

METHODS FOR ENHANCING FOREST INVENTORIES AT DIFFERENT SPATIAL SCALES USING AUXILIARY INFORMATION

METODER FOR Å FORBEDRE SKOGSTAKSERINGEN PÅ FORSKJELLIG ROMLIG SKALA VED
BRUK AV TILLEGGSINFORMASJON

LIVIU THEODOR ENE

Methods for enhancing forest inventories at different spatial scales using auxiliary information

Metoder for å forbedre skogstakseringen på forskjellig romlig skala ved bruk av tilleggsinformasjon

Philosophiae doctor (ph.d.) avhandling

Liviu Theodor Ene

Institutt for Naturforvaltning

Universitetet for miljø- og biovitenskap

Ås 2012



Avhandling nr 2012:19

ISSN 1503-1667

ISBN 978-82-575-1056-5

Supervisory group:

Professor Dr. Erik Næsset (main supervisor)
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
INA, Postbox 5003, 1432 Ås, Norway

Professor Dr. Terje Gobakken (co-supervisor)
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
INA, Postbox 5003, 1432 Ås, Norway

Evaluation committee:

Adjunct Professor Dr. Ron E. McRoberts
University of Minnesota, 15111 Elmcrest Avenue North, Hugo, Minnesota 55038, USA
E-mail: rmcroberts@fs.fed.us

Professor Dr. Annika S. Kangas
Department of Forest Sciences
University of Helsinki, PL 27, FI-00014 University of Helsinki, Finland
E-mail: annika.kangas@helsinki.fi

UMBs committee coordinator:

Professor Dr. Tron Eid
Norwegian University of Life Sciences
INA, Postbox 5003, 1432 Ås, Norway
E-mail: tron.eid@umb.no

Preface

This thesis has been submitted in completion of my PhD studies at the Norwegian University of Life Sciences, Department of Ecology and Natural Resource Management (project no. 166482/i10, project no. no. 184636/S30 and project no. 180645, founded by the Research Council of Norway).

Ås, March 1st, 2012

Contents

Preface	iii
Abstract.....	vi
1. Introduction	3
1.1 Remote sensing support for forest inventories	4
1.1.1 ALS-aided forest inventories	6
1.1.2 Mapping and estimation of forest resources using satellite imagery.....	8
1.2 Research objectives.....	9
2. Materials	11
2.1 Study areas.....	12
2.1 Field data.....	13
2.2 Auxiliary data.....	14
2.2.1 Laser scanning data.....	14
2.1.2 Satellite imagery data.....	15
2.2.3 Cartographical data.....	15
3. Methods.....	16
3.1 Assessing sampling strategies for large-area forest inventories	16
3.1.1 Creating the artificial population	16
3.1.2 Sampling strategy assessment	17
3.2 Small-area estimation	18
3.3 Estimation of tree-level attributes.....	19
4. Results and discussion	20
4.1 Assessing sampling strategies for large-area forest inventories	20
4.2 Small-area estimation	21
4.3 Estimation of tree-level attributes.....	22
5. Conclusions and future research	23
References	26

APPENDIX: Papers I - IV

Abstract

The increasing awareness on the importance of forests and forest management practices to modern society requires accurate and cost-effective methods for monitoring and assessment of forest ecosystems. The four studies included in the thesis are addressing a variety of aspects related to the estimation of forest resources using auxiliary information, considering both the design- and model-based inferential frameworks. More precisely, the case studies described in the present work are focusing on above ground biomass estimation using auxiliary information consisting of airborne laser scanning (ALS) measurements, satellite imagery, and other various cartographical products like elevation data from digital terrain models and land-use maps. The methods discussed in the thesis have a wide applicability, and can be used in relation to forest inventories conducted at national and regional scales, and down to small areas and individual tree level.

Large-scale inventories using ALS as sampling tool have the potential to provide timely and reliable estimates of forest characteristics. Such inventories rely on complex sampling designs for acquiring ground measurements and ALS data, thus assessing the validity of the inference cannot always be performed using analytical approaches. A possible solution in such cases is to rely on simulated sampling for assessing the behaviour of various estimators. A sampling simulator was created using empirical datasets, copula modelling and nearest neighbor imputations. The studies demonstrated that ignoring the underlying assumptions required by the estimators can seriously affect the precision of the estimates (e.g., nearly five times overestimation of standard errors). Besides, simulated sampling can provide the means for choosing the appropriate estimator (and even the right sampling strategy) to be used in a real application. Furthermore, the simulation results demonstrated

that using ALS as sampling tool can be a cost-efficient inventory method for large-area applications.

Estimation of forest resources at local scales is often difficult when the sample sizes are very small or missing entirely. Such situations were address within a model-based framework, where the superpopulation model was replaced by a canonical vine (C-vine) copula. Using simulated sampling from the copula function followed by nearest neighbor predictions, the approach demonstrated a higher accuracy compared to bootstrap resampling, the main improvement consisting in a significant bias reduction.

To meet the demand for detailed information at tree level required by intensive forest management activities, a novel method for tree top detection and extraction of individual tree attributes was developed. Using the stem number estimates provided by area-based inventories and under mild assumptions regarding the spatial process generating the spatial stem distribution, the algorithm demonstrated a robust behaviour by favourably balancing the omission and commission errors in a heterogeneous boreal forest. An important asset of the method is the potential to be seemingly integrated with area-based, ALS-aided operational forest inventories.

The strengths and drawbacks of the methods, as well as further improvements to be considered are also discussed.

Sammendrag

Den økende bevisstheten om betydningen av skogen og ulike forvaltningsmetoder for det moderne samfunnet krever nøyaktige og kostnadseffektive metoder for overvåking og vurdering av skogøkosystemer. De fire studier som inngår i avhandlingen, tar opp en rekke aspekter knyttet til estimering av skogressursene ved hjelp av tilleggsinformasjon, og vurderer både design- og modellbasert rammeverk. Mer presist, er de studiene beskrevet i dette arbeidet fokuserte på biomasseberegning ved hjelp av tilleggsinformasjon som består i målinger ved hjelp av flybåren laserskanning (ALS), satellittbilder og diverse kartografiske produkter, som for eksempel høydedata fra digitale terrengmodeller og arealbrukskart. Metodene som diskuteres i avhandlingen dekker et stort bruksområde, og de kan brukes i forhold til nasjonal og regional skogtaksering, og ned til estimering på små områder og for enkelttrær.

Storskala taksering av skog med ALS som potensial til å gi presise og pålitelige anslag over skogegenskaper. Slike takster krever komplekse samplingsmetoder for å skaffe feltmålinger og ALS data, og dermed kan gyldighetsvurderingen ikke alltid utføres med analytiske tilnærminger. En mulig løsning i slike tilfeller er å bruke simulert sampling for å vurdere oppførselen til ulike estimatorer. En samplingsimulator ble utviklet ved hjelp av empiriske data, copula modellering og nærmeste-nabo prediksjoner. Studiene viste at ignorering av underliggende forutsetninger som estimatorene bygger på, kan ha store effekter for presisjonen i anslagene (f. eks nesten fem ganger overvurdering av middelfeilen). Dessuten kan bruk av simulert sampling bidra til riktig valg av estimatorer (og til og med en riktig samplingstrategi) for anvendelse i reelle applikasjoner.

Simuleringsresultatene viste at bruk av ALS som samplingsverktøy kan være en kostnadseffektiv metode for taksering av store skogområder.

Beregning av skogressursene på lokalt nivå er ofte vanskelig når utvalget av feltobservasjoner er svært lite eller fraværende. Slike situasjoner kan adresseres innen et modellbasert rammeverk, der superpopulasjonsmodellen erstattes av en C-vine copula. Ved å bruke simulert sampling fra copula-funksjonen etterfulgt av nærmeste-nabo prediksjoner, ble det demonstrert en høyere nøyaktighet i forhold til bootstrap resampling, og at den største forbedring består i en signifikant reduksjon av systematisk feil.

For å møte etterspørselen etter detaljert informasjon på trenivå som kreves ved intensiv skogforvaltning, ble en ny metode for tretopp identifisering og ekstrahering av individuelle treattributter utviklet. Ved å bruke estimert stammeantall fra arealbaserte metoder og under milde antagelser om prosessen som genererer den romlige stammefordelingen, viste algoritmen en robust adferd som balanserte utelatelses- og inkluderingsfeilene som oppstod ved bruk i en inhomogen boreal skog. En viktig fordel med metoden er potensialet til integrasjon med arealbasert skogstaksering basert på ALS.

Styrker og ulemper ved metodene, samt ytterligere forbedringer som kan implementeres, er også omtalt.

List of papers:

- I. Ene, L., Næsset, E., Gobakken, T., Gregoire, T.G., Ståhl, G. & Nelson, R. (submitted). Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach.
- II. Ene, L., Næsset, E., Gobakken, T., Gregoire, T.G., Ståhl, G. & Holm, S. (submitted). A simulation approach for accuracy assessment of two-phase post-stratified estimation in large-area LiDAR biomass surveys.
- III. Ene, L., Næsset, E. & Gobakken, T. (submitted). Model-based inference for k-nearest neighbour predictions using a canonical vine copula.
- IV. Ene, L., Næsset, E. & Gobakken, T. (2012). Single tree detection in heterogeneous boreal forests using airborne laser scanning and area based stem number estimates. *International Journal of Remote Sensing*, 33, 5171-5193

Paper IV is reprinted with kindly permission of Taylor & Francis.

SYNOPSIS

1. Introduction

The demands for timely and accurately information regarding forest ecosystems have increased steadily during the past decades, moving from timber supply information towards multi-purpose assessment of forest and non-forest resources at various spatial scales (Kleinn, 2002; Köhl et al., 2006; Corona & Marchetti, 2007; Corona, 2010). National Forest Inventory (NFI) systems can provide information at national, regional and sub-regional level for strategic and tactical forest management and planning, while the operational stand-based forest inventories cover regions at small and intermediate scales. International collaborative efforts like the United Nations Food and Agriculture Organization Global Forest Resource Assessment, United Nations Framework Convention on Climate Change (UNFCCC) and the Kyoto Protocol are aiming towards compiling global scale information regarding the state of world's forests.

At either level, forest inventories rely on sampling techniques for collecting the ground observations necessary for estimation and inference for the underlying area of interest (Köhl et al., 2006). Using cartographical representations of forest resources (or forest mapping- Corona, 2010) makes possible to obtain spatially localized descriptions of the forest cover, which can serve operational, tactical and strategic forest planning goals. The synergy of traditional, ground based forest inventories and modern forest mapping technologies can result into enhanced estimation accuracies of forest resources required at different levels of decision making (Corona, 2010).

1.1 Remote sensing support for forest inventories

The purpose of a forest inventory is to provide quantitative and qualitative descriptions of the forest resources from a defined area required for forest planning and forest policy making

(Köhl et al., 2006; Corona, 2010). In most situations, forest inventories are sample surveys where the parameter estimates are obtained from a small amount of data collected using various sampling strategies (the combination of a sampling design and an estimator; Gregoire & Valentine (2008)). Two main approaches are commonly used for inferring values of target population parameters from the sampled data: probability-based inference and model-based inference. The probability-based inference (called also design-based inference) relies upon the probabilistic sampling design, the attributes of the population elements being considered fixed quantities. In the case of design-based, model-assisted inference, auxiliary information is used for increasing the precision of the estimates, but the validity of the inference is still ensured by the probabilistic design. Under model-based inference, all the elements and parameters of the population are considered random variables generated from a model (a superpopulation model; Cochran (1977)). Hence, the inference relies on the model without explicitly requiring probabilistic sampling, all the elements and parameters of the population are considered random variables generated from a model (a superpopulation model; Cochran (1977)). More detailed discussions regarding the design- and model-based inferences can be found in Gregoire (1998), Kangas (2006), and McRoberts et al. (2010*ab*).

Due to the technological and theoretical developments during the last decades, remote sensing (either active or passive) is nowadays accepted as a cost-effective auxiliary source of information for monitoring and mapping of natural resources (FAO, 2010). Passive remote sensing measures the backscattered sunlight using airborne or satellite platforms, the final product being an imagery product with various spatial and spectral resolutions. Airborne imagery and photo interpretation techniques have been used to support operational forest inventory for decades (Köhl et al., 2006; Corona, 2010). The emergence of easily accessible satellite imagery products (e.g. the Landsat program; Woodcock et al. (2008)) provided the

means for creating ‘wall-to-wall’ maps of forest attributes and for obtaining estimates at different spatial scales.

Active remote sensing like LIght Detection And Ranging (LiDAR) or Radio Detection And Ranging (RADAR) measures the elapsed time between the emission of a signal (a laser pulse for LiDAR and a radio signal in the case of RADAR) and the returned signal reflected by a target object, providing highly accurate determination of the sensor-to-target distance. Although promising results have been reported (Solberg et al., 2010), forestry applications using RADAR are still subject of intensive research. The capability of LiDAR technology for monitoring and mapping of natural resources at various geographical scales has been investigated using terrestrial, airborne and spaceborne systems, and nowadays the airborne laser scanning (ALS) technology is currently migrating from research to commercial applications (Hudak et al., 2009; Wulder et al., 2012). The strength of the ALS technology consists in the ability of laser pulses to penetrate through dense vegetation layers, thus providing accurate three-dimensional characterization of the vegetation structure. For natural resources applications, small-footprint (0.1-2.0m), discrete returns LiDAR systems are the most common, due to their versatility for uses in topographic mapping (Hyypä et al., 2008; McRoberts et al., 2010b).

1.1.1 ALS-aided forest inventories

According to Hyypä (2008), the main approaches for estimating forest attributes using ALS measurements are (1) individual tree detection and (2) using metrics derived from the 3D cloud of laser echo heights.

Individual tree detection and tree crown delineation methods - also called individual tree crown (ITC) methods – use the 3D information provided by the clouds of laser measurements for tree identification and variable extraction. Use of ITC methods does not

guarantee the detection of all the trees within the study area, the results depending heavily on the forest conditions (Hyypä et al., 2008; Vauhkonen et al., 2011). However, quantitative and qualitative attributes can be estimated for the detected trees, making the ITC methods attractive in forest management and timber procurement planning (Hyypä et al., 2008; Maltamo et al., 2009a). For instance, the ITC methods can provide tree lists containing direct estimates of tree heights, crown diameters, stem locations and help indentifying the presence of rare tree species (Hyypä et al., 2008; Maltamo et al., 2007). Also, stand-level estimates can be derived by aggregating the individual trees. Still, spatial information provided by the estimated stem locations may be used together with height and crown diameter information for estimating the diameters of the identified trees (Salas et al., 2010), or to derive competition indices (Maltamo et al., 2007; Pedersen et al., *submitted*). However, simply scaling up individual tree estimates does not guarantee unbiased estimation of stand attributes, especially due to the underestimation of the stem densities as a result of the inability of ITC methods to detect all the small and suppressed trees. A probability-based forest inventory approach using the segmented tree crowns and the associated ground measurements allowing for unbiased estimation of selected stand attributes (stem number, basal area, and the basal area - height product) was developed by Flewelling (2008).

With area-based ALS estimation, various metrics are extracted from the height distributions of the laser echoes recorded on the field plots, and the relationships between ALS metrics and plot-level forest variables the are usually described using regression models (Næsset, 2002; Næsset, 2004). Hence, the prediction of forest variables is possible on all locations where the ALS metrics are available. Area-based, ‘wall-to-wall’ ALS-aided surveys are nowadays commercially used in operational forest inventories in Nordic countries (Næsset, 2004; Maltamo et al., 2007a; Hyypä et al., 2008), targeting forest planning

activities at small to medium geographical scales (e.g. 50-2000 km²; Næsset, 2007; Næsset et al., 2009).

For surveys targeting large regions (e.g. districts, counties or countries), full-coverage with ALS data can become economically prohibitive (Wulder et al., 2012). A possible solution is using ALS as a strip-sampling tool, surveying only a subset of the target area (Næsset et al., 2009). Efforts to develop large-area forest inventories using ALS as a sampling tool have been addressed using both design-based and model-based inferential frameworks. Design-based approaches in the form of double-sampling surveys (Parker & Evans, 2004; Stephen et al., 2012) or using two-stage designs (Andersen et al., 2009; Gregoire et al., 2011; Gobakken et al., *in press*) have been proposed, and a novel theoretical contribution for design-based variance estimation for two-stage inventories was introduced by Gregoire et al. (2011). The properties of several regression estimators have been assessed by Andersen & Breidenbach (2007), and model-based approaches for large-area surveys were described by Andersen et al. (2011) and Asner et al. (2011). In the context of model-based inference, a two-phase variance estimator was developed by Ståhl et al. (2011). Recently, Gobakken et al. (*in press*) describes a comparative study including both the design- and model-based estimation and inference for large-area biomass estimation using a systematic sample of ALS strips and NFI data. A comprehensive overview of LiDAR applications for large-area forest surveys is given by Wulder et al. (2012).

1.1.2 Mapping and estimation of forest resources using satellite imagery

A widely used technique for coupling auxiliary information (e.g., remotely sensed data, DTM elevations, and land-use maps) is the k near-neighbours imputations (Tomppo, 1991; Tomppo et al., 2008, 2009; McRoberts, 2008). With k near-neighbour (k -NN) methods, the unknown attributes in a target data set are predicted as weighted means of the attributes from the k -

nearest reference observations found in a reference data set, the distances between the target and reference observations being calculated in the feature space spanned by the auxiliary variables. Due to their non-parametrical nature, the near-neighbours imputations do not require distributional assumptions about the data. Moreover, they permit simultaneous multivariate predictions of both continuous and categorical variables.

Coupling medium spatial resolution satellite (e.g. Landsat imagery) and k -NN methods are considered to be cost-efficient solutions for supporting NFI programs (Tomppo, 1991; Tomppo et al., 2008), either by mapping the forest attributes of regional and national scales, or for providing local estimates (LeMay & Temesgen, 2005; LeMay et al., 2008; Maltamo et al., 2009; McRoberts, 2008, 2011). Reviews of the nearest neighbor methods with applications in forestry are given by Tomppo et al. (2008), Eskelson et al. (2009), McRoberts et al. (2010a).

The estimation following k -NN imputations have been addressed under the design-based inferential framework for large-area surveys (Baffetta et al., 2009, 2011), the model-based inference being considered most appropriate for small-area estimation (Kim & Tomppo, 2006; McRoberts, 2006; McRoberts et al., 2007, 2011; Magnussen et al., 2009, 2010ab; McRoberts, 2011).

1.2 Research objectives

The main focus of the thesis is assessing the accuracy of forest attributes obtained from surveys performed at different spatial scales. The research considered design- and model-based inferential approaches related to large-area forest inventories covering regions and sub-regions (papers *I-II*), as well as model-based methods applicable for small-area estimation (paper *III*). At the finest scale, the studies were focusing on estimating quantitative single-tree variables (paper *IV*). A common topic of the analyses was usage of auxiliary information

providing either partial or full coverage of the target areas. The assessment methods were based on simulated sampling (papers *I-III*), and a novel single-tree detection algorithm combining area-based and single-tree methods was introduced in paper *IV*. The succession of the research papers included in the thesis is the following:

- *Paper I - "Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach"* - introduces the set up of the simulation approach for assessing estimators for design- and model-based sampling strategies, and illustrates the use of the method through a case study resembling a real sample survey.
- *Paper II - "A simulation approach for accuracy assessment of two-phase post-stratified estimation in large-area LiDAR biomass surveys"* - takes the analysis from paper *I* a step further, focusing on post-stratified estimation and considering also the variance estimation under two-phase systematic sampling. Moreover, a cost-efficiency analysis is performed to assess the gains of using auxiliary ALS data in the estimation.
- *Paper III - "Model-based inference for k -nearest neighbour predictions using a canonical vine copula"* shifts the focus from large-area inventories to small-domain estimation. A model-based approach using copulas is introduced and assessed against bootstrap resampling.
- *Paper IV - "Single tree detection in heterogeneous boreal forests using airborne laser scanning and area based stem number estimates"* - describes a single-tree detection algorithm to be used for enhancing area-based, ALS-aided forest inventories with single-tree level information.

The specific objectives of the thesis can be summarized as follows:

1. Aspects related to ALS-aided surveys covering large geographical regions and sub-regions:

- 1.1. to develop a framework for assessing the properties of the estimators used in surveys involving complex designs, and
- 1.2. to demonstrate the functionality of the platform by validating existing estimators, and assessing the estimation and cost-efficiency of various sampling strategies;
2. To introduce and to assess a true model-based inferential approach based on simulated sampling from a superpopulation model, which can be suitable for small-area estimation;
3. To develop a method for estimating attributes at the scale of individual trees, using empirical datasets as the large-area ALS-aided inventories.

A general description of the datasets used in the studies is provided in section 2, and the main methods employed by each paper are detailed in section 3. Section 4 presents and discusses the main results of the analyses. Finally, conclusions and future research opportunities related to the present work are given in section 5.

2. Materials

The material used for in the studies consists in field plots and auxiliary information consisting in remotely sensed data, a digital terrain model (DTM) and land-use maps. *Paper I-III* are based on field data, ALS and satellite datasets acquired in Hedmark County (HC), while the material for *paper 4* was collected independently of the HC datasets in Aurskog-Høland municipality (AH) and consists in plot measurements and ALS data. Both study areas were located in south eastern Norway (Figure 1), within the boreal forest vegetation zone (Olson et al., 2001). The material used in papers *I, II* and *III* was collected to support area-based forest inventory methods, while the material used in paper *IV* targeted the development of ITC.

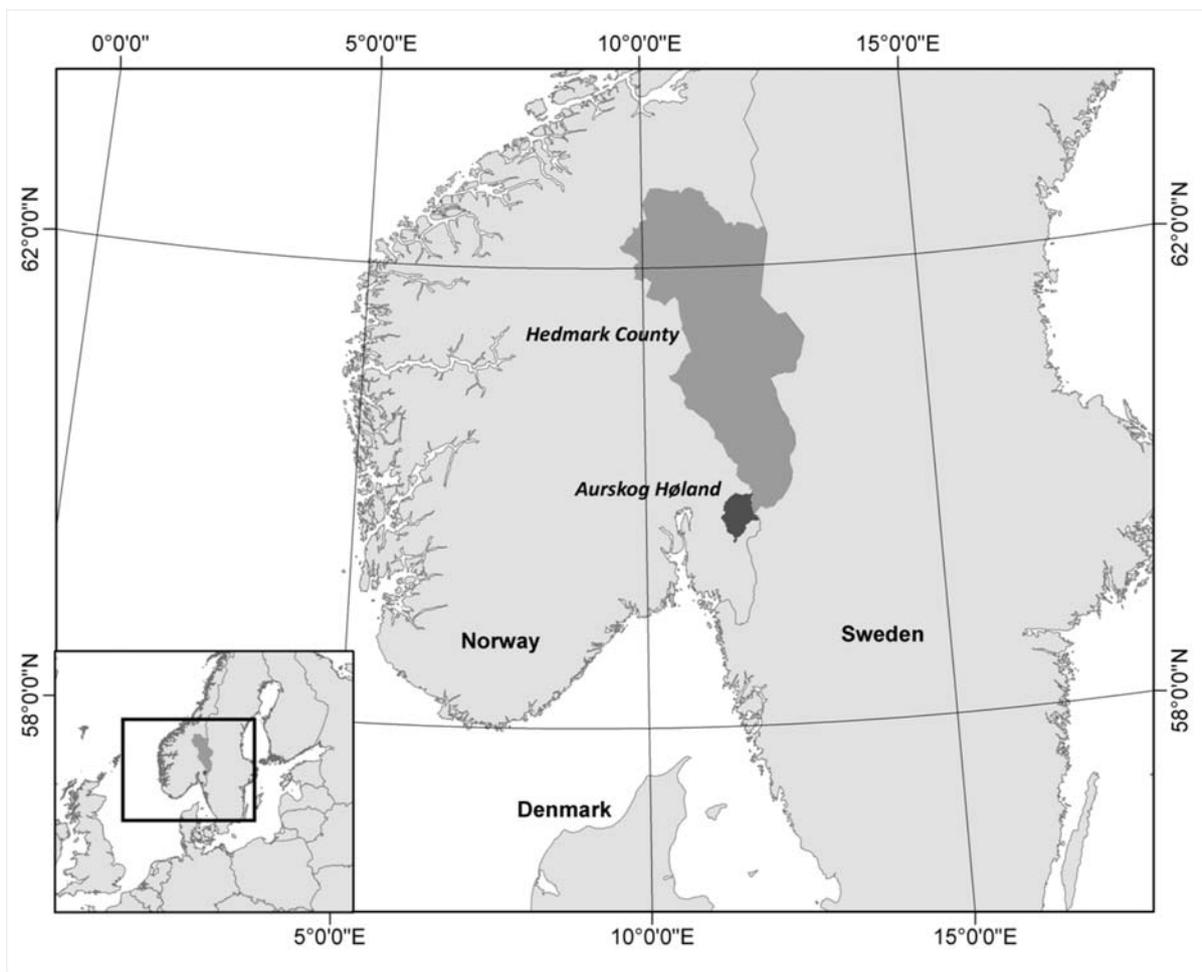


Figure 1 Geographical locations of the study area- Hedmark County and Aurskog Høland

2.1 Study areas

The AH study area covers 890 km² and the forested part consists of coniferous and mixed forest stands. Considering the local scale of the measurements, the forest structure is rather heterogeneous and the terrain steepness varies considerably among the plots.

The total area of HC is 27340 km², of which 53.7 % is covered by forests (Anon., 2004). The HC forests are spread over a broad range of growth conditions, being dominated by Norway spruce (*Picea abies* (L.) Karst.), and Scots pine (*Pinus sylvestris* L.) and with birch as the main deciduous species. The topography across HC displays large altitudinal variations, with a decreasing altitudinal trend from the mountain areas in the north towards the south.

2.1 Field data

In AH, the field measurements were collected from 40 circular plots during fall 2007 and winter of 2008. Four plots located in dense young forests were established using a radius of 12.62 m (500 m² area), the remaining 36 plots having a radius of 17.84 m (1000 m² area). The field measurements in HC were provided by the Norwegian National Forest Inventory (Tomter et al., 2010) and were collected between the years 2005 and 2007 from circular plots of 8.92 m radius (250 m² area).

On each location, the (x, y) plot centre coordinates were determined using differential post-processing of dual-frequency Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) measurements acquired by Topcon dual-frequency receivers, with reported average plot centre positioning error <0.12 m.

On each plot, tree species and stem diameter at breast height (*dbh*) were recorded for all trees with $dbh \geq 5$ cm, and height measurements were acquired from sample trees. In addition crown measurements were performed for sample trees on the AH plots. Different strategies for sample tree selection were followed on the AH and NFI plots, as the data served two different purposes. On all plots, the tree stem locations were also mapped.

In addition to the individual tree variables, area-based variables like plot basal area, mean height, Lorey's height and mean crown diameter were derived for the AH plots. For the NFI plots across HC, the above ground biomass (AGB) of living trees was obtained by summing up individual biomass estimates obtained using tree-species specific allometric equations (Marklund, 1988) for all living trees with $h \geq 1.3$ m. Generally, the AGB estimates showed a trend of increasing values from north to south, following the altitudinal trend present on HC.

2.2 Auxiliary data

2.2.1 Laser scanning data

Small-footprint airborne laser scanning (ALS) measurements were acquired in summer 2006. For both areas, the acquisition of ALS data was performed on east-west oriented parallel corridors containing the field plots. An overview of the ALS data is presented in Table 1.

Table 1 Airborne laser scanning data

Study area	Hedmark	Aurskog-Høland
<i>Paper</i>	<i>I,II,III</i>	<i>IV</i>
<i>Acquisition settings</i>		
acquisition date		2006
platform	PA31 Piper Navajo fixed wing	
canopy conditions		leaf-on
flying altitude (m)		800
flying speed (m s ⁻¹)		75
<i>Sensor settings</i>		
sensor		Optech ALTM 3100
no. of echoes		1-4
echo category for analysis		first and single
pulse repetition frequency (kHz)		100
scan frequency (Hz)	55	70
echo density (echoes m ⁻²)	2.8	7.4
footprint (m)	0.21	0.25

The ALS measurements were used to obtain canopy metrics for AGB prediction in papers *I* and *II*, and for individual tree identification and estimation of tree-level variables in paper *IV*.

2.1.2 Satellite imagery data

The satellite imagery consisting of three nearly cloud-free Landsat 5 TM images (Table 2) was required for the analysis in papers *I*, *II* and *III*. The Landsat images were delivered with a georeferencing error of $\leq 15\text{m}$ for 95% of the pixels. The imagery data was converted to at-satellite reflectance and atmospherically corrected. Only six of the seven TM bands (bands 1 to 5 and band 7) were used in the analysis.

