of cover density, type and configuration, aggressive noise filtering, or a combination thereof) a top-down approach to characterize cover (such as VF) could be advantageous, as it provides an indication of cover that is not dependent upon the efficacy of the laser penetration.

Retrieval of canopy cover with ICESat-2 is expected to depend on the radiometry (number of signal photons per laser pulse) of both ground and canopy, which vary depending on vegetation structure, terrain reflectivity, and atmosphere scattering (Neuenschwander et al., 2020a). The VF index presented herein provides some first insights on a possible method to retrieve canopy cover along the ICESat-2 ground tracks. However, further research focusing specifically on ICESat-2 canopy cover is needed to develop validated methods and characterize the effect of radiometry on the estimate.

5.4.3. Comparison of ATL03 and SPL100 height and height variability at the AU level

Most of the research reported to date has focused on the retrieval of canopy height with ICESat-2, which is defined as the 98th height percentile of canopy photons at the 100 m step by the ATL08 product (Neuenschwander and Pitts, 2019). However, lidar metrics characterizing height variability are consistently important predictors in models for estimating forest attributes, such as biomass (Lefsky et al., 2005; White

et al., 2013). For example, the regression model developed by Narine et al. (2019) to estimate ALS-derived canopy cover from simulated ICESat-2 predictors used the 25th and 50th canopy photons height percentiles as predictor variables.

Duncanson et al. (2020) observed that simulated ICESat-2 data underestimated canopy height percentile at the 100 m segment level when compared to GEDI, especially at the higher percentiles, with a bias of -5 m for the 95th percentile of canopy height. Results presented in this chapter also indicate a tendency of ICESat-2 to underestimate SPL100 height percentiles at the AU level (as well as the photon level). The tendency towards underestimation of top of canopy was also reported in Neuenschwander et al. (2020) in their validation of ICESat-2 ATL08 products against airborne linear-mode lidar in boreal forests. From the three forest groups examined in this study, AUs dominated by more complex structure, with greater variability in heights and mature to declining stands (group 1) consistently had a higher number of outliers, confirming findings at the individual photon level and AU level that such environments can be challenging for ICESat-2. In addition, limitations were observed across the three groups for retrieving the lowest height percentiles (25th and 50th).

The standard deviation of ICESat-2 canopy returns are relatively consistent with those of SPL100, indicating similarities in absolute height variability ($r = 0.64$ to 0.75 , $\text{RMSD}\% = 21.3\%–32.9\%$). In contrast, the coefficient of variation of height for ICESat-2 and SPL100 were markedly less similar ($r = 0.09–0.51$, $\text{RMSD}\% = 31.6\%–39.9\%$), suggesting the relative variability in canopy heights is not consistent between these data sources. In particular, the CV was larger for the SPL100 data for group 3, whereas ICESat-2 CV was greater for groups 1 and 2 (Figure 10). Given the low level of agreement between the SPL100 and ICESat-2 measures of CV that was observed, the use of this particular metric in any modelling approach that seeks to integrate these two data sources is not recommended.

5.4.4. Additional considerations

This study compares ICESat-2 data to an airborne single-photon lidar instrument for which there remain some uncertainties around the capacity to characterize the full vertical distribution of vegetation under all

possible forest conditions (White et al. 2021b). Chapter 3 and other studies have demonstrated the effectiveness of SPL100 data for area-based forest inventories across a range of boreal (Queinnec et al., 2021a; Wästlund et al., 2018; Yu et al., 2020) and mixedwood (White et al. 2021a) forest types, with results that are comparable to those attained using linear-mode lidar. Studies have also shown that the distribution of SPL100 returns may be skewed towards the upper canopy layer compared to linear-mode instruments (Brown et al., 2020; Yu et al., 2020) and that first and single returns dominate (White et al. 2021b). It should be noted though that discrete return systems can also be biased upwards as a function of leading edge ranging, and that photon counting systems have the potential to provide "unbiased range data that samples the height distribution of the surfaces illuminated by the laser pulse" (Harding, 2018). Li et al. (2016) and Brown et al. (2020) hypothesized that the lack of returns from vegetation under the upper canopy layer may partly be the result of aggressive noise filtering rather than inherent detector limitations. However, White et al. (2021b) found differences in penetration capabilities between leaf-on and leaf-off SPL100 data that were subjected to the same noise filtering procedures, suggesting there may be some limitations to the full characterization of forest vertical structure using SPL100 data under certain forest conditions. It is possible that the de-noised SPL100 data used herein may contain systematic sampling errors that impact vegetation structure retrieval. While the comparison of SPL acquisitions from both airborne and spaceborne platforms is important, the absolute validation of ICESat-2 terrain and vegetation products should ideally be performed against data acquired with linear-mode lidar instruments, as demonstrated in Neuenschwander et al. (2020), and for which the relationship between the point cloud and vegetation structure has been more extensively studied.

Throughout this study, the horizontal coordinates reported in the ATL03 data product were used to match individual top of canopy photons or continuous AUs to their coincident SPL100 point cloud. The geolocation errors associated to the ATL03 photons (Luthcke et al., 2021) could cause spatial mismatch between the true location of the ATL03 photons and the SPL100 returns they are compared against. Neuenschwander and Magruder (2019) found that the geolocation errors in a transect acquired over Finland

(ATL03 version 001) were larger in the along-track direction (~ 5 m) compared to the cross-track direction (0 m). ICESat-2 data processing algorithms continue to evolve and Luthcke et al. (2021) quantified geolocation errors using the version 3 ICESat-2 data release, reporting beam-dependent geolocation errors ranging from 2.5–4.4. m. Presumably, ICESat-2 geolocation will continue to improve with future data releases. The analysis of ICESat-2 data along tailored AU lengths requires additional processing compared to the direct use of the ATL08 100-m segments. However, findings reported herein show that the accuracy of ICESat-2 height and cover metrics depends on forest type and structure, and likewise the associated spatial variability in structure. A product such as ICESat-2 ATL08 provides consistent and transparent measures of global vegetation height that are critical for applications such as forest biomass estimation (Neuenschwander and Pitts, 2019). For more regional investigations and applications such as presented herein, the flexibility to integrate the ATL03 and ATL08 products is very useful, particularly in forests that structurally and spatially heterogeneous, where a tailored analysis unit is desirable.

5.4.5. Future directions and implications of findings

Findings of this study have implications on the future use of ICESat-2 data to provide estimates of forest structure and attributes, such as biomass, for both broad-scale monitoring and eventually to inform strategic management decisions. Duncanson et al. (2020), who focused on forested areas of Sonoma County in California, highlighted a tendency of ICESat-2 to markedly underestimate canopy height and therefore also biomass in dense and tall stands while providing a more reliable estimation in shorter and more open canopies. These findings at the Romeo Malette Forest indicate that ICESat-2 estimates of canopy height, cover, and variability were more similar to those derived from SPL100 in the sparser forest type (group 3). In forests with similar structural groups to the ones observed at the Romeo Malette Forest, results also suggest that it may be more challenging to use ICESat-2 structural metrics in older stands with complex structure and greater variability in heights (group 1).

The analysis of ICESat-2 data by detailed forest structural groups within a boreal forest environment as presented herein, provides a comprehensive assessment of ICESat-2 characterization of forest structure and

the capacity to further use these data for forest attribute estimation. ICESat-2 structural metrics have already been used as predictor variables to estimate biomass for the ground tracks (Duncanson et al., 2020; Montesano et al., 2015; Narine et al., 2019b). Considering the increasing availability of ALS data and EFIs (White et al. 2016), there is significant potential to utilise ICESat-2 in synergy with other datasets to spatially extrapolate detailed forest inventory estimates to areas that lack such information for monitoring purposes.

Chapter 6. Mapping canopy height over time using ICESat-2 and Landsat time-series products

6.1. Introduction

In Canada, sustainable forest management is supported by the periodic collection of airborne remote sensing data and field data, for example the acquisition of ALS data and subsequent establishment of a plot network to derive forest inventory products, as demonstrated at the Romeo Malette Forest in Chapter 3 and Chapter 4. Unmanaged northern boreal forests however, which account for about 33 % of Canada's forests (114 Mha; Stinson et al. 2019), are remote, not subject to extensive forest management activities and therefore not characterized by the same breadth of data and knowledge as southern managed forests. Nevertheless, Canada's unmanaged forests have critically important value for biodiversity conservation, large-scale ecological processes and ecosystem services such as carbon storage and water resources (Venier et al., 2018). In addition, shifts in forested ecosystem disturbance regimes are being observed globally (McDowell et al., 2020) and northern boreal forested ecosystems are expected to be particularly vulnerable to climate change (Gauthier et al., 2015; Kurz et al., 2013) due to an increase in the rate of disturbances including fire and insects (Wotton et al., 2010). Evidence of shifts in the trajectories of forests recovery after disturbances have also been demonstrated due to increase in atmospheric CO₂ concentration, temperature warming and altered precipitation regimes (Anderson-Teixeira et al., 2013). As a result, there is a need for spatially-explicit inventory of forest resources beyond Canada's managed forests. In particular, wall-to-wall estimates of AGB could enhance the characterization of Canada's unmanaged forests carbon budget, which are currently not included in nation-wide net carbon fluxes reporting because of the lack of spatially-explicit and accurate forest inventory information (Kurz et al., 2013).

Accurate estimates of forest structure, in particular canopy height, are crucial to model a set of forest attributes including AGB (Duncanson et al., 2022). The remoteness and large geographic extent of Canada's

unmanaged forests require the use of remote sensing measurements of forest structure, particularly spaceborne acquisitions. Long-term satellite optical imagery programs such as Landsat or MODIS, have enabled the characterization of forest cover and mapping of disturbances and recovery dynamics at regional and global scale (Hansen et al., 2013; Justice et al., 2002; Wulder et al., 2022). However, optical imagery does not provide a direct measure of forest structure and derived estimations of forest attributes have been limited by the saturation of reflectance as AGB values increase (Zhao et al., 2016). Lidar systems, conversely, provide direct measurements of vegetation structure. However, even though airborne and spaceborne lidar data are becoming increasingly available, wall-to-wall mapping at a global scale is currently not achievable. As a result, a range of data sources and methods have been used to extend airborne or spaceborne lidar acquisitions across broader extents and through time (Coops et al., 2021). Some examples relevant in the context of this study include time-series of canopy height and other forest inventory attributes derived from ALS transects acquired across the boreal forests of Canada (Matasci et al., 2018b, 2018a) as well as global forest canopy height maps produced by extrapolating GLAS measurements with spectral information collected with the Moderate Resolution Imaging Spectroradiometer (MODIS; Lefsky, 2010) or global climatic, topographic and tree cover products (Simard et al., 2011). More recently, Potapov et al., (2021) also produced a 30-m global canopy height map by integrating time-series of Landsat spectral metrics and GEDI data. Although, GEDI does not sample above 52°N latitude, Potapov et al., (2021) predicted canopy height in boreal forest above 52°N by calibrating their models using data acquired between 40°N and 52°N. Previous studies have also extrapolated either simulated or acquired ICESat-2 canopy height or modeled AGB data in non-boreal forest types with satellite multispectral imagery (Narine et al., 2019c, 2019a), synthetic aperture radar (Nandy et al., 2021; Silva et al., 2021) or a combination of both (Li et al., 2020).

The research undertaken in Chapter 5 explored the ability of ICESat-2 to characterize canopy height across a range of boreal forest types. This last research chapter builds on this knowledge to demonstrate how ICESat-2 data can be integrated with wall-to-wall satellite optical imagery to map canopy height across an

unmanaged boreal forested ecosystem of > 19.6 Mha annually between 1984 and 2021. The value of such products is then illustrated by monitoring canopy height recovery trends across areas that have been impacted by fire disturbance.

6.2. Methods

6.2.1. Study area

This chapter focuses on the NW Boreal study area, which is described in section 2.1.2 of this dissertation. The Romeo Malette Forest (RMF; described in section 2.1.1) is also used to validate the imputed canopy height estimates against the SPL100 data.

6.2.2. Wall-to-wall predictor variables

Landsat time series products as well as topographic and geographic metrics available across the NW Boreal and RMF study areas were used in order to extrapolate ICESat-2 measures of canopy height beyond sampled transects.

Landsat time-series products

Annual Landsat surface reflectance products generated from 1984 to 2021 for both regions using a best-available-pixel (BAP) compositing method (White et al., 2014) where data gaps were in-filled using a protocol considering spatio-temporal trends (Hermosilla et al., 2015a) were obtained. Only Landsat images acquired within ± 30 days of August 1st at each year of the time-series were considered to generate the gap-free BAP composites. This 30-m spatial resolution surface reflectance time-series was used to calculate annual Normalized Burn Ratio (NBR), Enhanced Vegetation Index (EVI) and Tasseled Cap Wetness (TCW), Brightness (TCB), Greenness (TCG), Angle (TCA) and Distance (TCD) as described in Table 6-1. Existing data products capturing changes in the NBR time-series and attributing them to a disturbance type (fire, harvesting, non-stand replacing, road or well site) using random forest classifiers (Hermosilla et al., 2015b) were also utilized. From these data, the number of years since change as well as the change type and change magnitude, persistence and rate of NBR values at each year in the time-series were calculated

for each cell of the BAP grid (more details provided in Table 6-1). Finally, land-cover classification time-series generated from the gap-free BAP composites using the Virtual Land Cover Engine (VLCE) framework in which detected change products are considered (Hermosilla et al., 2018) were also obtained. The VLCE framework considers four forested classes: coniferous, broadleaf, mixed wood and wetland-treed. Other non-forested land cover classes include water, snow/ice (not present in the NW Boreal and RMF study areas), rock/rubble, exposed/barren land, bryoid, shrubland, wetland and herbs. Only forested land cover classes, shrublands and herbs were considered.

Topographic and geographic metrics

The ASTER Global Digital Elevation Model Version 3 (GDEM 003), which provides a global digital elevation model (DEM) at a 30-meter resolution was obtained and resampled using bilinear interpolation in order to match the grid of the Landsat time-series products. Slope (in degrees) and topographic solar radiation index (TSRI; transformed measure of aspect) were also calculated from the DEM. The latitude of each 30-m pixel centroid was also included as a predictor variable for the imputation across the NW Boreal area which covers a large geographic extent.

Table 6-1. Description of the predictor variables included in the k-NN imputation models. Spectral indices, change metrics and land cover metrics were derived for each year between 1984 and 2021.

Type	Variable name	Description
Spectral indices	EVI	Enhanced Vegetation Index at the year of imputation.
	NBR	Normalized Burn Ratio at the year of imputation.
	TCW, TCB, TCG, TCA, TCD	Tasseled Cap Wetness, Brightness, Greenness, Angle ($TCA = \tan^{-1}(TCG/TCB)$) and Distance ($TCD = \sqrt{TCG^2 + TCB^2}$) at the year of imputation.
Change metrics	change_type	Type of change attributed to the detected change (if any). Can be one of “No change”, “Fire”, “Harvesting” or “Non-stand replacing”. All changes detected after the year of imputation are set to “No change”.
	yrs_since_change	Number of years between the year of the imputation and the year when the greatest change was detected. Set to 50 years if “No change” was detected before the year of imputation.
	change_magnitude*	Difference of NBR between the year of change and the year before change.
	pre_magnitude*	Difference of NBR between 1984 and the year prior change.
	pre_rate*	pre_magnitude divided by the number of years between 1984 and the year prior change.
	pre_mean; pre_iqr*	Mean and interquartile range of NBR between 1984 and the year prior change
	post_magnitude*	Difference of NBR between the year of imputation and the year of change
	post_rate*	post_magnitude divided by the number of years between the year of change and the year of imputation
Land cover	vlce	Land cover at the year of imputation
	pre_vlce*	Land cover at the year before change
	post_vlce*	Land cover at the year of change
Topography	DEM	Digital Elevation Model (ASTER GDEM v3)
	SLOPE	Slope derived from the DEM
	TSRI	Topographic Solar Radiation Index. Transformed measure of aspect derived from the DEM.
Location**	LAT	Latitude in decimal degrees (WGS84)

*If change_type is “No change”, pre_, post_ and change_ metrics are equal and calculated from 1984 to the target year.

**Latitude only included in the NW Boreal model

6.2.3. Generation of a gridded ICESat-2 canopy height product

ICESat-2 canopy height estimates were obtained from the version 005 of ICESat-2 Land and Vegetation Height Product (ATL08). The ATL08 provides several absolute and relative canopy height percentile metrics calculated from the photons classified as canopy or top of canopy. The ATL08 estimate of canopy height (h_{canopy}) corresponds to the 98th percentile of canopy or top of canopy photons height above ground. In previous versions of ATL08, canopy metrics were only delivered for 100-m segments along ICESat-2 ground track but version 005 also provide an estimate of canopy height along 20-m segments (h_{canopy_20m}). The 20-m segments were used as they matched the resolution of the 30-m grid of predictor variables better than the 100-m segments. In the remaining part of the manuscript, h_{canopy_20m} is referred to as h_{canopy} even though the 20-m h_{canopy} estimate was always used.

The *subset_te_flag* and *subset_can_flag* parameters of the ATL08 product indicate if the 20-m segment contain photons classified as ground or canopy, respectively. In the ATL08 product, if no canopy or ground photons are detected in a segment (*subset_can_flag* = 0 or *subset_te_flag* = 0) a “no data” value is assigned to h_{canopy} . Instead, a h_{canopy} value of 0 m was assigned to segments with no canopy photons (*subset_can_flag* = 0) but with ground photons (*subset_te_flag* = 1).

As done in chapter 5, the actual footprint of the ATLAS beams was simulated by applying a buffer with a width of 11 m around the ICESat-2 segments. The intersection between the buffered segments and the 30-m cells of the wall-to-wall predictor variables was then examined in order to generate a gridded 30-m h_{canopy} product. The h_{canopy} value in each cell of the 30-m grid intersecting the ICESat-2 segments was determined by calculating the average h_{canopy} value among all overlapping segments, weighted by their percentage of overlap. Overlap percentages between 30-m cells and buffered 20-m segments varied between < 1 % for segments located at the edge of a cell to 35 % for a segment fully contained within a cell.

6.2.4. Imputation of canopy height

Wall-to-wall maps of ICESat-2 *h_canopy* estimates were produced using a k-nearest neighbor (k-NN) approach, where prediction at a target is achieved by imputing the *h_canopy* values observed among the k-nearest reference observations, determined based on the predictor variables values. When $k = 1$, the response variable values of the nearest reference observation are imputed. When $k > 1$, the imputed response variables values are determined from the average or weighted average (used here) among the k-nearest reference observations. Different methods exist to compute the proximity between target and reference observations. The proximity measure was derived from a random forest (RF) model trained on the reference observations. When a prediction is made at a target, the values of the target predictor variables are ran through the RF and the number of times each reference observation falls into the same final node as the target (i.e. RF proximity matrix) is used to infer the proximity between the target and reference observations (Queinnec et al., 2021b).

