The Multisource Vegetation Inventory (MVI): A Satellite-based Forest Inventory for the Northwest Territories Taiga Plains

Supplementary Materials (S1): Validation of the Forest Attribute Rasters

DOI 10.3390/rs14051108

Guillermo Castilla [1,*], Ron J. Hall [1], Rob Skakun[1], Michelle Filiatrault[1], André Beaudoin[2], Michael Gartrell[1], Lisa Smith[3], Kathleen Groenewegen[3], Chris Hopkinson[4] and Jurjen van der Sluijs[5]

[1] Northern Forestry Centre; guillermo.castilla@canada.ca, ron.hall@canada.ca; rob.skakun@canada.ca, michelle.filiatrault@canada.ca, michael.gartrell@canada.ca
[2] Laurentian Forestry Centre; andre.beaudoin@canada.ca
[3] Forest Management Division, GNWT; lisa_smith@gov.nt.ca, kathleen_groenewegen@gov.nt.ca
[4] Department of Geography, University of Lethbridge, Canada; c.hopkinson@uleth.ca
[5] Northwest Territories Centre for Geomatics, GNWT; Jurjen_vanderSluijs@gov.nt.ca

Summary

Here we assess the accuracy of the Satellite Vegetation Inventory (SVI), a set of 7 forest attribute rasters at a 30-m resolution (forest type, stand height, crown closure, stand volume, total volume, aboveground biomass, and stand age) covering roughly 40 million hectares in the Northwest Territories. Several independent analyses were undertaken, each using a different reference dataset: National Forest Inventory (NFI) ground plots, independent airborne and spaceborne LiDAR (Light Detection and Ranging) datasets, helicopter close-up geotagged photos, and several existing conventional Forest Vegetation Inventories (FVI). The results from the NFI analysis indicate that the SVI estimates were in general more accurate in Phase 1 (mostly corresponding to the Mid-Boreal and High Boreal ecoregions of the Taiga Plains Ecozone, see Figure 1 in main article) than in Phase 2 (Low Subarctic Ecoregion). The attribute with the best pixel-wise relative root mean square error (%RMSE) corresponded to stand height (26% in both phases 1 and 2), and the worst corresponded to volume and aboveground biomass (AGB) estimates in Phase 2, where the %RMSE exceeded 100%. Stand age, which was estimated as a function of stand height, performed surprisingly well in Phase 1 (%RMSE of 35%, or 31 years), but not in Phase 2, where the underestimation bias could be as much as 90 years, owing to the lack of tree ring data in Phase 2 where trees can be very old and still have a low height. Analysis of unused spaceborne LiDAR plots showed that the k-NN imputation did not introduce bias to the forest attribute estimates, and the mean absolute difference for stand height, crown closure, stand volume, total volume, and AGB was 19%, 11%, 36%, 30%, and 29% of their mean value, respectively. The extensive coverage of independent airborne LiDAR data provides good insight into the spatial factors affecting the quality of SVI estimates. Performance was poorer in steep terrain, heterogeneous landcover types, and less common forest types (i.e., broadleaf). Comparison with FVI polygons corroborated that the mean absolute difference between the average SVI pixel and its encompassing FVI polygon was small (2 to 3 m in stand height, and 13 to 16 percentage points for crown closure). Per polygon analyses revealed that the SVI rasters provide better estimates for large homogeneous stands, but may do less well in complex multi-story stands. The error estimates presented here are conservative because of other factors contributing to the observed differences between the SVI estimates and the reference data used in the assessments (e.g., temporal and spatial mismatches, which will tend to inflate the differences). Despite the highlighted limitations, this validation corroborates the usefulness of our approach to scale up forest inventory information from ground plots to wall-to-wall estimates across a large region using geospatial features including synergistic optical, LiDAR, and L-band synthetic aperture radar (SAR) data.

www.mdpi.com/article/10.3390/rs14051108/s1

Contents

www.mdpi.com/article/10.3390/rs14051108/s1

S1. Introduction

The goal of the Multisource Vegetation Inventory (MVI) project, a long-standing collaboration between the Canadian Forest Service and the Government of NWT (GNWT), is to test, apply, and evaluate methods for deriving wall-to-wall information on poorly inventoried northern boreal forests in the Northwest Territories (NWT) using limited field data and a variety of remote sensing data, including optical (Landsat), light detection and ranging (LiDAR)[1], and Synthetic Aperture Radar (SAR) data. The study area, which is mostly located in the Taiga Plains Ecozone, is divided into two parts: Phase 1 (151,700 km^2), corresponding to the southern portion of the MVI project; and Phase 2 (244,000 km^2), which covers the northern portion. A key deliverable of the MVI project is a set of seven 30-m rasters called the Satellite Vegetation Inventory or SVI for the approximate[2] reference years 2007 (Phase 1) and 2010 (Phase 2). Each raster represents one of the following forest attributes: forest type (conifer, broadleaf, mixed), stand height (**Ht**)[3], crown closure (**CC**)[4], stand volume (**Vs**)[5], total volume (**Vt**)[6], aboveground tree biomass (**AGB**)[7], and stand age[8]. Except for the first and last, each of these variables is estimated using four main steps:

i. **ALS modelling**: The variable is first modelled on the basis of a limited set of 38 forest inventory (FI) ground plots near Fort Simpson where, in addition to ground measurements, there were airborne LiDAR data (also known as airborne laser scanning or ALS data) available; both sets of data were from the summer of 2007. The model uses as predictors point cloud metrics derived from the ALS data.

ii. **GLAS modelling**: The predicted values from the model in (i) are used as "observed" values in a second modelling exercise that involves a small set of 43 surrogate FI (forest inventory) plots consisting of sites near Fort Simpson with coincident airborne and spaceborne LiDAR. The latter came from the Geoscience Laser Altimeter (GLAS), a satellite waveform LiDAR instrument operational from 2003 to 2009 that emitted pulses with an average footprint diameter of 70 m spaced at 170 m. The predictors in this case are metrics derived from the GLAS data, either directly (for stand height and

[1] LiDAR data in this project include both airborne laser scanning (ALS) data acquired from a fixed-wing aircraft and spaceborne data from the Geoscience Laser Altimeter System (GLAS), a LiDAR instrument onboard ICESat.

[2] The actual time period to which these 5 rasters refer varies. For example, in Phase 1, the data employed to estimate the forest attributes in the surrogate plots are from 2003–2004, and the raster data used as input to select the 4 most similar surrogate plots to each 30-m cell to be imputed are from 2007–2009. However, the reference year can ultimately be pinned down to the year of the ground data at the base of the modelling chain, which is 2007. The assumption is that no significant disturbances occurred in the reference surrogate plots between 2003 and 2009, thus the models fitted with 2007 data implicitly account for normal growth and mortality during 2003–2007.

[3] **Stand height** is the average height (m) of dominant and codominant trees, i.e., with height ≥ Lorey's height. Lorey's height is the weighted (by cross-section) mean height of all trees with dbh ≥ 5 cm and taller than 1.3 m.

[4] **Crown closure** (%) is the percentage of the ground that is covered by the vertical projection of tree crowns.

[5] **Stand volume** (m^3/ha) is the total volume (see below) for trees with height ≥ Lorey's height.

[6] **Total volume** (m^3/ha) is the sum of the volume inside bark of the boles of trees with dbh ≥ 5 cm, from point of germination to tree top. It is estimated using species-specific taper equations ($v = f(h, dbh)$) used by the National Forest Inventory.

[7] **AGB** (t/ha) is the aboveground biomass of live trees with dbh ≥ 5 cm and ≥ 1.3 m tall. It is estimated using the same procedure as for total volume but with a set of species-specific equations that can be found in Lambert et al. [80] and Ung et al. [31].

[8] **Stand age** is the mean age that dominant and codominant trees in the stand had in 1990, which is the start year used in carbon flux modelling. Stand age was derived from models relating the mean age of cored trees in MVI ground plots with stand height, where the models were directly applied to the stand height raster.

crown closure), or indirectly from variables modelled as a function of the GLAS data (for the other forest attributes[9]).

iii. **Creation of the surrogate FI plot (reference) dataset**: The models developed in ii) are applied to a much larger set of GLAS footprints to provide an estimate for the forest attributes at each footprint. This larger set is itself a subset of all available GLAS footprints in the region, namely those that passed a number of filters based on factors such as acquisition date, terrain slope, and proximity to water or other footprints. The centroid of a valid GLAS footprint with its modelled forest attributes is used to extract (using bilinear interpolation) and assign the corresponding values of the geospatial features used in the k-NN mapping. These cells become the reference dataset (i.e., surrogate 30-m FI plots) used in the final step.

iv. **k-NN mapping**: Finally, each forest attribute is estimated over all forested area[10] using a k-nearest neighbour (k-NN) algorithm. This algorithm finds the k (in our case, k = 4) most similar reference cells to the target cell (i.e., its nearest neighbours in a multidimensional feature space) among those cells corresponding to 30-m surrogate FI plots. Similarity is measured as the Euclidean distance in the feature space defined by several remote sensing, topographical, and climatic variables that, unlike GLAS, are available wall-to-wall across the study area. The estimated value is the mean of those four surrogate plots.

Before they can be used, each raster of the SVI needs to be validated, that is, it needs to have its accuracy assessed. This report details the materials, methods, and results of the validation of the SVI for each phase of the MVI project.

S2. Materials

The following independent data were used for the assessment (Figure S1):

1. **National Forest Inventory (NFI) ground plots:** There are 33 NFI ground plots ([80][11]) that are treed (i.e., with ground-measured crown closure equal to or greater than 10%) in Phase 1 and 47 in Phase 2[12]. In addition, there are 19 NFI ground plots with a wetland landscape position in Phase 1 and 5 in Phase 2 (note that AGB was also computed for the Earth Observation for Sustainable Development of Forests (EOSD) wetland-treed class). The NFI ground data were used to derive an estimate of the pixel-wise bias or mean error (ME = SVI – NFI), mean absolute error (MAE), and root mean square error (RMSE) of each of the 5 SVI rasters, plus their respective 95% confidence intervals (CI95). The latter can be derived because the NFI, unlike other reference data used in other analyses, is a systematic-random survey that allows for design-based inference.

[9] Stand volume, total volume, AGB, and stand age are indirectly estimated as sole functions of height. Stand age is not imputed; instead, the model is directly applied to each forest pixel on the basis of its imputed height.

[10] Whether a raster cell is considered forested depends on its label on the forest type raster, which in reality is the landcover map described in section 3.2 of the main manuscript.

[11] For citations numbered 79 or less, please see the reference list in main manuscript.

[12] There were inconsistencies between the NFI forest type assignment and the EOSD. For phases 1 and 2 combined, 18 conifer plots were non-forested according to EOSD, 4 wetland plots were labelled as wetland shrub by EOSD, and the remaining plots were labelled as forest by EOSD although not necessarily the same forest type. We used the stand type determined from the NFI tree data (i.e., weighted by basal area).

2. **GLAS footprints not used in the k-NN imputation:** There are just under 6,000 valid GLAS footprints (i.e., GLAS footprints that passed the date and topography filters) in Phase 1 and almost 4000 in Phase 2, which were not used in the k-NN imputation because they were too close (i.e., within the range of spatial autocorrelation) to the footprints selected as the reference set for the k-NN imputation. We capitalized on these unused footprints to assess the part of the total pixel-wise error that can be attributed to the imputation.

3. **Boreal transect (BT) ALS point cloud metrics in 25-m cells:** The transect was acquired in summer 2010. It is approximately 750 m wide and runs along 1,800 km in Phase 1 and 950 km in Phase 2 ([40], [41]). We used these to estimate an "observed" value of some forest attributes in those cells using the Fort Simpson ALS models (S1i). Point cloud metrics for the 25-m cells used as predictors in those models were computed using LAStools (http://rapidlasso.com/LAStools).

4. **Boreal transect (BT) ALS-derived forest attributes resampled to 150-m cells:** To assess how errors change as a function of cell size, we resampled the "observed" BT values from 25 m to 150 m by simple averaging.

5. **GNWT ALS point cloud metrics in 25-m cells:** There were 6 ALS datasets acquired in 2010 near the Mackenzie Valley highway extending between the southern and northern boundaries of Phase 2:
   - Willow Lake – This dataset is for an area of approximately 36 km^2 situated in the southern portion of Phase 2 between Fort Simpson and Wrigley.
   - Little Smith – This dataset is for an area of approximately 50 km^2 located midway between Wrigley and Norman Wells.
   - Big Smith – This dataset is for an area of approximately 18 km^2 located just north of Little Smith.
   - Great Bear – This dataset is for an area of approximately 25 km^2 located near the town of Tulita.
   - Thunder River – This dataset is for an area of approximately 80 km^2 located near the northernmost boundary of Phase 2.
   - Mackenzie Valley Hwy – This dataset is for a 400 km long transect roughly 1 km wide running along the Mackenzie Valley highway between Wrigley and north of Norman Wells. It will be used to identify potential differences attributable to latitudinal gradient.

6. **Jean Marie, Axe Point, and Behchoko Forest Vegetation Inventories (FVI):** These FVIs cover 4,200 km^2, 5,900 km^2, and 4,200 km^2, respectively. The first two are located in Phase 1, and the third in Phase 2. These are modern inventories done with softcopy interpretation, so we wanted to assess how they compare with our forest attribute rasters.

7. **Ground-truthed landcover locations:** Forest and wetland reference data collected by aerial survey from Ducks Unlimited Canada (DUC) were used as reference data (among other high-resolution interpretations and oblique photography) for the accuracy assessment of the Phase 2 landcover map (Appendix SA2). These were interpreters' calls from a helicopter; for forested locations they included percent species composition and forest cover density.

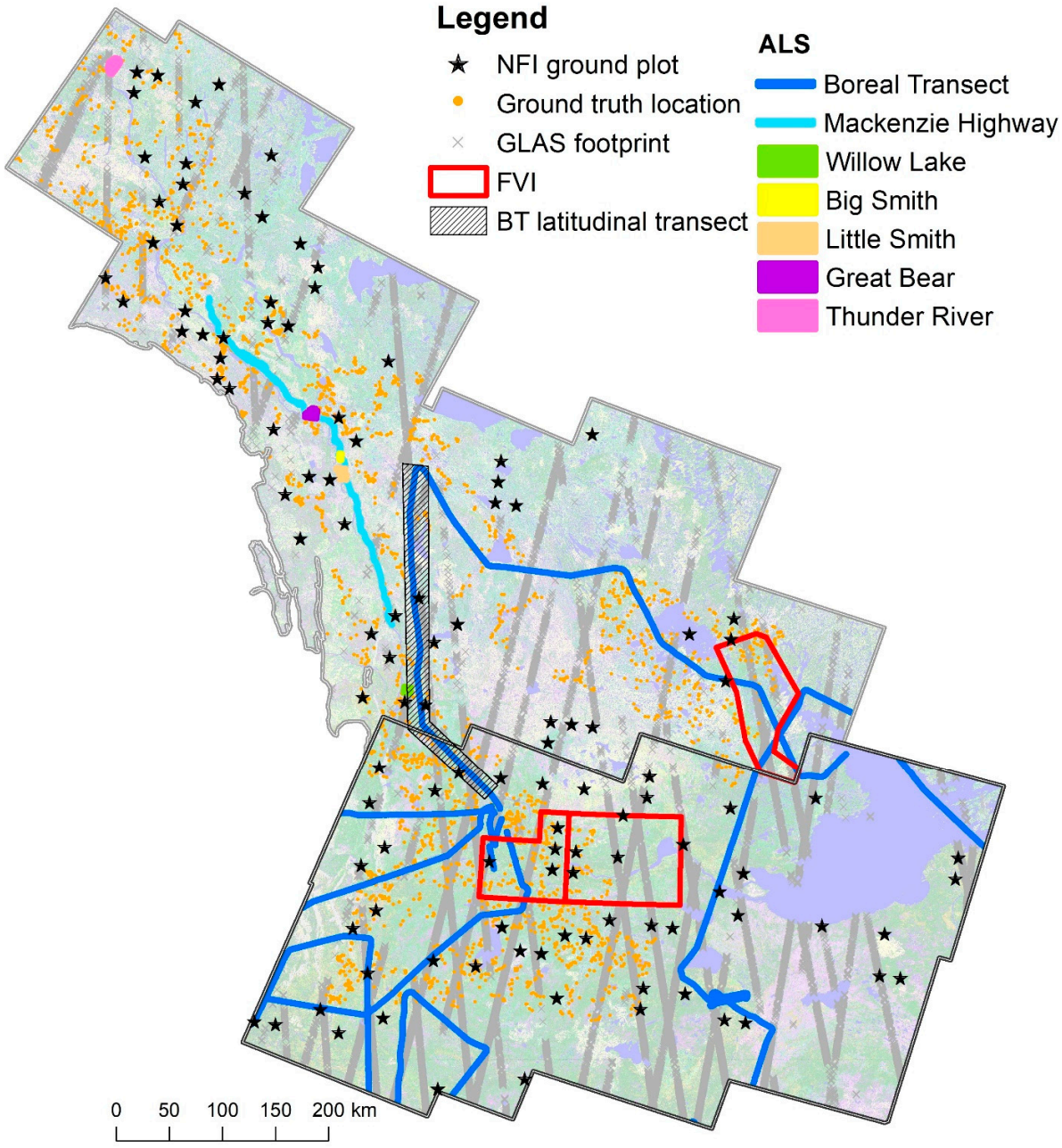


Figure S1. Location of all independent datasets used for the Satellite Vegetation Inventory validation within Phase 1 (darker grey boundary) and Phase 2 (lighter grey) of the Multisource Vegetation Inventory project area. ALS, airborne laser scanning datasets; BT, boreal transect; FVI, Forest Vegetation Inventory; GLAS, Geoscience Laser Altimeter; NFI, National Forest Inventory.

S3. Methods

S3.0. Assumptions and caveats

Assuming the true value of a given forest attribute in a given cell could be known for the reference year, any difference between that value and the SVI value could be explained by four main sources of error:

a) error due to k-NN imputation (i.e., the observed vs. predicted difference assuming the forest attributes from the surrogate FI plots used for imputation were themselves error-free);
b) error in the GLAS estimate of the attribute (i.e., residual error in the GLAS models assuming the "observed value" used to fit the model was error free);
c) error in the models used to produce a surrogate "observed value" for fitting the GLAS models (i.e., residual error in the ALS models, assuming the field measurements were error free), and other modelling errors (e.g., in the models relating AGB or volume to height); and
d) other errors (e.g., spatial and temporal mismatches, errors in ground data, sampling error).

In reality these sources of error are compounded and can hardly be separated, and the true value cannot be known, but they provide a framework to interpret the validation analyses, which require the following assumptions:

I. For the analyses involving NFI or FVI data, we assume that errors of type (d) are negligible (i.e., that the spatial mismatches in the footprint of the data, or the time difference between the NFI ground measurement and the reference year of the forest attribute rasters) have no impact; and that ground (or photo-interpreted, for FVI) measurements were performed and recorded with perfection, in similar forest stands and to the same specifications as the forest plots used for modelling. Therefore, we assess errors (a), (b), and (c) combined.
II. For the analyses involving airborne LiDAR data (BT or GNWT ALS), we assume that errors of type (c) and (d) are negligible (i.e., in addition to (d), that the RMSE of the ALS models is insignificant; and that for a given cell, different LiDAR datasets will yield the exact same point cloud metrics as the LiDAR instrument used for ALS modelling would have yielded if flown with the same specs). Therefore, we assess errors (a) and (b) combined.
III. For the analyses involving unused GLAS footprints, we assume that errors (b), (c), and (d) are negligible, hence we assess the error (a) due to the k-NN imputation.

We also note that there was a power issue during the acquisition of the ALS data used in the models developed with the 38 ground plots near Fort Simpson [8]. Hence the assumption regarding small residuals in the ALS models (b) is only reasonable for stand height, which shows the highest correlation with the ALS 95th percentile (Adj $R^2 = 0.89$, RMSE = 1.4 m). For crown closure, the residuals in the ALS model may be too high to consider the ALS-predicted value a reasonable "ground truth" from which to assess the k-NN estimates, because of low energy returns from the ground. Notwithstanding this limitation, k-NN imputed crown closure is still assessed with ancillary ALS data, because it may provide information about the spatial distribution of errors.