In order to obtain full coverage with imagery data of the HC area required by the analysis in papers *I* and *II*, an image mosaic with 15.81 m pixel resolution was created using nearest neighbour resampling such that the pixel area (approximately 250 m^2) matched the area of NFI plots in HC.

Table 2 Satellite imagery data

Satellite Image	Date	Path/row	Sun elevation (deg)	Sun azimuth (deg)	Spatial resolution (m)
1	03-June-2007	197/16	49.64	162.17	25
2	03-June-2007	197/17	50.00	162.00	30
3	10-June-2007	198/16	50.07	161.93	25

2.2.3 Cartographical data

Cartographical data was used for performing the analysis in papers *I*, *II* and *III*. A 25 m spatial resolution DTM produced by the Norwegian Mapping Authority from the official topographic map series and *AR5* land use maps (Bjørndal & Bjørkelo, 2006) were used to obtain a forest mask for HC at the same spatial resolution as the Landsat mosaic. In addition, the DTM elevations recorded at locations corresponding to the NFI plots in HC were used for the analysis in paper *III*.

3. Methods

3.1 Assessing sampling strategies for large-area forest inventories

The assessment of sampling strategies for large area forest inventories is addressed in papers *I* and *II*. The studies consider ALS-aided inventories using sampling strategies involving two-phase designs with generalized regression estimators, which are also compared with the ground-based surveys traditionally employed by the NFI program. The core of the studies consists in using a sampling simulation approach, where simulated sampling from an artificial population with known parameters is used for deriving the sampling distribution of the estimators.

Both papers addressed a general case resembling the large-area biomass survey performed in HC (Gregoire et al., 2011; Ståhl et al., 2011; Gobakken et al., *in press*), where the NFI field data was supplemented with a systematic sample of ALS measurements acquired in corridors overlapping every second NFI grid line (approximately 8% of the HC land area). *Paper I* introduces the simulation platform and demonstrates the utility of the simulated sampling for unstratified AGB estimation under design- and model-based inference. *Paper II* takes the analysis a step further, focusing on post-stratified estimation and considering also another class of variance estimators (successive difference estimators) for two-phase systematic sampling. Moreover, a cost-efficiency analysis is performed to assess the gains of using the ALS-aided estimation.

3.1.1 Creating the artificial population

In order to obtain valid simulation results, the artificial population has to resemble the original population targeted by the survey. Considering the scale of the simulated survey, the empirical dataset (ALS metrics, spectral information corresponding to the satellite bands and

AGB data) was rather scarce. Instead of acquiring more observations (a hopeless endeavour, however), a Gaussian copula (Nelsen, 2009) was used for modelling the joint multivariate distribution between the empirical variables. Hence, a much large sample of ALS metrics, spectral information and AGB data could be generated from the fitted copula function. Using the similarities between the spectral information in the Landsat TM mosaic and copula sample, the ALS metrics and the AGB were imputed on each of the mosaic pixels using the nearest neighbor technique. The result was a map of the HC forest area, where the population elements were represented by the pixels. Thus, the parameters of the artificial population (the total and the average AGB) could be calculated without error. The analyses also indicated that, due to the nearest neighbor imputations, the main trends in the empirical data were reasonably well preserved in the artificial population.

3.1.2 Sampling strategy assessment

Using the simulation platform, various designs like simple random sampling without replacement (SRSwoR) and systematic sampling (SYS) were implemented for both the ALS and the ground sample. The per-hectare AGB was estimated using design-based estimators like Horvitz-Thompson and generalized regression estimators, and by using the synthetic regression estimator under the model-based inferential framework.

Using simulated sampling, the bias and the true standard errors of the AGB estimators were obtained under SRSwoR and SYS. Moreover, the estimated (analytical) standard errors (Gregoire et al., 2011; Ståhl et al., 2011; Gobakken et al., *in press*) could be compared with the true standard errors obtained from the randomization distributions of the estimators. Hence, it was possible to obtain the estimation accuracies in terms of root mean squared errors, and to compare the relative efficiencies of SRSwoR and SYS. Moreover, the validity

of the inference was expressed by building 95% confidence intervals around the AGB estimates.

The assessment of sampling strategies included cost-efficiency analyses, where the relationships between the precision and the costs of simulated surveys have been estimated using various sampling intensities. The sampling intensities were defined using different settings for the design parameters like the spacing between the parallel ALS corridors and the distances between the field plots.

3.2 Model-based estimation for small domains

The analyses described in paper *III* focus on non-parametrical AGB estimation using nearest-neighbor imputations, under the model-based inferential framework. The model-based approach proposed in the paper assumes that the superpopulation model can be represented as a joint multivariate distribution of the auxiliary data and the ground-based AGB estimates. The multivariate dependencies among the variables were modelled using the hierarchical pair-copula decomposition approach, which allows building high-dimensional copula models from bivariate copula models and conditional marginal distributions, the copula-pairs being identified by the means of graphical representation called vines (Bedford & Cooke, 2001, 2002; Aas et al., 2009). The superpopulation model was constructed using a canonical vine-copula (called also C-vine copula, Bedford & Cooke, 2001; Aas et al., 2009), and it included auxiliary information from both reference and target observations.

The study was conducted using a simulation approach, where the empirical dataset containing auxiliary information (satellite imagery and elevation data) and ground-based AGB estimates was randomly split into reference and target observations, the C-vine copula being fitted each time new reference and target observations were created. Considering the C-vine copula as the superpopulation model, new samples of the same size as the reference

dataset were iteratively generated from the copula function and used with nearest neighbor imputations for predicting the average AGB of the target dataset. The number of neighbours (k) used for imputations was determined using a bias-minimizing criterion (McRoberts et al., 2011; McRoberts, 2011), and the results were compared to the case of using $k=1$. Hence, the sampling distribution of the nearest neighbor estimators could be constructed for determining the biases and the standard errors produced for each data split.

Finally, the accuracy and the validity of the inference for the nearest neighbor predictions using the copula-based approach was assessed against bootstrap resampling (McRoberts et al., 2011; McRoberts, 2011).

3.3 Estimation of tree-level attributes

A novel method for identification and delineation of individual trees using ALS measurements is described in *paper IV*. The method was developed assuming that the tree stems positions at plot level is the result of a spatial random process. Having the hypothesis of a random process confirmed by the \hat{l} estimator (Ripley 1979) on most of the plots in AH study area, the expected nearest neighbor distance (Clark & Evans, 1954) between trees was estimated using stem number predictions and the known plot areas. For each plot, the stem number was predicted using a regression model having selected ALS metrics as covariates.

The identification and delineation of individual tree crowns was performed using canopy height models (CHM) obtained for each plot by rasterizing the laser echo height measurements. The estimated nearest neighbor distances among trees obtained on each plot were used for (1) building up smoothing kernels for the CHMs, and (2) obtaining appropriate spatial resolutions for creating the CHMs. Image processing techniques could be applied individually on each smoothed CHM for detecting tree tops, extracting individual tree attributes (e.g. tree height and crown width) from ALS height measurements, and estimating

plot-level forest characteristics (e.g. mean height) by aggregating the individual tree estimates.

The performance of the new method was assessed against an established algorithm (Hyypä & Inkinen, 1999; Hyypä et al., 2001).

4. Results and discussion

4.1 Assessing sampling strategies for large-area forest inventories

The results obtained from the simulation studies (e.g. the estimated AGB per hectare and its standard error) cannot replace the estimates from a real survey. On the other hand, the simulation exercises demonstrated the use of simulated sampling for validating estimators for complex designs, and how the relative efficiency of various estimators can be assessed. Moreover, the simulation approach makes possible to assess the design effect when using complex multi-phase sampling designs.

The studies describe in detail a novel approach for constructing a sampling simulation platform, where empirical observations acquired from different sources are combined into a multivariate distribution using copulas and then generalized across a geographical area by the means of nearest neighbor imputations.

The case studies presented in papers *I* and *II* demonstrated that using systematic sampling designs can dramatically inflate the estimated (analytical) precision of the ALS-aided surveys compared to the empirical (true) precision. For instance, the simulation results suggested that the estimated standard errors of the per-hectare AGB under two-phase SYS can be approximately 5 times higher compared to the true errors. Although the variance inflation was expected from the general sampling theory, the magnitude of the design effect cannot be derived analytically when complex designs are involved (Gregoire et al., 2011).

Moreover, it can be expected that the variance inflation under the real survey to be even higher, given that the artificial population used during the simulations is expected to have less variability compared to the real target population (e.g., the forested areas in HC).

When the theoretical assumptions used for developing the estimators are fulfilled (e.g. SRSwoR, independence and invariance for subsampling, Särndal et al., 1992, p. 134-135), the variance estimators performed generally well, the differences between true and analytical values varying between 1 and 10%. The simulation exercise described in *paper II* indicated also the ability of the MB estimator to account for post-stratification and to provide very precise estimates. However, the simulation results indicated that the best standard error estimates under SYS were provided by the successive difference estimator.

According to the cost-efficiency analysis performed in *paper II*, ALS-aided surveys are in general preferable compared to ground-based surveys, which means that they have a better precision at comparable costs. However, due to the large variance inflation under SYS, this advantage cannot be accounted for when using most of the analytical estimators.

4.2 Small-domain estimation

Overall, the copula-based approach provided nearest neighbor estimates having approximately 15-20% higher accuracy (in terms of root mean squared errors) compared to bootstrap resampling. The accuracy improvements of the copula method were mainly due to a significant bias reduction (30-34%), while the standard errors were higher (6-39%) compared to bootstrap. Furthermore, using the copula method allowed constructing confidence intervals with coverages much closer to their nominal values.

4.3 Estimation of tree-level attributes

Using the method introduced in *paper IV*, the creation of CHMs and the estimation of individual tree attributes could be tailored at plot-level. The individual tree delineation results outperformed the benchmark method when applied appropriately, and the estimation accuracies obtained at tree- and plot level were in line with other studies performed in Scandinavia (approximately 50% detection rate and 5% commission errors). Furthermore, the robustness of the method has been tested in an exhaustive comparison study (Vauhkonen et al., 2011), and it was found to be the most stable algorithm (in terms of balancing the omission and commission errors) under various data sets and forest types analyzed.

Since the method is tuned using area-based stem number estimates, two immediate advantages can be mentioned. First, the method can be seemingly integrated into operational ALS-aided forest management inventories for extracting individual tree data. For instance, with two-stage ALS-aided forest inventory (Næsset, 2002) the prediction of forest variables (including stem numbers) is performed in a grid covering the entire study area. Hence, the CHM creation and the extraction of individual tree variables can be performed at grid cell level within each forest stand. However, for practical purposes, it can be reasonable to consider implementing the method at a larger spatial scale (e.g. directly at the stand level). Secondly, using the appropriate spatial resolution for creating the CHM avoids the need for accurate georeferencing of tree stem positions. However, stem coordinates might still be necessary for sample trees subjectively selected across the study area in order to perform e.g. tree-species classification (Dalponte et al., 2009).

5. Conclusions and future research

Overall, the results reported in the thesis responded to the main research objectives.

Papers I and II succeeded to integrate the joint efforts for developing a theoretical framework for uncertainty estimation of large-area, ALS-aided forest inventories. The newly developed simulation platform demonstrates that erroneous inference is prone to occur when the basic underlying assumptions required by the estimators are ignored. The simulation approach discussed here offers a solution to avoid making wrong inference by allowing the exploration of different sampling strategies, and thus making possible the comparison of various estimators. It was demonstrated that sophisticated estimators are not necessary performing better than simple ones when complex surveys are involved. Moreover, the sampling simulator allows assessing the risks of using biased estimators (e.g., successive variance estimators).

A drawback of the method consists in the way the artificial population was created. Using nearest neighbor imputations for mapping the reference observations does not guarantee the preservation of the short-range autocorrelation, making difficult for instance to perform simulations using cluster sampling or analyses considering various plot size and shapes.

Despite these problems, the functionality of the sampling simulator (*papers I and II*) should be expanded to become a decision support tool at the design-stage of an inventory (Tomppo, 2009; Tomppo et al., 2010). For doing so, the artificial population has to be constructed prior to sampling the empirical data, thus the possibility of calibrating other datasets acquired from similar forest conditions should be investigated. This would require, for instance, a data base of empirical data sets of field- and ALS measurements from which multivariate observations for constructing the artificial population can be selected using

various similarity criteria. Alternatively, the artificial population can be built at individual tree level from empirical crown segments produced by the algorithm introduced in *paper IV* and using a forest stand generator for creating tree lists and mapping the tree position. The method can be further improved by considering vine-copulas (as used in *paper III*) instead of using multivariate elliptical copulas for modeling the joint dependencies between the covariates. An accurate representation of the target population would allow for assessing stratification schemes following other criteria than administrative regions (e.g. by forest type or land-use category). Another issue which has not been considered in *papers I and II*, but which definitely deserves consideration, is using the simulating approach for change detection and estimation issues.

The model-based inference using the copula-based approach has been proven to be a viable alternative to bootstrap resampling, although more computationally intensive. This result is not surprising considering that sampling from the copula function provides reference datasets containing new observations, thus filling the gaps in the feature space and reducing the edge-biases characteristic for near-neighbor imputations. However, the copula-based approach deserves further research mainly for (1) prediction of multivariate responses, including categorical variables, (2) assessing bias-reduction methods for k-NN estimators (Magnussen et al., 2010), (3) comparison against parametric variance estimators, and (4) combining auxiliary data from different platforms (e.g. imagery and ALS data) for small-area estimation in the context of biomass and carbon sequestration.

The individual tree study presented in *paper IV* has successfully answered the 3rd research objective by integrated area-based and individual tree methods for estimating forest attributes. At this stage, the individual tree method is not mature yet for being considered as a individual tree inventory approach. Further development is required, such as sensor integration combining ALS and imagery data to improve the crown delineation and for tree

species classification (Hyyppä, 2008). Thus, the algorithm presented in paper *IV* can become useful in combination with area-based estimates to produce tree lists at stand level (Lindberg et al., 2010), which may be used for growth and yield modelling.

With minimal improvements, all the methods described in this thesis have the potential to be integrated in a GIS-based decision-support system for forest management and planning.

References

- Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 442, 182-198.
- Ackermann, F. (1999). Airborne laser scanning - present status and future expectations. *Journal of Photogrammetry and Remote Sensing* 54, 64-67.
- Andersen, H.-E. & Breidenbach, J. (2007). Statistical properties of mean stand biomass estimators in a LiDAR-based double sampling forest survey design. *International Archives of Photogrammetry and Remote Sensing*, Volume XXXVI, Part 3/W52. Espoo, Finland, September 12-14, 2007.
- Andersen, H.-E., Barret, T., Winterberger, K., Strunk, J. & Temesgen, H. (2009). Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using airborne lidar and field plot data in a model-assisted sampling design. (McRoberts, R. ed.), *IUFRO Symposium*, May 2009, Quebec City, Quebec, Canada, 5 pgs.
- Andersen, H.-E., Strunk, J. & Temesgen, H. (2011). Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the Upper Tanana valley of Interior Alaska. *Western Journal of Applied Forestry*, 26, 157-164.
- Anon. (2004). Norwegian Forest and Landscape Institute (*in Norwegian*). URL: http://www.skogoglandskap.no/artikler/2007/Landsskogdata_enkel_tabell (*last accessed: 07-Feb-2012*).
- Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M. et al. (2011). A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia*, 168, 1147-1160.

- Baffetta, F., Fattorini, L., Francheschi, S. & Corona, P. (2009). Design-based approach to k -nearest neighbours techniques for coupling field and remotely sensed data in forest surveys. *Remote Sensing of Environment*, 113, 463-475.
- Baffetta, F., Corona, P., Fattorini, L. (2011). Design-based diagnostics for k -NN estimators of forest resources. *Canadian Journal of Forest Research*, 40, 59-72.
- Bedford, T. & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modelled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.
- Bedford, T. & Cooke, R. M. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics* 30, 1031–1068.
- Bjørndal, I. & Bjørkelo, K., 2006, AR5 klassifikasjonssystem. Klassifikasjon av arealressurser. [AR5 classification system. Classification of areal resources]. Ås, Norway: Norwegian Forest and Landscape Institute. (in Norwegian with English summary).
- Clark, P. & Evans, E. (1954). Distance to the nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35, 445-453.
- Cochran, W.G., 1977, *Sampling Techniques*. New York: Wiley, 3rd Ed., 428 s.
- Corona, P. (2010). Integration of forest mapping and inventory to support forest management. *iForest*, 3, 59-64.
- Corona, P. & Marchetti, M. (2007). Outlining multi-purpose forest inventories to assess the ecosystem approach in forestry. *Plant Biosystems*, 141, 243-251.
- Dalponte, M., Bruzzone, L., Vescovo, L & Gianelle, D. (2009). The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, 11, 2345-2355.

- Eid, T., Gobakken, T & Næsset, E. (2004). Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. *Scandinavian Journal of Forest Research*, 19, 512-523.
- Ene, L., Næsset, E. & Gobakken, T. 2007, Simulating sampling efficiency in airborne laser scanning based forest inventory. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVI, Part 3/W52, September 12-14, Espoo, Finland.
- Ene, L., Næsset, E & Gobakken, T. (2012). Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates, *International Journal of remote Sensing*, 33, 5171-5193.
- Falkowski, M.J., Wulder, M.A., White, J.C. & Gillis, M.D., (2009). Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Progress in Physical Geography*, 33, 403-423.
- FAO (2010). Global Forest Resources Assessment 2010. Main Report. *Food and Agriculture Organization of the United Nations*, Rome, 2010. URL: <http://www.fao.org/docrep/013/i1757e/i1757e.pdf> (last accessed 15-Feb-2012).
- Flewelling, J. (2008). Probability models for individually segmented tree crown images in a sampling context. *In Proceedings of the SilviLaser 2008 Conference*. R.A., Rosette, J. and Suárez, J. (Eds). September 2008, Edinburgh, UK.
- Gobakken, T., Næsset, E. (2004). Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research*, 19, 529–542.
- Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., et al. (2011). Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. (*in press*).

- Gregoire, T.G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28, 1429-1447.
- Gregoire, T.G. & Valentine, H., 2008, Sampling strategies for natural resources and the environment. (Boca Raton, Fla.: Chapman & Hall/CRC), 474 p.
- Gregoire, T.G., Ståhl, G., Næsset, E., Gobakken, T. & Nelson, R., 2011, Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41, pp. 83-95.
- Hudak, A.T., Evans, J.S. & Smith, A.M.S. (2009). LiDAR utility for natural resource managers. *Remote Sensing*, 1, 934-951.
- Hyypä, J. & Inkinen, M. (1999). Detecting and estimating attributes for single trees using laser scanner. *The photogrammetric Journal of Finland*, 16, 27-42.
- Hyypä, J., Kelle, O., Lehikoinen, M. & Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 969-975.
- Hyypä, J., Hyypä, H., Leckie, D., Gougeon, F., Yu, X. & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29, 1339-1366.
- Kim, H.-J. & Tomppo, E. (2006). Model-based prediction error uncertainty estimation for k -nn method. *Remote Sensing of Environment*, 104, 257-263.
- Kleinn, C. (2002). New technologies and methodologies for national forest inventories. *Unasylva*, 53, 10-15.
- Lindberg, E., Holmgren, J., Olofsson, K., Wallerman, J. & Olsson, H. (2010). Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. *International Journal of Remote Sensing*, 31, 1175-1192.

- Magnussen, S., McRoberts, R.E. & Tomppo, E. (2009). Model-based mean-squared error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. *Remote Sensing of Environment*, 113, 473-488.
- Magnussen, S., Tomppo, E. & McRoberts, R.E. (2010a). A model-assisted k-nearest neighbor approach to remove extrapolation bias. *Scandinavian Journal of Forest Research*, 25, 174-184.
- Magnussen, S., McRoberts, R.E. & Tomppo, E. (2010b). A resampling variance estimator for the k nearest neighbours technique. *Canadian Journal of Forest Research*, 40, 648-658.
- Maltamo, M., Packalén, P., Peuhkurinen, J., Suvanto, A., Pesonen, A., & Hyyppä, J. (2007a). Experiences and possibilities of ALS based forest inventory in Finland. In P. Rönholm, H. Hyyppä, & J. Hyyppä (Eds.), *Proceedings of ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*. September 12–14, 2007, Espoo, Finland (pp. 270–279). IAPRS XXXVI Part 3 / W52. URL: http://www.commission3.isprs.org/laser07/final_papers/Maltamo_2007_keynote.pdf (last accessed 21-Feb-2012).
- Maltamo, M., Suvanto, A. & Packalén, P. (2007b). Comparison of basal area and stem frequency diameter distribution modeling using airborne laser scanner data and calibration estimation. *Forest Ecology and Management*, 247, 26–34.
- Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P. & Tokola, T. (2009a). Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. *Silva Fennica*, 43, 507-521.
- Maltamo, M., Næsset, E., Bollandsås, O.M., Gobakken, T. & Packalén, P. (2009b). Non-parametric prediction of diameter distributions using airborne laser scanner data. *Scandinavian Journal of Forest Research*, 24, 541-553.

- Marklund, L.G. (1988). Biomass functions for pine, spruce and birch in Sweden. Umeå: Swedish University of Agricultural Sciences, Department of Forest Survey (in Swedish).
- McRoberts, R.E. (2006). A model-based approach to estimating forest area. *Remote Sensing of Environment*, 103, 56–66
- McRoberts, R.E. & Tomppo, E.O. (2007). Remote sensing support for national forest inventories. *Remote Sensing of Environment*, 110, 412-419.
- McRoberts, E.R., Tomppo, E.O., Finley, A.O. & Heikkinen, J.H. (2007). Estimating areal means and variances of forest attributes using the k -nearest neighbors technique and satellite imagery. *Remote Sensing of Environment*, 111, 466-480.
- McRoberts, R.E., Cohen, W.B., Næsset, E., Stehman, S.V. & Tomppo, E.O. (2010a). Using remotely sensed data to construct and assess forest attribute maps and related spatial products. . *Scandinavian Journal of Forest Research*, 25, 340-367.
- McRoberts, R.E., Tomppo, E.O. & Næsset, E. (2010b). Advances and emerging issues in national forest inventories. *Scandinavian Journal of Forest Research*, 25, 368-381.
- McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbor techniques. *Forest Ecology and Management* (2011),
doi:10.1016/j.foreco.2011.06.039.
- McRoberts, R.E., Magnussen, S., Tomppo E.O. & Chirici, G. (2011). Parametric, bootstrap, and jackknife variance estimators for the k -nearest neighbours technique with illustrations using forest inventory and satellite image data. *Remote Sensing of Environment*, 115, 3165-3174.
- McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbor techniques. *Forest Ecology and Management* (2011),
doi:10.1016/j.foreco.2011.06.039.

- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning lidar using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88-99.
- Næsset, E. (2004). Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19, 164-179.
- Næsset, E. (2007). Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest research*, 22, 433-442.
- Næsset, E., Gobakken, T., Holmgren, J., Hyypä, J., Hyypä H., Maltamo, M., Nilsson M., Olsson, H., Persson, Å. & Söderman, U. (2004). Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, 19, 482-499.
- Næsset, E., Gobakken, T. & Nelson, R. (2009). Sampling and mapping forest volume and biomass using airborne LiDARs. *Proceedings of the eight Annual Forest Inventory and Analysis Symposium*, October 16-19, Monterey, CA, USA. General Technical Report WO-79, United States Department of Agriculture, Forest Service, Washington DC, pp. 297-301.
- Nelsen, R.B. (2009). An introduction to copulas, (2nd ed.). New York: Springer.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N, Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Colby, J.L., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. & Kassem, K.R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *Bioscience*, 51, 933-938.
- Parker, R.C & Evans, D.L. (2004). An application of LiDAR in a double-sample forest inventory. *Western Journal of Applied Forestry*, 19, 95-101.

- Pedersen, R.Ø., Bollandsås, O.M., Næsset, E. & Gobakken, T. Eliminating the plot edge bias using individual tree competition indices based on airborne laser scanning (submitted).
- Popescu, S. C., Wynne, R. H. & R. F. Nelson (2002). Estimating plot-level tree heights with LiDAR: local filtering with a canopy height based variable window size. *Computer and Electronics in Agriculture*, 37, 71-95.
- Ripley, B.D. (1979). Test of randomness for spatial point patterns. *Journal of the Royal Statistical Society* 41, p. 368-374.
- Salas, C., Ene, L., Gregoire, T.G., Næsset, E. & Gobakken, T. (2010). Modeling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models, *Remote Sensing of Environment*, 114, 1277-1285.
- 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, 2353-2360.
- Särndal, C.E., Swensson, B. & Wretman, J., 1992, *Model assisted survey sampling*. Springer, New York, 694 p.
- Stephens, P.R., Kimberly M.O, Beets, P.N., Paul, T.S.H., Searles, N., Bell, A., Brack, C., & Broadley, J. (2012). Airborne scanning LiDAR in a double sampling forest carbon inventory. *Remote Sensing of Environment*, 117, 348-357.
- Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., & Nelson, R. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest research*, 41, 96-107.
- Tomppo, E., Haakana, M., Katila, M. & Peräsaari, J. (2008). *Multi-Source National Forest Inventory Methods and Applications*, Springer Series Managing Forest Ecosystems, 374 p

- Tomppo, E. (2009). The Finnish National Forest Inventory. In McRoberts, R.E., Reams, G. A., Van Deusen, P. A. & McWilliams, W. H. (Eds.), *Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium*, Monterey, California, October 16-19, 2006 (USDA Forest Service Gen. Technical Report WO-79, pp. 39-46).
- Tomppo, E., Malimbwi, R., Katila, M., Chamuya, R., Mäkisara, K., Peräsaari, J., et al. (2010). Sampling design for NAFORMA-*Forest Monitoring and Assessment of Tanzania*. Presentation. Workshop on NAFORMA Sampling Design, November 30-December 1, 2009, Dar es Salaam, Tanzania.
- Tomter, S.M., Hysten, G. & Nilsen J-E. (2010). National Forest Inventory Reports: Norway. In Tomppo, E., M., Gschwantner, T., Lawrence, M. and McRoberts, R.E. (Eds.), *National Forest Inventories- Pathways for Common Reporting* (p. 411-424). Dordrecht: Springer Science+Business Media B.V.
- van Leeuwen, M., Hilker, T., Coops, N.N., Frazer, G., Wulder, M. , Newham, G.J. et al. (2011). Assessment of standing wood and fiber quality using ground and airborne laser scanning: a review. *Forest Ecology and Management*, 261, 1467-1478.
- Vauhkonen, J., Ene, L., Gupta, S., Heinzl, J., Holmgren, J., Pitkänen, J., et al. (2011a). Comparative testing of single-tree detection algorithms under different types of forests. *Forestry*, doi:10.1093/forestry/cpr051.
- Vauhkonen, J., Korpela, I., Maltamo, M & Tokola, T. (2011b). Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, 114, 1263–1276.
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W. et al. (2008). Free access to Landsat imagery. *Science*, 320, 1011.

Wulder, M.A., J.C. White, R.F. Nelson, E. Næsset, H.O. Ørka, N.C. Coops, et al. (*in press*).

Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*. (Submitted on Oct 3, 2011; Accepted Feb 9, 2012).

Paper I

Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach

Prepared by:

Liviu Theodor Ene^{1*}, Erik Næsset¹, Terje Gobakken¹, Timothy G. Gregoire², Göran Ståhl³, and Ross Nelson⁴

Address:

1. Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway, liviu.ene@umb.no;

Terje.Gobakken@umb.no; Erik.Naesset@umb.no.

2. School of Forestry and Environmental Studies, 360 Prospect St., Yale University, New Haven CT 06511-2189 USA, Timothy.Gregoire@yale.edu.

3. Dept. of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden, Goran.Stahl@slu.se.

4. 2.614.4/Biospheric Science Branch, NASA-Goddard Space Flight Center, Greenbelt MD 20771 USA, Ross.F.Nelson@nasa.gov

*Corresponding author

Liviu Theodor Ene

Email: liviu.ene@umb.no

Phone: +47-64965766

Fax: +47-64968890

Abstract

To meet the increasing need for reliable and timely timber resources and carbon stock estimates at intermediate and local decision levels, a sampling approach using airborne laser scanning (ALS) as a strip sampling tool has been proposed as a supplement to the conventional field-based National Forest Inventory system. This idea led to a large-scale biomass survey project undertaken in Hedmark County, Norway, an area encompassing 27390km². The field biomass estimates were provided by the Norwegian NFI, and the ALS measurements were acquired in parallel strips using a systematic (*SYS*) design. The ALS-based biomass estimation was performed using regression estimators under design-based and the model-based inferential frameworks. Assessing the validity of inference is not straightforward when complex designs are involved, a possible approach being using a sampling simulator where an artificial population represents the ‘ground truth’ and the properties of the estimators are investigated via simulated sampling. To create the artificial population, a large multivariate dataset containing NFI field observations and ALS metrics was generated using a copula function fitted to the empirical observations, and then it was generalized over the study area using satellite imagery and nearest-neighbor imputations. The properties of several design-based model-assisted and model-based variance estimators were investigated using simulated sampling and the accuracy of ALS-based and ground-based estimates under simple random sampling without replacement (*SRSwoR*) and *SYS* designs were compared. The simulation results indicated that the ALS-based survey produced valid inference under design-based and model-based frameworks. The variance estimators performed well under two-phase *SRSwoR*, but the real standard errors were overestimated approximately 4.7 times under two-phase *SYS*. Compared to the pure ground-based inventories, the estimated standard errors of the ALS-based estimates were approximately 1.8 times larger, while the real accuracy improved with 59%.