Reference observations consisted in the pixels of the gridded *h_canopy* product. Similarly to chapter 5, a series of filters were applied to the data, as described in Table 6-2, in order to form a pool of observations acquired in the most suitable conditions for vegetation mapping. Only observations that were acquired by ATLAS strong beams, at night and within ± 30 days of August 1st (matching acquisition dates of the Landsat images used to generate the BAP time-series products) were retained. Gridded *h_canopy* > 40 m were also removed and cloud and aerosol cover, the presence of water, land cover and change type were also considered to filter the data. After filtering, 9,979 and 97,173 observations were retained at the RMF and NW Boreal study area, respectively. At the NW Boreal study area, the sample of reference observations was further reduced to 10,000 observations by randomly sampling 1000 observations in each decile of the distribution of *h_canopy* values in order to keep computational performances within practical limits.

The R package *yaImpute* (Crookston and Finley, 2008) was used to perform the RF-based k-NN imputation of *h_canopy* across the RMF and the NW Boreal study area for the year 2018. One k-NN imputation model was developed for each of the study areas. A total of 500 regression trees were used for each RF model and

the *mtry* parameter which controls the number of variables randomly selected at each tree node was set to the default value of 8 (number of predictor variables divided by 3). The *yaImpute* package uses the RF implementation available in the *randomForest* package.

Table 6-2. Filters applied to the gridded *h_canopy* values at the RMF and NW Boreal study area in order to form a pool of reference observations acquired in optimal conditions.

Description of filter	Value used	Percentage of all observations (RMF; NW Boreal)
Percentage area of 30-m pixel covered by ICESat-2 segments	$\geq 35\%$ (corresponds to at least one 20-m segment fully contained within the pixel)	58.2%; 57.5%
Strength of the ATLAS beam (weak or strong). Derived from the ATL08 parameter <i>sc_orient</i> which tracks the orientation of ICESat-2.	Strong	63.2%; 59.9%
Indicates if data were acquired in day or night conditions (ATL08 <i>night_flag</i> parameter).	Night	44.4%; 43.7%
Maximum gridded <i>h_canopy</i> value.	< 40 m	99.8%; 99.9%
Number of cloud or aerosol layers at the time of acquisition as indicated by ATL08 <i>cloud_flag_atm</i> parameter. Ranged between 0 and 4 in the data used in this study.	≤ 1	90.9%; 91.9%
Day of year at the time of data acquisition.	213 ± 30	21.1%; 14.4%
Difference between the 100-m median terrain height and a reference digital elevation model height (GMTED2010; Danielson and Gesch, 2011)	Within ± 1 standard deviation from the mean (-1.3 ± 3.3 m at the RMF; -0.94 ± 1.9 m at the NW Boreal)	79.4%; 78.9%
ATL08 flag (<i>segment_watermask</i>) indicating the absence (0) or presence (1) of inland water in 100-m segments according to the Global Raster Water Mask (Carroll et al., 2009)	0	98.2%; 88.2%
Land cover class according to the VLCE at the year of acquisition	Coniferous, Broadleaf, Mixed Wood, Wetland-Treed, Shrubs, Herbs	95.4%; 78.3%
Change type according to changes that were detected between 1985 and the year of acquisition	No change, Fire, Harvesting, Non-stand replacing	96.2%; 98.4%
All pixels in a 3x3 neighborhood have the same land cover class.	-	39.6%; 38.6%
All pixels in a 3x3 neighborhood have the same change type and change year.	-	76.8%; 86.0%
All filters combined	-	1.7%; 0.4%

6.2.5. Comparison of imputed and observed canopy height

No dataset other than the grid of h_{canopy} observations was available at the NW Boreal study area to validate the imputed canopy height values. As highlighted in chapter 5, ICESat-2 measures are associated with relatively large uncertainties and should not be considered as reference data. It is therefore important to highlight here that only the relative accuracy of the k-NN imputation method could be assessed in that study area, as opposed to the accuracy of the imputed canopy height estimates. However, the availability of independent SPL100 reference observations at the RMF allowed for the assessment of the accuracy of the imputed canopy height estimates there.

At the NW Boreal study area, the accuracy of the k-NN imputation was assessed by comparing values imputed in 2018 to a sample of 10,000 observations of h_{canopy} values not used as reference observations (i.e., for model development) and selected using random sampling stratified by land cover, change type and years since change. The allocation of the number of sampled observations in each stratum was determined proportionally to the occurrence of those strata across the area. At the RMF, the k-NN imputation accuracy assessment was performed by holding-out 25% of the available h_{canopy} observations ($n = 2495$) using random sampling. In addition, the 95th percentile of the height of SPL100 returns classified as vegetation (SPL100 p95) within the 30-m cells of the Landsat grid were used as reference observations to assess the accuracy of the imputed canopy height estimates. An independent sample of 10,000 SPL100 p95 observations were selected using the same stratified random sampling scheme as described above for the NW Boreal area.

Regardless of the study area and validation method, accuracy was assessed by repeating the k-NN imputation 10 times for each value of k between 1 and 10. A new held-out sample was randomly selected at each repetition for the RMF validation against observed h_{canopy} values. The Pearson correlation coefficient (r), root-mean-square difference (RMSD) and mean difference (MD) between predicted (\hat{y}_i) and observed (y_i) values were used as summary statistics of the accuracy assessment. Relative RMSD (pRMSD) and relative MD (pMD) were also computed as percentages of the mean of observed values (\bar{y}).

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$MD = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n}$$

$$pRMSD = \frac{RMSD}{\bar{y}} \times 100$$

$$pMD = \frac{MD}{\bar{y}} \times 100$$

6.2.6. Trends in canopy height recovery after fire disturbance

The evolution of canopy height after fire disturbance was examined across a sample of 5,000 cells that were selected using random sampling stratified by latitude (10 groups of 0.5° increments between 50 – 55°) and change_magnitude (Table 6-1; 5 groups of 0.2 increments between -0.2 – -1.2). For each stratum, 100 cells where fire disturbance was detected (change_type; Table 6-1) during the time-series were randomly sampled. The imputation of *h_canopy* was performed at the sampled cells for each year between 1984 and 2021. Imputed *h_canopy* values were expressed as percentages of the *h_canopy* value imputed the year prior to the fire event. Canopy height decrease and recovery trends were examined by calculating the number of years until which the lowest % pre-fire canopy height value was observed and the % pre-fire canopy height 15, 20, 25 and 30 years after the fire event.

6.3. Results

6.3.1. Examination of reference observations

As expected when comparing the productive managed forest to the unmanaged northern boreal forested ecosystems, the distribution of *h_canopy* observations differed between the RMF (median = 14.0 m; q25 = 10.3 m; q75 = 16.9 m) and the NW Boreal study area (median = 8.5 m; q25 = 5.5 m; q75 = 12.1 m), as shown in Figure 6-1.

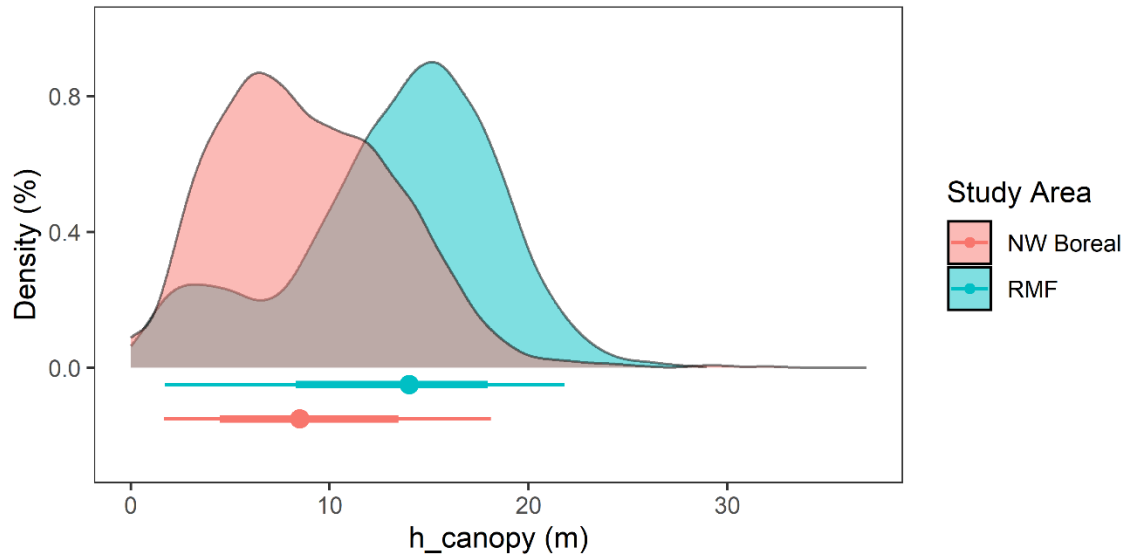


Figure 6-1. Distribution of gridded h_{canopy} values of the reference observations at the Romeo Malette Forest (9,979 observations) and at the NW Boreal study area (10,000 observations)

At the RMF, h_{canopy} values at the 9,905 reference observations where no change was detected between the SPL100 and ICESat-2 acquisitions had an $r = 0.92$, $RMSD = 2.13$ m ($pRMSD = 16.00\%$) and negligible bias ($MD = -0.06$ m) when compared to SPL100 p95 values. The two estimates aligned relatively well along the 1:1 line despite some outliers (Figure 6-2A) and their distribution were comparable (Figure 6-2B). The level of agreement between h_{canopy} and SPL100 p95 varied by land cover type (Table 6-3; Appendix E) and agreement between the two estimates was the poorest in the Herbs land cover ($r = 0.46$; $RMSD = 3.17$ m; $MD = 0.52$ m).

Table 6-3. Comparison of ICESat-2 and SPL100 estimates of canopy height (h_{canopy} and $p95$, respectively) in reference observations grouped by forest type (VLCE at the year of ICESat-2 data acquisition). $p\text{RMSD}$ and $p\text{MD}$ are calculated based on the mean values of SPL100 $p95$. Observations for which change was detected after 2018 were removed to account for the time difference between the SPL data acquisition (2018) and the ICESat-2 data acquisition (2019 – 2021).

Land cover	n	r	RMSD	MD	$p\text{RMSD}$	$p\text{MD}$
Broadleaf	162	0.94	2.1	-0.35	21.4	-3.59
Mixed Wood	2538	0.91	2.41	-0.66	15.34	-4.24
Coniferous	6090	0.85	1.77	0.09	12.45	0.65
Shrubs	6	0.85	1.09	0.64	31.69	18.52
Wetland-Treed	62	0.73	1.09	0.07	18.42	1.2
Herbs	1047	0.46	3.17	0.52	90.05	14.74
All	9905	0.92	2.13	-0.06	16	-0.47

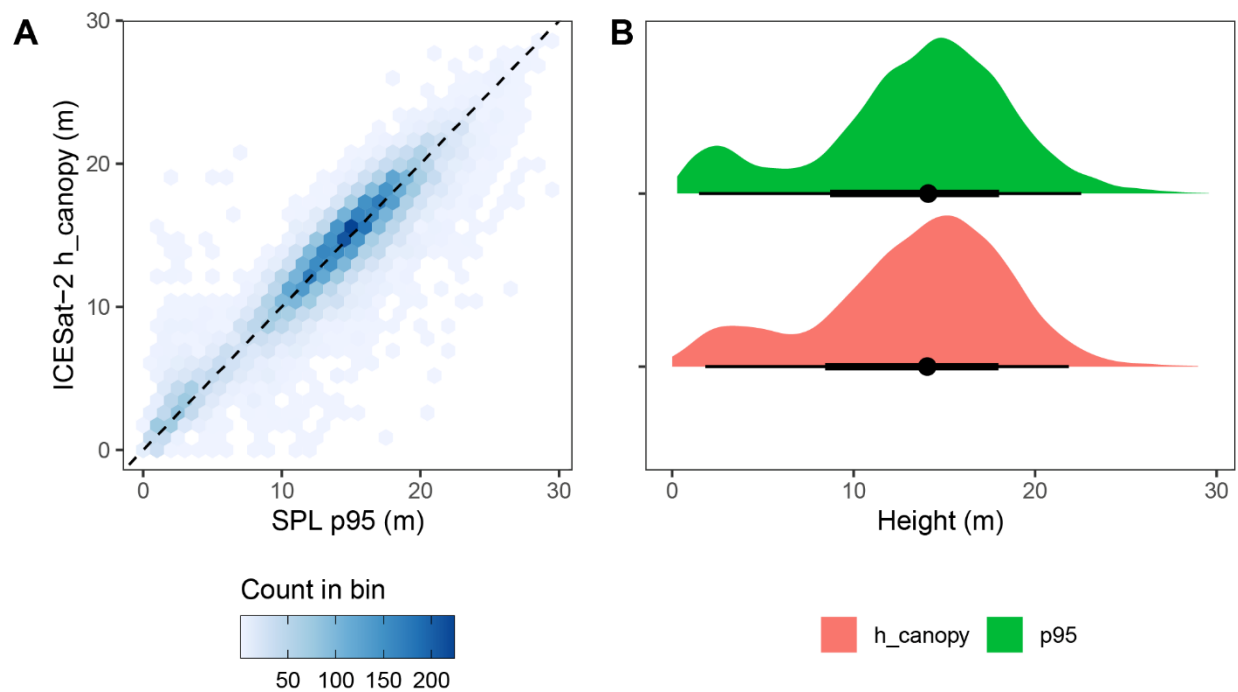


Figure 6-2. Scatter plot (A) and distribution (B) of $p95$ (SPL) and h_{canopy} (ICESat-2) estimates of canopy height in reference observations where no change was detected after 2018. In A, the scatterplot is visualized using 1-m hexagonal bins colored (shaded from light to dark blue) based on the number of observations contained within them. In B, the point indicates the median of observations and the thick and thin lines indicate intervals that contain 66% and 95% of the distribution around the median, respectively.

6.3.1. ICESat-2 canopy height imputation

An overview of the wall-to-wall imputation ($k = 5$) of h_{canopy} values at a 30-m spatial resolution in the NW Boreal study area in 2018 is provided in Figure 6-3. This map highlights a gradient of canopy height values, with a shift from larger canopy height values in the southern and south-western part of the area,

dominated by coniferous and mixed wood forests, towards lower canopy height in the northern and north-eastern areas where the landscape is more largely dominated by shrubs and treed wetlands (Figure 2-4; Figure 6-3).

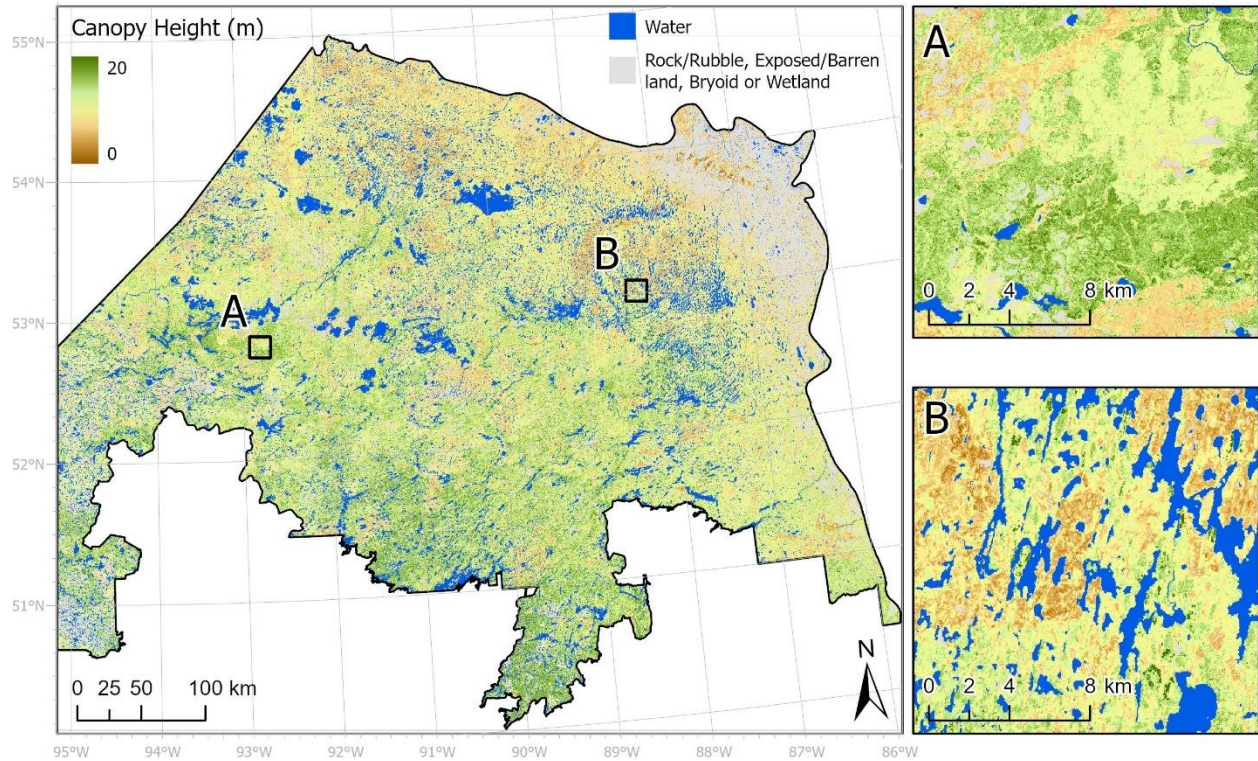


Figure 6-3. Map of ICESat-2 h_{canopy} estimates imputed across the study area located in the NW portion of the Boreal Shield Ecozone in Ontario, Canada in 2018.

Overall, the accuracy of the k -NN imputation improved as the number of nearest-neighbours (k) increased but a plateau in r , $p\text{RMSD}$ and $p\text{MD}$ values started to be observed for values of $k > 5$ (Figure 6-4).

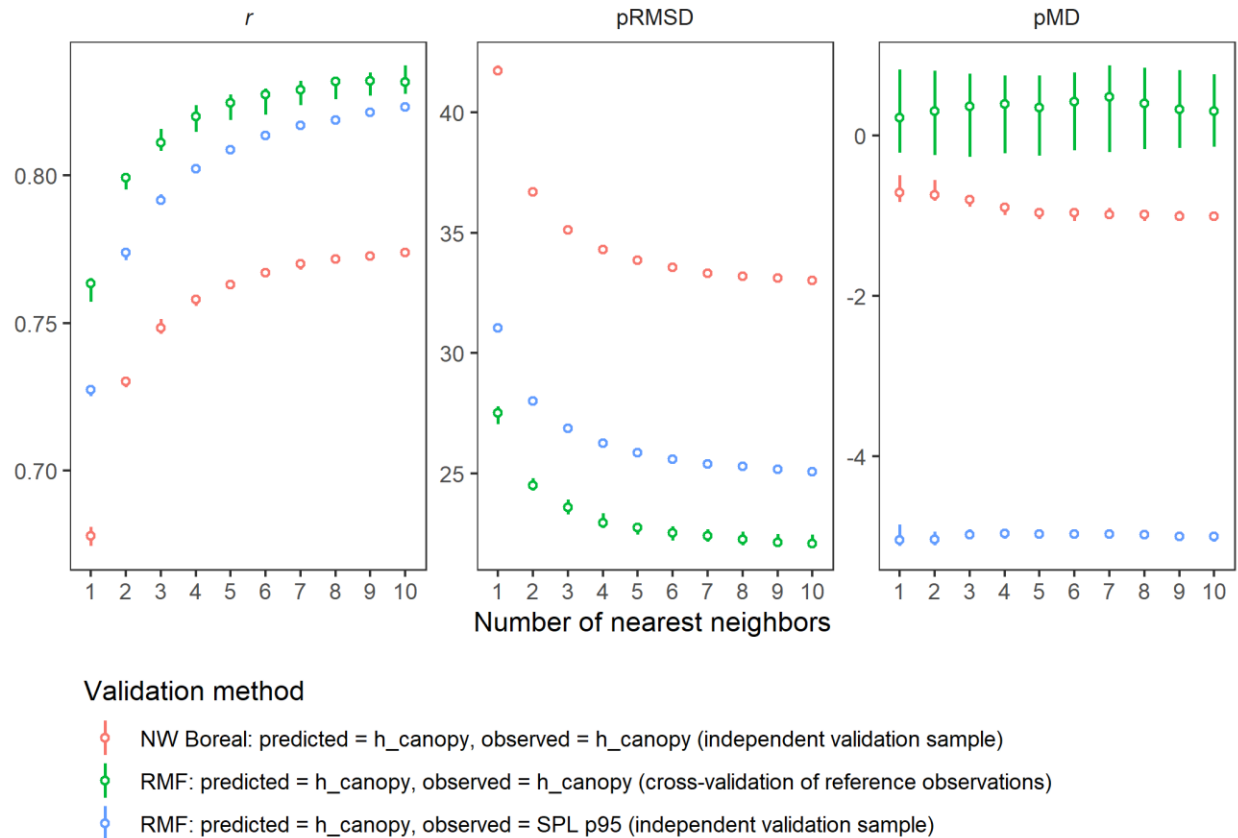


Figure 6-4. Influence of the number of nearest neighbors (k) considered in the k-NN imputation on the Pearson correlation (r), relative RMSD (pRMSD) and relative MD (pMD) for different validation methods. The imputation was performed 10 times for each number of nearest neighbor and validation method. Points and lines represent the median and 75% intervals around the median of the distributions across the 10 repetitions, respectively.

For $k = 5$, the comparison of the imputed h_{canopy} estimates with the observed h_{canopy} values of the independent validation data at the NW Boreal area ($n = 10,000$) yielded $r = 0.76$; $RMSD = 3.04$ m ($pRMSD = 33.9\%$) and $MD = -0.10$ m ($pMD = -1.0\%$). At the RMF, median values of $r = 0.82$; $RMSD = 2.99$ m ($pRMSD = 22.7\%$) and $MD = 0.05$ m ($pMD = 0.34\%$) were observed when comparing imputed h_{canopy} values to the observed ones across the held-out samples. Relatively similar r and $RMSD$ were observed when comparing imputed h_{canopy} values to SPL100 p95 values of the independent sample of 10,000 observations ($r = 0.81$; $RMSD = 3.65$ m; $pRMSD = 25.9\%$). However, MD values indicated that

imputed h_{canopy} tended to underestimate SPL100 p95 values by 0.70 m on average (MD = -0.70 m; pRMSD = -4.96%).

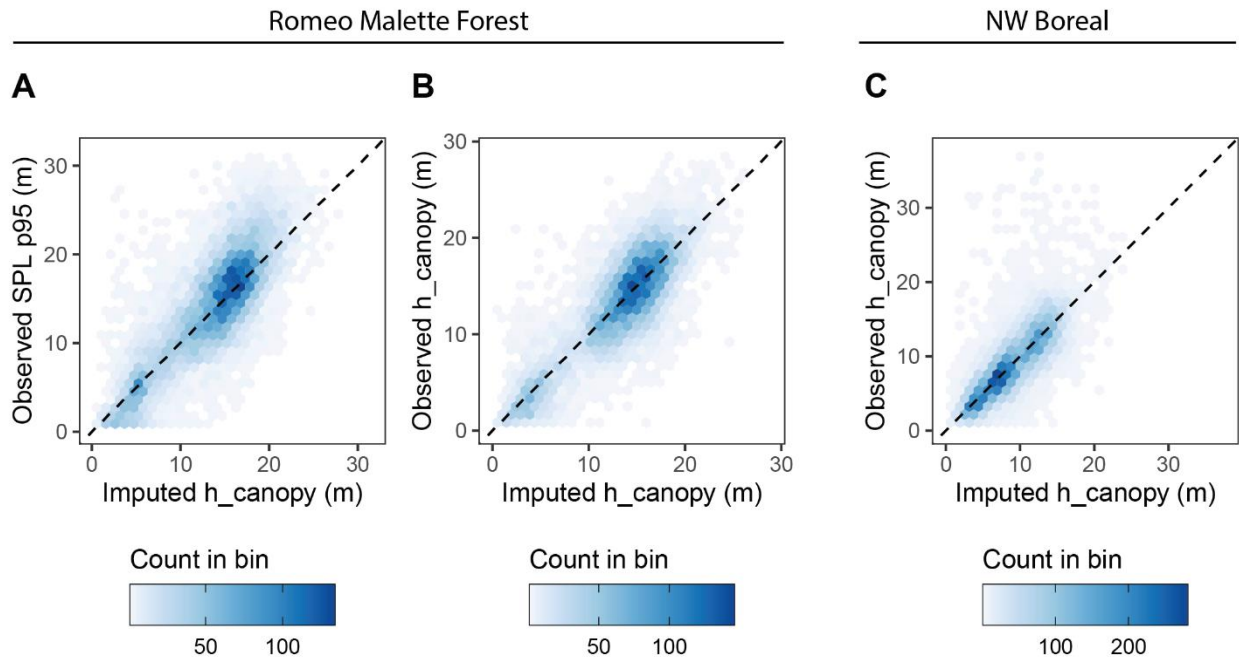


Figure 6-5. Scatterplots illustrating the results of the h_{canopy} k-NN imputation accuracy assessment at the Romeo Malette Forest and NW Boreal site. Points are aggregated into 1-m width hexagonal bins colored from light to dark blue based on point density.

6.3.2. Canopy height recovery after fire

The examination of canopy height trajectories in the 30 years following fire disturbance highlighted that canopy height decreased gradually in the 8 – 10 years following the event, regardless of the change magnitude (Table 6-4; Figure 6-6). However, the magnitude of the decrease in canopy height tended to increase with change magnitude. At the lowest point of the post-fire canopy height trajectory imputed h_{canopy} values ranged between 28.4 % (change_magnitude = 0.8 – 1) and 44.8% (change_magnitude = 0.2 – 0.4) of the pre-fire values. As a result, different % pre-fire canopy height values were observed as canopy height recovered. For example, imputed h_{canopy} values reached between 52.2% (change_magnitude = 0.8 – 1) and 73.7% (change_magnitude = 0.2 – 0.4) of pre-fire imputed h_{canopy}

values 20 years after fire disturbance. However, a ~20% increase in % pre-fire canopy height was observed between years 15 and 30 after fire regardless of change_magnitude, indicating similar recovery rates.

Table 6-4. Summary of the evolution of canopy height after fire disturbance expressed as a percentage of pre-fire canopy height. The median values calculated across all sampled pixels having been disturbed between 1984 and 2021 are reported.

Absolute change magnitude	Lowest % pre-fire canopy height	Years to lowest % pre-fire canopy height	% Pre-fire canopy height X years after fire			
			Y15	Y20	Y25	Y30
0.2 – 0.4	44.8	8	67.6	73.7	77	86.7
0.4 – 0.6	37.4	9	53.8	60.9	68.8	73.5
0.6 – 0.8	33.9	10	47.2	51.6	58.9	67.4
0.8 – 1	28.4	10	44.9	52.2	59.9	69.1
1 – 1.2	29	9	45.8	55.4	64.5	64.7

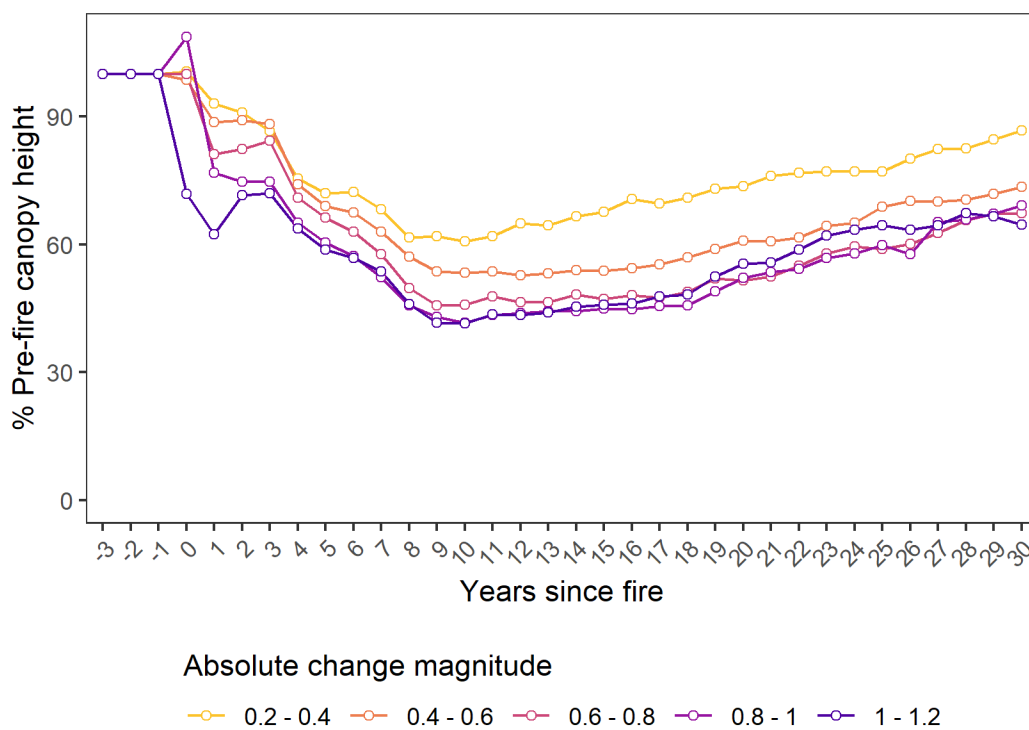


Figure 6-6. Percentage of pre-fire canopy height before (years -3 to 0) and after (years 1 to 30) a fire event grouped by change magnitude (dNBR) values. Each point represents the median value calculated across all the sampled pixels having been disturbed by fire between 1984 and 2021.

6.4. Discussion

ALS acquisitions provide detailed and accurate information about forest structure and resources. However, their acquisition is not systematic and often targeted towards actively managed forests to inform forest inventories. The extrapolation of lidar data with wall-to-wall datasets such as medium-resolution optical satellite imagery data enables the estimation of forest structure and resources across larger areas. Nevertheless, the accuracy of the extrapolation models depends partly on the representativeness of the reference lidar samples of the type of forest where they are applied. In Canada for example, ALS data acquired in actively managed forest units where harvesting and silvicultural treatments occur do not necessarily represent unmanaged forested areas where forest structure is driven by natural dynamics. Spaceborne lidar samples, such as the ones collected by ICESat-2 or GEDI, enable systematic sampling of forest structure in areas where ALS mapping is not necessarily feasible.

6.4.1. Accuracy of the imputation

Findings indicate that extrapolated ICESat-2 canopy height estimates at the RMF ($r = 0.81$; $\text{RMSD} = 3.65 \text{ m}$; $\text{pRMSD} = 25.9\%$) evaluated against SPL100 data yielded more accurate results than extrapolated GLAS and GEDI canopy height measurements reported in other studies. Bolton et al. (2013) compared extrapolated GLAS estimates (ICESat) derived by Simard et al. (2011) and Lefsky (2010) against ALS transects acquired across the Canadian boreal zone and reported $\text{RMSD} = 3.9 \text{ m}$ ($\text{pRMSD} = 33\%$) and $\text{RMSD} = 7.4 \text{ m}$ ($\text{pRMSD} = 62\%$), respectively. More recently, Potapov et al., (2021) produced a global canopy height map by extrapolating GEDI measurements with Landsat data and reported $\text{RMSD} = 9.07 \text{ m}$; $R^2 = 0.61$. It is important to note however that the geographic extent of the imputation at the RMF is smaller than the aforementioned studies and could therefore contribute to the improved level of accuracy. The validation in the larger NW Boreal study area was limited by the lack of reference ALS data to compare against. However, lower accuracy than what was reported herein ($r = 0.76$; $\text{RMSD} = 3.04 \text{ m}$; $\text{MD} = -0.10 \text{ m}$) would be expected when comparing imputed h_{canopy} against ALS data rather than observed h_{canopy} values.

6.4.2. Selection of reference observations and predictor variables

A careful selection of ICESat-2 canopy height reference observation was performed (Table 6-2). Moudrý et al. (2022) examined the influence of a range of flags on the accuracy of canopy and terrain retrieval and noted that a number of studies were performing too aggressive filtering on the data, hence unnecessarily greatly reducing the number of observations. The influence of different types of filtering on the accuracy of the imputation was not particularly examined in this study and the quality of the reference observations was prioritized to their amount. In this study, a sufficient number of reference observations were still available after filtering. However, less restrictive data filtering could be necessary in certain areas limited by the availability of observations in optimal conditions (e.g., effect of atmospheric scattering and / or snow cover).

The Landsat metrics that were used as predictor variables only represent a subset of metrics that could be used. For example, phenology metrics (e.g. Potapov et al., 2021) or texture metrics highlighting edges and area of high vegetation (Lang et al., 2019) may improve the accuracy of the imputation. Moreover, estimation of forest structure and attributes with Landsat data is inherently limited by the saturation of the reflectance measured in the different spectral channels as vegetation biomass increases (Zhao et al., 2016). In addition to multispectral information obtained from Landsat data, other wall-to-wall remote sensing data types, such as synthetic aperture radar (SAR) data could also be integrated. For example, Silva et al. (2021) produced regional estimates of above ground biomass using a fusion of simulated GEDI, ICESat-2 NASA-Indian Space Research Organization SAR (NISAR) data.

6.4.3. Multivariate imputation

This study only examined the imputation of ICESat-2 canopy height. However, a particular advantage of k-NN imputation approaches such as the one used herein is that they are multivariate as the imputation can be performed simultaneously if multiple response variables are available at the reference observations (Eskelson et al., 2009). Instead of modeling each response variable individually, multivariate imputation ensures that the variance-covariance structure and logical relationship between response variables is

retained, as long as $k = 1$. Multiple ALS metrics and modeled attributes such as basal area, volume and AGB were imputed simultaneously in Canada-wide 30-m layers of forest structure (Matasci et al., 2018b, 2018a; Zald et al., 2016). Similarly, other metrics directly obtained or derived from the ICESat-2 ATL03 and ATL08 products, such as proxies of vertical structure variability or canopy cover (Queinnec et al., 2021c) could be imputed along with canopy height. Other modeled attributes such as estimates of above ground biomass derived at the location of ICESat-2 samples with intermediate models (Narine et al., 2020, 2019b) could also be added as response variables. It is important to note however that the propagation of errors associated to the estimates when two-stage modeling approaches are used (i.e. response variables derived from intermediate models) should be accounted for (Saarela et al., 2016) but can be difficult to estimate particularly when non-parametric models such as k-NN are used.

6.4.4. Applications of canopy height monitoring products

Wall-to-wall regional or global canopy height products can ultimately be used for a range of applications. Here, the examination of canopy height recovery trends after fire disturbance agreed well with observations made by Bolton et al. (2015) who assessed post-fire canopy height, cover and height variability along the boreal ALS transects used in the NTEMS structural products. In particular, Bolton et al. (2015) also observed a decrease in lidar height metrics during the 10 – 15 years following fire due to gradual loss of residual forest structure. Canopy height trajectory curves presented in Figure 6-6 indicated that canopy height generally did not recover to pre-fire values after 30 years, which would indicate an overall loss of carbon stored in above-ground biomass. However, further research would be necessary in order to better characterize canopy succession and ecosystem resilience. In particular, examination of pixel-level trends as well as relationships with climatic, soil or geomorphological variables would be particularly insightful in order to identify the conditions leading to under-performing canopy height recovery. Previous research has also examined post-disturbance recovery trends with temporal trends of spectral indices (Frazier et al., 2015; Morresi et al., 2019; Pickell et al., 2016; White et al., 2022, 2018). Pickell et al. (2016) and White et al., (2022) found for example that Landsat spectral indices returned to 80% of pre-disturbance levels within

1.7 – 5.6 years (Pickell et al., 2016; multiple indices and disturbance types) or 10.6 years (White et al., 2022; for NBR and wildfire disturbance only) on average following disturbance. In the years following disturbance, spectral response is driven by the rapid return of short vegetation such as shrubs, which explain the rapid recovery of spectral indices such as NBR to pre-disturbance levels. The emergence of trees and their growth takes longer and eventually stem density, composition (coniferous vs. broadleaf), and disturbance severity all influence spectral recovery (White et al., 2022). It is important to note that previous research has shown that ICESat-2 is challenged to measure height in young boreal forests that are short and/or have low canopy cover, which would be conditions to expect following fire disturbance (Montesano et al., 2015). Beyond 10 years post-disturbance, the gradual increase in tree height could however overcome the uncertainty in ICESat-2 measures.

Canopy height is also a key metric to derive estimates of AGB, which are needed to improve the characterization of carbon fluxes across spatial scales (Boisvenue and White, 2019; Kurz et al., 2013). AGB models driven by GEDI measurements of relative canopy height are being developed and applied globally (Duncanson et al., 2022) and ICESat-2 could play an important role in complementing these estimates in northern boreal forests that are not directly sampled by GEDI (Duncanson et al., 2020). The time-series of canopy height derived in this chapter across the study area of about 20 Mha could be extended to the 114 Mha of Canada's unmanaged boreal forests by integrating additional local ICESat-2 samples. This is a particular strength of the imputation of spaceborne canopy height estimates as opposed to other approaches that often substitute ALS data from other locations that might not be representative of local forest types. In addition to spatial scalability, the fact that spaceborne lidar missions acquire systematic samples of data over time is particularly valuable in a context of a rapidly changing climate. Even though the record of spaceborne lidar observations is still limited, current (ICESat-2; GEDI) and future missions could start building long-term records of global canopy height, similarly to successful Earth observation programs such as Landsat.

Chapter 7. Conclusions

7.1. Dissertation objectives

The overall objective of this dissertation was to examine how information needs related to forest management and monitoring objectives in the boreal forests of Canada can be fulfilled with airborne and spaceborne SPL data.

This overall objective was addressed by answering the four research questions posed in Chapter 1 as detailed below.

Q1: What is the accuracy of key forest inventory attributes when derived from SPL100 data?

Approach: A complete EFI fully driven by SPL100 data was developed at the Romeo Malette Forest in Chapter 3. Key steps included the selection of plot locations to represent the range of forest conditions occurring across the area, the development of forest attribute models using SPL100 metrics, and their application across the forest.

Main findings: Chapter 3 demonstrated that the accuracy and bias of forest attributes predicted from the SPL100 at the Romeo Malette Forest were commensurate with those reported for a previous EFI conducted in the same area with a low-density linear-mode lidar sensor. The forest attribute predicted with the highest accuracy was LOR with $R^2 = 0.91$ and $RMSE\% = 8.5\%$. The attributes BA, QMDBH, WVOL, MVOL and AGB had R^2 values ranging between 0.79 and 0.9 and $RMSE\%$ ranging between 13.97% and 23.79%. DENS was predicted with markedly lower accuracy ($R^2 = 0.63$; $RMSE\% = 30.82\%$). Relative bias was $< 1\%$ for all forest attributes. The $RMSE\%$ was the lowest in black spruce-dominated plots for LOR (6.8%), BA (19.37%) and QMDBH (9.69%) but in plots with mixed conifers for MVOL (19.02%), WVOL (21.32%) and AGB (17.22%). DENS was predicted with the lowest $RMSE\%$ (24.77%) in mixedwoods plots. Jack pine-dominated plots had the highest $RMSE\%$ compared to the other species groups for LOR (10.65%), BA (22.29%), QMDBH (17.21%), DENS (39.31%) and WVOL (24.28%). Predictions at the

deciduous plots exhibited higher absolute bias% than for the other species groups for BA (9.01%), WVOL (6.80%), MVOL (7.10%) and AGB (7.99%).

In addition to accurate predictions across a range of forest types, Chapter 3 emphasized that the value of wall-to-wall SPL data for EFI also relates to its use to guide the establishment and optimization of a sample plot network through a SGS approach. The SGS process resulted in decreasing the number of plots in oversampled forest types covered by the existing plot network and increasing the number of plots in poorly represented forest types.

Q2: How can SPL100 data be used to include dominant tree species composition estimation in enhanced forest inventories?

Approach: In chapter 4, dominant tree species groups were classified across the Romeo Malette Forest by combining common area-based lidar metrics with metrics derived from ITD and ITC segmentation methods and spectral information obtained from Sentinel-2.

Main findings: Previous research has examined individual tree species classification with ALS data, often in combination with spectral data from high resolution imagery, but scaling this information up to large extents remains a challenge. Chapter 4 demonstrated across the Romeo Malette Forest that some of the species composition information provided by photo-interpreted inventories can be derived with relatively high accuracy using a combination of area-based and individual tree crown summary metrics of forest structure from SPL100 data as well as moderate resolution multispectral data from Sentinel-2. The spectral information of Sentinel-2, particularly in the near-infrared region of the electromagnetic spectrum ($\lambda = 740$ nm), the area and volume of upper tree crowns and the cumulative distribution of returns in the upper sections of the canopy were the most important metrics in the classification models. Area-based species grouping products such as the ones derived herein could be integrated operationally to inform regional forest management and monitoring as well as used to guide the derivation of more precise information (i.e., tree-level) where needed.

Q3: How do ICESat-2 and SPL100 estimates of canopy height, cover and height variability compare across a gradient of boreal forest structure?

Approach: Chapter 5 presented a comparison of SPL100 and ICESat-2 estimates of canopy height, canopy cover and height variability along transect of data that cover a range of boreal forest conditions.

Main findings: Chapter 5 showed that the height of ICESat-2 individual photons reflected by the top of canopy corresponded most strongly to the 90th percentile of SPL100 returns coincident to the ATLAS footprint. If compared against the SPL100 estimate of canopy height (99th percentile), this corresponded to an underestimation of canopy height by an average of 2.3 m. The least similarity of ICESat-2 structural metrics relative to SPL100 counterparts was observed in older forest stands, which had more complex vertical structure and greater variability in canopy heights. Estimates of canopy cover analogous to those made using SPL100 were challenging to retrieve with ICESat-2. As a result, the use of the vegetation fill index, which was originally conceived for use with data acquired from profiling LiDAR sensors, was proposed as a proxy for canopy cover. The utility of the vegetation fill index for estimating canopy cover with ICESat-2 data requires further investigation to fully consider ATLAS radiometry effects and applicability in other forested environments.

Q4: How can widespread ICESat-2 samples of forest structure be used to monitor canopy height over time?

Approach: In Chapter 6, maps of canopy height were produced at the NW Boreal study area (> 19 Mha) and the Romeo Malette Forest (630,000 ha) by imputing ICESat-2 canopy height estimates using wall-to-wall remote sensing products mainly derived from Landsat time-series. The imputation was performed over space and time (1984 – 2021) and enabled the examination of trends in canopy height recovery after fire disturbance.

Main findings: Gradients of canopy height structure mapped across the NW Boreal study area were consistent with expectations, with larger canopy height values in the southern and south-western part of the

area, dominated by coniferous and mixedwood forests, towards lower canopy height in the northern and north-eastern areas where the landscape is more largely dominated by shrubs and treed wetlands. The comparison of observed and imputed ICESat-2 estimates of canopy height across a sample of 10,000 observations selected at the NW Boreal study area yielded $r = 0.76$; $\text{RMSD} = 3.04 \text{ m}$ ($\text{pRMSD} = 33.9\%$) and $\text{MD} = -0.10 \text{ m}$ ($\text{pMD} = -1.0 \%$). The accuracy of the canopy height map was also assessed against SPL100 canopy height measures at the Romeo Malette Forest, where results yielded $r = 0.81$; $\text{RMSD} = 3.65 \text{ m}$ ($\text{pRMSD} = 25.87\%$); $\text{MD} = -0.70 \text{ m}$ ($\text{pRMSD} = -4.96\%$). The imputation of ICESat-2 canopy height estimates from 1984 to 2021 across areas of the NW Boreal study area that were impacted by fires also allowed to show different trends in canopy height recovery over time depending on the magnitude of the disturbance. Overall, findings show that canopy height tended to decrease in the 8 to 10 years following fire, likely due to the gradual falling of residual forest structure. These observations were consistent with findings of Bolton et al. (2015) who examined the evolution of ALS metrics after fire across different boreal regions of Canada.

7.2. Research innovations and significance

This dissertation provided several innovations contributing to enhancing the knowledge on the potential of airborne and spaceborne SPL data for characterizing forest structure and resources:

- As highlighted in the introduction of this dissertation, the ability of SPL100 data to characterize forest structure and model forest attributes was not well known at the start of this thesis. Results presented in Chapter 3 showed that area-based estimates of key forest inventory attributes modeled with SPL100 data could be estimated with an accuracy similar to what would be expected from conventional ALS systems. The work presented in this dissertation, in combination with other research published during the course of this dissertation (Brown et al., 2020; Raty et al., 2021; Wästlund et al., 2018; White et al., 2021b, 2021a; Yu et al., 2020), has contributed to increasing the depth of knowledge of the capabilities of the SPL100 system.

- The SGS method presented in Chapter 3 and implemented across the Romeo Malette Forest has further contributed toward the establishment of guidelines and frameworks that forest managers were missing. The interest in the approach has led to its implementation across additional forest management units of Ontario.
- Chapter 4 presented innovations in methods for integrating estimates of tree species composition into area-based EFIs using a combination of area-based and individual tree metrics as well as satellite optical imagery from Sentinel-2. In addition, few studies have examined the use of individual tree detection and crown delineation approaches with SPL100 data.
- Chapter 5 contributed to enhancing the knowledge of the performance of ICESat-2 to characterize a gradient of boreal forest structure with on-orbit data. Previous work prior to the launch of ICESat-2 had focused on the use and analysis of simulated data. In this chapter, the use of the vegetation fill index as a way to characterize canopy cover with ICESat-2 data was also proposed. Canopy cover retrieval was identified as a target attribute for the ATL08 Land and Vegetation product; however, estimates of cover have yet to be included in an ATL08 release, underscoring the challenges associated with deriving this attribute.
- Chapter 6 contributed to enhance the characterization of northern unmanaged boreal forests, which are particularly impacted by an increase in the frequency and severity of abiotic and biotic disturbances due to climate change. Despite an essential need of monitoring forest attributes in these remote regions, they are currently not regularly inventoried nor covered by systematic ALS acquisitions. ICESat-2 is the only sensor that provides systematic samples of forest structure in northern unmanaged boreal forests, and the work carried out in Chapter 6 provides significant contributions towards a better characterization of canopy height, and potentially height-related variables in these ecosystems. Products such as the ones derived in Chapter 6 could provide significant insights for policy-making or reporting, such as the inclusion of unmanaged boreal forest into Canada's carbon budget monitoring. In addition, by using local samples of canopy

structure measured by ICESat-2, the method also presents significant advantages over other approaches that often extrapolate forest structure from lidar acquisitions acquired in forests that are not necessarily representative of the area where estimates are needed.

7.3. Research limitations and future directions

Methods and products derived in this dissertation are associated with a range of limitations which are important to discuss. The main limitations that were encountered and possible future directions to address them are listed below.

- *Limitation:* The SGS used in Chapter 3 to establish a network of ground plots was representative by design of the range of structural conditions characterized by the SPL100. However, other types of information such as species composition, stand development stage and site index were not considered for stratification. Depending on the objectives of the ground plot selection, considering such attributes could be important, particularly if a focus on certain forest types is desired. Attributes from traditional stand-level forest inventory could have been integrated into the sampling process but have not in order to meet the objective of proposing a framework fully driven by SPL100 data

Future direction: Research is needed to provide a comprehensive overview of sampling methods particularly well-suited to the establishment of ground plot locations (Hawbaker et al., 2009; Junttila et al., 2013b). Frameworks and tools are needed to be made available to forest managers in order to ensure that existing or newly established ground plots are suitable for inventory or other applications. These research themes are currently being addressed with the development of the SGS R package (<https://github.com/tgoodbody/sgsR>; Goodbody et al., 2022), which will be accompanied by case studies demonstrating a range of applications for such sampling methods.

- *Limitation:* Chapter 3 and 4 focused on the Romeo Malette Forest, which despite its relatively large extent represents only 1.4% of Ontario's managed forests. In addition, non-boreal forest types occur in other regions of Ontario, such as in the Great Lakes-St. Lawrence mixewood forests and deciduous forest of south Ontario located at the border with the United States of America.

Future direction: Even though the EFI developed in Chapter 3 and 4 at the Romeo Malette Forest was successful, it will also be necessary to develop and assess EFI derived from SPL100 data in other types of forests. White et al. (2021a) have also demonstrated the operational implementation of a SPL100-driven EFI in a management unit of the Great Lakes-St. Lawrence mixewood forest but an Ontario-wide EFI has not been produced yet. Challenges associated to the extent and diversity of Ontario's forests, the cost of extensive field data collection programs and the harmonization of SPL100 products derived from different acquisitions are likely to be faced. Insights gained from research examining the optimization of plot location (Goodbody et al., 2022) as well as the transferability of forest attribute models (Tompalski et al., 2019) should be particularly relevant.

- *Limitation:* The SPL100 acquisition provides an accurate characterization of forest structure at a static point in time. Despite the reduction in ALS data acquisition costs associated with the SPL100, it is currently unlikely that such large-scale acquisitions will occur at the same rate as the periodic development of forest management plans (e.g., 10-year cycle in Ontario).

Future directions: Coops et al. (2022) highlighted the need for regular inventory updates and propose a near real-time forest inventory framework in which targeted fine-scale remotely sensed data and growth and yield models are used to update forest inventory attributes in areas having experienced disturbance. Coops et al. (2022) also suggest that growth inferred from systematic lidar samples acquired by spaceborne lidar missions such as ICESat-2 or GEDI could also provide an alternative to the use of growth models. Uncertainties associated with ICESat-2 and GEDI measures will likely limit the derivation of forest growth increments and their integration into

detailed inventory framework for managed forests. However, such concepts could be suitable in unmanaged forests where constraints on the precision and accuracy of forest resources estimates are not as strong as in managed forests.

- *Limitation:* Even though the wall-to-wall canopy height map derived in Chapter 6 is informative, there is a need to examine the value of such products to improve our understanding of the resilience and adaptation of northern boreal forest to disturbances. For example, canopy height recovery trends could be analyzed further by considering pixel-level trajectories and their relationship with climatic, spectral and geomorphological variables to name a few.

Future direction: The science objectives of both ICESat-2 and GEDI missions include the estimation of AGB and carbon balance (Dubayah et al., 2020; Neumann et al., 2019). As a result, a number of studies, utilizing a range of methods, are being carried out towards that objective (Duncanson et al., 2022, 2020; Potapov et al., 2021; Silva et al., 2021). Other fields aside from carbon accounting could also benefit from an increasing availability of spatially-explicit and wall-to-wall forest attributes products. Previous global canopy height maps produced by extrapolating GLAS measurements with MODIS data have been used for example to highlight relationships between canopy height and geophysical factors (Simard et al., 2018) or between forest structure and species richness (Gouveia et al., 2014). As illustrated in chapter 6, research examining the recovery of forests after disturbances could particularly benefit from the integration of structural information in addition to spectral trajectories that are currently used (White et al., 2022).

7.4. Closing statement

Spatially-explicit, timely and accurate estimates of forest attributes are needed in both managed and unmanaged forests of Canada to inform sustainable forest management and monitor the impact of climate

change on forested ecosystems. The uptake of ALS data into forest inventory programs has enhanced the three-dimensional characterization of forest structure. However, ALS acquisitions have been particularly targeted towards productive forest areas which are managed for timber procurement. The work carried out in this dissertation demonstrated the ability of the airborne SPL100 instrument to provide accurate estimates of key forest inventory attributes in a managed boreal forest of Ontario, Canada. Moreover, this dissertation assessed the performance of the ATLAS instrument onboard the ICESat-2 satellite to measure forest structure and demonstrated its value to monitor vegetation structure, particularly in unmanaged boreal forests. This dissertation shows that airborne and spaceborne SPL could have an important role to play in the future to (1) increase lidar data coverage both spatially and temporally, hence (2) having a better representation of poorly characterized forest types such as unmanaged boreal forests and (3) leading to the generation of novel products to address key information needs in both managed and unmanaged forests. My hope is that the work presented in this dissertation will be impactful and ultimately help taking informed decisions on future actions and policies in order to mitigate the impact of the changing climate on forests and adapt forest management practices accordingly.

Bibliography

- Achim, A., Moreau, G., Coops, N.C., Axelson, J.N., Barrette, J., Bédard, S., Byrne, K.E., Caspersen, J., Dick, A.R., D'Orangeville, L., Drolet, G., Eskelson, B.N.I., Filipescu, C.N., Flamand-Hubert, M., Goodbody, T.R.H., Griess, V.C., Hagerman, S.M., Keys, K., Lafleur, B., Girona, M.M., Morris, D.M., Nock, C.A., Pinno, B.D., Raymond, P., Roy, V., Schneider, R., Soucy, M., Stewart, B., Sylvain, J.-D., Taylor, A.R., Thiffault, E., Thiffault, N., Vepakomma, U., White, J.C., 2022. The changing culture of silviculture. *For. An Int. J. For. Res.* 95, 143–152. <https://doi.org/10.1093/FORESTRY/CPAB047>
- Adnan, S., Maltamo, M., Mehtätalo, L., Ammaturo, R.N.L., Packalen, P., Valbuena, R., 2021. Determining maximum entropy in 3D remote sensing height distributions and using it to improve aboveground biomass modelling via stratification. *Remote Sens. Environ.* 260, 112464. <https://doi.org/10.1016/j.rse.2021.112464>
- Alam, M.B., Shahi, C., Pulkki, R., 2014. Economic impact of enhanced forest inventory information and merchandizing yards in the forest product industry supply Chain. *Socioecon. Plann. Sci.* 48, 189–197. <https://doi.org/10.1016/j.seps.2014.06.002>
- Anderson-Teixeira, K.J., Miller, A.D., Mohan, J.E., Hudiburg, T.W., Duval, B.D., DeLucia, E.H., 2013. Altered dynamics of forest recovery under a changing climate. *Glob. Chang. Biol.* 19, 2001–2021. <https://doi.org/10.1111/GCB.12194>
- Andrew, M.E., Wulder, M.A., Coops, N.C., 2012. Identification of de facto protected areas in boreal Canada. *Biol. Conserv.* 146, 97–107. <https://doi.org/10.1016/J.BIOCON.2011.11.029>
- Astola, H., Häme, T., Sirro, L., Molinier, M., Kilpi, J., 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sens. Environ.* 223, 257–273. <https://doi.org/10.1016/J.RSE.2019.01.019>
- Axelsson, A., Lindberg, E., Olsson, H., 2018. Exploring Multispectral ALS Data for Tree Species Classification. *Remote Sens.* 2018, Vol. 10, Page 183 10, 183. <https://doi.org/10.3390/RS10020183>
- Banskota, A., Kayastha, N., Falkowski, M.J., Wulder, M.A., Froese, R.E., White, J.C., 2014. Forest Monitoring Using Landsat Time Series Data: A Review. *Can. J. Remote Sens.* <https://doi.org/10.1080/07038992.2014.987376>
- Bartels, S.F., Chen, H.Y.H., Wulder, M.A., White, J.C., 2016. Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. *For. Ecol. Manage.* 361, 194–207. <https://doi.org/10.1016/j.foreco.2015.11.015>
- Beaudoin, A., Bernier, P.Y., Villemaire, P., Guindon, L., Guo, X.J., 2018. Tracking forest attributes across Canada between 2001 and 2011 using a k nearest neighbors mapping approach applied to MODIS imagery. *Can. J. For. Res.* 48, 85–93. https://doi.org/10.1139/CJFR-2017-0184/SUPPL_FILE/CJFR-2017-0184SUPPLA.PDF
- Bilyk, A., Pulkki, R., Shahi, C., Larocque, G.R., 2021. Development of the ontario forest resources inventory: A historical review. *Can. J. For. Res.* <https://doi.org/10.1139/cjfr-2020-0234>
- Boisvenue, C., Smiley, B.P., White, J.C., Kurz, W.A., Wulder, M.A., 2016. Integration of Landsat time series and field plots for forest productivity estimates in decision support models. *For. Ecol. Manage.* 376, 284–297. <https://doi.org/10.1016/J.FORECO.2016.06.022>
- Boisvenue, Céline, Smiley, B.P., White, J.C., Kurz, W.A., Wulder, M.A., 2016. Improving carbon monitoring and reporting in forests using spatially-explicit information. *Carbon Balance Manag.* 11,

- 1–16. <https://doi.org/10.1186/s13021-016-0065-6>
- Boisvenue, C., White, J.C., 2019. Information needs of next-generation forest carbon models: Opportunities for remote sensing science. *Remote Sens.* 11. <https://doi.org/10.3390/rs11040463>
- Bolton, D.K., Coops, N.C., Wulder, M.A., 2015. Characterizing residual structure and forest recovery following high-severity fire in the western boreal of Canada using Landsat time-series and airborne lidar data. *Remote Sens. Environ.* 163, 48–60. <https://doi.org/10.1016/j.rse.2015.03.004>
- Bolton, D.K., Coops, N.C., Wulder, M.A., 2013. Investigating the agreement between global canopy height maps and airborne Lidar derived height estimates over Canada. *Can. J. Remote Sens.* 39. <https://doi.org/10.5589/m13-036>
- Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sens. Environ.* 240, 111685. <https://doi.org/10.1016/J.RSE.2020.111685>
- Brandt, J.P., 2009. The extent of the North American boreal zone. *Environ. Rev.* 17, 101–161. <https://doi.org/10.1139/A09-004/ASSET/IMAGES/A09-004T8H.GIF>
- Breidenbach, J., Astrup, R., 2014. The Semi-Individual Tree Crown Approach. Springer, Dordrecht, pp. 113–133. https://doi.org/10.1007/978-94-017-8663-8_6
- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., Solberg, S., 2010a. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sens. Environ.* 114, 911–924. <https://doi.org/10.1016/j.rse.2009.12.004>
- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., Solberg, S., 2010b. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sens. Environ.* 114, 911–924. <https://doi.org/10.1016/j.rse.2009.12.004>
- Breidenbach, J., Waser, L.T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S., Astrup, R., 2021. National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. *Can. J. For. Res.* 51, 365–379. <https://doi.org/10.1139/cjfr-2020-0170>
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brown, R., Hartzell, P., Glennie, C., 2020. Evaluation of SPL100 Single Photon Lidar Data. *Remote Sens.* 12, 722. <https://doi.org/10.3390/rs12040722>
- Budei, B.C., St-Onge, B., Hopkinson, C., Audet, F.A., 2018. Identifying the genus or species of individual trees using a three-wavelength airborne lidar system. *Remote Sens. Environ.* 204, 632–647. <https://doi.org/10.1016/J.RSE.2017.09.037>
- Carroll, M.L., Townshend, J.R., DiMiceli, C.M., Noojipady, P., Sohlberg, R.A., 2009. A new global raster water mask at 250 m resolution. <http://dx.doi.org/10.1080/17538940902951401> 2, 291–308. <https://doi.org/10.1080/17538940902951401>
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. <https://doi.org/10.1016/J.RSE.2018.09.002>
- Clifton, W.E., Steele, B., Nelson, G., Truscott, A., Itzler, M., Entwistle, M., 2015. Medium altitude airborne Geiger-mode mapping LIDAR system, in: Turner, M.D., Kamerman, G.W., Wasiczko Thomas, L.M.,

- Spillar, E.J. (Eds.), . International Society for Optics and Photonics, p. 946506. <https://doi.org/10.1117/12.2193827>
- Coops, N.C., Hermosilla, T., Wulder, M.A., White, J.C., Bolton, D.K., 2018. A thirty year, fine-scale, characterization of area burned in Canadian forests shows evidence of regionally increasing trends in the last decade. *PLoS One* 13, e0197218. <https://doi.org/10.1371/journal.pone.0197218>
- Coops, N.C., Tompalski, P., Goodbody, T.R.H., Achim, A., Mulverhill, C., 2022. Framework for near real-time forest inventory using multi source remote sensing data. *For. An Int. J. For. Res.* 1–19. <https://doi.org/10.1093/FORESTRY/CPAC015>
- Coops, N.C., Tompalski, P., Goodbody, T.R.H., Queinnec, M., Luther, J.E., Bolton, D.K., White, J.C., Wulder, M.A., van Lier, O.R., Hermosilla, T., 2021. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sens. Environ.* 260, 112477. <https://doi.org/10.1016/J.RSE.2021.112477>
- Coops, N.C., Wulder, M.A., Culvenor, D.S., St-Onge, B., 2004. Comparison of forest attributes extracted from fine spatial resolution multispectral and lidar data. *Can. J. Remote Sens.* 30, 855–866. <https://doi.org/10.5589/m04-045>
- Crookston, N.L., Finley, A.O., 2008. **yaImpute** : An R Package for k NN Imputation. *J. Stat. Softw.* 23, 1–16. <https://doi.org/10.18637/jss.v023.i10>
- D’Amato, A.W., Bradford, J.B., Fraver, S., Palik, B.J., 2011. Forest management for mitigation and adaptation to climate change: Insights from long-term silviculture experiments. *For. Ecol. Manage.* 262, 803–816. <https://doi.org/10.1016/J.FORECO.2011.05.014>
- Dalponte, M., Ørka, H.O., Ene, L.T., Gobakken, T., Næsset, E., 2014. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sens. Environ.* 140, 306–317. <https://doi.org/10.1016/J.RSE.2013.09.006>
- Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010), Open-File Report. <https://doi.org/10.3133/ofr20111073>
- DeFries, R., 2013. Why Forest Monitoring Matters for People and the Planet. *Glob. For. Monit. from Earth Obs.* 1–14. <https://doi.org/10.1201/b13040-1>
- Degnan, J.J., 2016. Scanning, multibeam, single photon lidars for rapid, large scale, high resolution, topographic and bathymetric mapping. *Remote Sens.* 8, 958. <https://doi.org/10.3390/rs8110958>
- Degnan, J.J., 2002. Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements. *J. Geodyn.* 34, 503–549. [https://doi.org/10.1016/S0264-3707\(02\)00045-5](https://doi.org/10.1016/S0264-3707(02)00045-5)
- Degnan, J.J., Field, C., Machan, R., Leventhal, E., Lawrence, D., Zheng, Y., Upton, R., Tillard, J., Howell, S., 2007. Recent Advances in Photon-Counting , 3D Imaging Lidars. pp. 8–13.
- Degnan, J.J., Field, C.T., 2014. Moderate to high altitude, single photon sensitive, 3D imaging lidars, in: *Advanced Photon Counting Techniques {VIII}*. International Society for Optics and Photonics, p. 91140H. <https://doi.org/10.1117/12.2049995>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* 120, 25–36. <https://doi.org/10.1016/J.RSE.2011.11.026>

- Dubayah, R.O., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L.I., Hancock, S., Jantz, P., Marselis, S., Patterson, P.L., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* 1, 100002. <https://doi.org/10.1016/j.srs.2020.100002>
- Duncanson, L.I., Dubayah, R.O., Cook, B.D., Rosette, J., Parker, G., 2015. The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation. *Remote Sens. Environ.* 168, 102–112. <https://doi.org/10.1016/J.RSE.2015.06.021>
- Duncanson, L.I., Kellner, J.R., Armston, J., Dubayah, R., Minor, D.M., Hancock, S., Healey, S.P., Patterson, P.L., Saarela, S., Marselis, S., Silva, C.E., Bruening, J., Goetz, S.J., Tang, H., Hofton, M., Blair, B., Luthcke, S., Fatoyinbo, L., Abernethy, K., Alonso, A., Andersen, H.E., Aplin, P., Baker, T.R., Barbier, N., Bastin, J.F., Biber, P., Boeckx, P., Bogaert, J., Boschetti, L., Boucher, P.B., Boyd, D.S., Burslem, D.F.R.P., Calvo-Rodriguez, S., Chave, J., Chazdon, R.L., Clark, D.B., Clark, D.A., Cohen, W.B., Coomes, D.A., Corona, P., Cushman, K.C., Cutler, M.E.J., Dalling, J.W., Dalponte, M., Dash, J., de-Miguel, S., Deng, S., Ellis, P.W., Erasmus, B., Fekety, P.A., Fernandez-Landa, A., Ferraz, A., Fischer, R., Fisher, A.G., García-Abril, A., Gobakken, T., Hacker, J.M., Heurich, M., Hill, R.A., Hopkinson, C., Huang, H., Hubbell, S.P., Hudak, A.T., Huth, A., Imbach, B., Jeffery, K.J., Katoh, M., Kearsley, E., Kenfack, D., Kljun, N., Knapp, N., Král, K., Krůček, M., Labrière, N., Lewis, S.L., Longo, M., Lucas, R.M., Main, R., Manzanera, J.A., Martínez, R.V., Mathieu, R., Memiaghe, H., Meyer, V., Mendoza, A.M., Monerri, A., Montesano, P., Morsdorf, F., Næsset, E., Naidoo, L., Nilus, R., O'Brien, M., Orwig, D.A., Papathanassiou, K., Parker, G., Philipson, C., Phillips, O.L., Pisek, J., Poulsen, J.R., Pretzsch, H., Rüdiger, C., Saatchi, S., Sanchez-Azofeifa, A., Sanchez-Lopez, N., Scholes, R., Silva, C.A., Simard, M., Skidmore, A., Stereńczak, K., Tanase, M., Torresan, C., Valbuena, R., Verbeeck, H., Vrska, T., Wessels, K., White, J.C., White, L.J.T., Zahabu, E., Zraggen, C., 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sens. Environ.* 270, 112845. <https://doi.org/10.1016/J.RSE.2021.112845>
- Duncanson, L.I., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C.A., Armston, J., Luthcke, S.B., Hofton, M., Kellner, J.R., Dubayah, R.O., 2020. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sens. Environ.* 242, 111779. <https://doi.org/10.1016/j.rse.2020.111779>
- Dwyer, J.L., Roy, D.P., Sauer, B., Jenkerson, C.B., Zhang, H.K., Lymburner, L., 2018. Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sens.* 2018, Vol. 10, Page 1363 10, 1363. <https://doi.org/10.3390/RS10091363>
- Ecological Stratification Working Group, 1996. A national ecological framework for Canada. Centre for Land and Biological Resources Research; Hull, Quebec.
- Eskelson, B.N.I., Temesgen, H., Lemay, V., Barrett, T.M., Crookston, N.L., Hudak, A.T., 2009. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. *Scand. J. For. Res.* 24, 235–246. <https://doi.org/10.1080/02827580902870490>
- FAO, 2016. Global Forest Resources Assessment 2015 - How are the world's forest changing? Rome.
- Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* 186, 64–87. <https://doi.org/10.1016/J.RSE.2016.08.013>
- Frazier, R.J., Coops, N.C., Wulder, M.A., 2015. Boreal Shield forest disturbance and recovery trends using Landsat time series. *Remote Sens. Environ.* 170, 317–327. <https://doi.org/10.1016/J.RSE.2015.09.015>

- Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., Philipson, C.D., Mikusiński, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J., Bengtsson, J., 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.* 2013 41 4, 1–8. <https://doi.org/10.1038/ncomms2328>
- Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A.Z., Schepaschenko, D.G., 2015. Boreal forest health and global change. *Science* (80-.). 349, 819–822. https://doi.org/10.1126/SCIENCE.AAA9092/SUPPL_FILE/GAUTHIER.SM.PDF
- Glenn, N.F., Neuenschwander, A., Vierling, L.A., Spaete, L., Li, A., Shinneman, D.J., Pilliod, D.S., Arkle, R.S., McIlroy, S.K., 2016. Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. *Remote Sens. Environ.* 185, 233–242. <https://doi.org/10.1016/J.RSE.2016.02.039>
- Gluckman, J., 2016. Design of the processing chain for a high-altitude, airborne, singlephoton lidar mapping instrument, in: Turner, M.D., Kamerman, G.W. (Eds.), *Laser Radar Technology and Applications XXI*. SPIE, p. 983203. <https://doi.org/10.1117/12.2219760>
- Gobakken, T., Korhonen, L., Næsset, E., 2013. Laser-assisted selection of field plots for an area-based forest inventory. *Silva Fenn.* 47. <https://doi.org/10.14214/sf.943>
- Goodbody, T.R.H., Coops, N.C., Queinnec, M., White, J.C., Tompalski, P., Hudak, A.T., Auty, D., Valbuena, R., Leboeuf, A., Sinclair, I., McCartney, G., Prieur, J.F., Woods, M.E., 2022. sgsR: A structurally guided sampling toolbox for lidar-based forest inventories.
- 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. Reports.* <https://doi.org/10.1007/s40725-019-00087-2>
- Goodwin, N.R., Coops, N.C., Wulder, M.A., Gillanders, S., Schroeder, T.A., Nelson, T., 2008. Estimation of insect infestation dynamics using a temporal sequence of Landsat data. *Remote Sens. Environ.* 112, 3680–3689. <https://doi.org/10.1016/J.RSE.2008.05.005>
- Gouveia, S.F., Villalobos, F., Dobrovolski, R., Beltrão-Mendes, R., Ferrari, S.F., 2014. Forest structure drives global diversity of primates. *J. Anim. Ecol.* 83, 1523–1530. <https://doi.org/10.1111/1365-2656.12241>
- Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sens.* 2019, Vol. 11, Page 1197 11, 1197. <https://doi.org/10.3390/RS11101197>
- Grafström, A., Ringvall, A.H., 2013. Improving forest field inventories by using remote sensing data in novel sampling designs. *Can. J. For. Res.* 43, 1015–1022. <https://doi.org/10.1139/cjfr-2013-0123>
- Gwenzi, D., Lefsky, M.A., 2014. Prospects of photon counting lidar for savanna ecosystem structural studies. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 40, 141–147. <https://doi.org/10.5194/isprsarchives-XL-1-141-2014>
- Hall, R.J., 2003. The Roles of Aerial Photographs in Forestry Remote Sensing Image Analysis. *Remote Sens. For. Environ.* 47–75. https://doi.org/10.1007/978-1-4615-0306-4_3
- Hansen, M.C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. *Science* (80-.). 342, 850–853. <https://doi.org/10.1126/science.1244693>

- Harding, D., 2018. Pulsed Laser Altimeter Ranging Techniques and Implications for Terrain Mapping, in: Topographic Laser Ranging and Scanning. CRC Press, Second edition. | Boca Raton : Taylor & Francis, CRC Press, 2018., pp. 201–220. <https://doi.org/10.1201/9781315154381-5>
- Harper, K.A., Macdonald, S.E., Burton, P.J., Chen, J., Brososke, K.D., Saunders, S.C., Euskirchen, E.S., Roberts, D., Jaiteh, M.S., Esseen, P.A., 2005. Edge Influence on Forest Structure and Composition in Fragmented Landscapes. *Conserv. Biol.* 19, 768–782. <https://doi.org/10.1111/J.1523-1739.2005.00045.X>
- Hartzell, P., Dang, Z., Pan, Z., Glennie, C., 2018. Radiometric evaluation of an airborne single photon LiDAR sensor. *IEEE Geosci. Remote Sens. Lett.* 15, 1466–1470. <https://doi.org/10.1109/LGRS.2018.2841811>
- Hawbaker, T.J., Keuler, N.S., Lesak, A.A., Gobakken, T., Contrucci, K., Radeloff, V.C., 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res. Biogeosciences* 114, n/a-n/a. <https://doi.org/10.1029/2008JG000870>
- Hemmerling, J., Pflugmacher, D., Hostert, P., 2021. Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sens. Environ.* 267, 112743. <https://doi.org/10.1016/J.RSE.2021.112743>
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., 2022. Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes. *Remote Sens. Environ.* 268, 112780. <https://doi.org/10.1016/J.RSE.2021.112780>
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2018. Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series. *Can. J. Remote Sens.* 44, 67–87. <https://doi.org/10.1080/07038992.2018.1437719>
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015a. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sens. Environ.* 158, 220–234. <https://doi.org/10.1016/j.rse.2014.11.005>
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015b. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. *Remote Sens. Environ.* 170, 121–132. <https://doi.org/10.1016/j.rse.2015.09.004>
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Campbell, L.B., 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. *Int. J. Digit. Earth* 9, 1035–1054. <https://doi.org/10.1080/17538947.2016.1187673>
- Holmgren, J., 2004. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scand. J. For. Res.* 19, 543–553. <https://doi.org/10.1080/02827580410019472>
- Holmgren, J., Persson, Å., 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sens. Environ.* 90, 415–423. [https://doi.org/10.1016/S0034-4257\(03\)00140-8](https://doi.org/10.1016/S0034-4257(03)00140-8)
- Holmgren, J., Persson, Å., Söderman, U., 2008. Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *Int. J. Remote Sens.* 29, 1537–1552. <https://doi.org/10.1080/01431160701736471>
- Honer, T.G., 1983. Metric timber tables for the commercial tree species of Central and Eastern Canada. *Can. For. Ser. Inst. Info FMR-X-5*, 139.
- Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4, 273–288. <https://doi.org/10.1080/01431168308948546>

- Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., Van Der Laan, M.J., 2006. Survival ensembles. *Biostatistics* 7, 355–373. <https://doi.org/10.1093/BIOSTATISTICS/KXJ011>
- Hovi, A., Raitio, P., Rautiainen, M., 2017. A spectral analysis of 25 boreal tree species. *Silva Fenn.* 51. <https://doi.org/10.14214/sf.7753>
- Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J., Guo, Q., 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sens.* 8, 565. <https://doi.org/10.3390/rs8070565>
- Hyypä, J., Hyypä, H., Leckie, D., 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, 1339–1366. <https://doi.org/10.1080/01431160701736489>
- Irwin, L., Coops, N.C., Queinnec, M., McCartney, G., White, J.C., 2021. Single photon lidar signal attenuation under boreal forest conditions. *Remote Sens. Lett.* 12, 1049–1060. <https://doi.org/10.1080/2150704X.2021.1962575>
- Isenburg, M., 2017. LAStools [WWW Document].
- Jarron, L.R., Coops, N.C., MacKenzie, W.H., Tompalski, P., Dykstra, P., 2020. Detection of sub-canopy forest structure using airborne LiDAR. *Remote Sens. Environ.* 244, 111770. <https://doi.org/10.1016/j.rse.2020.111770>
- Joyce, M.J., Erb, J.D., Sampson, B.A., Moen, R.A., 2019. Detection of coarse woody debris using airborne light detection and ranging (LiDAR). *For. Ecol. Manage.* 433, 678–689. <https://doi.org/10.1016/j.foreco.2018.11.049>
- Junttila, V., Finley, A.O., Bradford, J.B., Kauranne, T., 2013a. Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory. *For. Ecol. Manage.* 292, 75–85. <https://doi.org/10.1016/j.foreco.2012.12.019>
- Junttila, V., Finley, A.O., Bradford, J.B., Kauranne, T., 2013b. Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory. *For. Ecol. Manage.* 292, 75–85. <https://doi.org/10.1016/j.foreco.2012.12.019>
- Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., Roy, D.P., Morisette, J.T., 2002. An overview of MODIS Land data processing and product status. *Remote Sens. Environ.* 83, 3–15. [https://doi.org/10.1016/S0034-4257\(02\)00084-6](https://doi.org/10.1016/S0034-4257(02)00084-6)
- Kaartinen, H., Hyypä, J., Yu, X., Vastaranta, M., Hyypä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S.C., Solberg, S., Wolf, B.M., Wu, J.-C., 2012. An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. *Remote Sens.* 4, 950–974. <https://doi.org/10.3390/rs4040950>
- Kane, V.R., Gillespie, A.R., Mcgaughey, R., Lutz, J.A., Ceder, K., Franklin, J.F., 2008. Interpretation and topographic compensation of conifer canopy self-shadowing. <https://doi.org/10.1016/j.rse.2008.06.001>
- Korhonen, L., Vauhkonen, J., Virolainen, A., Hovi, A., Korpela, I., 2013. Estimation of tree crown volume from airborne lidar data using computational geometry. <http://dx.doi.org/10.1080/01431161.2013.817715> 34, 7236–7248. <https://doi.org/10.1080/01431161.2013.817715>
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, the R.C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., 2015. caret: Classification

- and Regression Training, <https://CRAN.R-project.org/package=caret>. R Packag. version.
- Kukkonen, M., Korhonen, L., Maltamo, M., Suvanto, A., Packalen, P., 2018. How much can airborne laser scanning based forest inventory by tree species benefit from auxiliary optical data? *Int. J. Appl. Earth Obs. Geoinf.* 72, 91–98. <https://doi.org/10.1016/J.JAG.2018.06.017>
- Kukkonen, M., Maltamo, M., Korhonen, L., Packalen, P., 2019. Multispectral Airborne LiDAR Data in the Prediction of Boreal Tree Species Composition. *IEEE Trans. Geosci. Remote Sens.* 57, 3462–3471. <https://doi.org/10.1109/TGRS.2018.2885057>
- Kurz, W.A., Shaw, C.H., Boisvenue, C., Stinson, G., Metsaranta, J., Leckie, D., Dyk, A., Smyth, C., Neilson, E.T., 2013. Carbon in Canada's boreal forest-A synthesis. *Environ. Rev.* 21, 260–292. <https://doi.org/10.1139/ER-2013-0041/ASSET/IMAGES/LARGE/ER-2013-0041F10.JPEG>
- Lang, N., Schindler, K., Wegner, J.D., 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sens. Environ.* 233, 111347. <https://doi.org/10.1016/J.RSE.2019.111347>
- Leckie, D., Gillis, M.D., 1995. Forest inventory in Canada with emphasis on map production. *For. Chron.* 71, 74–88. <https://doi.org/10.5558/tfc71074-1>
- Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., Shreenan, R., 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Can. J. Remote Sens.* 29, 633–649. <https://doi.org/10.5589/m03-024>
- Lefsky, M.A., 2010. A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system. *Geophys. Res. Lett.* 37. <https://doi.org/10.1029/2010GL043622>
- Lefsky, M.A., Hudak, A.T., Cohen, W.B., Acker, S.A., 2005. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sens. Environ.* 95, 517–531. <https://doi.org/10.1016/j.rse.2005.01.004>
- Lefsky, M.A., Keller, M., Pang, Y., De Camargo, P.B., Hunter, M.O., 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. spiedigitallibrary.org. <https://doi.org/10.1117/1.2795724>
- Leutner, B., Horning, N., Schwalb-Willmann, J., 2019. RStoolbox: Tools for Remote Sensing Data Analysis.
- Li, Q., Degnan, J.J., Barrett, T., Shan, J., 2016. First Evaluation on Single Photon-Sensitive Lidar Data. *Photogramm. Eng. Remote Sens.* 82, 455–463. <https://doi.org/10.4236/aces.2017.74031>
- Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., Chen, H., 2020. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *Int. J. Appl. Earth Obs. Geoinf.* 92, 102163. <https://doi.org/10.1016/J.JAG.2020.102163>
- Li, Y., Andersen, H.E., McGaughey, R., 2008. A Comparison of statistical methods for estimating forest biomass from light detection and ranging data. *West. J. Appl. For.* 23, 223–231. <https://doi.org/10.1093/wjaf/23.4.223>
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R news* 2, 18–22. <https://doi.org/10.1023/A:1010933404324>
- Lim, K., Hopkinson, C., Treitz, P., 2008. Examining the effects of sampling point densities on laser canopy height and density metrics. *For. Chron.* 84, 876–885. <https://doi.org/10.5558/tfc84876-6>
- Lim, K., Treitz, P., Wulder, M.A., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure.

- Prog. Phys. Geogr. 27, 88–106. <https://doi.org/10.1191/0309133303pp360ra>
- Lindberg, E., Holmgren, J., Olofsson, K., Wallerman, J., Olsson, H., 2013. Estimation of tree lists from airborne laser scanning using tree model clustering and k-MSN imputation. *Remote Sens.* 5, 1932–1955. <https://doi.org/10.3390/rs5041932>
- Luthcke, S.B., Thomas, T.C., Pennington, T.A., Rebold, T.W., Nicholas, J.B., Rowlands, D.D., Gardner, A.S., Bae, S., 2021. ICESat-2 Pointing Calibration and Geolocation Performance. *Earth Sp. Sci.* 8, e2020EA001494. <https://doi.org/10.1029/2020ea001494>
- Magruder, L.A., Brunt, K.M., Alonzo, M., 2020. Early ICESat-2 on-orbit Geolocation Validation Using Ground-Based Corner Cube Retro-Reflectors. *Remote Sens.* 12, 3653. <https://doi.org/10.3390/rs12213653>
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2 3. <https://doi.org/10.1117/12.2278218>
- Maltamo, M., Bollandsas, O.M., Naesset, E., Gobakken, T., Packalen, P., 2011. Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry* 84, 23–31. <https://doi.org/10.1093/forestry/cpq039>
- Maltamo, M., Packalen, P., 2014. Species-Specific Management Inventory in Finland. Springer, Dordrecht, pp. 241–252. https://doi.org/10.1007/978-94-017-8663-8_12
- Maltamo, M., Packalen, P., 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/ASSET/IMAGES/LARGE/CJFR-2020-0322F3.JPEG>
- Margolis, H.A., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.E., Wulder, M.A., 2015. Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. *Can. J. For. Res.* 45, 838–855. <https://doi.org/10.1139/cjfr-2015-0006>
- Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S.C., Shum, C.K., Schutz, B.E., Smith, B., Yang, Y., Zwally, J., 2017. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sens. Environ.* 190, 260–273. <https://doi.org/10.1016/j.rse.2016.12.029>
- Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. *Geosci. Remote Sens. Lett. IEEE* 3, 68–72.
- Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Bolton, D.K., Tompalski, P., Bater, C.W., 2018a. Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots. *Remote Sens. Environ.* 216, 697–714. <https://doi.org/10.1016/j.rse.2018.07.024>
- Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S.J., 2018b. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sens. Environ.* 209, 90–106. <https://doi.org/10.1016/j.rse.2017.12.020>
- McDowell, N.G., Allen, C.D., Anderson-Teixeira, K., Aukema, B.H., Bond-Lamberty, B., Chini, L., Clark,

- J.S., Dietze, M., Grossiord, C., Hanbury-Brown, A., Hurtt, G.C., Jackson, R.B., Johnson, D.J., Kueppers, L., Lichstein, J.W., Ogle, K., Poulter, B., Pugh, T.A.M., Seidl, R., Turner, M.G., Uriarte, M., Walker, A.P., Xu, C., 2020. Pervasive shifts in forest dynamics in a changing world. *Science* (80- .). 368. https://doi.org/10.1126/SCIENCE.AAZ9463/ASSET/4212D388-6B0E-473F-A252-1181403BCE3F/ASSETS/GRAPHIC/368_AAZ9463_FA.JPEG
- McGill, M., Markus, T., Scott, V.S., Neumann, T.A., 2013. The Multiple Altimeter Beam Experimental Lidar (MABEL): An Airborne Simulator for the *ICESat-2* Mission. *J. Atmos. Ocean. Technol.* 30, 345–352. <https://doi.org/10.1175/JTECH-D-12-00076.1>
- Michałowska, M., Rapiński, J., 2021. A review of tree species classification based on airborne lidar data and applied classifiers. *Remote Sens.* 13, 1–27. <https://doi.org/10.3390/rs13030353>
- Middleton, E.M., Chan, S.S., Rusin, R.I., Mitchell, S.K., 1997. Optical properties of black spruce and jack pine needles at BOREAS Sites in saskatchewan, Canada. *Can. J. Remote Sens.* 23, 108–119. <https://doi.org/10.1080/07038992.1997.10855193>
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* 32, 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>
- Ministry of Natural Resources and Forestry, 2016. Ontario Specifications for Lidar Acquisition [WWW Document]. https://www.bespatialontario.ca/resources/Presentations/2018_LiDAR/LiDARSpeccs/OntarioSpecificationsForLidarAcquisition2016.pdf URL
- Montesano, P.M., Rosette, J., Sun, G., North, P., Nelson, R.F., Dubayah, R.O., Ranson, K.J., Kharuk, V., 2015. The uncertainty of biomass estimates from modeled ICESat-2 returns across a boreal forest gradient. *Remote Sens. Environ.* 158, 95–109. <https://doi.org/10.1016/J.RSE.2014.10.029>
- Morresi, D., Vitali, A., Urbinati, C., Garbarino, M., 2019. Forest Spectral Recovery and Regeneration Dynamics in Stand-Replacing Wildfires of Central Apennines Derived from Landsat Time Series. *Remote Sens.* 2019, Vol. 11, Page 308 11, 308. <https://doi.org/10.3390/RS11030308>
- Moudrý, V., Gdulová, K., Gábor, L., Šárovcová, E., Barták, V., Leroy, F., Špatenková, O., Rocchini, D., Prošek, J., 2022. Effects of environmental conditions on ICESat-2 terrain and canopy heights retrievals in Central European mountains. *Remote Sens. Environ.* 279, 113112. <https://doi.org/10.1016/J.RSE.2022.113112>
- Næsset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scand. J. For. Res.* 19, 164–179. <https://doi.org/10.1080/02827580310019257>
- 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](https://doi.org/10.1016/S0034-4257(01)00290-5)
- Næsset, E., 1997. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* 52, 49–56. [https://doi.org/10.1016/S0924-2716\(97\)83000-6](https://doi.org/10.1016/S0924-2716(97)83000-6)
- Nandy, S., Srinet, R., Padalia, H., 2021. Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophys. Res. Lett.* 48, e2021GL093799. <https://doi.org/10.1029/2021GL093799>
- Narine, L., Popescu, S.C., Malambo, L., 2020. Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example. *Remote Sens.* 2020, Vol. 12, Page 1824 12, 1824. <https://doi.org/10.3390/RS12111824>

- Narine, L., Popescu, S.C., Malambo, L., 2019a. Synergy of ICESat-2 and Landsat for Mapping Forest Aboveground Biomass with Deep Learning. *Remote Sens.* 11, 1503. <https://doi.org/10.3390/rs11121503>
- Narine, L., Popescu, S.C., Neuenschwander, A., Zhou, T., Srinivasan, S., Harbeck, K., 2019b. Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. *Remote Sens. Environ.* 224, 1–11. <https://doi.org/10.1016/J.RSE.2019.01.037>
- Narine, L., Popescu, S.C., Zhou, T., Srinivasan, S., Harbeck, K., 2019c. Mapping forest aboveground biomass with a simulated ICESat-2 vegetation canopy product and Landsat data, *Ann. For. Res.* _.
- Natural Resources Canada, 2018. The State of Canada's Forests. Annual Report 2018. Ottawa.
- Nelson, R., 2013. How did we get here? An early history of forestry lidar ¹. *Can. J. Remote Sens.* 39, S6–S17. <https://doi.org/10.5589/m13-011>
- Nelson, R., Swift, R., Krabill, W., 1988. Using Airborne Lasers to Estimate Forest Canopy and Stand Characteristics. *J. For.* 86, 31–38. <https://doi.org/10.1093/JOF/86.10.31>
- Neuenschwander, A., Guenther, E., White, J.C., Duncanson, L.I., Montesano, P.M., 2020a. Validation of ICESat-2 terrain and canopy heights in boreal forests. *Remote Sens. Environ.* 251, 112110. <https://doi.org/10.1016/j.rse.2020.112110>
- Neuenschwander, A., Magruder, L., 2016. The Potential Impact of Vertical Sampling Uncertainty on ICESat-2/ATLAS Terrain and Canopy Height Retrievals for Multiple Ecosystems. *Remote Sens.* 8, 1039. <https://doi.org/10.3390/rs8121039>
- Neuenschwander, A., Magruder, L.A., 2019. Canopy and Terrain Height Retrievals with ICESat-2: A First Look. *Remote Sens.* 11, 1721. <https://doi.org/10.3390/rs11141721>
- Neuenschwander, A., Pitts, K., 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sens. Environ.* 221, 247–259. <https://doi.org/10.1016/J.RSE.2018.11.005>
- Neuenschwander, A., Popescu, S.C., Nelson, R.F., Harding, D., Pitts, K.L., Robbins, J., 2020b. ATLAS/ICESat-2 L3A Land and Vegetation Height, version 3. NASA Natl. Snow Ice Data Cent. DAAC. <https://doi.org/10.5067/ATLAS/ATL08.003>
- Neumann, T.A., Brenner, A., Hancock, D., Robbins, J., Luthcke, S.B., Harbeck, K., Lee, J., Gibbons, A., Saba, J., Brunt, K., 2020. ATLAS/ICESat-2 L2A Global Geolocated Photon Data, version 3. NASA Natl. Snow Ice Data Cent. DAAC. <https://doi.org/10.5067/ATLAS/ATL03.003>
- Neumann, T.A., Martino, A.J., Markus, T., Bae, S., Bock, M.R., Brenner, A.C., Brunt, K.M., Cavanaugh, J., Fernandes, S.T., Hancock, D.W., Harbeck, K., Lee, J., Kurtz, N.T., Luers, P.J., Luthcke, S.B., Magruder, L., Pennington, T.A., Ramos-Izquierdo, L., Rebold, T., Skoog, J., Thomas, T.C., 2019. The Ice, Cloud, and Land Elevation Satellite – 2 mission: A global geolocated photon product derived from the Advanced Topographic Laser Altimeter System. *Remote Sens. Environ.* 233. <https://doi.org/10.1016/j.rse.2019.111325>
- Nie, S., Wang, C., Xi, X., Luo, S., Li, G., Tian, J., Wang, H., 2018. Estimating the vegetation canopy height using micro-pulse photon-counting LiDAR data. *Opt. Express* 26, A520. <https://doi.org/10.1364/OE.26.00A520>
- Ontario Ministry of Natural Resources (OMNR), 2013. Enhancing Ontario's Forest Resources Inventory [WWW Document]. URL <http://www.forestryfutures.ca/upload/464883/documents/1F58FA93086BBFBD.pdf> (accessed 5.11.21).

- Ørka, H.O., Dalponte, M., Gobakken, T., Næsset, E., Ene, L.T., 2013. Characterizing forest species composition using multiple remote sensing data sources and inventory approaches. *Scand. J. For. Res.* 28, 677–688. <https://doi.org/10.1080/02827581.2013.793386>
- Palm, S., Yang, Y., Herzfeld, U., 2020. ICESat-2 Algorithm Theoretical Basis Document for Atmospheric Data ICESat-2 Algorithm Theoretical Basis Document for the Atmosphere, Part I: Level 2 and 3 Data Products, nsidc.org.
- Papa, D. de A., Almeida, D.R.A. de, Silva, C.A., Figueiredo, E.O., Stark, S.C., Valbuena, R., Rodriguez, L.C.E., d'Oliveira, M.V.N., 2020. Evaluating tropical forest classification and field sampling stratification from lidar to reduce effort and enable landscape monitoring. *For. Ecol. Manage.* 457, 117634. <https://doi.org/10.1016/j.foreco.2019.117634>
- Persson, M., Lindberg, E., Reese, H., 2018. Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sens.* 2018, Vol. 10, Page 1794 10, 1794. <https://doi.org/10.3390/RS10111794>
- Pickell, P.D., Hermosilla, T., Frazier, R.J., Coops, N.C., Wulder, M.A., 2016. Forest recovery trends derived from Landsat time series for North American boreal forests. *Int. J. Remote Sens.* 37, 138–149. <https://doi.org/10.1080/2150704X.2015.1126375>
- Pinto, F., Rouillard, D., Sobze, J.M., Ter-Mikaelian, M., 2007. Validating tree species composition in forest resource inventory for Nipissing Forest, Ontario, Canada. *For. Chron.* 83, 247–251. <https://doi.org/10.5558/tfc83247-2>
- Plakman, V., Janssen, T., Brouwer, N., Veraverbeke, S., 2020. Mapping Species at an Individual-Tree Scale in a Temperate Forest, Using Sentinel-2 Images, Airborne Laser Scanning Data, and Random Forest Classification. *Remote Sens.* 2020, Vol. 12, Page 3710 12, 3710. <https://doi.org/10.3390/RS12223710>
- Popescu, S.C., Wynne, R.H., 2004. Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogramm. Eng. Remote Sensing* 70, 589–604. <https://doi.org/10.14358/PERS.70.5.589>
- Popescu, S.C., Zhou, T., Nelson, R., Neuenschwander, A., Sheridan, R., Narine, L., Walsh, K.M., 2018. Photon counting LiDAR: An adaptive ground and canopy height retrieval algorithm for ICESat-2 data. *Remote Sens. Environ.* 208, 154–170. <https://doi.org/10.1016/j.rse.2018.02.019>
- Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., Armston, J., Dubayah, R.O., Blair, J.B., Hofton, M., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* 253, 112165. <https://doi.org/10.1016/J.RSE.2020.112165>
- Prieur, J.F., St-Onge, B., Fournier, R.A., Woods, M.E., Rana, P., Kneeshaw, D., 2022. A comparison of three airborne laser scanner types for species identification of individual trees. *Sensors* 22, 35. <https://doi.org/10.3390/s22010035>
- Queinnec, M., Coops, N.C., White, J.C., McCartney, G., Sinclair, I., 2021a. Developing a forest inventory approach using airborne single photon lidar data: from ground plot selection to forest attribute prediction. *For. An Int. J. For. Res.* <https://doi.org/10.1093/forestry/cpab051>
- Queinnec, M., Tompalski, P., Bolton, D.K., Coops, N.C., 2021b. FOSTER—An R package for forest structure extrapolation. *PLoS One* 16, e0244846. <https://doi.org/10.1371/JOURNAL.PONE.0244846>
- Queinnec, M., White, J.C., Coops, N.C., 2021c. Comparing airborne and spaceborne photon-counting LiDAR canopy structural estimates across different boreal forest types. *Remote Sens. Environ.* 262, 112510. <https://doi.org/10.1016/J.RSE.2021.112510>

- R Core Team, 2022. R: A Language and Environment for Statistical Computing.
- Ranghetti, L., Boschetti, M., Nutini, F., Busetto, L., 2020. “sen2r”: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Comput. Geosci.* 139, 104473. <https://doi.org/10.1016/J.CAGEO.2020.104473>
- Raty, J., Varvia, P., Korhonen, L., Savolainen, P., Maltamo, M., Packalen, P., 2021. A Comparison of Linear-Mode and Single-Photon Airborne LiDAR in Species-Specific Forest Inventories. *IEEE Trans. Geosci. Remote Sens.* 1–14. <https://doi.org/10.1109/TGRS.2021.3060670>
- Roussel, J.-R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R.H., Meador, A.S., Bourdon, J.-F., de Boissieu, F., Achim, A., 2020. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* 251, 112061. <https://doi.org/10.1016/j.rse.2020.112061>
- Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T.G., Nelson, R.F., Ståhl, G., 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. *Ann. For. Sci.* 73, 895–910. <https://doi.org/10.1007/s13595-016-0590-1>
- Schneider, F.D., Ferraz, A., Hancock, S., Duncanson, L.I., Dubayah, R.O., Pavlick, R.P., Schimel, D.S., 2020. Towards mapping the diversity of canopy structure from space with GEDI. *Environ. Res. Lett.* 15, 115006. <https://doi.org/10.1088/1748-9326/ab9e99>
- Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D., DiMarzio, J.P., 2005. Overview of the ICESat mission. *Geophys. Res. Lett.* <https://doi.org/10.1029/2005GL024009>
- Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M.J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T.A., Reyer, C.P.O., 2017. Forest disturbances under climate change. *Nat. Clim. Chang.* <https://doi.org/10.1038/nclimate3303>
- Shi, Y., Wang, T., Skidmore, A.K., Heurich, M., 2018. Important LiDAR metrics for discriminating forest tree species in Central Europe. *ISPRS J. Photogramm. Remote Sens.* 137, 163–174. <https://doi.org/10.1016/J.ISPRSJPRS.2018.02.002>
- Silva, C.A., Duncanson, L.I., Hancock, S., Neuenschwander, A., Thomas, N., Hofton, M., Fatoyinbo, L., Simard, M., Marshak, C.Z., Armston, J., Lutchke, S., Dubayah, R.O., 2021. Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. *Remote Sens. Environ.* 253, 112234. <https://doi.org/10.1016/J.RSE.2020.112234>
- Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V.H., Castañeda-Moya, E., Thomas, N., Van der Stocken, T., 2018. Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nat. Geosci.* 2018 121 12, 40–45. <https://doi.org/10.1038/s41561-018-0279-1>
- Simard, M., Pinto, N., Fisher, J.B., Baccini, A., 2011. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res.* 116, G04021. <https://doi.org/10.1029/2011JG001708>
- Stephens, P.R., Kimberley, M.O., Beets, P.N., Paul, T.S.H., Searles, N., Bell, A., Brack, C., Broadley, J., 2012. Airborne scanning LiDAR in a double sampling forest carbon inventory. *Remote Sens. Environ.* 117, 348–357. <https://doi.org/10.1016/j.rse.2011.10.009>
- Stinson, G., Thandi, G., Aitkin, D., Bailey, C., Boyd, J., Colley, M., Fraser, C., Gelhorn, L., Groenewegen, K., Hogg, A., Kapron, J., Leboeuf, A., Makar, M., Montigny, M., Pittman, B., Price, K., Salkeld, T., Smith, L., Viveiros, A., Wilson, D., 2019. A new approach for mapping forest management areas in Canada. *For. Chron.* 95, 101–112. <https://doi.org/10.5558/tfc2019-017>
- Stoker, J.M., Abdullah, Q.A., Nayegandhi, A., Winehouse, J., 2016. Evaluation of single photon and Geiger

- mode lidar for the 3D Elevation Program. *Remote Sens.* 8, 767. <https://doi.org/10.3390/rs8090767>
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinforma.* 2007 81 8, 1–21. <https://doi.org/10.1186/1471-2105-8-25>
- Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9, 1–11. <https://doi.org/10.1186/1471-2105-9-307/FIGURES/4>
- Sun, G., Ranson, K.J., Kimes, D.S., Blair, J.B., Kovacs, K., 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sens. Environ.* 112, 107–117. <https://doi.org/10.1016/j.rse.2006.09.036>
- Swatantran, A., Tang, H., Barrett, T., Decola, P., Dubayah, R.O., 2016. Rapid, high-resolution forest structure and terrain mapping over large areas using single photon lidar. *Sci. Rep.* 6, 28277. <https://doi.org/10.1038/srep28277>
- Tang, H., Swatantran, A., Barrett, T., DeCola, P., Dubayah, R.O., 2016. Voxel-based spatial filtering method for canopy height retrieval from airborne single-photon lidar. *Remote Sens.* 8, 771. <https://doi.org/10.3390/rs8090771>
- Ter-Mikaelian, M.T., Korzukhin, M.D., 1997. Biomass equations for sixty-five North American tree species. *For. Ecol. Manage.* 97, 1–24. [https://doi.org/10.1016/S0378-1127\(97\)00019-4](https://doi.org/10.1016/S0378-1127(97)00019-4)
- Thompson, I.D., Maher, S.C., Rouillard, D.P., Fryxell, J.M., Baker, J.A., 2007. Accuracy of forest inventory mapping: Some implications for boreal forest management. *For. Ecol. Manage.* 252, 208–221. <https://doi.org/10.1016/j.foreco.2007.06.033>
- Tompalski, P., Coops, N.C., Marshall, P.L., White, J.C., Wulder, M.A., Bailey, T., 2018. Combining Multi-Date Airborne Laser Scanning and Digital Aerial Photogrammetric Data for Forest Growth and Yield Modelling. *Remote Sens.* 2018, Vol. 10, Page 347 10, 347. <https://doi.org/10.3390/RS10020347>
- Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., 2014. Simulating the impacts of error in species and height upon tree volume derived from airborne laser scanning data. *For. Ecol. Manage.* 327, 167–177. <https://doi.org/10.1016/j.foreco.2014.05.011>
- Tompalski, P., White, J.C., Coops, N.C., Wulder, M.A., 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. *Remote Sens. Environ.* 227, 110–124. <https://doi.org/10.1016/j.rse.2019.04.006>
- Tompalski, P., White, J.C., Coops, N.C., Wulder, M.A., Leboeuf, A., Sinclair, I., Butson, C.R., Lemonde, M.O., 2021. Quantifying the precision of forest stand height and canopy cover estimates derived from air photo interpretation. *For. An Int. J. For. Res.* 94, 611–629. <https://doi.org/10.1093/FORESTRY/CPAB022>
- Treitz, P., Lim, K., Woods, M.E., Pitt, D., Nesbitt, D., Etheridge, D., 2012. LiDAR Sampling Density for Forest Resource Inventories in Ontario, Canada. *Remote Sens.* 4, 830–848. <https://doi.org/10.3390/rs4040830>
- Ullrich, A., Pfennigbauer, M., 2016. Linear LIDAR versus Geiger-mode LIDAR: impact on data properties and data quality, in: Turner, M.D., Kamerman, G.W. (Eds.), . International Society for Optics and Photonics, p. 983204. <https://doi.org/10.1117/12.2223586>
- Ung, C.-H., Bernier, P., Guo, X.-J., 2008. Canadian national biomass equations: new parameter estimates that include British Columbia data. *Can. J. For. Res.* 38, 1123–1132. <https://doi.org/10.1139/X07-224>

- Ung, C.H., Guo, X.J., Fortin, M., 2013. Canadian national taper models. *For. Chron.* 89, 211–224. <https://doi.org/10.5558/tfc2013-040>
- Valbuena, R., Maltamo, M., Mehtätalo, L., Packalen, P., 2017. Key structural features of Boreal forests may be detected directly using L-moments from airborne lidar data. *Remote Sens. Environ.* 194, 437–446. <https://doi.org/10.1016/j.rse.2016.10.024>
- Vastaranta, M., Kankare, V., Holopainen, M., Yu, X., Hyypä, J., Hyypä, H., 2012. Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data. *ISPRS J. Photogramm. Remote Sens.* 67, 73–79. <https://doi.org/10.1016/J.ISPRSJPRS.2011.10.006>
- Véga, C., Renaud, J.P., Durrieu, S., Bouvier, M., 2016. On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. *Remote Sens. Environ.* 175, 32–42. <https://doi.org/10.1016/j.rse.2015.12.039>
- Venier, L.A., Walton, R., Thompson, I.D., Arsenault, A., Titus, B.D., 2018. A review of the intact forest landscape concept in the Canadian boreal forest: its history, value, and measurement. <https://doi.org/10.1139/er-2018-0041> 26, 369–377. <https://doi.org/10.1139/ER-2018-0041>
- Wang, X., Pan, Z., Glennie, C., 2016. A Novel Noise Filtering Model for Photon-Counting Laser Altimeter Data. *IEEE Geosci. Remote Sens. Lett.* 13, 947–951. <https://doi.org/10.1109/LGRS.2016.2555308>
- Waring, R.H., Running, S.S., 2007. *Forest Ecosystems*. Elsevier. <https://doi.org/10.1016/B978-0-12-370605-8.X5001-4>
- Wästlund, A., Holmgren, J., Lindberg, E., Olsson, H., 2018. Forest variable estimation using a high altitude single photon lidar system. *Remote Sens.* 10, 1422. <https://doi.org/10.3390/rs10091422>
- Wehr, A., Lohr, U., 1999. Airborne laser scanning—an introduction and overview. *ISPRS J. Photogramm. Remote Sens.* 54, 68–82. [https://doi.org/10.1016/S0924-2716\(99\)00011-8](https://doi.org/10.1016/S0924-2716(99)00011-8)
- 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, 619–641. <https://doi.org/10.1080/07038992.2016.1207484>
- White, J.C., Hermosilla, T., Wulder, M.A., Coops, N.C., 2022. Mapping, validating, and interpreting spatio-temporal trends in post-disturbance forest recovery. *Remote Sens. Environ.* 271, 112904. <https://doi.org/10.1016/J.RSE.2022.112904>
- White, J.C., Penner, M., Woods, M.E., 2021a. Assessing single photon LiDAR for operational implementation of an enhanced forest inventory in diverse mixedwood forests. *For. Chron.* 97, 78–96. <https://doi.org/10.5558/tfc2021-009>
- White, J.C., Saarinen, N., Kankare, V., Wulder, M.A., Hermosilla, T., Coops, N.C., Pickell, P.D., Holopainen, M., Hyypä, J., Vastaranta, M., 2018. Confirmation of post-harvest spectral recovery from Landsat time series using measures of forest cover and height derived from airborne laser scanning data. *Remote Sens. Environ.* 216, 262–275. <https://doi.org/10.1016/J.RSE.2018.07.004>
- White, J.C., Tompalski, P., Vastaranta, M.A., Wulder, M.A., Saarinen, N.P., Stepper, C., Coops, N.C., 2017a. A model development and application guide for generating an enhanced forest inventory using airborne laser scanning data and an area-based approach.
- White, J.C., Woods, M.E., Krahn, T., Papasodoro, C., Bélanger, D., Onafrychuk, C., Sinclair, I., 2021b. Evaluating the capacity of single photon lidar for terrain characterization under a range of forest conditions. *Remote Sens. Environ.* 252, 112169. <https://doi.org/10.1016/j.rse.2020.112169>

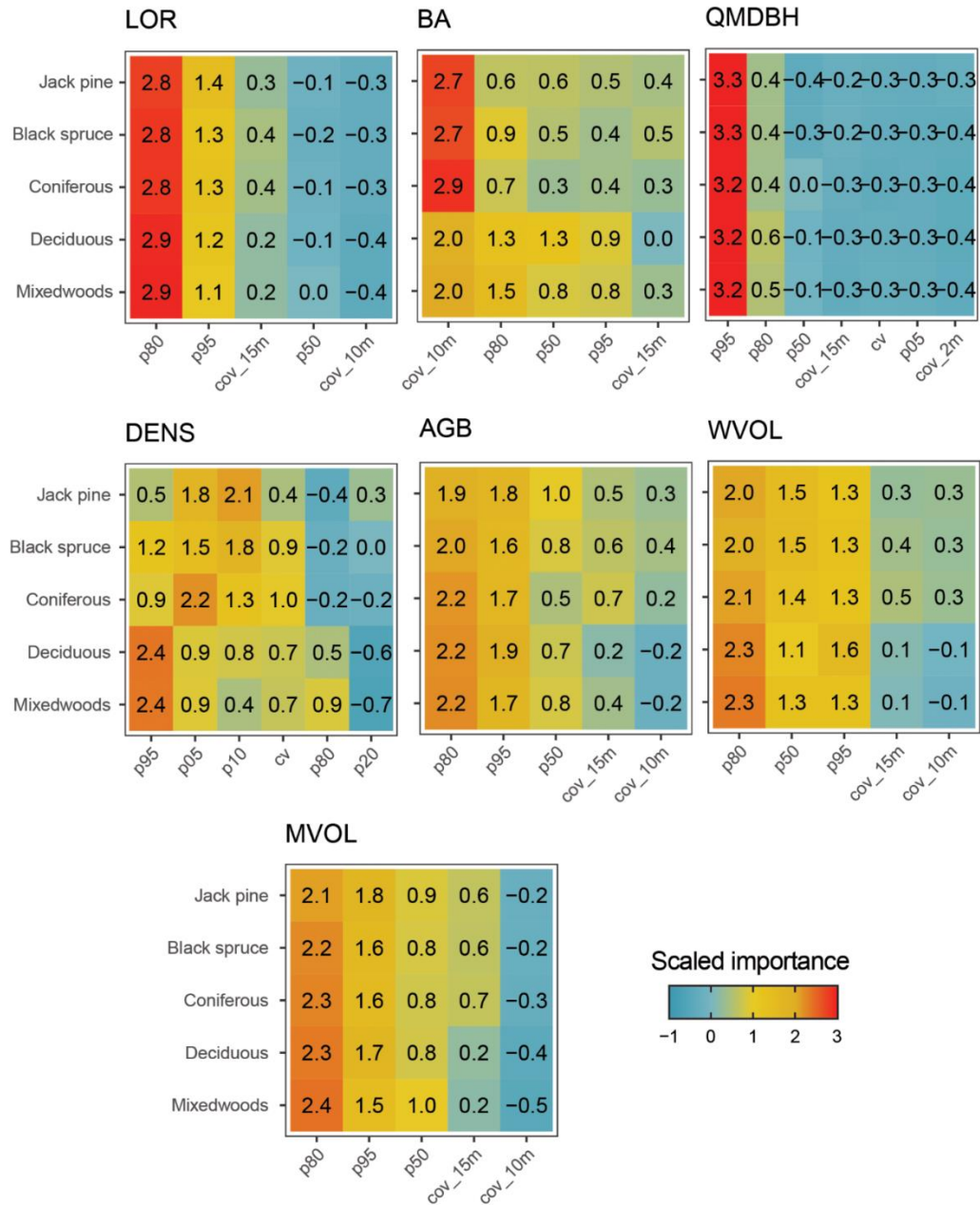
- White, J.C., Wulder, M.A., Hermosilla, T., Coops, N.C., Hobart, G.W., 2017b. A nationwide annual characterization of 25 years of forest disturbance and recovery for Canada using Landsat time series. *Remote Sens. Environ.* 194, 303–321. <https://doi.org/10.1016/j.rse.2017.03.035>
- White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., Guindon, L., 2014. Pixel-based image compositing for large-area dense time series applications and science. *Can. J. Remote Sens.* 40, 192–212. <https://doi.org/10.1080/07038992.2014.945827>
- 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. *For. Chron.* 89, 722–723. <https://doi.org/10.5558/tfc2013-132>
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free Access to Landsat Imagery. *Science* (80-.). 320, 1011. <https://doi.org/10.1126/SCIENCE.320.5879.1011A>
- Woods, M.E., Lim, K., Treitz, P., 2008. Predicting forest stand variables from LiDAR data in the Great Lakes - St. Lawrence forest of Ontario. *For. Chron.* 84, 827–839. <https://doi.org/10.5558/tfc84827-6>
- Woods, M.E., 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, 512–528. <https://doi.org/10.5558/tfc2011-050>
- Wotton, B.M., Nock, C.A., Flannigan, M.D., 2010. Forest fire occurrence and climate change in Canada. *Int. J. Wildl. Fire* 19, 253–271. <https://doi.org/10.1071/WF09002>
- Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., White, J.C., 2008a. The role of LiDAR in sustainable forest management. *For. Chron.* 84, 807–826. <https://doi.org/10.5558/tfc84807-6>
- Wulder, M.A., Campbell, C., White, J.C., Flannigan, M., Campbell, I.D., 2007. National circumstances in the international circumboreal community. *For. Chron.* 83, 539–556. <https://doi.org/10.5558/tfc83539-4>
- Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. <https://doi.org/10.1080/01431161.2018.1452075> 39, 4254–4284. <https://doi.org/10.1080/01431161.2018.1452075>
- Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J.D., Hostert, P., Hughes, M.J., Huntington, J., Johnson, D.M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T.A., Schaaf, C., Schott, J.R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J.C., Wynne, R.H., Zhu, Z., 2019. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* 225, 127–147. <https://doi.org/10.1016/J.RSE.2019.02.015>
- Wulder, M.A., Masek, J., Cohen, W., Loveland, T., Woodcock, C., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* 122, 2–10. <https://doi.org/10.1016/j.rse.2012.01.010>
- Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Healey, S., Zhu, Z., Scambos, T.A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C.J., Masek, J.G., Hermosilla, T., White, J.C., Belward, A.S., Schaaf, C., Woodcock, C.E., Huntington, J.L., Lymburner, L., Hostert, P., Gao, F., Lyapustin, A., Pekel, J., Strobl, P., Cook, B.D., 2022. Fifty years of Landsat

science and impacts. *Remote Sens. Environ.* 280, 113195.
<https://doi.org/10.1016/J.RSE.2022.113195>

- Wulder, M.A., White, J.C., Alvarez, F., Han, T., Rogan, J., Hawkes, B., 2009. Characterizing boreal forest wildfire with multi-temporal Landsat and LIDAR data. *Remote Sens. Environ.* 113, 1540–1555. <https://doi.org/10.1016/j.rse.2009.03.004>
- Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G., Dechka, J.A., 2008b. Monitoring Canada's forests. Part 1: Completion of the EOSD land cover project. *Can. J. Remote Sens.* 34, 549–562. <https://doi.org/10.5589/m08-066>
- Yu, X., Hyypä, J., Karjalainen, M., Nurminen, K., Karila, K., Vastaranta, M., Kankare, V., Kaartinen, H., Holopainen, M., Honkavaara, E., Kukko, A., Jaakkola, A., Liang, X., Wang, Y., Hyypä, H., Katoh, M., 2015. Comparison of laser and stereo optical, SAR and InSAR point clouds from air- and space-borne sources in the retrieval of forest inventory attributes. *Remote Sens.* 7, 15933–15954. <https://doi.org/10.3390/rs71215809>
- Yu, X., Kukko, A., Kaartinen, H., Wang, Y., Liang, X., Matikainen, L., Hyypä, J., 2020. Comparing features of single and multi-photon lidar in boreal forests. *ISPRS J. Photogramm. Remote Sens.* 168, 268–276. <https://doi.org/10.1016/j.isprsjprs.2020.08.013>
- Zald, H.S.J., Wulder, M.A., White, J.C., Hilker, T., Hermosilla, T., Hobart, G.W., Coops, N.C., 2016. Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sens. Environ.* 176, 188–201. <https://doi.org/10.1016/j.rse.2016.01.015>
- Zhang, J., Lin, X., 2017. Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. *Int. J. Image Data Fusion* 8, 1–31. <https://doi.org/10.1080/19479832.2016.1160960>
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., Yu, S., Tomppo, E., Mcroberts, R.E., Zhang, H., Chen, Q., Baghdadi, N., Thenkabail, P.S., 2016. Examining Spectral Reflectance Saturation in Landsat Imagery and Corresponding Solutions to Improve Forest Aboveground Biomass Estimation. *Remote Sens.* 2016, Vol. 8, Page 469 8, 469. <https://doi.org/10.3390/RS8060469>
- Zhen, Z., Quackenbush, L.J., Zhang, L., 2016. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sens.* 2016, Vol. 8, Page 333 8, 333. <https://doi.org/10.3390/RS8040333>
- Zhu, G., Ju, W., Chen, J.M., Liu, Y., 2014. A Novel Moisture Adjusted Vegetation Index (MAVI) to Reduce Background Reflectance and Topographical Effects on LAI Retrieval. *PLoS One* 9, e102560. <https://doi.org/10.1371/JOURNAL.PONE.0102560>
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94. <https://doi.org/10.1016/J.RSE.2011.10.028>

Appendices

Appendix A



Appendix A. Scaled variable importance (z-score) of the most important variables for each forest attribute and species group. Negative scaled importance indicate that the variable was less important than the average importance among all variables. Conversely, positive scaled importance indicates that the variable was more important than the average importance among all variables. See Table 3 for the description of species group.

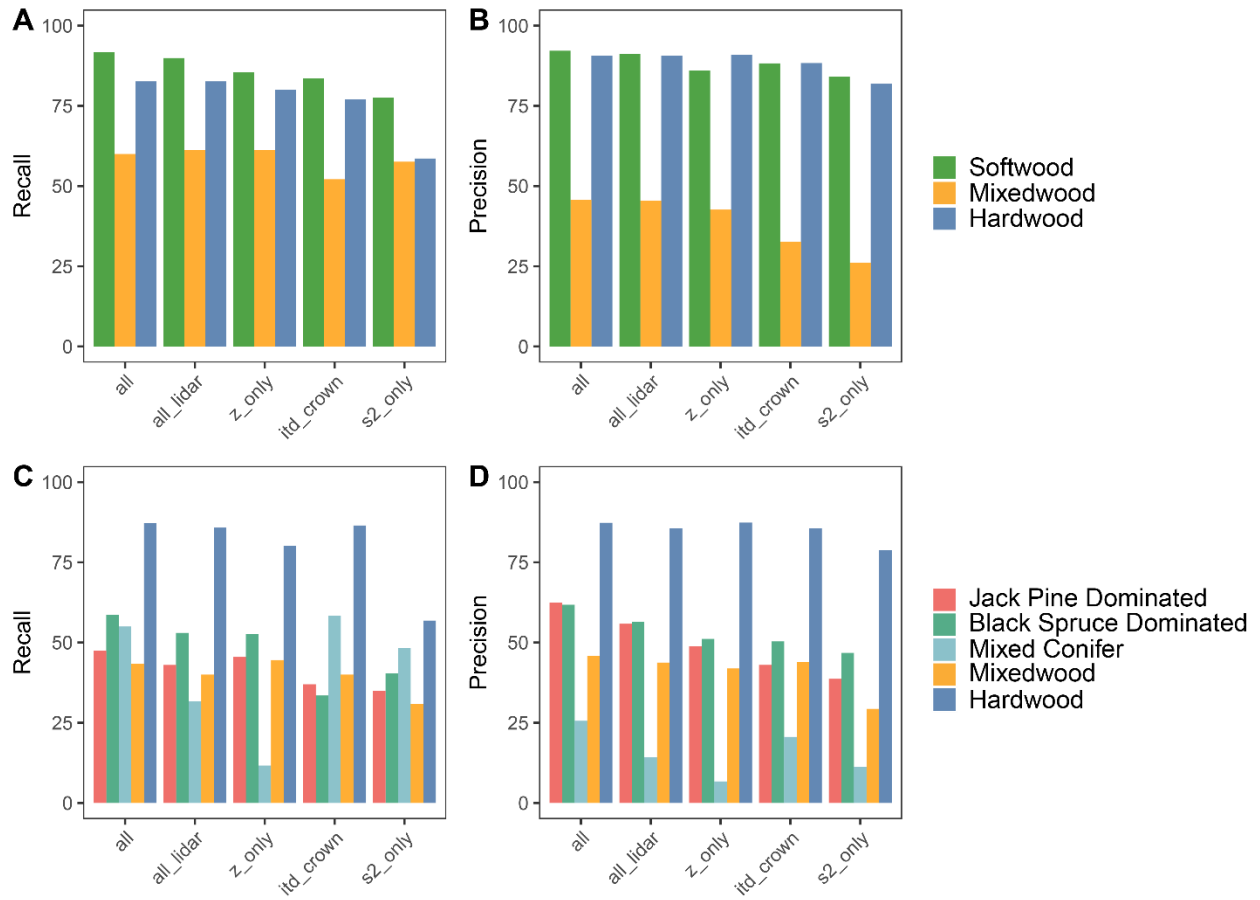
Appendix B

Appendix B. Median recall, precision and F-score of the individual tree detection algorithm by species group and local maxima filter window diameter.

Species group	Window Diameter	Recall	Precision	F-score
Jack Pine Dominated	0.5	0.93	0.26	0.42
Jack Pine Dominated	1	0.9	0.5	0.67
Jack Pine Dominated	1.5	0.88	0.57	0.73
Jack Pine Dominated	2	0.86	0.67	0.8
Jack Pine Dominated	2.5	0.79	0.76	0.85
Jack Pine Dominated	3	0.75	0.89	0.77
Jack Pine Dominated	3.5	0.69	0.93	0.76
Jack Pine Dominated	4	0.6	0.93	0.7
Jack Pine Dominated	4.5	0.6	0.93	0.67
Jack Pine Dominated	5	0.58	0.94	0.68
Black Spruce Dominated	0.5	0.94	0.47	0.63
Black Spruce Dominated	1	0.83	0.76	0.8
Black Spruce Dominated	1.5	0.8	0.88	0.82
Black Spruce Dominated	2	0.74	0.92	0.8
Black Spruce Dominated	2.5	0.68	0.92	0.77
Black Spruce Dominated	3	0.62	0.93	0.71
Black Spruce Dominated	3.5	0.54	0.96	0.68
Black Spruce Dominated	4	0.53	1	0.65
Black Spruce Dominated	4.5	0.45	1	0.57
Black Spruce Dominated	5	0.4	1	0.55
Mixed Conifer	0.5	0.98	0.52	0.67
Mixed Conifer	1	0.9	0.78	0.8
Mixed Conifer	1.5	0.79	0.92	0.82
Mixed Conifer	2	0.76	0.96	0.8
Mixed Conifer	2.5	0.58	0.98	0.69
Mixed Conifer	3	0.49	1	0.62
Mixed Conifer	3.5	0.46	1	0.6
Mixed Conifer	4	0.43	1	0.57
Mixed Conifer	4.5	0.35	1	0.5

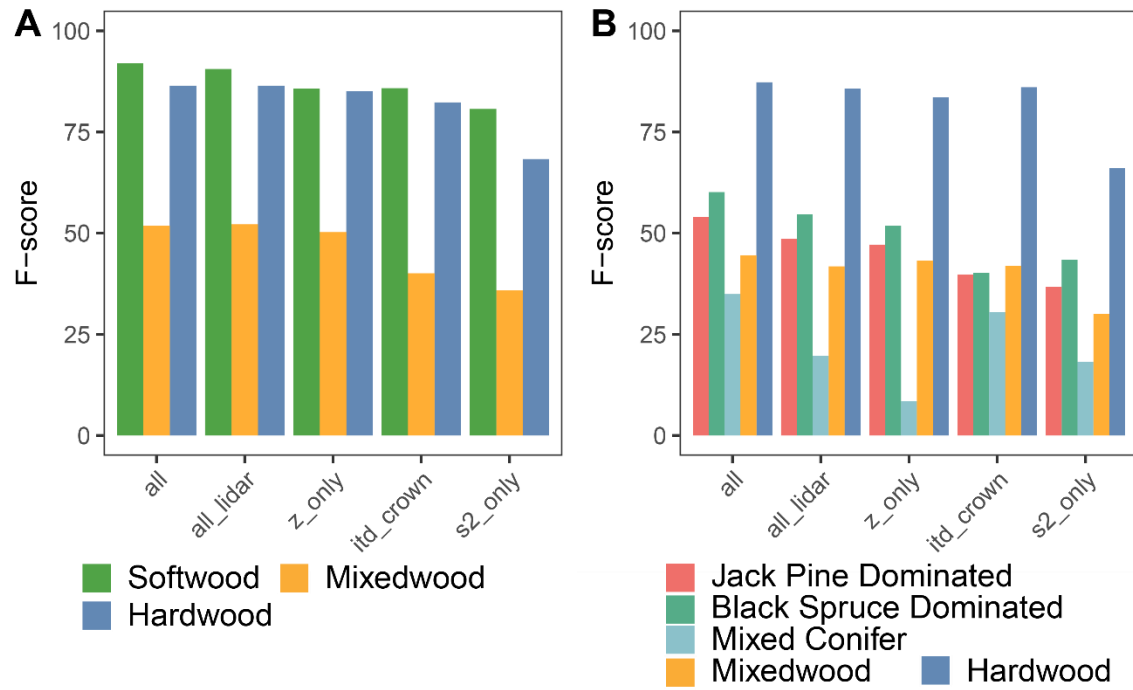
Mixed Conifer	5	0.32	1	0.47
Mixedwood	0.5	0.95	0.25	0.4
Mixedwood	1	0.77	0.41	0.56
Mixedwood	1.5	0.66	0.5	0.64
Mixedwood	2	0.58	0.64	0.72
Mixedwood	2.5	0.48	0.65	0.65
Mixedwood	3	0.39	0.62	0.56
Mixedwood	3.5	0.38	0.6	0.52
Mixedwood	4	0.38	0.67	0.48
Mixedwood	4.5	0.35	0.6	0.44
Mixedwood	5	0.31	0.67	0.43
Hardwood	0.5	0.93	0.11	0.2
Hardwood	1	0.9	0.23	0.36
Hardwood	1.5	0.86	0.36	0.51
Hardwood	2	0.85	0.48	0.61
Hardwood	2.5	0.79	0.64	0.72
Hardwood	3	0.74	0.8	0.75
Hardwood	3.5	0.67	0.79	0.73
Hardwood	4	0.64	0.82	0.67
Hardwood	4.5	0.55	0.86	0.68
Hardwood	5	0.5	0.9	0.63
All species groups	0.5	0.93	0.26	0.4
All species groups	1	0.86	0.41	0.55
All species groups	1.5	0.83	0.54	0.65
All species groups	2	0.81	0.64	0.72
All species groups	2.5	0.74	0.75	0.72
All species groups	3	0.68	0.86	0.72
All species groups	3.5	0.63	0.87	0.7
All species groups	4	0.57	0.88	0.65
All species groups	4.5	0.53	0.92	0.64
All species groups	5	0.46	0.93	0.59

Appendix C



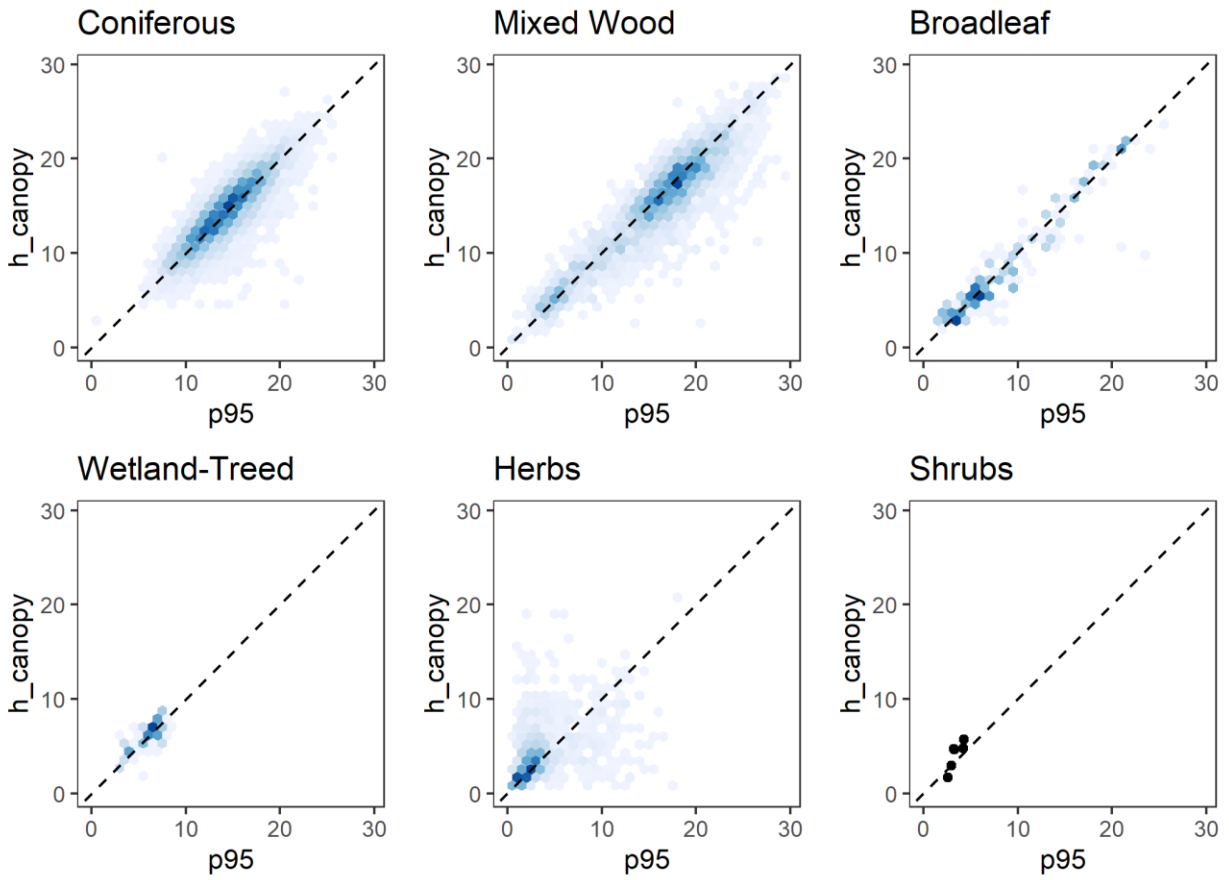
Appendix C. Recall and precision of the 3-group (A; B) and 5-group (C; D) species classification models including different sets of predictor variables.

Appendix D



Appendix D. F-score of the 3-group (A) and 5-group (B) species classification models including different sets of predictor variables.

Appendix E



Appendix E. Scatterplots of p95 (SPL100) and h_canopy (ICESat-2) estimates of canopy height at the Romeo Malette Forest in reference observations where no change was detected after 2018 grouped by land cover class. All scatterplots except “Shrubs” are represented using 1-m hexagonal bins colored (shaded from light to dark blue) based on the number of observations contained within them.