S3.1. NFI ground plots

Fifty-two NFI ground plots (400-m^2 circular plots where data were collected for all trees with diameter at breast height (dbh) ≥ 9 cm, plus a concentric 50-m^2 circular subplot where data were also collected for all small trees < 9 cm dbh) measured between 2001 and 2004 were used for this assessment in Phase 1, and 52 measured between 2002 and 2006 were used in Phase 2 [80]. The NFI plot centre coordinates were used to

identify the collocated pixel in the SVI raster from which the estimated value for the NFI plot was extracted using bilinear interpolation. Out of the 104 NFI plots, 65 were conifer, 6 were broadleaf, 9 were mixedwood, and 24 were assigned a landscape position of "wetland," which corresponds to EOSD class "wetland treed." Note that only AGB estimates were computed for the wetland treed class; therefore, those 24 plots (19 in Phase 1 and 5 in Phase 2) were not used in the assessment of the other attributes. The reference value in each plot for stand height, stand volume, total volume, and AGB was computed on the basis of NFI individual large and small tree data ≥ 5 cm dbh and ≥ 1.3 m in height, in the same manner as for the MVI field plots that were used for ALS modelling. For this, we assumed that the proportion and characteristics of trees between 5 and 9 cm dbh (which were only measured in the small tree subplot) were the same in the rest of the plot; therefore, we applied an expansion factor to the small tree estimates to account for the smaller size of the small tree subplot.

For stand height, stand volume, total volume, and AGB, the signed (SVI – NFI) and absolute differences (|SVI – NFI|) were computed between observed (NFI) and predicted (SVI) values, where the latter were obtained by bilinear interpolation of the 4 closest SVI pixels to each NFI plot centre. Assuming that the impact of temporal (difference between plot measurement year and SVI reference year) and spatial (positional errors, different spatial support – 22.6-m diameter NFI plot vs. 30-m SVI pixel) mismatches is negligible, these differences can be equated to the pixel-wise signed and absolute pixel-wise estimation error in the SVI rasters. The sample mean and its CI95 for the pixel–wise Bias (B) (i.e., ME) and MAE were calculated for each forest attribute. Basic descriptive statistics and linear model statistics/coefficients were also computed for each NFI/SVI attribute pair, including the p-value for the Breusch–Pagan heteroscedasticity test.

For crown closure, the SVI raster was classified into density classes as defined by NFI ocular calls of percent tree cover: sparse (10%–25%), open (26%–60%), and dense (61%–100%). A confusion matrix was produced and accuracy estimates were derived following [43].

The stand age raster was derived from a separate process wherein power functions were employed to estimate stand age as a function of stand height (section 3.7 in main paper). In terms of validation, for each non-wetland NFI ground plot in phases 1 and 2, the stand age value of the bilinear interpolation of the 4 closest pixels to the plot centre was compared with the NFI value as for the other attributes.

S3.2. GLAS footprints

The reference data used in this assessment were the forest attribute values derived from the GLAS models for those valid (i.e., meeting requirements regarding date, topography, etc.), forested GLAS footprints that were not used in the k-NN imputation: 4,617 footprints in Phase 1 and 1,621 in Phase 2. The k-NN-predicted value of the forest attributes in each footprint was the bilinearly interpolated value of the 4 closest SVI pixels to the footprint centre. For the "observed" value, GLAS models for crown closure, Lorey's mean height, and stand height were applied to the relevant GLAS metrics (main paper, Table 3), and for stand volume, total volume, and AGB, the obtained height value was applied to the field-based models that estimated those attributes as a function of stand or Lorey's height (main paper, Table 4a). For each forest attribute, a number of descriptive statistics were computed for the SVI value, GLAS value, and signed (SVI – GLAS) and absolute (|SVI – GLAS|) differences. Linear model statistics and coefficients were also computed for each GLAS/SVI attribute pair.

S3.3. Boreal transect ALS Part 1: Comparison of 30-m SVI and 25-m ALS

Since there was good correspondence between the Fort Simpson and the boreal transect (BT) ALS CI95 metric (Appendix SA1, Table SA1.1), and despite power issues with the Fort Simpson ALS acquisition that affected crown closure estimates, we applied the ALS models for stand height and crown closure (main paper, Table 2) to the 25-m BT cells and used the resulting estimates as reference data in this comparison

(note that the BT dataset was provided at 25-m cells). Since volume, AGB, and age estimates are a function of height, only stand height and crown closure were compared.

The BT 25-m dataset was filtered to exclude cells that were non-forested, on a slope > 5°, or with stand height > 35 m (the latter are outliers representing < 0.01% of the BT). The centroids of the remaining 25-m cells were used to point sample (by bilinear interpolation) the 30-m SVI rasters. For each forest attribute, a number of descriptive statistics were computed for the SVI value, BT-derived value, and signed (SVI – BT) and absolute (|SVI – BT|) differences. Linear model statistics/coefficients were also derived for each BT/SVI attribute pair.

S3.4. Boreal transect ALS Part 2: Comparison of 150-m SVI and 150-m ALS

The SVI rasters of stand height and crown closure were resampled to 150 m using the mean of the 30-m cells within (note that 150 is the lowest common multiple to the 25-m BT cells and the 30-m SVI cells). The mean BT-derived value of those 25-m BT filtered cells within each 150-m pixel was also computed (see item 4 in section S2). Only 150-m cells that had at least 75% coverage with valid, forested BT cells were used. Forest type (C: conifer, B: broadleaf, and M: mixedwood) was assigned using the EOSD, where a 150-m cell was considered pure (C or B) if at least 75% of the forested 30-m EOSD pixels within were C or B, otherwise it was set to M (mixedwood). For stand height and crown closure, the signed (SVI – BT) and absolute (|SVI – BT|) differences were computed in each forested 150-m cell within the BT. From these two parameters some descriptive statistics were derived, as well as linear model statistics/coefficients.

S3.5. Government of Northwest Territories ALS: Comparison of 30-m SVI and 25-m ALS

The GNWT ALS datasets were processed with LAStools using 25-m cells. Since there were no forest inventory ground data within the extent of the GNWT ALS datasets, we applied the Fort Simpson ALS models of stand height and crown closure to these cells. Each ALS dataset was analyzed independently in the same manner as the BT, where the datasets were filtered to exclude cells that were non-forested, on a slope > 5°, or with stand height > 35 m. The centroids of the remaining 25-m cells were used to point sample (by bilinear interpolation) the 30-m SVI rasters. For each forest attribute, a number of descriptive statistics were computed for the SVI value, GNWT ALS-derived value, and signed (SVI – GNWT ALS) and absolute (|SVI – GNWT ALS |) differences. Linear model statistics/coefficients were also derived for each GNWT ALS /SVI attribute pair.

We hypothesized that the accuracy of the SVI forest attributes would decrease moving northward because the forest types would become more ecologically dissimilar to those used in model development (i.e., Fort Simpson) and because the GLAS estimates would become less reliable as canopy height decreases moving north. To test this assumption, the Mackenzie Valley Highway ALS (hereafter MackHwy) dataset was used, which is the dataset with the longest latitudinal gradient, extending from roughly 63.2°N to 65.8°N latitude (350 km). A section of the BT was also used, which overlaps in latitudinal extent with the MackHwy ALS by roughly 150 km and extends further south by 100 km. Both datasets were used to assess whether a relationship exists between latitude and stand height, both observed (ALS) and predicted (SVI), as well as the mean unsigned relative and absolute difference between SVI and ALS stand heights. For that, we randomly selected 500 forest cells in each 25-km interval of latitude we defined, and we plotted the mean value of each height metric in one graph per ALS dataset.

S3.6. Forest Vegetation Inventory

The Jean Marie (reference year 2003; 420,176 ha forested), Axe Point (2003; 593,500 ha), and Behchoko (2011; 336,276 ha) FVI, produced by softcopy interpretation of digital aerial imagery, were selected for this analysis. The first 2 are located in Phase 1 and Behchoko is in Phase 2. Since the imagery used to create those inventories (2003 and 2011) is close to the reference year (2007 in Phase 1 and 2010 in Phase 2), in the

analysis we assumed that the impact of the time difference was negligible. FVI attributes HEIGHT[13] and CRWNCLOS[14] were used to compare SVI estimates of stand height and crown closure, respectively. All forested FVI polygons greater than 2 ha, and having at least 50% coverage with forested SVI cells, were used for analysis. The mean SVI value within an FVI polygon was used as the value to compare. The mean (weighted by polygon size) and other descriptive statistics were derived for the signed (SVI – FVI) and absolute (|SVI – FVI|) differences, as well as linear model statistics/coefficients.

To assess the accuracy of the Phase 1 (circa 2007) and Phase 2 (circa 2010) landcover maps regarding forest type (C: conifer, B: broadleaf, M: mixedwood), an error matrix was produced comparing FVI attribute TYPECLAS and the EOSD level 4 equivalent. For this, the forest EOSD classes were summarized by FVI polygon, and we used the same procedure that we used with the BT 150-m cells to assign a forest type. To compare the Behchoko Phase 2 analysis with the separate accuracy assessment of the Phase 2 landcover that was carried out with ground-truthed locations (Appendix SA2), a smaller confusion matrix for only the conifer, broadleaf, and mixedwood classes was prepared on the basis of those locations and presented beside the Behchoko matrix.

S4. Results and Discussion

S4.1. NFI ground plots

The estimated pixel-wise bias for stand height in Phase 1 was –0.7 m (95% confidence interval CI95 = [–2.5, 1.1]), and the mean pixel-wise absolute error MAE was 3.8 m (CI95 = [2.5, 5.0]) (Table S1a). The estimated pixel-wise bias for stand height in Phase 2 was –1.5 m (CI95 = [–2.2, –0.8]), and the mean pixel-wise absolute error was 1.6 m (CI95 = [1.4, 2.6]) (Table S1b). While stand height MAE was lower in Phase 2 than in Phase 1, in relative terms[15], they were similar (26% in Phase 1 and 24% in Phase 2), although the confidence intervals of these estimates were wide. Reducing the uncertainty around the error estimates to a narrower confidence interval would require a larger sample size of NFI ground plots, which unfortunately was not available.

Assessment of the volume and AGB SVI products provides an indication of how well the k-NN estimation and height-based functions fared for these attributes, with the caveat that the reference data were not directly measured but rather were obtained from allometric equations based on height and dbh. It is unclear whether there was under- or over-estimation for total and stand volume in Phase 1 (Bias CI95 [–32, 43] for total volume and [–27, 26] for stand volume, both in m^3/ha). However, in Phase 2, there was a clear overestimation (Bias CI95 = [18, 30] for total volume and [8, 16] for stand volume, both in m^3/ha). In terms of relative absolute errors, CI95 indicate these errors could be as low as 37% (total volume, in Phase 1) or as large as 187% (stand volume, Phase 2). AGB tended to be slightly overestimated in Phase 1, with an estimated bias of 6 t/ha (CI95 = [–6, 19] t/ha) and MAE of 29 t/ha (CI95 = [20, 39] t/ha). In Phase 2 the overestimation bias was greater, with a mean bias of 15 t/ha (CI95 = [11, 20] t/ha), while the MAE was 19 t/ha (CI95 = [16, 22] t/ha). Similar to the volume estimates, the confidence intervals in Phase 2 indicate that the relative absolute error in AGB was somehow larger (MAE CI95 = [90%, 126%]) than in Phase 1 (MAE CI95 = [60%, 117%]; this is probably a consequence of the lack of field data in unproductive areas covered

[13] FVI attribute HEIGHT is defined as the average height of dominant and co-dominant trees of the leading species and is recorded in 1-m increments. Hence, it is not exactly the same as how the SVI defines stand height but is very close.

[14] FVI attribute CRWNCLOS is defined as the proportion of ground area covered by vertically projected crowns of the tree layer within the polygon (hence the same as in SVI) and is recorded in 10% increments.

[15] Relative to the mean stand height in the relevant NFI plots. Whenever we refer to relative errors in this document, they are relative to the mean value of the forest attribute in the reference data, not to the estimated mean value of the error metric.

by small, sparse black spruce so common in Phase 2. Overall, the prediction of forest structure attributes was less accurate in Phase 2 than in Phase 1, which was expected because Phase 2 forests are less ecologically similar than those in Phase 1 used for modelling, in particular, the trees are shorter and occur on more open stands. This is also consistent with Nelson's [60] finding that "GLAS measurements become problematic with respect to height and AGB retrievals in the boreal forest when AGB values fall below 20 t/ha and when GLAS 75th percentile heights fall below 7 m." Both thresholds are close to the mean values of these attributes in Phase 2 NFI plots.

Table S1. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation; CI+, upper confidence limit, CI–, lower confidence limit) for the comparison between National Forest Inventory (NFI) ground plots and collocated Satellite Vegetation Inventory (SVI) pixels for stand height, total volume, stand volume, aboveground biomass (AGB), and stand age in (a) Phase 1 and (b) Phase 2, including the pixel-wise signed (Δ) and absolute (|Δ|) differences and the confidence intervals (CI) of their mean (Bias (B) and mean absolute error (MAE), respectively, in bold). NB. The last 5 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., NFI) linear model (y = a + bx): RMSE, root mean square error; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a; Breusch-Pagan p-value, if < 0.05, then the residuals are heteroscedastic.

a) Phase 1

Parameter	Stand height (m)				Stand volume (m^3/ha)				Total volume (m^3/ha)				AGB (t/ha)				Stand age (years)			
	NFI	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**
n	33	33	33	33	33	33	33	33	33	33	33	33	52	52	52	52	31	31	31	31
Min	5.0	4.9	-14.0	0.2	1.1	11.4	–238.1	1.6	2.6	26.0	–361.1	9.3	0.3	7.5	–182.8	0.7	44.8	59	–76.5	0.3
Max	31.5	22.4		14.0	361.0	198.5		238.1	545.3	278.7		361.1	300.1	166.3		182.8	205.5	139	52.8	76.5
Mean	14.5	13.8	**–0.7**	**3.8**	91.0	90.3	**–0.7**	**51.6**	132.1	137.3	**5.2**	**75.5**	59.3	65.7	**6.4**	**29.2**	88.7	91.3	**2.6**	**23.9**
Median	14.0	14.7	–1.1	2.5	39.6	89.2	6.1	33.3	65.9	139.3	22.1	46.5	32.9	44.0	11.0	16.9	83.3	85	6.3	19.3
SD	6.9	5.7	5.1	3.5	97.5	61.3	73.8	52.0	138.9	82.7	105.7	73.0	67.9	49.1	44.3	33.6	35.4	23.4	30.3	18.8
CV	0.5	0.4	–7.0	0.9	1.1	0.7	–102.5	1.0	1.1	0.6	20.3	1.0	1.1	0.7	6.9	1.2	0.4	0.3	11.6	0.8
CI+			1.1	5.0			25.5	70.1			42.7	101.4			18.7	38.6			13.9	30.9
CI–			–2.5	2.5			–26.9	33.1			–32.3	49.6			–6.0	19.9			–8.7	16.9
RMSE	5.1				72.7				104.2				44.3				30.96			
Adj. R^2	0.45				0.41				0.41				0.57				0.23			
Slope	0.8				1.0				1.1				1.1				0.8			
Intercept	3.1				–2.9				–18.0				–9.7				15.52			
BP p-value	0.07				0.05				0.06				0.01				0.05			

b) Phase 2

Parameter	Stand height (m)				Stand volume (m^3/ha)				Total volume (m^3/ha)				AGB (t/ha)				Stand age (years)			
	NFI	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**	**NFI**	**SVI**	**Δ**	**|Δ|**
n	47	47	47	47	47	47	47	47	47	47	47	47	52	52	52	52	24	24	24	24
Min	4.5	4.9	–10.1	0.0	0.1	11.3	–34.9	1.2	0.1	24.3	–58.7	7.1	0.2	20.7	–32.1	0.1	38	19	–226.2	1.6
Max	19.5	11.6	2.5	10.1	75.4	56.9	43.2	43.2	128.3	94.1	73.5	73.5	80.4	66.0	51.5	51.5	295.2	82	41	226.2
Mean	8.3	6.8	**–1.5**	**2.0**	10.1	22.3	**12.2**	**16.2**	17.9	41.9	**24.0**	**28.9**	17.2	32.6	**15.4**	**18.5**	130.3	69.8	**–60.5**	**68.7**
Median	7.5	6.3	–1.3	1.6	3.5	18.4	14.1	15.6	6.9	36.1	24.5	28.5	12.2	29.0	16.9	17.5	122.7	73	–50.7	50.7
SD	3.0	1.7	2.4	2.0	17.7	11.2	14.2	9.2	28.0	17.1	21.1	13.5	18.7	10.9	15.1	10.9	68.1	11.9	68.6	60.4
CV	0.4	0.2	–1.5	1.0	1.8	0.5	1.2	0.6	1.6	0.4	0.9	0.5	1.1	0.3	1.0	0.6	0.5	0.2	–1.1	0.9
CI+			–0.8	2.6			16.4	18.9			30.2	32.9			19.6	21.6			–31	94.8
CI–			–2.2	1.4			8.1	13.5			17.8	25.0			11.2	15.5			–90.1	42.7
RMSE	2.8				18.6				31.8				21.4				71.03			
Adj. R^2	0.36				0.34				0.42				0.33				0.00			
Slope	1.1				0.9				1.1				1.0				0.25			
Intercept	0.9				–11.1				–27.0				–15.9				113.24			
BP p-value	0.01				0.00				0.00				0.00				0.68			

There were not enough NFI plots to provide reliable statistics by forest stand type, but scatterplots of NFI vs. SVI for different forest attributes were made, distinguishing stand types (Figure S2). Conifer plots tended to be closer to the 1:1 line than broadleaf and mixedwood plots. There was a potential outlier in the upper end of NFI values for volume and AGB estimates (NFI plot 1087286). Upon investigation into ancillary NFI attributes, the only information that set this plot apart from the other plots was that it was experiencing disease at the time of measurement (2003). Although firm conclusions cannot be drawn from this information, this plot may have experienced a partial stand-level replacement between 2003 and 2007 (vintage of other SVI inputs). If this supposition is confirmed by a future ground remeasurement, it would provide justification for removing this plot from the analysis. Other plots exhibiting considerable difference between predicted and observed values were found to be on gently sloping terrain, or near a landcover boundary.

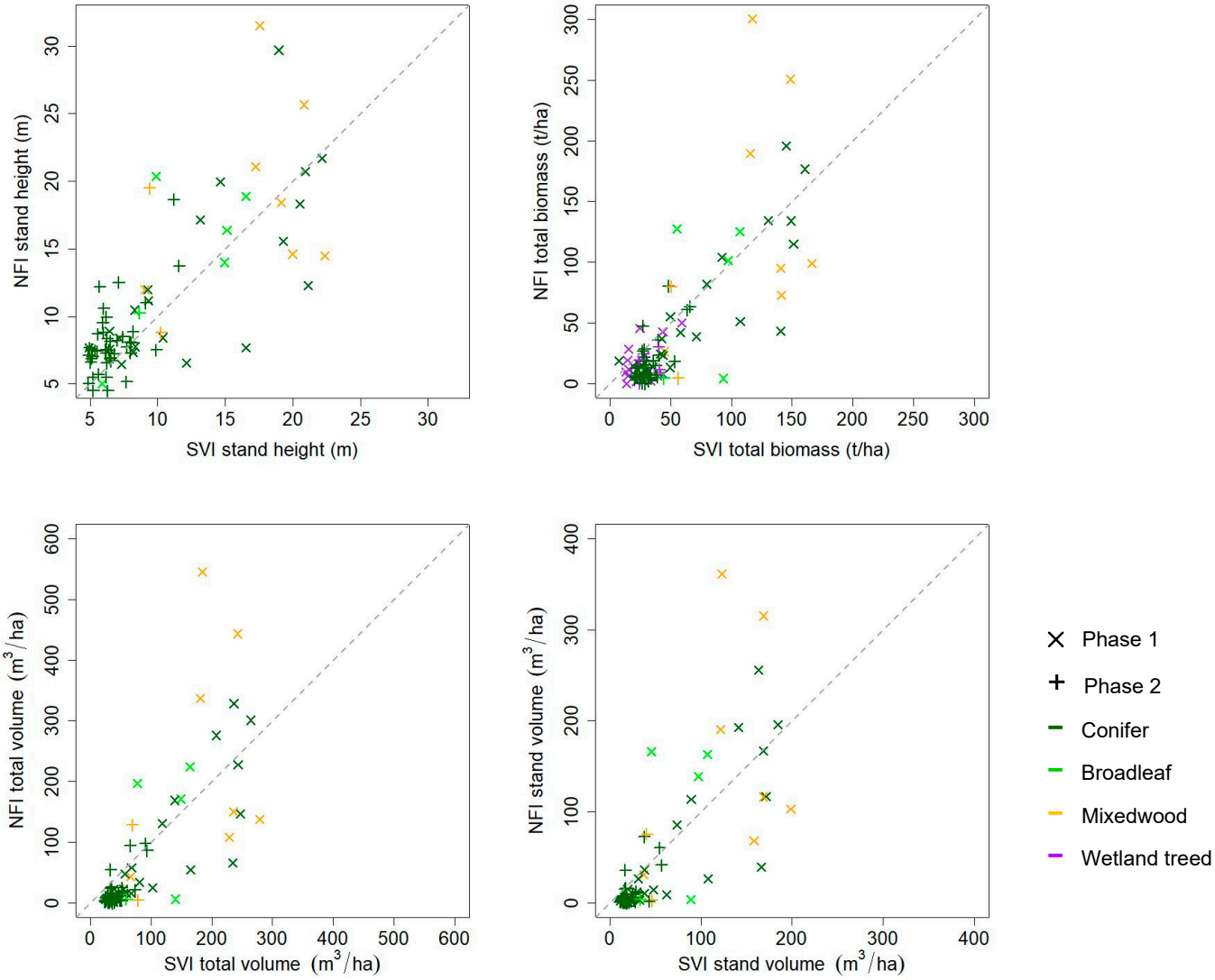


Figure S2. Plots of SVI attributes as a function of NFI attributes by forest stand type. NFI, National Forest Inventory; SVI, Satellite Vegetation Inventory.