1. Introduction

As one of the countries ratifying the Kyoto Protocol to the United Nations Framework Convention on Climate Change (UNFCCC 2008), Norway must quantify and report the greenhouse gas (GHG) inventories for the period 2008-2012. Large amounts of the principal GHG- carbon dioxide (CO₂)- are removed from the atmosphere and stored by forest ecosystems (IPCC, 2000), thus mitigating the magnitude of the global climate change (Gonzales et al., 2010; Fahey et al., 2010). In Norway, reporting of GHG from the Land Use, Land Use Change and Forestry (LULUCF) activities (IPCC 2006) has been using the estimates provided by Norwegian National Forest Inventory (NFI), which is the main source of information regarding carbon pool changes (Tomter et al., 2010). One of the most dynamic and largest carbon pools in forest ecosystems is the aboveground biomass (AGB) stored by living trees (Fahey et al., 2010). In general, carbon inventory systems relying exclusively on ground observations alone may not be feasible for large area surveys due to expense associated with installation and remeasurement of an exhaustive ground network. In addition, regional arrays of ground plots cannot always provide accurate local estimates, for instance at sub-regional and administrative unit level, or by land-use or cover classes (Fahey et al., 2010; Gonzales et al., 2010).

Auxiliary information provided by remote sensing systems has the potential to enhance the terrestrial surveys for forest carbon estimation (Streck & Scholz, 2006; Gonzales et al., 2010). Airborne Light Detection and Ranging (LiDAR) remote sensing has been successfully tested for biomass estimation in various forest types and for different sensors and platforms (Lefsky et al., 2002; Drake et al., 2003; Nelson et al., 2004, 2007; Boudreau et al., 2008; Næsset & Gobakken, 2008; Asner et al., 2009; Næsset, 2011). Airborne laser scanning (ALS) systems were deemed feasible for ‘wall-to-wall’ operational forest management for areas between 20-2000 km² (Næsset, 2009) based on results showing

marked improvements in precision relative to conventional surveys (Næsset 2002, 2004, 2007; Corona & Fattorini, 2008).

Because national-scale projects using ‘wall-to-wall’ ALS coverage might not be feasible, LiDAR measurements have to be integrated within a sampling framework for AGB and carbon stock estimation. Double-sampling applications using ALS measurements on sub-sampled field plots were discussed by Parker & Evans (2004), Andersen & Breidenbach (2007), Corona & Fattorini (2008) and Stephens et al. (*in press*). Profiling airborne lasers like the Portable Airborne Laser System (PALS; Nelson et al., 2003) have been used within a line-intersect sampling framework for areas up to 5000 km² for AGB and forest volume estimation (Nelson et al., 2004; Nelson et al., 2008), and a three-phase survey involving space borne lasers and PALS has been tested by Boudreau et al. (2008) and Nelson et al. (2009). Using ALS as strip sampling tool where the LiDAR measurements are collected along corridors several hundred meters wide has been suggested by Næsset (2005) and Gobakken et al. (2006). This approach has been applied by Andersen et al. (2009) in a two-stage survey for biomass estimation in Kenai Peninsula of Alaska, USA.

Demonstrations of the utility of airborne scanning and profiling lasers within a sampling framework for regional and sub-regional AGB estimation have been undertaken also in Norway, and a large-scale project covering approximately 30,000 km² was conducted in Hedmark County (HC) in the period 2005-2010, where both types of instruments were used as sampling tools along parallel flight lines following the NFI grid (Næsset et al., 2009). The complex inferential challenges posed by the HC sample survey were addressed from a two-stage, design-based model-assisted (DBMA) perspective by Gregoire et al. (2011) and using a two-phase, model-based (MB) approach by Ståhl et al. (2011). Under both approaches, the AGB was estimated using linear regression models. The DBMA inference relies on the randomization distribution of the estimator under a predefined design, wherein

auxiliary data (e.g. LiDAR metrics) are incorporated in the estimation using a sample-driven model. DBMA inference has the attractive feature of being approximately design-unbiased, the randomness being brought by the probabilistic design which validates the inference even when the model fits poorly (Särndal et al., 1992, p. 239). Under the MB framework, the inference relies on a trusted model without explicitly requiring probabilistic sampling, because the random component needed for statistical inference is introduced by the model itself. Further insights into design- and model-based inference can be found in Gregoire (1998), Schreuder et al. (2001), and Kangas (2006).

The endeavours of Gregoire et al. (2011) and Ståhl et al. (2011) succeeded in developing estimators and variance estimators tailored for forest sample surveys where partial coverage with auxiliary information is available. Their results corroborated findings reported by Andersen et al. (2009) and Gobakken et al. (*in submission*) which indicated that sampling strategies (*sensu* Hájek, 1981; Särndal et al., 1992; Gregoire & Valentine, 2008) involving LiDAR auxiliary data might be less precise compared to the conventional field-based surveys when using an equal number of sample plots. However, the inference in all these projects was based on simplifying assumptions which could influence the estimation unexpectedly. For instance, the variance estimators cannot account unbiasedly for the systematic design of both the ALS and the field samples, which would probably result in overestimating the variances (Gregoire et al., 2011; Ståhl et al., 2011; Gobakken et al., *in submission*). Moreover, the DBMA variance estimator can produce negative variance estimates and in such cases the inference becomes impossible (Gregoire et al., 2011; Gobakken et al., *in submission*). This problem compelled us to consider an alternative design-based variance estimator adapted from Särndal et al. (1992, p. 154), which guarantees positive variance estimates.

A direct comparison of strategies involving sampling plans with different structures is difficult (Gregoire et al., 2011), and not knowing the ground truth AGB further complicates the accuracy assessment. Thus, it has been suggested that the statistical properties of the DBMA and MB variance estimators ought to be explored via simulations (Gregoire et al., 2011; Gobakken et al., *in submission*). Using simulated sampling from a completely known finite population one can closely approximate (and in some cases fully describe) the sampling distribution of complex estimators, which otherwise would be difficult, or even impossible, to obtain analytically. The empirical sampling distribution of an estimator can be used to assess the bias and variance relative to the population parameter subject to estimation (Särndal et al., 1992, p. 277; Gregoire & Valentine, 2008, p. 32). Monte Carlo simulations are useful not only for exploring the properties of design-based estimators, but also for investigating the model-based estimators when the inference is directed toward quantifying finite population parameters (so-called descriptive inference by Särndal et al., 1992, p. 514; Kangas, 2006, p. 40). The use of simulations for assessing the properties of LiDAR-based estimators has been previously considered. Using small-scale empirical simulators, design-based strategies were investigated by Ene et al. (2007) and Marcell et al. (2009). The properties of several regression estimators have been assessed by Andersen & Breidenbach (2007) by repeatedly extracting small samples from a population consisting of a large set of field plots and associated LiDAR measurements. Andersen et al. (2011) considered resampling methods for estimating the influence of sampling and model errors on biomass estimates in a two-phase model-based sampling strategy.

This study introduces a sampling simulator created using empirical data (field measurements, ALS data, and satellite imagery). The use of the simulator is exemplified through a case study mimicking a simplified version of the HC survey, focusing on two main

objectives: (1) to assess the performance of variance estimators for ALS-aided inventories, and (2) to assess the relative gain in accuracy obtained using auxiliary ALS data.

2. Material

The study area incorporates all of Hedmark County, a 27340 km² area located in south-eastern Norway (Figure 1). The topography across HC displays large altitudinal variations, from 119 m to 2178 m a.s.l. The northern part of HC consists of mountain areas, with the highest altitudes occurring in the north-western region. The altitude decreases nearly linearly towards the south, while the east-west altitudinal differences are less pronounced.

Approximately 53.7 % of HC's land area is covered by forests (Anon., 2004). The dominant tree species present in HC are Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.).

The general sources of data were used to quantitatively describe and characterize the County: (1) field inventory data from the Norwegian NFI, (2) remote sensing data (ALS and satellite imagery), and (3) cartographical products in the form of a digital terrain model (DTM) as raster data and land-use maps (vector data). All datasets were georeferenced to the WGS84 reference frame, using the UTM (zone 32N) projection model.

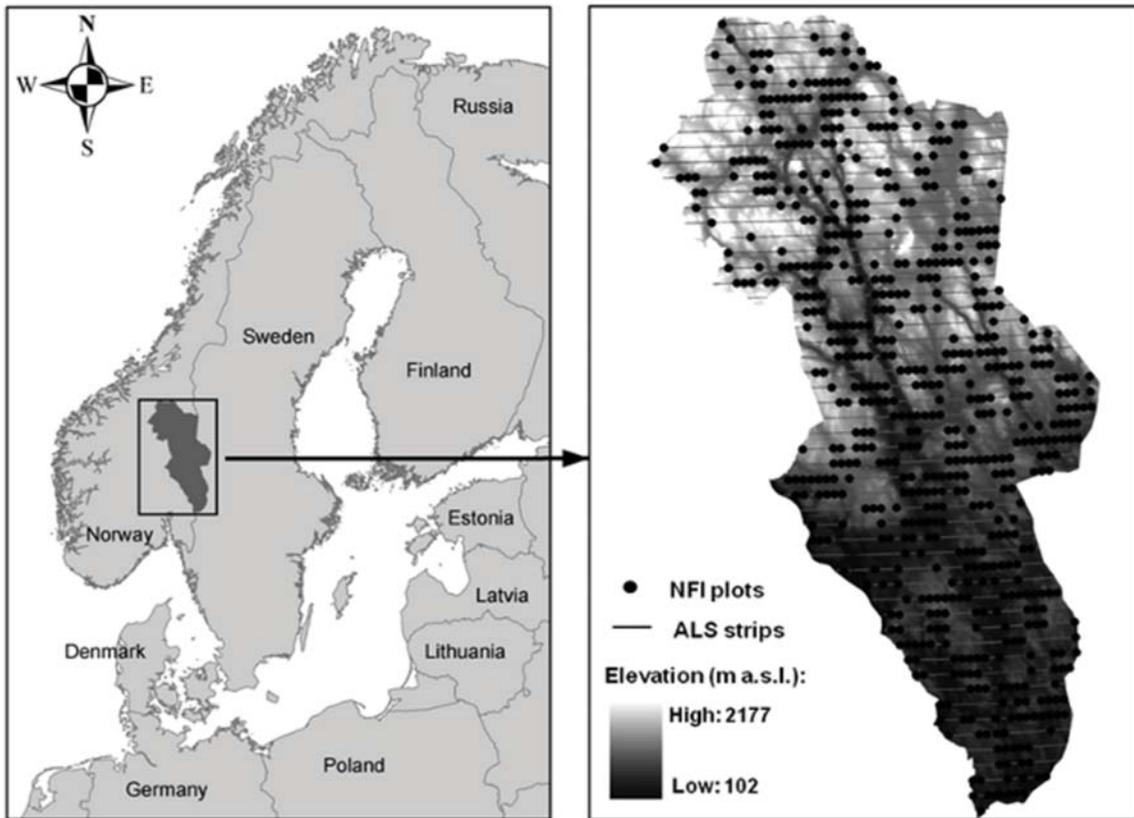


Figure 1 Spatial location of NFI plots and ALS strips in Hedmark County.

2.1 Satellite imagery data

The satellite data consists in three nearly cloud-free Landsat 5 TM images (bands 1 to 5 and 7) acquired in June 2007 (Table 1). The three Landsat images have a georeferencing error of $\leq 15\text{m}$ for 95% of the pixels. The images were atmospherically corrected using the COST model (Chavez, 1996).

Table 1 Satellite imagery metadata

Satellite image	Date	Path/row	Sun elevation (deg)	Sun azimuth (deg)	Spatial resolution (m)
1	03.06.2007	197/16	49.6	162.2	25
2	03.06.2007	197/17	50.0	162.00	30
3	10.06.2007	198/16	50.1	161.9	25

Since none of the three satellite images entirely coverage the HC, an image mosaic was created at 25 m spatial resolution and then resampled to 15.81 m resolution using nearest neighbour resampling, such that the pixel area matched the size of the field plots (250 m²). Although the images were acquired from the same contractor within a one week time interval, differences in surface reflectance due to rapid phenological changes were observed, especially at higher altitudes. Hence, we did not apply relative radiometric calibration or colour-balancing techniques when the Landsat mosaic was built (Koukal et al., 2007).

2.2 Map data

In addition to field and remote sensing data, a 25 m spatial resolution DTM produced by the Norwegian Mapping Authority using the official topographic map series, and *AR5* land use maps (Bjørndal & Bjørkelo, 2006) were used to create a forest mask. The DTM provided full coverage for HC, but the land use map data did not. Land-use categories as water and developed areas were well represented, which was considered sufficient for the objectives of this study. Vector data were converted to raster format using a 25 m pixel resolution. The Landsat mosaic and the DTM were aligned with the rasterized land-use maps to obtain congruent raster layers. The manipulation of map data (raster and vector) was performed using ArcGIS 9.3 software from the Environmental Systems Research Institute (Anon., 2011b).

2.3 Field data

The field measurements were obtained from the permanent Norwegian National Forest Inventory (NFI) grid. The Norwegian NFI is a continuous forest inventory system where the sampling units are fixed-area circular plots of 250m², located on a 3x3 km grid covering the entire country, except for mountain areas above the coniferous tree line, where the plots are located in a 3x9 km grid. The grid axes are oriented to the north-south and east-west directions. HC contains 2309 permanent NFI plots, and the plots are revisited every 5th year following the ordering provided by a country-level Latin square design with 45 x 45 km blocks (Tomter et al., 2010). Between 2005 and 2007, 1483 plots (including NFI and additional plots) were measured. Due to the Latin-square design, the 1483 field plots were not uniformly spread over the HC, nonetheless they cover the main north-south geomorphologic/altitudinal gradients.

For this study, only the plots from every second east-west oriented NFI grid line were used because they were located along the corridors designed for ALS data acquisition, resulting in 662 plots (Table 1) located 6 km apart along those parallel grid lines included in the analysis. More details regarding the field plot selection are given in Gobakken et al. (*in submission*).

On each plot, all trees having a breast height diameter (*dbh*) ≥ 5 cm were callipered, and tree height measurements (*h*) were taken for approximately ten sample trees selected proportional to stem basal area, using an adjustable basal area factor (Tomter et al., 2010). A more detailed description regarding the NFI estimation methods is given by Tomter et al. (2010) and Gobakken et al. (*in submission*). The total above ground dry biomass was predicted for each tree having $h \geq 1.3$ m using tree-species specific allometric models with *dbh* and *h* as predictor variables (Marklund, 1988). The tree level biomass estimates were

summed to obtain plot predictions of above ground biomass (AGB), which then were considered to be equivalent to actual AGB of live trees on the plot.

The plot centre coordinates were determined using differential post-processing of dual-frequency Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) measurements acquired using Topcon LegacyE receivers. The observation time was at least 15 minutes on each plot, at a 2 seconds logging rate. Six base stations were established at different locations across HC, such that the vector ranges between base stations and plots were less than 50 km. Vector differential corrections were calculated for each plot using the nearest base station as reference. The precision of plot centre positioning reported by the Pinnacle 1.0 post-processing software (Anon., 1999) varied between 0 and 2 m, with an average of 0.05 m.

The correlation between the altitudinal gradient in HC and plot biomass expressed using a robust surface fit $\hat{AGB}(x,y) = 1.053 - 0.01x - 0.2782y$ ($RMSE = 42.4 \text{ Mg ha}^{-1}$, $R^2_{\text{adj}} = 0.38$) indicated a dominant trend along the y-axis (north-south direction). Generally, the plot biomass seemed to be inversely correlated with the topographic trend, being lower on higher altitudes (in the northern part of HC) and increasing towards South.

Table 2 Biomass distribution by field plots

Cover class category	Plots by strata		Above ground biomass (Mg ha ⁻¹)			Altitude (m a.s.l.)		
	No	%	mean	SD ⁽¹⁾	CV ⁽¹⁾	mean	SD	CV
Productive forest:								
Low	138	21	50.9	38.99	76.6	611.9	207.39	0.3
Medium	105	16	96.7	57.53	59.5	445.7	146.1	0.3
High	46	7	129.3	77.06	59.6	335.1	108.7	0.3
Young forest	151	23	34.9	40.67	116	454.6	190.4	0.4
All	440	66	-	-	-	-	-	-
Nonproductive forest and non-forest:								
Nonproductive forest ⁽²⁾	107	16	28.6	28.65	100	645.7	207.4	0.3
Mountain areas	85	13	22.8	20.85		887.2	69.5	0.1
Developed areas	30	5	27.7	34.64	125	211.5	99.9	0.5
All	222	34	-	-	-	-	-	-
Total	662	100	51.7	53.66	104	553.1	232.73	0.42

⁽¹⁾ standard deviation (Mg ha⁻¹); ⁽²⁾ variation coefficient (%); ⁽²⁾ annual growth < 1m³ year⁻¹ ha⁻¹

2.4 ALS data

The ALS data acquisition was carried out during leaf-on conditions, between July and September 2006, using two Optech ALTM 3100 laser scanning systems (Optech, Canada) mounted on two PA31 Piper Navajo aircrafts. Approximately 8.4 % of the county's land area was sampled from fifty-three parallel flight lines equally spaced at a distance of 6 km in the north-south direction, and covering approximately 50% of the plots. The average flying height was at approximately 800 m above ground, at an average speed of 75 ms⁻¹. The laser scanners operated at a pulse repetition frequency of 100 kHz and a scan frequency of 55 Hz, which translated to an average density of 2.8 pulses m⁻². The echoes recorded at scanning angles greater than 17° were removed, obtaining an average strip width of approximately 500 m. The 3D adjustment, the classification (canopy and ground) of laser echoes and the digital terrain surface were performed using TerraSolid software (TerraSolid Ltd, Finland). After height normalization relative to the digital terrain surface, only the echoes having relative heights above 1.3 m were classified as canopy hits. Further in this analysis, we used only the laser echoes recorded as 'single' and 'first of many' by the ALTM 3100 sensors.

3. Methods

It was assumed that the design parameters are fixed, because they are related to the NFI system. We also assumed that the general form of the regression model for predicting the biomass as a function of ALS measurements is not known before sampling the data, which is a reasonable assumption for ALS-based surveys where the ALS measurements can be strongly influenced by a variety of factors (Næsset, 2009; Ørka et al., 2010).

Practically, the approach employed to construct the simulator requires that the empirical data (e.g. field data and auxiliary information) are first sampled from the original population following a stipulated sampling design. Then, an artificial population is created using the empirical data, following the steps presented in this section. Using the artificial population as ground-truth, sampling is then simulated following a design which resembles closely the original one, and the properties of several estimators are assessed. Simulation results, then, help to identify those sampling strategies that provide the most accurate and precise estimates. As important, the results can also help identify those assumptions that must be met in order to generate reliable, robust estimates of biomass.

The remainder of the section proceeds as follows: we start with building predictive regression models for AGB using the LiDAR metrics. Then, we describe the creation of the artificial population and the mechanisms for generating the samples. Finally, the accuracy assessment for several strategies is demonstrated through a case study.

3.1 Predictive LiDAR models for AGB

The relationship between plot-level AGB and LiDAR measurements was investigated using multiple linear regression. Twenty two LiDAR metrics were derived from the height

distributions of laser echoes (first returns only) covering each plot: the maximum height (H_{max}), the mean height (H_{mean}), the deciles ($H_{d, d=0.9}$) of the echo height distributions, and the canopy densities ($D_{d, d=0.9}$) obtained as described by Næsset (2004) and Gobakken et al. (*in submission*). Although we found an altitudinal gradient across HC, the height above sea level was not included among predictors because it produced many influential outliers.

We considered that AGB during the simulated sampling should be predicted using a fixed set of covariates obtained from the original AGB and LiDAR metrics, however the model parameters would be estimated from each simulated sample. For building the predictive model, we adopted a two-step strategy (Draper and Smith, 1998, p. 343-344). First, we selected a subset of first-order regressors among the LiDAR metrics using a backward stepwise variable selection based on Akaike Information Criterion (AIC) as stopping rule. In the second step, we build candidate models from the variables selected at the first step, adding also interaction and higher order terms following the hierarchy principle. The relationship between forest variables and LiDAR metrics is usually linearized via nonlinear transformations using logarithmic transformation of both the regressors and the response variable (Næsset 2002, 2004; Næsset & Gobakken, 2008; Li et al. 2008), or by the square root transformation of the response (Andersen & Breidenbach, 2007; Boudreau et al. 2008; Næsset, 2011). In order to avoid non-linear transformation of the response variable, Gregoire et al. (2008) advocates using generalized linear models (McCulloch & Nelder, 1989) instead of ordinary least square regression, thus eliminating the need for back-transforming the response which is prone to biases. We subscribed to this advice and we used a generalized regression model assuming normal distribution of errors and square-root link function. The model estimation was performed by iteratively re-weighted least square (McCulloch & Nelder, 1989)) using MATLAB[®] (Anon., 2011a). The model considered to the most appropriate was formulated as:

$$\begin{cases} g(\mu_{AGB}) = X^t \beta \\ AGB = \mu_{AGB} + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad \text{eq(1)}$$

where $X^t \beta$ is the linear predictor including the H_{max} , D_{10} and the interaction term $H_{max} \times D_{10}$, and $g(\cdot)$ is the square-root link function ($RMSE = 17.40 \text{ Mg ha}^{-1}$ (36.12%), $r^2 = 0.90$, where r is the linear correlation coefficient between original and predicted AGB values (Zheng & Agresti, 2000)). During simulations, the model parameters β of the linear predictor were estimated from each sample.

3.2 Creating the sampling simulator

The sampling simulator consists of an artificial population and a sample selection mechanism.

The artificial population should represent realistically the forested areas across HC. For simulating the ALS-based inventory, it is also necessary to generate a full coverage of LiDAR metrics which are consistent with the artificial forest. Although there is no straightforward way for modelling the population, two main approaches can be of interest: an individual tree level representation and an area-based approach. A simple, individual tree approach was used by Ene et al. (2007) to generate an artificial population using a forest stand generator, individual tree measurements and laser echoes delineated for individual trees using stem and crown locations. Marcell et al. (2009) built a ‘LiDAR-forest’ at tree level using the methodology of Popescu & Wynne (2004). The latter approach is not applicable in our case, since we only had partial coverage with ALS data for HC. It is not clear whether the individual tree modelling approach would result in a more accurate description of the forest across large areas like HC. Instead, we constructed the ‘LiDAR-forest’ considering a fully

tessellated representation of the HC area congruent to the satellite imagery raster, and distributing the plot-level AGB and LiDAR metrics via nearest neighbour imputations using the satellite imagery mosaic (Tomppo et al., 2008). The area-based approach also permits an easier integration and manipulation of different data sets. At this stage it should be noted though that the aim was not to re-create the true forest of HC, but to create an artificial population comprising the major geographical trends in HC while maintaining a fairly realistic overall variability.

When using nearest neighbor imputations, the 662 empirical plot-level observations (AGB and LiDAR metrics) would be replicated across the entire study area. Given the small sample size relative to the population size (approximately 10^7 elements), this would cause a small population variability, and large samples would have many duplicate observations. Since it was not possible to increase the number of observations by acquiring more data, we used copula functions for generating a large multivariate set of joint AGB, LiDAR, and multispectral observations.

The creation of the artificial population follows three main steps: (1) data preparation, (2) stochastic data generation using copulae, and (3) generalization of copula sample using nearest neighbour (*NN*) imputations. An overview of the process is presented in Figure 2.

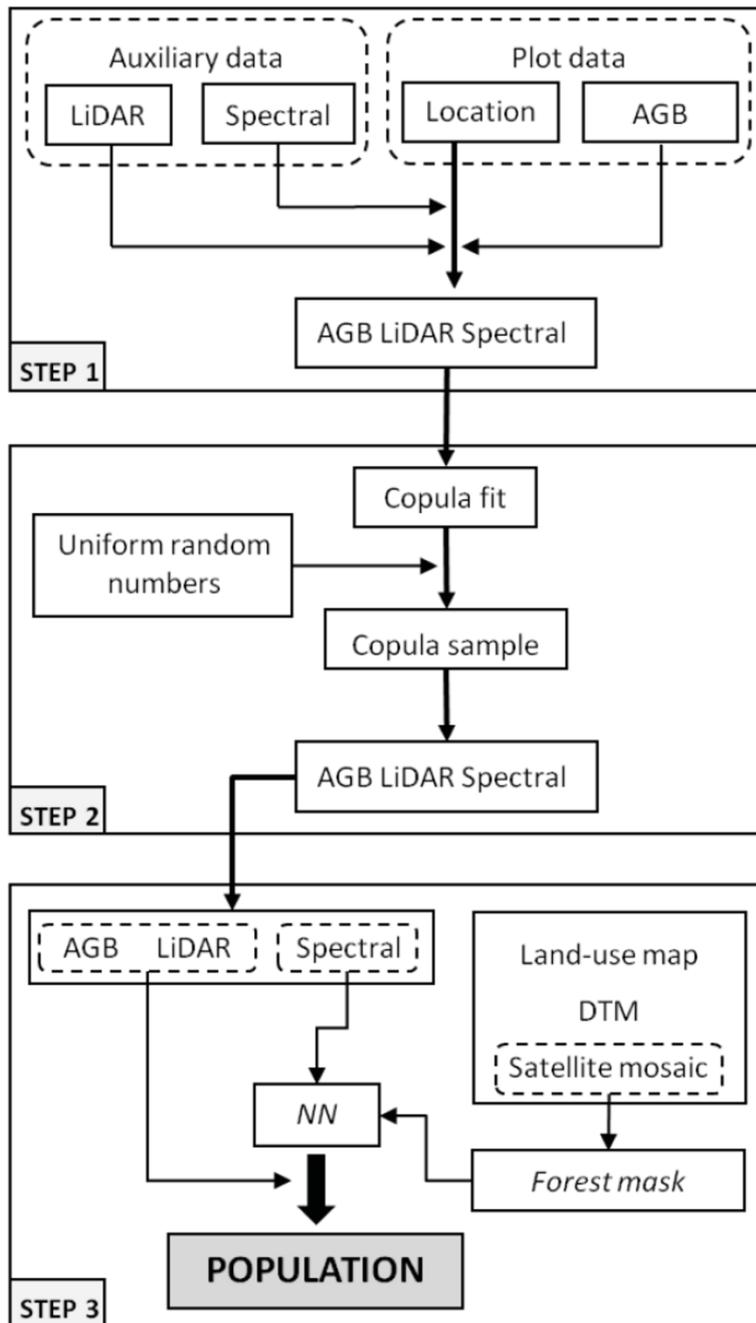


Figure 2 The main steps for generating the artificial population.

3.2.1 Step 1: Data preparation

Construction of the population requires that a set of LiDAR metrics, satellite multi-spectral records and field-based estimates of AGB can be assigned to each of the 250m² grid cells.

The empirical plot-level data set used for creating the artificial population comprised the

AGB estimates, the LiDAR metrics, and spectral information extracted from the satellite imagery mosaic pixels covering the plots. When extracting the plot-level spectral data, the possible registration errors between field plots and satellite imagery pixels were ignored.

3.2.2 Step 2: Generating the copula sample

Copulae are mathematical functions which are useful for modelling dependencies between random variables (Nelsen, 2006). According to Sklar's theorem (Nelsen 2006, p. 21), copulae link the joint probability distribution functions to their marginal distributions. The rank correlations between variables are preserved by their cumulative distribution functions (*cdf*), and copula functions are used to model these dependencies, making it possible to preserve non-linear relationships existing in the empirical observations. Thus, a joint probability distribution can be expressed using the marginal distributions and the dependency structure between marginals captured by the copula function.

The process of fitting copula functions and generating copula samples consists of several steps (Anon., 2011a):

- identify the univariate *cdfs* of the variables and independently transform each of these *cdfs* to uniform (0, 1) distributions;
- fit the appropriate copula function to the uniforms;
- generate uniform random numbers using the copula function, and
- retransform the uniform copula sample using the appropriate inverse *cdfs*.

The derivation of the *cdfs* and the retransformation steps were performed using a non-parametric density estimator with a Gaussian kernel function.

Although the copulae are used for stochastic simulations, copulae functions describe the correlation pattern between uniforms in a deterministic fashion. The stochasticity of the copula samples is obtained by evaluating the copula function using uniform pseudo-random numbers. For an in-depth treatment of copulae see Bouyé et al. (2000), Nelsen (2006), Kojadinovic & Yan (2010) and Schepsmeier & Brechmann (2011).

Copulae are popular modelling tools in many fields (e.g. econometrics), but have been little used in forestry applications. The Gaussian copula were found to perform well for bivariate modelling of tree diameters and height (Wang et al., 2008), and for trivariate modelling of tree diameter, height and volume (Wang et al., 2010). A seven dimensional normal copula was used by Miina and Heinonen (2008) in a multilevel multivariate stochastic modelling study of the regeneration establishment. Kershaw et al. (2010) used Gaussian copula for generating spatially correlated forest stand structures, and recently Gaussian copula were used by Eskelson et al. (2011) in a simulation study for estimating shrub cover in riparian forests.

For applications focusing on estimating the dependence structure, it is important to use the copula which best captures the dependence between variables. Although many copula functions have been documented for the bivariate case, only a few can cope with high dimensional datasets (Genest et al., 2009). New approaches for building multivariate copulae have been recently proposed (Aas et al., 2009; Schepsmeier & Brechmann, 2011), but for this study we used the Gaussian copula, with which it is easy to fit to high-dimensional data (Anon., 2011a). The Gaussian copula is derived from the multivariate normal distribution, thus it fully defines the dependence structure using the correlation matrix. Regardless of the correlation matrix, the Gaussian copula has zero tail dependence.

3.2.3 Step 3: Generalizing the copula sample by nearest neighbour imputations

Using the Gaussian copula, a large sample of 100,000 observations was generated from the AGB-LiDAR-Spectral dataset. Figure 3 illustrates the similarity between the copula-generated simulation pool and the actual ground observations for the combinations of ALS predictors and biomass.

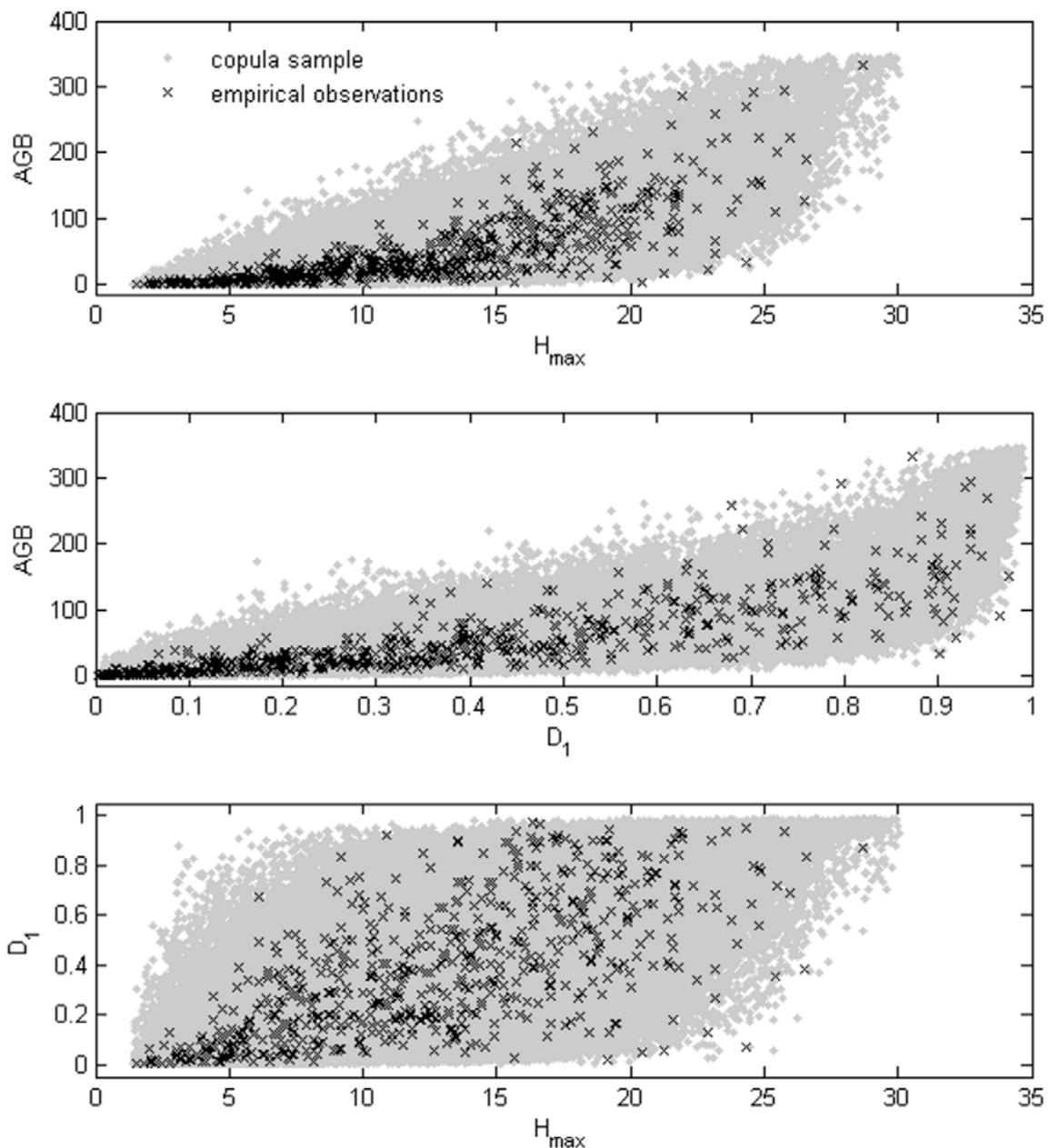


Figure 3 Dependencies between AGB (Mg ha⁻¹) and LiDAR metrics in original data and in the copula sample. See text for description of LiDAR metrics.

Compared to original data, the linear correlations between AGB and LiDAR height metrics in the copula sample (Figure 4) were slightly lesser, especially for the upper LiDAR deciles, while the correlation between AGB and canopy density metrics related to the upper canopy fractions was slightly greater. Also, the dependencies between AGB and the LiDAR included in the regression model were little altered.

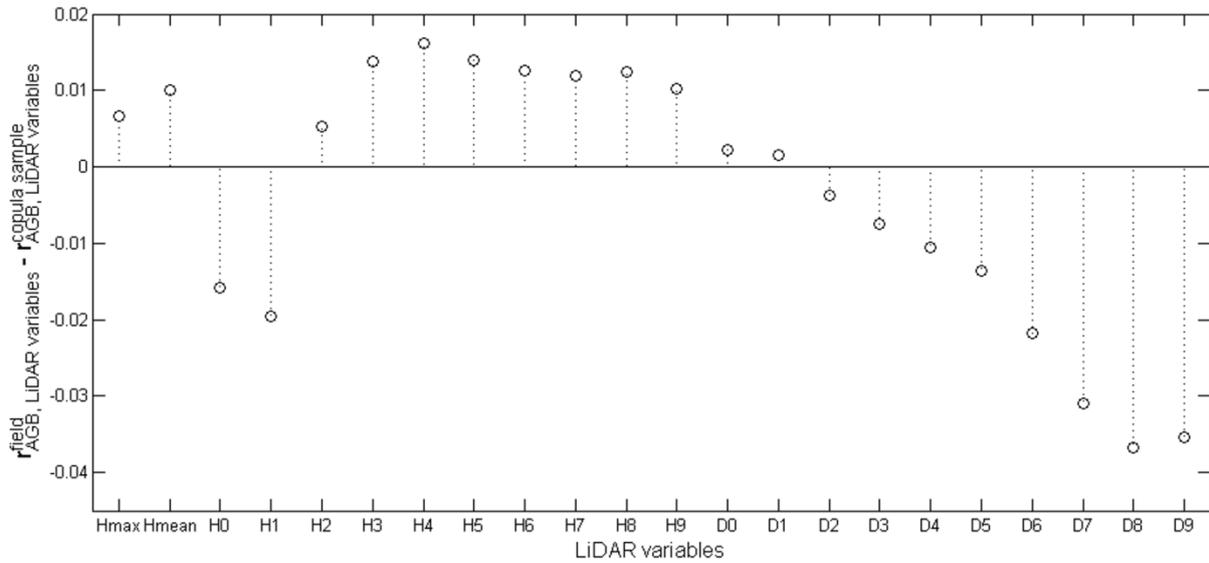


Figure 4 Linear correlations between AGB and LiDAR metrics in the original observations and copula-generated data. The LiDAR metrics derived from first returns consist of maximum echo height (Hmax), mean echo height (Hmean), echo height deciles (Hd, p=0-9), and the canopy densities ($D_d, d=0-9$).

The copula-generated AGB-LiDAR-Spectral dataset was merged with the original 662 plot observations, and we called this final dataset AGB-LiDAR-Spectral*. The observations from AGB-LiDAR-Spectral* were imputed across the forested area of HC using nearest neighbour imputations. In order to do so, a forest mask for HC was created using the DTM and land-use maps. The DTM was used for delineating the areas above 1150 m altitude, which corresponds roughly to the coniferous tree line in HC (Ørka et al., *accepted*). The land use maps were used for eliminating water and other types of non-forested areas. Because we had poor coverage with map data for forest vegetation located in developed areas and above 850 m a.s.l., we used a box-classifier and the multispectral information associated

with the field plots to separate the image mosaic pixels which might cover forest vegetation in this areas. The forest mask obtained represents about 86% of the tessellated HC area.

The nearest neighbour imputations were based on the Gaussian distances between the multispectral features (bands 1-5 and 7) from the copula dataset and the satellite imagery mosaic. A summary of the nearest neighbour imputations presented in Table 3 indicated that the average AGB in the population (μ_{AGB}) was lower compared to both the field sample and imputation data. However, the range, variability, skewness and kurtosis in the imputed AGB were slightly higher relative to the field data.

Table 3 Descriptive statistics for AGB in field data, copula sample and artificial population

AGB (Mg ha ⁻¹)	min	max	mean	SD	skewness
Population	0.121	345.764	48.185	55.338	1.846
Copula sample	0.121	346.307	53.128	56.743	1.713
Field sample	0.264	331.511	51.727	53.656	1.716

The preservation of main geographical trends identified in the field sample was analyzed by applying a robust surface fit in the form $AGB(x, y) = \beta_0 + \beta_x x + \beta_y y$ to the imputed AGB values ($R^2 = 0.56$, $RMSE = 36.44$ Mg ha⁻¹), obtaining estimated model parameters close to the values obtained using the field sample data

($\beta_x = -0.022$, $\beta_y = -0.331$). Hence, we concluded that the main geographical trends present in the original data were also reasonably well preserved in the artificial population.

The artificial population was stored in the form of a look-up table, where each cell is linked to a specific set of biomass and ALS metrics from among the AGB-LiDAR-Spectral* dataset. This look-up table was considered as a spatial sampling frame congruent to the raster representation of the HC area. During sampling, any unit can be identified by its linear in the look-up table, and points to an observation in the AGB-LiDAR-Spectral* dataset. The frame origin has the index (1, 1) and it was set as the upper-left grid cell. Using this sampling

frame, the distances between population elements could be expressed in number of grid cells. Hence, the distances (e.g. sampling intervals and strip width) were converted from meters to number of pixels by division to 15.81 m (pixel size) and rounding to integer value using the ceiling operator. Hereafter, the terms pixel and population elements are used interchangeably.

3.3 Case study

We conducted a case study involving simulated sampling related to the biomass survey in HC. The inputs required for running the sampling simulations can be grouped as: (1) design parameters (e.g. the swath, location and orientation of ALS strips, the number and location of the field plots), (2) the inferential approach (the estimators of the descriptive population parameter (μ_{AGB}) and the inferential framework), and (3) the number of sampling simulations.

Three estimation cum inferential scenarios were considered: design-based inference with Horvitz-Thompson (HT) estimator, model-assisted estimation with design-based inference and model-based estimation and inference. For the sake of easy reference, these are abbreviated as DBHT, DBMA and MB henceforth.

The DBHT strategy consists of systematic sampling (*SYS*) with HT estimator, and corresponds roughly to the current Norwegian NFI system. The inference relies upon the sampling design and makes no use of auxiliary information.

Under DBMA, the auxiliary information provided by ALS is incorporated into the estimation by the linear regression model developed in section 3.1, but the validity of the inference is ensured by the probabilistic design.

The MB inference for the finite population parameter $\hat{\mu}_{AGB}$ relies on a model without explicitly requiring probabilistic samples. The model-based inference would usually require full coverage with auxiliary data, such that predictions can be made for all the non-sampled population elements. For the HC survey, the coverage with ALS data was only partial, thus

requiring a probabilistic sample of auxiliary data, while the field plot selection could be either probabilistic or not. The model-based estimator proposed by Ståhl et al. (2011) tackled this situation by taking into consideration both sampling and model errors.

The sampling designs and the estimators are described in the remainder of this section, following Gregoire et al. (2011) and Ståhl et al. (2011).

3.3.1 One-phase systematic sampling with HT estimator

Under the one-phase sampling design with restricted selection in two dimensions (SYS_1), the first sample unit is drawn from the first $a_{x,l} \times a_{y,l}$ population elements, where $a_{x,l}$ and $a_{y,l}$ are the sampling intervals (in pixels) along the spatial frame's axes oriented towards the east and north directions. The rest of the sample units are selected systematically using predefined sampling intervals $a_{x,l}$ and $a_{y,l}$. The total number of possible samples is $a_{xy,l} = a_{x,l} \times a_{y,l}$, each sample being selected with the same probability according to the design $p_{SYS_1} = a_{xy,l}^{-1}$. The inclusion probability for the k^{th} population element in a sample is $\pi_k = a_{xy,l}^{-1}$. Under this design, the sample size is not fixed due to the irregular borders of HC. The sampling intervals along the east and north directions were $a_{x,l} = 3000\text{m}$ (190 pixels) and $a_{y,l} = 6000\text{m}$ (380 pixels). This design does not fully comply with the Norwegian NFI system because it disregards the continuous nature of the inventory, by assuming that all field plots distributed across HC are visited during one campaign. Such an intensive survey would also produce estimates having higher accuracy than the usual situation, when in fact only 20% of the plots are visited during a campaign (one field season). The total number of possible samples produced under the SYS_1 design was $a_{xy,l} = a_{x,l} \times a_{y,l} = 190 \times 380 = 72,200$.

The total AGB is unbiasedly estimated by the Horvitz-Thompson estimator from the k^{th} sample as (Gregoire & Valentine, 2008, p. 52):

$$\hat{t}_{HT} = a_{xy,1} \sum_{i=1}^n AGB_i \quad \text{eq(2)}$$

and the unbiased estimate of the average AGB per hectare is obtained as:

$$\hat{\mu}_{HT} = cf N^{-1} \hat{t}_{HT} \quad \text{eq(3)}$$

where N is the number of population elements and $cf = 1000 \text{ m}^2/250 \text{ m}^2 = 40$ is the per-hectare expansion factor. The design-based variance of $\hat{\mu}_{HT}$ was approximated using the variance estimator for simple random sampling without replacement as:

$$\hat{v}_{1.1} = cf^2 \left(1 - \frac{n}{N}\right) \frac{s_{AGB}^2}{n} \quad \text{eq(4)}$$

where s_{AGB}^2 is the sample variance. In equation 4, the finite population correction term $1 - n/N$ can be omitted because the samples size (n) is negligible relative to the number of population elements.

3.3.2 Two-phase systematic sampling with regression estimator

The sampling design used for the HC ALS-aided survey involves systematic sampling at both selection steps (SYS_2). Under this design, the subsampling is invariant since any primary sampling unit (PSU) belongs to only one sample but the selection of secondary sampling units (SSU) is not independent because it follows the NFI grid. Moreover, the MB framework we are considering here allows using two independent sources of auxiliary data (Ståhl et al., 2011). For this reason, this design can be considered a two-phase element sampling design according to Särndal et al. (1992, p. 344-345).

To perform the two-phase sampling, HC was partitioned into M mutually exclusive clusters, each cluster representing a sub region fully covered by an ALS strip, resulting in a number of 625 clusters oriented parallel to the east-west direction. The cluster width was set to 32 pixels (approximately 500m) to match the average ALS strip width in the HC survey.

Because the spatial frame was not an integer multiple of the cluster swath, the width of the last cluster was 29 pixels (approximately 485m). The clusters have also varying lengths, because the target population is distributed on an irregularly shaped area. The clusters contained varying numbers of sample frame units, and consequently different numbers of population elements. Using this selection mechanism eliminates the possible misalignments between ALS data and field plots which might occur in practical applications.

Using the 1-in- $a_{y,2}$ selection (where $a_{y,2}$ is the sampling interval for ALS strips) produces $a_{y,2}$ possible equal probability first-phase samples, and the n_1 PSUs in each sample are selected with equal probability $\pi_p = a_{y,2}^{-1}$ among the M clusters. In the actual ALS survey of HC, the flight lines were 6 km apart and each swath was parsed to a 500m width, which translates into a sampling interval $a_{y,2} = 12$ ($\approx M/n_1$). Henceforth, the number of pixels in the p^{th} PSU is denoted $n_{1,p}$, and the number of SSUs sampled from the p^{th} PSU was indicated as $n_{2,p}$. The subsampling was carried out as a two-dimensional systematic sampling with restricted selection of the first unit among the first $ALS_w \times a_{x,2}$ elements of the first PSU in the first-stage sample, where ALS_w represents the width of the ALS strips (32 pixels). The sampling intervals along the (x, y)-axis following the HC survey were $a_{x,2} = 3000\text{m}$ (190 pixels) and $a_{y,2} = 6000\text{m}$ (380 pixels). Thus, the number of all possible second-stage samples selected from every first-stage sample became $ALS_w \times a_{x,2} = 32 \times 190 = 6,080$. A graphical representation of the sample selection mechanism is presented in Figure 5.

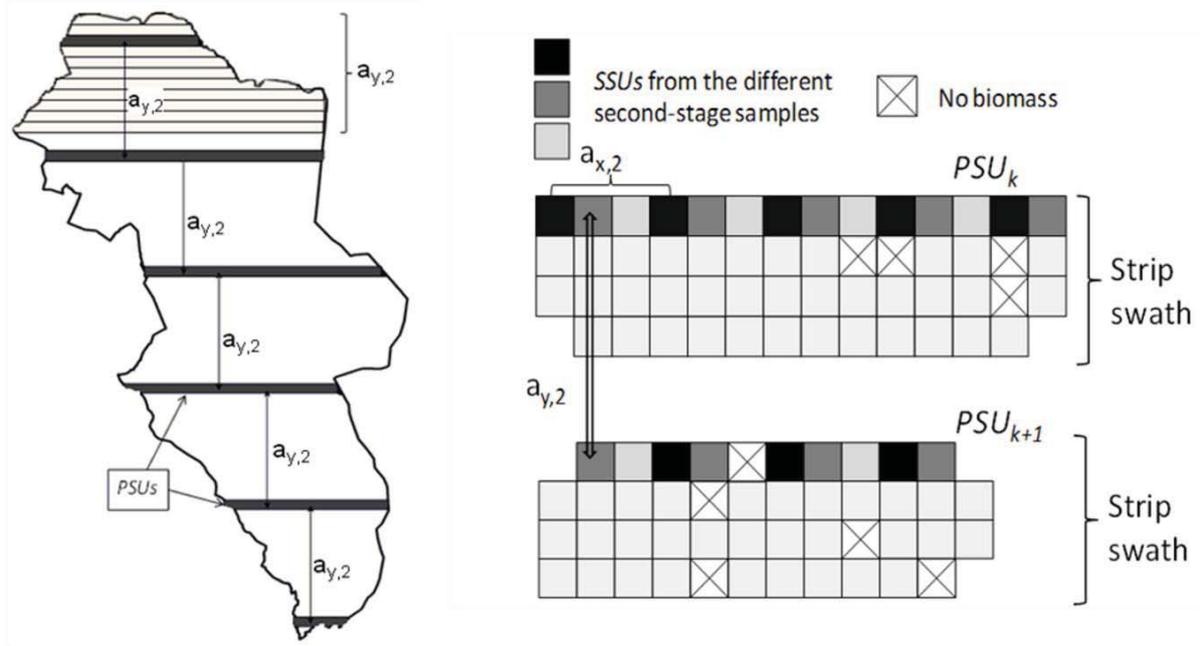


Figure 5 The sampling plan for the case study survey. The sampling interval for selecting the PSUs (the ALS strips) is $a_y = 6$ km. Within PSUs, the field plots (SSUs) were selected using a sampling interval $a_x = 3$ km.

The inference for μ_{AGB} was performed first using the estimator described by Gregoire et al. (2011), developed from the type-C regression estimator for two-stage element sampling proposed by Särndal et al. (1992, §8.9). The estimator assumes *SRSwoR* at both selection steps, independence and invariance for subsampling, conditions required by the two-stage element sampling in Särndal et al. (1992, p. 134-135).

The total AGB for the p^{th} PSU was estimated using the generalized regression estimator (GREG) described by equation 8.9.6 in Särndal et al. (1992, p. 323) and equation 6 in Gregoire et al. (2011). The AGB for the pixels contained in the first-phase sample were predicted using the model from equation 1 fitted to the second-phase samples, ignoring the cluster structure:

$$\hat{i}_{GREG,p} = \sum_{k=1}^{n_{1,p}} A\hat{G}B_k + \frac{n_{1,p}}{n_{2,p}} \sum_{j=1}^{n_{2,p}} (AGB_{p,j} - A\hat{G}B_{p,j}) \quad \text{eq(5)}$$

In equation 5, $A\hat{G}B_{p,j}$ is the predicted *AGB* value for the j^{th} SSU in the p^{th} PSU. The first term is the sum of the predictions for all population elements in a selected PSU, and the second

term is a probabilistic bias correction derived from the regression residuals and the inclusion probabilities $n_{1,p}/n_{2,p}$ of the population elements in the second-stage sample, conditioned on the selection of the p^{th} PSU at the first stage. Thus, the overall *AGB* total becomes

$\hat{t}_{GREG} = a_{y,2} \sum_{p=1}^{n_1} \hat{t}_{GREG,p}$, and the average per hectare *AGB* is unbiasedly estimated by:

$$\hat{\mu}_{GREG} = \frac{cf}{N_{pop}} \hat{t}_{GREG} . \quad \text{eq(6)}$$

An approximated variance estimator of $\hat{t}_{GREG,2}$ is given by Gregoire et al. (2011) as:

$$\hat{v}(\hat{t}_{GREG}) = M^2 \left(\frac{1}{n_1} - \frac{1}{M} \right) \left[s_{\hat{t}_\pi}^2 - \frac{1}{n_1} \sum_{p=1}^{n_1} n_{1,p}^2 \left(\frac{1}{n_{2,p}} - \frac{1}{n_{1,p}} \right) s_{y,p}^2 \right] + \frac{M^2}{n_1^2} \sum_{p=1}^{n_1} n_{1,p}^2 \left(\frac{1}{n_{2,p}} - \frac{1}{n_{1,p}} \right) s_{e,p}^2 \quad \text{eq(7)}$$

In equation 7, $s_{\hat{t}_\pi}^2$ estimates the variance between PSUs as $s_{\hat{t}_\pi}^2 = \frac{1}{n_1 - 1} \sum_{p=1}^{n_1} \left(\hat{t}_{\pi,p} - \bar{\hat{t}}_\pi \right)^2$,

$\hat{t}_{\pi,p} = \frac{n_{1,p}}{n_{2,p}} \sum_{j=1}^{n_{2,p}} AGB_{p,j}$ is the HT estimate of the p^{th} PSU total, and $\bar{\hat{t}}_\pi = \frac{1}{n_1} \sum_{p=1}^{n_1} \hat{t}_{\pi,p}$ is the

average of the estimated PSU totals. The $s_{y,p}^2$ is the intra-cluster sample variance estimated as

$s_{y,p}^2 = \frac{1}{n_{2,p}} \sum_{k=1}^{n_{2,p}} \left(AGB_{p,k} - \hat{AGB}_p \right)^2$, where $AGB_{p,k}$ is the *AGB* value of the k^{th} element in the

second-stage sample from the p^{th} PSU, and $\hat{AGB}_p = \frac{1}{n_{2,p}} \sum_{k=1}^{n_{2,p}} AGB_k$ is the average *AGB* of the

second-stage sample units within the p^{th} PSU. The variance due to subsampling is

approximated by the second term of equation 7, which prorates the cluster residual

variances $s_{e,p}^2 = \frac{1}{n_{2,k} - 1} \sum_{k=1}^{n_{2,p}} \left(e_{p,k} - \bar{e}_p \right)^2$, where $e_{p,k} = AGB_k - \hat{AGB}_k$ is the k^{th} regression

residual in the p^{th} PSU, and $\bar{e}_p = \frac{1}{n_{2,p}} \sum_{k=1}^{n_{2,p}} e_k$ is the average residual for the p^{th} PSU. From

equation 7, the first DBMA estimator of the per-hectare AGB variance becomes:

$$\hat{v}_{2.1} = \left(\frac{cf}{N_{pop}} \right)^2 \hat{v}(\hat{t}_{GREG}). \quad \text{eq(8)}$$

By construction, the $\hat{v}_{2.1}$ estimator can produce negative estimates (Gregoire et al., 2011) when large intra-cluster variances occur, especially due to a reduced number of field observations (Gobakken et al., *in submission*). For such situations, an alternative would be using the DBMA variance estimator for multistage sampling described by Särndal et al. (1992, p. 154) which assumes *SRSwoR* of PSUs:

$$\hat{v}^*(\hat{t}_{AGB,\pi}) = M^2 \left(\frac{1}{n_1} - \frac{1}{M} \right) \frac{\sum_{p=1}^{n_1} \left(\hat{t}_{\pi,p} - \frac{1}{n_1} \sum_{p=1}^{n_1} \hat{t}_{\pi,p} \right)^2}{n_1 - 1} \quad \text{eq(9)}$$

The \hat{v}^* estimator from equation 9 assumes that the cluster totals $\hat{t}_{p,\pi}$ are estimated using the Horvitz-Thompson estimator and it ignores subsampling. When used as an approximate variance estimator of \hat{t}_{GREG} , equation 9 can be written as:

$$\hat{v}^*(\hat{t}_{GREG}) = M^2 \left(\frac{1}{n_1} - \frac{1}{M} \right) \frac{\sum_{p=1}^{n_1} \left(\hat{t}_{GREG,p} - \frac{1}{n_1} \sum_{p=1}^{n_1} \hat{t}_{GREG,p} \right)^2}{n_1 - 1} \quad \text{eq(10)}$$

and the per-hectare variance can be expressed as:

$$\hat{v}_{2.2} = \left(\frac{cf}{N_{pop}} \right)^2 \hat{v}^*(\hat{t}_{GREG}) \quad \text{eq(11)}$$

Under the MB inference, the AGB total of the p^{th} PSU is predicted using the regression estimator (REG) becomes

$$\hat{t}_{REG,p} = \sum_{s_{1,p}-s_{2,p}} A\hat{G}B + \sum_{s_{2,p}} AGB, \quad \text{eq(12)}$$

where the predictions were performed using the same model as in the case of GREG. The second term in equation 12 is the summation of AGB over the subsample in the p^{th} PSU, and the first term is the summed predictions for rest of the population elements within the p^{th} PSU. Under 1-in- $a_{y,1}$ *SYS* of PSUs, the the estimate of the population total AGB is expressed as

$$\hat{t}_{REG} = c_f a_{y,2} \sum_{p=1}^{n_1} \hat{t}_{REG,p} \quad \text{eq(13)}$$

and the average per-pixel AGB becomes:

$$\hat{\mu}_{REG} = \frac{\hat{t}_{REG}}{N_{pop}} \quad \text{eq(14)}$$

Following Ståhl et al. (2011), the approximated variance estimator of $\hat{\mu}_{REG}$ assuming *SRSwoR* and finite population correction for the first phase can be written as:

$$\hat{v}_{2.3} = \frac{M^2}{N_{pop}^2} \left[\left(\frac{1}{n_1} - \frac{1}{M} \right) s_{\hat{t}_{REG}}^2 \right] + \frac{1}{N_G} \sum_{i=1}^k \sum_{j=1}^k \text{cov}(\hat{\beta}_i, \hat{\beta}_j) \hat{G}_i \hat{G}_j. \quad \text{eq(15)}$$

The first term in the brackets estimates the sampling variance due to the first-phase selection,

where $s_{\hat{t}_{REG}}^2 = \frac{1}{(n_1 - 1)} \sum_{p=1}^{n_1} \left(\hat{t}_{REG,p} - \frac{1}{n_1} \sum_{p=1}^{n_1} \hat{t}_{REG,p} \right)^2$ is the sample variance between the predicted

PSU totals $\hat{t}_{REG,p}$. The second term accounts for the uncertainty of the estimated model parameters,

and it includes the covariance matrix of the k regression parameters $\text{cov}(\hat{\beta}_i, \hat{\beta}_j)$ and the average

values of the first-order derivatives of the inverse g function (the square-root link) from equation 1.

3.3.3 Two-phase *SRSwoR* sampling with regression estimator

For this strategy, we assumed that both ALS and field data were collected using a *SRSwoR* design (*SRSwoR*₂). The size of the first-phase samples was set a priori as the expected size of the first-phase sample under *SYS*₂ ($n_1 = M/a_{y,2}=52$). The second-phase samples were selected randomly and independently within each PSU. In order to fulfil the invariance requirement,

the inclusion probabilities of the SSUs were calculated as the ratio of the PSU length (in pixels) and the sampling interval along the east-west direction ($a_{x,2} = 190$ pixels) used under SYS_2 . Hence, the size of the second-phase sample would equal in the long run the expected size of a second-phase sample selected by SYS_2 .

The $SRSwoR_2$ design complies with the theoretical assumptions considered by Gregoire et al. (2011) and Ståhl et al. (2011) for building the variance estimators for the total AGB. Thus, we could use simulated sampling following the $SRSwoR_2$ design for validating $\hat{v}_{2,1}$, $\hat{v}_{2,2}$ and $\hat{v}_{2,3}$ presented in section 3.3.2.

3.4 Accuracy assessment

For the accuracy assessment the aim was to evaluate the performance of the variance estimators and to assess the relative accuracy of the two-phase strategies compared to pure HT estimation under SYS_1 . The assessment was performed assessing the sampling distributions of the estimators derived from repeated sampling from the fixed artificial population. However, the sampling distributions of an estimator under design-based and genuine model-based inference are not identical (Särndal et al., 1992, p. 534). Under a design-based framework, the sampling distribution of an estimator can be approximated using simulated sampling from a fixed, finite population. The MB inference treats the population as a random sample, thus the population parameters are also random variables. However, in our study we can treat the finite population as a realization of the superpopulation model which is then considered fixed and the inference regards a descriptive parameter (μ_{AGB}) of this particular population (case (a) inference in Särndal et al., 1992, p. 514; Kangas, 2006, p. 40), rather than the mean of the generating model. Moreover, the MB variance estimator

$\hat{v}_{2,3}$ includes a sampling error term, blurring the distinctions between design-based and model-based inference.

For strategies involving systematic sampling, all possible samples were obtained and the complete sampling distributions of the estimators were available, so that we could actually assess the variance of the estimators as well as the expected values of the variance estimates. For SRS_{woR_2} , we run $1.0e^6$ simulations (10,000 first-phase samples with 100 second-phase samples each), ensuring that the errors due to the Monte Carlo simulations were very small.

3.4.1 The validity of variance estimation

The analytical standard errors (SE_{an}) were calculated as the square root of the average variance estimates, and observed standard error (SE_{obs}) was obtained as the standard deviation of the sampling distribution of the $\hat{\mu}_{AGB}$ estimates. In addition, the percentage SE relative to population mean was derived.

It was expected that the SE_{an} produced by the variance estimators would be close to the observed standard error SE_{obs} . In order to test whether the differences between SE_{an} and SE_{obs} were significantly different at a significance level $\alpha = 0.05$, we constructed 95% percentile-based confidence intervals from the sampling distributions of SE_{an} and tested if they covered SE_{obs} .

3.4.2 AGB estimation accuracy

For assessing if the design-based sampling strategies lead to valid inference for $\hat{\mu}_{AGB}$, we examined the coverage rates of 95% z-based confidence intervals (z-CI) built around $\hat{\mu}_{AGB}$. The actual coverage rates of the z-CIs was obtained as the percentage of times the z-CIs

included the population parameter μ_{AGB} during simulated sampling, and it was compared to the nominal 95% coverage. Despite the large second-phase samples, the skewness of the parent population may induce a positive correlation between $\hat{\mu}_{AGB}$ and SE_{an} , which can produce the failure of the confidence interval probability statements (Gregoire & Schabenberger, 1999). Thus, we also examined the from-above and from-below failure rates of the z-CIs, as described by Gregoire & Schabenberger (1999).

Under the MB framework, the $\hat{v}_{2,3}$ variance estimator of Ståhl et al. (2011) targets the variance of $\mu_{AGB} - \hat{\mu}_{AGB}$, where μ_{AGB} is considered a random variable. When the prediction model is unbiased, a sample based $1 - \alpha$ prediction interval for μ_{AGB} can be written as $\hat{\mu}_{AGB} \pm z_{(1-\alpha/2)} \times SE_{an}$. Theoretically, a prediction interval under MB inference is not equivalent to a design-based z-CI because the former seeks to provide a coverage of $\mu_{AGB} - \hat{\mu}_{AGB}$ for approximately $100(1-\alpha)\%$ of the realizations of the finite population vector for a fixed sample (Särndal et al., 1992, p. 534). In our case, the prediction intervals were constructed by using simulated sampling and considering μ_{AGB} as fixed. Finally, we compared the percentage of variance accounted for by each sampling phase by $\hat{v}_{2,1}$ and $\hat{v}_{2,3}$ estimators, and by the corresponding observed Monte Carlo variances.

3.4.3 The relative efficiency of AGB estimation

In this study, we aimed to assess the relative efficiency between strategies involving different designs, estimators and inferential frameworks. Although it is customary to express the accuracy of an estimator as the mean square error (Cochran, 1977, p.15-16; Gregoire & Valentine, 2008, p. 28), we used the root means square error (*RMSE*) because it accounts for potential biases as well. The *RMSE* values were calculated

as $RMSE = \sqrt{Bias(\hat{\mu}_{AGB})^2 + SE_{obs}^2}$, where $Bias(\hat{\mu}_{AGB}) = \mu_{AGB} - \bar{\hat{\mu}}_{AGB}$ is the bias of the estimator calculated as the difference between population mean and the average $\hat{\mu}_{AGB}$ estimates. Hence, the gain in accuracy (*eff*) obtained using two-phase systematic sampling (SYS_2) with regression estimator compared to systematic ground survey (SYS_1) can be expressed as the proportion between the observed *RMSE*s obtained under each sampling strategy:

$$eff_{SYS_1}^{SYS_2} = \frac{RMSE_{SYS_2}(\hat{\mu}_{AGB})}{RMSE_{SYS_1}(\hat{\mu}_{AGB})} \quad eq(16)$$

The efficiency of the SYS_2 design relative to $SRSwoR_2$ was expressed using the design effect (Cochran, 1977, p. 85; Särndal et al. 1992, p. 492) calculated as the ratio of the observed variances under the two designs:

$$deff(SYS_2, \hat{\mu}_{AGB}) = \frac{Var_{obs, SYS_2}(\hat{\mu}_{AGB})}{Var_{obs, SRSwoR_2}(\hat{\mu}_{AGB})} \quad eq(17)$$

4. Results

The average sample size under SYS_1 was 1309 population elements. The ALS samples under SYS_2 and $SRSwoR_2$ contained on average 52 PSUs, and the ground samples had on average 1295 pixels. The sampling distributions of the AGB estimators and of their standard error biases are presented in Figure 6 to 11. The Monte Carlo estimates for bias and standard errors of $\hat{\mu}_{AGB}$ are presented in Table 4. The coverage and failure rates for the z-CIs for μ_{AGB} and the percentile-based CI derived from the analytical standard error estimates are shown in Table 5. Details are presented in the following sections.

Table 4 Results obtained running simulated sampling following one- and two phase designs under design and model-based inference. The percentages relative to population mean (Mg ha^{-1}) are given in the brackets.

Sampling strategies				Inference	Estimates				Percentage SE by sampling phase			
Design	Estimators		\hat{v}		analytical		observed		SE_{an}		SE_{obs}	
	mean	variance			SE_{an}	Bias	SE_{obs}	RMSE _{obs}	phase 1	phase 2	phase 1	phase 2
One stage	SYS ₁	HT	$\hat{v}_{1.1}$	DBHT	1.5296	-8.8e ⁻¹³	1.4477	1.4477	x	x	x	x
					(3.17)	(0.00)	(3.00)	(3.00)				
Two stages	SRSwoR ₂	GREG	$\hat{v}_{2.1}$	DBMA	2.8032				80	20		
					(5.82)	-8.5e ⁻³	2.7859	2.7859			80	20
			$\hat{v}_{2.2}$		2.7925	(-0.02)	(5.78)	(5.78)	x	x		
					(5.80)							
		REG	$\hat{v}_{2.3}$	MB	2.8562	-87.6e ⁻³	2.5868	2.5883	66	34	78	22
					(5.93)	(-0.18)	(5.37)	(5.37)				
Two stages	SYS ₂	GREG	$\hat{v}_{2.1}$	DBMA	2.7986				80	20		
					(5.81)	6.9e ⁻³	0.5989	0.5990			6	94
		$\hat{v}_{2.2}$			2.8161	(0.01)	(1.24)	(1.24)	x	x		
					(5.84)							
	REG	$\hat{v}_{2.3}$	MB	2.8754	-49.7e ⁻³	0.6395	0.6414	67	33	12	88	
				(5.97)	(-0.10)	(1.32)	(1.33)					

Table 5 Coverage and failure rates (%) of z-based CI for μ_{AGB} following one- and two-phase designs, under design and model-based inference.

Sampling strategy				Coverage and failure rates (%) of z-CIs for population mean			Empirical p-CIs for SE_{obs} (Mg ha^{-1})			
sign	Estimators		\hat{v}	below	cover	above	$p_{2.5}$	$p_{97.5}$	SE_{obs}	
	mean	variance								
One phase	SYS ₁	HT	$\hat{v}_{1.1}$	2.32	96.1	1.58	1.4321	1.6270	1.4477	
		SRSwoR ₂	GREG	$\hat{v}_{2.1}$	2.43	94.46	3.11	2.2385	3.3376	2.7859
				$\hat{v}_{2.2}$	2.32	94.69	2.99	2.3620	3.1924	
Two phase	SYS ₂	REG	$\hat{v}_{2.3}$	1.95	95.22	2.83	2.4502	3.2312	2.5868	
		GREG	$\hat{v}_{2.1}$	0.00	100.00	0.00	2.4100	3.1849	0.5989	
			$\hat{v}_{2.2}$	0.00	100.00	0.00	2.6336	2.9827		
			$\hat{v}_{2.3}$	0.00	100.00	0.00	2.7396	2.9873		0.6395

4.1 Ground survey using systematic sampling with HT estimator

Under SYS_1 , the bias of the HT estimator was virtually zero (Table 4). The actual coverage rate of the z-CIs for μ_{AGB} (Table 5) was 96.1%, and the percentage failures from below (2.32%) exceeded the failure rates from above (1.58%) inasmuch as the μ_{AGB} and the standard error estimates were highly correlated (Pearson's $r = 0.72$). The observed and estimated

standard errors (Table 5) were close ($SE_{obs} = 1.4477 \text{ Mg ha}^{-1}$ and $SE_{an} = 1.5296 \text{ Mg ha}^{-1}$), differing by approximately 5.7 %. Still, the SE_{an} and SE_{obs} were not significantly different, the 95% empirical p-CI constructed from the sampling distribution of the SE_{an} (Figure 6) covering the SE_{obs} value (Table 5).

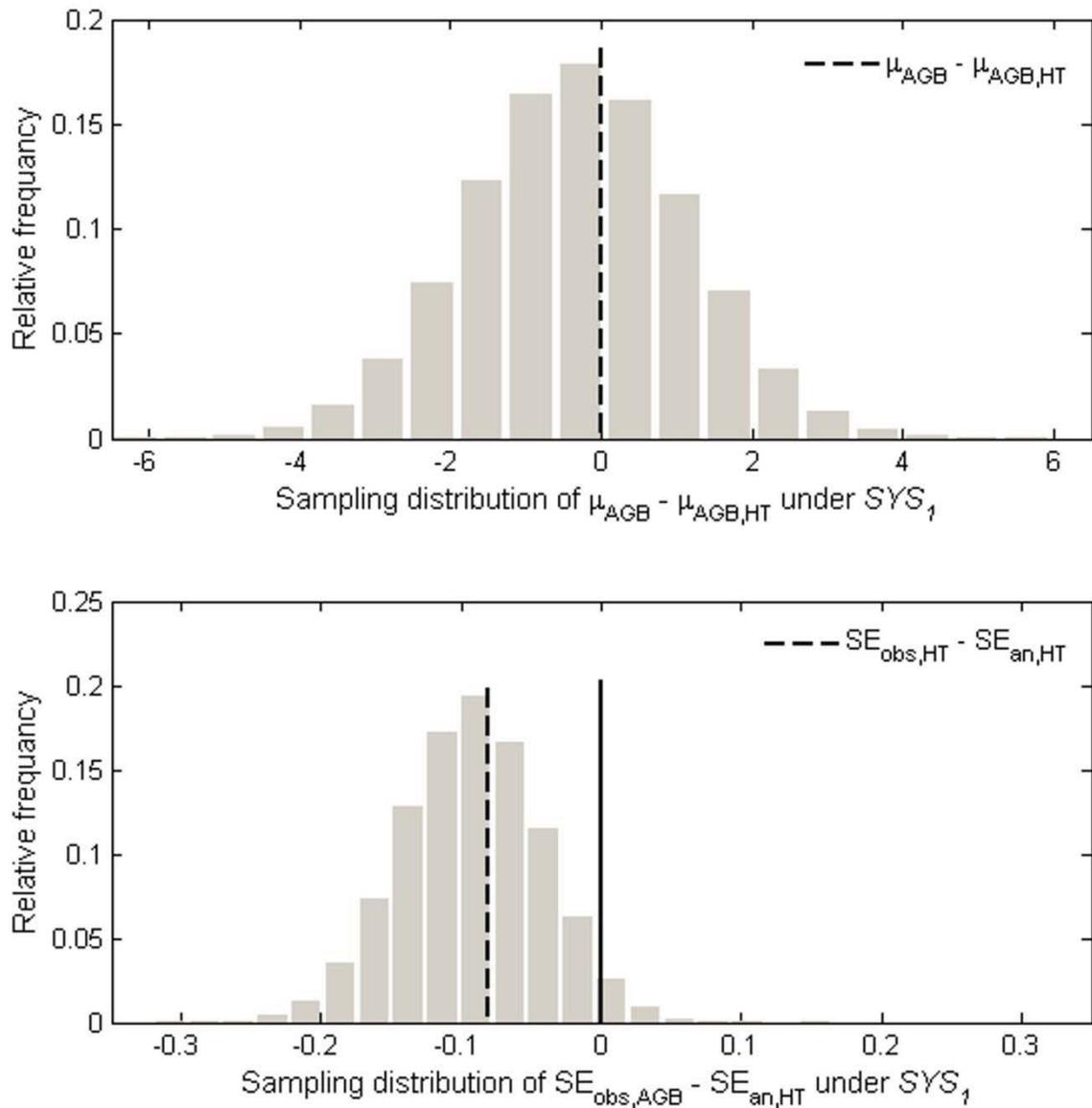


Figure 6 Sampling distributions for HT estimator for AGB and of its standard errors under SYS1. The dashed lines indicate the biases of the estimators.

4.2 Two-phase systematic sampling with regression estimators

The sampling distributions of GREG and REG estimators under SYS_2 are presented in Figure 7. The regression estimators over- and under estimated μ_{AGB} with approximately $6.9e^{-3} \text{ Mg ha}^{-1}$ (0.014%) and $-49.7e^{-3} \text{ Mg ha}^{-1}$ (0.101%) for GREG and REG, respectively (Table 4). The sampling distribution of the deviation between observed and analytical standard errors ($SE_{obs} - SE_{an}$) under SYS_2 is presented in Figure 8.

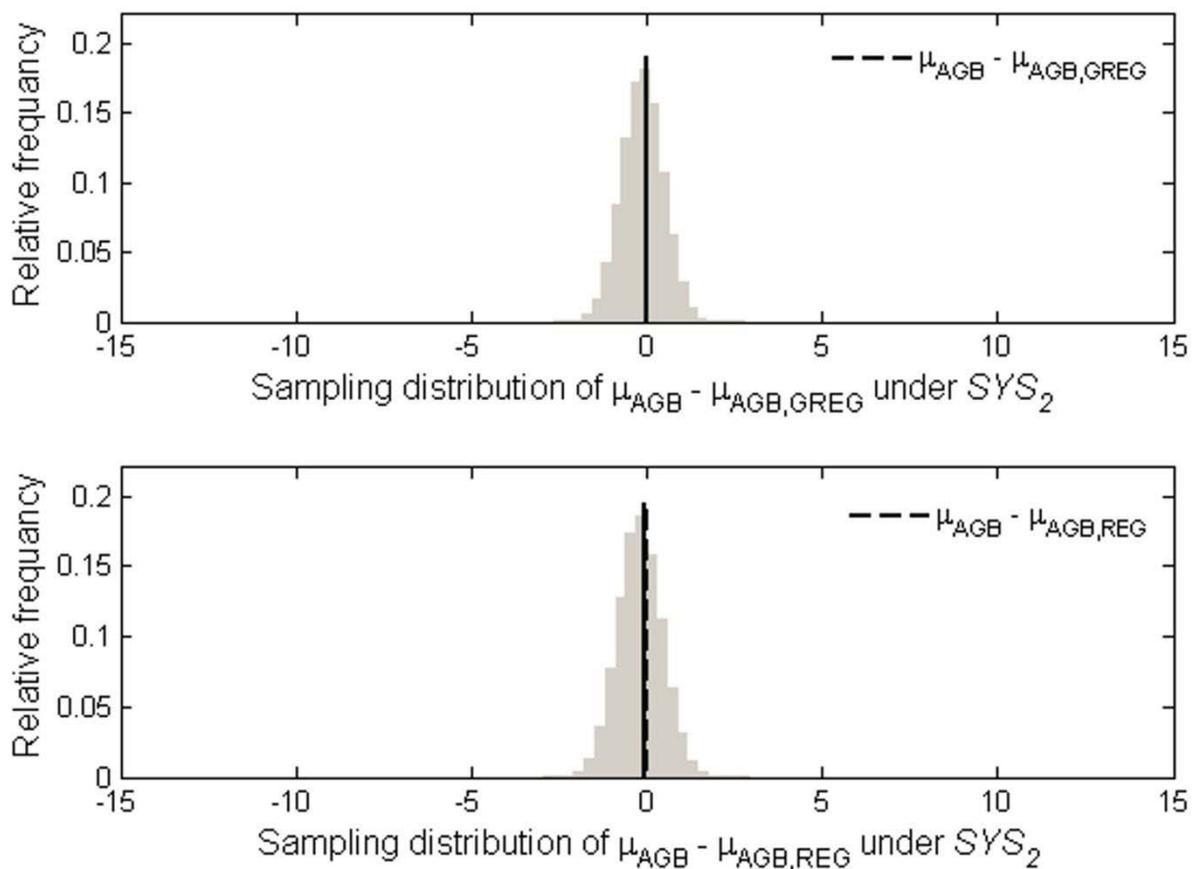


Figure 7 Sampling distributions for GREG and REG estimators for AGB under SYS_2 . The vertical solid lines represent the reference values of zero bias, and the dashed lines indicate the biases of the estimators.

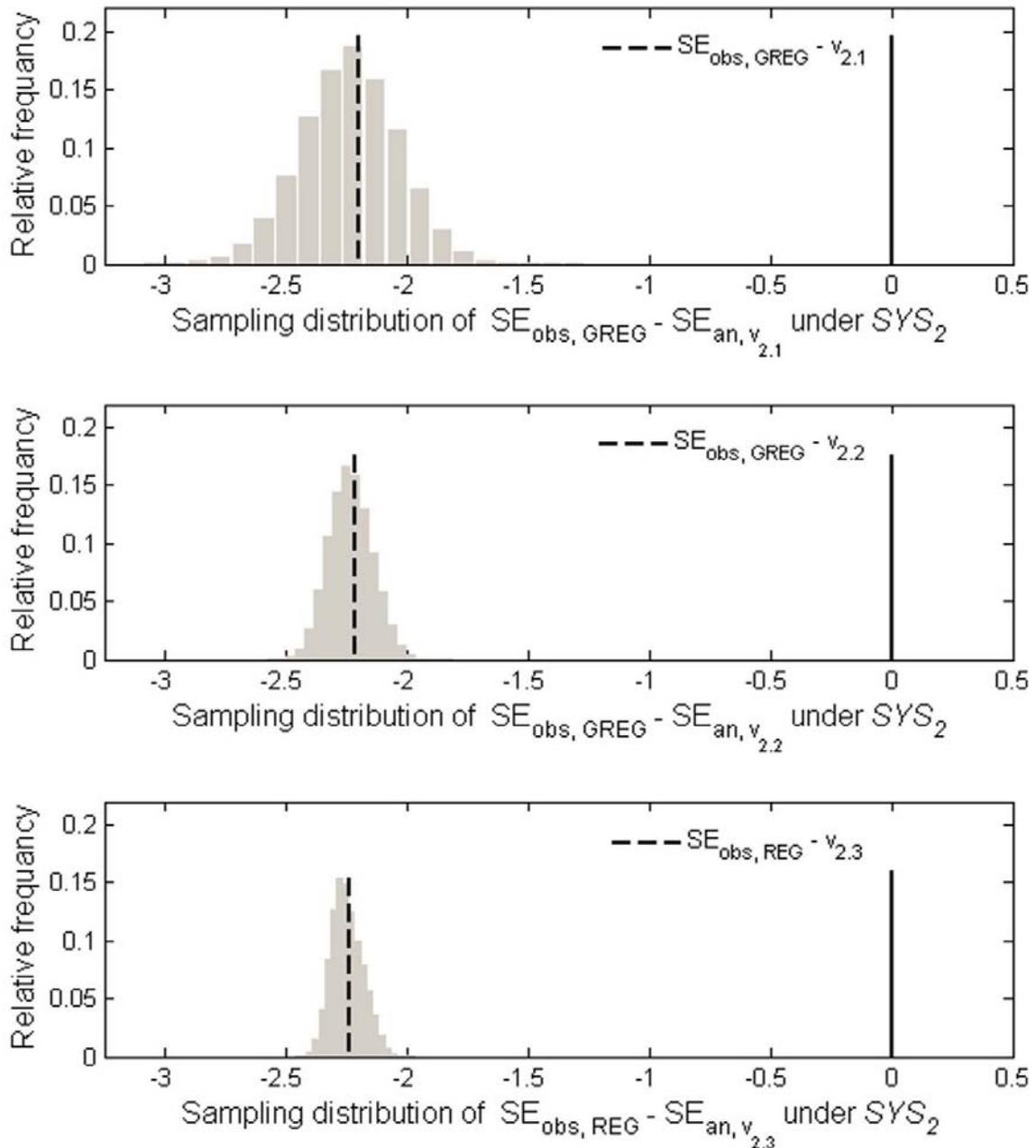


Figure 8 Sampling distributions for the estimated SE of GREG and REG estimators for AGB under SYS2. The vertical solid lines represent the reference values of zero bias, and the dashed lines indicate the biases of the estimators.

The analytical SE estimates (SE_{an}) were approximately 4.7 times larger than SE_{obs} (Table 4). For GREG, the SE_{an} obtained using the $\hat{v}_{2.1}$ estimator attained a value of 2.7986 Mg ha^{-1} (5.81%), while SE_{obs} was 0.5989 Mg ha^{-1} (1.24%). The $\hat{v}_{2.2}$ estimator produced a slightly higher SE_{an} (2.8161 Mg ha^{-1}) comparing to $\hat{v}_{2.1}$. Compared to GREG, the analytical standard error of REG (the $\hat{v}_{2.3}$ estimator) was slightly higher (2.8754 Mg ha^{-1}), but the

GREG and REG estimators performed close in terms of $RMSE$ (1.24% and 1.33%, respectively). However, the standard error estimates produced by $\hat{v}_{2.1}$, $\hat{v}_{2.2}$ and $\hat{v}_{2.3}$ were significantly different from the SE_{obs} , the empirical p-CIs failing to cover the SE_{obs} for any of the variance estimators (Table 5).

For GREG and REG estimators, the error component due to the first-phase sample represented approximately 6% of the total SE_{obs} , while the second-phase sample accounted for 94% (Table 4). The largest components of SE_{an} for $\hat{v}_{2.1}$ and $\hat{v}_{2.3}$ were due to the first-phase samples (80 % and 67%, respectively), while the second-phase sampling error ($\hat{v}_{2.1}$) and the model-error component ($\hat{v}_{2.3}$) accounted for 20% and 33%. The coverage of the z-based CIs for μ_{AGB} was 100% (Table 5), and a positive linear correlations were observed between SE_{an} and $\hat{\mu}_{AGB}$ ($r = 0.22$ for $\hat{v}_{2.1}$, 0.52 for $\hat{v}_{2.2}$ and 0.72 for $\hat{v}_{2.3}$).

4.3 Two-phase $SRSwoR$ with regression estimator

Both REG and GREG estimators slightly overestimated μ_{AGB} (Table 4). Although small, the bias of REG under $SRSwoR_2$ was approximately ten times higher than the bias under GREG ($-8.5e^{-3}$ Mg ha⁻¹ and $88.6e^{-3}$ Mg ha⁻¹, respectively). The sampling distribution of GREG and REG estimators under $SRSwoR_2$ are presented in Figure 9.

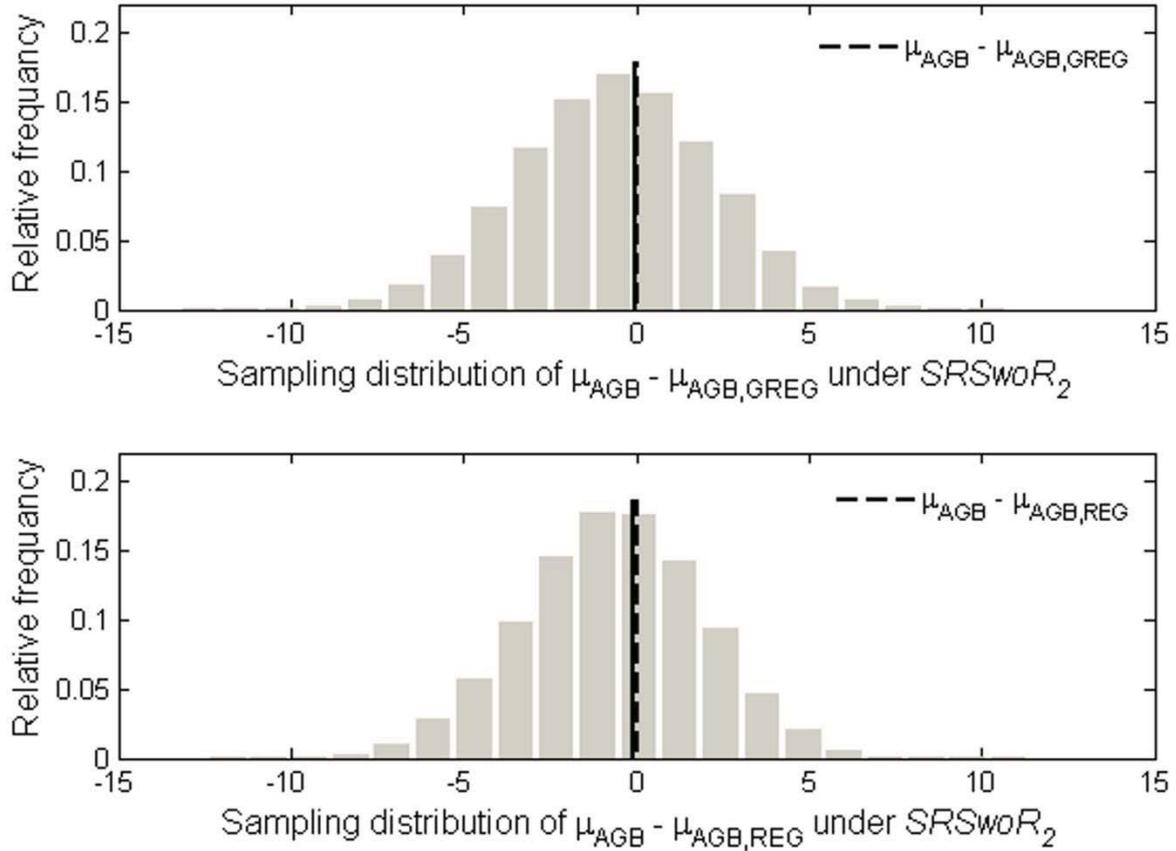


Figure 9 Sampling distributions for GREG and REG estimators for AGB under $SRSWOR_2$. The vertical solid lines represent the reference values of zero bias, and the dashed lines indicate the biases of the estimators.

The SE_{an} for GREG under $SRSWOR_2$ obtained using $\hat{v}_{2.1}$ and $\hat{v}_{2.2}$ estimators (2.8032 $Mg\ ha^{-1}$ and 2.7925 $Mg\ ha^{-1}$) were close to the SE_{obs} (2.7859 $Mg\ ha^{-1}$), the differences between the analytical and observed estimates being below 1.0% (Table 4). The analytical standard error estimate produced by the $\hat{v}_{2.3}$ estimator (2.8562 $Mg\ ha^{-1}$) overestimated SE_{obs} (2.5868 $Mg\ ha^{-1}$) by nearly 10% (Table 4). However, the empirical 95% p-CIs obtained from the sampling distribution of $\hat{v}_{2.1}$, $\hat{v}_{2.2}$ and $\hat{v}_{2.3}$ estimators under $SRSWOR_2$ covered the Monte Carlo estimate for SE_{obs} , indicating that the differences were not significant. The $RMSE$ of GREG was approximately 8% than for REG (2.7859 $Mg\ ha^{-1}$ and 2.5883 $Mg\ ha^{-1}$). The sampling distribution of the $SE_{obs} - SE_{an}$ ($Mg\ ha^{-1}$) under $SRSWOR_2$ is presented in Figure 10.

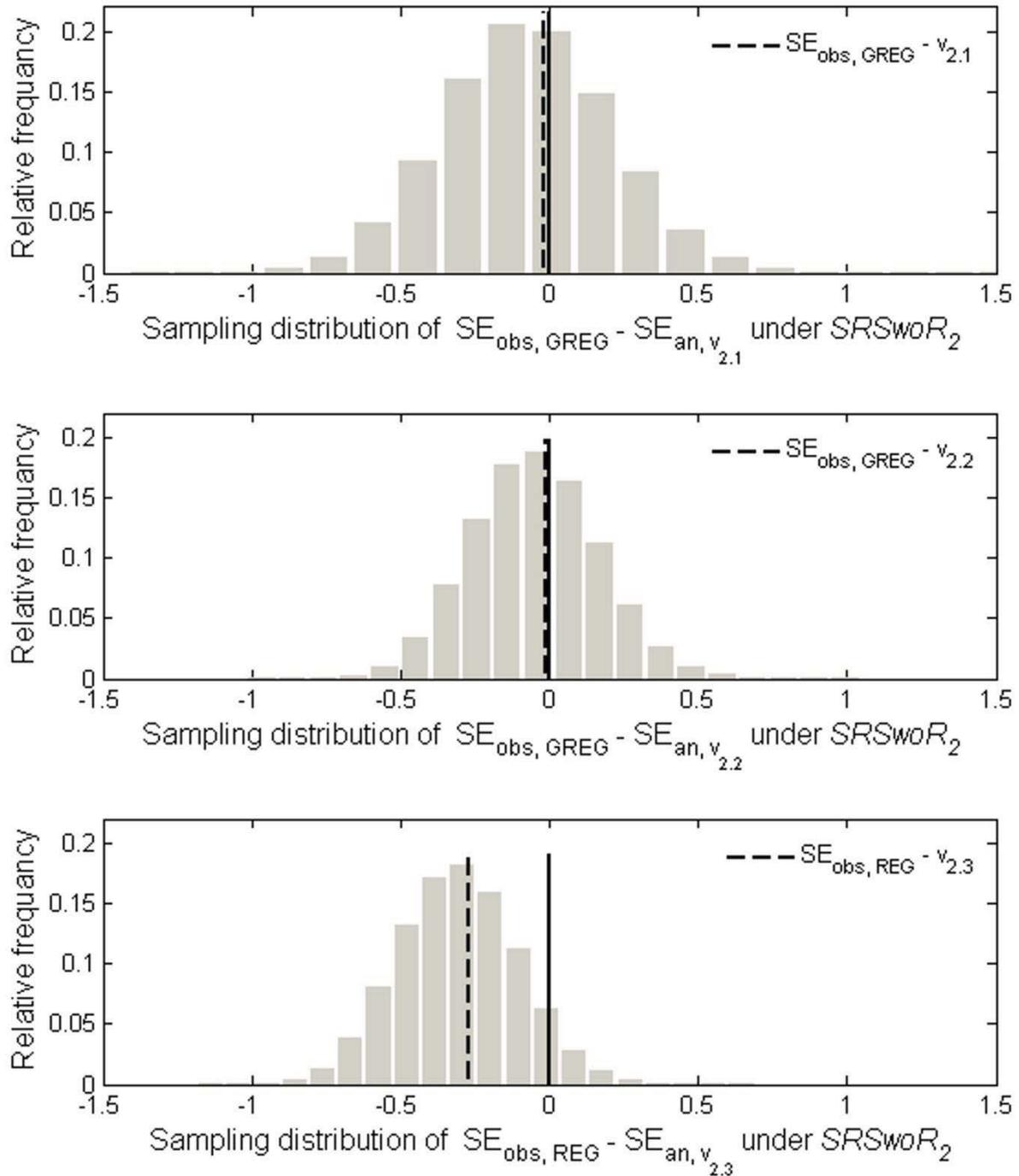


Figure 10 Sampling distributions for the estimated SE of GREG and REG estimators for AGB under SRSWOR2. The vertical solid lines represent the reference values of zero bias, and the dashed lines indicate the biases of the estimators.

For both GREG and REG estimators, the largest proportion of the observed sampling error (SE_{obs}) came from the first-phase sample (80% and 78%, respectively) while the contributions of the second-phase samples and model-error component were around 20%

and 22% (Table 4). The analytical results obtained using the $\hat{v}_{2.1}$ and $\hat{v}_{2.3}$ estimators indicated that the first-phase sampling error accounted for 80% and 66% of SE_{an} , respectively, while the second-phase samples and the model-error component contributed with 20% and 34%, respectively (Table 4).

The coverage rates of the z-CIs for μ_{AGB} obtained using $\hat{v}_{2.1}$ and $\hat{v}_{2.2}$ estimators (94.46% and 94.69%, respectively) were slightly below the nominal rate of 95% (Table 5), while the coverage rate provided by the $\hat{v}_{2.3}$ estimator (95.22%) matched well with the 95% nominal rate. The failure rates were asymmetric having an opposite behaviour comparing to the ground-sample case, the from-above failure rate being higher. However, the z-CIs for GREG were slightly more balanced. A weak but significant negative linear correlation was observed between SE_{an} and $\hat{\mu}_{AGB}$ ($r = -0.07$ for $\hat{v}_{2.1}$, -0.10 for $\hat{v}_{2.2}$ and 0.24 for $\hat{v}_{2.3}$).

4.4 Estimation accuracy and the design effect

The accuracy (in terms of $RMSE$) of ALS-based systematic sampling was 59% higher compared to the field-based systematic survey (equation 16). The design effect of SYS_2 relative to SRS_{woR_2} (equation 17) was approximately 0.046 for both GREG and REG estimators, which corresponds to nearly five times smaller observed standard errors when using systematic sampling.

5. Discussion

This study described a simulation approach for assessing the performance of various estimators and to compare the accuracy of complex designs. Using an empirical dataset containing field observations, ALS measurements, satellite imagery and cartographic

products, an artificial population was created and served as ‘ground truth’ for performing simulated sampling.

The empirical validation of the variance estimators for GREG and REG indicated that all the three estimators we investigated performed well when their underlying assumptions were fulfilled. Indeed, under $SRSwoR_2$ the analytical and observed standard errors for GREG and for REG were not significantly different. Violating these assumptions however (as in the case of SYS_2), still allows valid inference for both GREG and REG estimators, but the analytic variance estimates were inflated almost five times (e.g., 2.7986/0.5989 from Table 4). The reason for obtaining better observed precision under SYS_2 is that the systematic design produced more balanced samples compared to $SRSwoR_2$, thus the variability between different samples was much reduced. Compared to the field-based survey, the $RMSE$ of the ALS-aided inventories was improved by approximately 59% (e.g., 1-1.4477/0.5990 from Table 4), while the analytical results indicated that the accuracy of the ALS-based SE estimates was deteriorating approximately 1.8 times (e.g., 2.7986/1.5296 from Table 4).

Under $SRSwoR_2$, the SE_{an} and SE_{obs} produced consistent results in the sense that both accounted the first-phase sample as the largest source of error. This can also explain the good performance of the $\hat{v}_{2,2}$ estimator in our study. Under SYS_2 , the opposite situation was observed, namely the $\hat{v}_{2,1}$ and $\hat{v}_{2,3}$ estimators continuing to assign much higher proportion of the standard error to first-phase sample, while in reality the highest amount of uncertainty was related to the second-phase sample.

Gregoire et al. (2011) reported a higher precision of the ground-based survey compared to DBMA estimation, while Ståhl et al. (2011) found the model-based standard errors for ALS-based surveys slightly superior to the traditional field survey. For equal number of sample plots, Gobakken et al. (*in submission*) obtained comparable estimated accuracies of the ground-based and ALS-based surveys, and a better accuracy of the ground-

based inventory when increasing the field sample size. Their results were obtained using a post-stratified estimation by cover-classes, different prediction models and different second-phase sample sizes. Thus, a direct comparison with the results reported in the current study is difficult. Still, in the light of their results, it would have been difficult to favour the ALS-based inventories because the variance overestimation couldn't be detected. Using our simulation approach can reveal such problems and may eventually offer the tools for testing alternative solutions.

The failure rates of the z-CIs were asymmetric for all cases, as it would be expected when sampling from a skewed population (Gregoire & Schabenberger, 1999). However, the asymmetry might become important especially for ground-based inventory if one-sided CIs are required. A possible solution for correcting the asymmetric failure rate might be, in this case, constructing asymmetric CIs (Gregoire & Schabenberger, 1999). While the AGB estimates were positively correlated to their estimated standard errors under systematic sampling, this correlation was negative under SRS_{woR_2} for reasons still unclear.

Under SYS_2 , the estimated standard errors due to the first-phase sample dominated the total sampling error- for all estimators. This suggests accommodating the variance estimators to systematic sampling by applying alternative methods like successive differencing (Cochran, 1997) between PSUs, as suggested by Ståhl et al. (2011) and Gregoire et al. (2011).

We did not find any significant differences between the performance of DBMA and MB estimators in terms of bias. Using large second-phase samples for model building, avoiding the retransformation bias and automatic variable screening methods can explain the small bias of the MB estimator (REG). It might be reasonable thus to use the sampling simulator also for evaluating model building strategies, especially under the MB framework

which depend on the quality of the prediction models, although full protection against model biases cannot be obtained.

The systematic sampling with restricted selection for clusters with unequal sizes might not be the optimal choice in terms of precision (Valentine et al., 2009), but allowed to derive the full sampling distribution of our estimators. Using unrestricted systematic sampling with selection proportional with the number of NFI plots within each ALS strips might be a more efficient design and it merits further investigation.

Using copulae makes possible creating a large artificial population even when the empirical dataset is relatively limited with respect to number of observations. However, the copula-based approach requires an empirical sample which describes well the underlying multivariate population distribution. Our copula modelling approach can be affected by the relatively high dimensionality of the dataset, especially when the sample size is small. On the other hand, dimensionality reduction would reduce the complexity of the interactions existing in the empirical data. As long as the simulation results are not directly inferred to the real population, we consider that the differences between the real and the artificial population can be neglected. Recently, hierarchical copulae modelling approaches more suitable for high-dimensional data (Aas et al., 2009) were made available for users by the ‘CDVine’ package (Schepsmeier & Brechmann, 2011) of the open-source statistical environment R (R Development Core Team, 2011). This new methods will be further investigated for improving the quality of the artificial population.

Another possible drawback of the methodology used for creating the population is that the nearest neighbour imputations alter the short-range autocorrelation (Barth et al., 2009; McRoberts, 2009). This may affect the assessment of multistage cluster sampling strategies and also analyses involving assessment of various field plot configurations (e.g. shape, size, orientation). This issue might be alleviated to some extent by employing computationally

intensive methods to obtain spatially consistent nearest neighbor imputations (Barth et al., 2009). However, we do not believe that such effects influenced our results since the SSUs were displaced in a rather coarse grid (6x3km).

Moreover, when performing simulation studies, the flight parameters and ALS settings cannot vary, since the population is based on an empirical dataset with fixed settings. Still, it still should be possible to thin the original LiDAR data and then to build a new artificial populations for simulating the effect of using a higher flight height or a lower pulse density.

Another shortcoming of our approach is that the artificial population depends on the empirical observations gathered using some sampling design, which means that the results pertain only to the sampled population and should only be generalized with caution.

It is too soon to assert that the strip sampling approach using ALS is a feasible solution for seasonal or annual large area timber resource, biomass, and carbon inventories. We did not include the inventory costs in our analysis and we did not have a well defined precision target for the survey. However, the simulation approach can be used to solve allocation problems specific to complex surveys and the technique might be employed as a decision-support tool for planning large-area surveys.

Acknowledgments

This research has been funded by the Research Council of Norway (project no. 166482/i10, no. 184636/S30 and no. 180645). The authors would like to thank the Norwegian National Forest Inventory for collection of field data and Blom Geomatics, Norway, for collection and processing of the airborne laser scanner data.

References

- Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44 , 182-198.
- Andersen, H.-E. & Breidenbach, J. (2007). Statistical properties of mean stand biomass estimators in a LiDAR-based double sampling forest survey design. *International Archives of Photogrammetry and Remote Sensing*, Volume XXXVI, Part 3/W52. Espoo, Finland, September 12-14, 2007.
- Andersen, H.-E., Barrett, T., Winterberger, K., Strunk, J., & Temesgen, H. (2009). Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using airborne lidar and field plot data in a model-assisted sampling design. *Proceedings of the IUFRO Division 4 Conference: "Extending Forest Inventory and Monitoring over Space and Time,"* Quebec City, Canada, May 19-22, 2009.
- Andersen, H.-E., Strunk, J., & Temesgen, H. (2011). Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the upper Tanana valley of interior Alaska. *Western Journal of Applied Forestry*, 26, 157-164.
- Anon. (1999). Pinnacle user's manual. San Jose, CA: Javad Positioning Systems.
- Anon. (2004). Norwegian Forest and Landscape Institute (in Norwegian) (http://www.skogoglandskap.no/artikler/2007/Landsskogdata_enkel_tabell) (last accessed 16-June-2011).
- Anon. (2011a). MATLAB Online Help (<http://www.mathworks.com/help/toolbox/stats/>) (last accessed 11-July-2011).
- Anon. (2011b). ArcGIS Desktop: Release 9.3. Redlands, CA: Environmental Systems Research Institute (<http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html>) (last accessed 11-July-2011).

- Asner G.P, Hughes R.F, Varga T.A, Knapp D.E, & Kennedy-Bowdoin T. (2009). Environmental and biotic controls over aboveground biomass throughout a rainforest *Ecosystems*, 12, 261–78.
- Barth, A., Wallerman, J. & Ståhl, G. (2009). Spatially consistent nearest neighbour imputation of forest stand data. *Remote Sensing of Environment*, 113, 546-553.
- Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., & Kimes, D.S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, 112, 3876-3890.
- Bjørndal, I. & Bjørkelo, K. (2006). AR5 klassifikasjonssystem. Klassifikasjon av arealressurser. [*AR5 classification system. Classification of areal resources*]. Ås, Norway: Norwegian Forest and Landscape Institute. (In Norwegian with English summary).
- Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G. & Roncalli, T. (2000). Copulas for Finance: A Reading Guide and Some Applications (Working Paper). Groupe de Recherche Operationnelle, Credit Lyonnais.
- Chavez, P.S. (Jr.) (1996). Image-based atmospheric corrections- revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62, 1025-1036.
- Cochran, W.G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.
- Corona, P., Fattorini, L. (2008). Area-based lidar-assisted estimation of forest standing volume. *Canadian Journal of Forest Science*, 38, 2911-2916.
- Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography*, 12, 147-159.
- Draper, N.R. & Smith, H. (1998). *Applied regression analysis* (3rd ed). New York: Wiley.

- Ene, L., Næsset, E., & Gobakken, T. (2007). Simulating sampling efficiency in airborne laser scanning based forest inventory. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVI, Part 3/W52, Espoo, Finland.
- Eskelson, B.N.I., Madsen L., Hagar J.C., & Temesgen, H. (2011). Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, 57, 212-221.
- Fahey, T.J., Woodbury, P.B., Battles, J.J., Goodale, C.L., Hamburg, S.P., Ollinger, S.V., & Woodall, C.W. (2010). Forest carbon storage: ecology, management, and policy. *Frontiers in Ecology and the Environment*, 8, 245-252.
- Genest, C., Gerber, H.U., Goovaerts M.J., & Laeven, R.J.A. (2009). Editorial to the special issue on modelling and measurement of multivariate risk in insurance and finance. *Insurance: Mathematics and Economics* 44, 143-145.
- Gobakken, T., Næsset, E., & Nelson, R. (2006). Developing regional forest inventory procedures based on scanning LiDAR. *Proceedings of the International Conference Silvilaser 2006* (p. 99-104). Matsuyama, Japan: Japan Society of Forest Planning, Forestry and Forest Products Research Institute and Ehime University.
- 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 (*in press*).
- Gonzales, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., & Palace, M. (2010). Forest carbon densities and uncertainties from LiDAR, QuicBird, and field measurements in California. *Remote Sensing of Environment*, 114, 1561-1575.
- Gregoire, T.G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28, 1429-1447.

- Gregoire, T.G. & Schabenberger, O. (1999). Sampling-skewed biological populations: behaviour of confidence intervals for the population total. *Ecology*, 80, 1056-1065.
- Gregoire, T.G. & Valentine, H. (2008). Sampling strategies for natural resources and the environment. Boca Raton, Fla.: Chapman & Hall/CRC.
- Gregoire, T.G., Lin, Q.F., Boudreau, J., & Nelson, R. (2008). Regression estimation following the square-root transformation of the response. *Forest Science*, 54, 597-606.
- Gregoire, T.G., Ståhl, G., Næsset, E., Gobakken, T., & Nelson, R. (2011). Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41, 83-95.
- IPCC (2000). Land-use, land-use change, and forestry: A special report of the IPCC. Cambridge and New York: Cambridge University Press.
- IPCC (2006) Good practice Guidance for Land Use, Land-Use Change and Forestry. (<http://www.ipcc-nggip.iges.or.jp/public/gpplulucf/gpplulucf.html>) (last accessed 17-July-2011).
- Kangas, A. (2006). Model-based inference. In A. Kangas, & M. Maltamo, (Eds.), *Forest Inventory: Methodology and Applications* (pp. 39-52). Dordrecht: Springer.
- Kershaw, J.A. (Jr.), Richards, E.W., McCarter, J.B., & Oborn, S. (2010). Spatially correlated forest structures: a simulation approach using copulas. *Computers and Electronics in Agriculture*, 74, 120-128.
- Kojadinovic, I. & Yan, J. (2010). Modelling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34. (<http://www.jstatsoft.org/v34/i09/paper>) (last accessed 29-June-2011)
- Koukal, T., Suppan, F., & Schneider, W. (2007). The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. *Remote Sensing of Environment*, 110, 431-437.

- Lawrence M., McRoberts, R.E., Tomppo E., Gschwantner, T., & Gabler, K. (2010). Comparison of National Forest Inventories. In Tomppo, E., M., Gschwantner, T., Lawrence, M. and McRoberts, R.E. (Eds.), *National Forest Inventories- Pathways for Common Reporting* (pp. 19-32). Dordrecht: Springer Science+Business Media B.V.
- Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., & Brower, S.T. (2002). Lidar remote sensing of aboveground biomass in three biomes. *Global Ecology and Biogeography*, 11, 393-400.
- Li Y., Andersen, H-E. & McGaughey, R. (2008). A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *Western Journal of Applied Forestry*, 23, 223-231.
- Marcell, W., Eriksson, M., & Popescu, S. (2009). Systematic sampling of scanning lidar swaths. In Popescu, S., Nelson, R., Kaiguang, Z. and Neusenschwander, A. (Ed.), *Proceedings of SILVILASER 2009* (p. 184-192), 14-16 October 2009, Texas A&M University, College Station, USA.
- Marklund, L.G. (1988). Biomass functions for pine, spruce and birch in Sweden. Umeå: Swedish University of Agricultural Sciences, Department of Forest Survey (In Swedish).
- McCulloch, C.E. & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London, UK: Chapman and Hall.
- McRoberts, R. (2009). Diagnostic tools for nearest neighbour techniques when used with satellite imagery. *Remote Sensing of Environment*, 113, 489-499.
- Miina, J. & Heinonen, J. (2008) Stochastic Simulation of Forest Regeneration Establishment Using a Multilevel Multivariate Model. *Forest Science* 54, 206-219.
- Næsset, E. (2002). Predicting forest stand characteristics with airborne laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88-99.

- Næsset, E. (2004). Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19, 164-179.
- Næsset, E. (2005). Towards a laser-scanner based biomass monitoring system. In K. Hobbelstad (Ed.), *Forest Inventory and Planning in the Nordic Countries. Proceedings of the SNS Meeting* (p. 117-119). Sjusjøen, Norway: Report 09/05 of the Norwegian Institute of Land Inventory.
- Næsset, E. (2007). Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research*, 22, 433-422.
- Næsset, E. (2009). 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 Sensing of Environment*, 113, 148-159.
- Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing*, 32, 473-501.
- Næsset, E., Gobakken, T., Holmgren, J., Hyypä, J., Hyypä H., Maltamo, M., Nilsson M., Olsson, H., Persson, Å., & Söderman, U. (2004). Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, 19, 482-499.
- Næsset, E. & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112, 3079-3090.
- Næsset, E., Gobakken, T., & Nelson, R. (2009). Sampling and mapping forest volume and biomass using airborne LIDARs. *Proceedings of the Eight Annual Forest Inventory and Analysis Symposium* (p. 297-301), October 16-19, 2006, Monterey, CA, USA. General Technical Report WO-79, United States Department of Agriculture, Forest Service, Washington DC.

- Nelsen, R.B. (2009). An introduction to copulas, (2nd ed.). New York: Springer.
- Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, 24, 247–267.
- Nelson, R., Parker, G. & Hom, M. (2003). A portable airborne laser system for forest inventories. *Photogrammetric Engineering & Remote Sensing*, 69, 267-273.
- Nelson, R., Short, A., & Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LiDAR. *Scandinavian Journal of Forest Research*, 19, 500-511.
- Nelson, R., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M. L., Campbell, R., et al. (2007) Investigating RaDAR–LiDAR synergy in a North Carolina pine forest. *Remote Sensing of Environment*, 110, 98–108.
- Nelson, R., Næsset, E., Gobakken, T., Ståhl G., & Gregoire, T.G. (2008). Regional forest inventory using an airborne profiling LiDAR. *Journal of Forest Planning*, 13, 287-294.
- Nelson, R., Boudreau, J., Gregoire, T.G, Margolis, H., Næsset, E., Gobakken, T., & Ståhl, G. (2009). Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, 39, 862-881.
- Nelson, R., Gobakken, T., Næsset, E., Gregoire, T.G., Ståhl, G., Holm, S., & Flewelling, J., (2011). Lidar Sampling - Using an Airborne Profiler to Estimate Forest Biomass in Hedmark County, Norway. Accepted for publication, *Remote Sensing of Environment*.
- Ørka, H.O., Næsset, E., & Bollandsås, O.M. (2010). Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. *Remote Sensing of Environment*, 114, 1445-1461.

- Ørka, H.O., Wulder, M.A., Gobakken, T., & Næsset, E. (2011). Subalpine zone delineation using LiDAR and Landsat imagery (Accepted for publication, *Remote Sensing of Environment*).
- Parker, R.C. & D.L. Evans (2004). An application of lidar in a double-sample forest inventory. *Southern Journal of Applied Forestry*, 29, 40-47.
- Popescu, S.C. & Wynne, R.H. (2004). Seeing the trees in the forest: using LiDAR and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing*, 70, 589-604.
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Schepsmeier, U. & Brechmann, E.C. (2011). CDVine: Statistical inference of C- and D-vine copulas. R package version 1.1 (<http://CRAN.R-project.org/package=CDVine>) (last accessed 27-Aug-2011).
- Schreuder, H.T., Gregoire, T.G., & Weyer, J.P. (2001). For what applications can probability and non-probability sampling be used?. *Environmental Monitoring and Assessment*, 66, 281-291.
- Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., & Nelson, R. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest research*, 41, 96-107.
- Stephens, P.R., Kimberly M.O, Beets, P.N., Paul, T.S.H., Searles, N., Bell, A., Brack, C., & Broadley, J., Airborne scanning LiDAR in a double sampling forest carbon inventory. *Remote Sensing of Environment* (2011), doi: 10.1016/j.rse.2011.10.009.
- Streck, C. & Scholz, S.M. (2006). The role of forests in global climate change: Whence we come and where we go. *International affairs*, 82, 861-897.

- Särndal, C.E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.
- Tomppo, E., Haakana, M., Katila, M., & Peräsaari, J. (2008). Multi-Source National Forest Inventory. Methods and Applications. Dordrecht : Springer Science + Business Media B.V.
- Tomter, S.M., Hysten, G., & Nilsen J-E. (2010). National Forest Inventory Reports: Norway. In Tomppo, E., M., Gschwantner, T., Lawrence, M. and McRoberts, R.E. (Eds.), National Forest Inventories- Pathways for Common Reporting (pp. 411-424). Dordrecht: Springer.
- UNFCCC (2008). Kyoto protocol reference manual on accounting of emissions and assigned amounts. (http://unfccc.int/kyoto_protocol/items/3145.php) (last accessed 27 May 2011).
- Valentine, H.T., Affleck, D.L.R., & Gregoire, T.G. (2009). Systematic sampling of discrete and continuous populations: sample selection and the choice of estimator. *Canadian Journal of Forest Research*, 39, 1061-1068.
- Zheng, B. & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in medicine*, 19, 1171-1781.
- Wang, M., Rennolls, K., & Shouzheng, T. (2008). Bivariate distribution modelling of tree diameters and heights: dependency modelling using copulas. *Forest Science*, 54, 284-293.
- Wang, M, Upadhyay, A., & Zhang, L. (2010). Trivariate distribution modelling of tree diameter, height, and volume. *Forest Science*, 53, 290-300.

Paper II

A simulation approach for accuracy assessment of two-phase post-stratified estimation in large-area LiDAR biomass surveys

Prepared by:

Liviu Theodor Ene^{1*}, Erik Næsset¹, Terje Gobakken¹, Timothy G. Gregoire², Göran Ståhl³
and Sören Holm³

Address:

1. Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, PO Box 5003, NO-1432 Ås, Norway,

Liviu.Ene@umb.no; Terje.Gobakken@umb.no; Erik.Naesset@umb.no.

2. School of Forestry and Environmental Studies, 360 Prospect St., Yale University, New Haven CT 06511-2189 USA, Timothy.Gregoire@yale.edu.

3. Dept. of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden, Goran.Stahl@slu.se.

*Corresponding author

Liviu Theodor Ene

Email: Liviu.Ene@umb.no

Phone: +47-64965766

Fax: +47-64968890

Abstract

Auxiliary information provided by airborne laser scanners (ALS) is expected to increase the accuracy of biomass estimation in large-scale surveys. Because acquisition of “wall-to-wall” ALS data over large areas is not economically feasible, a systematic sampling approach using ALS as a strip sampling tool was used to supplement a conventional field-based inventory in a large-scale biomass survey in Hedmark County (HC), Norway. For the type of complex designs that had to be applied in this case, properties of estimators (e.g. bias and variance) cannot be determined analytically. In such cases, an alternative is to analyze the statistical properties of the estimators using simulated sampling from an artificial population. Through this approach, estimators with desirable properties can be identified and used for inference in real applications. By combining biomass estimates from Norwegian National Forest Inventory plots in HC, ALS measurements and Landsat 5 TM imagery, an artificial population at the scale of HC was created. Using this artificial population as “ground-truth”, we demonstrate how simulated sampling can be used for assessing the statistical properties of regression estimators and of their variance estimators under two-phase post-stratified systematic sampling (*SYS*) and simple random sampling without replacement (*SRS_{woR}*) designs, considering design- and model-based inferential frameworks. The results were assessed using a purely ground-based systematic design with a Horvitz-Thompson (HT) estimator as benchmark. The real overall precision of the ALS-aided systematic survey was nearly five times overestimated when using the design-based variance estimators developed for *SRS_{woR}*, while under model-based inference the overestimation of the real standard errors was around 40%. Compared to ground-based inventory, the estimated standard errors of the systematic ALS survey doubled while in reality the standard errors were 55% lower. Using successive differences variance estimators greatly improved the precision of the

systematic ALS-aided survey and produced valid 95% confidence intervals under the design-based inference. The most satisfactory results for the ALS-aided survey terms of analytical variances occurred under design-based inference with successive difference variance estimator, closely followed by the model-based estimators. Using simulations, the cost efficiency of the ground based and ALS-aided surveys was assessed by evaluating accuracy against inventory cost for various sampling intensities. The results indicated that the ALS-aided surveys can be a cost-efficient alternative to traditional field inventories.

1. Introduction

With increasing concerns regarding global climate change, accurate estimation of forest carbon pools using remotely sensed auxiliary information has become a very active research field. Forest ecosystems are known to store large amounts of carbon (IPCC, 2000); one of the most dynamic and largest carbon pools is the aboveground biomass (AGB) pool of living trees (Fahey et al., 2010). In Norway, reporting of carbon stock changes from the Land Use, Land Use Change and Forestry (LULUCF) sector (IPCC, 2006) required by the Kyoto Protocol has been undertaken by the Norwegian National Forest Inventory (NFI) (Tomter et al., 2010). In tropical countries, an emerging interest for large-scale carbon inventories is driven by the initiatives to reduce emissions from deforestation and degradation (REDD) (Gibbs et al., 2007; Asner, 2009).

However, terrestrial carbon inventory systems may not be economically feasible for large area surveys, and may not provide reliable local estimates, for instance at sub-regional and administrative unit level or by land-use or cover classes (Fahey et al., 2010; Gonzales et al., 2010). Combining auxiliary information provided by remote sensing systems with terrestrial surveys has the potential to enhance the precision of forest carbon estimation (Gonzales et al., 2010; Næsset et al., 2011). Having the ability to describe the three-dimensional canopy structure, Airborne Light Detection and Ranging (LiDAR) is a promising remote sensing tool for enhancing biomass inventories in various forest types (Lefsky et al., 2002; Drake et al., 2003; Nelson et al., 2004, 2007; Boudreau et al., 2008; Næsset & Gobakken, 2008; Asner et al., 2009; Næsset, 2011; Næsset et al. 2011). However, “wall-to-wall” LiDAR surveys are still expensive for large areas, and economically feasible emerging technologies like the Interferometric Synthetic Aperture Radar (InSAR) (Gama et al., 2010;

Solberg et al., 2010a,b; Næsset et al., 2011) or fusion of LiDAR and InSAR (Sun et al., 2011) have been tested as an alternative for LiDAR-based AGB estimation.

An alternative to the expensive “wall-to-wall” surveys is the use of LiDAR measurements within a sampling framework. For instance, double-sampling applications using airborne laser scanning (ALS) measurements on sub-sampled field plots were discussed by Parker & Evans (2004), Andersen & Breidenbach (2007), Corona & Fattorini (2008), and Stephens et al. (2012). Line-intersect sampling for AGB and forest volume estimation using profiling airborne lasers such the Portable Airborne Laser System (PALS; Nelson et al., 2003) was tested in areas up to 5000 km² in size (Nelson et al., 2004, 2008). Three-phase surveys using spaceborne laser, PALS and field plots as the first, second, and third phase samples, respectively, were described by Boudreau et al. (2008) and Nelson et al. (2009).

An alternative design was suggested by Næsset (2005) and Gobakken et al. (2006), considering ALS as a strip sampling tool, where the LiDAR measurements are collected along corridors of several hundred meters wide aligned with the ground plot network. Under this design, the LiDAR data becomes itself a sample, introducing additional uncertainty while reducing the data acquisition costs. Andersen et al. (2009) applied this design in a two-stage survey for biomass estimation in Kenai Peninsula of Alaska, USA. In Norway, this design was tested in a large-scale biomass survey covering nearly 30,000 km² in Hedmark County (HC), in the period 2005-2010, using ALS, PALS, Norwegian NFI data and supplementary ground plots (Næsset et al., 2009).

The estimation in the HC survey was performed using regression estimators under two inferential frameworks: (1) two-stage, design-based model-assisted (DBMA) by Gregoire et al. (2011), and (2) two-phase, model-based (MB) by Ståhl et al. (2011). Gregoire (1998), Schreuder et al. (2001), and Kangas (2006) present more details regarding the design- and model-based inferential frameworks.

The estimation results obtained in the HC project (Gregoire et al., 2011; Gobakken et al., in press; Ståhl et al., 2011) asserted that LiDAR-based estimates attained a precision level close to the pure ground-based sampling. However, for complex surveys like the HC project where systematic sampling is used for collecting both the LiDAR and the field measurements, there is not an unbiased estimator that can be used to estimate uncertainty. To overcome this problem, Ene et al. (*in submission*) introduced a Monte-Carlo sampling simulation approach for assessing the statistical properties of the DBMA and MB estimators used in HC survey. Using simulated sampling from an artificial population allows describing the sampling distribution of complex estimators when analytical solutions do not apply, thus facilitating better efficiency assessments. Moreover, the behaviour of several estimators under complex designs can be assessed simultaneously, and the results can help identify which estimator has the most desirable properties under a given design.

The results presented by Ene et al. (*in submission*) indicated that violating the assumptions the variance estimators were built upon (e.g., using systematic sampling (*SYS*) instead of simple random sampling without replacement (*SRSwoR*)) greatly overestimated the uncertainty of the AGB estimates. Compared to the field-based survey, the simulation results indicated that the precision (in terms of standard error) of the ALS-based estimation improved by 59%, while the estimated standard errors indicated the opposite, i.e. that the standard errors were 1.8 times higher. Moreover, the estimated standard errors due to the first-phase sampling dominated the total estimated sampling error. This suggests adopting variance estimators to systematic sampling that apply alternative methods like successive differences (Cochran, 1997; Wolter, 2007), as proposed by Gregoire et al. (2011) and Ståhl et al. (2011).

In the current study, the simulation methodology introduced by Ene et al. (*in submission*) where unstratified designs were assumed, is taken several steps further by

considering post-stratified designs and by testing variance estimators better suited for systematic sampling. The objectives of the study were:

1. to assess the bias of the variance estimators proposed by Gregoire et al. (2011) and Ståhl et al. (2011) for post-stratified AGB estimation under SRSwoR and SYS designs;
2. to assess the performance of successive differences estimators under two-phase systematic sampling designs;
3. to assess the efficiency of ALS-aided and pure ground-based AGB surveys;
4. to assess the design effects of the two-phase systematic and simple random sampling designs, and
5. to evaluate the relative costs of the ALS aided survey and field inventory under systematic sampling.

The analyses are exemplified through a case study involving post-stratified AGB estimation, where the four administrative units represented the post-strata.

2. Material

The study area was Hedmark County located in southeastern Norway (Figure 1). The county has a land area of 27 399.72 km² and contains four main administrative units (AU). HC is rather heterogeneous with regard to forest types and geomorphology. The dominant tree species present in HC are Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.).

The material comprises three datasets: (1) field inventory data, (2) remote sensing data (ALS and satellite imagery), and (3) cartographical products in the form of a digital terrain model (DTM) as raster data and land-use maps. For a more detailed presentation of the datasets, see Ene *et al.* (*in submission*) and Gobakken *et al.* (*in press*).

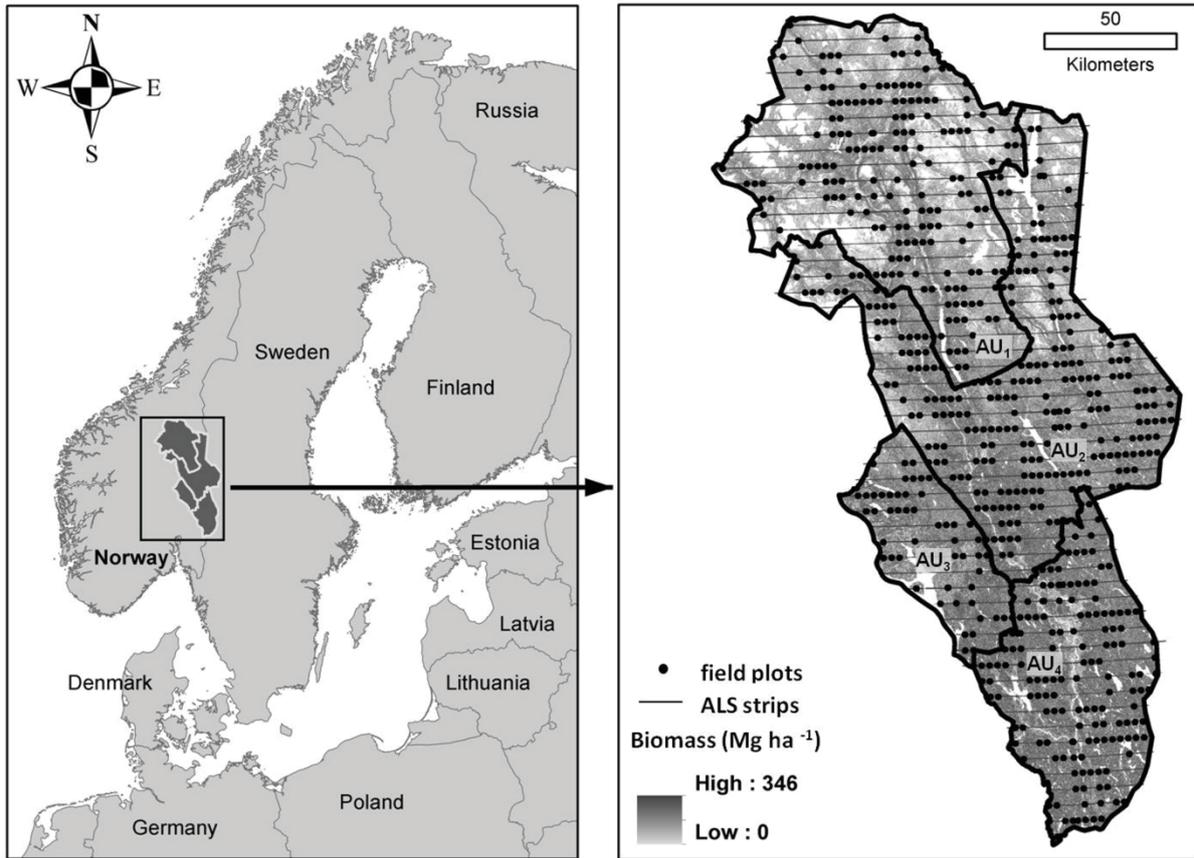


Figure 1 Spatial location of field plots and laser scanning strips across the administrative units in Hedmark County.

2.1 Satellite imagery and map data

An image mosaic providing full coverage of HC was build using the bands 1 to 5 and 7 of three nearly cloud-free Landsat 5 TM images. The spatial resolution of the mosaic was 15.81m such that the area covered by a pixel was 250 m².

The map data consisted in a 25 m spatial resolution DTM produced by the Norwegian Mapping Authority from the official topographic map series, and *AR5* land use maps (Bjørdal & Bjørkelo 2006). Using the DTM and the land-use map, a forest mask for the entire HC was created. The DTM was used for delineating the coniferous tree line in HC (Ørka et al., 2012), and the land-use maps were used for eliminating water and other types of non-forested areas. A forest vegetation mask representing about 86% of the tessellated area of HC was created for the developed areas and for the regions between 850 and 1150 m a.s.l.

2.3 Field data

The field observations were acquired from the permanent plots of the Norwegian National Forest Inventory. The Norwegian NFI is a continuous forest inventory system having the sampling units represented by fixed-area circular plots of 250m² displaced in 3x3 km grid covering the entire country except the areas above the coniferous tree line where the grid spacing is 3x9 km. The grid axes are oriented to the north-south and east-west directions. Due to the Latin square design used for deciding the revisiting order of the field plots across the country (Tomter et al., 2010), approximately 20% of the 2309 NFI plots located in HC are measured every year. Totally, 1483 plots (including NFI plots and additional plots subjectively located in developed and mountain areas) were measured during three field campaigns in the period 2005-2007. Among the 1483 plots, this study is using a subset of 662 plots (Table 1) located on every second east-west oriented NFI grid line, because they were

aligned with the corridors selected for ALS data acquisition. Although the field plots were not uniformly distributed over HC, they covered the main vegetation gradients in HC (Ene et al., *in submission*). More details regarding the field plot selection can be found in Gobakken et al. (*in press*).

On each plot, all trees having a breast height diameter (*dbh*) ≥ 5 cm were callipered, and approximately ten sample trees were selected proportional to stem basal area for height measurements (*h*). A detailed description regarding the NFI estimation methods is given by Tomter et al. (2010) and Gobakken et al. (*in press*). The total aboveground dry biomass of each living tree having *h* ≥ 1.3 m was predicted using tree-species specific allometric models (Marklund, 1988), and tree level AGB estimates were summed up to obtain plot estimates which then were considered to be equivalent to actual AGB of live trees on the plot.

Table 1. AGB distribution (Mg ha^{-1}) by field plots and administrative units (AU)

AU	NFI plots by AU		AGB distribution by plots		
	no.	%	mean	std.dev	range
AU ₁	178	26.9	33.0	32.7	0.47 - 141.98
AU ₂	261	39.4	49.2	51.5	0.26 - 294.83
AU ₃	54	8.2	61.4	62.4	0.58 - 290.85
AU ₄	169	25.5	72.2	63.6	0.65 - 331.51
Total	662	100	51.7	53.7	0.26 - 331.51

The field plot positions were determined using differential post-processing of dual-frequency Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) measurements acquired using Topcon LegacyE receivers. The observation time was at least 15 minutes on each plot, at a 2 seconds logging rate. Six base stations were established at different locations across HC, such that the vector ranges between base stations and plots were less than 50 km. Vector differential corrections were calculated for each plot using the nearest base station as reference. The precision of plot centre positioning reported

by the Pinnacle 1.0 post-processing software (Anon., 1999) varied between 0 and 2 m, with an average of 0.05 m.

2.4 ALS data

The ALS data acquisition was performed during the period July-September 2006 along fifty-three parallel flight lines, using two Optech ALTM 3100 laser-scanning systems (Optech, Canada) mounted on two PA31 Piper Navajo aircrafts. The ALS strips were equally spaced at a distance of 6 km in the north-south direction, covering approximately 8.4 % of the land area in HC. The ALS acquisition parameters (average flying height approximately 800 m above ground, average speed 75 ms^{-1} , 100 kHz pulse repetition frequency, 55 Hz scan frequency, maximum scanning angle of 17°) translated to an average density of $2.8 \text{ echoes m}^{-2}$, and an average swath width of approximately 500 m. The digital terrain surface was created using the TerraSolid software (TerraSolid Ltd, Finland). The laser echoes recorded as ‘single’ and ‘first of many’ by the ALTM 3100 sensors were used for further analysis after height normalization relative to the digital terrain surface.

3. Methods

The study was performed using the simulation approach described in Ene et al. (in submission). Briefly, this approach requires creating first an artificial population using the empirical observations. The creation of the artificial population starts with fitting a Gaussian copula (Nelsen, 2006) to the empirical data (field observations and remote sensing information). Then, a large sample generated from the copula model is imputed to the image mosaic covering the HC area using the nearest neighbor method. Thus, the resulting dataset with imputed ALS metrics and AGB values for every image pixel in HC is considered our population. Using this artificial population as ground-truth, Monte-Carlo sampling simulation

is conducted following the same design used to collect the empirical data. The properties of the estimators are assessed from their sampling distributions derived using simulated sampling.

Following the outline of the HC project, three sampling strategies (*sensu* Särndal et al., 1992, p. 30; Gregoire & Valentine, 2008, p. 1) were considered. The first strategy corresponds roughly to the current Norwegian NFI system and uses the Horvitz-Thompson estimator followed by design-based inference (DBHT). The second strategy considered model-assisted estimation with design-based inference (DBMA) and the third consisted of model-based estimation and inference (MB). The inference under DBHT relies upon the sampling design and makes no use of auxiliary information. The DBMA inference uses the auxiliary information provided by ALS for estimation through the linear regression model developed in section 3.1, and the validity of the inference is based on the probabilistic design. Typically, the model-based inference does not require a probabilistic sample. In our case, this applies only to the ground sample which can be probabilistic or not, and it allows using any external model considered appropriate for the data. However, the LiDAR data acquisition follows the same design as for the DBMA, and the MB approach in fact mixes the design- and model-based inference.

In order to overcome the true standard error overestimation reported by Ene et al. (*in submission*), two variance estimators based on successive differences were tested.

The sampling designs, the estimators and the assessment procedures are introduced in the remainder of this section. The methodology is demonstrated through a case study resembling the HC survey, considering a post-stratified estimation by treating the administrative units of HC as poststrata.

3.1 Predictive ALS-based model for AGB

A generalized linear model (McCulloch & Nelder, 1989) was used to describe the entire population using the combined observations from all groups, the group effects (different intercepts) being represented using dummy variables (Särndal *et al.*, 1992, p. 261-262). The common-slope group regression model was formulated as:

$$\begin{cases} g(\mu) = X\beta' \\ AGB = \mu + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad \text{eq(1)}$$

where the $g(\cdot)$ function is the square-root link, $X\beta'$ is the linear predictor of the expected AGB value μ , X is the design matrix and β is the vector of parameters (McCulloch & Nelder, 1989, p. 26-27). The predictor variables included in the model were the maximum echo height (H_{max}), a canopy density metric (D_{10}), the interaction term $H_{max} \times D_{10}$, and three dummy regressors corresponding to AU₂₋₄. For obtaining the canopy density metrics, the ranges of the laser echo heights were first defined for each plot as the difference the 95% percentile of the echo height distributions and a minimum threshold height of 1.3 m. Then, these ranges were partitioned into ten fractions of equal height and the densities of laser echoes above each fraction (D_0 to D_{90}) were calculated (Næsset 2004; Gobakken and Næsset, 2008). The overall RMSE of the model was 16.5 Mg ha⁻¹ (32.0%), with $r^2 = 0.91$, where r is the linear correlation coefficient between the original and predicted AGB values (Zheng & Agresti, 2000). The group RMSEs were, in order, 12.3 Mg ha⁻¹ (37.2%), 14.6 Mg ha⁻¹ (29.7%), 27.0 Mg ha⁻¹ (44.0%) and 19.8 Mg ha⁻¹ (27.4%) for AU 1 to 4. The model parameters β were estimated from the second-phase samples selected during simulations. The derivation of the ALS metrics used for building the predictive models for AGB is described in Ene *et al.* (*in submission*) and Gobakken *et al.* (*in press*).

3.2 Creating the sampling simulator

The sampling simulator described by Ene et al. (*in submission*) consists of an artificial population and a sample selection mechanism.

The artificial population serving as ground-truth during simulations was constructed as follows:

- a copula model was fitted to a multivariate empirical dataset containing plot-level measurements (field observations, ALS metrics and spectral information);
- a large multivariate sample was generate from the fitted copula model using uniform pseudo-random numbers;
- the large-sample data was generalized across the study area using satellite imagery data and nearest neighbor imputations.

We emphasize that the aim was not to re-create the true forest of HC, but to produce an artificial population retaining the major geographical trends in HC while maintaining a realistic overall variability. The AGB statistics for the artificial population are presented in Table 2.

Table2. AGB distribution (Mg ha^{-1}) in the artificial population by administrative unit (AU)

AU	AGB distribution by AU			
	% area	mean	std.dev	range
AU ₁	32.1	29.4	38.7	0.12 - 343.11
AU ₂	37.9	45.3	51.9	0.12 - 345.76
AU ₃	10.0	63.3	61.4	0.12 - 345.76
AU ₄	20.0	76.3	66.6	0.12 - 345.76
Total	100	48.2	55.3	0.12 - 345.76

The spatial sampling frame used to identify the sample units was in the form of a look-up table following the raster representation of the HC area, such that each cell corresponds to a sampling unit. The geographical distances were converted from meters to

number of pixels through division by the pixel size (15.81 m) and rounding to integer value using the ceiling operator. Hereafter, the terms pixel and population elements are used interchangeably.

3.4 Sampling designs

The simulated sampling was performed using designs described in Ene et al. (*in submission*): one-phase systematic sampling (SYS_1), two-phase simple random sampling without replacement ($SRSwoR_2$) and two-phase systematic sampling (SYS_2). The SYS_1 design resembles the Norwegian NFI system and it was chosen as a benchmark. The SYS_2 design replicates the one used in the HC survey and it may represent a feasible solution for future operational applications. The use of the $SRSwoR_2$ design is justified because it complies with the theoretical assumptions required by several variance estimators described in section 3.5.2.

The ALS-aided survey in HC involves systematic sampling at both selection stages (SYS_2). Under this design, the subsampling is invariant but the selection of secondary sampling units (SSUs) is not independent. On the other hand, the $SRSwoR_2$ design fulfils both the invariance and independency requirements. However, the MB framework allows using independent sources of auxiliary data (Ståhl et al., 2011). Hence, the SYS_2 and the $SRSwoR_2$ designs were viewed as two-phase element sampling designs according to Särndal et al. (1992, p. 344-345).

To perform the two-phase sampling, HC was partitioned into $M=625$ mutually exclusive clusters following the raster representation spatial frame. The clusters represent the ALS corridors and are oriented parallel to the east-west direction. The cluster width was set to 32 pixels (approximately 500m) to match the average ALS swath width in the HC survey. However, the width of the last cluster in the frame was 29 pixels (approximately 485m), because the spatial frame dimension along the y-axis (north-south direction) was not an

integer multiple of the cluster swath. Due to the irregularly shaped area of HC (Figure 1), the clusters have varying lengths and contain different numbers of sample units. During simulated sampling, the clusters represent the primary sampling units (PSU), and the pixels subsampled within each PSU are the secondary sampling units (SSUs). Henceforth, the number of pixels in the p th PSU is denoted n_{1p} , and the number of SSUs sampled from the p th PSU was denoted as n_{2p} .

3.4.1 One-phase SYS

The ground-based inventory (SYS_1) was viewed as a one-phase systematic sampling design (Köhl et al., 2006, p. 81) with restricted selection in two dimensions, where each sample is selected with the same probability $(a_{x,l} \times a_{y,l})^{-1}$. The first sample unit is drawn randomly among the first $a_{x,l} \times a_{y,l}$ population elements with the inclusion probability $\pi_k = (a_{x,l} \times a_{y,l})^{-1} = a_{xy,l}^{-1}$, where $a_{x,l}$ and $a_{y,l}$ are the sampling intervals (in pixels) along the spatial frame's axes. The remaining sample units are selected systematically at the predefined sampling intervals $a_{x,l}$ and $a_{y,l}$, thus the sample size varies due to the irregular borders of HC. Using the raster sampling frame of the artificial population, the sampling intervals along the east and north directions were $a_{x,l} = 190$ pixels and $a_{y,l} = 380$ pixels, corresponding to approximately 3000 m and 6000m, respectively. The total number of possible samples produced under the SYS_1 design was $a_{xy,l} = a_{x,l} \times a_{y,l} = 190 \times 380 = 72,200$.

However, this design does not fully comply with the Norwegian NFI system because it ignores every second east-west grid line and it assumes that all field plots distributed across HC are measured during one field campaign rather than over a five-year cycle.

3.4.2 Two-phase SYS

Following the ALS survey of HC, the PSUs were spaced 6 km (380 pixels) apart, and each swath was parsed to a 500m width (32 pixels), which translates into a sampling interval of $a_{y2} = 12$ ($\approx M/n_I$). Hence, the total number of first-phase samples was $a_{y2} = 12$, of which the first sample contains $n_I=53$ PSUs and the rest of the 11 samples contain $n_I=52$ PSUs. The sampling intervals along the (x, y)-axis for the two-dimensional grid were $a_{x2} = 3000\text{m}$ (190 pixels) and $a_{y2} = 6000\text{m}$ (380 pixels), the first SSU being selected among the first $ALS_w \times a_{x2}$ units of the first PSU, where ALS_w is the width of the ALS strips (32 pixels). The number of all possible second-stage samples selected from every first-stage sample was equal to $ALS_w \times a_{x2} = 32 \times 190 = 6,080$.

3.4.3 Two-phase SRSwoR

Under $SRSwoR_2$, the size of the first-phase samples was set as the expected size of the first-phase sample under SYS_2 ($n_I=52$). The SSUs were selected independently within each PSU. The sampling invariance requirement was fulfilled by setting the inclusion probabilities for the SSUs in each PSU as the average PSU length (in pixels) divided by the sampling interval along the east-west direction ($a_{x2}=190$ pixels) used under SYS_2 . This would also allow the expected size of the second-phase samples under $SRSwoR_2$ and SYS_2 to match each other.

The $SRSwoR_2$ design complies with the theoretical assumptions considered by Gregoire et al. (2011) and Ståhl et al. (2011) for building the variance estimators for the total AGB. Thus, we could use simulated sampling following the $SRSwoR_2$ design for validating the estimators presented in section 3.5.2.

3.5 Post-stratified estimation

Post-stratification is the process of partitioning the observations into a set of discrete strata after the sample has been selected (Gregoire & Valentine, 2008, p. 152). Theoretically, any sample can be post-stratified, but the post-stratified estimation becomes meaningful only if the strata sizes or the proportion of each stratum relative to the population (the strata weights) are known (Zhang, 2000; Gregoire & Valentine, 2008, p. 152). With post-stratification by administrative units, i.e., by treating the AUs as our post-strata, the sampling is not performed independently within each AU, the sample units being assigned to the AUs after selection. Hence, the strata sample sizes are random variables because they are known only after the sample selection (Särndal et al., 1992, p. 261; Gregoire & Valentine, 2008, p. 153).

The systematic sampling designs used for the HC survey introduces dependencies between post-strata estimates because the grid lines (for SYS_1) and ALS corridors (SYS_2) often extend over several AUs.

These dependencies are accounted for under the model-based inference (Ståhl et al., 2011), while the DBMA estimator does not incorporate a between-stratum covariance term (Gregoire et al., 2011; Gobakken et al., *in press*).

We used u to index the U administrative units in HC, n for the ground sample size and N for the population size. Hence, the size of the u th AU will be denoted as N_u and the number of field observations as n_u . Considering the post-stratum sizes known without error from the GIS system, the post-stratum weights were calculated as $W_u = N_u/N$.

3.5.1. Post-stratified estimation following SYS_1

Assuming SRS_{woR} and conditioning on the selected sample size n_u , the poststratification estimator of total AGB (Mg ha^{-1}) in the u th AU (Gregoire & Valentine, 2008, p. 154-155 and

p.162-163) is the Horvitz-Thompson estimator $\hat{t}_{HT,u} = cf N_u \left(\frac{1}{n_u} \sum_{i=1}^{n_u} AGB_{iu} \right)$, where $cf = 1000$

$m^2/250 m^2 = 40$ is the per-hectare expansion factor. The average per-hectare AGB in stratum

AU_u is given by $\hat{\mu}_{AGB,u} = \frac{1}{N_u} \hat{t}_{HT,u}$ and the design-based variance of $\hat{\mu}_{HT,u}$ is estimated as

$$\hat{v}_{1.1,u} = cf^2 \left(1 - \frac{n_u}{N_u} \right) s_{HT,u}^2 \quad \text{eq(2)}$$

where $s_{\hat{\mu}_{HT,u}}^2 = (n_u - 1)^{-1} \sum_{i=1}^{n_u} (AGB_{i,u} - \hat{\mu}_{HT,u})^2$ is the stratum sample variance and the $1 -$

n_u/N_u term is the finite population correction (negligible given the small sample size

compared to the stratum size).

The average AGB per-hectare across AUs was estimated as $\hat{\mu}_{AGB} = \sum_{u=1}^U W_u \hat{\mu}_{HT,u}$ and

the conditional variance estimator for $\hat{\mu}_{AGB}$ is given by (Gregoire and Valentine, 2008, p.

155):

$$\hat{v}_{1.1} = cf^2 \sum_{u=1}^U W_u^2 \frac{s_{\hat{\mu}_{HT,u}}^2}{n_u} \quad \text{eq(3)}$$

3.5.2. Post-stratified estimation following *SRSwoR*₂ and *SYS*₂

The DBMA inference for μ_{AGB} was performed using two alternative estimators, i.e., (1) the two-stage element sampling variance estimator described by Gregoire et al. (2011) which requires *SRSwoR* at both selection stages, and (2) a simplified variance estimator (Särndal et al., 1992, p. 154) which assumes *SRSwoR* of PSUs. In the simulation study presented by Ene et al. (in submission), these estimators performed equally well for unstratified estimation under *SRSwoR*₂ and *SYS*₂.

Using the generalized regression estimator (GREG) from equation 8.9.6 in Särndal et al. (1992, p. 323) and equation 6 in Gregoire et al. (2011), the estimated total for the p th PSU

intersecting the u th AU is $\hat{t}_{GREG,pu} = \sum_{k=1}^{N_{pu}} A\hat{G}B_{pu,k} + \frac{N_p}{n_p} \sum_{j=1}^{n_{pu}} e_{pu,j}$. The first term is the

summation over predicted AGB for population elements in a selected PSU, and the second term represents the bias correction estimated from the regression residuals $e_{pu,j}$ for the SSUs

within the particular PSU. The AGB total for the u th AU becomes $\hat{t}_{GREG,u} = a_{y,2} \sum_{p=1}^{n_1} \hat{t}_{GREG,pu}$,

and the average AGB per hectare for the u th AU is estimated as $\hat{\mu}_{GREG,u} = cf N_u^{-1} \hat{t}_{GREG,u}$.

Following Gregoire et al. (2011), an approximated variance estimator of per-hectare $\hat{\mu}_{GREG,u}$ is:

$$\hat{v}_{2.1u} = \left(\frac{cf}{N}\right)^2 \left\{ M^2 \left(\frac{1}{n_1} - \frac{1}{M}\right) \left[s_{tu}^2 - \frac{1}{n_1} \sum_{p=1}^{n_{1u}} N_p^2 \left(\frac{1}{n_{2p}} - \frac{1}{N_p}\right) s_{pu}^2 \right] + \left(\frac{M}{n_1}\right)^2 \sum_{p=1}^{n_{1u}} N_p^2 \left(\frac{1}{n_{2p}} - \frac{1}{N_p}\right) s_{epu}^2 \right\} \quad \text{eq(4)}$$

In equation 4, $s_{tu}^2 = (n_{1u} - 1)^{-1} \sum_{p=1}^{n_{1u}} (\hat{t}_{pu\pi} - \bar{\hat{t}}_{u\pi})^2$ estimates the between-PSU variability,

where $\hat{t}_{pu\pi} = \frac{N_p}{n_{2p}} \sum_{j=1}^{n_{2pu}} AGB_{pu,j}$ is the HT estimate of the p th PSU total, and $\bar{\hat{t}}_{u\pi} = n_{1u}^{-1} \sum_{p=1}^{n_{1u}} \hat{t}_{pu\pi}$ is the

average of the estimated PSU totals. The variances between the SSUs contained by the p th

PSU are estimated as $s_{pu}^2 = (n_{2p} - 1)^{-1} \sum_{k=1}^{n_{2p}} \left(AGB_{pk} - \hat{A}\hat{G}B_p \right)^2$ where AGB_{pk} is the AGB value of

the k th SSU from the p th PSU, and $\hat{A}\hat{G}B_p = n_{2pu}^{-1} \sum_{k=1}^{n_{2pu}} AGB_{puk}$ is the average AGB of the

second-stage sample units within the p th PSU. The second term approximates the variance due to subsampling by prorating the cluster residual

variances $s_{epu}^2 = (n_{2pu} - 1)^{-1} \sum_{k=1}^{n_{2pu}} (e_{puk} - \bar{e}_{pu})^2$, where $e_{puk} = AGB_{puk} - \hat{AGB}_{puk}$ is the k^{th}

regression residual in the p^{th} PSU, and $\bar{e}_{pu} = n_{2pu}^{-1} \sum_{k=1}^{n_{2pu}} e_{puk}$ is the average residual for the p^{th}

PSU.

The across-strata estimate of total AGB $\hat{t}_{GREG} = \sum_{u=1}^U \hat{t}_{GREG,u}$ is obtained by collapsing the

post-stratum total estimates, and the across-strata per-hectare AGB estimate

becomes $\hat{\mu}_{GREG} = cf N^{-1} \hat{t}_{GREG}$. The variance estimator of $\hat{\mu}_{GREG}$ is expressed as (Gregoire et al., 2011):

$$\hat{v}_{2.1} = \left(\frac{cf}{N}\right)^2 M^2 \left(\frac{1}{n_1} - \frac{1}{M}\right) \left[s_i^2 - \frac{1}{n_1} \sum_{p=1}^{n_1} N_p^2 \left(\frac{1}{n_{2p}} - \frac{1}{N_p}\right) s_p^2 \right] + \left(\frac{M}{n_1}\right)^2 \sum_{p=1}^{n_1} N_p^2 \left(\frac{1}{n_{2p}} - \frac{1}{N_p}\right) s_{ep}^2 \quad \text{eq(5)}$$

To avoid obtaining negative variance estimates when using the $\hat{v}_{2.1}$ estimator (Gregoire et al., 2011; Gobakken et al., *in press*), Ene et al. (in submission) used the simplified estimator described by Särndal et al. (1992, p. 154), which assumes *SRSwoR* for PSU selection:

$$\hat{v}_{2.2u} = \frac{1}{n_{1u} - 1} \sum_{p=1}^{n_{1u}} \left(\hat{t}_{GREG,p} - \frac{1}{n_{1u}} \sum_{p=1}^{n_{1u}} \hat{t}_{GREG,p} \right)^2 \quad \text{eq(6)}$$

The per-hectare variance estimator is expressed as:

$$\hat{v}_{2.2u} = \left(\frac{cf \cdot M}{N}\right)^2 \left(\frac{1}{n_{1u}} - \frac{1}{M}\right) \hat{v}_{2.2u} \quad \text{eq(7)}$$

Under the MB inference, the AGB total of the p^{th} PSU predicted using the regression estimator (REG) becomes $\hat{t}_{REG,p} = \sum_{s_{1pu}} \hat{AGB} + \sum_{s_{2pu}} AGB$. The second term in the equation is the summation of the sample AGB over in the p^{th} PSU, and the first term is sum of the predictions for the rest of the population elements within the PSU. The estimate of AGB per-

hectare in the u th AU is expressed as the ratio of the total and average PSU size (Ståhl et al., 2011, equation 11):

$$\hat{\mu}_{REG,u} = \left(\frac{n_{1pu}}{n_{1u}^{-1} \sum_{p=1}^{n_{1pu}} \hat{t}_{REG,pu}} \right) \left(\frac{n_{1pu}}{n_{1u}^{-1} \sum_{p=1}^{n_{1pu}} N_{pu}} \right)^{-1} = \hat{t}_{REG,u} \bar{N}_u^{-1}.$$

The approximated variance estimator of $\hat{\mu}_{REG,u}$ assuming *SRSwoR* and finite population correction for the first phase sampling can be written as (Ståhl et al., 2011, equation 15):

$$\hat{v}_{2.3u} = \left(\frac{1}{n_1} - \frac{1}{M} \right) \frac{1}{\bar{N}_u^2} \frac{\sum_{p=1}^{n_1} (\hat{t}_{REG,u} - \hat{\mu}_{REG,u} N_{pu})^2}{n_1 - 1} + \frac{1}{\bar{N}_u^2} \sum_{j_1}^m \sum_{j_2}^m \hat{C}ov_{S_2}(\hat{\beta}_i, \hat{\beta}_j) \hat{G}'_{j_1 h} \hat{G}'_{j_2 k} \quad \text{eq(8)}$$

The first term in equation 8 estimates the error due to the first-phase sampling. The second term accounts for the uncertainty of the model parameters estimated from the second-phase sample, and it includes the covariance matrix of the m model parameters ($\hat{\beta}$) and the average values of the first-order derivatives of the $g^{-1}(\cdot)$ function from equation 1. The across-stratum estimates for per-hectare AGB was calculated as (Ståhl et al., 2011, equation 12)

$$\hat{\mu}_{REG} = \sum_{u=1}^U W_u \hat{\mu}_{REG,u}, \text{ with an approximated variance estimator (Ståhl et al., 2011, equation 14)}$$

given as:

$$\hat{v}_{2.3} = \frac{1}{n_1} \sum_{h=1}^U \sum_{t=1}^U \frac{W_h W_t}{\bar{N}_h \bar{N}_t} \frac{\sum_{p=1}^{n_1} (\hat{t}_{REG,ph} - \hat{\mu}_{REG,h} N_{ph})(\hat{t}_{REG,pt} - \hat{\mu}_{REG,t} N_{pt})}{n_1 - 1} + \sum_{h=1}^U \sum_{t=1}^U \frac{W_h W_t}{\bar{N}_h \bar{N}_t} \sum_{j_1}^m \sum_{j_2}^m \hat{C}ov_{S_2}(\hat{\beta}_i, \hat{\beta}_j) \hat{G}'_{j_1 h} \hat{G}'_{j_2 t} \quad \text{eq(9)}$$

The stratum dependencies are accounted by the cross-stratum covariances denoted by the double summations in equation 9.

3.5.3. Post-stratified variance estimation under SYS_2 using successive differences estimators

When using systematic sampling, there is not an unbiased, design-based estimator variance estimator (Cochran 1977, §8.1; Gregoire & Valentine, 2008, §3.2.2). However, Cochran (1977, §8) describes how a superpopulation model can be used in order to obtain variance estimators approximately unbiased for the model, suggesting that gathering information about the population structure is recommended at both planning and estimation phases. For the case of equal-probability, one-dimensional systematic sampling, Wolter (2007, §8.3.2) discuss the properties of several successive differences variance estimators assuming various types of superpopulation models. Historically, successive differences estimators were used with line survey sampling in early forest inventories conducted in Nordic countries (Heikkinen, 2006).

Because the successive differences estimators described in Wolter (2007) assume a one-dimensional sampling, the artificial population had to be considered as consisting of M non-overlapping clusters, so the contrasts between clusters could be calculated. At this point, as underlined by Wolter (2007, p. 323), expert judgment can be applied to investigate if there is an underlying model describing the spatial distribution of the AGB in the M clusters, in order to select an appropriate variance estimator. More precisely, we tried to detect the presence of trends, stratification effects and auto-correlation between cluster AGB totals and averages using the artificial population as a proxy for the real forest in HC. The results of this exploratory analysis are presented in Figure 2.

A second-order trend ($R^2 = 0.76$) was found among cluster totals (Figure 2) and a linear trend ($R^2 = 0.94$) between the cluster means (Figure 3). Moreover, the Durbin-Watson test revealed significant autocorrelation among the residuals obtained after trend removal from the cluster totals and means (Figure 2 and Figure 3, respectively). Henceforth, we considered the cluster means for calculating the successive differences estimators.

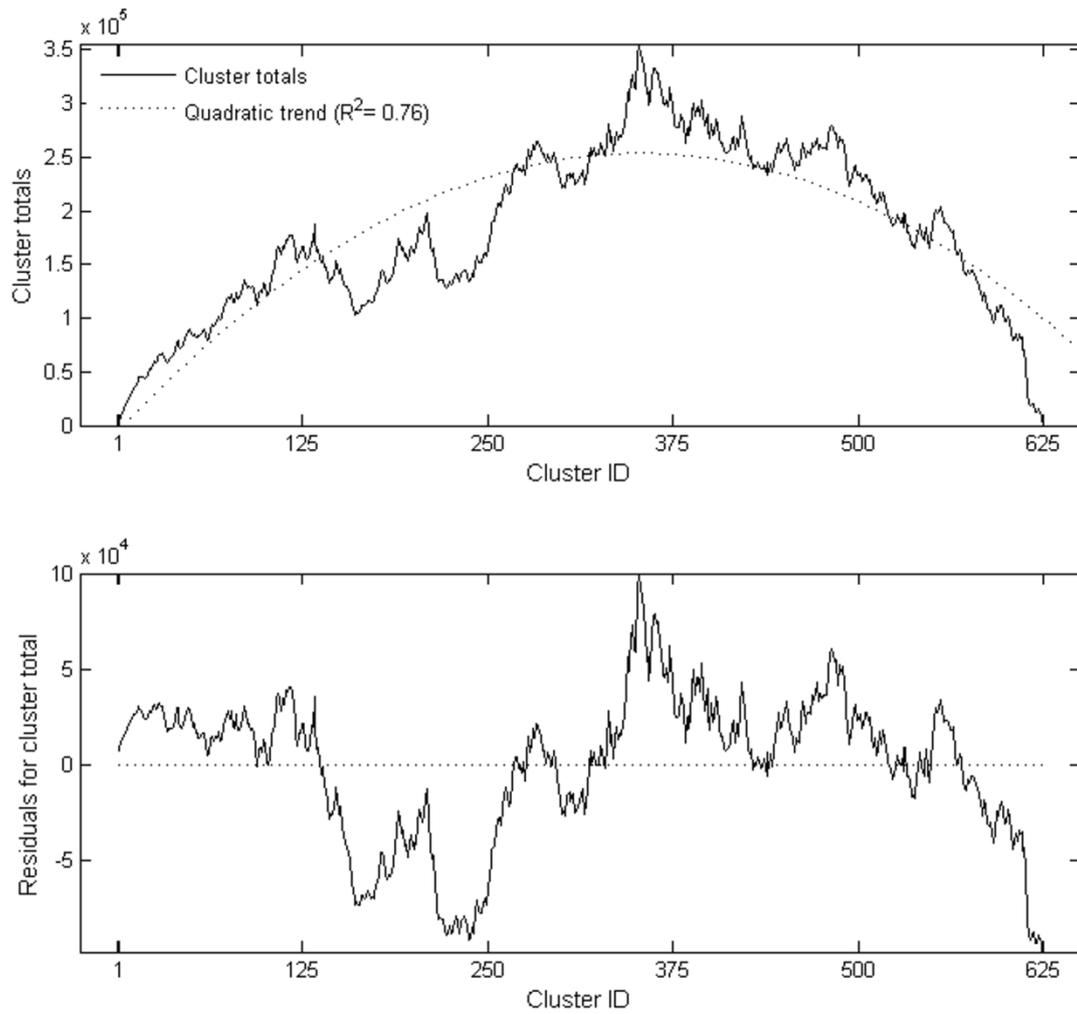


Figure 2 The north-south quadratic trend among the cluster totals in the artificial population (Figure 2a), and the distribution of the residuals after trend removal (Figure 2b). The values on x-axis indicates the cluster ordering from 1 to M , where $M = 625$.

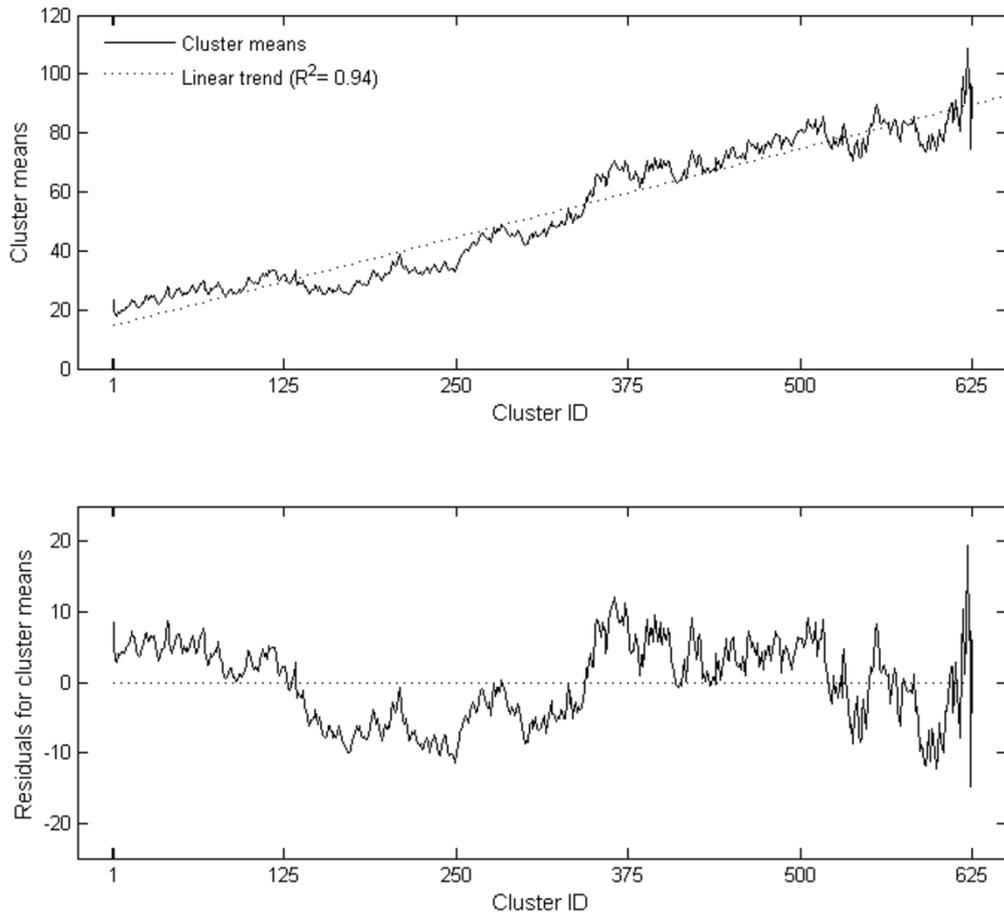


Figure 3 The north-south linear trend among the clusters means in the artificial population (Figure 3a), and the distribution of the residuals after trend removal (Figure 3b). The values on x-axis indicates the cluster ordering from 1 to M , where $M = 625$.

It is more realistic to assume that the structure of the artificial population is in fact a combination of several effects. Beside the trends and auto-correlation, stratification effects are also present as indicated by the plateaus observed in the residual distributions (Figure 2b and 3b).

However, analyzing all 12 possible systematic samples of clusters we did not find indications of any trend among their means, which may suggest that the trend is not the dominating effect in the population of clusters. Hence, we decided to use the v_4 -estimator described in Wolter (2007, §8.3) which was found to provide protection against trend, stratification and autocorrelation effects (Wolter 2007, §8.3 p. 331). Although Marcell et al. (2009) found this estimator satisfactory, they advice using an estimator which accounts for

auto-correlation effects (the v_8 -estimator in Wolter, 2007, §8.3), even when linear trends are present. However, we did not consider the v_8 -estimator here because it was found too optimistic and provided mediocre results in populations exhibiting stratification effects (Wolter, 2007, p. 332).

The v_4 -estimator described in Wolter (2007) does not consider systematic sampling of clusters of unequal sizes, as we are faced with in the HC survey. For this reason, we used the weighted equivalent of the v_4 -estimator described in equation 3 in Nelson et al. (2008), formulated as:

$$\hat{v}_{NT} = \left(1 - \frac{n_1}{M}\right) \frac{n_1}{96(n_1 - 2)} \sum_{p=2}^{n_{1u}} (w_{p-1,u} + 2w_{pu} + w_{p+1,u})^2 (\hat{\mu}_{p-1,u} - 2\hat{\mu}_{pu} + \hat{\mu}_{p+1,u})^2$$

Here, the cluster weights are calculated as $w_{pu} = n_{pu} \left(\sum_{p=1}^{n_{1u}} n_{pu} \right)^{-1}$ and the cluster

averages were obtained as $\hat{\mu}_{pu} = N_{pu}^{-1} \hat{t}_{GREG,pu}$ when used with GREG, and

as $\hat{\mu}_{pu} = N_{pu}^{-1} \hat{t}_{REG,pu}$ for REG. Thus, the systematic variances under DBMA inference were

estimated as:

$$\hat{v}_{NT,u}^{DBMA} = \left(1 - \frac{n_1}{M}\right) \frac{n_1}{96(n_1 - 2)} \sum_{p=2}^{n_{1u}} w_{NT,pu}^2 (\hat{\mu}_{GREG,p-1,u} - 2\hat{\mu}_{GREG,pu} + \hat{\mu}_{GREG,p+1,u})^2 \quad \text{eq(10)}$$

For the case of MB inference, the successive differences estimator was used for estimating the sampling error due the first-phase sample, as:

$$\hat{v}_{NT,u}^{REG} = \left(1 - \frac{n_1}{M}\right) \frac{n_1}{96(n_1 - 2)} \sum_{p=2}^{n_{1u}} w_{NT,pu}^2 (\hat{\mu}_{REG,p-1,u} - 2\hat{\mu}_{REG,pu} + \hat{\mu}_{REG,p+1,u})^2 + \hat{D}_{2,REG,u} \quad \text{eq(11)}$$

where $\hat{D}_{2,REG,u}$ is the model-error term of the $\hat{v}_{2.3,u}$ estimator in equation 8.

The across-stratum variance estimator under DBMA was constructed by collapsing the strata within the same PSU:

$$\hat{v}_{NT}^{GREG} = \left(1 - \frac{n_1}{M}\right) \frac{n_1}{96(n_1 - 2)} \sum_{p=2}^{n_1} w_{NT,p}^2 \left(\hat{\mu}_{GREG,p-1} - 2\hat{\mu}_{GREG,p} + \hat{\mu}_{GREG,p+1}\right)^2 \quad \text{eq(12)}$$

where $w_{NT,p} = w_{p-1} + 2w_p + w_{p+1}$, $\delta_{NT,p} = \hat{\mu}_{GREG,p-1} - 2\hat{\mu}_{GREG,p} + \hat{\mu}_{GREG,p+1}$,

$$w_p = n_p \left(\sum_{p=1}^{n_1} n_p \right)^{-1} \quad \text{and}$$

$$\hat{\mu}_p = N_p^{-1} \hat{t}_{GREG,p}.$$

For the MB inference, the $\hat{v}_{2,3}$ estimator considered the covariances between PSUs, since they can extend over several strata. Hence, we proposed a variance estimator which includes the covariance of the successive differences between clusters, and the model-error term from $\hat{v}_{2,3}$, formulated as:

$$\hat{v}_{NT}^{REG} = \left(1 - \frac{n_1}{M}\right) n_1 \sum_{h=1}^U \sum_{t=1}^U W_h W_t \frac{\sum_{p=2}^{n_1} \left(w_{NT,ph} \delta_{NT,ph}^{REG} \right) \left(w_{NT,pt} \delta_{NT,pt}^{REG} \right)}{96(n_1 - 2)} + \hat{D}_{2,REG} \quad \text{eq(13)}$$

The $\hat{D}_{2,REG}$ term in equations 13 represents the model-error estimate used for the $\hat{v}_{2,3}$ estimator (equation 9).

3.6 Accuracy assessment

The sampling distributions of an estimator under design-based and model-based inference are not identical (Särndal et al., 1992, p. 534). However, the inference following the designs presented in section 3.4 requires probability sampling, which justifies using simulated sampling from the artificial population as described in Ene et al. (in submission) for assessing the properties of the estimators introduced in section 3.2.

The strategies using systematic sampling allowed extracting all possible samples and the complete sampling distributions of the estimators could be obtained. Under SRS_{woR_2} , the

sampling distributions of the estimators were approximated from 5,000 first-phase samples with 100 second-phase samples each.

3.6.1 Assessing the AGB estimation

The standard error of the sampling distribution of $\hat{\mu}_{AGB}$ constituted the observed standard error (SE_{obs}). The analytical standard errors (SE_{an}) were calculated as the square root of the average variance estimates obtained from each sample during the simulated sampling. In addition, the coefficients of variation relative to the group and population means were for also calculated.

For assessing the agreement between observed and analytical standard errors, we built percentile-based confidence intervals using the 2.5 and 97.5 percentiles of the sampling distribution of SE_{an} , and checked if they covered SE_{obs} .

To assess the inference for the estimated AGB, 95% z-based confidence intervals (CI) were built around $\hat{\mu}_{AGB}$ using SE_{an} estimates, and the coverage rates of the z-CIs obtained from simulated sampling were compared to the nominal 95% coverage. Having a skewed parent population, the from-above and from-below failure rates of the z-CIs we also examined as described by Gregoire & Schabenberger (1999).

3.6.2 Relative efficiency of the AGB estimation

The relative efficiency between sampling strategies was calculated using the root mean square errors $RMSE = \sqrt{Bias(\hat{\mu}_{AGB})^2 + SE_{obs}^2}$ of the estimators, where $Bias(\hat{\mu}_{AGB})$ is the estimated bias calculated as the difference between the Monte Carlo $\hat{\mu}_{AGB}$ estimates and the strata and population values.

Hence, the accuracy (*eff*) of SYS_2 with regression estimator relative to SYS_1 with HT estimator was expressed as the proportion between the observed *RMSE* obtained under each sampling strategy:

$$eff_{SYS_1}^{SYS_2} = \frac{RMSE_{SYS_2}(\hat{\mu}_{AGB})}{RMSE_{SYS_1}(\hat{\mu}_{AGB})} \quad \text{eq(14)}$$

The relative efficiency of the SYS_2 and SRS_{woR_2} designs was expressed using the design effect (Cochran, 1977, p. 85; Särndal et al., 1992, p. 492) as the ratio of the observed variances under the two designs:

$$deff(SYS_2, \hat{\mu}_{AGB}) = \frac{Var_{obs, SYS_2}(\hat{\mu}_{AGB})}{Var_{obs, SRS_{woR_2}}(\hat{\mu}_{AGB})} \quad \text{eq(15)}$$

3.7 Cost efficiency analysis

The sample allocation scheme in the poststratified HC survey resembles the proportional-to-size scheme, larger strata being sampled more intensively. Considering this type of sample allocation, our interest was to perform a cost efficiency analysis of the ALS-aided survey relative to the pure ground-based surveys, and to investigate of the survey costs and precision for various sampling and subsampling rates.

In order to perform this analysis, the cost of the ALS data was expressed as a function of the field inventory costs. Simplifying, the total cost of the ALS-aided inventory in HC can be formulated as:

$$C_{Total} = C_{ALS} + C_{Field} \quad \text{eq(16)}$$

where C_{ALS} and C_{Field} are the overall costs associated with the ALS acquisition and field inventory, respectively. The total cost of the ALS acquisition was expressed as $C_{ALS} = n_1 \bar{L} C_{ALS}^0$, where \bar{L} is the average length of the ALS corridors (approximately 75 km

in the HC survey) and C_{ALS}^0 is an average cost per strip-kilometer. Similarly, the total cost of the field survey can be written as $C_{Field} = nC_{Field}^0$, where C_{Field}^0 is an average cost per plot.

Using the notation $k = \frac{C_{ALS}^0}{C_{Field}^0}$, equation 17 becomes:

$$C_{Total} = n_1 \bar{L}(kC_{Field}^0) + nC_{Field}^0 = C_{Field}^0(n_1 \bar{L}k + n) \quad \text{eq(17)}$$

The total cost of the ALS-aided survey formulated in equation (17) is proportional to the average cost of a field plot, the term C_{Field}^0 being a constant which can be neglected in the subsequent analysis.

The relationship between C_{Total} and SE_{obs} (%) was obtained using simulated sampling following one- and two-phase equal probability systematic designs, and considering the three sampling intervals of 3, 6 and 9km along the north-south and east-west directions. For each of the nine combinations, the observed standard errors of the HT, GREG and REG estimators were obtained by performing simulated sampling.

To illustrate the cost-efficiency improvements of ALS-aided inventories over time, the inventory costs were calculated according to equation 17, where k was assigned the values of 1/20, 1/10, and 1/5, respectively. The value $k=1/10$ was obtained from the empirical evidence accumulated with commercial ALS-based forest inventories performed in Norway during the last years (including the HC project) which suggests that the average cost of a field plot is approximately ten times higher than the per-kilometer cost of an ALS strip. The value $k=1/5$ was used to express the situations in the past when the ALS data were highly expensive, and $k=1/20$ is meant to be a rough prediction for the midterm future when the price of ALS data are expected to decrease even more relative to the cost of field measurements.

The relationship between SE_{obs} (%) obtained from simulations and the corresponding C_{Total} values were used to analyze the relative cost efficiency of ALS-aided surveys. In order

to assess the relationship between survey costs and the various error estimates, the relative costs were transformed to the logarithmic scale and the relationship between the standard errors (%) and log-transformed costs was investigated using simple linear regression. Hence, the slope of the regression line indicates the expected change in precision when the costs change by a fixed percentage. Having the costs directly related to the sample size, this allows analyzing also the relationship between precision and sample size.

4. Results

The one-phase samples under SYS_7 contained in average 1309 population elements. The average sizes of the first- and second phase samples were 52 PSUs and 1295 SSUs, respectively. The performance of the estimators in terms of bias and standard errors is presented in Table 3. The coverage properties of the confidence intervals are indicated in Table 4. Further details are presented in the following sections.

Table 3. Results obtained running simulated sampling following one- and two phase designs under design and model-based inference. The percentages relative to population mean are given in the brackets.

Design	Estimators		Poststrata				Total		
	mean	variance	AU ₁	AU ₂	AU ₃	AU ₄			
SYS ₁	HT	observed	Bias	Mg ha ⁻¹	-0.0024	-0.0008	0.0026	0.0002	0.0008
				%	(-8.2e ⁻³)	(-1.8e ⁻³)	(4.1e ⁻³)	(0.3e ