Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:96
Faculty of Forest Sciences

Forest Attribute Prediction and Mapping using 3D Remote Sensing Data

Ritwika Mukhopadhyay

Forest Attribute Prediction and Mapping using 3D Remote Sensing Data

Ritwika Mukhopadhyay
Faculty of Forest Sciences
Department of Forest Resource Management
Umeå

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Umeå 2024

Acta Universitatis Agriculturae Sueciae
2024:96

Cover: Boreal forests, Sweden.
(photo: Ritwika Mukhopadhyay)

ISSN 1652-6880
ISBN (print version) 978-91-8046-423-9
ISBN (electronic version) 978-91-8046-431-4
https://doi.org/10.54612/a.7q6n063f5u

Swedish University of Agricultural Sciences, Department of Forest Resource Management, Umeå, Sweden

Print: SLU Grafisk service, Uppsala 2024

Forest Attribute Prediction and Mapping using 3D Remote Sensing Data

Abstract

Forest inventory enables collection of essential data on forest attributes such as volume (VOL), aboveground biomass (AGB), species composition, age, and forest health. Knowledge about these attributes are vital for strategic and tactical forest management purposes, including planning timber harvests, conserving biodiversity, estimating carbon sequestration, and forecasting future yields. Forest inventory practices have evolved significantly over the past century along with the development of remote sensing (RS) assisted inventory approaches. This thesis focuses on using 3D RS data acquired from different platforms and with different remote sensors, for example - airborne laser scanning (ALS), digital aerial photogrammetry, and synthetic aperture radar. The individual papers focused on different forest regions and different spatial extents of acquired RS data for the prediction, estimation, and mapping of forest attributes such as, VOL and AGB, for various cases of model-based inference. The included papers have shown that, 3D RS data can be successfully integrated as auxiliary data and reference data within model-based inference frameworks. A combination of dense and sparse ALS data can be used for forecasting forest VOL growth through VOL models. Several methods are also employed in the individual papers to quantify uncertainty, including root mean square error, confidence intervals, and prediction intervals. Overall, this thesis concludes that 3D RS data is efficient for accurate forest attribute prediction, supporting cost-effective forest monitoring and management solutions. The integration of RS data into forest inventory practices continues to evolve, offering new opportunities for large-scale forest monitoring, resource management, and biodiversity conservation.

Keywords: aboveground biomass, airborne laser scanning, digital aerial photogrammetry, forest inventory, synthetic aperture radar, uncertainty, volume, 3D remote sensing.

Forest Attribute Prediction and Mapping using 3D Remote Sensing Data

Abstract

Skogsinventering möjliggör insamling av viktig data om skogsattribut som volym (VOL), biomassa ovan mark (AGB), artsammansättning, ålder och skogshälsa. Kunskap om dessa attribut är viktiga för strategisk och taktisk skogsförvaltning, inklusive planering av avverkning, bevarande av biologisk mångfald, skattning av koldioxidinlagring och prognoser för framtida avkastning. Skogsinventeringspraxis har utvecklats avsevärt under det senaste århundradet i takt med utvecklingen av fjärranalys (RS)-assisterade inventeringsmetoder. Denna avhandling fokuserar på användningen av 3D RS-data som samlats in från olika sensorer, till exempel laserskanning (ALS), fotogrammetri och syntetisk aperturradar, och från olika plattformar. De enskilda artiklarna fokuserade på olika skogsområden och olika rumsliga utsträckningar av insamlade RS-data för att förutsäga, skatta och kartlägga skogsattribut såsom VOL och AGB i olika fall av modellbaserad inferens. De inkluderade artiklarna har visat att 3D RS-data framgångsrikt kan integreras som hjälpdata i modellbaserade inferensramverk. Kombinationen av tät och gles ALS-data kan användas för att prognostisera skogsvolymtillväxt genom volymmodeller. Flera metoder har också använts i de enskilda artiklarna för att kvantifiera osäkerhet, inklusive medelkvadratfel, konfidensintervall och prediktionsintervall. Sammanfattningsvis konstaterar denna avhandling att 3D RS-data är effektiva för att skatta olika skogliga attribut, vilket stöder kostnadseffektiva lösningar för skogsövervakning och förvaltning. Integrationen av RS-data i skogsinventeringsmetoder fortsätter att utvecklas och erbjuder nya möjligheter för storskalig skogsövervakning, resursförvaltning och bevarande av biologisk mångfald.

Nyckelord: biomassa, luftburen laserskanning, fotogrammetri, skogsinventering, syntetisk aperturradar, osäkerhet, volym, 3D fjärranalys.

Dedication

To all my well-wishers.

"*When you come out of the storm, you won't be the same person who walked in. That's what this storm's all about.*"

~ Haruki Murakami, *Kafka on the Shore.*

Contents

List of publications

This thesis is based on the work in the following papers, referred to by Roman numerals in the text:

I. Mukhopadhyay R., Næsset E., Gobakken T., Mienna I.M., Bielza J.C., Austrheim G., Persson H.J., Ørka H.O., Roald B-E., Bollandsås O.M. (2023). Mapping and Estimating Aboveground Biomass in an Alpine Treeline Ecotone under Model-Based Inference. *Remote Sensing*, 15 (14), 3508. https://doi.org/10.3390/rs15143508

II. Mukhopadhyay R., Ekström M., Lindberg E., Persson H.J., Saarela S., Nilsson M. (2024). Computation of prediction intervals for forest aboveground biomass predictions using generalized linear models in a large-extent boreal forest region. *Forestry: An International Journal of Forest Research*, cpae006, 1-11. https://doi.org/10.1093/forestry/cpae006

III. Mukhopadhyay R., Nilsson M., Ekström M., Lindberg E., Persson H.J. (2023). Comparing TanDEM-X InSAR Forest Stand Volume Prediction Models Trained Using Field and ALS Data. *Proceedings of IGARSS 2023 - IEEE International Geoscience and Remote Sensing Symposium proceedings*, 3253-3256. ISBN: 979-8-3503-3174-5, eISBN: 979-8-3503-2010-7

IV. Mukhopadhyay R., Lindberg E., Olofsson K., Holmgren J., Nilsson M., Ståhl G., Ekström M., Persson H.J. Large scale forest inventory using digital methods. (manuscript)

V. Mukhopadhyay R., Lindberg E., Wallerman J., Mensah A.A., Nilsson M., Persson H.J. Forest Volume Growth Forecasting with Fusion of Dense and Sparse Airborne Laser Scanning Data. (manuscript)

Papers I-III are reproduced with the permission of the publishers.

The contribution of Ritwika Mukhopadhyay to the papers included in this thesis was as follows:

I. Contributed in the formal analysis, data visualisation, writing a major part of the manuscript with the last co-author, reviewing and editing the manuscript.

II. Contributed in data analysis and visualisation, writing a major part of the manuscript, reviewing and editing the manuscript.

III. Planned the study with the main supervisor, carried out data analysis for ALS data, writing a major part of the manuscript, reviewing and editing the manuscript.

IV. Planned the study with the main supervisor and co-authors, carried out data analysis, writing a minor part of the manuscript, reviewing and editing the manuscript.

V. Developed the study with the main supervisor and co-authors, carried out data analysis, writing a major part of the manuscript, reviewing and editing the manuscript.

List of tables

List of figures

Abbreviations

3D	Three-dimensional
ABA	Area-based approach
AGB	Aboveground biomass
ALS	Airborne laser scanning
CI	Confidence interval
DAP	Digital aerial photogrammetry
DBH	Diameter at breast height
DEM	Digital elevation model
DTM	Digital terrain model
FI	Forest inventory
GCP	Ground control points
GLM	Generalized linear model
InSAR	Interferometric synthetic aperture radar
ITC	Individual tree crown
k-MSN	k-most similar neighbour
k-NN	k-nearest neighbour
LD	Low density
LiDAR	Light detection and ranging
LS	Laser scanning

LUT	Look-up table
MLR	Multiple linear regression
MLS	Mobile laser scanning
NFI	National forest inventory
NLS	National laser scanning
NN	Nearest neighbour
PI	Prediction interval
RMSE	Root mean square error
RS	Remote sensing
SAC	Spatial auto-correlation
SAR	Synthetic aperture radar
SCA	Svenska Cellulosa Aktiebolaget
SHORT	Short vegetation
SLU	Swedish University of Agricultural Sciences
sqrt	Square root
SRS	Simple random sampling
TALL	Tall vegetation
TLS	Terrestrial laser scanning
UAV	Unmanned aerial vehicle
VHD	Very high density
VHR	Very high resolution
VOL	Volume

1. Introduction

1.1 Forests in Sweden

The land covered by forests in Scandinavia is majorly boreal forest in the mainlands and alpine forest towards the mountainous regions. The boreal forests in Sweden are mostly dominated by Scots pine (*Pinus sylvestris* L., 39.9%) and Norway spruce (*Picea abies* (L.) Karst., 39%), and rest covered with broadleaf and deciduous trees (Skogsdata 2024, 2024). Boreal forests act as the largest terrestrial carbon sink by covering almost 30% of the global forest area (Högberg, 2021). In Sweden, approximately 69% of the land area is covered by forests, and 58% of the toal land area is covered by productive forests (Roberge et al., 2023). In Sweden, 70% of the forests are under private ownership including private forest owners and companies (*Skogsdata 2024*, 2024). The management practices have been developed over the years aiming towards sustainability, i.e., the goal to achieve forest resources being used efficiently without compromising their ecological integrity. Sustainable forest management practices have been of importance, as forests act as carbon sinks.. Reliable forest data are necessary for fulfilling international reporting requirements and demonstrating compliance with emission reduction commitments, such as the Kyoto Protocol. Boreal forests are home to a wide variety of flora and fauna, some of which are endemic or endangered, which makes assessing and monitoring of such areas important. As per the commitment towards sustainable forestry practices, a comprehensive forest inventory (FI) helps ensure that annual total volume (VOL) loss due to cuttings and mortality should not exceed the total growth.

1.2 Forest inventory

Accurate measurements from FIs of the size, species composition, age, and health of forest stands, enable foresters to make informed decisions about when and how much to harvest, ensuring the long-term health and productivity of the forest ecosystem. The wood-based forest industries include - timber, pulp, and paper production, all of which rely on accurate inventory data to optimize operations and plan for future supply chains. FI is an important component of sustainable forest management aiding in identifying critical habitats and biodiversity hotspots for conservationists to implement protective measures to preserve species and their habitats, and balancing commercial activities, providing detailed data for calculating carbon sequestration and emissions, and forecasting timber yields, assessing market conditions, and planning infrastructure investments.

Over a century, since 1923 the **national forest inventory (NFI)** has been the main source of FI data collected over entire Sweden at different scales covering gradients of management practices and climate zones aiming mainly to describe forest state and change conditions. The NFI is located at the Department of Forest Resource Management at the Swedish University of Agricultural Sciences (SLU). The NFI performs an annual inventory of around 11,000 circular field sample plots, consisting of about 40% temporary (7 m radius) and 60% permanent (10 m radius) plots distributed systematically throughout the country. Trees with a diameter at breast height of at least 10 cm are measured on plots with 7 and 10 m radius. Trees with a diameter between 4 and 10 cm are measured on plots with 3.5 m radius, and trees below 4 cm diameter are measured on two small plots (of 1 m radius). The permanent plots were introduced later in the NFI, in the year 1983, which lead to more precise estimation of temporal changes. FI of attributes, such as, **VOL** and **aboveground biomass (AGB)**, are an important source of information for a variety of strategic, operational, and tactical forest management purposes. However, it is not practical for manual field inventories to be conducted contiguously across large areas.

1.3 Statistical inference

Design-based and model-based inferences are the two main statistical frameworks that has been used in forest survey. The NFI surveys use **design-based inference** for acquiring probability sample plots in a systematic

manner over Sweden (Fridman et al., 2014). The design-based inference does not assume any distribution for the population of the forest attribute of interest, unlike a model-based inference.

A **model-based inference** used within forest survey presented by Matérn (1960), is an important example. It assumes an underlying distribution of the population or realization that is drawn from a super-population, i.e., a hypothetical population of which the reference dataset being observed is one realization. Model parameters are estimated and variables of interest are predicted based on the assumed underlying distribution and relationship between the data and the variable of interest (Brewer, 1963; Royall, 1970; Cassel et al., 1977; Gregoire, 1998).

A special case of model-based inference has been implemented in forest attribute prediction and estimations, when the model- and design-based inference features are combined and has been referred to as a **hybrid inference** framework (Corona et al., 2014; Ståhl et al., 2011a, 2016).

1.4 Remote sensing data in forest inventory

Model-based inference integrating remote sensing (RS) data have been and are being developed and applied to predict and map forest attributes, with the goal of providing an accurate, spatially continuous, and detailed information base for practitioners of forestry and ecosystem management. RS has played an important role over the last decades to acquire more information with fewer field sample plots and has been of interest to integrate RS data within FI practices ever since.

The use of RS as auxiliary data in model-based inference was implemented for prediction of target variables by developing a link between the field reference target variable and information extracted from the RS data. Forest variables such as AGB and VOL are related to tree height which can be directly derived as three-dimensional (3D) information from remote sensors. Various RS data for extraction of 3D information have been tested for decades within model-based inference such as, **digital aerial photogrammetry (DAP)**, **laser scanning (LS)**, and **synthetic aperture radar (SAR)**.

LS (light detection and ranging (LiDAR)) is a source of active RS data. It measures the distance between the sensor and the object based on the time recorded between the emitted and reflected light pulse to reach the sensor.

Airborne LS (ALS) has been extensively used for forest mapping purposes, whereas, **terrestrial** and **mobile LS** (**TLS** and **MLS**, respectively) have been mostly used for research and construction industry. Data from ALS are stored in two different ways: 1) Discrete returns or echoes captured as point clouds with each individual point or return having a 3D coordinate in space, and 2) Full waveform data captured as a profile of all the returns for each emitted pulse (Mallet & Bretar, 2009). ALS can be used to measure the height of vegetation from the top returns, the multiple returns, and the ground returns from the reflected pulses.

DAP is based on image matching of aerial photographs, using the principle of stereo photogrammetry for deriving relative position of an object in the vertical plane (i.e., height of vegetation) from overlapping images by defining triangulated lines of sight (Kraus, 2007; Rahlf, 2017). Automated matching of corresponding points from overlapping images are done using various image matching algorithms (Haala et al., 2014). Forest parameters estimated from DAP combined with an accurate digital terrain model (DTM) was observed by Bohlin et al. (2012), Nurminen et al. (2013), and Persson (2016), to be comparably accurate to estimates from ALS data (Rahlf, 2017).

SAR is often acquired from spaceborne platforms and is yet another RS data source from which 3D information of forest heights can be derived. Height information can be derived using interferometric SAR (InSAR) or radargrammetry, two 3D reconstruction techniques using SAR, and they have been used for modelling and prediction of forest variables such as height and AGB (Solberg, Astrup, Bollandsås, et al., 2010; Solberg, Astrup, Gobakken, et al., 2010; Persson & Fransson, 2014; Soja et al., 2015b, 2015a; Solberg et al., 2013, 2015; Persson & Fransson, 2017). SAR data has gained importance in forest variable prediction and mapping due to being an active sensor able to penetrate clouds and overcome light constraints, providing large area coverage when acquired as satellite images, and their frequent acquisitions enabling multitemporal analysis. Phase height derived from the phase difference of two slightly differently positioned SAR images has been used for predicting vegetation height and other correlated attributes such as, VOL and AGB. Interferometry is also used for generation of digital elevation models (DEMs) (Balzter, 2001).

RS data can be implemented as auxiliary data in different approaches, where the **area-based approach (ABA)** and **individual tree crown (ITC) approach** are commonly used in model-based inference.

1.4.1 Area-based approach

ABA for forest attribute prediction is based on the statistical relation of field measured reference data of field sampling units (multiple trees and other vegetation) and the independent variables extracted from the RS auxiliary data from the corresponding areas. The size of the field sampling unit is assumed to agree with the size of the area where the RS metrics are averaged. AGB and VOL can be determined for example, by using the heights derived from ALS data (Kellndorfer et al., 2010). Initial studies using an ABA, for mean height and VOL predictions using laser returns, were, Nilsson (1996) and Næsset (1997a, 1997b). Næsset (2002) estimated various forest stand characteristics - mean tree height, dominant height, mean diameter, stem number, basal area, and timber VOL using an ABA for regression analysis of canopy height and density metrics derived from small-footprint laser scanning across young and mature forest stands. The proposed two-stage method successfully predicted stand-level characteristics with most predictions being unbiased, and reasonable accuracy was achieved, as indicated by standard deviations of the differences between predicted and ground-truth values. It has been shown that forest attributes in boreal and alpine ecosystems can be predicted from ALS data using linear regression models (Næsset, 2004; Hollaus et al., 2009; Ståhl et al., 2011b; Bohlin et al., 2011; Vastaranta et al., 2011; Lindberg, 2012; Saarela, 2015; Nilsson et al., 2017), non-linear regression models (Saarela et al., 2016, 2018, 2020), non-parametric approaches such as, nearest neighbour (NN), k-nearest neighbours (k-NN), and k-most similar neighbour (k-MSN) (Maltamo et al., 2006; Packalén & Maltamo, 2007), and machine learning algorithms (Esteban et al., 2019, 2020). Model-based prediction using ALS data in an ABA has been studied for alpine ecosystems (Nyström et al., 2012, 2013; Næsset et al., 2019; Noordermeer et al., 2023), and similarly the use of DAP has been assessed (Naesset et al., 2021). A similar concept of large-area model-based predictions using an ABA and mapping of various forest attributes has also been tested with space-borne acquisitions of SAR images, example for forest height (Persson et al., 2013; Soja et al., 2015b; Persson & Fransson, 2017), VOL (Persson et al., 2017), and AGB (Persson & Fransson, 2014; Soja et al., 2015a; Persson et al., 2017). However, the use of ALS data as auxiliary RS data compared to X-band SAR data for the same scale of measurement has been proven to be more accurate for predicting AGB and VOL (Vastaranta et al., 2014; Yu et al., 2015; Persson & Fransson, 2017).

Yet, the use of readily available satellite imageries are more cost- and time-effective when compared to acquiring wall-to-wall ALS data for large areas.

1.4.2 Individual tree crown approach

Instead of working on ABAs, the use of **ITCs** allows the corresponding work to be carried out for single trees. Finding the local tree maxima in canopy height models followed by tree crown segmentation where the ITC boundaries are defined. The RS data variables are then extracted for ITC segments to be used in the prediction models.

Hyyppä & Inkinen (1999) presented the segmentation of tree crowns from an ALS canopy height model which enabled segmentation of coniferous trees with 40-50% accuracy. Single tree diameter at breast height (DBH) and ITC height are important variables commonly derived in ITC approaches, which can be used to describe stem form and estimate VOL. When working with ALS, a higher density of the ALS data has been an advantage for estimating the single tree attributes (Holmgren & Soderman, 2002; Solberg et al., 2006; Vauhkonen et al., 2012; Lindberg, 2012; Lindberg et al., 2012; Holmgren & Lindberg, 2013; Strîmbu & Strîmbu, 2015; Eysn et al., 2015; Holmgren et al., 2022), and tree species (Holmgren & Persson, 2004; Holmgren et al., 2008; Persson et al., 2012; Vauhkonen et al., 2014; Lindberg et al., 2021; Persson et al., 2022; Axelsson et al., 2023).

Both area-based and single-tree based models can be used for predicting wall-to-wall maps when the RS auxiliary data is available wall-to-wall. These maps of different forest attributes such as, VOL and AGB can be used for monitoring forest health and damage assessment, taxation of forest owners, planning of forest management practices, mapping of forest dynamics and other purposes by forest stakeholders. For cases when the auxiliary RS data is not available wall-to-wall, the RS based predictions can be used in hybrid inference to estimate forest attributes for the total/mean of an aggregated area. The VOL is an important variable for forest companies to have an idea of the total standing stock of forest, which can be used as timber for commercial purposes. The goal of a FI is to quantify the amount and type of forest resources and related attributes in a given area while the goal of forest mapping is to depict the spatial distribution of forest resources and related attributes (Corona et al., 2014).

1.5 Uncertainty estimation

Detailed and accurate information about forest attributes are important for an effective management of forests and to avoid sub-optimal decisions and poor conclusions. Yet, errors are an inherent part of data collection both in manual field inventories and from remote sensors. These errors may arise due to various factors such as, measurement errors in the instruments, mistakes such as missing trees from inventory personnel, and poor environmental conditions during data acquisitions. Errors may also arise from misspecified models and the limitations due to the statistical design. Understanding and accounting for such errors is important and they can be separated into systematic and random errors.

Systematic errors, also known as bias, are consistent errors that occur in the same direction every time. These errors may arise from flaws in the measurement process, calibration of instruments, or underlying assumptions in the model hence affecting the accuracy of the predicted and estimated values. These errors can be corrected by instrument calibration or using correction factors derived from a comparison of measured and true values from prediction models.

Whereas, **random errors** describe the inherent variability in the data that cannot be described in the model. Unlike systematic errors, random errors do not have a consistent behaviour and tend to average out over many measurements. Bias and variance estimators are statistical approaches to account for systematic and random errors, respectively.

Some uncertainty estimators follow:

1. **Bias –** a quantification of the systematic errors that hence is possible to compensate for. For example, probabilistic sampling designs help mitigate bias when measuring reference data for training models in model-based inference. By ensuring each sample has a known probability of selection, these designs provide a more representative dataset of the super-population.

2. **Root mean square error (RMSE)** – is an aggregated accuracy estimate for model predictions computed as the square-root of the average squared differences between the reference and the predicted values of a forest attribute of interest, integrating both systematic and random errors.

3. **Confidence interval (CI)** – provides a range for the estimated mean predicted values of forest attribute of interest. CI accounts for the variability due to the randomness in the estimation of the model parameters, e.g., the regression coefficients.

4. **Prediction interval (PI)** – provides a range for new predicted values of forest attribute of interest accounting for the random variability of estimating the model parameters, as well as the random variance in the prediction of the new observation, making PIs wider compared to the range of CIs.

5. **Sampling variance** – accounts for the variance due to the randomness in a sample in design-based inference from a finite population. It measures how much an estimate from one sample might differ from another sample, even if both are drawn from the same finite population. The larger the sample, the smaller the sampling variance tends to be, because larger samples provide more information and reduce the impact of randomness within the data.

2. Aims and objectives

The overall aim of this thesis was to evaluate the use of model-based inference in association with multiple 3D RS data acquired from different sensors, RS platforms, and at different spatial extents, for predicting state and growth of forest VOL and related forest attributes for large-area operational forest management.

The specific objectives in the different papers included in this thesis were:

1. To evaluate and compare the precision of AGB estimates from a model-based inference in treeline ecotone obtained using ALS and DAP as auxiliary data and to assess the possibilities of using ABA (**Paper I**).
2. To compute and compare pixel-level prediction uncertainties in the form of PIs for Gamma and Tweedie exponential dispersion models implemented within a model-based inference framework for AGB prediction using ALS data, and to compare methodologies for computation of PIs (**Paper II**).
3. To use model-based inference for prediction and mapping of stand-level forest VOL using TanDEM-X SAR images as auxiliary data across a test site and evaluating the performance of models trained with stand-level VOLs estimated from FI data and volumes estimated based on combination of Swedish national laser scanning (NLS) and NFI data, respectively (**Paper III**).
4. To demonstrate and evaluate an entirely RS-based two-phase FI approach using sampled VHR ALS data and TLS data for mean VOL estimations for a case of model-based inference when model- and design-based features are combined (i.e., hybrid inference framework) (**Paper IV**).
5. To present a large-extent RS-based forest VOL growth forecasting workflow for operational use by evaluating transferability of VOL

models using dense and sparse ALS data between two timepoints (**Paper V**).

3. Materials and methods

3.1 Materials

3.1.1 Study area

The three test sites used in this thesis were located in a treeline ecotone in Norwegian mountainous region (Paper I), and in boreal forests in south-central Sweden (Paper II), and central Sweden (Paper III, IV, and V). The site 1 (***S1***) has been used before for herbivory studies (Mienna et al., 2020, 2022). The main tree species in *S1* was mountain birch (*Betula pubescens* ssp. *czerepanovii*), along with a few individuals of rowan (*Sorbus aucuparia*), aspen (*Populus tremula*), Scots pine (*Pinus sylvestris*), and Norway spruce (*Picea abies*). The shrub species considered in this study were *Salix lapponum*, *Salix glauca* subsp. *glauca*, and *Betula nana*. Site 2, and 3 (***S2*** and ***S3***) were productive boreal forests mainly dominated by Scots pine and Norway spruce with some extent of birch and other deciduous tree species. *S2* has been an area of interest for mapping and monitoring of forests due to a significant ownership of the forest lands focusing on wood production and also due to past mining activities for copper, iron and silver ore in this region (International Model Forest Network » Bergslagen Model Forest, n.d.). *S3* is a test-site owned and managed by Svenska Cellulosa Aktiebolaget (SCA), which has been of interest for testing RS-based methodologies for frequent monitoring of the forest state and change within the test-site.

The summary and location of the sites have been presented in Table 1 and Figure 1.

Table 1. Summary of the study area

Paper	Site number	Site name	Lat, Long	Area (in ha)	Forest type
I	***S1***	Hol	60º42'N, 7º56'E	300	Treeline
II	***S2***	Bergslagen	60º0'N, 15º0'E	500,500	Boreal
III, IV, V	***S3***	Laxsjo	62º40'N, 17º0'E	50,000	Boreal

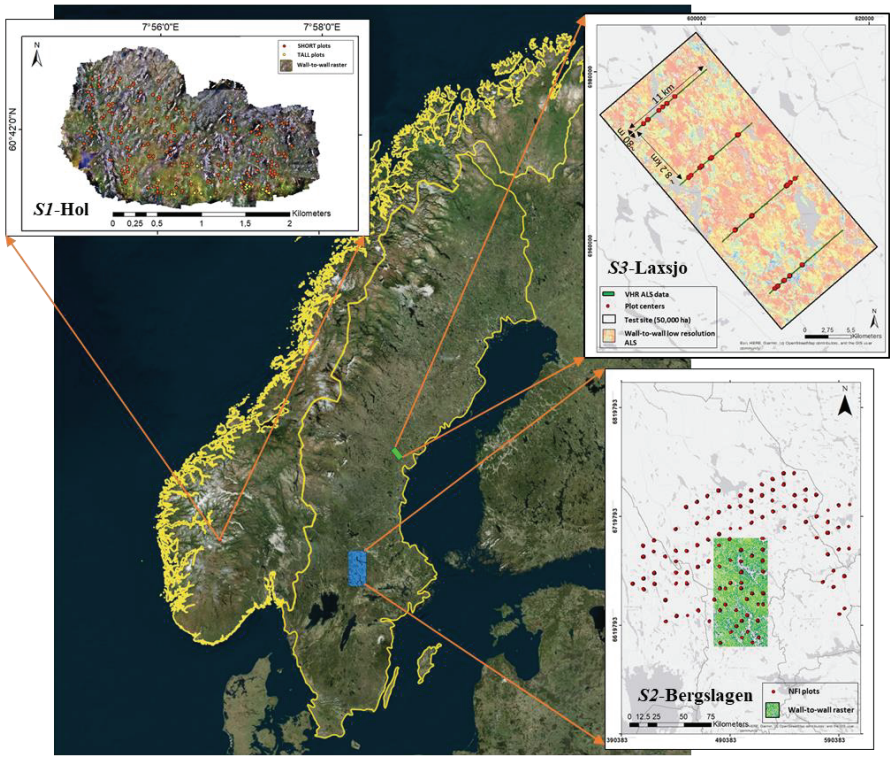


Figure 1. Study areas (*S1*, *S2*, and *S3*) used in this thesis.

3.1.2 Field inventory data

FI data were collected in the test sites following different sampling protocols. The summary of the acquired field datasets have been presented in Table 2. All trees in the sample plots in *S1*, *S2*, and *S3*, as presented in Table 1 (*Section 3.1.1*), were measured for DBH, whereas, a few sampled trees were also measured for height.

For *S1*, the FI data (***F1***) were collected by tessellating the area into 100 m^2 grid cells and the average height (h_{mean}) from the wall-to-wall ALS data was calculated for each grid cell. Grid cells were divided into 'short woody vegetation' (SHORT) ($h_{mean} < 1$ m), and 'tall vegetation' (TALL) ($h_{mean} \geq 1$ m). Twenty evenly distributed TALL grid cells were selected and circular field plots of 5.64 m radius were established in the centre of each grid cell. For the SHORT stratum, 180 circular plots of 1.5 m radius were field surveyed. For *S2*, field data (***F2***) from 500 Swedish NFI plots were used. The sample plots had a radius of 10 m and they were aggregated into clusters of 8 plots, systematically distributed across the area. For *S3*, the FI data (***F3***) was acquired by SCA where approximately 8 circular sample plots of 8 m radius were laid systematically within each of the 30 stands. The distance between the sample plots differed depending on the size of the respective stands. The field data ***F4***, was gathered in *S3* by the research team involved in MISTRA Digital Forest task 1.2. Within the 4 ALS systematic strips, 32 circular sample plots with a 10 m radius were randomly established, with approximately 8 plots per ALS strip.

The harvester data (***F5***) was acquired for 151 stands within *S3* during logging operations. The sensors installed in the harvesters measure tree DBH and length until the top cut. The tree species is selected manually by the operator, and the final height is estimated by a combination of the measured length and estimates of the top length. The volume is estimated using species-specific model functions.

Table 2. Summary of the field data used.

Data-set	Paper	Site	Acquir-ed by	Plot size (radius in m)	Number of plots/ stands	Inventory year	Use*
F1	**I**	*S1*	Research team	5.64 (TALL); 1.50 (SHORT)	20 plots (TALL); 182 plots (SHORT)	2019	T_r, T_e
F2	**II**	*S2*	Swedish NFI	10	500 plots	2009-2011	T_r
F3	**III**	*S3*	SCA	8	30 stands	2019	T_r
F4	**IV**	*S3*	Research team	10	32 plots**	2019	T_r

Data-set	Paper	Site	Acquir-ed by	Plot size (radius in m)	Number of plots/ stands	Inventory year	Use*
F5	**III**	*S3*	Harvester	—	151 stands	2019-2022	T_e

*where, 'T_r' indicates dataset used for training prediction models, and 'T_e' indicating dataset used for testing or validating prediction models.

**used for TLS scans and no manual field inventory measurements were used from these plots.

3.1.3 Remote sensing data

The different RS data acquired from various platforms and sensors were utilised for computing area-based and individual tree metrics. A summary of the RS datasets have been presented in Table 3.

Optical aerial images (***R1***) captured by an unmanned aerial vehicle (UAV) in *S1*, were acquired in parallel flight lines with 80% overlap both laterally and longitudinally. Ground control point (GCP) markers were established throughout the test site to derive the positions and orientations of the images and point cloud generation using 'lidR' R package for implementing DAP.

The Norwegian NLS data (***R2***), was collected by Terratec AS with a point density of 2 points/m² in *S1*. The point clouds were normalized and classified. All vegetation echoes without applying any threshold for returns closest to ground were included in order to include vegetation growing parallel to the ground. The Swedish NLS data (***R3a, R3b,*** and ***R3c***) were acquired wall-to-wall by the Swedish National Land Survey (Lantmäteriet, 2019), with a point density of 0.5–1 points/m², covering the years 2009–11, 2012, and 2019, respectively. Area-based metrics for the corresponding plots (*F2*) and stands (*F3*) were extracted using the Fusion software (McGaughey, 2020). The raster generated from the point clouds for *R3a* were 18 × 18 m² spatial resolution, which approximately corresponds to the area covered by the permanent NFI plots (circular with 10 m radius).

The very high resolution (VHR) ALS data (***R4***) was collected by Riegl LMS-Q680i sensor with a pulse repetition frequency of 400 KHz and a scanning frequency of 135 Hz. This data was acquired from a helicopter and with a point density varying between 490 and 654 points/m², and across 4 systematic strips (11 km × 80 m) laid 8.2 km apart.

The TLS data (***R5***) scanned with a Trimble TX 8 laser scanner in Level 2 mode, with a wavelength of 1.5 μm, was used for single scans (~3 minutes) of trees from the centre of each plot. The TLS scans were conducted for *F4* in each strip in *S3*.

A pair of TanDEM-X satellite images (***R6***) were acquired in horizontally transmitted and horizontally received (HH) polarization in strip-map mode. InSAR processing technique was implemented to derive the phase height metric for *F3* in *S3*.

The field measured attributes and the metrics computed from the RS datasets were used for developing models under regression analysis for predicting forest attributes.

Table 3. Summary of the RS data used.

Data-set	Paper	Site	RS data	Sensor	Resolution	Acqu-isition year	Use*
Aerial images							
R1	**I**	*S1*	UAV DAP	Sensefly S.O.D.A camera	55 points/m^2	2019	T_r
ALS							
R2	**I**	*S1*	Norweg-ian NLS	Riegl VQ-1560i	2 points/m^2 (λ=1064 nm) **	2018	T_r
R3a	**II**	*S2*	Swedish NLS	Leica, Optech	0.5-1 points/m^2 (λ=1064 nm)	2009-2011	T_r
R3b	**V**	*S3*	Swedish NLS	Leica, Optech	0.5-1 points/m^2 (λ=1064 nm)	2012	T_r
R3c	**III, V**	*S3*	Swedish NLS	Leica ALS80	1.5 points/m^2 (λ=1064 nm)	2019	T_r, T_e
R4	**IV, V**	*S3*	VHR ALS	Riegl LMS-Q680i	490-654 points/m^2 (λ=1550 nm)	2019	T_r
TLS							
R5	**IV**	*S3*	TLS	Trimble TX 8	(λ=1500 nm)	2019	T_r

Data-set	Paper	Site	RS data	Sensor	Resolution	Acqu-isition year	Use*
Satellite images							
R6	**III**	*S3*	TanDEM-X	X-band SAR (HH)***	2.5 (slant) × 3.3 (azimuth) m^2 (Resampled to 10 m^2)	2015	T_r

*where, 'T_r' indicates dataset used for training prediction models, and 'T_e' indicating dataset used for testing or validating prediction models.

**λ is the wavelength.

***horizontal co-polarization was used.

3.2 Methods

3.2.1 Model-based inference overview

The implementation of model-based inference in association with RS datasets can be described overall as, incorporating a pre-processed reference dataset of the target variable, which can be acquired from manual field inventories or RS data (e.g., TLS, wall-to-wall rasters of Skogliga grunddata (low-density (LD) ALS) estimates) from the real population that can be seen as a single realization from a larger, hypothetical super-population, as represented in Figure 2. For past decades RS data have been used as auxiliary dataset. Similarly, in the current methodologies of the papers included in this thesis, 3D RS data has been used as auxiliary dataset. Followed by developing a regression model by linking the reference data and the auxiliary data for prediction of the target variable (e.g., AGB and VOL in this case). RS data with high spatial resolution enable models to make predictions at fine spatial scales given the reference data is available at the same fine spatial resolution. It also facilitates in understanding spatial patterns and heterogeneity within the study area, which is vital for ecological studies, land-use planning, and resource management. Regression models involving RS datasets can integrate data from multiple remote sensors (e.g., optical, radar, LiDAR) along with other anciliary information, e.g., climate data, soil type data, or topographic data in order to improve the accuracy and reliability of predictions.

The predicted values can be validated and evaluated for accuracy assessments, i.e., estimates of uncertainty, or presented as statistical plots. The individual papers deal with how the model-based inference was implemented or modified depending on the spatial extent, spatial resolution, and acquisition platform of the RS datasets. The model performances were evaluated using different statistical approaches, such as, computation of estimates and predictions together with corresponding CIs and PIs, respectively. Further improving model performances by introducing new reference datasets, combining probability sampling of the auxiliary variables

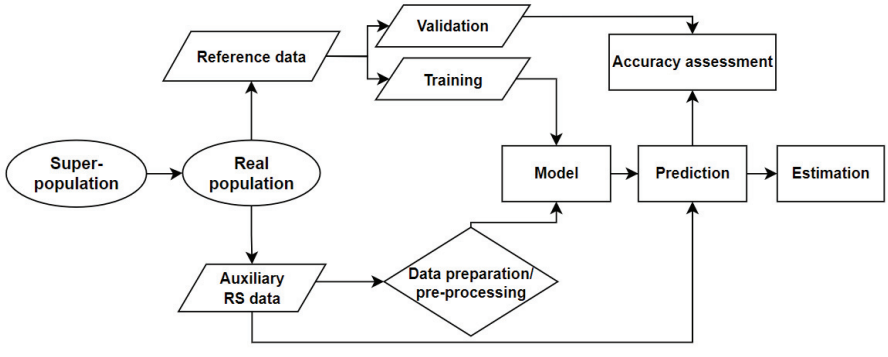


Figure 2. Conceptual view of a model-based inference implemented in this thesis.

3.2.2 Implementation of model-based inference and uncertainty estimation

In **Papers I-II,** regression models were developed based on field reference data and wall-to-wall auxiliary 3D RS data for AGB predictions in treeline and boreal ecosystems, respectively. In **Paper I**, model-based inference using a multiple linear regression (MLR) model was implemented by combining and comparing field reference data with auxiliary DAP and ALS datasets using an ABA. In **Paper II**, generalized linear models (GLMs) with different underlying exponential dispersion models namely, Gamma and Tweedie, were implemented within a model-based inference using ABA. In **Papers III-V**, MLR models were implemented to estimate VOL. Different transformations of the response and the explanatory variables were used to linearize the relationships. In **Paper III**, MLR models for InSAR data acquired wall-to-wall were developed and assessed by training the models using different reference datasets. In **Paper IV**, a MLR model was

implemented for hybrid inference, which is a special case of model-based inference. In **Paper V**, the model involved in the VOL growth forecasting approach was evaluated for transferability between two timepoints by manipulating the properties of the explanatory variables.

The uncertainty estimators described in *Section 1.5* were incorporated in the different papers as:

1. The **bias** in observations were handled by implementing probability sampling techniques for acquiring reference data (**Papers I-V**) and RS auxiliary data (**Paper IV, V**).

2. The model accuracies for the developed MLR models (**Papers I-IV**) were evaluated based on the RMSE, expressed as,

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

   where, n is the number of observations, y_i are the reference values, and $\hat{y}_i$ are the predicted values. As the VOL model used in **Paper V** was pre-developed for LD ALS estimations, therefore, the model accuracy was not evaluated explicitly.

3. The accuracy of the mean AGB estimates $\left(\overline{\widehat{\text{AGB}}}\right)$ for a treeline ecotone (in **Paper I)** was expressed by computing CIs around the mean AGB estimates, represented as,

$$\text{CI} = \overline{\widehat{\text{AGB}}} \pm (\widehat{se} \times 1.96) \qquad (2)$$

   where, $\widehat{se}$ is the standard error and 1.96 indicates a 95% CI.

4. PI derived based on Hattab (2016), was implemented to account for the uncertainties for predicted AGB values (in **Paper II**) for a GLM assuming Gamma distributions in a boreal forest stand, represented as,

$$\text{PI} = \left(\frac{\hat{G}_{(0.025)}\left(x_0'\hat{\beta}\right)^2}{2\hat{r}}, \frac{\hat{G}_{(0.975)}\left(x_0'\hat{\beta}\right)^2}{2\hat{r}}\right) \qquad (3)$$

where, $\hat{G}_{(0.025)}$ and $\hat{G}_{(0.975)}$ are the quantiles of a Chi-squared distribution (i.e., a special case of a Gamma distribution), $\hat{r}$ is the estimated shape parameter of Gamma regression model, x_0 is the explanatory variable, and $\hat{\beta}$ is the estimated vector of the model coefficients.

5. Sampling variance was computed for the design-phase ($\widehat{Var}_{dp}$) within the hybrid inference framework (in **Paper IV**), that can be represented generically as,

$$\widehat{Var}_{dp} = \widehat{Var}\left(\hat{V}_{\text{Site}}\right) = \frac{1}{J}\hat{s}^2_{\hat{T}_J} \quad (4)$$

where, $\hat{V}_{\text{Site}}$ is the predicted VOL per hectare for the entire test site, J is the total number of strips, and $\hat{s}^2_{\hat{T}_J}$ is the estimated standard deviation for the estimated total VOL.

3.2.3 Mapping and Estimating Aboveground Biomass in an Alpine Treeline Ecotone under Model-Based Inference (Paper I)

In **Paper I**, wall-to-wall auxiliary RS data from both DAP (*R1*) and ALS (*R2*), as presented in Table 3 (*Section 3.1.3*), were acquired for the test site, *S1*, as presented in Table 1 (*Section 3.1.1*), along with plot-level field reference data (*F1*), as presented in Table 2 (*Section 3.1.2*). A case study using linear regression models within a model-based inference with an ABA was implemented for prediction and mapping of AGB for SHORT and TALL vegetation strata.

The DAP point clouds were corrected for negative height values due to the inaccuracy of the DAP system and positioning of the GCP markers. This correction was done by interpolating ALS points for the DAP GCP locations and deriving a correction factor from the difference between the ALS and DAP ground points, which was then used for correcting all the DAP point clouds. The area-based metrics were derived from both ALS and DAP for TALL and SHORT strata separately, which included height percentiles, densities, standard deviation of the heights, skewness, and kurtosis.

Linear models for both sets of RS data and vegetation strata were fitted between the field reference AGBs and the metrics derived from ALS and DAP point clouds. A scale-independent metric, i.e., mean vegetation height (h_{mean}) was selected as the explanatory variable for the models for SHORT

stratum, whereas, 90th height percentile (h_{90}) was selected as the explanatory variable for the models for TALL stratum, represented in equation (5) and (6), respectively,

$$\mathrm{AGB} = \alpha_0 + \alpha_1 h_{\mathrm{mean}} + \varepsilon \quad (5)$$

$$\mathrm{AGB} = \beta_0 + \beta_1 h_{90} + \delta \quad (6)$$

where, α_i and β_i are the model coefficients, i is the number of model parameters, and ε and δ are the random error terms.

The mean estimated AGB $\left(\overline{\widehat{\mathrm{AGB}}}\right)$ for both strata individually and also for the entire area was computed as presented below,

$$\overline{\overline{\mathrm{AGB}}}_H = \frac{\overline{\overline{\mathrm{AGB}}}_{\mathrm{TALL}} \cdot A_{\mathrm{TALL}} + \overline{\overline{\mathrm{AGB}}}_{\mathrm{SHORT}} \cdot A_{\mathrm{SHORT}}}{A_H} \quad (7)$$

where, A is the area, and the subscripts TALL, SHORT, and H denote the TALL stratum, SHORT stratum, and the entire area, respectively.

The uncertainty of the mean estimates were presented as $\widehat{se}$ derived using parametric bootstrapping for 50,000 bootstrap samples. Parametric bootstrapping is a flexible non-parametric method that does not require assumptions about the underlying distribution of the data. It can be applied in situations where obtaining an independent validation dataset is difficult or expensive. This method also helps to assess the variability of a model by resampling the original data multiple times, which can enhance the model reliability. The number of iterations (corresponding to the number of bootstrap samples ($\mathrm{N_{PB}}$)) were checked with a $\widehat{se}$ stabilization indicator. The $\widehat{se}$ were computed for individual strata and the entire area by the following equations (8), and (9),

$$\widehat{se} = \sqrt{\frac{1}{\mathrm{N_{PB}}-1} \sum_{k=1}^{\mathrm{N_{PB}}} \left(\overline{\overline{\mathrm{AGB}}}_k - \overline{\overline{\mathrm{AGB}}}_{\mathrm{PB}}\right)^2} \quad (8)$$

$$\widehat{se}_H = \sqrt{\frac{\left(\widehat{se}^2_{\mathrm{TALL}} \cdot A^2_{\mathrm{TALL}} + \widehat{se}^2_{\mathrm{SHORT}} \cdot A^2_{\mathrm{SHORT}}\right)}{A^2_H}} \quad (9)$$

where, $\overline{\widehat{\mathrm{AGB}}}_{\mathrm{PB}}$ is the mean of the AGB prediction from k iterations of bootstrap samples $\left(\overline{\widehat{\mathrm{AGB}}}_{k}\right)$.

3.2.4 Computation of Prediction Intervals for Forest Aboveground Biomass Predictions using Generalized Linear Models in a Large-extent Boreal Forest Region (Paper II)

In **Paper II**, a case study with model-based inference using exponential dispersion models with an ABA was implemented for prediction and mapping of AGB for a large-extent boreal forest test site. Wall-to-wall auxiliary RS data from low-density ALS (*R3a*), as presented in Table 3 (*Section 3.1.3*), was acquired for the test site *S2*, as presented in Table 1 (*Section 3.1.1*), along with plot-level field reference data (*F2*), as presented in Table 2 (*Section 3.1.2*).

The wall-to-wall ALS data was processed following point cloud classification, height normalization, and feature extraction where area-based metrics similar to Paper I were derived. Gamma and Tweedie distributions were assumed for the generalized linear regression models (GLMs) to incorporate positive, continuous, and right-skewed AGB distribution. The GLMs were developed relating to the ALS metrics, as expressed in equation (10),

$$\eta_{\mathrm{AGB}_i} = \beta_0 + \beta_1 h_{90} + \beta_2 v + \beta_3 (h_{90} \times v) + \varepsilon \quad (10)$$

where, h_{90} is the 90$^{\text{th}}$ height percentile, v is the vegetation ratio, $h_{90} \times v$ is the interaction term between the former ALS metrics, $\eta_{\mathrm{AGB}_i} = (E(AGB_i))^{1/2}$ is the 'sqrt' (squareroot) link function, β_i are the model coefficients, i is the number of model parameters, and ε is the random error term.

Followed by the model development, PIs were computed using two methods for the Gamma GLM with a 'sqrt' link function: 1) R package '*ciTools*', and 2) PIs derived through asymptotic theory, as presented in equation (3). Whereas, only an extended version of the R package '*ciTools*' was used for computing PIs for the Tweedie GLM.

3.2.5 Comparing TanDEM-X InSAR Forest Stand Volume Prediction Models Trained using Field and ALS data (Paper III)

In **Paper III**, wall-to-wall auxiliary RS data from TanDEM-X SAR was acquired for the test site *S3*, as presented in Table 1 (*Section 3.1.1*), along with stand-level reference data collected through FI (*F3*), as presented in Table 2 (*Section 3.1.2*), and VOL estimates based on NFI and low-density ALS data (*R3c*), as presented in Table 3 (*Section 3.1.3*). VOL estimates have been of great importance to forest owners for measuring forest products such as timber, and acquisition of FI data is more time and labour intensive, especially in remote areas. A case study using model-based inference with an ABA was implemented for prediction and mapping of VOL using regression models to enable frequent upscaling of the estimates for larger spatial extents using satellite images to obtain results from digital FI for uses in forest management practices. The regression models were evaluated based on different sets of reference data – VOL estimates from FI (*F3*) (model A) and VOL estimates based on NFI and LD ALS data (*R3c*) (model B). The VOL estimated from ALS and NFI data for training model B allows us to use data that is available for entire Sweden, unlike collecting reference data from separate FIs, that might limit the implementation of the VOL regression models in different test sites in Sweden. Use of TanDEM-X images make the model predictions convenient to upscale to larger areas and giving the possibility of updating the predictions more frequently.

The VOL values were extracted for the corresponding stand areas which were then used for training the VOL regression model B. The explanatory variables were derived from products of InSAR processing of *R6*, and only phase height was selected for the regression analysis (models A and B), as represented in equation (11),

$$\mathrm{VOL} = \beta_0 + \beta_1 ph^{0.5} + \varepsilon \qquad (11)$$

where, VOL is the stand-level volume (*F3* and *R3c*), β_i are the model coefficients, i is the number of model parameters, ph is the interferometric phase height, and ε is the random error term.

The regression models A and B were validated using a set of harvester data (*F5*), as presented in Table 2 (*section 3.1.2*), available over 151 independent stands in *S3* which were priorly categorized into thinned (28

stands) or clear-felled (123 stands) classes, and only the clear-felled stands were considered in the validation dataset.

3.2.6 Large scale Forest Inventory using Digital Methods (Paper IV)

In **Paper IV**, the use of digital FI for a special case of model-based inference in the test site *S3*, as presented in Table 1 (*Section 3.1.1*), was evaluated for estimation of VOL. The regression model was implemented in combination with probability sampling based acquisition of auxiliary RS data *R4*, as presented in Table 3 (*Section 3.1.3*). Digital FI has been of interest for use in frequent updates of forest attribute maps without the need to involve manual field measurements. In some circumstances RS data can be acquired as probability samples instead of wall-to-wall, for example, when acquisition of wall-to-wall high density ALS data is costly, enabling only limited estimations of the mean and total of a forest stand. A methodology of large scale digital FI for the estimation of VOL was demonstrated within a hybrid inference framework using ITC approach.

In the hybrid-inference framework, four strips of dense ALS data (*R4)* served as auxiliary RS data. The *R4* was processed similarly for point cloud classification, and normalization, as in Papers I and II, followed by implementation of a watershed segmentation algorithm for ITC approach and feature extraction for the ITC segments. Tree-level model was established by linking TLS-derived variables from *R5* (acquired only within *F4*) to the metrics derived from *R4*, as expressed in equation (12),

$$(\mathrm{VOL}_{\mathrm{TLS}})^{0.2} \sim \alpha_0 + \alpha_1 (h_{95} \times C_{\mathrm{dns}}) + \alpha_2 C_{\mathrm{area}} \qquad (12)$$

where α_i are the model coefficients, i is the number of model parameters, h_{95} is the 95$^{\mathrm{th}}$ height percentile derived from *R4*, C_{dns} is the ITC density computed as the ratio of number of points above height cut-off/number of all returns, and C_{area} is the area of the ITC segment projections.

A modified design for sub-dividing the ALS scanned strips was exhibited in this paper, as shown in Figure 3. The predicted VOL values were up-scaled using a ratio-to-size estimator and the mean and total VOL values were estimated for the entire test site (*S3*). The accuracy of the estimated mean VOL was expressed as $\widehat{se}$. The design-phase variance was estimated using a simple random sampling (SRS) variance estimator, and a variance estimator by Heikkinen (2006).

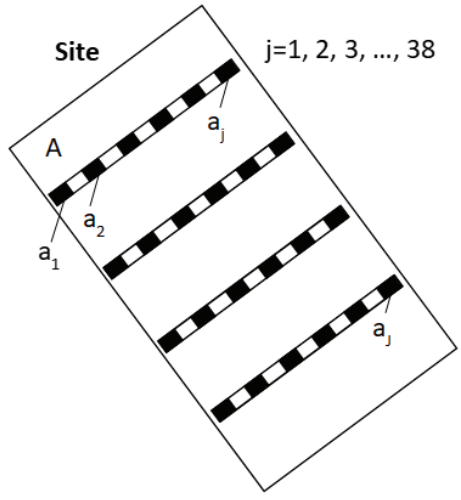


Figure 3. Shows the division of the strips into smaller grid cells at regular spatial intervals, selected systematic samples of 38 grid cells represented in black.

3.2.7 Forest Volume Growth Forecasting with Fusion of Dense and Sparse Airborne Laser Scanning Data (Paper V)

As mentioned in the previous Papers III-IV, the importance of VOL as a forest attribute of interest for stakeholders, making up-to-date VOL maps available enables making sound decisions and forest management plans. In **Paper V**, a methodology for the use of regression models for forest VOL growth prediction in terms of updated VOL maps for a large-area test site (*S3*), as presented in Table 1 (*Section 3.1.1*), was demonstrated. The independent variables, i.e., the ALS metrics from the LD ALS Skogliga grunddata VOL regression models were forecasted by combining systematic scanning strips of up-to-date dense ALS data (*R4*) with wall-to-wall pre-dated ALS data (*R3b*), as presented in Table 3 (*Section 3.1.3*), using histogram matching approach. The forecasted independent variables were incorporated in the LD ALS VOL regression models, as presented in equation (13). The LD ALS VOL regression models were evaluated for the accuracy of generating wall-to-wall forecasted VOL maps.

$$\sqrt{\mathrm{VOL}} = \hat{q}_{80} + \hat{p}_{95} + \hat{s} \qquad (13)$$

where, $\hat{q}_{80}$ is the product of estimated 80th height percentile ($\hat{p}_{80}$) and the vegetation ratio above 1.5 m, $\hat{p}_{95}$ is the 95th height percentile, and $\hat{s}$ is the vertical standard deviation of all values above 1.5 m.

The processed *R4* data from Paper IV was used as timepoint-2 (2019) data and *R3b* was used as timepoint-1 (2012) data, with the 2 timepoints being 7 years apart. The RS data *R3c* from 2019 was used as the validation dataset. Point clouds from all three datasets from 2012 and 2019 were re-processed with similar parameters, such as, height cut-off value ranging between 1.5 – 35 m above ground and scan angles ranging between ±19°.

A forest mask layer was used to mask out all non-forest areas from all the RS data. In the developed workflow, histogram matching was implemented between LD ALS 2012 and dense ALS 2019 data by accounting for the growth of 7 years. The look-up table (LUT) generated from the histogram matching step was then used to forecast the entire wall-to-wall rasters for each ALS metrics that would correspond to the growth up to timepoint-2. The forecasted ALS metrics were then used in the VOL regression models in order to predict wall-to-wall forecasted VOL map.

The above-mentioned workflow for forecasting VOLs was tested for: (i) histograms of every individual metric, e.g., $\hat{q}_{80}$, $\hat{p}_{95}$, and $\hat{s}$ extracted from LD ALS TP2 were matched with the histograms of the same metrics derived from LD ALS TP1 data, and (ii) histograms of every individual metric, e.g., $\hat{q}_{80}$, $\hat{p}_{95}$, and $\hat{s}$ extracted from very high density (VHD) ALS data from TP2 were matched with the histograms of the same metrics derived from LD ALS TP1, for each: (a) case A and (b) case B. The cases A and B stand for: A) forest change mask implemented, and B) forest change mask not implemented for LD ALS and VHD ALS rasters. Where, the forest change mask consisted of the clear-felled and harvested forest patches.

4. Results and discussion

The papers included in this thesis involve evaluation of model-based inference implemented for various cases of auxiliary 3D RS data acquisitions and for use in different spatial extents applicable for the use in forest management planning and practices. In **Paper I**, the accuracy of AGB regression model was observed to be higher by 2.7% for model-based case using wall-to-wall auxiliary ALS data compared to the model using DAP data for TALL vegetation stratum, but 36% lower for the SHORT vegetation stratum in a treeline ecotone. In **Paper II**, a model-based case with wall-to-wall ALS data for AGB prediction was tested for a larger productive boreal forest area with an accuracy of 25-26%. Followed by evaluating ALS as reference data for a model-based case in **Paper III**, when used in combination with height information extracted from TanDEM-X SAR data. It was observed to further improve the accuracy of the regression model by 3.6%, compared to the results from Paper II. The improved accuracy might be due to field measurements being averaged over a larger area, i.e., a forest stand instead of a plot, which might result in reduced local variation, measurement errors, and effects of spatial auto-correlation (SAC). In **Paper IV**, a dense ALS data acquired as probability samples in the first-phase sampling was used as auxiliary data for VOL regression modelling, in a special case of model-based inference referred to as, hybrid inference framework. The overall accuracy when the VOL regression model was used in combination with probability samples of RS data increased further by 3.9-17.4% compared to the results from Paper III. The increase in accuracy of estimated total VOL also varied between the naïve design and the modified approach. In **Paper V**, the VOL regression models from LD ALS were evaluated when forecasted independent variables from histogram matching process were incorporated in the models. The summary of statistics from the different papers are presented in Table 4.

Table 4. Summary of statistics for Papers I-IV.

Paper	Model	RS data	Approach	R^2	RMSE	rel. RMSE
AGB (Mg ha^{-1})						
I	TALL	ALS	ABA	0.47	12.4 Mg ha^{-1}	41.1%
		DAP	(5.6 m radius)	0.43	12.8 Mg ha^{-1}	43.8%
	SHORT	ALS	(1.5 m radius)	0.15	2.47 Mg ha^{-1}	154%
		DAP		0.27	2.28 Mg ha^{-1}	118%
II	Gamma	ALS	ABA	0.84	22.6 Mg ha^{-1}	25.9%
	Tweedie	ALS	(10 m radius)	0.88	21.7 Mg ha^{-1}	25.0%
VOL (m^3 ha^{-1})						
III	A*	SAR	ABA	0.58	46.5 m^3 ha^{-1}	22.3%
	B**	SAR	(8.0 m radius in 30 stands)	0.60	44.6 m^3 ha^{-1}	21.4%
IV	VOL	ALS	ITD	—	$\widehat{se}$ = 4.65 m^{3}ha^{-1}	$\widehat{se}$% = 4.04%

*model trained using volumes measured from *F3*.

**model trained using volumes computed based on volume estimates from *R3c*.

Evaluation of conventional models trained with manual FI data in **Papers I-II** lead the step ahead to the need to implement and test the use of RS-based estimates as reference data for regression models. In **Paper III**, the model-based inference case was evaluated by combining reference data extracted from available VOL estimates, derived using low-density ALS scans and NFI data, and InSAR phase height information. It was found to be performing similarly well compared to conventional ABA model-based inferences involving manual field measured data. A progress towards a completely RS-based digital FI workflow for the future of the forest industry without compromising on the accuracy of the forest attribute predictions was indicated. Even when shifting towards a digital FI workflow, there might be limitations regarding the cost of data acquisition, which is more significant for acquiring wall-to-wall large-area ALS data for use in a model-based inference framework, when compared to that of satellite images. Considering the higher accuracy of ALS data but also the high cost of acquiring wall-to-wall dense ALS data for large areas, the hybrid inference approach in **Paper IV,** is proposed for when the auxiliary RS data is not acquired wall-to-wall. The modified systematic sampling design was observed to have a positive

impact on the accuracy of forest VOL predictions. The impact of SAC was handled by empirically analysis conducted at different spatial intervals for predicted forest VOL values. Overall, the RS-based two-phase FI methodology demonstrated improved and precise predictions for canopy structure-related variables at an operational level. Since, the state of forest might change between longer intervals of field or RS data acquisition, regular and frequently updated maps are of great importance for the purpose of forest monitoring and management operations. The efficiency of model-based inference was further explored when the LD ALS VOL models were implemented with forecasted independent variables derived from recent ALS scans in **Paper V**. The sampled dense ALS data was combined with an earlier version of LD ALS data for forecasting growth of forest VOL. The use of the sampled dense ALS data was seen to provide similar forecasted mean VOL estimates when compared with VOL estimates from wall-to-wall ALS data. The specific results and the corresponding discussions of the individual papers are presented in *Sections 4.1-4.5*.

4.1 Mapping and Estimating Aboveground Biomass in an Alpine Treeline Ecotone under Model-Based Inference (Paper I)

ALS data has been used for prediction and mapping of forest attributes due to its metrics being strongly related to forest attributes such as, VOL and AGB. **Paper I** was a pilot case study where ALS and DAP were evaluated for using them as auxiliary data for inventorying AGB of treeline ecotone shrubs and low growing trees with a model-based inference. Due to the presence of low and irregularly growing vegetation the prediction uncertainties were higher, as shown in the scatterplots in Figure 4, compared to other studies (Ståhl et al., 2011a, 2011b; Maltamo et al., 2016; Nilsson et al., 2017).

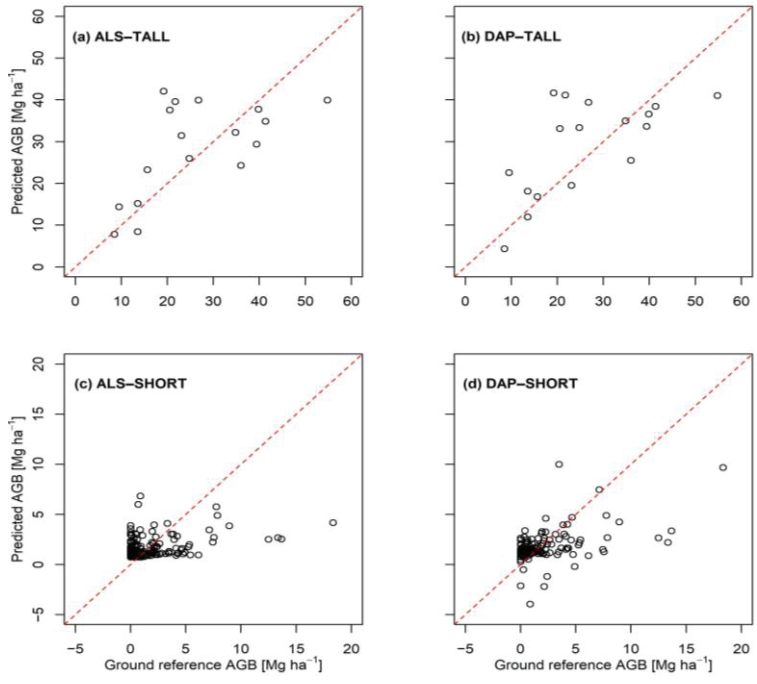


Figure 4. Scatterplots representing the relation between the ground reference AGB and predicted AGB values for: (a)ALS-TALL, (b) DAP-TALL, (c) ALS-SHORT, and (d) DAP-SHORT.

The inclusion or elimination of trees growing around the plot edges also seemed to have an impact on the accuracy of the predictions especially for the SHORT stratum because of the smaller plot size. It has also been observed before that the increase in positional errors decrease the estimation accuracies (Saarela, 2015; Persson et al., 2022). The number of iterations for the parametric bootstrapping was also set to a very large number, i.e., 50,000, since the AGB prediction models were weakly related even though the $\widehat{se}$.s stabilised between 30,000 and 40,000 iterations for SHORT and TALL, respectively, as shown in Figure 5. The overall $\widehat{se}$.s for the $\widehat{\text{AGB}}$.s derived from the parametric bootstrapping for the case of DAP was 0.03-0.05 Mg ha^{-1} lower compared to ALS. However, the CIs were promising for operationalising AGB estimation in the treeline ecotone. The 95% CIs for AGB estimates from ALS and DAP overlapped, indicating no significant difference between the two methods.

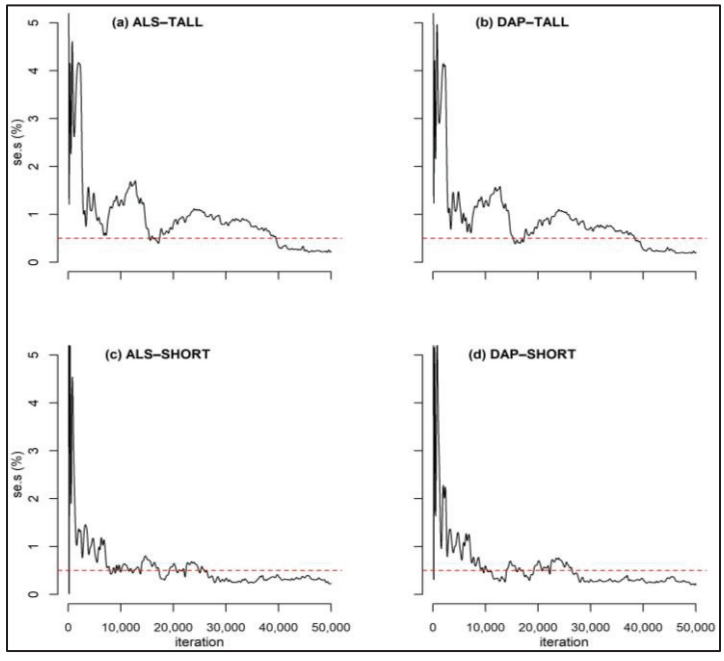


Figure 5. The $\widehat{se}$.s stabilisation indicator for plotted overs 50,000 iterations: (a)ALS-TALL, (b) DAP-TALL, (c) ALS-SHORT, and (d) DAP-SHORT.

4.2 Computation of Prediction Intervals for Forest Aboveground Biomass Predictions using Generalized Linear Models in a Large-extent Boreal Forest Region (Paper II)

Low density ALS data has been used extensively for large area FI purposes over the last few decades, but evaluation of prediction accuracies at the prediction unit-level has hardly been discussed and looked into. In **Paper II**, the computation of 95% PIs for assessing uncertainties in forest AGB using Gamma and Tweedie regression models with a 'sqrt' link function was introduced. Previous studies have estimated AGB and its uncertainty over large areas, but few have focused on prediction uncertainties using exponential dispersion models for wall-to-wall predictions. Our analysis showed that the Tweedie GLM provided a better fit than the Gamma GLM based on residual plots, though scatter plots of predicted versus observed AGB showed no distinct difference. PIs were estimated at the pixel level

using the R package '*ciTools*', which performed well regarding coverage%, and offered faster computations when compared to the PIs derived through an asymptotic theory.

As seen in Figure 6, the results indicate that PIs are narrower for lower predicted AGB values and widen with higher values, reflecting uniform growing pattern in young forests while increased uncertainty in older forests. The wider PIs for higher values of AGB might also result due to having less number of reference plots covering the higher AGB range or even due to low point density of the ALS data (*R3a*). The relative prediction uncertainties were consistent across different AGB ranges for the Gamma GLM. Whereas, the relative uncertainties for Tweedie GLM was observed to be inversely impacted by the magnitude of the predicted AGB values. The decrease in the relative uncertainties with increasing AGBs can be due to the proportion of the uncertainties compared to such high AGB values. The Tweedie regression had a lower average relative uncertainty (56.3%) compared to the Gamma regression (60.3%).

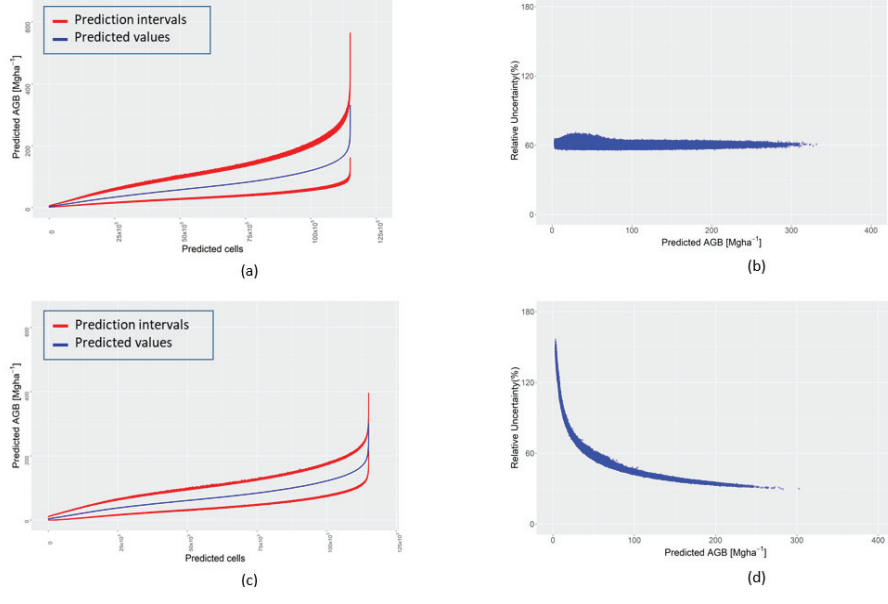


Figure 6. Plots of PIs and relative uncertainties for: (a-b) Gamma GLM, and (c-d) Tweedie GLM, respectively.

4.3 Comparing TanDEM-X InSAR Forest Stand Volume Prediction Models Trained using Field and ALS data (Paper III)

Model-based inference has been widely used for developing models based on field response data and auxiliary RS data. In **Paper III** for the first time, the efficiency of a model-based inference for an RS-based training dataset (Model B) was evaluated against a conventional model developed for the relation between field reference and RS auxiliary data (Model A). The results indicate that in both models A and B, the expected stand volumes were dependent on the variable phase height raised to the power of 0.5, as presented in equation (11). Although the model coefficients for phase height (ph) and corrected coherence (c_coh) were statistically significant, c_coh did not improve model prediction accuracy and was thus excluded from the final model. From the statistical summary as presented in Table 4 (*Section 4*), it can be observed that Model B's predictions were 0.7% more accurate compared to the predctions from Model A. Even though no distinct difference in the trends between the two models can be observed form the scatterplots presented in Figure 7, when comparing predicted stand volumes against observed stand volumes. Stand volumes were overestimated in both cases for the lower range of VOL values, likely due to the effect of 'regression towards the mean'. The relationship between predicted and observed volumes was nearly linear for stand volumes between 150 m^3ha^{-1} and 400 m^3ha^{-1}.

The models were validated with harvester data available for independent stands. It was observed that the performance of model B (44.6 m^3ha^{-1} (21.4%)) was slightly better compared to model A (46.5 m^3ha^{-1} (22.3%)) based on adjusted R^2 and RMSE values. It can be concluded that a RS-based estimated approach for prediction and mapping of stand volumes would be as promising as a method based on FI data along with being more frequent due to the availability of wall-to-wall ALS estimates over entire Sweden.

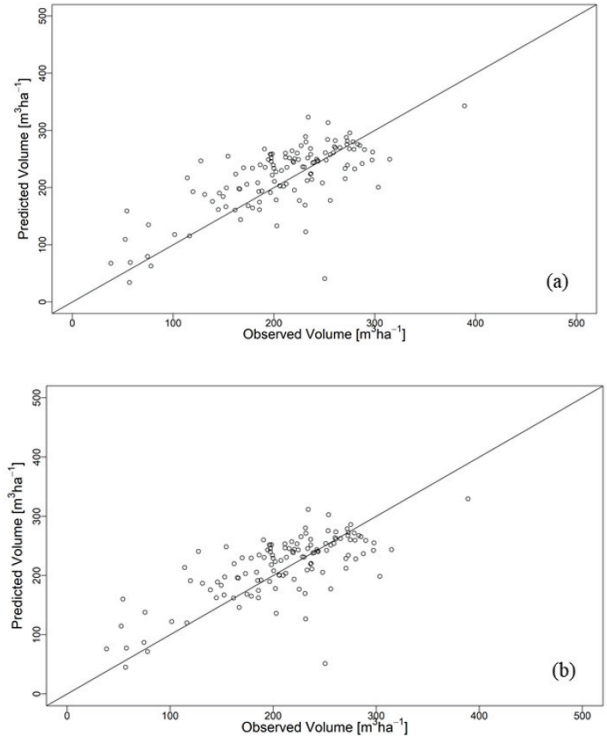


Figure 7. Scatterplots of reference VOL v/s predicted VOL for: (a) Model A, and (b) Model B.

4.4 Large scale Forest Inventory using Digital Methods (Paper IV)

In **Paper IV**, the developed VOL regression model in the model-phase had an RMSE of 0.15 m^3, as shown in Figure 8. The overall accuracy of 4% in this study appears promising with a design that covers a test site smaller than a region, yet considerably larger than stand-level. The first-phase, i.e., the design-phase with systematic sampling appeared as a useful contribution to lower the uncertainties of the estimates. It is also well suited for digital FI cases which can be implemented easily for ALS scans along with two days of TLS scanning period. Acquisition of systematic samples of very dense ALS data will also enable collection of detailed information about representative samples of large forest stands. The Matérn's estimator may underestimate the true variance, since the level of a single strip is cancelled out in differential approaches when strips in only one direction are used. The

total predicted VOL values for the fourth strip was higher by 5700-6700 m^3 when compared to the other three strips. Such differences in-between strips and the spatial variation can be captured better by adding at least one perpendicular flight strip.

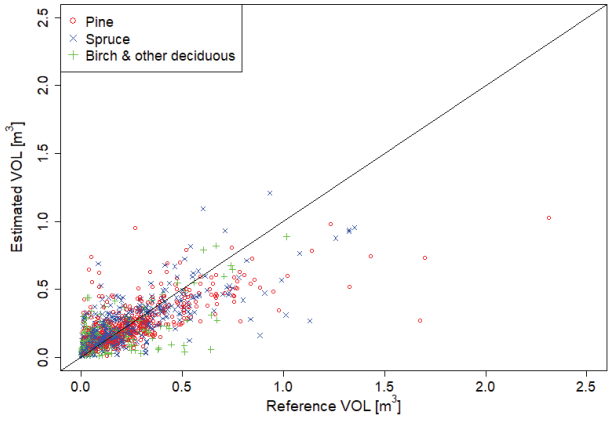
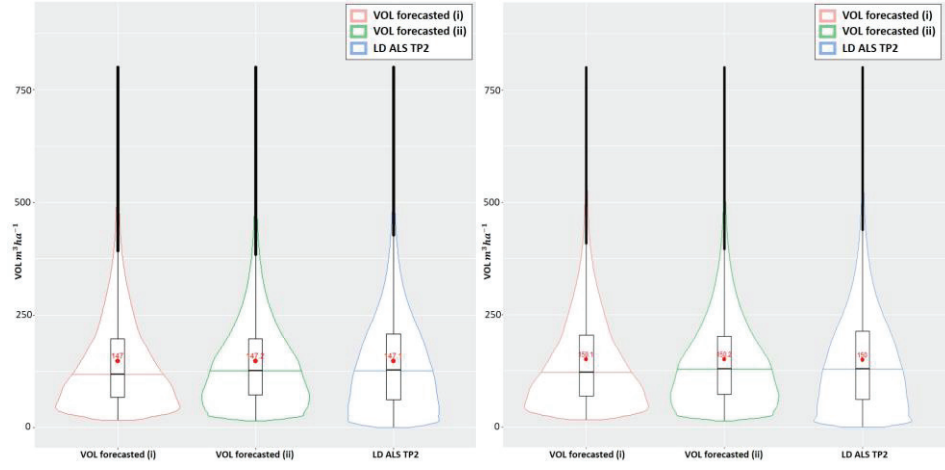


Figure 8. Scatterplots of reference v/s estimated VOL [m^3].

4.5 Forest Volume Growth Forecasting with Fusion of Dense and Sparse Airborne Laser Scanning Data (Paper V)

In **Paper V**, the histogram matching process was implemented to the independent variables of VOL regression model, i.e., the individual ALS metrics between the two timepoints. The forecasted VOLs from case A, case B, and the validation VOL data from *R3c*, respectively, have been presented in the violin plots in Figure 9 (a-b). The mean predicted VOL values were around 147 m^3ha^{-1} and 150 m^3ha^{-1} for case A and B, respectively. The forecasted VOL values for sub-cases (i) and (ii) for each case A and B were observed to be similar to the mean VOL value from the validation data, *R3c*. The differences in the distributions in the violin plot in Figure 9, might result due to difference in techniques of data filtering methods involved in LD ALS estimations which were not considered in our workflow. Nevertheless, the

LD ALS VOLs are itself predicted using VOL functions which can involve errors from field measurements.



(a) (b)

Figure 9. Violin plots representing the distribution of forecasted VOLs for: (i) histograms of every individual metric extracted from LD ALS 2019 (R3c) were matched with the histograms of the same metrics derived from LD ALS 2012 (R3b), and (ii) histograms of every individual metric extracted from VHD ALS 2019 data (*R4*) were matched with the histograms of the same metrics derived from LD ALS 2012 (*R3b*), for each case: (a) A (with mean VOLs around 147 m^3ha^{-1}), and (b) B (with mean VOLs around 150 m^3ha^{-1}), as described in *Section 3.2.7*.

5. Conclusions

RS has proven to be an irreplaceable tool in various fields of research. Over several years of application of RS in the field of forestry, 3D data acquired remotely have gained more interest and enabled further improvement of the forest management activities and tactical planning for forest owners and stakeholders. The main objective of this thesis was to evaluate the implementation of model-based inference for different cases of 3D RS data acquisitions from different sensors, platforms, and at different spatial extents for the prediction of state and growth of forest VOL and related forest attribute such as, AGB, for large-area operational forest management purposes.

Specific conclusions drawn from the individual papers included in this thesis are:

1. Regression models formulated in **Paper I** integrating 3D metrics derived from ALS data was found to be more accurate compared to 3D metrics derived from DAP for AGB predictions for TALL vegetation stratum, though no statistically significant differences were identified.

2. The findings in **Paper I** also indicated that DAP can be used operationally for monitoring of treeline ecotones with acceptably accurate measurements of forest attributes, such as AGB or similarly for VOL.

3. In **Paper II**, Tweedie regression was observed to be useful when planning for harvesting purposes due to its lower uncertainty for higher predicted AGB values, whereas, the Gamma regression will be useful for mapping young forests when planning for thinning activities due to its lower uncertainty for lower range of predicted AGB values.

4. Accuarcy assessment of the VOL predictions from regression models presented in **Paper III**, indicated the efficient use of RS and NFI based estimates as reference data for model-based inference instead of manually collected FI data.

5. InSAR images used in **Paper III**, acquired from TanDEM-X can be a reliable source for prediction and upscaling of forest VOL values over large-area forest stands.

6. The design modification, presented in **Paper IV**, for the design-phase incorporated in hybrid inference framework can have an impact on the magnitude of the uncertainty estimates.

7. The presented methodology in **Paper V**, indicates that histogram matching can be opted as a feasible and quick approach to incorporate forecasted independent variables in VOL regression models for updating VOL maps combining low density or sparse regional ALS scans and sampled dense ALS scans.

Overall, from this thesis it can be concluded that RS data acquired from different sensors, RS platforms, and at different spatial extents can be successfully integrated within model-based inference for prediction and estimation of VOL and AGB. 3D RS data is beneficial for accurate prediction of forest attributes for large-area operational forest management activities when combined with either field or RS-based reference datasets. Moreover, entirely RS-based workflows for forest attribute prediction and forecasting are practical for cost- and time-effective digital FI and monitoring purposes, though further research is required.

6. Future research

The current findings from this thesis gives scope for further investigations such as, assessing if the proposed model-based approach implemented in the treeline ecotone can support repeated surveys for the purpose of AGB change estimation (**Paper I**). Due to changing climate, the need to monitor changes in transition zone for example, treeline ecotones will be of high importance in the near future. Further evaluations can be done to compare predicted AGB and uncertainty maps in association with species distribution maps from sources like the National Land Cover Database or Sentinel-2 imagery to better understand AGB distribution and the uncertainties (**Paper II**). Testing the models developed in **Paper III**, for transferability and robustness over different test site locations, conditions and species composition will contribute in updating VOL maps over entire Sweden with readily available RS data. The quick approach of forecasting of metrics using histogram matching (**Paper V**) can be tested for producing forecasted VOL maps from RS based reference data and TanDEM-X InSAR images. Further simulation-based investigation is required for the results presented in **Paper IV**, in order to validate the precision of the variance estimators. A three-phase FI workflow can be developed as an extension to the two-phase digital FI approach (**Paper IV**), by combining the TLS scanned data, dense ALS scans, and a potential wall-to-wall RS data, for example, satellite images or low-density ALS scans.

References

Axelsson, C. R., Lindberg, E., Persson, H. J., & Holmgren, J. (2023). The use of dual-wavelength airborne laser scanning for estimating tree species composition and species-specific stem volumes in a boreal forest. *International Journal of Applied Earth Observation and Geoinformation*, *118*, 103251. https://doi.org/10.1016/J.JAG.2023.103251

Balzter, H. (2001). Forest mapping and monitoring with interferometric synthetic aperture radar (InSAR). *Http://Dx.Doi.Org/10.1177/030913330102500201*, *25*(2), 159–177. https://doi.org/10.1177/030913330102500201

Bohlin, I., Maltamo, M., Peltola, H., Pippuri, I., Kallio, E., & Packalén, P. (2011). *Exploring horizontal area-based metrics to discriminate the spatial pattern of trees using ALS*. https://www.researchgate.net/publication/266591940

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

Brewer, K. R. W. (1963). Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process. *Australian Journal of Statistics*, *5*(3), 93–105. https://doi.org/10.1111/J.1467-842X.1963.TB00288.X

Cassel, C., Särndal, C.-E., & Wretman, J. H. (1977). *Foundations of inference in survey sampling*. Wiley.

Corona, P., Fattorini, L., Franceschi, S., Scrinzi, G., & Torresan, C. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: Model-based, Design-based, And hybrid perspectives. *Canadian Journal of Forest Research*, *44*(11), 1303–1311. https://doi.org/10.1139/CJFR-2014-0203/ASSET/IMAGES/LARGE/CJFR-2014-0203F3.JPEG

Esteban, J., McRoberts, R. E., Fernández-Landa, A., Tomé, J. L., & Marchamalo, M. (2020). A Model-Based Volume Estimator that Accounts for Both Land Cover Misclassification and Model Prediction Uncertainty. *Remote Sensing 2020, Vol. 12, Page 3360*, *12*(20), 3360. https://doi.org/10.3390/RS12203360

Esteban, J., McRoberts, R. E., Fernández-Landa, A., Tomé, J. L., & Næsset, E. (2019). Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data. *Remote Sensing 2019, Vol. 11, Page 1944*, *11*(16), 1944. https://doi.org/10.3390/RS11161944

Eysn, L., Hollaus, M., Lindberg, E., Berger, F., Monnet, J. M., Dalponte, M., Kobal, M., Pellegrini, M., Lingua, E., Mongus, D., & Pfeifer, N. (2015). A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space. *Forests 2015, Vol. 6, Pages 1721-1747*, *6*(5), 1721–1747. https://doi.org/10.3390/F6051721

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A. H., & Ståhl, G. (2014). Adapting National Forest Inventories to changing requirements - The case of the Swedish National Forest Inventory at the turn of the 20th century. *Silva Fennica*, *48*(3). https://doi.org/10.14214/SF.1095

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, *28*(10), 1429–1447. https://doi.org/10.1139/cjfr-28-10-1429

Haala, N., Pfeifer, N., & Franzen, M. (2014). *Dense Image Matching | EuroSDR*. https://www.eurosdr.net/research/project/dense-image-matching

Hattab, M. W. (2016). A derivation of prediction intervals for gamma regression. *Journal of Statistical Computation and Simulation*, *86*(17), 3512–3526. https://doi.org/10.1080/00949655.2016.1169421

Heikkinen. (2006). *Forest Inventory*. *10*. https://doi.org/10.1007/1-4020-4381-3

Högberg, P. (2021). *Sustainable boreal forest management-challenges and opportunities for climate change mitigation Report from an Insight Process conducted by a team appointed by the International Boreal Forest Research Association (IBFRA)*.

Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B., & Maier, B. (2009). Operational wide-area stem volume estimation based on

airborne laser scanning and national forest inventory data. *International Journal of Remote Sensing*, *30*(19), 5159–5175. https://doi.org/10.1080/01431160903022894

Holmgren, J., & Lindberg, E. (2013). Tree crown segmentation based on a geometric tree crown model for prediction of forest variables. *Canadian Journal of Remote Sensing*, *39*(SUPPL.1). https://doi.org/10.5589/M13-025

Holmgren, J., Lindberg, E., Olofsson, K., & Persson, H. J. (2022). Tree crown segmentation in three dimensions using density models derived from airborne laser scanning. *International Journal of Remote Sensing*, *43*(1), 299–329. https://doi.org/10.1080/01431161.2021.2018149

Holmgren, J., & Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, *90*(4), 415–423. https://doi.org/10.1016/S0034-4257(03)00140-8

Holmgren, J., Persson, Å., & Söderman, U. (2008). Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*, *29*(5), 1537–1552. https://doi.org/10.1080/01431160701736471

Holmgren, J., & Soderman, U. (2002). Detecting and Measuring Individual Trees Using an Airborne Laser Scanner. *Photogrammetric Engineering & Remote Sensing*, *68*(9), 925–932.

Hyyppä, J., & Inkinen, M. (1999). Detecting and estimating attributes for single trees using laser scanner. Photogramm J Finl. *The Photogrammetric Journal of Finland*, *16*(2). https://www.researchgate.net/publication/258707378_Detecting_and_estimating_attributes_for_single_trees_using_laser_scanner_Photogramm_J_Finl

International Model Forest Network » Bergslagen Model Forest. (n.d.). Retrieved July 14, 2024, from https://imfn.net/regional-networks/bergslagen-model-forest/

Kellndorfer, J. M., Walker, W. S., LaPoint, E., Kirsch, K., Bishop, J., & Fiske, G. (2010). Statistical fusion of lidar, InSAR, and optical remote sensing data for forest stand height characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM+, and ancillary data sets. *Journal of Geophysical Research: Biogeosciences*, *115*(G2), 0–08. https://doi.org/10.1029/2009JG000997

Kraus, K. (2007). Photogrammetry: geometry from images and laser scans / Karl Kraus ; translated by Ian Harley, Stephen Kyle. *Photogrammetry: Geometry from Images and Laser Scans*. https://www.adlibris.com/se/bok/photogrammetry-9783110190076

Lantmäteriet. (2019). *Lantmäteriet - we know every single place in Sweden. | Lantmäteriet*. https://www.lantmateriet.se/sv/

Lindberg, E. (2012). *Estimation of canopy structure and individual trees from laser scanning data*. Swedish University of Agricultural Sciences.

Lindberg, E., Holmgren, J., Olofsson, K., & Olsson, H. (2012). Estimation of stem attributes using a combination of terrestrial and airborne laser scanning. *European Journal of Forest Research*, *131*(6), 1917–1931. https://doi.org/10.1007/S10342-012-0642-5/FIGURES/11

Lindberg, E., Holmgren, J., & Olsson, H. (2021). Classification of tree species classes in a hemi-boreal forest from multispectral airborne laser scanning data using a mini raster cell method. *International Journal of Applied Earth Observation and Geoinformation*, *100*, 102334. https://doi.org/10.1016/J.JAG.2021.102334

Mallet, C., & Bretar, F. (2009). Full-waveform topographic lidar : State-of-the-art. *Isprs Journal of Photogrammetry and Remote Sensing*, *64*(1), 1–16. https://doi.org/10.1016/J.ISPRSJPRS.2008.09.007

Maltamo, M., Bollandsås, O. M., Gobakken, T., & Næsset, E. (2016). Large-scale prediction of aboveground biomass in heterogeneous mountain forests by means of airborne laser scanning. *Canadian Journal of Forest Research*, *46*(9), 1138–1144. https://doi.org/10.1139/CJFR-2016-0086

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A., & Kangas, J. (2006). Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Https://Doi.Org/10.1139/X05-246*, *36*(2), 426–436. https://doi.org/10.1139/X05-246

Matérn, B. (1960). *Spatial Variation* (Vol. 49, Issue 5). Springer New York. https://doi.org/10.1007/978-1-4615-7892-5

McGaughey, B. (2020). *FUSION/LDV LIDAR analysis and visualization software*. Pacific Northwest Research Station USDA Forest Service. http://forsys.cfr.washington.edu/fusion/fusion_overview.html

Mienna, I. M., Austrheim, G., Klanderud, K., Bollandsås, O. M., & Speed, J. D. M. (2022). Legacy effects of herbivory on treeline dynamics along

an elevational gradient. *Oecologia*, *198*(3), 801–814. https://doi.org/10.1007/S00442-022-05125-8/FIGURES/5

Mienna, I. M., Speed, J. D. M., Klanderud, K., Austrheim, G., Næsset, E., & Bollandsås, O. M. (2020). The relative role of climate and herbivory in driving treeline dynamics along a latitudinal gradient. *Journal of Vegetation Science*, *31*(3), 392–402. https://doi.org/10.1111/JVS.12865

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

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

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, *80*(1), 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5

Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, *19*(2), 164–179. https://doi.org/10.1080/02827580310019257

Naesset, E., Gobakken, T., Jutras-Perreault, M.-C., Naesset Ramtvedt, E., Escarcena, C., Delgado García, J., & João Sousa, J. (2021). Comparing 3D Point Cloud Data from Laser Scanning and Digital Aerial Photogrammetry for Height Estimation of Small Trees and Other Vegetation in a Boreal–Alpine Ecotone. *Remote Sensing 2021, Vol. 13, Page 2469*, *13*(13), 2469. https://doi.org/10.3390/RS13132469

Næsset, E., Gobakken, T., & McRoberts, R. E. (2019). A Model-Dependent Method for Monitoring Subtle Changes in Vegetation Height in the Boreal–Alpine Ecotone Using Bi-Temporal, Three Dimensional Point Data from Airborne Laser Scanning. *Remote Sensing 2019, Vol. 11, Page 1804*, *11*(15), 1804. https://doi.org/10.3390/RS11151804

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, *56*(1), 1–7. https://doi.org/10.1016/0034-4257(95)00224-3

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., & Olsson, H. (2017). A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sensing of Environment*, *194*, 447–454. https://doi.org/10.1016/j.rse.2016.10.022

Noordermeer, L., Bielza, J. C., Saarela, S., Gobakken, T., Bollandsås, O. M., & Næsset, E. (2023). Monitoring tree occupancy and height in the Norwegian alpine treeline using a time series of airborne laser scanner data. *International Journal of Applied Earth Observation and Geoinformation*, *117*, 103201. https://doi.org/10.1016/J.JAG.2023.103201

Nurminen, K., Karjalainen, M., Yu, X., Hyyppä, J., & Honkavaara, E. (2013). Performance of dense digital surface models based on image matching in the estimation of plot-level forest variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, *83*, 104–115. https://doi.org/10.1016/J.ISPRSJPRS.2013.06.005

Nyström, M., Holmgren, J., & Olsson, H. (2012). Prediction of tree biomass in the forest–tundra ecotone using airborne laser scanning. *Remote Sensing of Environment*, *123*, 271–279. https://doi.org/10.1016/J.RSE.2012.03.008

Nyström, M., Holmgren, J., & Olsson, H. (2013). Change detection of mountain birch using multi-temporal ALS point clouds Change detection of mountain birch using multi-temporal ALS point clouds. *Remote Sensing Letters*, *4*(2), 190–199. https://doi.org/10.1080/2150704X.2012.714087

Packalén, P., & Maltamo, M. (2007). The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, *109*(3), 328–341. https://doi.org/10.1016/J.RSE.2007.01.005

Persson, Å., Holmgren, J., Söderman, U., & Olsson, H. (2012). Tree Species Classification of Individual Trees in Sweden by Combining High Resolution Laser Data with High Resolution Near-infrared Digital Images. *Inetrnational Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*(8).

Persson, H., & Fransson, J. E. S. (2014). Forest Variable Estimation Using Radargrammetric Processing of TerraSAR-X Images in Boreal Forests.

Remote Sensing 2014, Vol. 6, Pages 2084-2107, *6*(3), 2084–2107. https://doi.org/10.3390/RS6032084

Persson, H. J. (2016). Estimation of Boreal Forest Attributes from Very High Resolution Pléiades Data. *Remote Sensing 2016, Vol. 8, Page 736*, *8*(9), 736. https://doi.org/10.3390/RS8090736

Persson, H. J., Ekström, M., & Ståhl, G. (2022). Quantify and account for field reference errors in forest remote sensing studies. *Remote Sensing of Environment*, *283*, 113302. https://doi.org/10.1016/J.RSE.2022.113302

Persson, H. J., & Fransson, J. E. S. (2017). Comparison between TanDEM-X and ALS based estimation of above ground biomass and tree height in boreal forests. *Scandinavian Journal of Forest Research*, *32*(4), 306–319. https://doi.org/10.1080/02827581.2016.1220618

Persson, H. J., Olofsson, K., & Holmgren, J. (2022). Two-phase forest inventory using very-high-resolution laser scanning. *Remote Sensing of Environment*, *271*, 112909. https://doi.org/10.1016/J.RSE.2022.112909

Persson, H. J., Olsson, H., Soja, M. J., Ulander, L. M. H., & Fransson, J. E. S. (2017). Experiences from Large-Scale Forest Mapping of Sweden Using TanDEM-X Data. *Remote Sensing 2017, Vol. 9, Page 1253*, *9*(12), 1253. https://doi.org/10.3390/RS9121253

Persson, H., Wallerman, J., Olsson, H., & Fransson, J. E. S. (2013). Estimating forest biomass and height using optical stereo satellite data and a DTM from laser scanning data. *Canadian Journal of Remote Sensing*, *39*(3), 251–262. https://doi.org/10.5589/M13-032

Rahlf, J. (2017). *Forest resource mapping using 3D remote sensing: Combining national forest inventory data and digital aerial photogrammetry Kartlegging av skogressurser med 3D fjernmåling: Kombinering av data fra landskogtakseringen og*.

Roberge, C., Nilsson, P., Wikberg, P.-E., & Fridman, J. (2023). *Skogsdata 2023*. https://www.slu.se/globalassets/ew/org/centrb/rt/dokument/skogsdata/skogsdata_2023_webb.pdf

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, *57*(2), 377–387. https://doi.org/10.1093/BIOMET/57.2.377

Saarela, S. (2015). Use of remotely sensed auxiliary data for improving sample-based forest inventories. *Dissertationes Forestales*, *201*. https://doi.org/10.14214/df.201

Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T. G., Nelson, R. F., & Ståhl, G. (2016). Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science*, *73*(4), 895–910. https://doi.org/10.1007/s13595-016-0590-1

Saarela, S., Holm, S., Healey, S. P., Andersen, H. E., Petersson, H., Prentius, W., Patterson, P. L., Næsset, E., Gregoire, T. G., & Ståhl, G. (2018). Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and landsat data. *Remote Sensing*, *10*(11), 1832. https://doi.org/10.3390/rs10111832

Saarela, S., Wästlund, A., Holmström, E., Mensah, A. A., Holm, S., Nilsson, M., Fridman, J., & Ståhl, G. (2020). Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. *Forest Ecosystems*, *7*(1). https://doi.org/10.1186/s40663-020-00245-0

Skogsdata 2024. (2024). www.slu.

Soja, M. J., Persson, H. J., & Ulander, L. M. H. (2015a). Estimation of forest biomass from two-level model inversion of single-pass InSAR data. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(9), 5083–5099. https://doi.org/10.1109/TGRS.2015.2417205

Soja, M. J., Persson, H. J., & Ulander, L. M. H. (2015b). Estimation of forest height and canopy density from a single InSAR correlation coefficient. *IEEE Geoscience and Remote Sensing Letters*, *12*(3), 646–650. https://doi.org/10.1109/LGRS.2014.2354551

Solberg, S., Astrup, R., Bollandsås, O. M., Næsset, E., & Weydahl, E. (2010). Deriving forest monitoring variables from X-band InSAR SRTM height. *Canadian Journal of Remote Sensing*, *36*(1), 68–79. https://doi.org/10.5589/M10-025

Solberg, S., Astrup, R., Breidenbach, J., Nilsen, B., & Weydahl, D. (2013). Monitoring spruce volume and biomass with InSAR data from TanDEM-X. *Remote Sensing of Environment*, *139*, 60–67. https://doi.org/10.1016/J.RSE.2013.07.036

Solberg, S., Astrup, R., Gobakken, T., Næsset, E., & Weydahl, D. J. (2010). Estimating spruce and pine biomass with interferometric X-band SAR.

Remote Sensing of Environment, *114*(10), 2353–2360. https://doi.org/10.1016/J.RSE.2010.05.011

Solberg, S., Gizachew, B., Næsset, E., Gobakken, T., Bollandsås, M. O., Mauya, W. E., Olsson, H., Malimbwi, R., & Zahabu, E. (2015). Monitoring forest carbon in a Tanzanian woodland using interferometric SAR: A novel methodology for REDD+. *Carbon Balance and Management*, *10*(1), 1–14. https://doi.org/10.1186/S13021-015-0023-8/TABLES/4

Solberg, S., Naesset, E., & Martin Bollandsas, O. (2006). *Single Tree Segmentation Using Airborne Laser Scanner Data in a Structurally Heterogeneous Spruce Forest*.

Ståhl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., & Nelson, R. (2011a). Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, *41*(1), 83–95. https://doi.org/10.1139/X10-195

Ståhl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., & Nelson, R. (2011b). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark county, Norway. *Canadian Journal of Forest Research*, *41*(1), 96–107. https://doi.org/10.1139/X10-161/ASSET/IMAGES/X10-161E43H.GIF

Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S. P., Patterson, P. L., Magnussen, S., Næsset, E., McRoberts, R. E., & Gregoire, T. G. (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems*, *3*(1), 1–11. https://doi.org/10.1186/S40663-016-0064-9/FIGURES/2

Strîmbu, V. F., & Strîmbu, B. M. (2015). A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *104*, 30–43. https://doi.org/10.1016/J.ISPRSJPRS.2015.01.018

The forest and sustainable forestry - Swedish Wood. (2024). https://www.swedishwood.com/wood-facts/about-wood/wood-and-sustainability/the-forest-and-sustainable-forestry/

Vastaranta, M., Holopainen, M., Karjalainen, M., Kankare, V., Hyyppa, J., & Kaasalainen, S. (2014). TerraSAR-X stereo radargrammetry and airborne scanning LiDAR height metrics in imputation of forest aboveground biomass and stem volume. *IEEE Transactions on*

Geoscience and Remote Sensing, *52*(2), 1197–1204. https://doi.org/10.1109/TGRS.2013.2248370

Vastaranta, M., Holopainen, M., Yu, X., Haapanen, R., Melkas, T., Hyyppä, J., & Hyyppä, H. (2011). Individual Tree Detection and Area-based Approach in Retrieval of Forest Inventory Characteristics from Low-pulse Airborne Laser Scanning Data. *The Photogrammetric Journal of Finland*, *22*(2).

Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K. M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T., & Maltamo, M. (2012). Comparative testing of single-tree detection algorithms under different types of forest. *Forestry: An International Journal of Forest Research*, *85*(1), 27–40. https://doi.org/10.1093/FORESTRY/CPR051

Vauhkonen, J., Ørka, H. O., Holmgren, J., Dalponte, M., Heinzel, J., Koch, B., Vauhkonen, J., Ørka, H. O., Holmgren, J., Dalponte, M., & Heinzel, J. (2014). *Tree Species Recognition Based on Airborne Laser Scanning and Complementary Data Sources* (pp. 135–156). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8663-8_7

Yu, X., Hyyppä, J., Karjalainen, M., Nurminen, K., Karila, K., Vastaranta, M., Kankare, V., Kaartinen, H., Holopainen, M., Honkavaara, E., Kukko, A., Jaakkola, A., Liang, X., Wang, Y., Hyyppä, H., & Katoh, M. (2015). Comparison of Laser and Stereo Optical, SAR and InSAR Point Clouds from Air- and Space-Borne Sources in the Retrieval of Forest Inventory Attributes. *Remote Sensing 2015, Vol. 7, Pages 15933-15954*, *7*(12), 15933–15954. https://doi.org/10.3390/RS71215809

Popular science summary

Forests play a major role in our environment, acting as carbon sinks, biodiversity hotspots, and resources for timber production. To manage forests sustainably, we need to monitor important attributes like tree volume, aboveground biomass, and forest health. Historically, this information was gathered through ground-based forest inventories, but advancements in technology for collecting data from sensors remotely have transformed the process, making it faster, equally or more accurate, and frequently repeatable. These sensors can collect data from different platforms such as, the ground, air, and from the space.

The main focus of this thesis was to look into the use of three-dimensional (3D) information extracted from remote sensing data collected from different sensors, for example, airborne laser scanning, digital aerial photogrammetry, and synthetic aperture radar, for measuring forest attributes. These tools allow forest managers to gather detailed forest data over large areas, improving the efficiency of forest monitoring and mapping.

Sweden's National Forest Inventory (NFI), which started in 1923, is an example of how practices of forest measurements and data collection have evolved over time. The NFI uses a mix of permanent and temporary sample plots to collect forest data, and methods based entirely on the manually collected data is used for estimating forest attributes. However, using models with information extracted from remote sensing data, offers a way to predict forest attributes more comprehensively across larger areas.

The integration of 3D remote sensing data into different cases of model development for prediction and estimation of attributes like volume and aboveground biomass have been looked into (in Papers I-IV). Also, combining dense and sparse airborne laser scanning data to forecast forest growth through volume models was also explored (in Paper V).

Uncertainty estimation is a required aspect of forest inventory, particularly when combining field and remote sensing data. Errors in forest attribute estimates arise from various sources, including sensor noise, instrument inaccuracies, and model parameter estimations. These errors can be quantified using different methods. A few of the uncertainty estimation methods used in this thesis were, root mean square error, confidence intervals, and prediction intervals. Root mean square error measures the aggregated accuracy of model predictions by calculating the square root of the average squared differences between predicted and observed values. Confidence intervals provide a range for the estimated mean predicted values, while prediction intervals account for both the random variability in the model parameters and new observations, making them wider than confidence intervals, are used to assess the reliability of these predictions.

Overall, this thesis highlights that 3D remote sensing technology used in forest inventory practices not only saves time and costs but also provides more accurate data for forest planning and management activities implemented at all levels such as, strategic, operational, and tactical. As these technologies advance, they hold great promise for improving forest resource management.

Populärvetenskaplig sammanfattning

Skogar spelar en viktig roll i vår miljö genom att fungera som kolsänkor, hotspots för biologisk mångfald och resurser för timmerproduktion. För att kunna hantera skogar på ett hållbart sätt behöver vi övervaka viktiga attribut som trädvolym, biomassa ovan mark, och skogens hälsa. Historiskt sett har denna information samlats in genom markbaserade skogsinventeringar, men tekniska framsteg för att samla in data från sensorer på distans har förändrat processen, gjort den snabbare, mer exakt och oftare repeterbar. Dessa sensorer kan samla in data från olika plattformar, såsom marken, luften och rymden.

Huvudfokus i denna avhandling var att undersöka användningen av tredimensionell (3D) information som extraherats från fjärranalysdata insamlade från olika sensorer, som exempelvis flygburen laserskanning, digital flygfotogrammetri, och syntetisk aperturradar, för att mäta skogsattribut. Dessa verktyg gör det möjligt för skogsförvaltare att samla detaljerade skogsdata över stora områden, vilket förbättrar effektiviteten i skogsövervakning och kartläggning.

Sveriges Riksskogstaxering (NFI), som startade 1923, är ett exempel på hur metoder för skogsmätning och datainsamling har utvecklats över tid. NFI använder en blandning av permanenta och tillfälliga provytor för att samla in skogsdata, och metoder som helt och hållet bygger på manuellt insamlade data används för att uppskatta skogsattribut. Däremot erbjuder modellering baserad på information extraherad från fjärranalysdata bättre möjlighter att ta fram detaljerade beskrivningar av skogsattribut över större områden.

Integrationen av 3D RS-data i olika fall av modellutveckling för skattning av attribut som trädvolym och biomassa ovan mark har undersökts (i Papper I-IV). Även kombinationer av tät och gles flygburen laserskanning -data

utforskades för att skatta skogstillväxt med trädvolym -modeller (i Papper V).

Osäkerhetsskattning är en nödvändig aspekt av skogsinventering, särskilt när man kombinerar fält- och fjärranalys-data. Fel i skogsmätningar kan uppstå från olika källor, inklusive sensorbrus, instrumentfel och modellfel. Dessa fel kan kvantifieras med olika metoder. Några av de osäkerhetsskattningsmetoder som användes i denna avhandling var kvadratroten av medelkvadratfelet, konfidensintervall, och prediktionsintervall. Medelkvadratfelet mäter modellens samlade noggrannhet genom att beräkna kvadratroten av medelvärdet av de kvadrerade skillnaderna mellan skattade och observerade värden. Konfidensintervall ger ett intervall för de uppskattade medelvärdena, medan prediktionsintervall tar hänsyn till både slumpmässig variation i modellparametrarna och nya observationer, vilket gör dem bredare än konfidensintervall och används för att bedöma tillförlitligheten hos dessa skattningar.

Sammanfattningsvis belyser denna avhandling att användningen av 3D fjärranalys-teknik i skogsinventeringspraxis inte bara sparar tid och kostnader utan också ger mer exakta data för storskalig skogsförvaltning, t.ex. i strategisk, taktisk som operativ planering. När dessa teknologier utvecklas vidare har de stor potential att förbättra förvaltningen av skogsresurser.

Acknowledgements

"*When eating fruit, remember the one who planted the tree*"
~ Le Chi Thao; translated by Jan Dodd.

This beautiful quote from the Vietnamese culture struck me when I looked back and thought of every individual who have played a part in my journey of the PhD studies over the past 4 years. There were many who have supported me and morally stood by me throughout this journey, and without whose presence the path wouldn't have been as pleasant and meaningful as it has turned out to be. Firstly, I would like to thank my mother, brother, grandparents, Krypto, and my childhood friends, who I am extremely grateful for always being there for me just a call away and ready to listen to any good or demotivating updates from my day-to-day life. Also, for the efforts to make me feel included in every occasion and celebration that I have missed over the past years.

Surely, this thesis wouldn't have been successfully completed without the constant and immense support, and insightful feedbacks that I have received from my supervisors – Henrik J. Persson, Eva Lindberg, Magnus Ekström, and Mats Nilsson – who were very patient, motivating, and never disregarded any small, big or iterative queries that I have had. I would also like to thank all of the co-authors of the individual studies included in this thesis. I would like to express my gratitude towards the research team from the Norwegian University of Life Sciences, Norway, especially, Prof.Erik Næsset, Terje Gobakken, and Ole M. Bollandsås, for guiding me during and after my PhD exchange period to their group.

Thank you to all of my colleagues from the Department of Forest Resource Management, who have contributed in making this journey noteworthy and knowledgeable. I would like to extend my gratitude towards

the administration staff of the department. Also, thank the housekeeping staff whose efforts make our everyday worklife easier.

A heartfelt gratitude and lots of love to all the former and present PhD and Post-doc colleagues for the endless laughs, ‘intriguing’ lunch-hour conversations, guidance, and enjoyment.

A special appreciation and thanks to the amazing dancers for the nice lessons and dances that I have had at Swingum and U&Me dance academy. It helped me embrace every fun moment and helped restoring positive vibes and energy to keep going on.

On a final note, I extend my deepest gratitude to everyone who has supported and inspired me throughout this journey. Your unwavering belief in me and your invaluable contributions have been a constant source of motivation and strength.

I

remote sensing

MDPI

Technical Note

Mapping and Estimating Aboveground Biomass in an Alpine Treeline Ecotone under Model-Based Inference

Ritwika Mukhopadhyay [1,*], Erik Næsset [2], Terje Gobakken [2], Ida Marielle Mienna [2,3], Jaime Candelas Bielza [2], Gunnar Austrheim [4], Henrik Jan Persson [1], Hans Ole Ørka [2], Bjørn-Eirik Roald [2] and Ole Martin Bollandsås [2]

[1] Department of Forest Resource Management, Swedish University of Agricultural Sciences, 90183 Umeå, Sweden; henrik.persson@slu.se
[2] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, 1433 Ås, Norway; erik.naesset@nmbu.no (E.N.); terje.gobakken@nmbu.no (T.G.); ida.marielle.mienna@nmbu.no (I.M.M.); jaime.candelas.bielza@nmbu.no (J.C.B.); hans-ole.orka@nmbu.no (H.O.Ø.); bjorn-eirik.roald@nmbu.no (B.-E.R.); ole.martin.bollandsas@nmbu.no (O.M.B.)
[3] Geo-Ecology Research Group, Natural History Museum, University of Oslo, 0318 Oslo, Norway
[4] Department of Natural History, Norwegian University of Science and Technology, 7491 Trondheim, Norway; gunnar.austrheim@ntnu.no
* Correspondence: ritwika.mukhopadhyay@slu.se

Abstract: Due to climate change, treelines are moving to higher elevations and latitudes. The estimation of biomass of trees and shrubs advancing into alpine areas is necessary for carbon reporting. Remotely sensed (RS) data have previously been utilised extensively for the estimation of forest variables such as tree height, volume, basal area, and aboveground biomass (AGB) in various forest types. Model-based inference is found to be efficient for the estimation of forest attributes using auxiliary RS data, and this study focused on testing model-based estimations of AGB in the treeline ecotone using an area-based approach. Shrubs (*Salix* spp., *Betula nana*) and trees (*Betula pubescens* ssp. *czerepanovii*, *Sorbus aucuparia*, *Populus tremula*, *Pinus sylvestris*, *Picea abies*) with heights up to about five meters constituted the AGB components. The study was carried out in a treeline ecotone in Hol, southern Norway, using field plots and point cloud data obtained from airborne laser scanning (ALS) and digital aerial photogrammetry (DAP). The field data were acquired for two different strata: tall and short vegetation. Two separate models for predicting the AGB were constructed for each stratum based on metrics calculated from ALS and DAP point clouds, respectively. From the stratified predictions, mean AGB was estimated for the entire study area. Despite the prediction models showing a weak fit, as indicated by their R^2-values, the 95% CIs were relatively narrow, indicating adequate precision of the AGB estimates. No significant difference was found between the mean AGB estimates for the ALS and DAP models for either of the strata. Our results imply that RS data from ALS and DAP can be used for the estimation of AGB in treeline ecotones.

Keywords: aboveground biomass; airborne laser scanning; image matching; model-based inference; treeline vegetation; uncertainty estimation

Citation: Mukhopadhyay, R.; Næsset, E.; Gobakken, T.; Mienna, I.M.; Bielza, J.C.; Austrheim, G.; Persson, H.J.; Ørka, H.O.; Roald, B.-E.; Bollandsås, O.M. Mapping and Estimating Aboveground Biomass in an Alpine Treeline Ecotone under Model-Based Inference. *Remote Sens.* **2023**, *15*, 3508. https://doi.org/10.3390/rs15143508

Academic Editor: Markus Hollaus

Received: 10 May 2023
Revised: 30 June 2023
Accepted: 10 July 2023
Published: 13 July 2023

1. Introduction

Forests play a major role globally as carbon sinks. Hence, afforestation may be important for climate change mitigation through carbon sequestration [1]. However, there are large differences in the capacity of different forest types to sequester carbon and produce biomass. These differences depend on key environmental factors such as the available nutrients in the soils and the climatic conditions. In the Nordic countries, the lowland and low-latitude forests constitute the most productive areas and the largest pools of tree biomass, but even high-elevation ecosystems are known to store large amounts of carbon in vegetation and the soil [2–4].

Climate changes affect the establishment, growth, and mortality of trees and other woody vegetation [5–8], and the effects might be more pronounced in the transition zones between the boreal and alpine zone (i.e., the treeline ecotone), where trees grow at their tolerance limit in terms of temperature compared to productive forests at lower elevations [9,10]. Therefore, in the treeline ecotone, even a moderate increase in temperature or change in precipitation might lead to the increased growth of existing trees and promote the establishment of pioneer trees in currently treeless areas [11,12]. According to Bryn et al. [13], the alpine areas in Norway where pioneer trees potentially could establish constitutes between 25 and 30% of the total land area (excluding bare rock and barren areas). These are potential areas for forest expansion where the potential impact of carbon sequestration is substantial. An increase in tree vegetation into the current alpine areas will also have an impact on albedo [14–16], especially during winter as previously entirely white surfaces will have scattered dark areas which absorb more incoming solar radiation.

There are several other factors apart from climate change that could induce changes in the treeline ecotone, such as herbivory [17,18]. In Norway, where the current study was carried out, it has been common in the past for domestic animals to graze and browse in montane areas during the summer months (summer farming), in some areas more intensively than others. Considering the large areas in play, the decline in summer farming and herbivory can have a substantial effect on the biomass and carbon stocks in high-elevation forests. Previous research seeking to disentangle and quantify the importance of the different factors affecting the treeline ecotone has highlighted the complexity of the causal relationships of treeline dynamics [5,17–19]. Even though both climate and herbivory have been suggested as important factors, they only explain a small proportion of the variation in the observed changes in the treeline ecotone. This likely means that the climatic and herbivory variables fail to fully represent the true effects of these factors and that there are context-dependent local factors that go unaccounted for in such analyses [9,10,20–23]. Since climate-induced responses in the treeline ecotone are difficult to predict, the future development of biomass stock is highly uncertain. A monitoring system where objective data are collected on a regular basis that enables the estimation of actual biomass and carbon stocks is therefore important to fulfil national obligations with regard to carbon reporting.

Accurately estimating and monitoring changes in forest biomass and carbon content is crucial for meeting the requirements of both the Kyoto Protocol and the Paris Agreement [24]. In Norway, the national forest inventory (NFI) is the primary source for biomass and carbon estimates. For forest areas, the sampling grid of the NFI is 3 km × 3 km, but for montane regions, it is 3 km × 9 km [25], which is sparse in terms of providing precise local and overall country-wide estimates. However, with the use of auxiliary remotely sensed (RS) data, the acquisition of data with wall-to-wall coverage for large regions has, in many cases, been shown to be effective in terms of obtaining precise estimates of variables of interest [26–41]. Compared to pure field-based samples, RS data offer the advantage of quickly providing coverage over large areas, and wall-to-wall RS data are often available over the entire area of interest (AOI). Another advantage of RS data is their ease of use in remote or inaccessible areas, which is often the case where treeline ecotones are found.

Numerous studies have demonstrated the versatility of three-dimensional (3D) point cloud data obtained through airborne laser scanning (ALS) for estimating the properties of forest and vegetation (e.g., [34,37,39,42–47]). Næsset and Nelson [48] found that almost all treeline trees with a height of 1 m or taller could be detected using ALS data. However, trees below 1 m in height were often not detected because a laser pulse needs to intercept a minimum surface area before an echo is triggered. With point data, the probability of detecting trees will also depend on the point density. They also found that ground features like rocks could yield echoes with positive height values because they were misclassified as vegetation echoes. Thus, although detecting the smallest trees in the treeline ecotone using ALS presents challenges, there is potential for developing an effective monitoring system for AGB based on ALS, particularly for trees and shrubs taller than 1 m. Despite

the steep decline in detection rates of single trees shorter than 1 m, an area-based approach may still be a viable option for estimating AGB. Some studies have already used bi- and multi-temporal ALS for estimating changes in height among treeline trees [49,50], and the results from these studies are encouraging, showing the potential utility of ALS for monitoring purposes.

The use of 3D point cloud data from digital aerial photogrammetry (DAP) is an alternative to ALS, especially because of the potential finer spatial resolution. Aerial images are cheaper to acquire and can therefore be more frequently collected compared to ALS data [31]. Therefore, DAP data has emerged as an alternative to ALS in operational forest inventories when both cost and utility are taken into account [30,51,52]. Studies that have applied DAP for detection and estimation purposes related to small trees are still few, although Puliti et al. [53] studied the use of aerial images from a drone platform for the estimation of biophysical properties in productive forests under regeneration. In the latter study, it was found that the height values of DAP point clouds tended to underestimate tree height more than those from ALS. It was also found that solitary single trees were more likely to be smoothed from the DAP point cloud. Whether or not these results were directly related to the settings and algorithms used in the matching of the images was not identified. Therefore, further investigation into the use of DAP for vegetation attribute estimations in such transitional zones was recommended.

The uncertainties associated with estimates of state and changes in the AGB using auxiliary RS data have been studied for mature forest stands [29,54–60], montane forests [46], and young forests [61]. The estimation of AGB using auxiliary RS data can be accomplished through design-based or model-based inferences, which have been discussed in detail in Ståhl et al. [33]. The model-based inference is independent of the probability sampling assumptions, which makes it advantageous for inaccessible and remote study sites. However, the accuracy of the estimates is solely dependent on the applied model. Thus, correctly specified models are important to avoid systematic model prediction errors [33,43,62,63]. For informed decision making, it is also important that estimates of precision are provided along with the estimates of AGB. In a study by Næsset et al. [31], the height of treeline vegetation was estimated, and the efficiency of ALS and DAP RS data was compared by characterising the uncertainty in the height estimations using root mean squared error. However, to the best of our knowledge, there have been no studies where the uncertainty of AGB estimates in treeline ecotone sites has been estimated. Therefore, further research is needed to better understand the uncertainties associated with estimating AGB in this region.

Objective

In the current study, AGB for a treeline ecotone site was estimated with corresponding estimates of precision using an area-based approach and model-based inference. Models for field measured AGB were constructed using metrics calculated from both ALS and DAP data. The study area was divided into two strata (tall and short vegetation), and separate models were fitted for each stratum with both ALS and DAP metrics. Along with the estimates of total AGB based on both sets of metrics, standard error estimates were obtained by means of parametric bootstrapping (PB). The main objective was to evaluate and compare the precision of AGB estimates obtained utilising ALS and DAP as auxiliary data and to assess the possibilities of using the area-based approach aided by these RS data sources to estimate AGB in treeline ecotones.

2. Materials and Methods

2.1. Study Area

The AOI (7°58′E, 60°42′N; Figure 1) is located in the municipality of Hol, Norway, at elevations between 1050 and 1320 m above sea level [21]. The dominant treeline species in this region is mountain birch (*Betula pubescens* ssp. *czerepanovii*), along with a few individuals of rowan (*Sorbus aucuparia*), aspen *(Populus tremula*), Scots pine *(Pinus sylvestris*),

and Norway spruce (*Picea abies*). The shrub species considered in this study were *Salix lapponum*, *Salix glauca* subsp. *glauca*, and *Betula nana*. Other woody dwarf shrubs, such as *Vaccinium* spp. and *Empetrum* sp., are present in the area but not considered in the current study. For an overview of all plant species and their abundances within the study area, see [64].

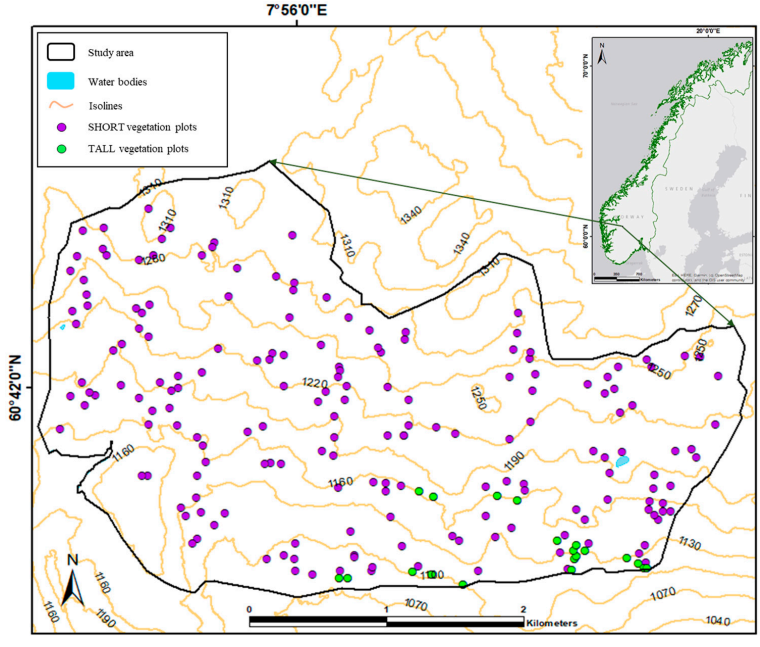


Figure 1. Map of the location of the AOI (Hol, Norway). Locations of field plots (see explanation in Section 2.2—Field methods) used for calibration of models appear as coloured circles along with the isolines.

A graphical overview of the workflow is displayed in the chart in Figure 2. The data collection, modelling, and estimation procedures are discussed further in the following sections.

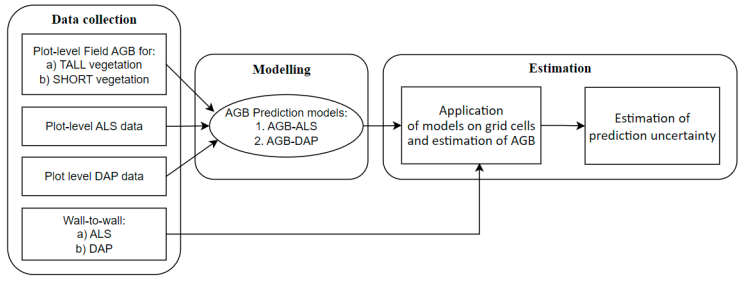


Figure 2. Overview of the methodological workflow for the TALL and SHORT vegetation plots.

2.2. Field Methods

There are both destructive and non-destructive methods available for measuring the aboveground biomass (AGB) of trees; they differ in terms of resource requirements and operational considerations [65,66]. In this study, our decision to employ a non-destructive method for measuring AGB values in the field was primarily driven by resource limitations. Destructive sampling, while providing more precise biomass estimates, would have required significant additional resources that were not available to us. Furthermore, in an operational setting, where resource allocation is constrained, the use of destructive sampling would not be feasible or practical. Lastly, it is worth noting that, in this specific case, we also faced restrictions and did not have permission to perform destructive sampling. As a result, we utilised a predictive approach, leveraging field measurements to estimate the total AGB for each plot. Despite the inherent limitations, this approach enabled us to derive biomass estimates effectively while avoiding the challenges associated with destructive sampling. The field data were collected partly using wall-to-wall ALS data to guide the location of the sample plots. The area was tessellated into 100 m^2 grid cells, and the average height (H_{mean}) of first echoes of the ALS pulses within each grid cell was calculated. Based on the H_{mean} values, the study area was split into two strata. Grid cells where $H_{mean} < 1$ m were categorised as 'short woody vegetation' (SHORT), and grid cells where $H_{mean} \geq 1$ m were categorised as 'tall vegetation' (TALL). After classification into SHORT and TALL, the SHORT grid cells were further tessellated into 16 equally sized squares (6.25 m^2).

Twenty TALL grid cells with H_{mean} values evenly distributed over the entire range of H_{mean} values were selected to establish field reference plots to initiate modelling. The selection process involved dividing the H_{mean} range between 1 m and the maximum value into 10 equally sized bins. Then, we randomly selected two grid cells from each bin. The fieldwork was carried out in summer 2019. A Topcon HiPer SR geodetic-grade GNSS receiver in real-time kinematic (RTK) mode was used to navigate to the centre of each selected grid cell, where a circular plot with a radius of 5.64 m (area = 100 m^2) was established. Within each plot, all trees with diameter at breast height (dbh) > 0 were calipered. Breast height was defined at 1.3 m above ground. Tree height was measured using a folding rule or a Vertex hypsometer on two subjectively selected height-sample trees in each plot. The selection was made so that the 40 selected trees (two trees per plot) covered a range of dbh values. In addition, a 1.5 m radius (7.07 m^2) sub-plot was established at the centre of each plot to sample shrubs and tree saplings with a height (h) < 1.3 m. All tree species were sampled, including juniper (*Juniperus communis*), as well as the shrubs species willow (*Salix* spp.) and dwarf birch (*Betula nana*). The sub-plot was divided into four quadrants by two perpendicular lines that intersected at the plot centre in north–south and east–west directions (Figure 3). In each quadrant, the tree sapling or shrub closest to each of the points 1 m from the plot centre in the directions that were 45, 135, 225, and 315 degrees relative to north (Figure 3) was measured for height and diameter at root collar. In the southeast and northwest quadrants, the number of saplings and shrubs were counted.

The field measurement for the SHORT stratum took advantage of 180 previously established vegetation monitoring plots [20]. The original vegetation plots were squares of 0.25 m^2, but we established 1.5 m radius plots in the NW corner of each square. Two additional plots were purposely established in tall *Salix* vegetation, since this was poorly covered by the 180 prepositioned plots. The same measurements as those described for the sub-plots of the 20 plots of the tall vegetation were carried out. The position at the centre of each plot was registered with the same Topcon receiver described above in RTK-mode.

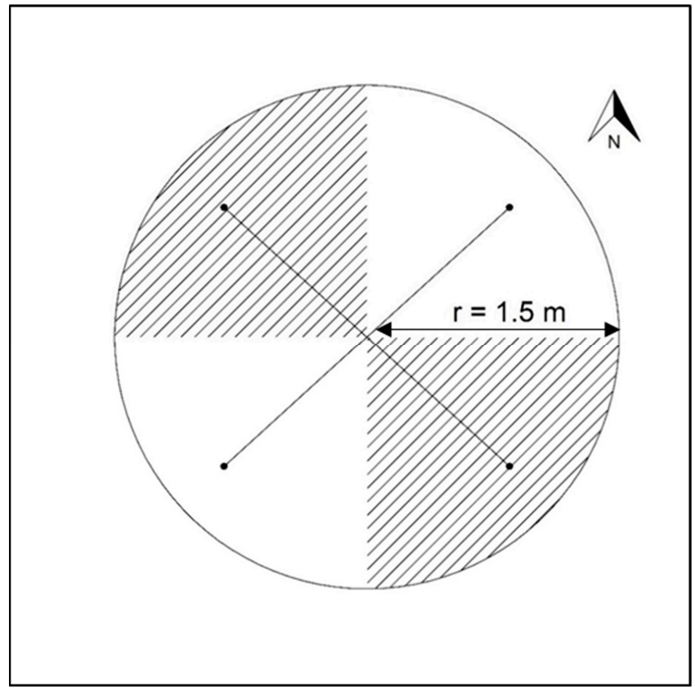


Figure 3. Visual representation of the 1.5 m radius plot for the sampling of shrubs and tree saplings with h < 1.3 m. In the NW and SE quadrants (hatched areas), all shrubs and saplings were counted. Four heights were sampled, one in each quadrant, as the closest individual to a point 1 m from the plot centre in directions 45, 135, 225, and 315 degrees, respectively.

For each plot, total AGB was predicted as the sum of individual tree AGB, following the same procedure as in [67,68]. For the TALL plots, the predictions were separate for trees with h < 1.3 m and those with h $\geq$ 1.3 m. For trees with h $\geq$ 1.3 m, the base heights were first predicted using the base height models of Fitje et al. [69]. Then, single-tree biomass models [70] with field-measured dbh and predicted base height as inputs were used to predict the base biomass values (b1) of each tree. For each of the height-sample trees, a second biomass (b2) value was predicted using measured dbh and measured h. Then, a common correction factor k for the base-biomass was calculated from all the height sample trees as the ratio between the sum of b2 and the sum of b1. Biomass for each tree was predicted as b1 $\times$ k. Single-tree biomass predictions were summed for each plot and scaled to a per hectare value. For trees and shrubs with h < 1.3 m observed on the 1.5 m radius sub-plots, a mean diameter at the root collar and a mean height were calculated from the measurements in each plot. These mean values were used as inputs in the single-tree biomass model of Kolstad et al. [71] to predict mean-tree biomass and then multiplied with the number of individuals and scaled to a per hectare value. The predicted biomass for trees with h < 1.3 m was added to the biomass prediction for the taller trees. For the SHORT vegetation plots, the total biomass was calculated using the same procedure used for the 1.5 m radius sub-plots of the TALL vegetation stratum. A summary of field reference data is presented in Table 1.

Table 1. Summary of field reference data collected in 2019.

Stratum	*n*	Mean AGB (Mg ha^{-1})	SD (Mg ha^{-1})	Min (Mg ha^{-1})	Max (Mg ha^{-1})
TALL	20	30.2	17.5	8.48	65.8
SHORT	182	1.61	2.68	0.00	18.3

2.3. Remotely Sensed Data

2.3.1. Data Acquisition and Initial Processing

The sensor and flight information for the RS data is provided in Table 2. ALS data were collected by the contractor, Terratec AS, as part of national scanning, with a point density of 2 points/m^2. The raw point clouds were pre-processed by the contractor, and laser echoes were classified as "ground", "unclassified", "noise", "bridge", or "snow". Planimetric coordinates and orthometric heights were computed for all echoes. A triangulated irregular network (TIN) was then generated from the laser echoes classified as "ground" using the R package "lidR" [72,73]. Since the study aimed to provide estimates of the biomass of shrubs and trees whose lower stems often grow parallel to the ground due to snow pressure, all vegetation echoes were included without setting a threshold on echo height. Thresholds of <2 m have commonly been used in previous studies to omit falsely classified vegetation echoes [50,74,75].

Table 2. Summary of sensor and flight information for ALS and DAP.

	ALS	DAP
Sensor system	Riegl VQ-1560i	Sensefly S.O.D.A. camera
Platform	Piper PA-31-350 Chieftain	Sensefly eBee
Acquisition dates	8 and 25 June 2018	7–10 July 2019
Flight altitude (m a.g.l) *	3400	120
Flight speed (m s^{-1})	NA	15
Point repetition frequency (KHz)	350	NA
Scan frequency (Hz)	162	NA
Point density (points m^{-2})	2	55
Half scan angle (degrees)	20	NA

* above ground level.

The DAP data were collected following a flight plan with perpendicular flight lines with lateral and longitudinal overlaps of 80%. Across the entire study area, 43 orange wooden crosses (30 cm width) were distributed uniformly and positioned and used as ground control points (GCP) to enable computation of the position and orientation of the images. The DAP point cloud was thinned from a mean density of 55 points m^{-2} to 38 points m^{-2} using the R package "lidR" [72,73] to obtain a uniform point density over the entire area. Examples of ALS and DAP point clouds are presented in Figure 4 for two grid cells, one in each of the TALL and the SHORT strata. The higher point density of the DAP point clouds is clearly visible.

2.3.2. Correction of the DAP Point Cloud

An initial inspection of the normalised DAP point cloud revealed many points with negative normalised height values. Regions with no observable vegetation, meaning they should have heights of 0 m, were observed to contain points with both positive and negative heights. These were observed to be in the magnitude of 0–2 m. This indicated that the error was not simply a global vertical shift and that this error was too large to be solely due to the inaccuracy of the DAP system. The most likely culprit was the positioning accuracy of the GCP markers. Positioning errors could result in either rotation or translation in the DAP registration and thus errors in both the horizontal and vertical directions.

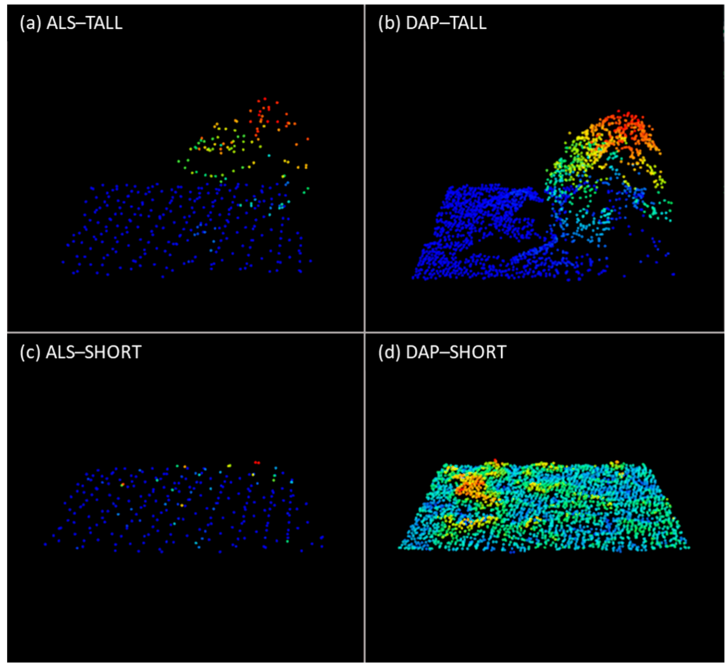


Figure 4. The point clouds for one 100 m^2 area in each stratum with a side-by-side comparison of point clouds obtained by ALS and DAP, respectively. The heights above ground for the points are illustrated using a colour scale from blue (ground) to red (highest point). The different panels display (**a**) ALS–TALL, (**b**) DAP–TALL, (**c**) ALS–SHORT, and (**d**) DAP–SHORT.

The ALS and DAP point clouds did not share any GCP markers. Furthermore, the density of the ALS data made the extraction of natural terrain features that could be applied as GCPs too inaccurate. As a result, the DAP point cloud registration could not be corrected by simply using shared tie points in the ALS and DAP point clouds. As an alternative to this approach, the local vertical errors between the point clouds were estimated and used to correct the original DAP point cloud. This approach can be summarised in three steps: (i) ground points in the DAP point cloud were classified using the software TerraScan [76], (ii) the ALS ground point heights were interpolated at the XY coordinates of the DAP ground points so that the vertical shift between ALS and DAP ground heights could be computed, and (iii) the correction of the entire DAP point cloud was interpolated from the results of the previous step. The interpolation was performed using the "scipy.interpolation.griddata" Python package. If the point to be interpolated was located inside the convex hull of the input points, linear interpolation was used; otherwise, nearest-neighbour interpolation was employed.

2.3.3. Computation of Metrics

For both the ALS and DAP point clouds, metrics that represented the height distribution and the density of the point cloud for each plot and grid cell were computed using the echoes classified as "first of many" and "single". The normalised ALS and DAP point clouds were clipped to the spatial extent of the plots and grid cells, and the points with negative heights were removed from further processing for the TALL stratum. For the SHORT stratum, the DAP points with negative heights were kept to account for small height shifts

still present in the point cloud after correction. We considered this to be important because, for the SHORT stratum, most of the AGB was growing close to the ground. Removing negative heights could possibly equalise the variation in the DAP point cloud in areas where AGB actually varies. The metrics representing height were heights at the deciles of the echo height distribution. The densities were the ratios between the number of echoes above 10 different height-levels and the total number of echoes. In addition, the standard deviation of the heights, skewness, and kurtosis were computed; all together, 23 metrics for each of the ALS and DAP point clouds were considered.

2.4. Model Construction

Since AGB was small within our study area, the values of both the ALS and DAP metrics were zero for many of the field plots, which was a limitation concerning variable transformations and model forms. After some preliminary tests with different modelling approaches, such as zero-inflated and log-transformation of the response variable, we chose to construct linear models fitted using ordinary least squares. Linear models were suitable in this case because the training dataset contained extreme AGB values, which reduced the need for extrapolation. The models were constructed separately for the TALL and SHORT strata. For the TALL stratum, the explanatory variables for the ALS and DAP datasets were separately selected through a stepwise procedure based on the Bayesian information criterion (BIC) and adj-R^2 [77]. The general form of the models for the TALL stratum was as follows:

$$\mathrm{AGB} = \beta_0 + \beta_1 X_1 + \cdots + \beta_j X_j + \varepsilon \quad (1)$$

where j is the number of explanatory variables (X), β_0, β_1, ..., β_j are the parameter estimates, and ε is the random error term.

As indicated in Section 2.2, there was a difference in the area of the field plots for the short vegetation (7.07 m^2) and the area of the grid cells (6.25 m^2). Thus, a scale-independent explanatory variable was purposely chosen for the models for the SHORT stratum. The metric chosen for both the ALS and DAP models was the mean point height (H_{mean}), which was retained even if the parameter estimate was statistically non-significant.

For the SHORT stratum, the models were formulated as follows:

$$\mathrm{AGB} = \beta_0 + \beta_1 H_{mean} + \varepsilon \quad (2)$$

where H_{mean} is the average height of either the ALS or DAP points, β_0 and β_1 are the model coefficients, and ε is the random error term.

The models' training accuracies were assessed using the root mean square residual error (RMSE) and relative RMSE (rel.RMSE) [78]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{\mathrm{AGB}}_i - \mathrm{AGB}_i\right)^2}{n}} \quad (3)$$

$$\mathrm{rel.RMSE} = \frac{\mathrm{RMSE}}{\sum_{i=1}^{n}(\mathrm{AGB}_i)/n} \times 100 \quad (4)$$

where $\widehat{\mathrm{AGB}}_i$ is the model-predicted AGB for field plot i, AGB_i is the corresponding field reference AGB, and n is the number of observations in the reference data used to train the models.

2.5. Estimation of Mean AGB

The final AGB—ALS and AGB—DAP models were applied to the grid cells belonging to the stratum for which they were constructed. Mean AGB was estimated separately for both strata and with both models ($\overline{\widehat{\mathrm{AGB}}}_{TALL}$, $\overline{\widehat{\mathrm{AGB}}}_{SHORT}$) as the means of the respective stratified AGB predictions. Mean AGB values for the entire area ($\overline{\widehat{\mathrm{AGB}}}_{H}$) using both models

were estimated as area weighted means for both strata individually for ALS and DAP as [79]:

$$\overline{\widehat{\mathrm{AGB}}}_H = \frac{\overline{\widehat{\mathrm{AGB}}}_{TALL} \cdot A_{TALL} + \overline{\widehat{\mathrm{AGB}}}_{SHORT} \cdot A_{SHORT}}{A_H} \quad (5)$$

where A is the area, and the subscripts *TALL*, *SHORT*, and H denote the TALL stratum, SHORT stratum, and the entire area, respectively.

2.6. Variance Estimation via Parametric Bootstrapping

Estimation of the standard error (*se*) for the AGB estimates was carried out using parametric bootstrapping (PB). PB is based on Monte Carlo simulation and is convenient within the model-based inference framework [80]. It enables the statistical inference of the variable of interest when the true distribution is unknown [81], and the Monte Carlo errors become negligible when the PB samples are large [46]. In this study, the number of bootstrap samples (N_{PB}) was set to 50,000, and for each iteration (k), the mean of AGB predictions ($\overline{\widehat{\mathrm{AGB}}}_k$) was calculated. The standard error for the AGB predictions ($\widehat{se}$) was obtained separately for each stratum as [46]:

$$\widehat{se} = \sqrt{\frac{1}{N_{PB}-1}\sum_{k=1}^{N_{PB}}\left(\overline{\widehat{\mathrm{AGB}}}_k - \overline{\widehat{\mathrm{AGB}}}_{PB}\right)^2} \quad (6)$$

where $\overline{\widehat{\mathrm{AGB}}}_{PB}$ is the mean of the $\overline{\widehat{\mathrm{AGB}}}_k$ values of the N_{PB} samples. To assess whether the number of iterations was sufficient to accurately estimate the uncertainty of AGB predictions, we calculated a standard error stabilisation indicator (se.s). This indicator evaluates the change in the standard deviation of mean predicted AGB as additional bootstrap iterations are conducted. The se.s value was determined after each iteration as the maximum difference between the standard deviation of mean predicted AGBs obtained during the last 50% of iterations and the standard deviation of all mean predicted AGBs obtained in the final iteration. According to the authors of [80], an se.s of <0.5% of the *se* after the last iteration indicates that the *se* has stabilised sufficiently.

The area-weighted standard error for the entire study area ($\widehat{se}_H$) combining both stratum estimates can be estimated as [79]:

$$\widehat{\mathrm{se}}_H = \sqrt{\frac{\left(\widehat{\mathrm{se}}^2_{TALL} \cdot A^2_{TALL} + \widehat{\mathrm{se}}^2_{SHORT} \cdot A^2_{SHORT}\right)}{A^2_H}} \quad (7)$$

The 95% confidence intervals (CI) were obtained for the mean estimate of predicted AGB for the respective strata and the total area as [82]:

$$\mathrm{CI} = \overline{\widehat{\mathrm{AGB}}} \pm (\widehat{se} \times 1.96) \quad (8)$$

$$\mathrm{CI}_H = \overline{\widehat{\mathrm{AGB}}}_H \pm (\widehat{se}_H \times 1.96) \quad (9)$$

3. Results

For both of the TALL stratum models, the 9th height decile (H_{90}) was selected as the only explanatory variable, while for the SHORT stratum, the scale-independent H_{mean} was selected. The estimated parameter coefficients for all models are displayed in Table 3 along with the corresponding adjusted R^2 (adj-R^2) and root mean squared error (RMSE). For all models, the adj-R^2 values were relatively small, indicating moderate to weak model fit. The relatively large proportion of unexplained AGB variation can also be seen in Figure 5, where field-observed AGB is plotted against the model fitted values for all four models. There were no large differences in model fit between the models based on ALS and DAP

data, but substantial model fit differences between the strata were observed. Table 3 also shows that the slope parameter values for the prediction models are smaller for the DAP model for both strata compared to those of the ALS models, indicating that the two point clouds have different height distributions.

Table 3. The table displays the explanatory variables, model coefficients, adj-R^2, RMSE, and relative RMSE (rel.RMSE) of the AGB prediction models for both strata and data sources.

Stratum	Model	Explanatory Variable	Prediction Model	adj-R^2	RMSE (Mg ha^{-1})	rel.RMSE (%)
TALL	AGB-ALS	H_{90}	$-15.09 + 11.64H_{90}$	0.47	12.4	41.1
	AGB-DAP	H_{90}	$-0.388 + 8.25H_{90}$	0.43	12.8	43.8
SHORT	AGB-ALS	H_{mean}	$0.74 + 32.3H_{mean}$	0.15	2.47	154.2
	AGB-DAP	H_{mean}	$0.96 + 28.41H_{mean}$	0.27	2.28	118.1

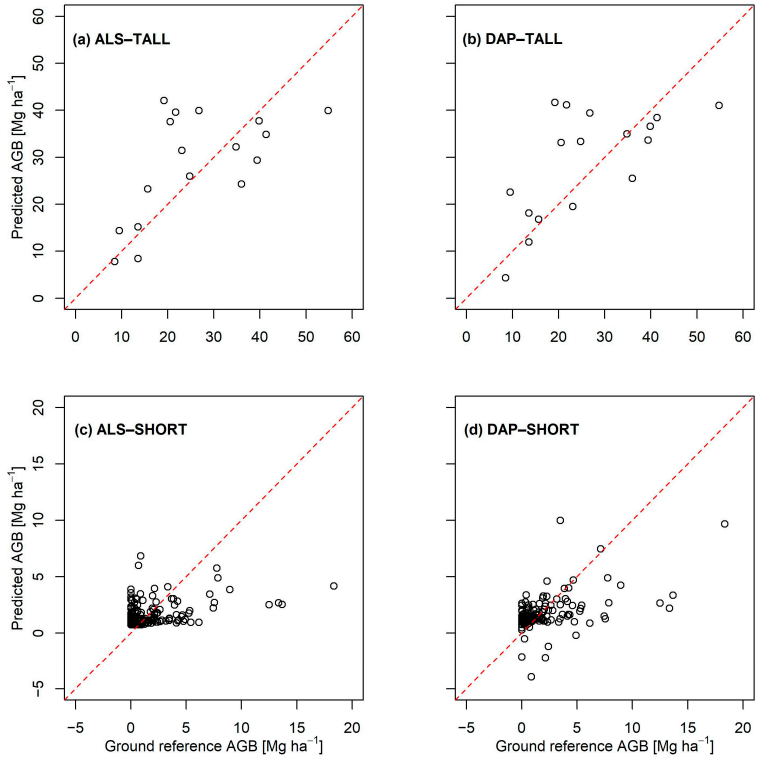


Figure 5. Ground reference AGB plotted against model-predicted AGB for all models: (**a**) ALS–TALL, (**b**) DAP–TALL, (**c**) ALS–SHORT, and (**d**) DAP–SHORT.

Figure 6 displays the se.s after each bootstrap iteration for both model types and both strata. In all data types and stratum combinations, the se.s dropped below 0.5% before reaching 50,000 iterations. However, in some instances and to a varying degree, se.s increased in value after the stabilisation criterion was first reached; however, at the final iteration of all four simulations, the se.s remained below the threshold of 0.5% for at least 10,000 of the last iterations.

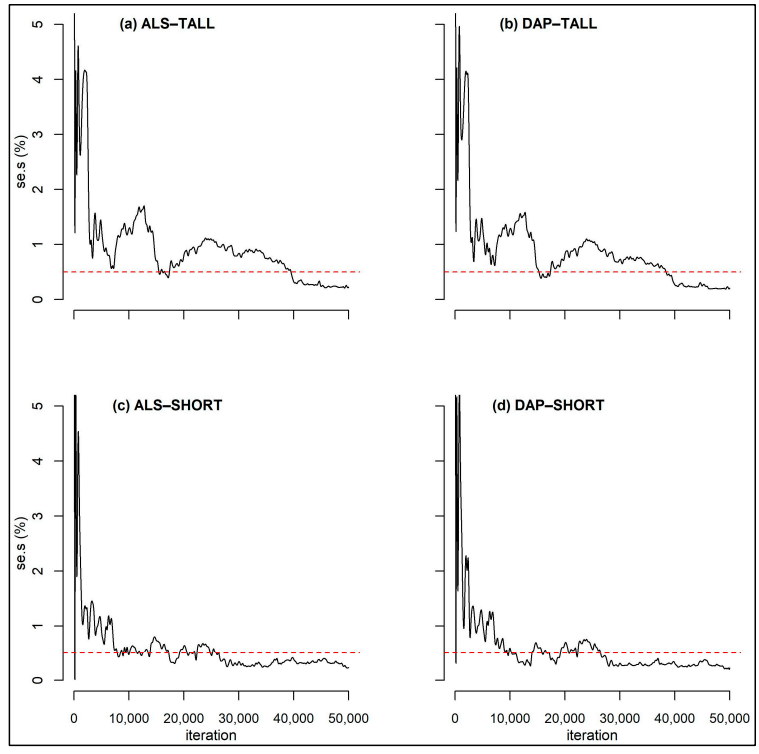


Figure 6. The standard error stabilisation indicator (se.s) after each of the 50,000 bootstrap iterations for (**a**) ALS–TALL, (**b**) DAP–TALL, (**c**) ALS–SHORT, and (**d**) DAP–SHORT. To enhance the display of se.s close to the critical value of se.s stabilisation, the *y*-axis has been truncated at se.s = 5% and the red horizontal line represents 0.5%.

Table 4 displays the results from the estimations for each stratum and model, as well as for the entire study area. There were no significant differences between the values of $\widehat{\overline{AGB}}$ between models or within strata. Due to the large difference in the size of the area regarding the two strata, the area-weighted values of $\widehat{\overline{AGB}}$ for the entire study area were similar to those for the SHORT stratum.

Table 4. Estimation results of the AGB prediction models. $\widehat{\overline{AGB}}$ is the estimated mean, $\hat{se}$ the corresponding standard error estimate, and $\widehat{CI}$ the 95% confidence interval.

Stratum *	Model	$\widehat{\overline{AGB}}$ (Mg ha^{-1})	$\hat{se}$ (Mg ha^{-1})	$\widehat{CI}$ (Mg ha^{-1})
TALL	AGB-ALS	26.5	3.16	23.4–29.7
A = 0.97 ha	AGB-DAP	29.2	3.13	22.9–35.5
SHORT	AGB-ALS	2.05	0.20	1.66–2.45
A = 270.75 ha	AGB-DAP	1.93	0.17	1.59–2.27
Total	AGB-ALS	2.14	0.39	1.75–2.53
A = 271.72 ha	AGB-DAP	2.03	0.34	1.68–2.37

* A = area in ha.

4. Discussion

This study employed a model-based approach with data from ALS and DAP as auxiliary information to estimate AGB and its corresponding precision. In contrast to other comparable studies [46,74], our AOI was exclusively covered by shrubs and short trees. As a result, this research has expanded the boundaries of RS technology in terms of supporting AGB estimation. Despite the small AGB values in our AOI, the widths of the CIs were encouraging in terms of operationalising AGB estimation in the treeline ecotone using both ALS and DAP as auxiliary information. The 95% CIs around the estimated mean AGBs overlapped for ALS and DAP in both strata (Table 4 and Figure 5), indicating that the estimates derived from ALS and DAP were not statistically significantly different. However, the relative uncertainties were larger compared to those obtained in previous studies for productive (e.g., [40,41,54,83]) and montane forests [46] due to the irregular shapes of stems and tree crowns in our AOI, which created weaker relationships between AGB and the RS data and led to higher relative measurement errors. Compared to productive and montane forests, where trees are substantially taller than in our AOI, the relative magnitude of the measurement errors of the RS data was also expected to be larger.

Preprocessing of the DAP point cloud was carried out to correct varying degrees of shifts in point heights or horizontal errors over the study area. Since our AOI comprised mostly short vegetation, this preprocessing was deemed highly important. The corrected values of the original DAP ground points were centred around 0, and both negative and positive corrections were carried out, indicating that the approach was at least partly successful at correcting erroneous height values at ground level. Errors still present after correction were considered to be random. A weakness of our approach was that it assumed vertical errors only. However, errors in the XY directions also must be expected, although they were most likely small. The impact of such potential horizontal errors would have been larger with a reduced size of the field reference plots used for modelling [84,85] and would have been manifested by larger uncertainties in the model parameters. The uncertainty will increase with increasing spatial heterogeneity of the vegetation over the AOI. In the current study, there is reason to believe that such errors had the greatest impact on the SHORT stratum since the field plot size was small. In the prediction phase, the horizontal errors will be less important since there is no direct requirement of spatial coherence. However, the precision of the predictions would still be affected by the errors associated with the model.

Judging by adj-R^2, the model relationships were weaker in the current study compared to studies carried out on boreal productive forests (e.g., [34]) mainly composed of conifer trees with regular conical crown shapes. Due to the irregularity and large variability of stem and crown shapes in our AOI, in addition to our population comprising mainly of shrub species and small individual trees, this difference was not surprising. In the SHORT stratum, most of the trees and shrubs were short and extended their stems partly along the ground, whereas the trees that constituted the TALL stratum were often snow-bent with stems that were both crooked and growing partly along the ground. As a result, a large part of the AGB was found within the uncertainty range of the TIN surface, which is often around 20–30 cm [86].

The positional errors of the plots would have had similar effects as those from the horizontal shift in the point cloud, as discussed above. However, in the open montane landscape of our AOI, the conditions for GNSS positioning were generally excellent, reducing the potential for positional errors. Nonetheless, potential positional errors are of greater importance for the SHORT stratum than the TALL stratum due to the smaller plot size used.

Furthermore, vegetation near the plot boundary can cause edge effects [85] as trees and shrubs rooted just outside the plots may intersect with the plots. In such cases, the trees or shrubs would not be measured in the field; therefore, they would not be part of the measured AGB, but they would still affect the properties of the point clouds for these plots. Conversely, trees and shrubs inside the plots close to the plot edge could lean out of the

plot and still be included in the AGB estimate, affecting the properties of the point clouds of these plots but only to a small degree. Therefore, such edge effects may generally be larger in the treeline ecotone due to the irregularly shaped stems that are seldom vertical. Like the positional errors, edge effects will be relatively larger with decreasing plot size. The effects from trees leaning in and out of the plot, as well as growing inside and outside the plot, will decrease with an increase in the spatial homogeneity of the trees and shrubs.

The graphical display of the AGB se.s (%) at each iteration (Figure 5a–d) indicates that the number of iterations were sufficient for both the TALL and SHORT strata. For all model and stratum combinations, the se.s oscillated around the 0.5% stabilisation criterion after around 10,000 iterations. However, for the SHORT and TALL strata, the se.s values were not stable until after iterations 30,000 and 40,000, respectively. As many as 50,000 bootstrap iterations are quite rare in comparable studies where biomass or other forest attributes are estimated [29,87,88]. The reason is for this is partly because many studies do not use an evaluation criterion for standard error stabilisation. However, the required number of iterations depends on several factors such as the sample size and variability of the original dataset used to fit the model, the complexity of the prediction model used, the desired level of precision of the se estimate, and the uncertainty of the model parameters [89]. In our study, the estimates of the model parameters were characterised by a high level of uncertainty, resulting in a large number of necessary iterations.

Previous studies carried out in productive forests [30,50] have shown that models relying on ALS metrics generally outperform those relying on DAP metrics [52] in terms of prediction accuracy. However, based on the 95% confidence intervals, our study did not find evidence of such a difference. While the R^2 values indicated weaker relationships between the AGB and DAP metrics compared to the corresponding relationships with the ALS metrics, these weaker relationships were not reflected in the uncertainty estimates. Additionally, the smaller slope parameter values of the DAP models for both strata compared to those of ALS suggest that the DAP point cloud primarily represents the vegetation surface without penetrating the canopies [30,31].

For the modelling, the challenges with the site conditions discussed above meant that we had to choose a robust and simple modelling approach [74]. Even though models for biomass based on point cloud metrics usually include metrics representing both forest height and density, our models only included one statistically significant explanatory variable (H_{90} or H_{mean}). Thus, the variation in tree and shrub density, as expressed by the density metrics, was weakly correlated to AGB, and most of the variation could be explained by just the height metrics of our AOI. This might make the models less useful in cases where extrapolation outside the range of variability of the reference data is needed. In our case, however, extrapolation was not extensive as extreme observations were purposely included in our reference data by the way the plots were selected using the ALS data as prior information to guide the sample selection. This was especially important since we relied on linear models [90]. Such a strategy could be adopted in any study area where RS data are available prior to fieldwork.

5. Conclusions

This study indicates that both ALS and DAP data can be used as auxiliary information in area-based AGB estimation with respect to treeline ecotones. We found no significant differences in terms of precision using the two data sources, both at the individual stratum level and for the total area. Although, based on the RMSE and rel.RMSE values, the models for the TALL stratum were more precise than those of the SHORT stratum, the 95% CIs for the models had similar widths in relative terms. This research indicates that area-based biomass estimation can also be carried out operationally in treeline ecotones. However, further studies should be carried out to assess if the proposed method can support repeated surveys for the purpose of AGB change estimation.

Author Contributions: Data acquisition, G.A., I.M.M. and O.M.B.; conceptualisation, E.N., O.M.B. and T.G.; methodology, O.M.B.; formal analysis, O.M.B. and R.M.; data curation, B.-E.R., H.O.Ø., J.C.B., O.M.B. and T.G.; writing—original draft preparation, R.M.; writing—review and editing, B.-E.R., E.N., G.A., H.J.P., H.O.Ø., I.M.M., J.C.B., O.M.B., R.M. and T.G.; visualisation, O.M.B., J.C:B. and R.M.; supervision, H.J.P. and O.M.B.; project administration, O.M.B.; funding acquisition, E.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Research Council of Norway, grant number 281066.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank the three anonymous reviewers for their useful comments that helped improve this article. We would also like to thank Mats Nilsson, Eva Lindberg, and Magnus Ekström from SLU for providing feedback on the first draft of the article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

FAO. *The State of World's Forests 2018—Forest Pathways to Sustainable Development*; FAO: Rome, Italy, 2018; ISBN 9789251305614.
Serreze, M.C.; Walsh, J.E.; Chapin, F.S.; Osterkamp, T.; Dyurgerov, M.; Romanovsky, V.; Oechel, W.C.; Morison, J.; Zhang, T.; Barry, R.G. Observational Evidence of Recent Change in the Northern High-Latitude Environment. *Clim. Chang.* **2000**, *46*, 159–207. [CrossRef]
Hassol, S.J. *Impacts of a Warming Arctic-Arctic Climate Impact Assessment*; Cambridge University Press: New York, NY, USA, 2005.
Speed, J.D.M.; Martinsen, V.; Hester, A.J.; Holand, Ø.; Mulder, J.; Mysterud, A.; Austrheim, G. Continuous and Discontinuous Variation in Ecosystem Carbon Stocks with Elevation across a Treeline Ecotone. *Biogeosciences* **2015**, *12*, 1615–1627. [CrossRef]
Setten, G.; Austrheim, G. Changes in Land Use and Landscape Dynamics in Mountains of Northern Europe: Challenges for Science, Management and Conservation. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* **2012**, *8*, 287–291. [CrossRef]
Menezes-Silva, P.E.; Loram-Lourenço, L.; Alves, R.D.F.B.; Sousa, L.F.; da Almeida, S.E.S.; Farnese, F.S. Different Ways to Die in a Changing World: Consequences of Climate Change for Tree Species Performance and Survival through an Ecophysiological Perspective. *Ecol. Evol.* **2019**, *9*, 11979. [CrossRef]
Hartmann, H.; Bastos, A.; Das, A.J.; Esquivel-Muelbert, A.; Hammond, W.M.; Martínez-Vilalta, J.; Mcdowell, N.G.; Powers, J.S.; Pugh, T.A.M.; Ruthrof, K.X.; et al. Climate Change Risks to Global Forest Health: Emergence of Unexpected Events of Elevated Tree Mortality Worldwide. *Annu. Rev. Plant Biol.* **2022**, *73*, 673–702. [CrossRef] [PubMed]
Taccoen, A.; Piedallu, C.; Seynave, I.; Gégout-Petit, A.; Gégout, J.C. Climate Change-Induced Background Tree Mortality Is Exacerbated towards the Warm Limits of the Species Ranges. *Ann. For. Sci.* **2022**, *79*, 23. [CrossRef]
Körner, C.; Paulsen, J. A World-Wide Study of High Altitude Treeline Temperatures. *J. Biogeogr.* **2004**, *31*, 713–732. [CrossRef]
Gray, S.T.; Betancourt, J.L.; Jackson, S.T.; Eddy, R.G. Role of Multidecadal Climate Variability in a Range Extension of Pinyon Pine. *Ecology* **2006**, *87*, 1124–1130. [CrossRef]
Kullman, L. Tree Line Population Monitoring of Pinus Sylvestris in the Swedish Scandes, 1973–2005: Implications for Tree Line Theory and Climate Change Ecology. *J. Ecol.* **2007**, *95*, 41–52. [CrossRef]
Kullman, L. Late Holocene Reproductional Patterns of Pinus Sylvestris and Picea Abies at the Forest Limit in Central Sweden. *Can. J. Bot.* **2011**, *64*, 1682–1690. [CrossRef]
Bryn, A.; Strand, G.H.; Angeloff, M.; Rekdal, Y. Land Cover in Norway Based on an Area Frame Survey of Vegetation Types. *Nor. Geogr. Tidsskr.* **2018**, *72*, 131–145. [CrossRef]
Loranty, M.M.; Berner, L.T.; Goetz, S.J.; Jin, Y.; Randerson, J.T. Vegetation Controls on Northern High Latitude Snow-Albedo Feedback: Observations and CMIP5 Model Simulations. *Glob. Chang. Biol.* **2014**, *20*, 594–606. [CrossRef] [PubMed]
Te Beest, M.; Sitters, J.; Ménard, C.B.; Olofsson, J. Reindeer Grazing Increases Summer Albedo by Reducing Shrub Abundance in Arctic Tundra. *Environ. Res. Lett.* **2016**, *11*, 125013. [CrossRef]
Ramtvedt, E.N.; Bollandsås, O.M.; Næsset, E.; Gobakken, T. Relationships between Single-Tree Mountain Birch Summertime Albedo and Vegetation Properties. *Agric. For. Meteorol.* **2021**, *307*, 108470. [CrossRef]
Mienna, I.M.; Speed, J.D.M.; Klanderud, K.; Austrheim, G.; Næsset, E.; Bollandsås, O.M. The Relative Role of Climate and Herbivory in Driving Treeline Dynamics along a Latitudinal Gradient. *J. Veg. Sci.* **2020**, *31*, 392–402. [CrossRef]
Mienna, I.M.; Austrheim, G.; Klanderud, K.; Bollandsås, O.M.; Speed, J.D.M. Legacy Effects of Herbivory on Treeline Dynamics along an Elevational Gradient. *Oecologia* **2022**, *198*, 801–814. [CrossRef]
Hansson, A.; Dargusch, P.; Shulmeister, J. A Review of Modern Treeline Migration, the Factors Controlling It and the Implications for Carbon Storage. *J. Mt. Sci.* **2021**, *18*, 291–306. [CrossRef]
Speed, J.D.M.; Austrheim, G.; Hester, A.J.; Mysterud, A. Experimental Evidence for Herbivore Limitation of the Treeline. *Ecology* **2010**, *91*, 3414–3420. [CrossRef] [PubMed]

Speed, J.D.M.; Austrheim, G.; Hester, A.J.; Mysterud, A. Growth Limitation of Mountain Birch Caused by Sheep Browsing at the Altitudinal Treeline. *For. Ecol. Manag.* **2011**, *261*, 1344–1352. [CrossRef]
Speed, J.D.M.; Austrheim, G.; Hester, A.J.; Mysterud, A. Browsing Interacts with Climate to Determine Tree-Ring Increment. *Funct. Ecol.* **2011**, *25*, 1018–1023. [CrossRef]
Austrheim, G.; Solberg, E.J.; Mysterud, A. Spatio-Temporal Variation in Large Herbivore Pressure in Norway during 1949-1999: Has Decreased Grazing by Livestock Been Countered by Increased Browsing by Cervids? *Wildlife Biol.* **2011**, *17*, 286–298. [CrossRef] [PubMed]
UNFCCC. UNFCCC Kyoto Protocol—Targets for the First Commitment Period. Available online: https://unfccc.int/process-and-meetings/the-kyoto-protocol/what-is-the-kyoto-protocol/kyoto-protocol-targets-for-the-first-commitment-period (accessed on 12 May 2021).
Breidenbach, J.; Granhus, A.; Hylen, G.; Eriksen, R.; Astrup, R. A Century of National Forest Inventory in Norway—Informing Past, Present, and Future Decisions. *For. Ecosyst.* **2020**, *7*, 46. [CrossRef]
Næsset, E. Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [CrossRef]
Gobakken, T.; Næsset, E. Estimation of Diameter and Basal Area Distributions in Coniferous Forest by Means of Airborne Laser Scanner Data. *Scand. J. For. Res.* **2006**, *19*, 529–542. [CrossRef]
Riihimäki, H.; Heiskanen, J.; Luoto, M. The Effect of Topography on Arctic-Alpine Aboveground Biomass and NDVI Patterns. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *56*, 44–53. [CrossRef]
Esteban, J.; McRoberts, R.E.; Fernández-Landa, A.; Tomé, J.L.; Næsset, E. Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data. *Remote Sens.* **2019**, *11*, 1944. [CrossRef]
Noordermeer, L.; Bollandsås, O.M.; Ørka, H.O.; Næsset, E.; Gobakken, T. Comparing the Accuracies of Forest Attributes Predicted from Airborne Laser Scanning and Digital Aerial Photogrammetry in Operational Forest Inventories. *Remote Sens. Environ.* **2019**, *226*, 26–37. [CrossRef]
Næsset, E.; Gobakken, T.; Jutras-Perreault, M.-C.; Naesset Ramtvedt, E. Comparing 3D Point Cloud Data from Laser Scanning and Digital Aerial Photogrammetry for Height Estimation of Small Trees and Other Vegetation in a Boreal–Alpine Ecotone. *Remote Sens.* **2021**, *13*, 2469. [CrossRef]
Persson, H.J.; Olofsson, K.; Holmgren, J. Two-Phase Forest Inventory Using Very-High-Resolution Laser Scanning. *Remote Sens. Environ.* **2022**, *271*, 112909. [CrossRef]
Ståhl, G.; Saarela, S.; Schnell, S.; Holm, S.; Breidenbach, J.; Healey, S.P.; Patterson, P.L.; Magnussen, S.; Næsset, E.; McRoberts, R.E.; et al. Use of Models in Large-Area Forest Surveys: Comparing Model-Assisted, Model-Based and Hybrid Estimation. *For. Ecosyst.* **2016**, *3*, 1. [CrossRef]
Næsset, E.; Gobakken, T. Estimation of Above- and below-Ground Biomass across Regions of the Boreal Forest Zone Using Airborne Laser. *Remote Sens. Environ.* **2008**, *112*, 3079–3090. [CrossRef]
Næsset, E.; Gobakken, T.; Solberg, S.; Gregoire, T.G.; Nelson, R.; Ståhl, G.; Weydahl, D. Model-Assisted Regional Forest Biomass Estimation Using LiDAR and InSAR as Auxiliary Data: A Case Study from a Boreal Forest Area. *Remote Sens. Environ.* **2011**, *115*, 3599–3614. [CrossRef]
Bollandsås, O.M.; Gregoire, T.G.; Næsset, E.; Øyen, B.-H. Detection of Biomass Change in a Norwegian Mountain Forest Area Using Small Footprint Airborne Laser Scanner Data. *Stat. Methods Appl.* **2012**, *22*, 113–129. [CrossRef]
Gobakken, T.; Næsset, E.; Nelson, R.; Bollandsås, O.M.; Gregoire, T.G.; Ståhl, G.; Holm, S.; Ørka, H.O.; Astrup, R. Estimating Biomass in Hedmark County, Norway Using National Forest Inventory Field Plots and Airborne Laser Scanning. *Remote Sens. Environ.* **2012**, *123*, 443–456. [CrossRef]
Næsset, E.; Bollandsås, O.M.; Gobakken, T.; Gregoire, T.G.; Ståhl, G. Model-Assisted Estimation of Change in Forest Biomass over an 11year Period in a Sample Survey Supported by Airborne LiDAR: A Case Study with Post-Stratification to Provide "Activity Data". *Remote Sens. Environ.* **2013**, *128*, 299–314. [CrossRef]
McRoberts, R.E.; Næsset, E.; Gobakken, T. Inference for Lidar-Assisted Estimation of Forest Growing Stock Volume. *Remote Sens. Environ.* **2013**, *128*, 268–275. [CrossRef]
Maltamo, M.; Bollandsås, O.M.; Gobakken, T.; Næsset, E. Large-Scale Prediction of Aboveground Biomass in Heterogeneous Mountain Forests by Means of Airborne Laser Scanning. *Can. J. For. Res.* **2016**, *46*, 1138–1144. [CrossRef]
Nilsson, M.; Nordkvist, K.; Jonzén, J.; Lindgren, N.; Axensten, P.; Wallerman, J.; Egberth, M.; Larsson, S.; Nilsson, L.; Eriksson, J.; et al. A Nationwide Forest Attribute Map of Sweden Predicted Using Airborne Laser Scanning Data and Field Data from the National Forest Inventory. *Remote Sens. Environ.* **2017**, *194*, 447–454. [CrossRef]
Hollaus, M.; Dorigo, W.; Wagner, W.; Schadauer, K.; Höfle, B.; Maier, B. Operational Wide-Area Stem Volume Estimation Based on Airborne Laser Scanning and National Forest Inventory Data. *Int. J. Remote Sens.* **2009**, *30*, 5159–5175. [CrossRef]
McRoberts, R.E.; Bollandsås, O.M.; Næsset, E. Modeling and Estimating Change. In *Forestry Applications of Airborne Laser Scanning*; Springer: Dordrecht, The Netherlands, 2014; pp. 293–313.
Økseter, R.; Bollandsås, O.M.; Gobakken, T.; Næsset, E. Modeling and Predicting Aboveground Biomass Change in Young Forest Using Multi-Temporal Airborne Laser Scanner Data. *Scand. J. For. Res.* **2015**, *30*, 458–469. [CrossRef]
Hansen, E.H.; Gobakken, T.; Bollandsås, O.M.; Zahabu, E.; Næsset, E. Modeling Aboveground Biomass in Dense Tropical Submontane Rainforest Using Airborne Laser Scanner Data. *Remote Sens.* **2015**, *7*, 788–807. [CrossRef]

Bollandsås, O.M.; Ene, L.T.; Gobakken, T.; Næsset, E. Estimation of Biomass Change in Montane Forests in Norway along a 1200 Km Latitudinal Gradient Using Airborne Laser Scanning: A Comparison of Direct and Indirect Prediction of Change under a Model-Based Inferential Approach. *Scand. J. For. Res.* **2018**, *33*, 155–165. [CrossRef]
Dalponte, M.; Ene, L.T.; Gobakken, T.; Næsset, E.; Gianelle, D. Predicting Selected Forest Stand Characteristics with Multispectral ALS Data. *Remote Sens.* **2018**, *10*, 586. [CrossRef]
Næsset, E.; Nelson, R. Using Airborne Laser Scanning to Monitor Tree Migration in the Boreal–Alpine Transition Zone. *Remote Sens. Environ.* **2007**, *110*, 357–369. [CrossRef]
Næsset, E.; Gobakken, T.; McRoberts, R.E. A Model-Dependent Method for Monitoring Subtle Changes in Vegetation Height in the Boreal–Alpine Ecotone Using Bi-Temporal, Three Dimensional Point Data from Airborne Laser Scanning. *Remote Sens.* **2019**, *11*, 1804. [CrossRef]
Noordermeer, L.; Bielza, J.C.; Saarela, S.; Gobakken, T.; Bollandsås, O.M.; Næsset, E. Monitoring Tree Occupancy and Height in the Norwegian Alpine Treeline Using a Time Series of Airborne Laser Scanner Data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103201. [CrossRef]
Næsset, E. Effects of Different Sensors, Flying Altitudes, and Pulse Repetition Frequencies on Forest Canopy Metrics and Biophysical Stand Properties Derived from Small-Footprint Airborne Laser Data. *Remote Sens. Environ.* **2009**, *113*, 148–159. [CrossRef]
Kangas, A.; Gobakken, T.; Puliti, S.; Hauglin, M.; Næsset, E. Value of Airborne Laser Scanning and Digital Aerial Photogrammetry Data in Forest Decision Making. *Silva Fenn.* **2018**, *52*, 9923. [CrossRef]
Puliti, S.; Solberg, S.; Granhus, A. Use of UAV Photogrammetric Data for Estimation of Biophysical Properties in Forest Stands Under Regeneration. *Remote Sens.* **2019**, *11*, 233. [CrossRef]
Ståhl, G.; Holm, S.; Gregoire, T.G.; Gobakken, T.; Næsset, E.; Nelson, R. Model-Assisted Estimation of Biomass in a LiDAR Sample Survey in Hedmark County, Norway. *Can. J. For. Res.* **2011**, *41*, 83–95. [CrossRef]
Skowronski, N.S.; Clark, K.L.; Gallagher, M.; Birdsey, R.A.; Hom, J.L. Airborne Laser Scanner-Assisted Estimation of Aboveground Biomass Change in a Temperate Oak-Pine Forest. *Remote Sens. Environ.* **2014**, *151*, 166–174. [CrossRef]
Magnussen, S.; Næsset, E.; Gobakken, T. Lidar-Supported Estimation of Change in Forest Biomass with Time-Invariant Regression Models. *Can. J. For. Res.* **2015**, *45*, 1514–1523. [CrossRef]
Cao, L.; Coops, N.C.; Innes, J.L.; Sheppard, S.R.J.; Fu, L.; Ruan, H.; She, G. Estimation of Forest Biomass Dynamics in Subtropical Forests Using Multi-Temporal Airborne LiDAR Data. *Remote Sens. Environ.* **2016**, *178*, 158–171. [CrossRef]
Saarela, S.; Holm, S.; Healey, S.P.; Andersen, H.E.; Petersson, H.; Prentius, W.; Patterson, P.L.; Næsset, E.; Gregoire, T.G.; Ståhl, G. Generalized Hierarchical Model-Based Estimation for Aboveground Biomass Assessment Using GEDI and Landsat Data. *Remote Sens.* **2018**, *10*, 1832. [CrossRef]
Saarela, S.; Wästlund, A.; Holmström, E.; Mensah, A.A.; Holm, S.; Nilsson, M.; Fridman, J.; Ståhl, G. Mapping Aboveground Biomass and Its Prediction Uncertainty Using LiDAR and Field Data, Accounting for Tree-Level Allometric and LiDAR Model Errors. *For. Ecosyst.* **2020**, *7*, 43. [CrossRef]
Duncanson, L.; Neuenschwander, A.; Hancock, S.; Thomas, N.; Fatoyinbo, T.; Simard, M.; Silva, C.A.; Armston, J.; Luthcke, S.B.; Hofton, M.; et al. Biomass Estimation from Simulated GEDI, ICESat-2 and NISAR across Environmental Gradients in Sonoma County, California. *Remote Sens. Environ.* **2020**, *242*, 111779. [CrossRef]
Næsset, E. Estimating Above-Ground Biomass in Young Forests with Airborne Laser Scanning. *Int. J. Remote Sens.* **2011**, *32*, 473–501. [CrossRef]
Andersen, H.-E.; Strunk, J.; Emesgen, H.T. Using Airborne Light Detection and Ranging as a Sampling Tool for Estimating Forest Biomass Resources in the Upper Tanana Valley of Interior Alaska. *West. J. Appl. For.* **2011**, *26*, 157–164. [CrossRef]
McRoberts, R.E.; Næsset, E.; Gobakken, T.; Chirici, G.; Condés, S.; Hou, Z.; Saarela, S.; Chen, Q.; Ståhl, G.; Walters, B.F. Assessing Components of the Model-Based Mean Square Error Estimator for Remote Sensing Assisted Forest Applications. *Can. J. For. Res.* **2018**, *48*, 642–649. [CrossRef]
Austrheim, G.; Mysterud, A.; Pedersen, B.; Halvorsen, R.; Hassel, K.; Evju, M. Large Scale Experimental Effects of Three Levels of Sheep Densities on an Alpine Ecosystem. *Oikos* **2008**, *117*, 837–846. [CrossRef]
Pearce, H.G.; Anderson, W.R.; Fogarty, L.G.; Todoroki, C.L.; Anderson, S.A.J. Linear Mixed-Effects Models for Estimating Biomass and Fuel Loads in Shrublands. *Can. J. For. Res.* **2010**, *40*, 2015–2026. [CrossRef]
Catchpole, W.R.; Wheeler, C.J. Estimating Plant Biomass: A Review of Techniques. *Aust. J. Ecol.* **1992**, *17*, 121–131. [CrossRef]
de Lera Garrido, A.; Gobakken, T.; Hauglin, M.; Næsset, E.; Bollandsås, O.M. Accuracy Assessment of the Nationwide Forest Attribute Map of Norway Constructed by Using Airborne Laser Scanning Data and Field Data from the National Forest Inventory. *Scand. J. For. Res.* **2023**, *38*, 9–22. [CrossRef]
Ørka, H.O.; Bollandsås, O.M.; Hansen, E.H.; Næsset, E.; Gobakken, T. Effects of Terrain Slope and Aspect on the Error of ALS-Based Predictions of Forest Attributes. *For. An Int. J. For. Res.* **2018**, *91*, 225–237. [CrossRef]
Fitje, A.; Vestjordet, E. Stand Height Curves and New Tariff Tables for Norway Spruce. *Meddelelser Fra Nor. Inst. Skogforsk.* **1977**, *34*, 23–68.
Marklund, L.G. *Biomass Functions for Pine, Spruce and Birch in Sweden*; Department of Forest Survey, Swedish University of Agricultural Sciences: Umeå, Sweden, 1988. (In Swedish)

Kolstad, A.L.; Austrheim, G.; Solberg, E.J.; Venete, A.M.A.; Woodin, S.J.; Speed, J.D.M. Cervid Exclusion Alters Boreal Forest Properties with Little Cascading Impacts on Soils. *Ecosystems* **2018**, *21*, 1027–1041. [CrossRef]
Roussel, J.R.; Auty, D.; Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; Meador, A.S.; Bourdon, J.F.; de Boissieu, F.; Achim, A. LidR: An R Package for Analysis of Airborne Laser Scanning (ALS) Data. *Remote Sens. Environ.* **2020**, *251*, 112061. [CrossRef]
Roussel, J.R.; Auty, D. *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications [R Package LidR Version 4.0.2]*; Comprehensive R Archive Network (CRAN): Canada, 2022.
Nyström, M.; Holmgren, J.; Olsson, H. Prediction of Tree Biomass in the Forest–Tundra Ecotone Using Airborne Laser Scanning. *Remote Sens. Environ.* **2012**, *123*, 271–279. [CrossRef]
Domingo, D.; Lamelas, M.T.; Montealegre, A.L.; García-Martín, A.; de la Riva, J. Estimation of Total Biomass in Aleppo Pine Forest Stands Applying Parametric and Nonparametric Methods to Low-Density Airborne Laser Scanning Data. *Forests* **2018**, *9*, 158. [CrossRef]
Terrasolid. UAV—Terrasolid. Available online: https://terrasolid.com/products/terrasolid-uav/?utm_source=search&utm_medium=cpc&utm_campaign=nettisivuliikenne-01-2023&gclid=Cj0KCQjw8qmhBhClARIsANAtboeZ4Bmgv_jxqD0LogIBqpAelgRj8fS6vyubS5TOMYwwmmNgeLPiCD4aAq3KEALw_wcB (accessed on 3 April 2023).
Lumley, T.; Miller, A. Regression Subset Selection. 2022. Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c61985341c077574872dfcd64c8c743f48c77f4e (accessed on 22 March 2023).
Hyndman, R.J.; Koehler, A.B. Another Look at Measures of Forecast Accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
Everit, B.S.; Skrondal, A. *The Cambridge Dictionary of Statistic*, 4th ed.; Cambridge University Press: Cambridge, UK, 2006; ISBN 9780521766999.
McRoberts, R.E.; Næsset, E.; Saatchi, S.; Quegan, S. Statistically Rigorous, Model-Based Inferences from Maps. *Remote Sens. Environ.* **2022**, *279*, 113028. [CrossRef]
Novkaniza, F.; Notodiputro, K.A.; Sartono, B. Bootstrap Confidence Interval of Prediction for Small Area Estimation Based on Linear Mixed Model. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *187*, 012040. [CrossRef]
Chatterjee, S.; Hadi, A.S. *Regression Analysis by Example*; Wiley: Hoboken, NJ, USA, 2015; ISBN 978-1-118-45624-8.
Ståhl, G.; Holm, S.; Gregoire, T.G.; Gobakken, T.; Næsset, E.; Nelson, R. Model-Based Inference for Biomass Estimation in a LiDAR Sample Survey in Hedmark County, Norway. *Can. J. For. Res.* **2011**, *41*, 96–107. [CrossRef]
Gobakken, T.; Næsset, E. Assessing Effects of Positioning Errors and Sample Plot Size on Biophysical Stand Properties Derived from Airborne Laser Scanner Data. *Can. J. For. Res.* **2009**, *39*, 1036–1052. [CrossRef]
Mauya, E.W.; Hansen, E.H.; Gobakken, T.; Bollandsås, O.M.; Malimbwi, R.E.; Næsset, E. Effects of Field Plot Size on Prediction Accuracy of Aboveground Biomass in Airborne Laser Scanning-Assisted Inventories in Tropical Rain Forests of Tanzania. *Carbon Balance Manag.* **2015**, *10*, 10. [CrossRef] [PubMed]
Hodgson, M.E.; Bresnahan, P. Accuracy of Airborne Lidar-Derived Elevation: Empirical Assessment and Error Budget. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 331–339. [CrossRef]
McRoberts, R.E.; Westfall, J.A. Propagating Uncertainty through Individual Tree Volume Model Predictions to Large-Area Volume Estimates. *Ann. For. Sci.* **2016**, *73*, 625–633. [CrossRef]
Hou, Z.; Mehtätalo, L.; McRoberts, R.E.; Ståhl, G.; Tokola, T.; Rana, P.; Siipilehto, J.; Xu, Q. Remote Sensing-Assisted Data Assimilation and Simultaneous Inference for Forest Inventory. *Remote Sens. Environ.* **2019**, *234*, 111431. [CrossRef]
Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1994; Volume 21, ISBN 9780429246593.
McRoberts, R.E.; Næsset, E.; Gobakken, T.; Bollandsås, O.M. Indirect and Direct Estimation of Forest Biomass Change Using Forest Inventory and Airborne Laser Scanning Data. *Remote Sens. Environ.* **2015**, *164*, 36–42. [CrossRef]

II

OXFORD Institute of Chartered Foresters

Forestry: An International Journal of Forest Research, 2024, 1–11
https://doi.org/10.1093/forestry/cpae006
Original Article

Computation of prediction intervals for forest aboveground biomass predictions using generalized linear models in a large-extent boreal forest region

Ritwika Mukhopadhyay [1,*], Magnus Ekström [1,2], Eva Lindberg [1], Henrik J. Persson [1], Svetlana Saarela [1,3], Mats Nilsson [1]

[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå 90183, Sweden
[2]Department of Statistics, USBE, Umeå University, Umeå 90187, Sweden
[3]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås 1433, Norway
*Corresponding author. E-mail: ritwika.mukhopadhyay@slu.se

Abstract

Remotely sensed data have an important application for estimation of forest variables, e.g. height, volume, and aboveground biomass (AGB). The increased use of remotely sensed data implemented along with model-based inference has shown improved efficiency in prediction and mapping of such forest variables. In this study, plot-level airborne laser scanning data and Swedish National Forest Inventory field reference data were used to predict AGB using generalized linear models (GLMs) assuming Gamma and Tweedie distributions for the field observed AGB. The GLMs were selected considering the convenience of not correcting transformation bias as it is required in other regression models with transformed response variable. To overcome the challenge in providing reliable uncertainty estimates for the estimated forest variable map products at individual pixel-scale, we focused on computing 95% prediction intervals (PIs) for Gamma and Tweedie GLMs with a square root link function. The relative uncertainties were computed as the ratio between the half-width of the PIs and the predicted AGBs. The AGB-airborne laser scanning models were developed with root mean square error values of 22.6 Mgha^{-1} (26%) and 21.7 Mgha^{-1} (25%), respectively, for the Gamma and Tweedie GLMs. Two methods were applied to compute PIs for the Gamma GLM, one using the R package 'ciTools' and another derived through asymptotic theory. It was found that the 95% PIs computed using 'ciTools' had the most accurate coverage probability in comparison to the other method. An extended version of these PIs was also utilized for the Tweedie GLMs. The range of PIs associated with the prediction of AGB were narrower for lower predicted AGB values compared with the length of higher predicted AGB values. Comparing the two fitted models, the Gamma GLM showed lower relative uncertainties for the lower range of predicted AGBs, whereas the Tweedie GLM showed lower relative uncertainties for the higher range of predicted AGBs. Overall, the Tweedie GLM provided a better model fit for AGB predictions.

Keywords: aboveground biomass; Gamma distribution; generalized linear model; model-based inference; Tweedie distribution; uncertainty estimation

Introduction

Forests play a crucial role in the global carbon cycle because of their ability to sequester carbon which is considered an important tool for mitigating climate change (FAO 2018). Forest biomass is therefore an important variable to describe the state and structure of forests. The need to acquire information about forest state and change has been increasing globally (IPCC 2022). The United Nations Framework Convention on Climate Change (UNFCCC) aims at achieving stabilized greenhouse gas concentrations in the atmosphere amongst others by protecting and increasing forest areas as they represent large carbon sinks (UNFCCC 2021). Field inventory data have been a primary source for collecting information about forests for a long time. The field data collected by National Forest Inventories (NFIs) can be used for the estimation of the state and change of forest variables, e.g. aboveground biomass (AGB) and growing stock volume at regional and national level (Tomppo *et al.* 2010). However, the collection of field data is time-consuming, expensive and unfeasible in inaccessible terrains.

Using remotely sensed (RS) data as auxiliary information has proven to be an efficient method to overcome many challenges in spatially estimating forest variables (McRoberts *et al.* 2010; Næsset *et al.* 2011; Vauhkonen *et al.* 2014; Nilsson *et al.* 2017; Saarela *et al.* 2020). Previous studies have combined RS data and field data at the design phase of sampling experiments (e.g. McRoberts *et al.* 2002; Saarela *et al.* 2015; Haakana *et al.* 2019), whereas other studies have combined RS data and field data for the estimation of forest variables of interest using model-based approaches, which is more straightforward (Andersen *et al.* 2005; Hudak *et al.* 2008; Wulder *et al.* 2008; Nilsson *et al.* 2017; Persson and Fransson 2017; Saarela *et al.* 2020; Persson *et al.* 2020).

Airborne laser scanning (ALS) has been combined with both parametric and non-parametric methods for producing accurate wall-to-wall predictions of forest variables, namely, tree height,

Handling editor: Bogdan Strimbu
Received: April 28, 2023. **Revised:** January 21, 2024. **Accepted:** January 23, 2024

biomass, and volume (Nilsson 1996; Næsset 2002; Holmgren 2006; Lim and Treitz 2006; Packalén and Maltamo 2007; Ståhl et al. 2011a; Ståhl et al. 2011b; Vauhkonen et al. 2014; Maltamo et al. 2016). Within model-based inference, target variables are predicted for individual map units based on the established model linking the available auxiliary data to the target variables. The model can be referred to as a super-population model (Särndal et al. 1978; Thompson et al. 1982; Gregoire 1998; Saarela et al. 2020). Model-based inference has been used for the prediction of AGB and other related forest variables such as, mean tree height, dominant tree height, canopy height, mean diameter, stem number, basal area, and timber volume, using ALS as auxiliary RS data at plot and species level (e.g. Nelson, Krabill and Tonelli 1988; Lefsky et al. 2002; Næsset, Bollandsås and Gobakken 2005; Gobakken and Næsset 2006; Maltamo et al. 2016). Various other studies using ALS as auxiliary data in the estimation phase of model-based inferences have been conducted (e.g. Næsset and Gobakken 2008; Ståhl et al. 2011b; Næsset 2011; Gobakken et al. 2012). Some have compared machine learning algorithms, like random forest (RF), support vector machine, and artificial neural network, to generalized linear model (GLM) for the prediction of AGB (Schuh et al. 2020). The use of RS data allows the mapping of the variable of interest across larger regions and providing additional information about the spatial distribution of the variables (e.g. Franco-Lopez et al. 2001; Tomppo et al. 2008; Hollaus et al. 2009; Nord-Larsen and Schumacher 2012; Nilsson et al. 2017; Esteban et al. 2019, 2020; Saarela et al. 2020).

The model-based inference can be further utilized for predicting the target variable along with the estimation and mapping of uncertainties related to the target variable. There have been a few studies conducted for exploiting model-based inference methods for estimating uncertainties and mapping of target variables. Babcock et al. (2018) developed and examined the 'coregionalization' modelling approach for the estimation of AGB with a combination of field inventoried data, ALS strip samples, and Landsat data to accommodate spatially structured errors for a more reliable characterization of uncertainty, for a study region in Tanana Valley, Alaska. Saarela et al. (2020) introduced generalized non-linear least square models within the hierarchical model-based (HMB) inference framework linking ALS metrics and field measured AGB for the prediction and mapping of AGB for a study area located in south-central Sweden. The uncertainty propagation was estimated both at tree-level and plot-level and mapped for all individual pixels for the wall-to-wall ALS data. Uncertainty estimation studies have also been performed for non-parametric regression methods, such as RF (e.g. Zhang et al. 2019; Esteban et al. 2019, 2020). We have not found any studies within the model-based inference framework implementing GLMs for prediction of the target variable AGB, which do not require the correction of transformation bias as is required in cases for transformed response variables (Hudak et al. 2006). Computation of prediction intervals (PIs) as a form of uncertainty estimation is rarely discussed. Computation of PIs would provide a better understanding of the map products and the uncertainties for more effective practical uses (Fassnacht et al. 2023). Providing the prediction uncertainties for individual cases, i.e. the individual pixels in our case, adds value to the map products, as lack of reliable uncertainty estimates has been identified as one of the challenges in RS by Fassnacht et al. (2023). Therefore, in this current study, we present model-based inference for prediction and mapping of AGB and its associated prediction uncertainties for the Gamma and Tweedie distribution families. The main reason for using these two exponential dispersion models was that they are useful for modelling positive data, such as our field observed AGB data, that have positively skewed distributions.

Table 1. Summary statistics for the 500 NFI field plots.

Data set	Mean	Minimum	Maximum	SD
AGB ($Mgha^{-1}$)	86.35	1.627	360.8	59.14
DBH* (cm)	17.68	0.5358	73.49	8.676
Height (m)	13.83	0.4312	36.50	6.016

The main objective of this study was to compute and compare pixel-level prediction uncertainties in the form of PIs for Gamma and Tweedie exponential dispersion models implemented within a model-based framework for AGB prediction. We utilized plot-level AGB values from the Swedish NFI as the response variable and ALS metrics as the explanatory variables for fitting regression models for wall-to-wall prediction of AGB in a study area located in southern central Sweden. A particular objective of the study was also to compare two available methods for computing PIs, one available in the R package 'ciTools' (Haman 2017) and an extension of Hattab's (2016) method from a 'log' link to a 'sqrt' link, implemented for Gamma GLMs.

Materials and methods

Study area

The study area is located in south-central Sweden with an overall area of 500 500 ha of which 390 900 ha is covered by forests. The forest-covered areas were selected using land-use maps available from the Swedish National Land Survey (Lantmäteriet 2019). The NFI plots from the entire study area are illustrated in Fig. 1a.

The NFI plots were used for training the models and the AGB was predicted for the subset outlined in Fig. 1b with black solid line.

Swedish NFI field data and ALS data

Field data for 500 Swedish NFI plots in the study area were selected for the period 2009–11. The NFI plots are circular with a radius of 10 m and systematically sampled as clusters of plots over Sweden; for each plot, diameter at breast height (DBH) was measured for all trees, whereas height was available for only a few sampled trees (Valinger et al. 2019). The summary statistics for the NFI field plots are presented in Table 1 and the distribution of the plot-level AGB is shown in Fig. 2. Wall-to-wall ALS data with a point density of 0.5–1 points/m^2 were available for 2009–11 from a national ALS survey carried out by the Swedish National Land Survey (Lantmäteriet 2019). ALS metrics for the corresponding plots were extracted using the Fusion software (McGaughey 2020). Five plots were identified and discarded from the data set since they were found to be clear cut between the field inventory and the scanning dates and had effect from trees located outside the plot boundaries. The wall-to-wall rasters were generated from the point clouds with raster cells of 18×18 m^2, which roughly conforms with the area of the NFI plots.

Modelling

Assuming the field reference AGB to be approximately Gamma or Tweedie distributed, both Gamma and Tweedie regression models were implemented to link the plot AGB with ALS metrics. The ALS metrics were selected as explanatory variables of the models from a set of 100 candidate variables using the *leaps::regsubsets* variable selection algorithm (Lumley and Miller 2022).The regression model parameters were estimated using the maximum likelihood estimation method. Predicted AGB and a corresponding PI were

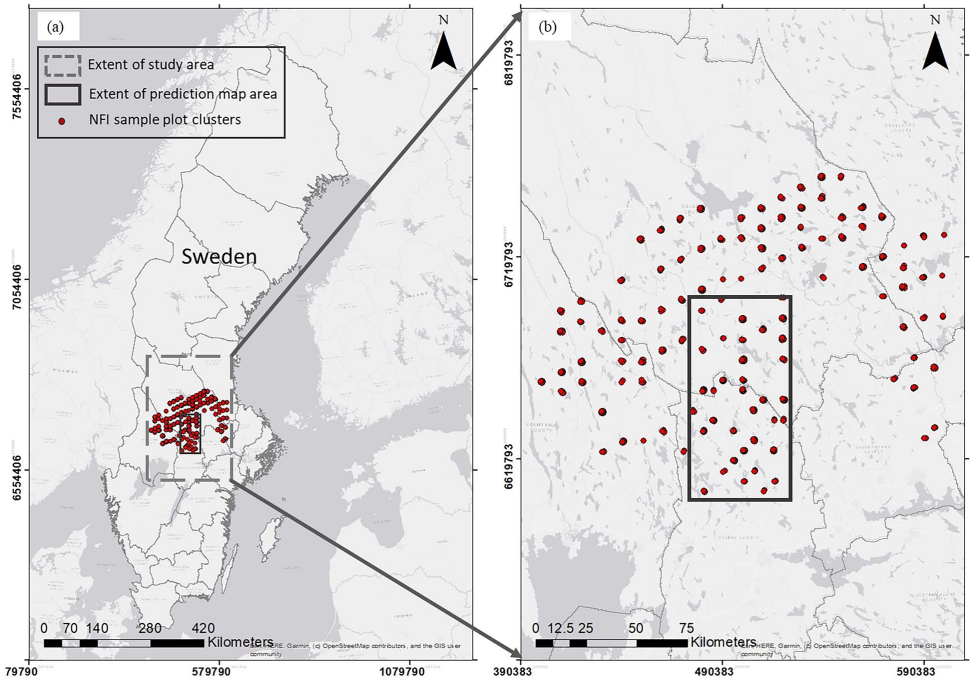


Figure 1. (a) The study area in south-central Sweden and the cluster of NFI sample plots used in the study, and (b) the sub-area for producing the prediction maps.

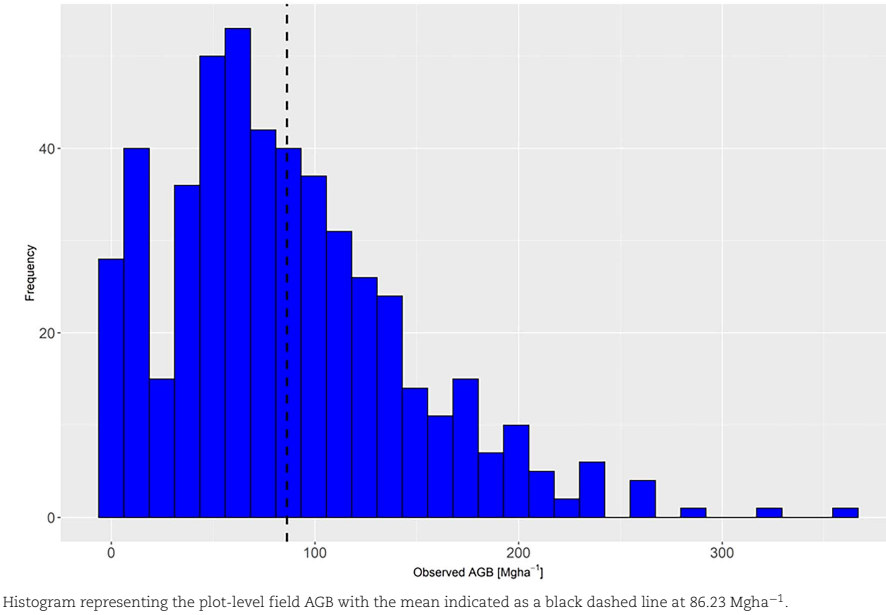


Figure 2. Histogram representing the plot-level field AGB with the mean indicated as a black dashed line at 86.23 Mgha^{-1}.

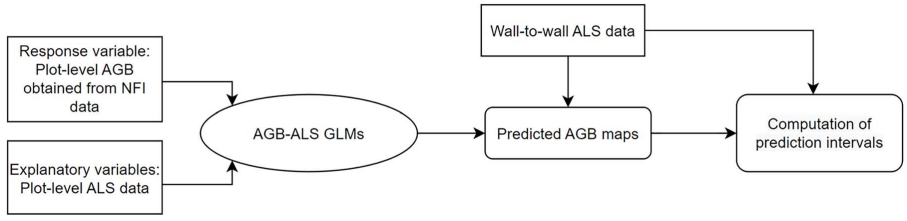


Figure 3. Methodological overview of the study.

computed for each pixel in the wall-to-wall raster generated using the Gamma or Tweedie model. The methodological overview is shown in Fig. 3.

AGB-ALS Generalized Linear Models

The GLMs are an extension of the traditional linear models but in the case of GLMs the distribution of the response variable does not have to be Gaussian. For fitting models, we used the *glm* function in the R package 'stats' (R Core Team 2021). The families 'Gamma' and 'Tweedie' distributions were used for the response variable, i.e. AGB, as the field observed AGB from the NFI was non-normal and had positive continuous values. Tweedie distributions are defined as exponential dispersion models with

$$\text{var}\,(\text{AGB}_i) = \phi E(\text{AGB}_i)^q, \tag{1}$$

where ϕ is the dispersion parameter and q is the variance power. By comparing Akaike information criterion (AIC) values for competing Tweedie GLMs, q was set to 1.25. Tweedie distributions with $1<q<2$ are sometimes called Poisson–Gamma distributions (Dunn and Smyth 2018).

Using Cook's distance and a threshold equal to 1 (Zuur *et al.* 2009), and checking with historical NFI entries for the permanent plots, five out of the 500 sample plots were identified as outliers and were discarded from the data set in the modelling phase, as mentioned above. Different link functions, like 'identity', 'sqrt' (square root), 'log', and 'inverse', were tested for the AGB-ALS Gamma and Tweedie GLMs and the best models were selected based on hypothesis testing on individual model parameters, AIC values, and by analyzing the distribution of the quantile residuals. The pseudo-R^2 for both models was computed by subtracting the ratio of the residual deviance and the null deviance of the GLM model from 1 (McFadden 1974; Hu, Shao and Palta 2006).

The AGB-ALS models were used for predicting AGB for the entire study area. The rasters of the explanatory variables were checked thoroughly before implementing the wall-to-wall predictions. For the h_{90} raster, all pixel values above 35 m were set to 35 m based on the histogram of the h_{90} pixel values and considering terrain model errors.

Estimation of PIs for the AGB-ALS model

A PI is an estimate of an interval in which a single future response observation will fall with some specified probability $(1-\alpha)\bullet 100\%$, where α is some fraction between 0 and 1 (Millard and Neerchal 2000). The value of α is typically set to 0.05. PIs are often used in multiple linear regression, but for Gamma and Tweedie regression, they are seldom used and discussed.

R package 'ciTools'

For Gamma regression, PIs can be computed with the R package 'ciTools'. The algorithm in Figure 4 describes the 'ciTools' package

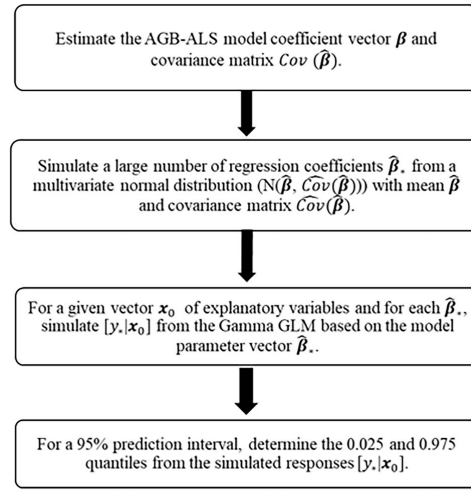


Figure 4. Algorithm describing the method of estimating the PIs by the 'ciTools' package.

implementation for computing PIs (Haman 2017), which we used in our study. The package does not yet have the Tweedie distribution implemented, so we implemented it for our computations. Essentially, Tweedie was included in the set of supported families of distributions in 'add_pi.glm' and we made it possible to simulate Tweedie responses in 'get_sim_response'. For the latter, we used the R function 'rTweedie' from the 'mgcv' package. As the Tweedie GLM was computationally more complicated and heavy to be handled by the R package 'ciTools', only a systematic sample (no. of pixels = 466 719) of the predicted AGB pixels were extracted from the predicted AGB map (Fig. 6c) and used for the computation of the 95% PIs.

In Hattab (2016), another method for computing PIs was derived for Gamma regression models with the 'log' link function. In the present study, we derived and evaluated corresponding PIs for the case when the Gamma regression models were based on the 'sqrt' link function.

Let $\hat{\boldsymbol{\beta}}$ denote the estimated vector of model coefficients and $\hat{r}$ the estimated shape parameter of the Gamma regression model. The latter estimate can be obtained using the 'gamma.shape' function in the R package MASS (Venables and Ripley 2004). Our proposed 95% PI for a future response observation, for a given vector of explanatory variables $\mathbf{x}_0$, is given by

$$\left(\frac{\hat{G}_{(0.025)}(\mathbf{x}_0'\hat{\boldsymbol{\beta}})^2}{2\hat{r}},\ \frac{\hat{G}_{(0.975)}(\mathbf{x}_0'\hat{\boldsymbol{\beta}})^2}{2\hat{r}}\right), \tag{2}$$

where $\hat{G}_{(0.025)}$ and $\hat{G}_{(0.975)}$ are estimated 0.025 and 0.975 quantiles of the distribution of

$$G = \frac{\chi^2_{2\hat{r}}}{z\sqrt{4(\mathbf{x}_0'\hat{\boldsymbol{\beta}})^{-2}\mathbf{x}_0' I^{-1}(\hat{\boldsymbol{\beta}})\mathbf{x}_0 + 1}} \quad (3)$$

where χ^2_{2r} denotes a chi-squared distributed random variable with degrees of freedom = 2r, z a standard normal random variable, and $I(\hat{\boldsymbol{\beta}})$ the information matrix. The random variables $\chi^2_{2\hat{r}}$ and z are assumed to be independent. The estimated quantiles may be obtained as the sample quantiles obtained from a large number of realizations of the random variable G. We refer to the Supplementary Material for the derivation of the PI. If the sample size n of training data is small, it might be advantageous to replace the standard normal random variable z in (2) with a random variable from a t distribution with $n - p$ degrees of freedom, where p is the number of explanatory variables in the model (cf. Hattab 2016).

Comparative analysis of the PI estimation methods

The two methods for computing PIs for the Gamma GLMs were compared using Monte Carlo simulations (Supplementary Material) in combination with: (i) the extension of Hattab's (2016) method from a 'log' link to a 'sqrt' link and (ii) the method available in the R package 'ciTools'. In these simulations, we compared the estimated coverage probabilities and average lengths of PIs for the two methods across different sample sizes ($n = 495$, 50, and 10). Coverage probability is the probability that the interval covers a future observation of the response variable, given the values of the explanatory variables (Casella *et al.* 2002).

The GLMs were applied to the wall-to-wall raster to generate maps of predicted AGB and prediction uncertainties, covering a subset of the study region as shown in Fig. 6a–d. To generate the uncertainty map for the Tweedie GLM (Fig. 6d), the uncertainties for the sampled predicted AGB pixels were interpolated for the entire range of the predicted AGB raster using a linear relationship between the predicted AGBs and uncertainties. The relative uncertainty was calculated as the half-width of the PIs relative to the predicted AGB value and multiplied by 100 (Sims *et al.* 2021) (Fig. 7b and d), for Gamma and Tweedie GLM, respectively, and overlaid with a layer representing the boundaries of felling activities that were registered after 2011, which is the year the map represents (Skogsstyrelsen 2023).

Results

The selected explanatory variables were the following ALS metrics: vegetation ratio (v), the 90th height percentile (h_{90}), and the interaction term between h_{90} and v ($h_{90} \times v$). The v metric is calculated as the proportion of the first laser returns above 1.5 m to all first returns. The Gamma and Tweedie regression models formulated for the estimation of AGB were

$$\eta_{\mathrm{AGB}_i} = \beta_0 + \beta_1 h_{90} + \beta_2 v + \beta_3 (h_{90} \times v), \quad (4)$$

where $\eta_{\mathrm{AGB}_i} = (E(\mathrm{AGB}_i))^{1/2}$, i.e. the 'sqrt' link function was used. The estimated model coefficients $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are presented in Tables 2 and 3 for Gamma and Tweedie regression, respectively. The P-values have been represented for each estimated model coefficients for the Gamma and the Tweedie GLM models in Tables 2 and 3, respectively, and a low P-value (i.e. $\leq .05$) indicates that the corresponding true coefficient is significantly different from zero.

Table 2. Estimated model parameters for AGB-ALS Gamma GLM model.

Model coefficients	Estimated values	Standard error	P-value
$\hat{\beta}_0$	1.51	0.076	$<.002 \times 10^{-13}$
$\hat{\beta}_1$	0.159	0.018	$<.002 \times 10^{-13}$
$\hat{\beta}_2$	0.032	0.233×10^{-2}	$<.002 \times 10^{-13}$
$\hat{\beta}_3$	0.002×10^{-2}	0.226×10^{-3}	$<.002 \times 10^{-13}$

Table 3. Estimated model parameters for AGB-ALS Tweedie GLM model.

Model coefficients	Estimated values	Standard error	P-value
$\hat{\beta}_0$	1.32	0.170	60.6×10^{-13}
$\hat{\beta}_1$	0.227	0.021	$<.002 \times 10^{-13}$
$\hat{\beta}_2$	0.034	0.284×10^{-2}	$<.002 \times 10^{-13}$
$\hat{\beta}_3$	0.157×10^{-2}	0.241×10^{-3}	$<.002 \times 10^{-13}$

Table 4. Statistics for the GLM fit assessment.

Model statistics	Gamma GLM	Tweedie GLM
AIC	4395	4215
Pseudo-R^2	0.8454	0.8844
RMSE	22.59 Mgha^{-1}	21.74 Mgha^{-1}
Relative RMSE	25.99%	25.02%

The obtained AIC, pseudo-R^2, root mean square error (RMSE), and relative RMSE values are listed in Table 4. The lowest AIC value for Gamma GLM and better residual and quantile–quantile (QQ)-plots were obtained for the 'sqrt' link function. Figure 5a and c shows plots of quantile residuals, obtained using the R package DHARMa (Hartig 2022). If a GLM model is valid, these quantile residuals will approximately follow a uniform distribution on the interval (0,1). The results of the Kolmogorov–Smirnov test of uniformity, the dispersion test, and the outlier test given in the QQ-plot of the quantile residuals suggest that there are deviations from the Gamma GLM assumptions. In the scatterplot of quantile residuals against (rank transformed) predictions of AGB, empirical 0.25, 0.5, and 0.75 quantiles in y direction (solid lines) are compared with the corresponding theoretical 0.25, 0.5, and 0.75 quantiles (dashed lines), and the plot suggests that there are significant deviations from the expected quantiles. In practical applications, the assumptions are never exactly true. However, the deviations from the Gamma GLM assumptions suggested by the quantile residual plots indicate that it might be possible to find a better model for the AGB data. Such a model is provided by the Tweedie GLM. In comparison, no significant problems were detected in the quantile residual plots for the Tweedie GLM. Also, from the different model statistics presented in Table 4, it can be observed that the Tweedie GLM with RMSE value of 21.74 Mgha^{-1} (25.02%) is better than the Gamma GLM with RMSE value of 22.59 Mgha^{-1} (25.99%), indicating that the Tweedie regression provided a better fit than the Gamma regression.

For the largest sample size of $n = 495$ (the entire field sample in our study), the two methods perform equally well in terms of coverage probability and average length of PIs (we refer to the Supplementary Material for details). An advantage with the PIs from 'ciTools' is that they were faster to compute.

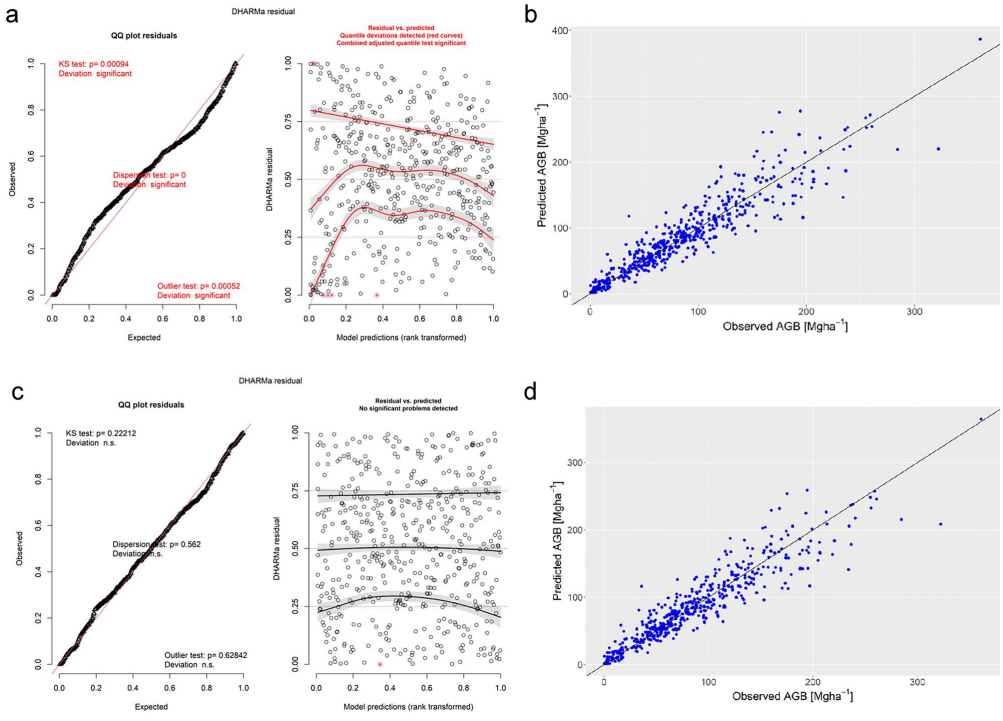


Figure 5. Plots for the Gamma (a), (b) and Tweedie GLMs (c), (d): (a, c) a QQ-plot of quantile residuals and a scatterplot of quantile residuals against (rank transformed) predictions of AGB; (b, d) predicted AGB (Mgha^{-1}) versus field reference AGB (Mgha^{-1}).

The predicted AGB values ranged from 3.11 to 331 Mgha^{-1} and from 2.82 to 324 Mgha^{-1} for the Gamma and Tweedie GLM, respectively. The associated prediction uncertainties (the difference between the 95% PIs) ranged from 3.53 to 387 Mgha^{-1} and from 14.9 to 252 Mgha^{-1}, as seen in Fig. 6b and d, for the Gamma and Tweedie GLM, respectively. Each prediction uncertainty was calculated as the difference between the endpoints of 95% PIs (Fig. 7a and c), for Gamma and Tweedie GLM, respectively. A subset of the predicted AGB map and prediction uncertainty map for the Gamma GLM has been presented for a small sub-region within the study area in Fig. 8a and b, respectively. The polygons in the predicted AGB map in Fig. 8a represent the mature forests delineated for felling activities, which tend to also correspond to the regions with high uncertainty values in the map of prediction uncertainties in Fig. 8b.

Discussion

In this study, we have introduced the computations of 95% PIs to account for prediction uncertainties in Gamma and Tweedie regression with 'sqrt' link function implemented for modelling and mapping of forest AGB. Studies on model-based estimation of AGB, mapping over large areas and uncertainty estimation have been done previously (e.g. Næsset and Gobakken 2008; McRoberts 2010; Ståhl *et al.* 2011; Maltamo *et al.* 2016; Nilsson *et al.* 2017; Saarela *et al.* 2020), but estimation of prediction uncertainties in the form of PIs has not been widely discussed and used in case of model-based estimation of AGBs with exponential dispersion models and for the wall-to-wall prediction maps. Residual plots in Fig. 5a and c suggest that the Tweedie GLM provides a better fit than the Gamma GLM. Whereas, in the scatter plots of the predicted AGB versus the field reference AGB in Fig. 5b and d, there appears to be no such distinct difference between the Gamma and the Tweedie GLM. Gamma regression models with different link functions and weights were tested in order to reduce the impact of some of the large residual values (which can be identified from the scatterplots in Fig. 5a), but no improvements were achieved. These observations could be a result of positional errors between the field plots and the ALS data, as the ALS data were extracted from exact circular plots that often do not correspond to the tree crowns of the trees included in the field plots (Mauya *et al.* 2015). The higher residuals might also occur because of measurement errors in the field data or because of the modelling approach, which might not capture the complex relationship between the response variable and the auxiliary data.

In Fig. 7a and c, it was observed that the PIs are narrower in the lower range of the predicted AGB values and gradually widen towards the higher predicted AGB values, i.e. the difference between the 95% PIs increases with increasing predicted AGB values. Thus, the range of the associated uncertainties increase with increasing predicted AGB values (typically older forests) meaning that higher predicted AGB values have a broader range of uncertainty associated with them as compared with the lower predicted AGB values for both the Gamma and Tweedie GLMs. This might be because younger trees have more homogeneous growth conditions and therefore exhibit more similar growth

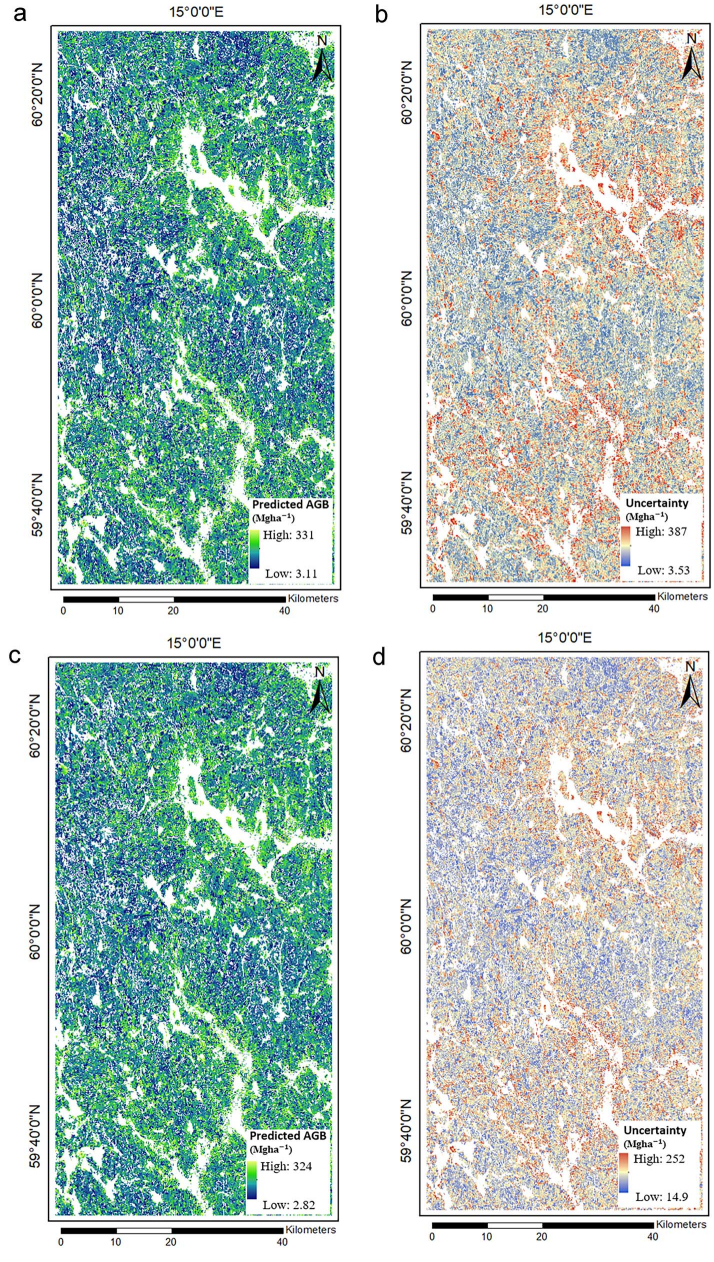


Figure 6. (a, c) AGB map predicted for the entire study area; (b, d) map of prediction uncertainties for the entire study area, where the uncertainty was computed as the difference between the endpoints of the 95% PIs for the respective models.

rates which result in lower variability amongst the trees in the stands. Whereas, forests with higher AGB values, usually indicating mature forest, may have more variability because of more complex growth conditions induced by long periods of competition and the occurrence of disturbances.

PIs account for uncertainties by estimating the mean along with the uncertainty associated with prediction of the individual units, i.e. pixels in our case (Hattab 2016). The trend between uncertainty and predicted AGBs can also be observed from the raster maps in Fig. 6a–d. The higher predicted AGBs represented

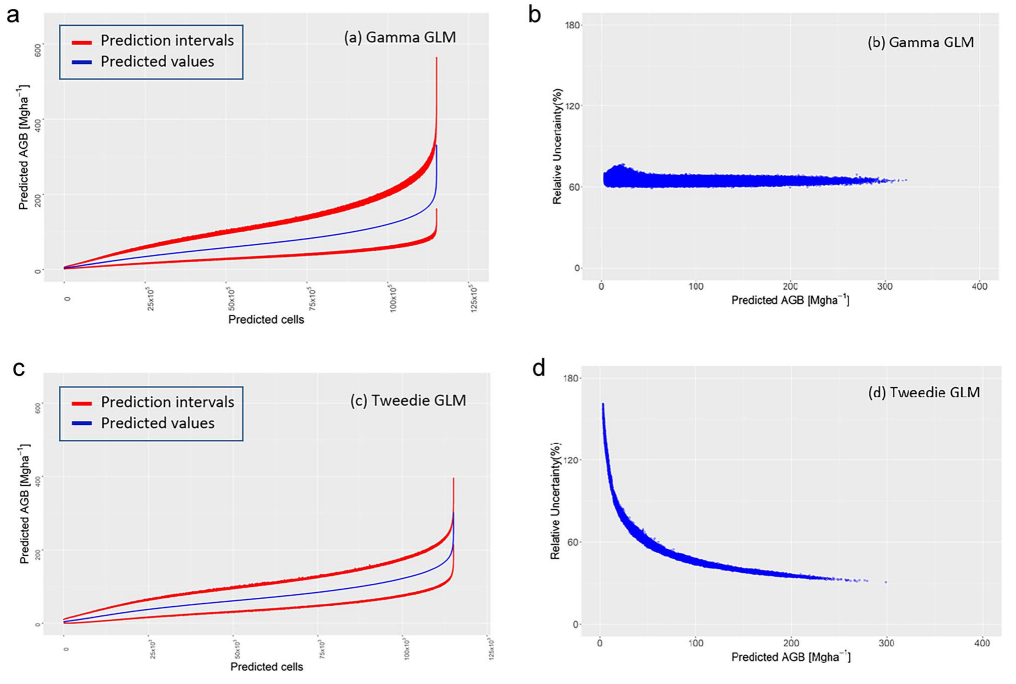


Figure 7. (a, c) Graphical representation of the predicted AGB map values (in Mgha^{-1}) with their corresponding 95% PIs; (b, d) relative uncertainty versus the predicted AGB (Mgha^{-1}) for the Gamma and Tweedie GLMs, respectively.

as the yellow or lighter green pixels in Fig. 6a and c correspond to the pixels with higher uncertainties (represented with red in Fig. 6b and d), and similarly the lower predicted AGB values correspond to the pixels with lower uncertainties in the raster maps generated from the Gamma and Tweedie GLMs, respectively. This trend can be seen in Fig. 8a and b for a sub-region within the study area. Therefore, the uncertainty maps can be used as orientation when planning future field inventories by, for example, conducting more intense sampling in mature stands in order to get more precise predictions.

The relative uncertainties represented in the scatterplots Fig. 7b for the Gamma GLM were more or less constant between 55% and 70%, whereas the relative uncertainties obtained from the Tweedie GLM (Fig. 7d) were higher at the lower range of predicted AGB values. The average relative uncertainty value for the Tweedie regression (56.3%) was lower compared with that for the Gamma regression (60.3%). If AGB corresponds directly to stand age, then we can say that the Tweedie GLM would be more effective in mature or woody stands.

The overall accuracy in terms of relative RMSE of the AGB-ALS GLMs was similar or higher than the relative RMSEs reported in other studies of AGB estimation conducted in Sweden (e.g. Nilsson *et al.* 2017; Saarela *et al.* 2020). Nilsson *et al.* (2017) implemented robust regression models for the large-scale prediction and mapping of stem volume, tree height, stem diameter, basal area, and biomass, and Saarela *et al.* (2020) implemented a generalized non-linear regression model for the modelling and mapping of AGB along with the uncertainties within an HMB framework. Based on these studies, we have elaborated the large-scale mapping of AGB and the prediction uncertainties further in terms of 95% PIs for the introduced Gamma and Tweedie regression models. These studies also bring forward the need of such statistically sound uncertainty analyses of prediction maps of forest variables, which can have useful impact on the decision-making of forest management activities, such as taxation to forest owners. The uncertainty maps can also be used as an auxiliary product for designing future inventories indicating more sampling units to be laid out in forests areas with higher uncertainty.

In future works, it would be interesting to compare the generated predicted AGB and uncertainty rasters with available maps providing information about species distribution dependent on land use land cover classes (e.g. National Land Cover Database 2020) or species maps generated using Sentinel-2 imageries (e.g. Axelsson *et al.* 2021), to understand the distribution of biomass levels and ranges of prediction uncertainty dependent on the distribution of individual species. It would also be interesting to account for the uncertainty contributions from the design and model phase by using a hybrid-inference framework (Gobakken *et al.* 2012; Corona *et al.* 2014; Ståhl *et al.* 2016; Condés and McRoberts 2017; Persson *et al.* 2022; Saarela *et al.* 2022).

Conclusion

In this study, the computation of prediction uncertainties as 95% PIs for Gamma and Tweedie GLMs for the prediction and mapping of AGB for a large, forested area (390 900 ha) in Sweden was demonstrated. The Tweedie GLM model resulted in lower RMSE and relative RMSE values than the Gamma GLM. Two methods for computing 95% PIs for Gamma GLMs with a 'sqrt' link function

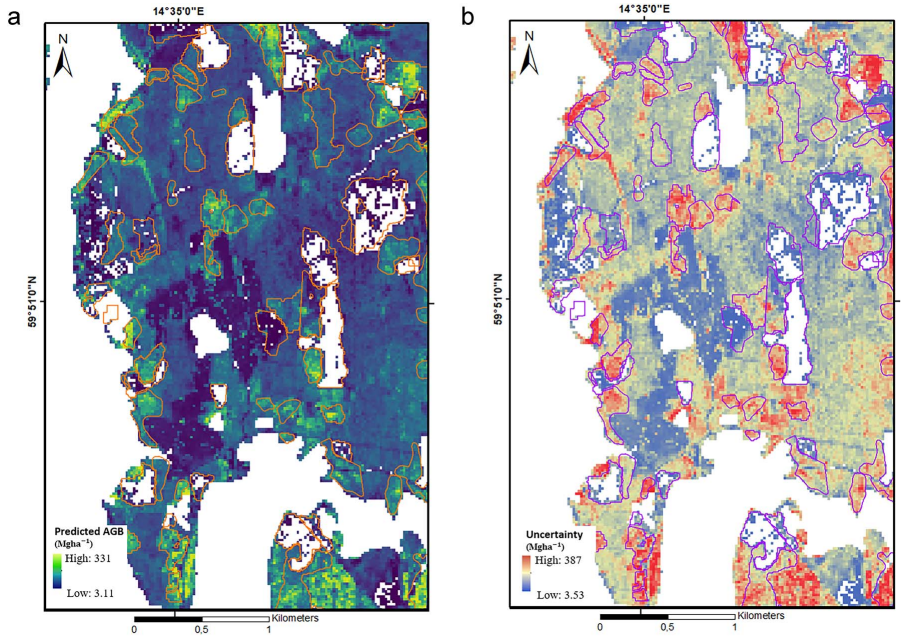


Figure 8. (a) Map of predicted AGB and (b) map of prediction uncertainties—for a sub-region within the study area overlaying the felling boundaries (represented by the delineated polygons) for the case of Gamma GLM.

were compared, i.e. the method available in the R package 'ciTools' and an extension of the method suggested by Hattab (2016). For larger sample sizes it was found that both methods performed equally well in terms of coverage probability and average length of PIs, but that PIs from the R package 'ciTools' were faster to compute. Therefore, the R package 'ciTools' was used for the Gamma GLM as well as for computation of the 95% PIs for the Tweedie GLM, in the latter case only for a systematically sampled set of predicted AGB pixels as the Tweedie GLM was computationally more complicated and expensive. It was observed that the range of uncertainties associated with the prediction of AGB was narrower for lower predicted AGB values (i.e. typically young forests) compared with the range of higher predicted AGB values for both the GLMs. From the relative uncertainty scatterplots, it could be stated that the Tweedie GLM was observed to be more accurate for the higher range of predicted AGB values. But broadly, the importance of such uncertainty analysis and maps, as has also been presented in earlier studies, is to provide information about the accuracy of the predicted forest variable, which might affect the decision-making of future forest management and planning.

Acknowledgements

We acknowledge the Swedish NFI for providing the field data used in the study. The authors would like to thank the editorial team and the anonymous reviewers for their useful comments that helped improve this article. We would also like to thank Sven Adler from SLU for providing feedback on the article.

Author contributions

Ritwika Mukhopadhyay (Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Magnus Ekström (Conceptualization, Formal analysis, Methodology, Software, Supervision, Writing—original draft, Writing—review & editing), Eva Lindberg (Supervision, Writing—review & editing), Henrik J. Persson (Funding acquisition, Project administration, Software, Supervision, Writing—review & editing), Svetlana Saarela (Conceptualization, Funding acquisition, Methodology), and Mats Nilsson (Conceptualization, Data curation, Methodology, Software, Supervision, Writing—review & editing)

Supplementary data

Supplementary data are available at *Forestry Journal* online.

Conflict of interest

None declared.

Funding

This work was supported by Formas grant (FR-2019/0007 to S.S.), the Bo Rydin Foundation for Scientific Research (award no F06/21 to H.J.P.), and Mistra Digital Forest (DIA 2017/14 #6 to H.J.P.).

Data availability

The R script with the modifications made to the existing R package 'ciTools' is available. The data and other R scripts underlying this article will be shared on reasonable request to the authors.

References

Andersen HE, McGaughey RJ, Reutebuch SE. Estimating forest canopy fuel parameters using LIDAR data. *Remote Sens Environ* 2005;**94**:441–9. https://doi.org/10.1016/j.rse.2004.10.013.

Axelsson A, Lindberg E, Reese H. *et al.* Tree species classification using Sentinel-2 imagery and Bayesian inference. *Int J Appl Earth Obs Geoinf* 2021;**100**:102318. https://doi.org/10.1016/j.jag.2021.102318.

Babcock C, Finley AO, Andersen HE. *et al.* Geostatistical estimation of forest biomass in interior Alaska combining Landsat-derived tree cover, sampled airborne lidar and field observations. *Remote Sens Environ* 2018;**212**:212–30. https://doi.org/10.1016/j.rse.2018.04.044.

Casella G, Berger RL, Santana D. *Statistical Inference*. USA: Cengage Learning, 2002.

Condés S, McRoberts RE. Updating national forest inventory estimates of growing stock volume using hybrid inference. *For Ecol Manage* 2017;**400**:48–57. https://doi.org/10.1016/j.foreco.2017.04.046.

Corona P, Fattorini L, Franceschi S. *et al.* Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Can J For Res* 2014;**44**:1303–11. https://doi.org/10.1139/cjfr-2014-0203.

Dunn PK, Smyth GK. *Generalized Linear Models with Examples in R, Springer Texts in Statistics*. New York, NY: Springer New York, 2018.

Esteban J, McRoberts RE, Fernández-Landa A. *et al.* A model-based volume estimator that accounts for both land cover misclassification and model prediction uncertainty. *Remote Sens (Basel)* 2020, 2020;**12**:3360.

Esteban J, McRoberts RE, Fernández-Landa A. *et al.* Estimating forest volume and biomass and their changes using random forests and remotely sensed data. *Remote Sens (Basel)* 2019, 2019;**11**:1944.

FAO. *The State of World's Forests 2018 - Forest Pathways to Sustainable Development*. Rome: FAO, 2018.

Fassnacht FE, White JC, Wulder MA. *et al.* Remote sensing in forestry: current challenges, considerations and directions. *For An Int J For Res* 2023;**97**:11–37. https://doi.org/10.1093/forestry/cpad024.

Franco-Lopez H, Ek AR, Bauer ME. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens Environ* 2001;**77**:251–74. https://doi.org/10.1016/S0034-4257(01)00209-7.

Gobakken T, Næsset E. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scand J For Res* 2006;**19**:529–42.

Gobakken T, Næsset E, Nelson R. *et al.* Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sens Environ* 2012;**123**:443–56. https://doi.org/10.1016/j.rse.2012.01.025.

Gregoire TG. Design-based and model-based inference in survey sampling: appreciating the difference. *Can J For Res* 1998;**28**:1429–47. https://doi.org/10.1139/x98-166.

Haakana H, Heikkinen J, Katila M. *et al.* Efficiency of post-stratification for a large-scale forest inventory-case Finnish NFI. *Ann For Sci* 2019;**76**. https://doi.org/10.1007/s13595-018-0795-6.

Haman, J. Generalized Linear Models with ciTools (WWW Document). 2017. https://cran.r-project.org/web/packages/ciTools/vignettes/ciTools-glm-vignette.html (22 April 2022, date last accessed).

Hartig F. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models*. R package version 0.4.6. 2022. http://florianhartig.github.io/DHARMa/.

Hattab MW. A derivation of prediction intervals for gamma regression. *J Stat Comput Simul* 2016;**86**:3512–26. https://doi.org/10.1080/00949655.2016.1169421.

Holmgren J. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scand J For Res* 2006;**19**:543–53. https://doi.org/10.1080/02827580410019472.

Hollaus M, Dorigo W, Wagner W, Schadauer K, Höfle B, Maier B. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *Int J Remote Sens* 2009;**30**:5159–75.

Hu B, Shao J, Palta M. Pseudo-R2 in logistic regression model. *Stat Sin* 2006;**16**:847–60.

Hudak AT, Crookston NL, Evans JS. *et al.* Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sens Environ* 2008;**112**:2232–45. https://doi.org/10.1016/j.rse.2007.10.009.

Hudak AT, Crookston NL, Evans JS. *et al.* Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Can J Remote Sens* 2006;**32**:126–38.

IPCC. Climate Change 2022: Impacts, Adaptation, and Vulnerability. In: Pörtner H-O, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A, Craig M, Langsdorf S, Löschke S, Möller V, Okem A, Rama B (eds.), *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022, pp. 3056. https://doi.org/10.1017/9781009325844.

Lantmäteriet. Lantmäteriet - we know every single place in Sweden. 2019. | Lantmäteriet (WWW Document). https://www.lantmateriet.se/sv/ (8 September 2021, date last accessed).

Lefsky MA, Cohen WB, Harding DJ. *et al.* Lidar remote sensing of above-ground biomass in three biomes. *Glob Ecol Biogeogr* 2002;**11**:393–9.

Lim KS, Treitz PM. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scand J For Res* 2006;**19**:558–70. https://doi.org/10.1080/02827580410019490.

Lumley T, Miller A. *leaps: Regression Subset Selection*, 2022.

Maltamo M, Bollandsås OM, Gobakken T. *et al.* Large-scale prediction of aboveground biomass in heterogeneous mountain forests by means of airborne laser scanning. *Can J For Res* 2016;**46**:1138–44. https://doi.org/10.1139/cjfr-2016-0086.

Mauya EW, Hansen EH, Gobakken T. *et al.* Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. *Carbon Balance Manag* 2015;**10**:1–14. https://doi.org/10.1186/s13021-015-0021-x.

McFadden DL. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. New York: Academic Press, 1974; 105–42.

McGaughey B. *FUSION/LDV LIDAR Analysis and Visualization Software*. Pacific Northwest Res. Stn. USDA For. Serv, 2020.

McRoberts RE. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sens Environ* 2010;**114**:1017–25.

McRoberts RE, Cohen WB, Erik N. *et al.* Using remotely sensed data to construct and assess forest attribute maps and related spatial products. *Scand J For Res* 2010;**25**:340–67.

McRoberts RE, Wendt DG, Nelson MD. *et al.* Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sens Environ* 2002;**81**:36–44. https://doi.org/10.1016/S0034-4257(01)00330-3.

Millard SP, Neerchal NK. *Environmental Statistics with S-PLUS*. CRC Press, 2000.

Næsset E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field

data. *Remote Sens Environ* 2002;**80**:88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.

Næsset E. Estimating above-ground biomass in young forests with airborne laser scanning. *Int J Remote Sens* 2011;**32**:473–501. https://doi.org/10.1080/01431160903474970.

Næsset E, Bollandsås OM, Gobakken T. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sens Environ* 2005;**94**:541–53. https://doi.org/10.1016/j.rse.2004.11.010.

Næsset E, Gobakken T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens Environ* 2008;**112**:3079–90. https://doi.org/10.1016/j.rse.2008.03.004.

Næsset E, Gobakken T, Solberg S. *et al.* Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: a case study from a boreal forest area. *Remote Sens Environ* 2011;**115**:3599–614. https://doi.org/10.1016/j.rse.2011.08.021.

National Land Cover Database. *The Swedish Environmental Protection Agency*, 2020.

Nelson R, Krabill W, Tonelli J. Estimating forest biomass and volume using airborne laser data. *Remote Sens Environ* 1988;**24**:247–67. https://doi.org/10.1016/0034-4257(88)90028-4.

Nilsson M. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens Environ* 1996;**56**:1–7. https://doi.org/10.1016/0034-4257(95)00224-3.

Nilsson M, Nordkvist K, Jonzén J. *et al.* A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sens Environ* 2017;**194**:447–54. https://doi.org/10.1016/j.rse.2016.10.022.

Nord-Larsen T, Schumacher J. Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. *Remote Sens Environ* 2012;**119**:148–57.

Packalén P, Maltamo M. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sens Environ* 2007;**109**:328–41. https://doi.org/10.1016/j.rse.2007.01.005.

Persson HJ, Fransson JES. Comparison between TanDEM-X and ALS based estimation of above ground biomass and tree height in boreal forests. *Scand J For Res* 2017;**32**:306–19. https://doi.org/10.1080/02827581.2016.1220618.

Persson HJ, Olofsson K, Holmgren J. Two-phase forest inventory using very-high-resolution laser scanning. *Remote Sens Environ* 2022;**271**:112909. https://doi.org/10.1016/j.rse.2022.112909.

Persson HJ, Soja MJ, Fransson JES. *et al.* National Forest Biomass Mapping Using the two-level model. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2020;**13**:6391–400. https://doi.org/10.1109/JSTARS.2020.3030591.

Saarela S, Grafström A, Ståhl G. *et al.* Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. *Remote Sens Environ* 2015;**158**:431–40. https://doi.org/10.1016/j.rse.2014.11.020.

Saarela S, Holm S, Healey SP. *et al.* Comparing frameworks for biomass prediction for the global ecosystem dynamics investigation. *Remote Sens Environ* 2022;**278**:113074. https://doi.org/10.1016/j.rse.2022.113074.

Saarela S, Wästlund A, Holmström E. *et al.* Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. *For Ecosyst* 2020;**7**. https://doi.org/10.1186/s40663-020-00245-0.

Särndal C-E, Thomsen I, Hoem JM. *et al.* Design-based and model-based inference in survey sampling (with discussion and reply). *Scand J Stat* 1978;**5**:27–52.

Schuh M, Favarin JAS, Marchesan J. *et al.* Machine learning and generalized linear model techniques to predict above-ground biomass in Amazon rainforest using LiDAR data. *J Appl Remote Sens* 2020;**14**:034518. https://doi.org/10.1117/1.JRS.14.034518.

Sims NC, Newnham GJ, England JR. *et al.* Good practice guidance proportion of land that is degraded over total land area. United Nations Convention to Combat Desertification (UNCCD). 2021.

Skogsstyrelsen. *Skogsstyrelsens Geodata - Beskrivningar (WWW Document)*. 2023. https://www.skogsstyrelsen.se/sjalvservice/karttjanster/skogsstyrelsens-geodata/.

Ståhl G, Holm S, Gregoire TG. *et al.* Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Can J For Res* 2011a;**41**:83–95. https://doi.org/10.1139/X10-195.

Ståhl G, Holm S, Gregoire TG. *et al.* Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark county, Norway. *Can J For Res* 2011b;**41**:96–107. https://doi.org/10.1139/X10-161.

Ståhl G, Saarela S, Schnell S. *et al.* Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *For Ecosyst* 2016;**3**:1–11. https://doi.org/10.1186/s40663-016-0064-9.

Team, R.C. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team, 2021.

Thompson ME, Cassel C-M, Sarndal C-E. *et al.* Foundations of inference in survey sampling. *J Am Stat Assoc* 1982;**77**:212. https://doi.org/10.2307/2287794.

Tomppo E, Olsson H, Ståhl G. *et al.* Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens Environ* 2008;**112**:1982–99.

Tomppo E, Schadauer K, McRoberts RE. *et al. National Forest Inventories, National Forest Inventories: Pathways for Common Reporting*. The Netherlands: Springer, 2010.

UNFCCC. Kyoto Protocol - Targets for the first commitment period | UNFCCC (WWW Document). 2021. https://unfccc.int/process-and-meetings/the-kyoto-protocol/what-is-the-kyoto-protocol/kyoto-protocol-targets-for-the-first-commitment-period (12 May 2021, date last accessed).

Valinger E, Kempe G, Fridman J. Impacts on forest management and forest state in southern Sweden 10 years after the storm Gudrun. *For An Int J For Res* 2019;**92**:481–9. https://doi.org/10.1093/forestry/cpz005.

Vauhkonen J, Maltamo M, McRoberts RE. *et al. Introduction to Forestry Applications of Airborne Laser Scanning*. Dordrecht: Springer, 2014. https://doi.org/10.1007/978-94-017-8663-8_1.

Venables B, Ripley BD. *Statistical Analysis of Financial Data in S-Plus, Statistical Analysis of Financial Data in S-Plus*. vol. **27**. New York: Springer, 2004; 1–16.

Wulder MA, White JC, Fournier RA. *et al.* Spatially explicit large area biomass estimation: three approaches using forest inventory and remotely sensed imagery in a GIS. *Sensors* 2008;**8**:529–60. https://doi.org/10.3390/s8010529.

Zhang H, Zimmerman J, Nettleton D. *et al.* Random forest prediction intervals. *Am Stat* 2019;**74**:392–406. https://doi.org/10.1080/00031305.2019.1585288.

Zuur AF, Ieno EN, Walker N. *et al. Mixed Effects Models and Extensions in Ecology with R, Statistics for Biology and Health*. New York, NY, Springer New York, 2009.

Supplementary material

Appendix

Prediction intervals for gamma regression with a square-root link

A fitted regression model is often used for prediction of a new individual response y_0 when the explanatory variable vector is equal to $\boldsymbol{x}_0$. Such a predicted value is often supplemented by a prediction interval. For gamma regression models with a square-root link, we will investigate two methods for computing prediction intervals. One of the methods is provided through the R package ciTools [Haman, 2017] while the other method is obtained below by extending the method suggested by [Hattab, 2016] from a log link to a square-root link.

In a generalized linear model (GLM), the response variable y is assumed to follow a particular distribution in the exponential family of distributions. Members of this family of distributions have probability density functions that can be expressed in the form

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right),$$

where $a(\cdot), b(\cdot)$, and $c(\cdot)$ are functions, and θ and ϕ are parameters [e.g., Myers et al., 2013]. One member of this family is the gamma distribution, which has the density

$$f(y) = \frac{1}{\Gamma(r)\lambda^r} y^{r-1}\exp\left(-\frac{y}{\lambda}\right),$$

where $\Gamma(r)$ is the gamma function. The mean of y is $\mu = r\lambda$ and the variance is $Var(y) = r\lambda^2$. That the density function of the gamma distribution is in the form of the exponential family follows by setting $\theta = -(\lambda r)^{-1}, a(\phi) = r^{-1}, b(\theta) = -\ln(-\theta)$, and $c(y,\phi) = r\ln r - \ln\Gamma(r) + (r-1)\ln y$, which implies that

$$\mu = r\lambda$$

and

$$\mu = r\lambda$$

To denote that y is gamma distributed with parameters r and λ, we write $y \sim \text{Gamma}(r,\lambda)$.

With independent response observations, $y_1, \ldots, y_n$, a square-root link gives the following model for the mean,

$$\mu_i^{1/2} = \boldsymbol{x}_i'\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is the vector of coefficients of the regression model and $\boldsymbol{x}_i$ is a vector of explanatory variables, $i = 1, \ldots, n$. The asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ using this link is given by

$$I^{-1}\left(\widehat{\boldsymbol{\beta}}\right) = (\boldsymbol{X}'\Delta\boldsymbol{V}\Delta\boldsymbol{X})^{-1}[a(\phi)]^2,$$

where $I\left(\widehat{\boldsymbol{\beta}}\right)$ is the information matrix,

$$\Delta = \text{diag}\left\{\frac{\partial\theta_i}{\partial\boldsymbol{x}_i'\boldsymbol{\beta}}\right\} = \text{diag}\left\{\frac{2}{\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)^3}\right\} = \text{diag}\left\{\frac{2}{\mu_i^{3/2}}\right\},$$

and

$$\boldsymbol{V} = \text{diag}\{Var(y_i)\} = \text{diag}\left\{\frac{\mu_i^2}{r}\right\}$$

[e.g., Myers et al., 2013]. Thus, with a square-root link we have

$$\Delta \boldsymbol{V} \Delta = \text{diag}\left\{\frac{4}{r\mu_i}\right\}.$$

Let y_0 denote the response value for a given explanatory vector $\boldsymbol{x}_0$. The corresponding predicted value is $\hat{y}_0 = \left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^2$. By Lewis, Montgomery, and Myers [2001], the asymptotic distribution of

$$\frac{\hat{y}_0 - (\boldsymbol{x}_0'\boldsymbol{\beta})^2}{\sqrt{\boldsymbol{d}_0' I^{-1}(\widehat{\boldsymbol{\beta}})\boldsymbol{d}_0}}$$

is $N(0,1)$, where

$$\boldsymbol{d}_0 = \frac{\partial \hat{\mu}_0}{\partial \widehat{\boldsymbol{\beta}}}$$

and, by the chain rule,

$$\frac{\partial \mu_0}{\partial \boldsymbol{\beta}} = \frac{\partial \mu_0}{\partial \theta_0}\frac{\partial \theta_0}{\partial \boldsymbol{x}_0'\boldsymbol{\beta}}\frac{\partial \boldsymbol{x}_0'\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \frac{Var(y_0)}{a(\phi)}\frac{2}{(\boldsymbol{x}_0'\boldsymbol{\beta})^3}\boldsymbol{x}_0 = (\boldsymbol{x}_0'\boldsymbol{\beta})^4\frac{2}{(\boldsymbol{x}_0'\boldsymbol{\beta})^3}\boldsymbol{x}_0 = 2(\boldsymbol{x}_0'\boldsymbol{\beta})\boldsymbol{x}_0.$$

By the consistency of $\widehat{\boldsymbol{\beta}}$ and Slutsky's theorem [e.g., Gut 2013],

$$\frac{\hat{y}_0(\boldsymbol{x}_0'\boldsymbol{\beta})^{-2} - 1}{\sqrt{4\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^{-2}\boldsymbol{x}_0' I^{-1}(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_0}}$$

is approximately normally distributed with mean 0 and variance 1, and $\hat{y}_0(\boldsymbol{x}_0'\boldsymbol{\beta})^{-2}$ has approximately the same distribution as

$$z\sqrt{4\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^{-2}\boldsymbol{x}_0' I^{-1}(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_0} + 1$$

where z is a standard normal random variable.

By our assumptions,

$$y_0 \sim \text{Gamma}(r, (\boldsymbol{x}_0'\boldsymbol{\beta})^2/r),$$

which implies that $2ry_0(\boldsymbol{x}_0'\boldsymbol{\beta})^{-2}$ is chi-squared distributed with $2r$ degrees of freedom [Bain, 2006]. Finally, we proceed in a similar way as in Hattab [2016], and set

$$G = \frac{\chi^2_{2\hat{r}}}{z\sqrt{4\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^{-2}\boldsymbol{x}_0' I^{-1}(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_0} + 1}, \tag{1}$$

where $\chi^2_{2\hat{r}}$ denotes a chi-squared distributed random variable with $2\hat{r}$ degrees of freedom, independent of z. Let $G_{(0.025)}$ and $G_{(0.975)}$ denote the 0.025 and 0.975 quantiles of the distribution of G. Then,

$$P\left(G_{(0.025)} \leq \frac{2ry_0(\boldsymbol{x}_0'\boldsymbol{\beta})^{-2}}{\hat{y}_0(\boldsymbol{x}_0'\boldsymbol{\beta})^{-2}} \leq G_{(0.975)}\right) = P\left(\frac{G_{(0.025)}\hat{y}_0}{2r} \leq y_0 \leq \frac{G_{(0.975)}\hat{y}_0}{2r}\right) \approx 0.95$$

for large sample sizes n. Estimates $\hat{G}_{(0.025)}$ and $\hat{G}_{(0.975)}$ of the quantiles can be found using simulations, and an approximate 95% prediction interval is then obtained as

$$\left(\frac{\hat{G}_{(0.025)}\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^2}{2\hat{r}}, \frac{\hat{G}_{(0.975)}\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^2}{2\hat{r}}\right) \tag{2}$$

By arguing as in Hattab [2016], we will also consider a slight modification of the prediction interval in (2), obtained by replacing z in (1) by t_{n-p}, which denotes a t-distributed random variable with $n-p$ degrees of freedom. That is, let

$$H = \frac{\chi^2_{2r}}{t_{n-p}\sqrt{4\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^{-2}\boldsymbol{x}_0' I^{-1}\left(\widehat{\boldsymbol{\beta}}\right)\boldsymbol{x}_0 + 1}}, \tag{3}$$

and let $\hat{H}_{(0.025)}$ and $\hat{H}_{(0.975)}$ denote estimates of the 0.025 and 0.975 quantiles of the distribution of H. The modified prediction interval is then given by

$$\left(\frac{\hat{H}_{(0.025)}\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^2}{2\hat{r}}, \frac{\hat{H}_{(0.975)}\left(\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}\right)^2}{2\hat{r}}\right), \tag{4}$$

A simulation study

In this Monte Carlo simulation study, we compared three different methods for computing prediction intervals for gamma regression models with a square-root link:

- $\text{PI}_{\text{ciTools}}$, which is the prediction interval obtained through the R package ciTools;
- PI_{norm}, which is the interval given by (2); and
- PI_{t}, which is the interval given by (4).

In the simulations, response data were generated according to the fitted model in Table 2 in the main text. Prediction intervals were computed for the following cases:

Case 1: h_{90} = 5.144 and v = 45.66474

Case 2: h_{90} = 6.536 and v = 60.16260

Case 3: h_{90} = 9.050 and v = 68.68687

Case 4: h_{90} = 10.670 and v = 75.54348

Case 5: h_{90} = 12.364 and v = 85.42714

Case 6: h_{90} = 14.488 and v = 88.34951

Case 7: h_{90} = 15.270 and v = 94.41624

Case 8: h_{90} = 17.076 and v = 102.5210

Case 9: h_{90} = 20.414 and v = 109.3407

For Case j, the value of h_{90} was chosen to be approximately equal to the jth decile of the values of h_{90} in our sample data from the 500 NFI plots, where $j = 1, \ldots, 9$. The corresponding values of v were chosen likewise. The explanatory variables used in the models were $x_1 = h_{90}$, $x_2 = v$ and $x_3 = h_{90} \times v$.

The following algorithm for performing the simulations was considered:

1. From the fitted model in Table 2, generate a response observation corresponding to case j.
2. Generate a random subsample of size m without replacement from the set of n observations of $\boldsymbol{x} = (x_1, x_2, x_3)$.
3. For each $i = 1, \ldots, 1000$:
   a) From the fitted model in Table 2, generate a response observation corresponding to each observation in the subsample from step 2.
   b) Based on the generated data in 2 and 3a), fit a gamma regression model with a square-root link and with x_1, x_2, and x_3 as explanatory variables.
   c) Compute 95% prediction intervals $PI_{ciTools}$, PI_{norm}, and PI_t for case j.
4. For each case j and each prediction interval method in 3c), compute the proportion of times the response observation in step 1 is covered by the interval.
5. For each case j and each prediction interval method in 3c), compute the average width of the interval.

As the n observations of $\boldsymbol{x} = (x_1, x_2, x_3)$, we used all but five of the observations corresponding to the 500 NFI plots. The removed observations were identified as outliers. Thus, $n = 495$. Estimated quantiles of the distributions of G and H were in each case obtained from 1 million replicates of G and H, respectively. Ideally, the proportions computed in step 4 of the algorithm should be as close as possible to the nominal level 0.95.

The results of the simulations are presented in Tables A and B.

Table A. Estimated coverage probabilities from Algorithm 2.

		Case								
PI	m	1	2	3	4	5	6	7	8	9
$PI_{ciTools}$	495	0.955	0.950	0.955	0.953	0.949	0.956	0.952	0.957	0.953
PI_{norm}	495	0.954	0.952	0.951	0.956	0.945	0.958	0.951	0.956	0.950
PI_t	495	0.954	0.952	0.950	0.956	0.945	0.957	0.951	0.956	0.950
$PI_{ciTools}$	50	0.935	0.951	0.953	0.929	0.943	0.938	0.930	0.943	0.939
PI_{norm}	50	0.905	0.936	0.940	0.917	0.938	0.926	0.921	0.937	0.922
PI_t	50	0.905	0.935	0.940	0.917	0.938	0.926	0.922	0.937	0.922
$PI_{ciTools}$	10	0.889	0.891	0.894	0.902	0.911	0.900	0.901	0.881	0.902
PI_{norm}	10	0.434	0.697	0.776	0.804	0.811	0.798	0.801	0.761	0.766
PI_t	10	0.434	0.697	0.778	0.804	0.812	0.798	0.801	0.761	0.766

Table B. Average widths of prediction intervals from Algorithm 2.

		Case								
PI	m	1	2	3	4	5	6	7	8	9
$PI_{ciTools}$	495	23.8	41.4	58.3	73.8	100.1	124.0	148.6	186.8	241.7
PI_{norm}	495	23.7	41.3	58.2	73.6	100.2	123.8	148.2	186.5	241.6
PI_t	495	23.7	41.3	58.2	73.6	100.2	123.8	148.2	186.5	241.6
$PI_{ciTools}$	50	24.6	42.4	58.5	74.7	100.6	124.5	148.6	187.8	247.7
PI_{norm}	50	22.7	39.6	55.1	70.7	95.6	118.9	141.4	178.1	231.4
PI_t	50	22.7	39.6	55.1	70.7	95.6	118.9	141.4	178.1	231.5
$PI_{ciTools}$	10	62.0	56.7	62.8	75.1	101.7	124.7	151.6	192.5	266.4
PI_{norm}	10	19.4	31.5	43.2	55.0	75.5	92.3	110.8	137.2	179.8
PI_t	10	19.4	31.6	43.3	55.1	75.6	92.4	110.9	137.3	180.4

For the larger sample size, $m = 495$, $PI_{ciTools}$, PI_{norm}, and PI_t all have estimated coverage probabilities close to the nominal level 0.95 (Table A), and are equally good with respect to average widths of intervals (Table B). For the medium sample size, $m = 50$, PI_{norm} and PI_t produce shorter prediction intervals on average than $PI_{ciTools}$, but their estimated coverage probabilities are a bit too low. In comparison, the estimated coverage probabilities for $PI_{ciTools}$ are the better ones when $m = 50$, and much better when $m = 10$. In the latter case, the estimated coverage probabilities for $PI_{ciTools}$ are about 0.90, which is too low, but the corresponding results for PI_{norm} and PI_t are worse. Another advantage with the intervals $PI_{ciTools}$ is that they are faster to compute.

References

Bain, L., 2006. Gamma Distribution. Kotz S, Balakrishnan N, Read C, Vidakovic B (editors), Encycl. Stat. Sci. (2nd Ed. Hoboken NJ John Wiley Sons.

Gut, A., 2013. Probability: A Graduate Course, Springer Texts in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-4708-5

Haman, J., 2017. Generalized Linear Models with ciTools [WWW Document]. URL https://cran.r-project.org/web/packages/ciTools/vignettes/ciTools-glm-vignette.html (accessed 4.22.22).

Hattab, M.W., 2016. A derivation of prediction intervals for gamma regression. J. Stat. Comput. Simul. 86, 3512–3526. https://doi.org/10.1080/00949655.2016.1169421

Lewis, S.L., Montgomery, D.C., Myers, R.H., 2001. Examples of designed experiments with nonnormal responses. J. Qual. Technol. 33, 265–278. https://doi.org/10.1080/00224065.2001.11980078

Myers, J.L., Well, A.D., Lorch, R.F., 2013. Research design and statistical analysis, third edition, Research Design and Statistical Analysis, Third Edition. Taylor and Francis. https://doi.org/10.4324/9780203726631

III

COMPARING TANDEM-X INSAR FOREST STAND VOLUME PREDICTION MODELS TRAINED USING FIELD AND ALS DATA

Ritwika Mukhopadhyay[1], Mats Nilsson[1], Magnus Ekström[1,2], Eva Lindberg[1], Henrik J. Persson[1]

[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden
[2]Department of Statistics, USBE, Umeå University, Sweden

ABSTRACT

Remote sensing (RS) techniques have been used for mapping forest variables, such as stem volume (important for forest management activities associated with timber production), over large areas which can be updated more frequently than with field inventory (FI) data. In this study, wall-to-wall TanDEM-X synthetic aperture radar images were used as auxiliary RS data for model-based prediction of stand-level volumes for two models, trained using volumes computed from FI (A) and airborne laser scanning estimations (B), respectively. The models were validated with harvester data available for independent stands. It was observed that the performance of model B was slightly better compared to model A based on adjusted R^2 and root mean squared error values. Therefore, it can be concluded that a completely RS based approach for prediction and mapping of stand volumes would be as promising as a method based on FI data along with being cost- and labour-efficient.

Index Terms— Airborne laser scanning, harvester data, stand volume, synthetic aperture radar, TanDEM-X.

1. INTRODUCTION

The use of auxiliary remote sensing (RS) data has been increasing over the past decades. RS data have been paired with field reference datasets for estimating forest variables such as, height, volume and aboveground biomass (AGB), in, e.g., [1], [2], [3]–[10], [11]–[17]. Stand- and tree-level volume estimates have been more extensively used by forest owners for measuring merchantable timber and forest products [18]. The collection of field inventory data is more cost and labour intensive when compared to acquiring aerial and spaceborne RS data, especially, in remote and inaccessible terrains.

Airborne laser scanning (ALS) data have been previously used for large scale mapping of AGB and volume [19], [20], but, using a wall-to-wall satellite image for large scale mapping of such forest variables would be even more time efficient [21], [22]. The German synthetic aperture radar (SAR) mission – TanDEM-X constitutes of a pair of satellites (launched in 2007 and 2010) carrying X-band SAR sensors, flying as an interferometer with global coverage. Several studies have been conducted using TanDEM-X for model-based inferences of forest AGB, volume and tree heights, in, e.g., [18], [23]–[29]. The use of TanDEM-X enables extraction of the vegetation height using the interferometric phase height (ph) and coherence [22]. The ph contains information about tree heights as well as forest density, which can be correlated to AGB and volume estimations [18], [22]. In some previous studies by [18], [30], these interferometric variables have been used to estimate forest volume at the stand level. Until now, manual field inventory data have been used as reference data for model-based estimations of volume, except for a handful of studies [2]. Still, models entirely based on RS data have not been implemented for large-scale mapping of forest variables.

Therefore, the main objective of this study was to use wall-to-wall TanDEM-X SAR data for modelling and mapping of stand-level forest volume across the entire test site and evaluating the performance of models A and B trained with stand-level volumes estimated from forest inventory data and volumes estimated based on ALS and national forest inventory (NFI) data, respectively. Stand volumes for independent forest stands within the same test site, estimated from harvesters during felling activities, were used for validating the models.

2. MATERIALS AND METHOD

2.1. Study area

The test site located in central Sweden, covers an area of 50,000 ha approximately (as shown in Fig 1a-b), majorly dominated by Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), Lodgepole pine (*Pinus contorta*) and other deciduous tree species.

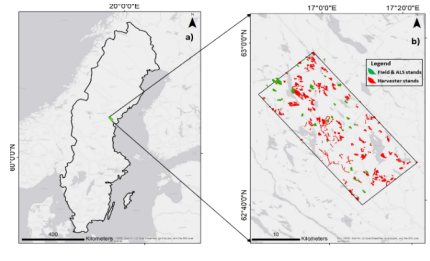


Figure 1. Represents the test site constituting the forest stands used for training models A and B (marked with 'green') and forest stands used for validation of the models available from the harvester data (marked with 'red').

Pine constituting around 50%, spruce around 44% and deciduous and other tree species constituting about 6% of the total species composition [2].

2.2. Field and Remote Sensing data

The field inventory data were acquired in 2019. Thirty stands were inventoried within the test site (marked with 'green' in Fig 1b). The inventory was done with an average of 8 circular plots (with 8m radius) distributed systematically across the stands. The distance between the plots and the number of plots varied for each stand depending on the stand area. The field inventory data were used as reference to select these 30 stands as training data since the field inventory data constituted observations over entire range of age-classes in the stands.

The ALS data were acquired in 2019 by a Leica ALS80 sensor from a 3000 m flight height and having an average point density of 1.5 points/m^2 scanning over entire Sweden. The mean volumes (in m^{3}ha^{-1}) for the same individual 30 stands were extracted from the volumes estimated using the ALS metrics and the corresponding NFI plots over the scanned regions.

The TanDEM-X dataset was acquired on 14th November 2015 for HH (horizontally transmitted and horizontally received) polarization in strip-map mode. The SAR data specifications have been mentioned in Table 1.

Table 1. TanDEM-X SAR data specifications.

Polariz-ation	Pixel resolution		Multilook factor
	SLC*	Resampled	
HH	2.5(slant)×3.3 (azimuth) m^2	10×10 m^2	5×5

*SLC – single look complex.

The pre-processing was done as explained in [22] and similar image variables were derived, namely backscatter, ph and corrected coherence (c_coh). The complex interferogram was obtained as

$$\tilde{\gamma} = \frac{E[s_1 s_2^*]}{\sqrt{E[|s_1|^2]E[|s_2^*|^2]}} \quad (1)$$

where, $\tilde{\gamma}$ is the complex correlation co-efficient, $E[.]$ is the expectation value, * is the complex conjugate and s_1 and s_2 are the Hermitian product of the two complex SAR images [18], [31].

A minimum cost flow function was used for unwrapping the phase followed by a phase-to-height sensitivity raster to obtain the height from the interferometric phase information [22].

The validation dataset consisted of 151 stands with volumes estimated from the harvester data (marked with 'red' in Fig 1b) acquired between 2019 and 2022. The average stand-level volume for the entire test site is 172.5 m^{3}ha^{-1} based on the harvester data accounting for only matured trees in stands. These 151 stands were checked and categorised into thinned (28) and clear-felled (123) stands. The thinned stands were discarded from the dataset to avoid representing over-estimated stand-volumes for such stands.

2.3. Volume estimation models

All the parameters derived from the TanDEM-X data were tested for statistical significance as model co-efficients of the explanatory variables for both model A and B. The final regression models A and B, represented in Eq. 2,

$$vol = \beta_0 + \beta_1 ph^{0.5} + \varepsilon \quad (2)$$

where, 'vol' represents the response variable (stand volume in this case), β_0 and β_1 are the model coefficients and ε is the random error.

The ph values ranged between [-5, 28]. Models A and B were compared based on adjusted co-efficient of determination (adj-R^2) and root mean squared error (RMSE) values with

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

where y is the reference values, $\hat{y}$ is the predicted values, and n is the number of stands in the validation dataset.

3. RESULTS, DISCUSSION AND CONCLUSIONS

The results show that for both models A and B, expected value of stand volumes had dependency on the ph with a power of 0.5, as represented in Table 2. The power value of 0.5 for ph was a slight deviation compared to previous studies, e.g., [18], [22], [28]. The corresponding model coefficients of ph and c_coh were statistically significant in the models but c_coh did not contribute in improving the model prediction accuracy, as observed similar to [22], and was therefore, not included in the models. Also, both models A and B were formulated with intercepts non-significantly different from 0. The statistical summary of model validation has been presented in Table 3. The predictions of model B were slightly more accurate based on the adj-R^2 and RMSE values when compared to that of Model A. From Fig 2, there is no distinct difference between models A and B in the trend of the two plots representing the relation between the predicted stand volumes against the observed stand volumes. The predicted values in the lower range of volume were over-estimated in both cases. The overestimation of these stands with low

stand volumes might be due to that thinning activities were carried out which could not be filtered out during the categorising of the validation dataset or might be due to measurement errors in the harvester volumes. The relationship between the predicted and the observed volumes are almost linear for stand volumes ranging between 150 m^3ha^{-1} and 400 m^3ha^{-1} . Stand volume maps for the entire test site were predicted for both the models, as presented in Fig 3.

The approach based on laser scanning estimates (method B) appeared as accurate and promising as the field inventory based approach (method A) for mature stands. Therefore, this study indicates that dedicated field inventories intended to train models to estimate forest volume wall-to-wall based on TanDEM-X could be eliminated and replaced by estimates from low-resolution LiDAR and NFI data. This reduces costs and makes forest planning and decision making more efficient without compromising the accuracy of the stand volume estimates. The approach needs to be tested further in other sites and the temporal robustness of the model parameters needs to be further analysed.

Table 2. Summary of model parameters.

Model	**Intercept**	**$ph^{0.5}$**
A	-82.37$^{\bullet}$	113.2***
B	-62.30$^{\bullet}$	104.3***

$^{\bullet}$= p>0.05 and ***=p≤0.001.

Table 3. Summary of statistics of volume prediction model validation.

Model	**adj-R^2**	**RMSE (m^3ha^{-1})**	**n**
A	0.58	46.5 (22.3%)	123
B	0.60	44.6 (21.4%)	123

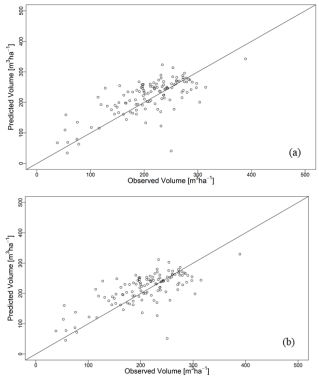


Figure 2. Scatterplots of observed volume vs predicted volume for: (a) model A and (b) model B.

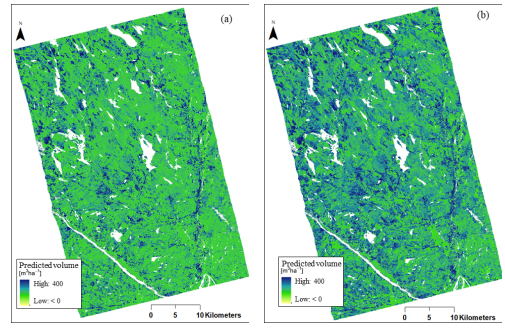


Figure 3. Wall-to-wall prediction maps of volume: (a) model A and (b) model B.

4. ACKNOWLEDGEMENTS

The authors would like to acknowledge the SCA for providing the field inventory and the harvester data and the German Aerospace Center (DLR) for providing the SAR data. We would also like to acknowledge the Bo Rydin Foundation for Scientific Research and the Mistra Digital Forest for funding this work.

5. REFERENCES

[1] J. Holmgren, “Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning,” *Scand. J. For. Res.*, vol. 19, no. 6, pp. 543–553, 2006, doi: 10.1080/02827580410019472.

[2] H. J. Persson, K. Olofsson, and J. Holmgren, “Two-phase forest inventory using very-high-resolution laser scanning,” *Remote Sens. Environ.*, vol. 271, p. 112909, Mar. 2022, doi: 10.1016/J.RSE.2022.112909.

[3] M. J. Soja, H. J. Persson, and L. M. H. Ulander, “Estimation of forest biomass from two-level model inversion of single-pass InSAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5083–5099, 2015, doi: 10.1109/TGRS.2015.2417205.

[4] S. Saarela *et al.*, “Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information,” *Remote Sens. Environ.*, vol. 158, pp. 431–440, Mar. 2015, doi: 10.1016/J.RSE.2014.11.020.

[5] R. O. Dubayah *et al.*, “Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica,” *J. Geophys. Res. Biogeosciences*, vol. 115, no. G2, p. n/a-n/a, Jun. 2010, doi: 10.1029/2009JG000933.

[6] M. Egberth *et al.*, “Combining airborne laser scanning and Landsat data for statistical modeling of soil carbon and tree biomass in Tanzanian Miombo woodlands,” *Carbon Balance Manag.*, vol. 12, no. 1, pp. 1–11, Apr. 2017, doi: 10.1186/S13021-017-0076-Y.

[7] R. Økseter, O. M. Bollandsås, T. Gobakken, and E.

Næsset, “Modeling and predicting aboveground biomass change in young forest using multi-temporal airborne laser scanner data,” *Scand. J. For. Res.*, vol. 30, no. 5, pp. 458–469, Jul. 2015, doi: 10.1080/02827581.2015.1024733.

[8] T. Gobakken, O. M. Bollandsås, and E. Næsset, “Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data,” *Scand. J. For. Res.*, vol. 30, no. 1, pp. 73–86, Jan. 2014, doi: 10.1080/02827581.2014.961954.

[9] E. Næsset, “Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data,” *Remote Sens. Environ.*, vol. 80, no. 1, pp. 88–99, Apr. 2002, doi: 10.1016/S0034-4257(01)00290-5.

[10] J. Esteban, R. E. McRoberts, A. Fernández-Landa, J. L. Tomé, and E. Næsset, “Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data,” *Remote Sens. 2019, Vol. 11, Page 1944*, vol. 11, no. 16, p. 1944, Aug. 2019, doi: 10.3390/RS11161944.

[11] F. Garestier, P. C. Dubois-Fernandez, D. Guyon, and T. Le Toan, “Forest biophysical parameter estimation using L- and P-band polarimetric SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3379–3388, Oct. 2009, doi: 10.1109/TGRS.2009.2022947.

[12] M. A. Stelmaszczuk-Górska, M. Urbazaev, C. Schmullius, and C. Thiel, “Estimation of above-ground biomass over boreal forests in Siberia using updated in Situ, ALOS-2 PALSAR-2, and RADARSAT-2 data,” *Remote Sens.*, vol. 10, no. 10, Oct. 2018, doi: 10.3390/rs10101550.

[13] M. J. Soja, H. J. Persson, and L. M. H. Ulander, “Estimation of forest height and canopy density from a single InSAR correlation coefficient,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 646–650, 2015, doi: 10.1109/LGRS.2014.2354551.

[14] J. Carreiras, J. Melo, and M. Vasconcelos, “Estimating the Above-Ground Biomass in Miombo Savanna Woodlands (Mozambique, East Africa) Using L-Band Synthetic Aperture Radar Data,” *Remote Sens.*, vol. 5, no. 4, pp. 1524–1548, Mar. 2013, doi: 10.3390/rs5041524.

[15] T. Mette, K. Papathanassiou, and I. Hajnsek, “Biomass estimation from polarimetric SAR interferometry over heterogeneous forest terrain,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2004, vol. 1, pp. 511–514, doi: 10.1109/igarss.2004.1369076.

[16] E. Næsset *et al.*, “Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area,” *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3599–3614, Dec. 2011, doi: 10.1016/J.RSE.2011.08.021.

[17] C. Thiel and C. Schmullius, “The potential of ALOS PALSAR backscatter and InSAR coherence for forest growing stock volume estimation in Central Siberia,” *Remote Sens. Environ.*, vol. 173, pp. 258–273, Feb. 2016, doi: 10.1016/J.RSE.2015.10.030.

[18] H. J. Persson and J. E. S. Fransson, “Comparison between TanDEM-X and ALS based estimation of above ground biomass and tree height in boreal forests,” *Scand. J. For. Res.*, vol. 32, no. 4, pp. 306–319, 2017, doi: 10.1080/02827581.2016.1220618.

[19] M. Nilsson *et al.*, “A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory,” *Remote Sens. Environ.*, vol. 194, pp. 447–454, Jun. 2017, doi: 10.1016/j.rse.2016.10.022.

[20] M. Maltamo, O. M. Bollandsås, T. Gobakken, and E. Næsset, “Large-scale prediction of aboveground biomass in heterogeneous mountain forests by means of airborne laser scanning,” *Can. J. For. Res.*, vol. 46, no. 9, pp. 1138–1144, 2016, doi: 10.1139/CJFR-2016-0086.

[21] H. J. Persson, M. J. Soja, J. E. S. Fransson, and L. M. H. Ulander, “National Forest Biomass Mapping Using the Two-Level Model,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6391–6400, 2020, doi: 10.1109/JSTARS.2020.3030591.

[22] H. J. Persson, H. Olsson, M. J. Soja, L. M. H. Ulander, and J. E. S. Fransson, “Experiences from Large-Scale Forest Mapping of Sweden Using TanDEM-X Data,” *Remote Sens. 2017, Vol. 9, Page 1253*, vol. 9, no. 12, p. 1253, Dec. 2017, doi: 10.3390/RS9121253.

[23] J. Praks, M. Hallikainen, O. Antropov, and D. Molina, “Boreal forest tree height estimation from interferometric TanDEM-X images,” *Int. Geosci. Remote Sens. Symp.*, pp. 1262–1265, 2012, doi: 10.1109/IGARSS.2012.6351309.

[24] F. Kugler, D. Schulze, I. Hajnsek, H. Pretzsch, and K. P. Papathanassiou, “TanDEM-X Pol-InSAR performance for forest height estimation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6404–6422, 2014, doi: 10.1109/TGRS.2013.2296533.

[25] J. I. H. Askne, J. E. S. Fransson, M. Santoro, M. J. Soja, and L. M. H. Ulander, “Model-Based Biomass Estimation of a Hemi-Boreal Forest from Multitemporal TanDEM-X Acquisitions,” *Remote Sens. 2013, Vol. 5, Pages 5574-5597*, vol. 5, no. 11, pp. 5574–5597, Oct. 2013, doi: 10.3390/RS5115574.

[26] A. Olesk, J. Praks, O. Antropov, K. Zalite, T. Arumäe, and K. Voormansik, “Interferometric SAR coherence models for Characterization of hemiboreal forests using TanDEM-X dssata,” *Remote Sens.*, vol. 8, no. 9, Sep. 2016, doi: 10.3390/RS8090700.

[27] K. Karila, M. Vastaranta, M. Karjalainen, and S. Kaasalainen, “Tandem-X interferometry in the prediction of forest inventory attributes in managed boreal forests,” *Remote Sens. Environ.*, vol. 159, pp. 259–268, Mar. 2015, doi: 10.1016/J.RSE.2014.12.012.

[28] S. Solberg, R. Astrup, J. Breidenbach, B. Nilsen, and D. Weydahl, “Monitoring spruce volume and biomass with InSAR data from TanDEM-X,” *Remote Sens. Environ.*, vol. 139, pp. 60–67, Dec. 2013, doi: 10.1016/J.RSE.2013.07.036.

[29] S. Abdullahi, F. Kugler, and H. Pretzsch, “Prediction of stem volume in complex temperate forest stands using TanDEM-X SAR data,” *Remote Sens. Environ.*, vol. 174, pp. 197–211, Mar. 2016, doi: 10.1016/J.RSE.2015.12.012.

[30] J. Rahlf, J. Breidenbach, S. Solberg, E. Næsset, and R. Astrup, “Comparison of four types of 3D data for timber volume estimation,” *Remote Sens. Environ.*, vol. 155, pp. 325–333, Dec. 2014, doi: 10.1016/J.RSE.2014.08.036.

[31] J. Sen Lee and E. Pottier, *Polarimetric radar Imaging - from Basics to Applications*. New York, USA: Taylor & Francis Group, 2009.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:96

This thesis explores the use of 3D remote sensing (RS) data from airborne laser scanning, digital aerial photogrammetry, and synthetic aperture radar to predict forest attributes like volume and aboveground biomass. It demonstrates the efficiency of 3D RS data for accurate forest monitoring, supporting large-scale forest management operations, while quantifying uncertainties in predictions and estimations.

Ritwika Mukhopadhyay received her doctoral education at the Department of Forest Resource Management at the Swedish University of Agricultural Sciences (SLU), Umeå. In 2020 she was awarded her Master of Science in Geo-information Science and Earth Observation from University of Twente, The Netherlands, and Indian Institute of Remote Sensing (IIRS-ISRO), India.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-423-9
ISBN (electronic version) 978-91-8046-431-4