In Phase 1, homoscedasticity occurred for all attributes with the exception of AGB ($p = 0.01$) on the basis of the Breusch–Pagan test for heteroscedasticity ($p = 0.05$) (Table S1a). In Phase 2, all attributes generated statistically significant p-values with the exception of stand age ($p = 0.68$), meaning that the error variance of the predicted values was not constant across the range of the stand attribute values (Table S1b). In general, the scatterplots of NFI vs. SVI indicate that there is a typical trend for SVI to underestimate high values of the forest attributes and to overestimate low values, consistent with results reported by Beaudoin

et al. [37]. With nearest neighbour techniques such as k-NN, reliable extrapolations are not possible because predictions are constrained by the input range of the independent or predictor variables [81]. (Figure S3).

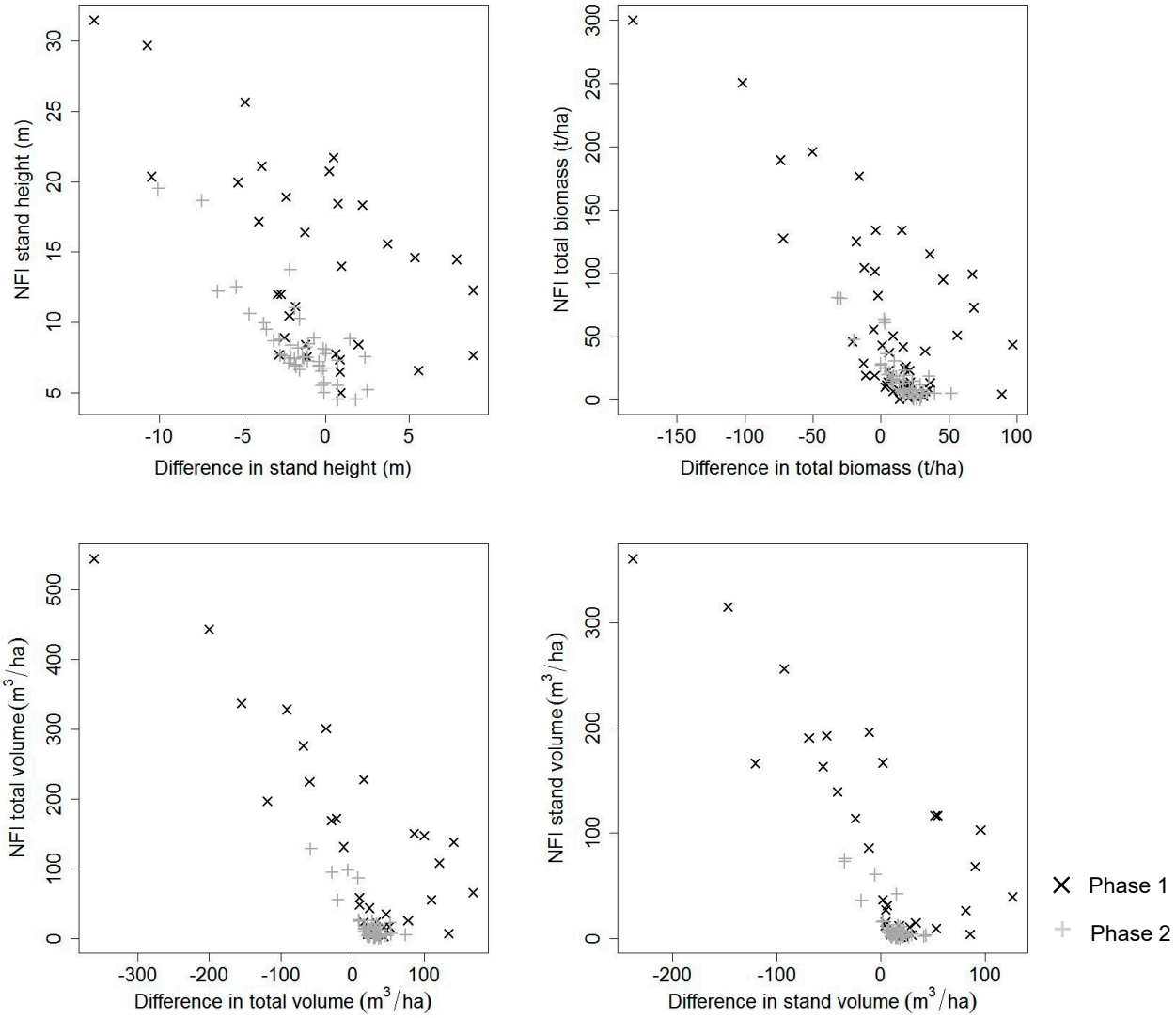


Figure S3. Trend between the difference in SVI–NFI attribute and NFI attribute. NFI, National Forest Inventory; SVI, Satellite Vegetation Inventory.

The crown closure confusion matrix yielded an overall accuracy estimate of 67% in Phase 1 and 51% in Phase 2, with three classes (dense, open, and sparse; Table S2). However, the range in the SVI was very limited, with 97% of plots classified as open, which together with the small sample size may explain the wide 95% confidence interval (the true overall accuracy may have ranged anywhere from 51% to 83% in Phase 1 and from 34% to 69% in Phase 2).

Table S2. Confusion matrix comparing classified Satellite Vegetation Inventory (SVI) crown closure with National Forest Inventory (NFI) density classes in the 40 NFI ground plots in Phase 1 and Phase 2 (first 2

horizontal blocks of the table), and results of normalized error matrix weighted by proportion of SVI density class present (last horizontal block of the table). CI95, 95% confidence interval.

	NFI density					
	Phase 1			Phase 2		
SVI density class	**Dense**	**Open**	**Sparse**	**Dense**	**Open**	**Sparse**
Dense	0	2	0	0	0	0
Open	6	21	3	6	23	15
Sparse	0	0	0	0	0	0
Overall accuracy	0.67			0.51		
Standard error	0.08			0.08		
CI95	(0.5, 0.8)			(0.3, 0.7)		

The stand age product for phases 1 and 2 was compared with NFI estimates in 31 plots in Phase 1 and 24 in Phase 2 (last block in Tables S1a and S1b). For Phase 1 the range of MVI age was within the full range of total age values of the NFI plots. There was a slight overestimation bias of 2.6 years; the MAE was just 23.9 years, and the RMSE was 31.0. For Phase 2 there was a large underestimation bias (61 years) and a MAE of 68 years. These poor results reflect the lack of field data in Phase 2 to develop independent stand age models for the low-height, unproductive forest stands that were more prevalent than in Phase 1.

To sum up, the NFI analysis indicates the SVI was relatively accurate for stand height (pixel-wise relative MAE ~25%) and less so for the other attributes. In general, there was a trend for SVI to underestimate high values of the forest attributes and to overestimate low values, which made relative errors worse in Phase 2, where forests of low height and density were more prevalent. We should note, though, that these error estimates are conservative (i.e., they are probably larger than the actual error) because other factors contributed to the observed differences, namely (i) the time lag between the ground measurements and the reference year of the SVI estimates; (ii) compounded positional errors in NFI and SVI; and (iii) the discrepancy in the spatial support of both sources (30-m pixels vs. 22.6-m diameter plots).

S4.2. GLAS footprints

There were 4,617 forested GLAS footprints in Phase 1 and 1,621 in Phase 2 that met the filter criteria outlined by Mahoney et al. [8] and that were not used in the k-NN imputation. We used the values of the forest attributes estimated directly or indirectly by the GLAS models (main paper, Tables 3 and 4) in these footprints as reference data to assess the SVI estimates. The GLAS-based analysis provides an indication of the amount of error introduced by the k-NN imputation for each phase assuming the reference data were error free.

There was hardly any bias introduced by the imputation, and the mean absolute differences ranged from 10% to 40% of the GLAS-estimated mean value depending on the attribute and phase, demonstrating that the k-NN algorithm worked reasonably and consistently well (Tables S3 to S5; percent values not shown but can be derived). The root mean squared difference (RMSD) of k-NN-predicted (SVI) vs. observed (GLAS) values was high for volume and AGB estimates, indicating the imputation of these attributes performed slightly worse than for the other attributes. When comparing the absolute difference to the observed mean for a given attribute, percent differences were similar in the two phases (difference of < 5 percentage points between phases for all attributes), despite the fact that the predictor variables selected for imputation were different for each phase.

Table S3. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between (i) unused valid Geoscience Laser Altimeter (GLAS) footprints and (ii) Satellite Vegetation Inventory (SVI) pixels corresponding to their bilinearly interpolated centroids, for stand height and crown closure in phases 1 and 2, including pixel-wise signed (Δ) and absolute (|Δ|) differences). NB. The last 4 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., GLAS) linear model (y = a + bx): RMSD, root mean square difference; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a.

	Stand height (m)								**Crown closure (%)**							
	Phase 1				**Phase 2**				**Phase 1**				**Phase 2**			
Parameter	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**
n	4146	4146	4146	4146	1309	1309	1309	1309	4146	4146	4146	4146	1309	1309	1309	1309
Min	3.9	4.1	–13.3	0.0	4.1	4.4	–13.0	0.0	22.0	21.3	–23.4	0.0	22.1	24.0	–41.3	0.0
Max	31.6	31.0	18.2	18.2	23.4	20.1	12.2	13.0	64.0	64.7	23.0	23.4	71.3	58.2	15.1	41.3
Mean	9.4	9.3	**–0.1**	**1.8**	7.1	6.9	**–0.2**	**1.2**	40.5	39.9	**–0.6**	**4.2**	37.0	36.2	**–0.8**	**4.3**
Median	7.9	7.8	0.0	1.1	6.4	6.4	0.1	0.7	40.1	39.0	–0.5	3.2	35.5	35.5	–0.1	3.0
SD	4.6	4.4	2.7	2.0	2.5	1.9	1.8	1.4	8.9	8.0	5.5	3.6	8.2	6.1	6.1	4.5
CV	0.5	0.5	–36.2	1.1	0.4	0.3	–11.0	1.2	0.2	0.2	–8.7	0.9	0.2	0.2	–7.6	1.0
RMSD	2.7				1.9				5.5				6.2			
Adj. R^2	0.67				0.48				0.63				0.44			
Slope	0.8				0.9				0.9				0.9			
Intercept	1.5				0.8				5.6				4.7			

Table S4. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between (i) unused valid Geoscience Laser Altimeter (GLAS) footprints and (ii) Satellite Vegetation Inventory (SVI) pixels corresponding to their bilinearly interpolated centroids, for stand and total volume in phases 1 and 2, including pixel-wise signed (Δ) and absolute (|Δ|) differences. NB. The last 4 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., GLAS) linear model (y = a + bx): RMSD, root mean square difference; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a.

	Stand volume (m^3/ha)								**Total volume (m^3/ha)**							
	Phase 1				**Phase 2**				**Phase 1**				**Phase 2**			
Parameter	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**
n	4146	4146	4146	4146	1309	1309	1309	1309	4146	4146	4146	4146	1309	1309	1309	1309
Min	7.5	8.1	–184.2	0.0	8.2	9.3	–154.4	0.0	18.9	20.1	–236.0	0.0	18.8	20.8	–208.5	0.0
Max	351.0	341.9	275.4	275.4	202.0	157.5	130.1	154.4	467.4	456.0	347.0	347.0	289.1	230.4	179.9	208.5
Mean	44.4	45.7	**1.3**	**16.5**	24.7	23.6	**–1.1**	**8.0**	74.8	76.0	**1.2**	**23.1**	45.3	43.7	**–1.7**	**12.2**
Median	27.5	27.8	0.8	7.2	18.7	19.0	0.8	3.8	52.2	52.0	1.0	11.1	36.7	37.1	1.2	6.2
SD	43.8	44.9	29.3	24.3	18.8	14.4	14.8	12.5	60.8	61.3	39.4	31.9	28.1	21.5	21.7	18.0
CV	1.0	1.0	22.6	1.5	0.8	0.6	–13.8	1.6	0.8	0.8	33.4	1.4	0.6	0.5	–12.9	1.5
RMSD	29.3				14.8				39.4				21.7			
Adj. R^2	0.61				0.40				0.63				0.42			
Slope	0.8				0.8				0.8				0.8			
Intercept	9.5				5.2				15.1				8.3			

Table S5. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between (i) unused valid Geoscience Laser Altimeter (GLAS) footprints and (ii) Satellite Vegetation Inventory (SVI) pixels corresponding to their bilinearly interpolated centroids, for aboveground biomass (AGB) in phases 1 and 2, including pixel-wise signed (Δ) and absolute (|Δ|) differences. NB. The last 4 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., GLAS) linear model (y = a + bx): RMSD, root mean square difference; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a.

	AGB (t/ha)							
	Phase 1				Phase 2			
Parameter	**GLAS**	**SVI**	**Δ**	**|Δ|**	**GLAS**	**SVI**	**Δ**	**|Δ|**
n	4617	4617	4617	4617	1621	1621	1621	1621
Min	1.7	1.9	–126.2	0.0	16.7	18.1	–116.1	0.0
Max	262.6	256.8	181.4	181.4	173.9	142.3	103.4	116.1
Mean	48.5	48.6	**0.1**	**14.6**	33.6	32.7	**–0.9**	**7.3**
Median	36.8	35.8	–0.3	9.0	28.6	29.0	0.7	3.9
SD	36.8	36.4	22.6	17.2	16.8	12.8	12.6	10.2
CV	0.8	0.7	227.4	1.2	0.5	0.4	–14.1	1.4
RMSD	22.6				12.6			
Adj. R^2	0.65				0.45			
Slope	0.8				0.9			
Intercept	8.7				4.7			

The difference in stand height and crown closure for both phases 1 and 2 followed a normal distribution, with the mean difference close to zero and more than half of the footprints showing an absolute difference of < 2 m in height or < 10% in crown closure (Figure S). Plotting the difference between SVI and GLAS values against the GLAS value clearly revealed the aforementioned tendency to underestimate the attribute at high values and overestimate it at low values (Figure S), consistent with previous studies that reported results from the k-NN algorithm [37]. GLAS footprints exhibiting a greater than 100 t/ha difference in AGB were visually inspected, and most were found in heterogeneous cover types. Other factors affecting the observed differences between GLAS and SVI included the following: disturbance (i.e., road, clear-cut, fire), where the disturbance could have occurred in the period between the GLAS acquisition and the date of the datasets used in the k-NN feature space (Figure S, left); and poor GLAS estimation because of sloping terrain or variable landcover within (Figure S6, right).

www.mdpi.com/article/10.3390/rs14051108/s1

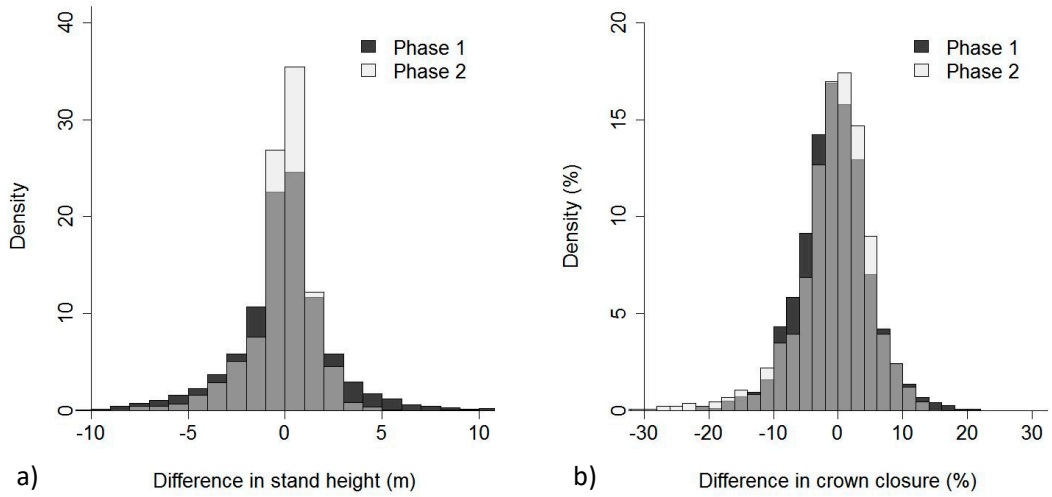


Figure S4. Histograms of the difference between SVI–GLAS estimates of (a) stand height (m) and (b) crown closure (%). GLAS, Geoscience Laser Altimeter; SVI, Satellite Vegetation Inventory.

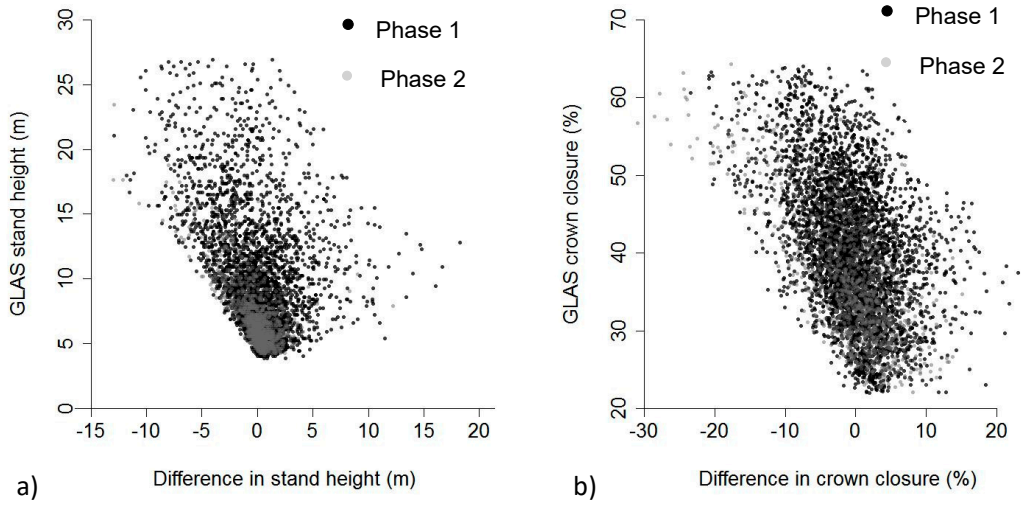


Figure S5. The difference between SVI–GLAS estimates as a function of the GLAS estimate for (a) stand height (m) and (b) crown closure (%). GLAS, Geoscience Laser Altimeter; SVI, Satellite Vegetation Inventory.

www.mdpi.com/article/10.3390/rs14051108/s1

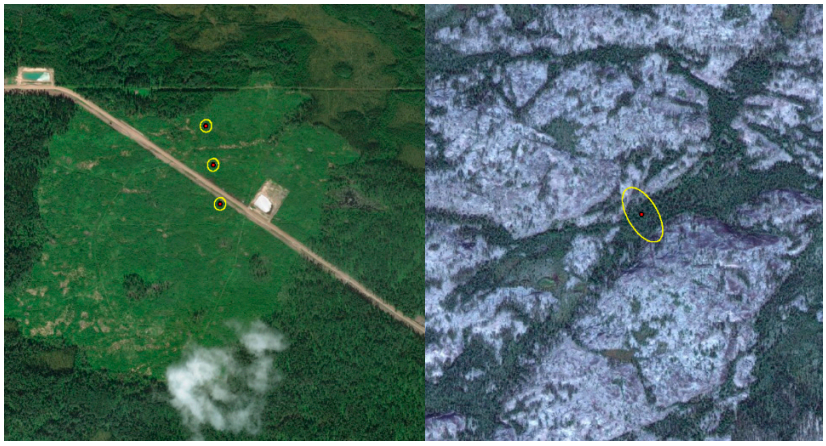


Figure S6. Sample GLAS footprints where the absolute difference in aboveground biomass (AGB) between SVI and GLAS exceeded 100 t/ha, highlighting potential sources of error. Left: GLAS footprints over a clear-cut and road that may have been cleared after the GLAS acquisition. Right: GLAS footprint containing heterogeneous cover type on sloped terrain. GLAS, Geoscience Laser Altimeter; SVI, Satellite Vegetation Inventory.

Stand height and crown closure estimates were analyzed by forest stand type to further understand how the k-NN estimation performed by forest type (Tables S6 and Table, respectively). For both stand height and crown closure the k-NN imputation performed worse in broadleaf and mixedwood stands. These results point at challenges in the estimation over broadleaf and mixedwood stands, which could in part be due to the following factors: the varying degrees of broadleaf composition between conifer-dominated and broadleaf-dominated stands; low representation of these stands in model development; higher volume and AGB values in these stand types, inducing greater underestimation and spectral/backscatter saturation; and classification confusion in the landcover product used for stratification.

Table S6. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of pixel-wise signed (Δ) and absolute (|Δ|) differences between (i) unused valid Geoscience Laser Altimeter (GLAS) footprints and (ii) Satellite Vegetation Inventory (SVI) pixels corresponding to their centroids, for stand height in phases 1 and 2, by forest stand type.

	Phase 1						Phase 2					
	Conifer		Broadleaf		Mixedwood		Conifer		Broadleaf		Mixedwood	
	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|
n	3401	3401	172	172	573	573	1179	1179	85	85	45	45
Min	–13.3	0.0	–9.3	0.0	–10.5	0.0	–13.0	0.0	–10.6	0.0	–6.1	0.0
Max	18.2	18.2	16.1	16.1	16.7	16.7	12.2	13.0	4.7	10.6	4.6	6.1
Mean	**–0.2**	**1.6**	**0.9**	**2.5**	**0.2**	**2.7**	**–0.2**	**1.2**	**0.0**	**1.1**	**0.0**	**1.3**
Median	0.0	1.0	0.4	1.7	0.1	2.0	0.1	0.7	0.1	0.7	0.3	0.8
SD	2.4	1.8	3.5	2.7	3.8	2.6	1.8	1.4	1.9	1.6	2.0	1.5
CV	–14.3	1.1	3.8	1.1	20.7	1.0	–9.9	1.2	–974.4	1.4	–65.1	1.1

Table S7. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of pixel-wise signed (Δ) and absolute (|Δ|) differences between (i) unused valid Geoscience Laser Altimeter (GLAS) footprints and (ii) Satellite Vegetation Inventory (SVI) pixels corresponding to their centroids, for crown closure in phases 1 and 2, by forest stand type.

	Phase 1						Phase 2					
	Conifer		Broadleaf		Mixedwood		Conifer		Broadleaf		Mixedwood	
	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|	Δ	|Δ|
n	3401	3401	172	172	573	573	1179	1179	85	85	45	45
Min	–20.6	0.0	–23.4	0.1	–20.8	0.0	–35.9	0.0	–41.3	0.3	–23.9	0.4
Max	16.3	20.6	20.8	23.4	23.0	23.0	13.7	35.9	12.9	41.3	15.1	23.9
Mean	**–0.9**	**3.8**	**–5.2**	**6.9**	**2.5**	**5.6**	**–0.6**	**4.1**	**–4.6**	**6.5**	**1.5**	**5.1**
Median	–0.7	3.0	–4.9	5.6	2.4	4.3	0.0	2.8	–4.5	5.4	3.0	3.9
SD	4.9	3.2	7.0	5.3	6.8	4.5	5.9	4.3	7.4	5.7	6.9	4.7
CV	–5.3	0.8	–1.4	0.8	2.7	0.8	–9.5	1.0	–1.6	0.9	4.5	0.9

S4.3. Boreal transect ALS Part 1: Comparison of 30-m SVI and 25-m ALS

In contrast to what was found in the NFI analysis, stand height underestimation was nil when the reference data were derived from the boreal transect (BT) ALS (mean difference of –10 cm for Phase 1 and –40 cm for Phase 2, **Error! Not a valid bookmark self-reference.**). The mean absolute difference was also smaller (3 m in Phase 1 and 1.6 m in Phase 2, whereas in NFI it was 4 m in Phase 1 and 2 m in Phase 2; **Error! Not a valid bookmark self-reference.**). These seemingly improved results could stem from the fact that the reference data for this comparison were derived by applying the same model that is at the start of the chain leading to SVI estimates. Although this model was developed with the Fort Simpson ALS dataset, it seems it transferred well to the BT. Indeed, the direct comparison of the 95th percentile metric between the Fort Simpson and BT ALS data in areas of overlap revealed a good agreement between the 2 datasets (Adj. R^2 = 0.93; RMSD = 1.2 m), with a mean difference of 0.1 m, 0.6 m, and 0.5 m in conifer, broadleaf, and mixedwood stands, respectively (Appendix SA1).

By forest type, there was height underestimation in broadleaf and mixedwood stands, as previously observed with GLAS, and a negligible overestimation in conifer stands (**Error! Not a valid bookmark self-reference.**). This could also be a consequence of the leaf-off conditions prevalent during the GLAS acquisition available for modelling, where the sensor would receive lower energy back from the canopy than in leaf-on conditions (and thus lead to lower height estimates). In terms of the linear regression between predicted (SVI) and observed (BT ALS) stand height values, the adjusted R^2 in conifer stands was double that for broadleaf in Phase 1 and about the same in Phase 2, while the RMSD was higher in broadleaf for both phases (**Error! Not a valid bookmark self-reference.**). This suggests that the SVI stand height estimates were best for conifer stands 10 to 15 m tall, which were more prevalent in Phase 1, while the performance for broadleaf stands was not as good but was more uniform across the project area. The histogram of differences (SVI–BT) in stand height followed a symmetric normal distribution (**Error! Reference source not found.**a), and the distribution of stand height values in the SVI was similar to that of BT estimates, with SVI exhibiting a slightly narrower range in Phase 1 (**Error! Reference source not found.**b), and an even narrower range in Phase 2, where the majority of the cells had stand height < 10 m (Figure S7c).

Although the mean differences were minor, particularly in conifer stands, which represented over 70% of the forests in the project area, there could be fairly large over- or under-estimation errors in some pixels. When we inspected the areas with the largest absolute errors, there appeared to be at least 4 common factors:

1. heterogeneous landcover
2. rough terrain
3. k-NN bias at lower (overestimation) and upper (underestimation) ends of attribute distribution
4. landcover changes in the period between the 2 datasets

Areas with heterogeneous landcover (Figure) had a lot of variability in the difference between predicted and observed stand height (Figure). The west-central portion of Figurec displayed an area with variable stand height. The same area in Figured revealed that the k-NN imputation had a smoothing effect and reduced the variability of estimates in that area, thus increasing the absolute differences in sites taller or shorter than the average local site. Also, the western portion of Figurea corresponds to a tall stand and the eastern portion to a short stand. The corresponding regions in Figuree show that the SVI underestimated the tall stand and overestimated the short stand. This example suggests that the SVI estimates may be more reliable in homogenous stands of uniform height.

Table S8. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between boreal transect (BT) derived values of stand height in 25-m cells and their k-NN counterparts, including pixel-wise signed (Δ) and absolute (|Δ|) differences for (a) Phase 1 and (b) Phase 2. Results are provided both for all cells and broken down by forest type. NB. The last 4 rows are the statistics for the predicted (x, i.e., Satellite Vegetation Inventory, SVI) vs. observed (y, i.e., BT) linear model (y = a + bx): RMSD, root mean square difference; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a.

a) Phase 1

	All				Conifer				Broadleaf				Mixedwood			
	BT	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**
n	1091056	1091056	1091056	1091056	838606	838606	838606	838606	103711	103711	103711	103711	148739	148739	148739	148739
Min	2.5	3.7	–27.1	0.0	2.5	3.7	–27.1	0.0	2.6	4.0	–26.4	0.0	2.6	4.0	–25.4	0.0
Max	35.0	31.0	23.9	27.1	34.8	30.8	23.9	27.1	35.0	31.0	23.9	26.4	34.9	30.9	23.6	25.4
Mean	11.5	11.5	**–0.1**	**3.0**	10.0	10.1	**0.1**	**2.6**	17.3	16.3	**–1.1**	**4.5**	16.4	15.8	**–0.6**	**4.2**
Median	9.5	9.9	0.0	2.0	8.4	8.6	0.1	1.8	17.3	16.6	–1.0	3.7	16.1	15.9	–0.4	3.3
SD	6.1	5.3	4.1	2.9	4.9	4.5	3.6	2.6	6.1	5.2	5.6	3.5	6.7	5.3	5.4	3.5
CV	0.5	0.5	–51.5	1.0	0.5	0.4	27.2	1.0	0.4	0.3	–5.3	0.8	0.4	0.3	–9.0	0.8
RMSD	4.1				3.6				5.7				5.4			
Adj. R^2	0.55				0.51				0.27				0.38			
Slope	0.8				0.8				0.6				0.8			
Intercept	1.9				2.1				7.4				4.2			

b) Phase 2

	All				Conifer				Broadleaf				Mixedwood			
	BT	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**
n	415256	415256	415256	415256	391289	391289	391289	391289	7999	7999	7999	7999	15968	15968	15968	15968
Min	2.6	4.3	–17.1	0.0	2.6	4.3	–17.1	0.0	2.6	4.4	–13.9	0.0	2.7	4.6	–14.3	0.0
Max	26.6	22.0	11.5	17.1	26.6	21.5	11.5	17.1	24.3	22.0	9.4	13.9	25.6	22.0	9.8	14.3
Mean	7.6	7.2	**–0.4**	**1.6**	7.6	7.2	**–0.4**	**1.6**	8.8	7.9	**–0.9**	**2.3**	8.3	8.0	**–0.3**	**1.9**
Median	7.2	6.9	–0.2	1.2	7.2	6.8	–0.2	1.2	8.1	7.3	–0.6	1.8	7.7	7.3	–0.2	1.4
SD	2.4	1.7	2.1	1.4	2.3	1.6	2.1	1.4	3.3	2.6	2.9	2.0	3.0	2.3	2.5	1.7
CV	0.3	0.2	–5.1	0.9	0.3	0.2	–5.1	0.9	0.4	0.3	–3.3	0.9	0.4	0.3	–7.8	0.9
RMSD	2.1				2.1				3.1				2.5			
Adj. R^2	0.26				0.25				0.27				0.34			
Slope	0.7				0.7				0.7				0.8			
Intercept	2.5				2.5				3.6				2.1			

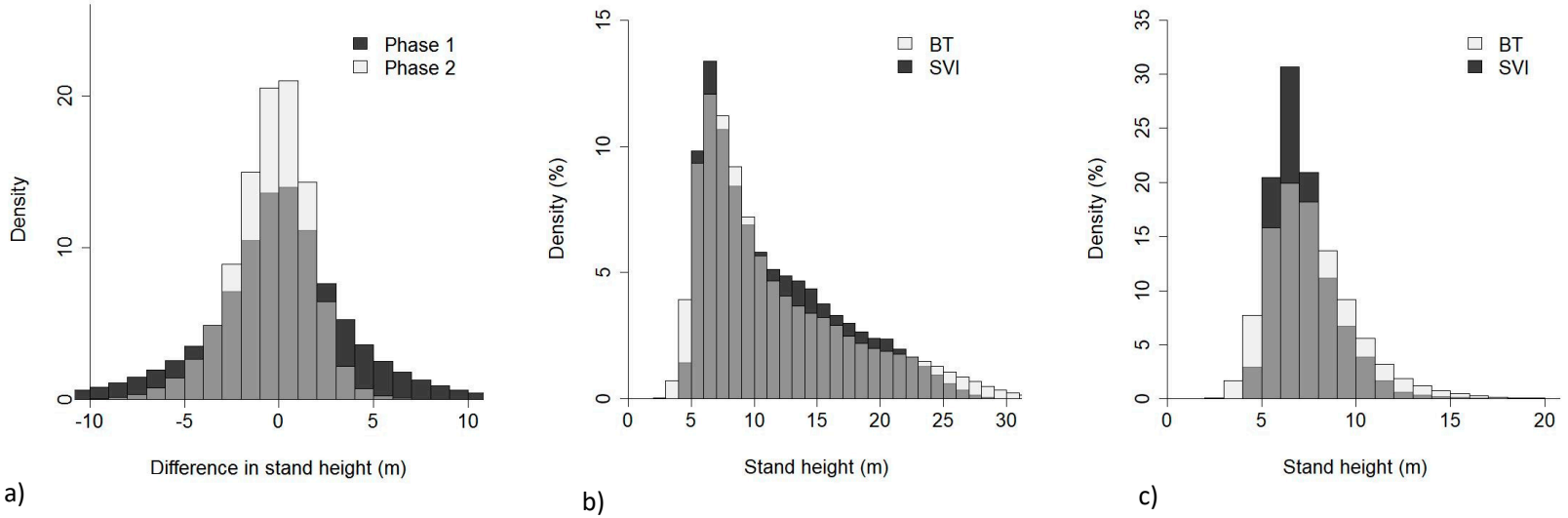


Figure S7. Histogram of the distribution in (a) stand height difference, and stand height values between the boreal transect (25 m) and SVI for (b) Phase 1 and (c) Phase 2. BT, boreal transect; SVI, Satellite Vegetation Inventory.

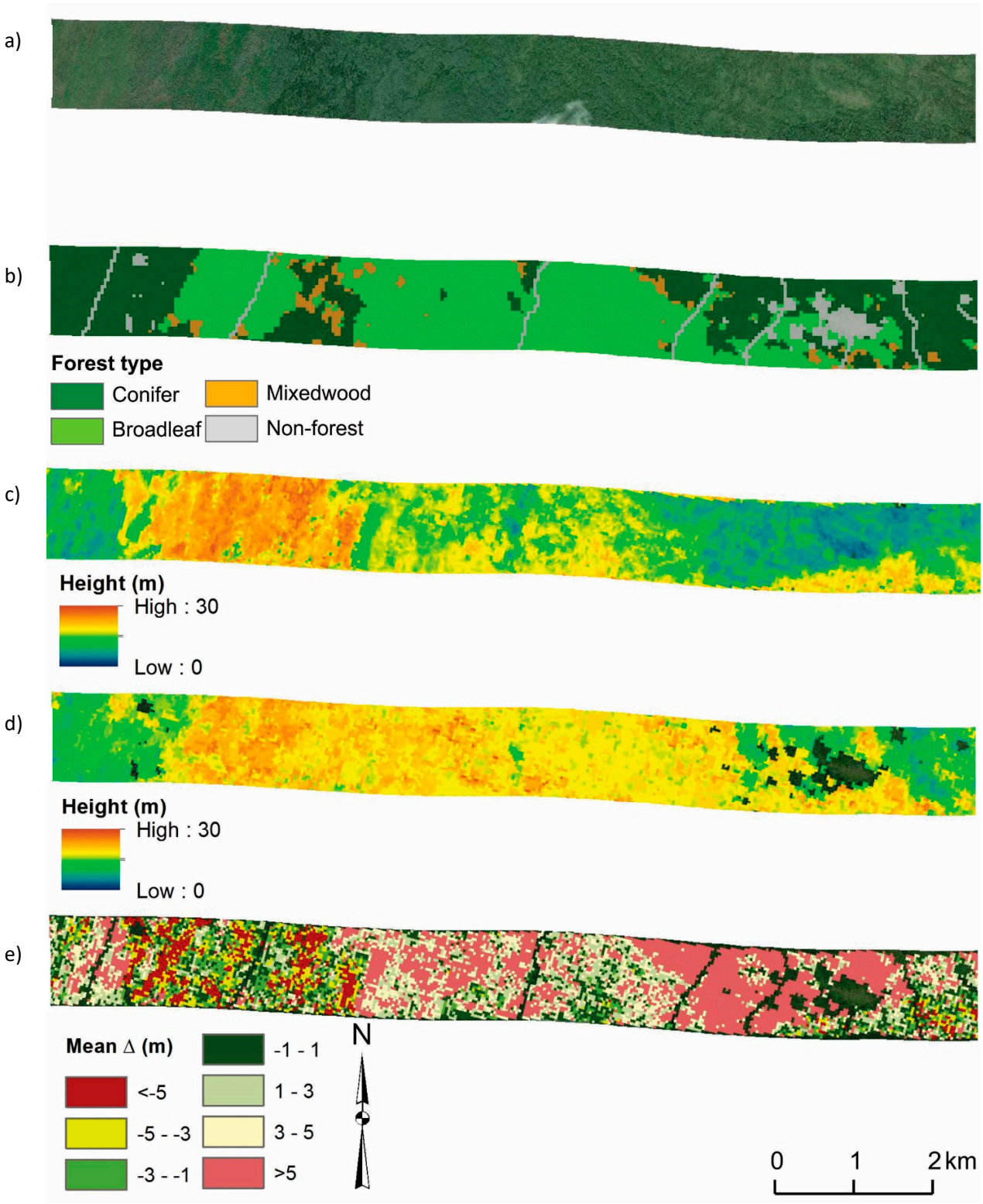


Figure S8. Effect of heterogeneous landcover on estimates of stand height in a 10-km portion of the boreal transect (BT). (a) Worldview 3 true colour composite (2019); (b) forest type; (c) reference stand height based on the BT airborne

laser scanning (ALS) data; (d) Satellite Vegetation Inventory (SVI) stand height, (e) predicted (SVI) minus observed (BT) differences in stand height.

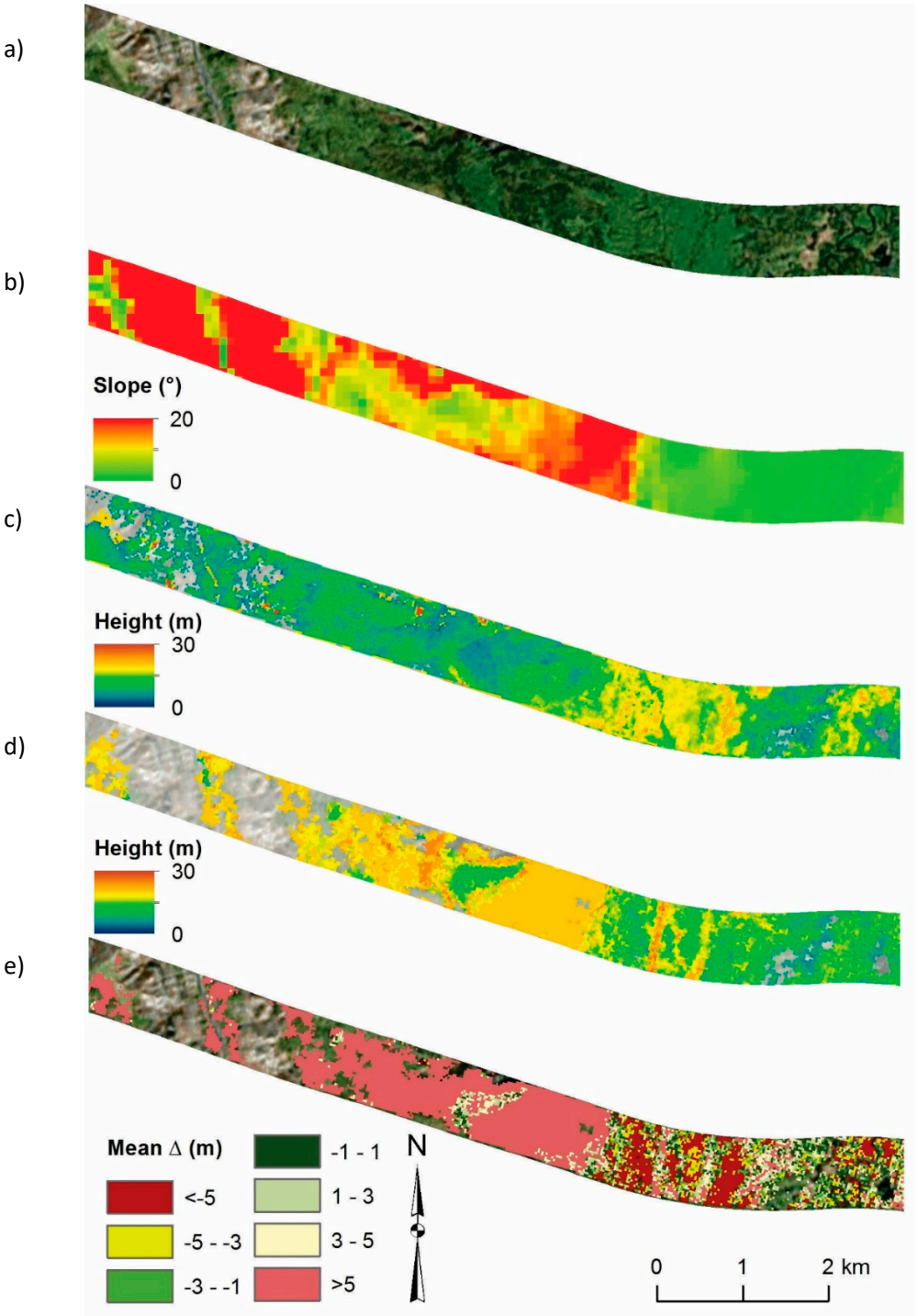


Figure S9. Effect of rough terrain on estimates of stand height in a 10-km portion of the boreal transect (BT). (a) Landsat true colour composite; (b) slope product (CDEM 90 m); (c) BT stand height (25 m); (d) Satellite Vegetation Inventory

(SVI) stand height; (e) stand height predicted minus observed difference (SVI – BT). Stand height is grossly overestimated in high slope terrain.

Rough terrain is in general troublesome for remote sensing data: GLAS waveforms will be skewed in high slope; ALS ground point classification is less reliable, which confounds the resulting percentile metrics; radar backscatter is sensitive to variations in local incidence angle, which can cause foreshortening, overlay, and shadows over steeper terrain; and optical reflectance data are also sensitive to local incidence angle and will display shadows on north-facing slopes, resulting in landcover misclassification. Figure illustrates how the SVI stand height estimate can be affected by rough terrain. Forested areas with high slope (>10°) (centre of Figure b) are misrepresented in the SVI raster (Figure d) where there is a uniform stand height (roughly 20 m) in comparison with the BT-derived stand height (Figure c), which shows variable height estimates. This could be a consequence of the lack of valid GLAS footprints in rugged terrain, coupled with the aforementioned effects on the ancillary data used to decide what footprints are used in the k-NN imputation.

For crown closure, the relative differences between SVI and BT were even smaller (13% in Phase 1 and 16% in Phase 2, (Table). The mean difference was very similar between forest types (not shown). The difference in crown closure between SVI and BT showed a normal distribution with a slight negative bias (Figure 10a). SVI estimates had a narrower range than the BT estimates (Figure b and S10c) and did not reflect denser areas well. It is unclear, however, if the denser areas that appear in the BT crown closure estimates are an artifact caused by the power issues experienced in the 2007 Fort Simpson ALS survey and propagated by applying the Fort Simpson model to the BT ALS. This suspicion was supported when we compared the Lz metric (main paper 4.2) between the Fort Simpson and BT ALS datasets in overlapping areas (

Figure **S4**.1; Appendix SA1), which showed a poor relationship between the Lz metric (used to model crown closure) of both ALS acquisitions (Adj. $R^2 = 0.36$, RMSE = 0.14). In addition, GLAS prediction of crown closure is not as reliable in broadleaf and mixedwood stands (because of leaf-off conditions), which will be propagated to the SVI crown closure product. Interestingly, the saturation at high crown closure values in the BT was not as evident in Phase 2, presumably because of the actual lack of dense stands in Phase 2. Phase 2 also had an overall narrower range in crown closure than Phase 1 (Table).

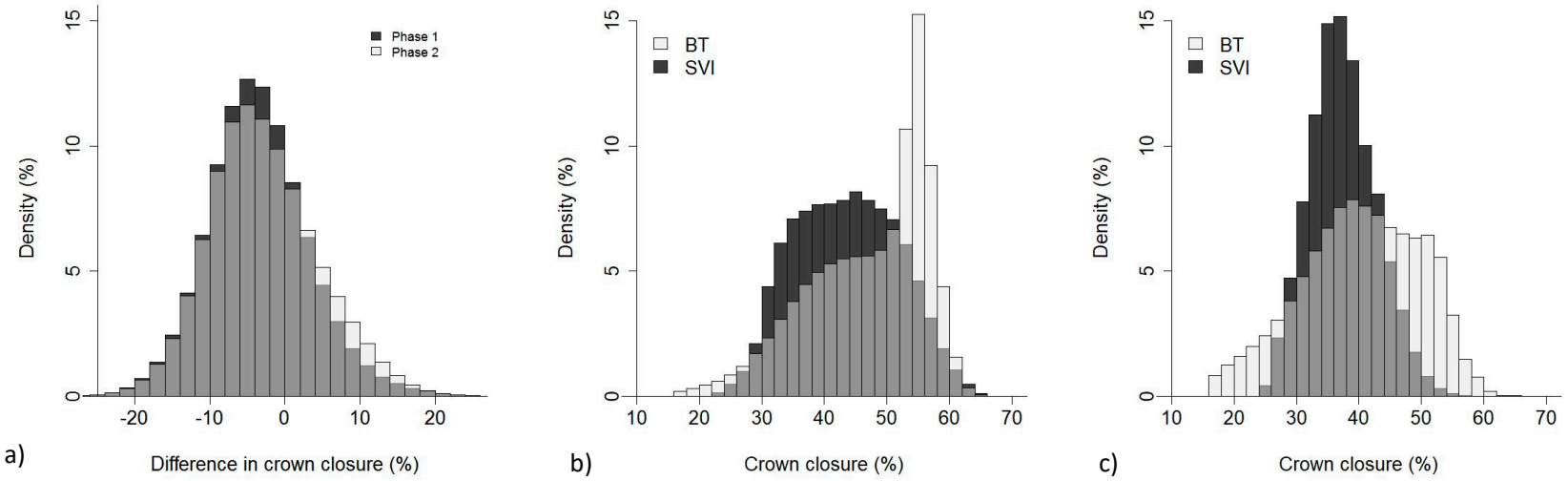


Figure S10. Histogram of the distribution in (a) crown closure difference, and crown closure values between the boreal transect (25 m) and SVI for (b) Phase 1 and (c) Phase 2. BT, boreal transect; SVI, Satellite Vegetation Inventory.

Table S9. Descriptive statistics (n, number of observations; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between boreal transect (BT) derived values of crown closure in 25-m cells and their Satellite Vegetation Inventory (SVI) counterparts, including pixel-wise signed (Δ) and absolute (|Δ|) differences for (a) Phase 1 and (b) Phase 2. NB. The last 2 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., BT) linear model (y = a + bx): RMSD, root mean square difference; Adj. R^2, adjusted coefficient of determination.

	Phase 1				Phase 2			
	BT	SVI	Δ	|Δ|	BT	SVI	Δ	|Δ|
n	1091056	1091056	1091056	1091056	415256	415256	415256	415256
Mean	47.1	43.5	–3.5	6.2	40.3	37.5	–2.8	6.4
Median	49.4	43.5	–3.9	5.4	40.6	37.1	–3.4	5.7
SD	9.4	8.3	6.9	4.6	9.5	5.3	7.3	4.5
CV	0.2	0.2	–1.9	0.7	0.2	0.1	–2.6	0.7
RMSD	7.7				7.8			
Adj. R^2	0.50				0.41			

S4.4. Boreal Transect ALS Part 2: Comparison of 150-m SVI and 150-m ALS[16]

As expected, the predicted vs. "observed" differences for stand height at 150-m resolution were smaller than those at 30-m resolution (Table). The bias remained close to 0, and the mean absolute difference was reduced from 3 m at 30 m to 2.2 m at 150 m in Phase 1, and from 1.6 m at 30 m to 1.2 m at 150 m in Phase 2. The mutual cancellation of local underestimations and overestimations resulted in smaller differences and reduced the range of variation of the pixel-wise differences at 150 m (Figure S11). As with the comparison at 25-m resolution, the mean difference in stand height was worse in broadleaf and mixedwood stands than in conifer stands for Phase 1. The same pattern almost held true for Phase 2; however, the small sample size for broadleaf and mixedwood stands precluded a clear conclusion. In terms of the linear regression between predicted (SVI) and "observed" (BT ALS) stand height values for 150-m cells, the RMSE was lower, the adjusted R^2 increased, the slope was closer to 1, and the intercept was lower than the regression for 30-m cells. This indicates that at a coarser resolution, pixel-wise accuracy improved as also shown in Beaudoin et al. [37], but there was somehow less variability in the predicted (SVI) values at 150 m, and there was also a decrease in the frequency of cells at the low and high ends of stand height and a concomitant increase for middle values, in both phases 1 and 2 (Figure S12). This shift may have been the result of the sporadic occurrence of stand heights < 5 m and > 25 m where these estimates got averaged out within a 150-m cell. Choosing a cell size that balances this and other information losses (e.g., loss of spatial detail and increased proportion of mixed pixels) against the gains in accuracy would require further analysis.

[16] Given that volume and biomass estimates are functions of height, the assessments using the boreal transect ALS data only include stand height. Crown closure was not assessed at 150 m because of the ALS power issues described in the previous section.

Table S10. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between boreal transect (BT) derived values of stand height resampled at 150-m cells and their Satellite Vegetation Inventory (SVI) counterparts, including pixel-wise signed (Δ) and absolute (|Δ|) differences for (a) Phase 1 and (b) Phase 2. NB. The last 4 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., BT) linear model (y = a + bx): RMSE, root mean square error; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a.

a) Phase 1

	All				Conifer				Broadleaf				Mixedwood			
	BT	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**
n	22303	22303	22303	22303	15569	15569	15569	15569	1365	1365	1365	1365	5369	5369	5369	5369
Min	4.1	4.6	–18.7	0.0	4.3	4.6	–13.8	0.0	4.3	5.3	–18.7	0.0	4.1	5.0	–17.9	0.0
Max	32.9	28.5	16.7	18.7	29.1	28.0	16.7	16.7	32.9	28.5	13.3	18.7	32.2	27.5	14.9	17.9
Mean	11.9	11.8	–0.1	2.2	9.9	10.0	0.1	1.8	17.5	16.6	–0.9	3.4	16.3	15.8	–0.5	3.1
Median	10.0	10.5	0.0	1.4	8.6	8.8	0.1	1.2	17.2	17.0	–1.0	2.9	15.7	15.7	–0.3	2.3
SD	5.5	4.9	3.1	2.3	4.1	3.9	2.6	1.9	5.1	4.7	4.3	2.7	5.5	4.5	4.0	2.6
CV	0.5	0.4	–28.7	1.0	0.4	0.4	29.1	1.1	0.3	0.3	–4.9	0.8	0.3	0.3	–8.2	0.9
RMSE	3.1				2.6				4.4				4.0			
Adj. R^2	0.68				0.63				0.37				0.49			
Slope	0.9				0.8				0.7				0.9			
Intercept	1.1				1.5				6.4				2.7			

b) Phase 2

	All				Conifer				Broadleaf				Mixedwood			
	BT	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**	**BT**	**SVI**	**Δ**	**|Δ|**
n	5413	5413	5413	5413	4931	4931	4931	4931	20	20	20	20	462	462	462	462
Min	4.1	4.8	–8.5	0.0	4.1	4.8	–8.0	0.0	5.7	5.8	–8.4	0.4	4.9	5.3	–8.5	0.0
Max	20.8	17.5	5.5	8.5	19.1	17.5	5.5	8.0	17.4	12.2	3.3	8.4	20.8	15.4	4.6	8.5
Mean	8.2	7.8	–0.4	1.2	8.1	7.7	–0.4	1.2	11.8	8.8	–3.0	3.7	8.7	8.3	–0.4	1.5
Median	7.9	7.5	–0.2	0.9	7.9	7.4	–0.2	0.9	12.3	8.7	–3.0	3.3	8.1	7.8	–0.2	1.2
SD	2.0	1.5	1.6	1.1	1.9	1.5	1.6	1.1	3.4	1.7	3.4	2.6	2.5	1.9	1.9	1.3
CV	0.2	0.2	–4.1	0.9	0.2	0.2	–4.1	0.9	0.3	0.2	–1.1	0.7	0.3	0.2	–4.8	0.9
RMSE	1.7				1.6				4.5				1.9			
Adj. R^2	0.35				0.34				0.00				0.45			
Slope	0.8				0.7				0.4				0.9			
Intercept	2.2				2.3				8.1				1.4			

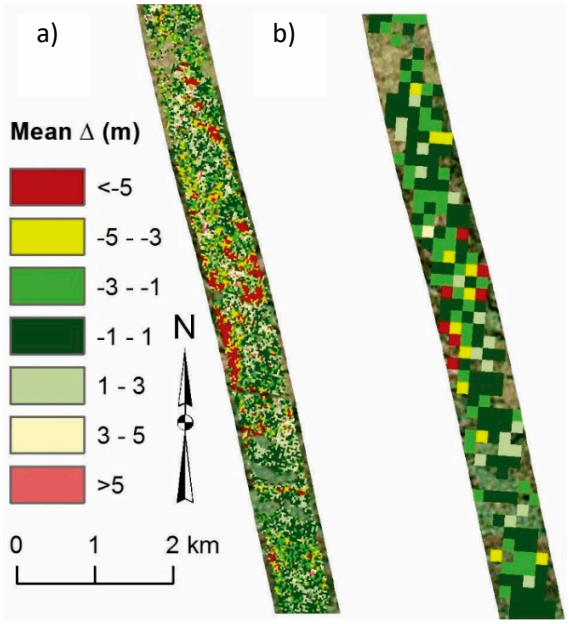


Figure S11. Sample boreal transect portion illustrating how the variability and differences seen at (a) 25 m are reduced at (b) 150 m.

www.mdpi.com/article/10.3390/rs14051108/s1

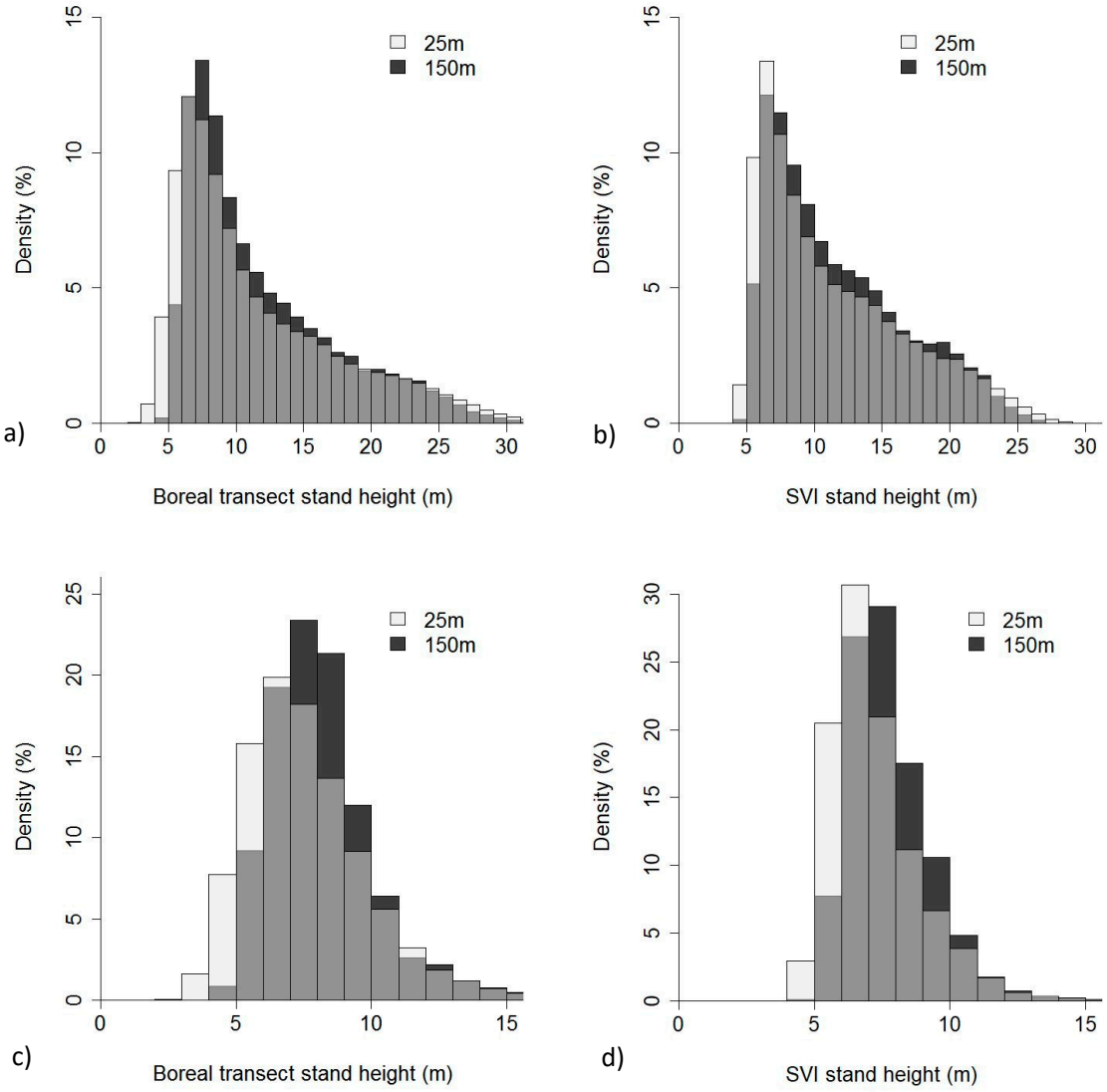


Figure S12. Comparing the distribution in stand height between 25 m and resampled to 150 m: (a) Phase 1 boreal transect; (b) Phase 1 SVI; (c) Phase 2 boreal transect; and (d) Phase 2 SVI. SVI, Satellite Vegetation Inventory.

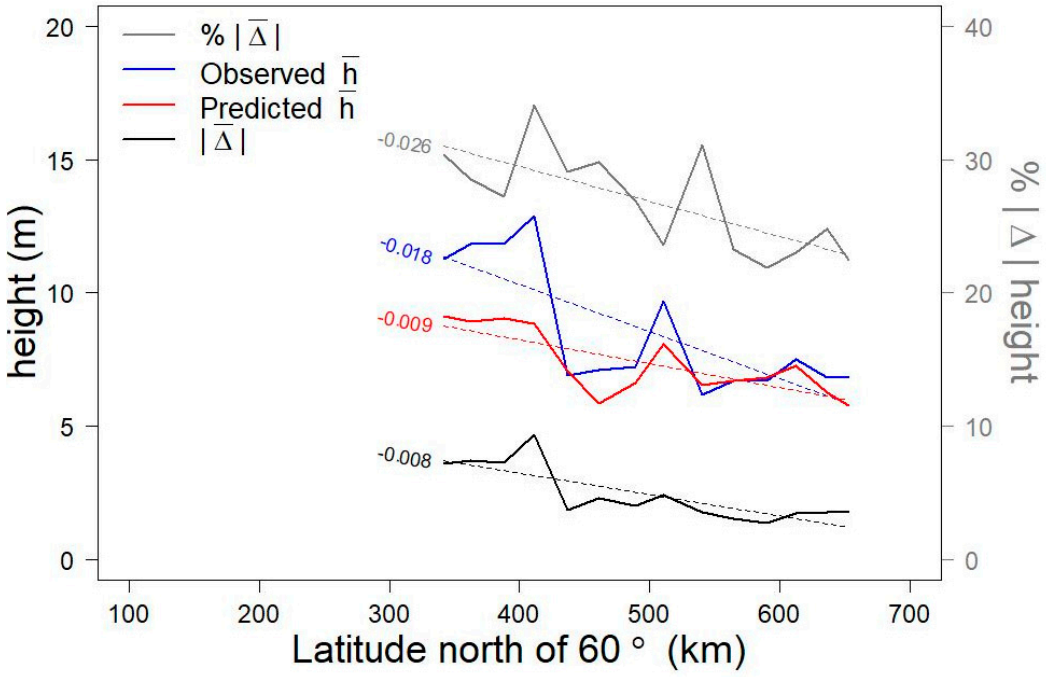
S4.5. GNWT ALS: Comparison of 30-m SVI and 25-m ALS

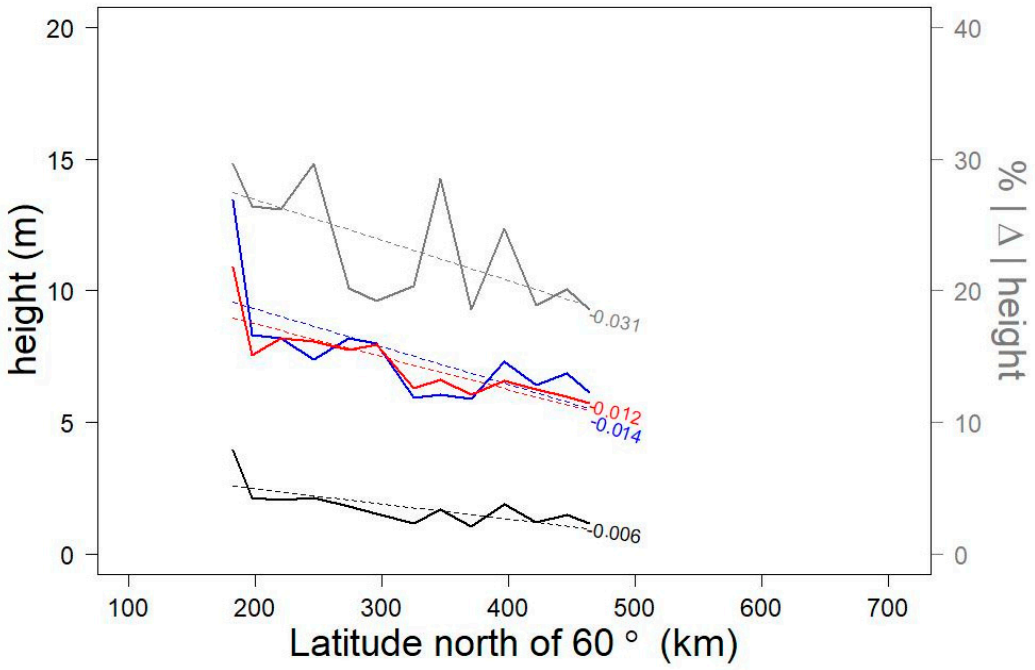
All Government of Northwest Territories (GNWT) ALS datasets were located in Phase 2 (Figure S1). The bias in stand height was more variable (–0.9 m to 0.7 m) than for the Phase 2 boreal transect, but the mean absolute error was similar (1.3 m to 2.4 m, Table 1). Much like the findings for the BT dataset, the mean and median were similar between a given ALS dataset and the SVI, with the range of SVI much narrower than that of the ALS (with the exception of Great Bear). The adjusted R^2 values were low (0.4 to 0.2) to very low (< 0.1), indicating poor agreement between the ALS estimates and the SVI. There are several possible explanations for this:

i) Median stand height in most of the GNWT ALS datasets was < 7 m, and low stand height is problematic for GLAS stand height prediction.
ii) No comparison was available between the Fort Simpson ALS and these datasets, meaning that transferability of the Fort Simpson ALS model could be an issue.
iii) There was frequent local variability in the GNWT ALS datasets, a situation where k-NN performs less well.

The GNWT ALS datasets were accompanied by 20-cm orthophotos, which enabled us to conduct an additional investigation into sources of error. When visually inspecting the spatial locations of high error, we found a similar trend to that in the BT, where the highest error occurred in areas of heterogeneous landcover, high slope, disturbance, and broadleaf vegetation. Regarding the latter, Figure S13 shows 2 examples of situations that may contribute to larger SVI errors in broadleaf areas: broadleaf shrubs may be misclassified as broadleaf forest in the landcover map, causing confusion in the k-NN imputation where nearest neighbours in the feature space could be tall broadleaf GLAS footprints (Figure S13a); conversely, a broadleaf forest with few standing trees may produce a signature similar to that of GLAS footprints where the underestimation error is more severe (Figure S13b) .

Because of their extended latitudinal range and north–south orientation, we used the Mackenzie Valley Highway (MackHwy) and a segment of the BT ALS (dashed portion in Figure S1) to assess whether the difference between predicted (SVI) and observed (MackHwy, BT) stand height was a function of latitude. There was a weak relationship (adjusted R^2 0.05 to 0.41) between latitude and all stand height metrics: observed, predicted, the absolute difference, and the percent absolute difference (a)





b)

Figure S14. Relationship between latitude and various stand height metrics (observed, i.e., derived from the ALS; predicted in SVI; and predicted minus observed, both absolute and relative) for (a) the Mackenzie Valley highway ALS and (b) the BT ALS. Each data point is the mean value of 500 randomly selected forest cells within a 25-km interval of latitude centred at the abscissa of the point. P-values for all slopes (slope value shown beside trend line) are significant ($p < 0.05$). ALS, airborne laser scanning; BT, boreal transect.

). However, all trends showed that both stand height and its associated error significantly decreased with increasing latitude. Mean stand height decreases 1.4 to 1.8 m each 100 km we go north, whereas the mean absolute and relative errors decrease less, about 70 cm for absolute error and 2.6 percentage points for relative error. This is contrary to our a priori assumption that errors would increase with latitude. While the two latitudinal profiles showed a similar trend, the MackHwy ALS covered a more productive region than the BT segment and exhibited greater variability in both predicted and observed stand height, as well as associated error.

Table S11. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between stand height values derived from different GNWT ALS data in 25-m cells and their SVI counterparts, including the mean pixel-wise signed (Δ) and absolute (|Δ|) differences. NB. The last 4 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., ALS) linear model (y = a + bx): RMSE, root mean square error; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a. ALS, airborne laser scanning; GNWT, Government of the Northwest Territories; SVI, Satellite Vegetation Inventory.

	Willow River (centre ~62.7°N)		**Little Smith** (centre ~64.4°N)		**Big Smith** (centre ~64.6°N)		**Great Bear** (centre 64.9°N)		**Thunder River** (centre ~67.5°N)		**Mackenzie Highway** ~(63.2°N to 65.8°N)	
	ALS	**SVI**	**ALS**	**SVI**	**ALS**	**SVI**	**ALS**	**SVI**	**ALS**	**SVI**	**ALS**	**SVI**
n	41857	41857	47633	47633	9706	9706	36832	36832	32020	32020	359175	359175
Min	2.5	4.5	2.5	4.4	2.5	4.4	2.5	4.5	2.5	4.4	2.5	4.3
Max	24.0	19.7	25.8	20.2	26.7	14.7	17.3	22.2	34.9	16.3	27.8	20.8
Mean	8.8	8.4	7.5	6.6	6.5	6.8	5.0	5.7	6.5	6.1	8.3	7.4
Median	8.0	8.5	6.9	6.0	6.0	6.2	4.7	5.4	5.9	5.7	6.9	6.7
SD	3.7	2.1	2.9	1.9	2.5	1.6	1.5	1.0	2.4	1.4	4.1	2.1
CV	0.4	0.2	0.4	0.3	0.4	0.2	0.3	0.2	0.4	0.2	0.5	0.3
Mean Δ	**–0.3**		**–0.9**		**0.2**		**0.7**		**–0.4**		**–0.9**	
Mean |Δ|	**2.4**		**1.9**		**1.8**		**1.3**		**1.9**		**2.3**	
RMSE	3.2		2.7		2.4		1.7		2.5		3.3	
Adj. R^2	0.26		0.29		0.17		0.08		0.05		0.41	
Slope	0.9		0.8		0.7		0.4		0.4		1.2	
Intercept	1.3		2.1		2.1		2.6		4.1		–0.9	

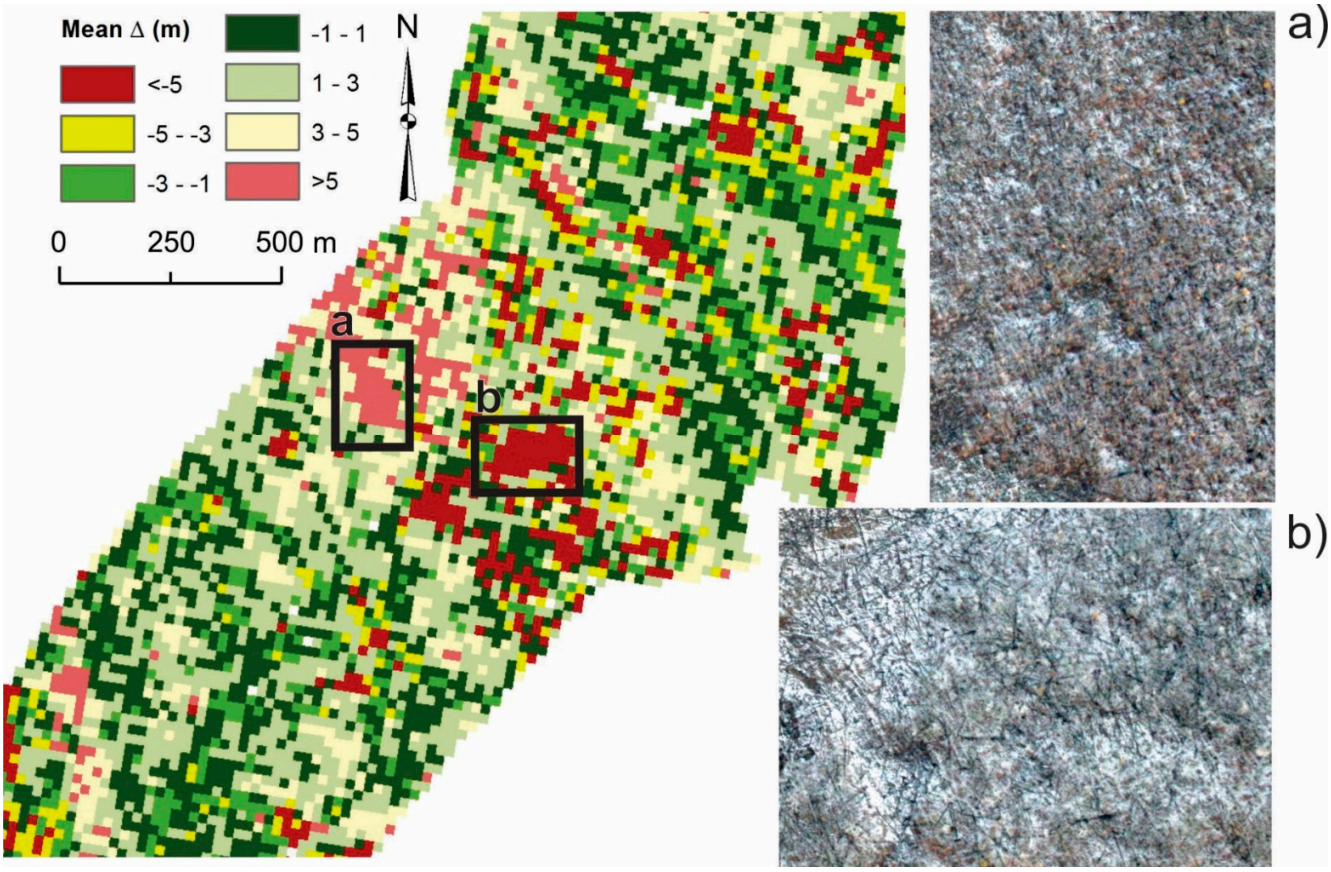
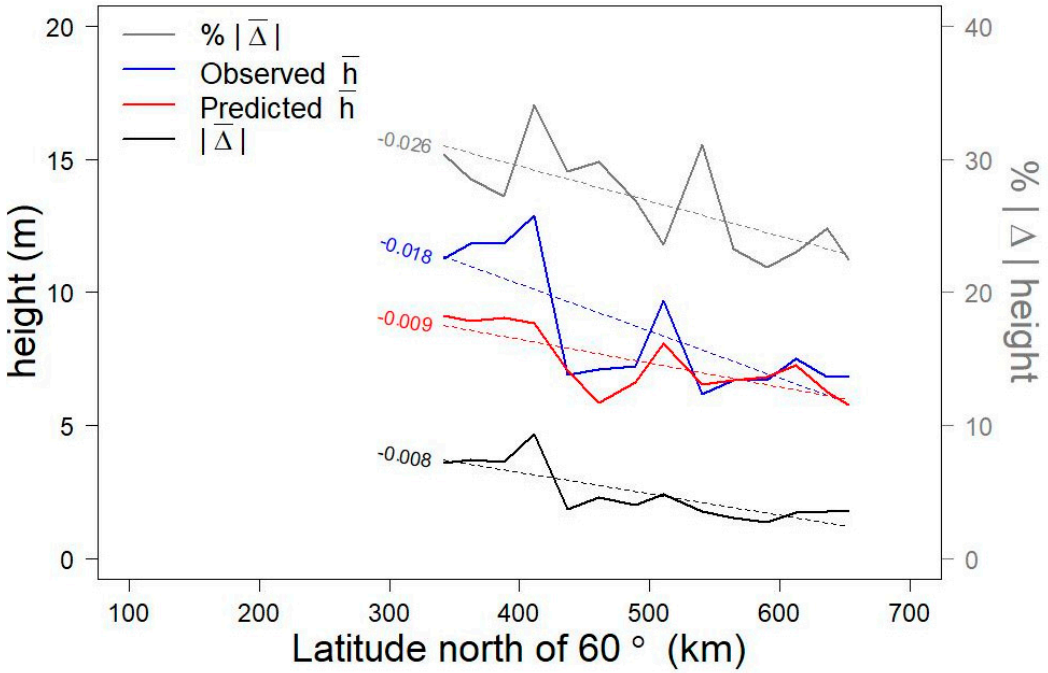
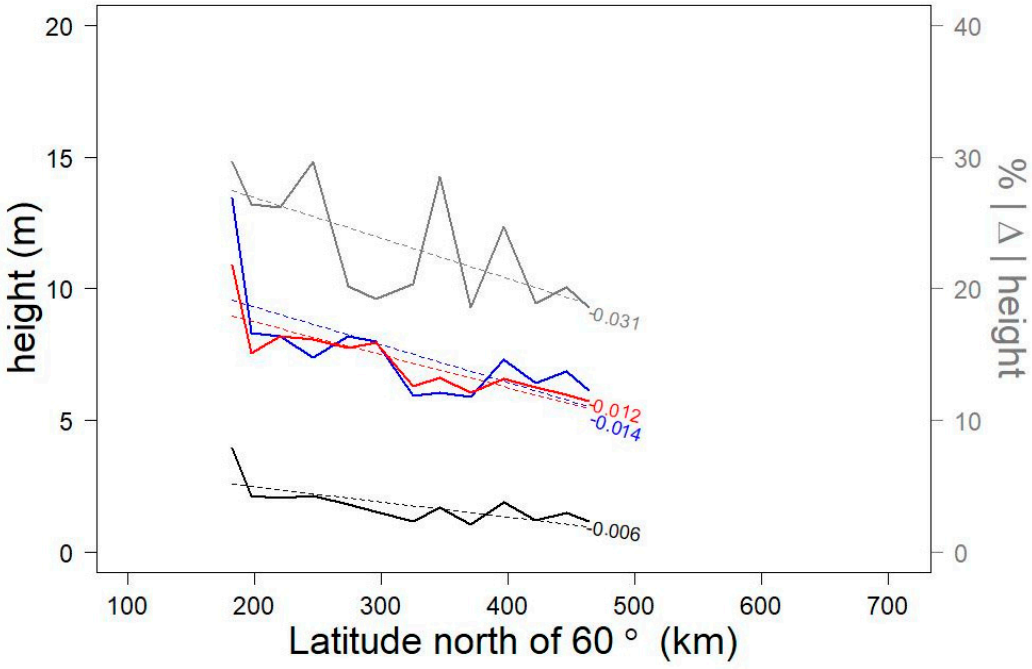


Figure S13. Investigation into sources of large differences in stand height estimates between the GNWT ALS dataset and the SVI: (a) a shrubby area displaying overestimation in heterogeneous broadleaf landcover; and (b) a deciduous forest stand with lots of deadfall where SVI underestimated stand height. ALS, airborne laser scanning; GNWT, Government of the Northwest Territories; SVI, Satellite Vegetation Inventory.

www.mdpi.com/article/10.3390/rs14051108/s1



a)



b)

Figure S14. Relationship between latitude and various stand height metrics (observed, i.e., derived from the ALS; predicted in SVI; and predicted minus observed, both absolute and relative) for (a) the Mackenzie Valley highway ALS and (b) the BT ALS. Each data point is the mean value of 500 randomly selected forest cells within a 25-km interval of latitude centred at the abscissa of the point. P-values for all slopes (slope value shown beside trend line) are significant ($p < 0.05$). ALS, airborne laser scanning; BT, boreal transect.

S4.6. Forest Vegetation Inventory

The mean SVI estimate of stand height within FVI polygons was in good agreement with the polygon-wise stand height in the Jean Marie and Axe Point inventories, both located in Phase 1 (Table a). As with the NFI, stand height was underestimated by SVI and the calculated value (–0.7 m when weighted by the area of the polygons) was close to the better end of the 95% confidence interval derived from the NFI sample (CI95 = [–3.74, –0.61]). As with the BT, the underestimation mostly occurred in tall broadleaf and mixedwood stands. The SVI distribution of stand height was much narrower than in the FVI (Figure), particularly in Phase 2 (Table b). The distribution of the difference (SVI – FVI) in stand height showed a small negative bias for Phase 1, which was emphasized in Phase 2 (Figure a). The negative bias was most likely attributable to the underestimation in tall broadleaf stands, as indicated by the larger underestimation error seen in the Jean Marie inventory where these stands are more frequent (Figure S16. FVI inventories and respective difference in stand height (SVI – FVI). When the mean difference was weighted by polygon size, the differences decreased, indicating that small polygons tend to perform worse than large polygons. This is akin to the trend observed in the BT when resampled to 150 m, where a coarser resolution (or larger polygon in the case of FVI) will average out local variability and produce a better mean estimate.

The SVI estimate of crown closure showed general agreement with the FVI, with the mean difference worse in broadleaf and mixedwood stands for both phases 1 and 2 (Table). As with stand height, the SVI provided a much narrower range in values than the FVI (Figure), with all values mostly occurring between 30% and 60% compared with between 10% and 90% in the FVI. In an attempt to assess classification accuracy according to NFI classes, we classified FVI and SVI crown closure according to NFI standards and produced an error matrix. The overall accuracy computed was 55%, but 99.9% of all polygons were classified as open according to SVI and 55.7% were classified as open in the FVI, which corroborates the SVI problems with dense stands.

To assess differences between the Jean Marie, Axe Point and Behchoko inventories, the weighted (by polygon size) mean signed difference (MSD) and mean absolute difference (MAD) between SVI and FVI estimates were compared for both stand height and crown closure (Table). Differences in crown closure metrics were small among all inventories, and the same can be said for MSD in stand height, but for MAD, Behchoko has a smaller MAD than the two FVIs in Phase 1, which could be a consequence of Bechoko having less variation in height (coefficient of variation CV= 16%) than the other two inventories (CV= 30%).

Being more detailed than the SVI, the FVI can be used to assess the accuracy of broad forest stand type (conifer, broadleaf, and mixedwood). A forest stand type confusion matrix was produced between SVI and FVI; the results showed an overall accuracy of 82% for the SVI in Phase 1 (Table) and 89% in Phase 2 (Table). Therefore, the SVI provided a fairly good representation of broad forest type at least over the area of these FVIs. When compared with the 3-class confusion matrix of Phase 2 based on ground-truthed locations (Appendix SA2, Table SA2.3), overall accuracy was lower (83% vs. 89%). This could be explained because the FVI assessment was based on the majority label of SVI pixels inside each FVI polygon, and that label tends to coincide more with the reference label when the reporting units (polygons) are larger than individual pixels. If this reasoning is correct, the overall accuracy for individual pixels in Phase 1, which we could not assess because all ground-truthed locations in Phase 1 were used for training, can be expected to be just a few percentage points lower than what is reported for the polygons of the Phase 1 FVI inventories, which is still high.

Table S12. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison between FVI polygons and the mean of the SVI pixels inside them for stand height, including polygon-wise signed (Δ) and absolute (|Δ|) differences for (a) Phase 1 and (b) Phase 2. Mean (Aweight) is the mean weighted by the area of the polygon. NB. The last 4 rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., FVI) linear model (y = a + bx): RMSE, root mean square error; Adj. R^2, adjusted coefficient of determination; slope, b; intercept, a. FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

a) Phase 1

	All				Conifer				Broadleaf				Mixedwood			
	FVI	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**
n	56721	56721	56721	56721	40059	40059	40059	40059	8556	8556	8556	8556	8106	8106	8106	8106
Min	1.0	4.3	–69.4	0.0	1.0	4.3	–18.9	0.0	1.0	4.4	–17.7	0.0	1.0	4.4	–69.4	0.0
Max	76.0	24.3	18.0	69.4	32.0	24.3	18.0	18.9	28.0	22.1	14.3	17.7	76.0	22.8	13.7	69.4
Mean	11.6	10.3	–1.2	3.5	10.5	9.8	–0.7	3.2	14.4	11.6	–2.8	4.1	13.7	11.3	–2.4	4.1
Mean (Aweight)			**–0.7**	**3.1**			**–0.2**	**2.8**			**–2.6**	**3.9**			**–2.2**	**4.2**
Median	11.0	10.0	–0.7	2.8	9.0	9.5	–0.1	2.6	15.0	12.0	–2.8	3.5	14.0	11.6	–2.4	3.5
SD	5.6	3.1	4.3	2.8	5.1	2.8	4.1	2.6	6.0	3.6	4.3	3.0	5.8	3.2	4.5	3.0
CV	0.5	0.3	–3.4	0.8	0.5	0.3	–5.9	0.8	0.4	0.3	–1.5	0.7	0.4	0.3	–1.9	0.7
RMSE	4.4				4.1				5.1				5.1			
Adj. R^2	0.44				0.38				0.51				0.41			
Slope	1.2				1.1				1.2				1.2			
Intercept	–0.9				–0.6				0.7				0.3			

b) Phase 2

	All				Conifer				Broadleaf				Mixedwood			
	FVI	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**
n	14772	14772	14772	14772	13891	13891	13891	13891	181	181	181	181	700	700	700	700
Min	1.0	4.5	–17.6	0.0	1.0	4.5	–17.6	0.0	2.0	5.5	–11.5	0.1	2.0	5.3	–13.1	0.0
Max	28.0	17.8	7.4	17.6	28.0	14.4	7.4	17.6	21.0	17.8	6.9	11.5	22.0	14.6	6.0	13.1
Mean	9.9	8.3	–1.6	3.1	9.8	8.3	–1.6	3.1	12.5	9.4	–3.1	3.7	10.9	9.0	–1.9	3.1
Mean (Aweight)			**–0.9**	**2.2**			**–1.1**	**2.8**			**–2.8**	**3.4**			**–1.6**	**2.9**
Median	10.0	8.2	–1.5	2.6	10.0	8.1	–1.4	2.6	12.0	9.1	–3.2	3.4	11.0	8.9	–2.0	2.7
SD	4.0	1.4	3.5	2.4	4.0	1.3	3.6	2.4	3.4	2.1	3.1	2.3	3.7	1.5	3.3	2.2
CV	0.4	0.2	–2.2	0.8	0.4	0.2	–2.3	0.8	0.3	0.2	–1.0	0.6	0.3	0.2	–1.7	0.7
RMSE	3.9				3.9				4.4				3.8			
Adj. R^2	0.22				0.22				0.18				0.20			
Slope	1.4				1.4				0.7				1.1			
Intercept	–1.5				–1.8				5.8				0.9			

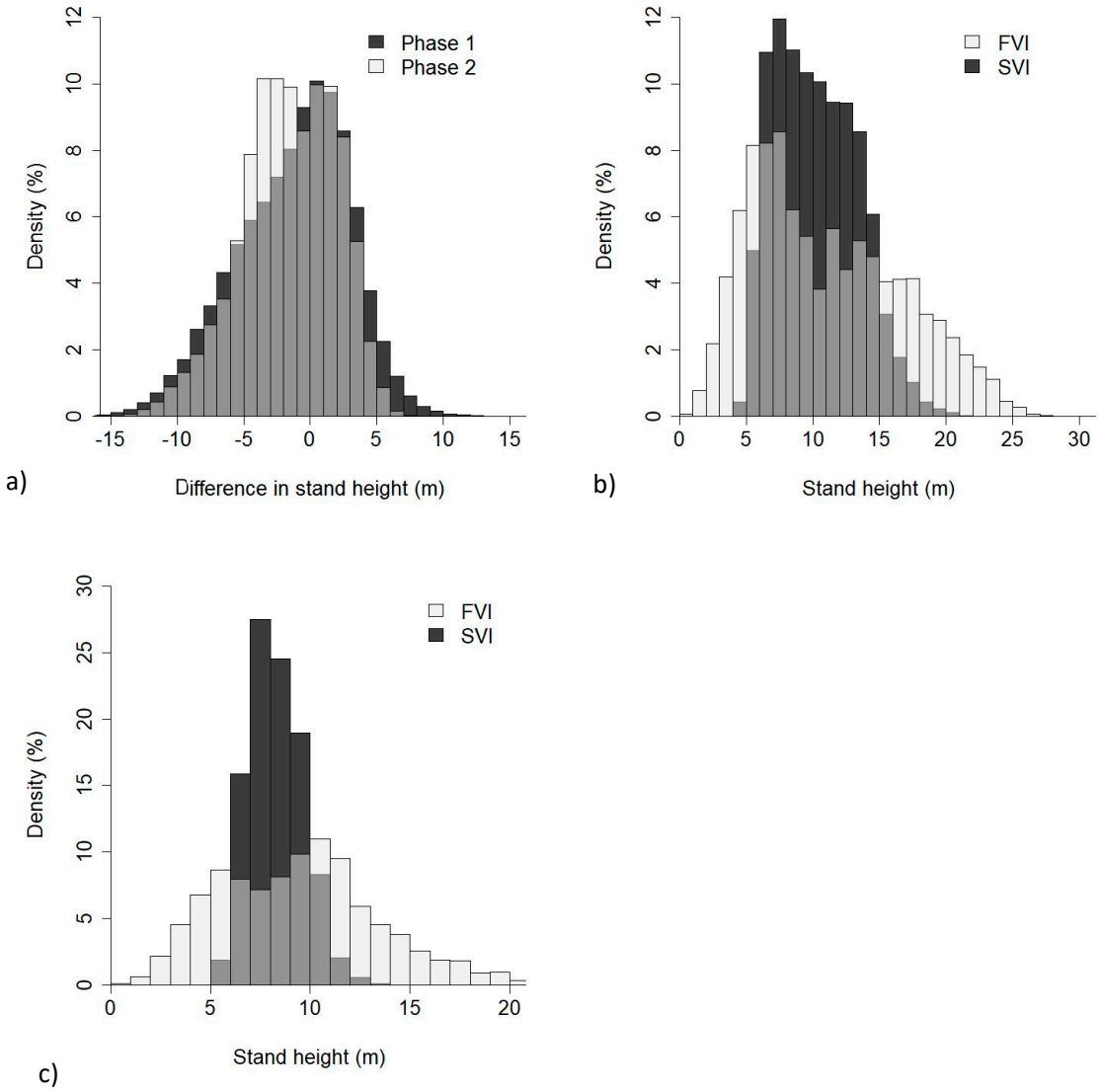


Figure S15. Histogram of the distribution in (a) stand height difference, and stand height values between the FVI(s) and SVI for (b) Phase 1 and (c) Phase 2. FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

www.mdpi.com/article/10.3390/rs14051108/s1

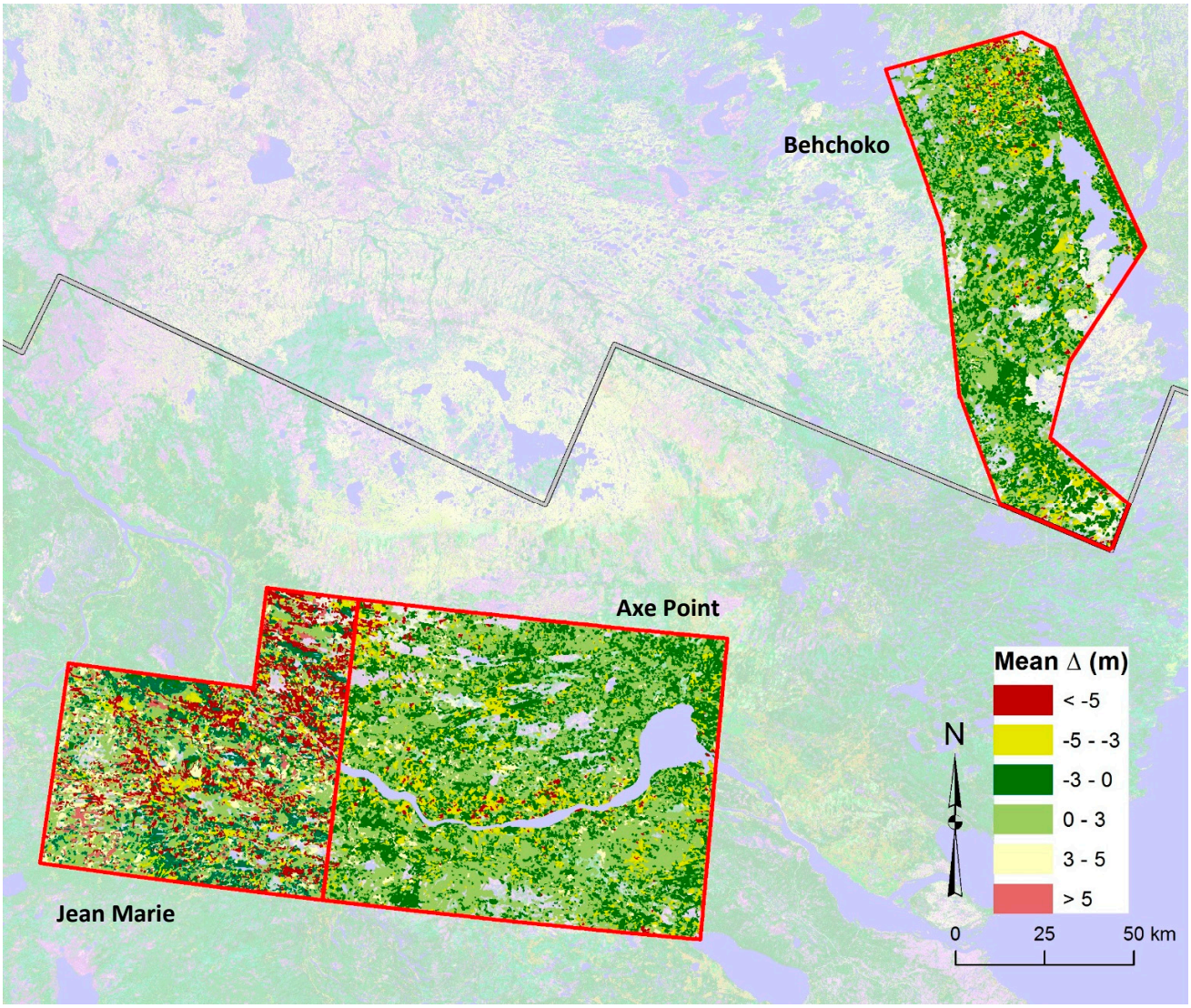


Figure S16. FVI inventories and respective difference in stand height (SVI – FVI). FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

Table S13. Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) of the comparison for crown closure between FVI polygons and the mean of SVI pixels inside them, including polygon-wise signed (Δ) and absolute (|Δ|) differences for (a) Phase 1 and (b) Phase 2. NB. The last 4 rows in the table are the statistics for the predicted (x) vs. observed (y) linear model (y = a + bx). Results are broken down by forest type. Mean (Aweight) is the mean weighted by the area of the polygon. NB. The last two rows are the statistics for the predicted (x, i.e., SVI) vs. observed (y, i.e., FVI) linear model (y = a + bx): RMSE, root mean square error; Adj. R^2, adjusted coefficient of determination. FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

a) Phase 1

	All				Conifer				Broadleaf				Mixedwood			
	FVI	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**
n	56,721	56,721	56,721	56,721	40,059	40,059	40,059	40,059	8,556	8,556	8,556	8,556	8,106	8,106	8,106	8,106
Min	6.0	23.5	–60.6	0.0	6.0	24.0	–58.3	0.0	10.0	23.5	–60.6	0.0	10.0	24.6	–56.2	0.0
Max	95.0	60.9	46.2	60.6	90.0	60.1	46.2	58.3	95.0	59.9	46.1	60.6	90.0	60.9	44.8	56.2
Mean	49.0	41.9	–7.2	16.7	46.3	41.2	–5.1	15.6	56.5	43.5	–13.0	21.3	54.5	43.3	–11.2	17.3
Mean (Aweight)			**–2.3**	**15.6**			**0.5**	**14.5**			**–14.3**	**21.7**			**–9.9**	**16.8**
Median	50.0	42.1	–10.0	15.7	50.0	41.4	–7.1	14.5	60.0	43.8	–18.0	21.4	60.0	43.6	–14.7	16.9
SD	19.9	5.6	18.5	10.7	19.5	5.4	18.0	10.3	21.3	6.1	20.5	11.6	17.0	5.4	16.5	9.8
CV	0.4	0.1	–2.6	0.6	0.4	0.1	–3.5	0.7	0.4	0.1	–1.6	0.5	0.3	0.1	–1.5	0.6
RMSE	19.8				18.7				24.2				19.9			
Adj. R^2	0.15				0.17				0.08				0.06			

b) Phase 2

	All				Conifer				Broadleaf				Mixedwood			
	FVI	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**	**FVI**	**SVI**	**Δ**	**|Δ|**
n	14772	14772	14772	14772	13891	13891	13891	13891	181	181	181	181	700	700	700	700
Min	10.0	25.6	–48.3	0.0	10.0	25.6	–48.3	0.0	10.0	30.9	–40.9	0.0	10.0	28.9	–41.9	0.1
Max	90.0	54.5	42.0	48.3	90.0	54.5	42.0	48.3	80.0	51.0	34.8	40.9	80.0	49.4	38.2	41.9
Mean	39.0	40.5	1.5	16.4	38.4	40.4	2.0	16.3	47.0	42.5	–4.5	18.0	50.4	42.4	–8.0	16.2
Mean (Aweight)			**–8.8**	**13.0**			**0.2**	**16.3**			**–3.5**	**19.9**			**11.6**	**17.9**
Median	40.0	40.5	–2.0	16.3	40.0	40.3	–1.5	16.3	50.0	43.0	–9.1	18.1	60.0	43.0	–13.2	15.9
SD	19.7	3.8	19.0	9.7	19.6	3.8	18.9	9.8	21.1	4.1	19.9	9.6	17.4	3.6	16.7	9.0
CV	0.5	0.1	12.8	0.6	0.5	0.1	9.3	0.6	0.4	0.1	–4.4	0.5	0.3	0.1	–2.1	0.6
RMSE	19.0				19.0				20.4				18.6			
Adj. R^2	0.07				0.07				0.14				0.07			

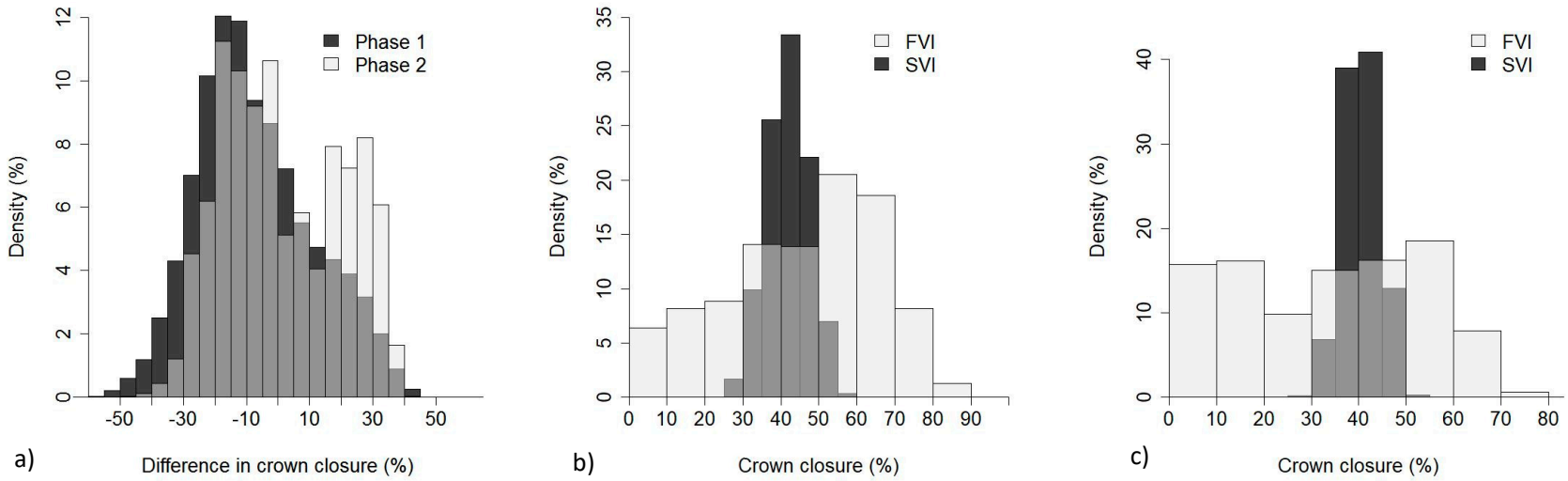


Figure S17. Histogram of the distribution in (a) crown closure difference, and crown closure values between the FVI(s) and SVI for (b) Phase 1 and (c) Phase 2. FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

Table S14. Weighted (by polygon size) mean value of the polygon-wise signed (Δ = SVI – FVI) and absolute (|Δ|) differences between FVI polygons and the mean of SVI pixels inside them in stand height and crown closure, by FVI source. FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

			Stand height (m)		Crown closure (%)	
Phase	**FVI**	**Parameter**	**Δ**	**|Δ|**	**Δ**	**|Δ|**
1	**Axe Point**	Mean (weight)	–0.4	2.8	–4.1	16.2
1	**Jean Marie**	Mean (weight)	–1.0	3.4	0.4	14.7
2	**Behchoko**	Mean (weight)	–0.9	2.2	–0.1	13.0

Table S15. Error matrix of SVI landcover (EOSD) and FVI polygon attribute TYPECLAS by area (ha) in (a) Phase 1 and (b) Phase 2. NB. Overall accuracy was computed weighing by the proportion of each class in the corresponding phase. EOSD, Earth Observation for Sustainable Development of Forests; FVI, Forest Vegetation Inventory; SVI, Satellite Vegetation Inventory.

a) Phase 1

SVI forest type	FVI forest type			Marginal proportion	Overall accuracy
	Broadleaf	Conifer	Mixedwood		
Broadleaf	23,496	989	2,548	0.068	0.82
Conifer	17,547	634,472	45,271	0.801	
Mixed	73,071	154,278	61,906	0.131	

b) Phase 2

SVI forest type	FVI forest type			Marginal proportion	Overall accuracy
	Broadleaf	Conifer	Mixedwood		
Broadleaf	31	17	14	0.096	0.89
Conifer	1,297	310,164	7,154	0.861	
Mixed	383	14,847	2,360	0.043	

S5. Conclusions

From all of the analyses presented, it appears that the SVI rasters provide reasonable estimates of all 6 forest attributes given the uncertainties associated with the remote sensing data, the propagation of errors at successive modelling stages, and the paucity of forest inventory ground data in the region. In particular, the results corroborate the well-known limitations of the k-NN algorithm, which tends to narrow the predictive range for all attributes, leading to an overestimation of low values and underestimation of high values.

The probability-based design of the NFI survey allows us to affirm with great (95%) confidence that stand height was underestimated by at most an average of 4 m for the 30-m forest cells, but probably by less than a metre (given the FVI comparison). Also, the mean pixel-wise absolute error was at most 4 m, or up to 26% of the mean stand height in Phase 1, and 24% of the mean height in Phase 2. Overall accuracy for crown closure was better than 50% when based on 3 density classes (and much better for the most common density class, "open"), with the best results being observed in Phase 1 (70% accuracy). The mean pixel-wise absolute error for stand volume, total volume, and AGB in Phase 1 was at worst (upper CI95 limit) 77%, 77%, and 65% of the mean value of these 3 variables, respectively, but it could be as good (lower CI95) as 36%, 38%, and 34%, respectively. In Phase 2, however, volume and AGB estimates are less reliable because of difficulties in accurately predicting low values with GLAS. For example, the mean pixel-wise absolute errors for stand volume, total volume, and AGB were at most 16 m^3/ha, 33 m^3/ha, and 22 t/ha, respectively; however, in percentages relative to the mean value of these 3 variables in Phase 2, these become 187%, 184%, and 125%, respectively.

The analysis of unused valid GLAS footprints provided a good understanding of how the k-NN imputation influences the estimation of forest attributes. The imputation does not seem to introduce bias to the estimates; however, it does lead to underestimation of tall and/or dense stands and overestimation of short and/or open stands. The mean absolute differences between k-NN-imputed and GLAS-derived values for stand height, crown closure, stand volume, total volume, and AGB were 19%, 10%, 37%, 31%, and 30% of their mean values in Phase 1, respectively, and 17%, 12%, 32%, 27%, and 22% of their mean values in Phase 2, respectively. The greater difference

observed in volume and AGB suggests that perhaps there is a need for other k-NN feature variables that are better related to these attributes. k-NN performed worse in broadleaf stands, which may be due to (i) potential leaf-off conditions at the time of GLAS acquisition resulting in lower height estimation by GLAS in these forest types; (ii) the low representation of GLAS footprints in broadleaf stands used in the imputation (n = 257 vs. n = 4580 for conifer when phases 1 and 2 were combined); (iii) the low occurrence of that stand type in the study area (< 7%); and (iv) the higher volume and AGB in these stand types, which would have worsened k-NN underestimation because of spectral and backscatter saturation above 100–150 t/ha. Most of the footprints where AGB estimates differed greatly (> 100 t/ha difference) corresponded to areas of heterogeneous forest type or near the border of a disturbance (i.e., road, clear-cut, water body). This may point out limitations of the k-NN imputation near ecological boundaries or in complex forest stands.

The boreal transect and GNWT ALS datasets provided good insights about the spatial factors affecting the accuracy of SVI estimates. As with the GLAS analysis, SVI accuracy was poorer in broadleaf stands, with higher underestimation than in other stand types, which again may be related to leaf-off conditions during GLAS acquisition, the use of substandard ALS data for model development, or poor transferability of Fort Simpson models to alternate ALS datasets. Heterogeneity of landcover affects SVI estimates, where it is expected that SVI performs best in homogenous stands of even height and/or crown closure. Estimates in rough terrain are not reliable because of the lack of valid GLAS data in those areas, as well as greater variability in the k-NN feature variables in steep areas. The importance of using datasets of similar vintage was also highlighted in areas subject to stand-replacing disturbances, where such disturbances may not be present in all input datasets.

The SVI estimates for reporting-units larger than a 30-m cell (i.e., in 150-m cells, or in FVI polygons) were on average closer to the reference values, but the predictive range was also further narrowed, which may be adequate for large homogenous stands but has negative implications in complex multi-story stands or in areas consisting of a mosaic of small stands different from each other.

Overall, the SVI estimates were reasonable for both stand height and crown closure in open, conifer stands, which occupy almost 60% of the forested area within the study area. Estimates of all forest attributes were within many reported estimates predicted using ALS data [41]. The poorest estimates were for dense broadleaf stands; however, they cover < 5% of the forested area within the study area. There was high variability in volume and AGB estimates, with decreasing accuracy in low-AGB stands. Although the bias and mean absolute error were lower in Phase 2 than in Phase 1, the relative errors were in general larger, and the agreement with the reference datasets was spatially inconsistent. Error in Phase 2 is probably attributable to the lower mean stand height found in Phase 2, which poses challenges for GLAS height estimation. However, the latitudinal trend analysis revealed that contrary to expectations, mean error, as well as mean height, decreased significantly with increasing latitude.

Finally, we note that our error estimates are conservative (i.e., our error estimates are probably larger than the actual error) because there are other factors contributing to the observed differences between the SVI estimates and the reference data used in our assessments (e.g., temporal and spatial mismatches), which will tend to inflate the differences. The limitations outlined above highlight the need for improved forest structure attribute models in less represented and/or productive forest stand types. Our results indicate that estimation error will probably be reduced by exploring better suited surrogate plots than those from GLAS, especially for short open stands found in the more northern reaches of Taiga Plains. Other scaling up techniques, such as random forest or convolutional neural networks, appear promising as an alternative to the traditional k-NN imputation. Between 2016 and 2019, three ALS surveys were conducted with multispectral LiDAR over a large extent of the study area and covering roughly 200 ground plots that were also remeasured in that period. These could be used to develop new ALS models with which to populate a large number of surrogate plots along those new transects, which we believe can lead to a more accurate, updated Multisource Vegetation Inventory for the Taiga Plains Ecozone.

www.mdpi.com/article/10.3390/rs14051108/s1

APPENDICES

APPENDIX SA1. Comparison of point cloud metrics derived from the Fort Simpson ALS and the boreal transect ALS

Table SA1.1 Descriptive statistics (n, number of observations; Min, minimum value across n observations; Max, maximum; Mean; Median; SD, standard deviation; CV, coefficient of variation) between the Fort Simpson (FS) and boreal transect (BT) airborne laser scanning (ALS) datasets for 4 variables (95% percentile of points 2 m aboveground, stand height, cumulative projected foliage area index (Lz), and crown closure) and their signed (Δ) and absolute (|Δ|) differences. Only forested cells with a total point count > 100, slope < 5 and valid crown closure values were used for comparison. The last 5 rows are the statistics for the predicted (x, i.e., BT) vs. observed (y, i.e., FS) linear model (y = a + bx): RMSE, root mean square error; Adj. R2, adjusted coefficient of determination; slope, b; intercept, a; BP p-value, p-value of the Breusch-Pagan test (if < 0.05 then heteroscedasticity is assumed).

	95th percentile (m)				Stand height (m)				Lz (unitless)				Crown closure (%)			
	FS	BT	Δ	|Δ|	FS	BT	Δ	|Δ|	FS	BT	Δ	|Δ|	FS	BT	Δ	|Δ|
n	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103	23103
Min	2.20	2.19	–6.00	0	2.64	2.63	–5.76	0	0	0	–1.77	0	13.31	15.01	–48.94	0
Max	25.88	26.71	11.48	11.48	25.37	26.17	11.02	11.02	2.12	0.42	0.24	1.77	75.76	50.69	22.47	48.94
Mean	10.18	10.26	0.08	0.88	10.31	10.38	0.07	0.84	0.18	0.17	–0.01	0.09	37.40	38.17	0.77	5.71
Median	9.81	9.46	0.03	0.67	9.95	9.61	0.03	0.65	0.13	0.18	0.02	0.06	37.37	40.67	2.07	5.00
SD	4.50	4.51	1.22	0.85	4.32	4.33	1.17	0.81	0.17	0.10	0.14	0.11	9.48	7.82	7.11	4.31
CV	0.44	0.44	15.82	0.96	0.42	0.42	15.82	0.96	0.98	0.59	–14.87	1.16	0.25	0.20	9.23	0.76
RMSE	1.20				1.16				0.14				6.97			
Adj. R^2	0.93				0.93				0.36				0.46			
Slope	0.96				0.96				1.05				0.82			
Intercept	0.33				0.34				0.00				6.02			
BP p-value	0.12				0.12				0.00				0.00			

www.mdpi.com/article/10.3390/rs14051108/s1

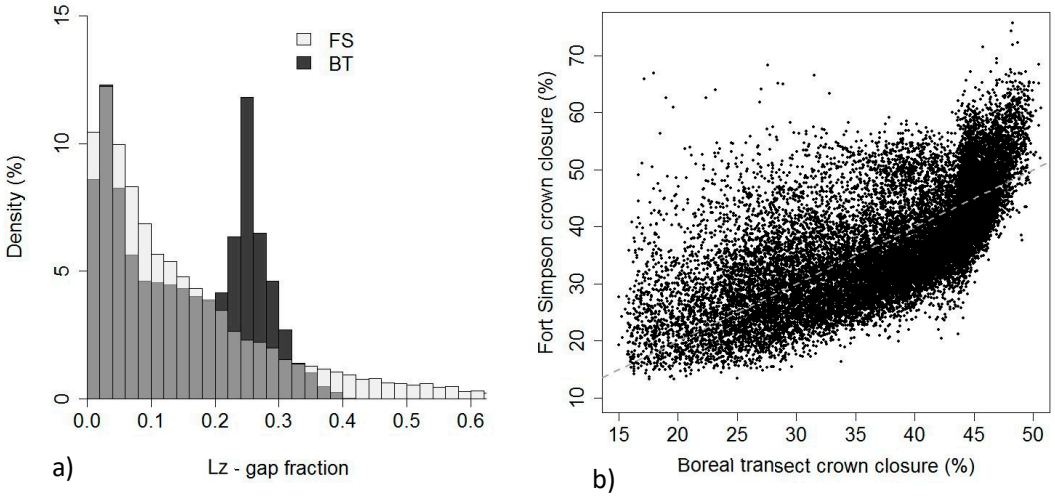


Figure S4.1 (a) Histogram showing the distribution of cumulative projected foliage area index (Lz) values in the geographic area of overlap between the Fort Simpson (FS) and boreal transect (BT) airborne laser scanning (ALS) surveys, and (b) the relationship between FS and BT crown closure in the area of overlap.

APPENDIX SA2. Accuracy assessment of the landcover maps[17]

SA2.1. Phase 1 accuracy assessment

Phase 1 agreement was assessed using 2 GNWT forest vegetation inventories (FVI; Jean Marie River and Axe Point) based on a subset of the EOSD level 4 classification schema, namely only those polygons of broadleaf, conifer and mixedwood forest type classes. All forested FVI polygons greater than 2 ha and having at least 50% coverage with forested EOSD cells were used for analysis. For each candidate FVI polygon the EOSD equivalent was considered pure (broadleaf B, or conifer C) if at least 75% of the forested 30-m EOSD pixels within were B or C, otherwise it was set to mixedwood (M). The overall accuracy for the 3 forest classes was 84% (Table SA2.1).

Table SA2.1. Contingency matrix of Phase 1 classification with existing Forest Vegetation Inventory (FVI) data for conifer, broadleaf, and mixed forest types.

Map prediction	**Validation reference (no. of FVI polygons)**				**User accuracy**
	Broadleaf	*Conifer*	*Mixedwood*	**Total**	
Broadleaf	**23,496**	989	2,548	27,033	87%
Conifer	17,547	**634,472**	45,271	697,290	91%
Mixedwood	73,071	154,278	**61,906**	289,255	21%
Total	114,114	789,739	109,725	1,013,578	
Producer accuracy	21%	80%	56%		

Note: overall weighted accuracy (proportional to area): 83.9%.

SA2.2 Phase 2 accuracy assessment (level 4)

Phase 2 agreement with independent validation locations was assessed through a 25% stratified random sample of landcover observations from a spatially extensive network of northern boreal vegetation type and density conditions. A total of 954 observations were retained for validation. Following Wulder et al. [83] in the use of vector polygons for accuracy assessment of EOSD-LC, accuracy was assessed at the polygon level where a generalized landcover dataset was compared with field observations. Specifically, the mode (i.e., majority) of all the landcover pixels found within the polygon in which the sample unit was found was compared with validation observations. This generalization effectively matches the level of generality associated with the NFI. The original EOSD-LC schema was generalized to vegetation type classes without density subclasses (level 4; 9 classes). For forest inventory purposes the level 5 wetland treed class was assigned to the generalized level 4 "conifer" class instead of the "wetland" class. Overall agreement with field data improved from 51% (circa 2000 EOSD-LC) to 71% using the same independent data for validation (Table SA2.2).

[17] Natural Resources Canada and Government of Northwest Territories. Monitoring forests in the Northwest Territories – Earth Observation for Sustainable Development of Forests (EOSD) landcover updates Phase 1 and 2 of the Taiga Plains Ecozone. 2017. Data access through the Data Coordinator, NWT Centre for Geomatics: gnwtmaps_admin@gov.nt.ca

Table SA2.2. Contingency matrix of Phase 2 classification (level 4).

Map prediction	Validation reference (no. of polygons)										User accuracy
	Water	*Exposed*	*Bryoids*	*Shrub*	*Wetland*	*Herb*	*Conifer*	*Broadleaf*	*Mixedwood*	**Total**	
Water	**67**	0	0	0	8	0	0	0	0	75	89%
Exposed	0	**60**	0	1	4	0	0	1	0	66	91%
Bryoids	0	0	**44**	1	0	0	2	0	0	47	94%
Shrub	0	1	3	**97**	30	3	20	5	1	160	61%
Wetland	0	0	0	6	**64**	1	30	3	1	105	61%
Herb	0	0	0	0	2	**2**	0	1	0	5	40%
Conifer	0	0	0	15	44	1	**272**	25	22	379	72%
Broadleaf	0	1	0	3	6	0	9	**29**	1	49	59%
Mixedwood	1	0	0	3	3	0	15	2	**23**	47	49%
Unknown	1	7	2	3	0	5	3	0	0	21	
Total	69	69	49	129	161	12	351	66	48	954	
Producer accuracy	97%	87%	90%	75%	40%	17%	77%	44%	48%		

Note: overall weighted accuracy (proportional to area): 70.9%.

To directly compare the agreement of Phase 1 and Phase 2 with validation data, the level 4 broadleaf, conifer and mixedwood classes were retained to also generate a contingency matrix using the treed-only classes of level 4. Overall accuracies between Phase 1 and Phase 2 were comparable (84% and 83%, respectively; Tables SA2.1 and SA2.3). For both phases the conifer class was classified with the highest producer (80%–92%) and user (85%–91%) accuracies among the generalized level 4 treed classes. Overall agreement with field data improved from 78% (circa 2000 EOSD-LC) to 83% using the same independent data for validation (Table SA2.3).

Table SA2.3. Contingency matrix of Phase 2 classification for conifer, broadleaf, and mixedwood forest types.

Map prediction	Validation reference (no. of polygons)				User accuracy
	Broadleaf	*Conifer*	*Mixedwood*	**Total**	
Broadleaf	**29**	9	1	39	74%
Conifer	25	272	22	319	85%
Mixedwood	2	15	23	40	58%
Total	56	296	46	398	
Producer accuracy	52%	92%	50%		

Note: overall weighted accuracy (proportional to area): 83.4%.

SA2.3. Phase 2 accuracy assessment (level 2)

Phase 2 agreement with independent validation locations was determined through a 25% stratified random sample of landcover observations, using a method similar to that described in Section SA.2.2. The original EOSD-LC schema was generalized to broad landcover type classes (level 2; 2 classes). For forest inventory purposes all level 5 treed classes were assigned to the generalized level 2 "treed" class, with all other classes assigned to level 2 "non-treed". Overall agreement with field data improved from 74% (circa 2000 EOSD-LC) to 86% using the same independent data for validation (Table SA2.4).

Table SA2.4. Contingency matrix of Phase 2 classification (level 2).

Map prediction	Validation reference (no. of polygons)			User accuracy
	Nontreed	*Treed*	**Total**	
Nontreed	**397**	61	458	87%
Treed	74	**401**	475	84%
Unknown	18	3	21	
Total	489	465	954	
Producer accuracy	81%	86%		

Note: overall weighted accuracy (proportional to area): 85.5%.

SA2.4. Phase 2 Accuracy Assessment (level 5)

Phase 2 agreement with independent validation locations was assessed through a 25% stratified random sample of landcover observations using a method similar to that described in Section SA.2.2 albeit for the original EOSD-LC class schema (vegetation type classes with density subclasses). Overall agreement with field data improved from 28% (circa 2000 EOSD-LC) to 50% using the same independent data for validation (Table SA2.5).

Table SA2.5. Contingency matrix of Phase 2 classification (level 5).

Map prediction	Validation reference (no. of polygons)																		User accuracy (%)
	Water	*Rock*	*Exp*	*Bryd*	*S-T*	*S-L*	*WT-T*	*WT-S*	*WT-H*	*Herb*	*Con-D*	*Con-O*	*Con-S*	*Brl-D*	*Brl-O*	*Mix-D*	*Mix-O*	**Total**	
Water	**67**	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	75	89
Rock	0	**20**	5	0	1	0	0	0	1	0	0	0	0	0	1	0	0	28	71
Exp	0	0	**35**	0	0	0	0	0	3	0	0	0	0	0	0	0	0	38	92
Bryd	0	0	0	**44**	0	2	0	0	0	0	0	0	2	0	0	0	0	48	92
S-T	0	0	0	0	**15**	3	1	1	1	0	0	0	1	1	3	0	0	26	58
S-L	0	0	1	3	13	**67**	8	20	9	3	1	1	8	1	1	0	1	137	49
WT-T	0	0	0	0	4	4	**21**	4	4	1	3	12	12	7	2	0	2	76	28
WT-S	0	0	0	0	1	3	12	**32**	4	0	1	3	4	3	0	0	1	64	50
WT-H	0	0	0	0	1	0	3	4	**27**	1	6	3	0	0	0	0	0	45	60
Herb	0	0	0	0	0	0	0	2	1	**2**	0	0	0	1	0	0	0	6	33
Con-D	0	0	0	0	0	0	4	1	3	0	**29**	14	0	0	0	4	0	55	53
Con-O	0	0	0	0	1	4	11	5	10	0	34	**44**	7	10	2	5	5	138	32
Con-S	0	0	0	0	0	1	17	6	7	0	1	24	**33**	1	2	2	2	96	34
Brl-D	0	1	0	0	1	0	0	2	1	0	2	4	1	**18**	3	0	1	34	53
Brl-O	0	0	0	0	2	0	0	1	1	0	0	1	2	4	**4**	0	1	16	25
Mix-D	1	0	0	0	0	0	2	0	3	0	5	1	0	1	0	**9**	4	26	35
Mix-O	0	0	0	0	3	0	3	0	0	0	2	4	1	1	0	1	**10**	25	40
Unknown	1	6	1	2	0	3	0	0	0	5	0	0	3	0	0	0	0	21	
Total	69	27	42	49	42	87	82	78	83	12	84	111	74	48	18	21	27	954	
Producer accuracy (%)	97	74	83	90	36	77	26	41	33	17	35	40	45	38	22	43	37		

Note: overall weighted accuracy (proportional to area): 49.5%.
Landcover classes: Exposed (Exp), Bryoids (Bryd), Shrub Tall (S-T), Shrub Low (S-L), Wetland Treed (WT-T), Wetland Shrub (WT-S), Wetland Herb (WT-H), Conifer Dense (Con-D), Conifer Open (Con-O), Conifer Sparse (Con-S), Broadleaf Dense (Brl-D), Broadleaf Open (Brl-O), Mixedwood Dense (Mix-D), Mixedwood Open (Mix-O).

References (numbering continues from main manuscript)

80. Natural Resources Canada. *Canada's National Forest Inventory Ground Sampling Guidelines, Version 5.0*. Natural Resources Canada, Canadian Forest Service, **2008**. Available online: https://nfi.nfis.org/en/ground_plot (accessed on 27 November 2020).
81. McRoberts, R.E. Diagnostic tools for nearest neighbors techniques when used with satellite imagery. *Remote Sens. Environ.* **2009**, 113, 489–499. https://doi.org/10.1016/j.rse.2008.06.015.
82. Lambert, M.-C.; Ung, C.-H.; Raulier, F. Canadian national tree aboveground biomass equations. *Can. J. For. Res.* **2005**, 35, 1996–2018. https://doi.org/10.1139/x05-112.
83. Wulder, M.A.; White, J.C.; Luther, J.E.; Strickland, G.; Remmel, T.K.; Mitchell, S.W. Use of vector polygons for the accuracy assessment of pixel-based landcover maps. *Can. J. Remote Sens.* **2006**, 32, 268–279.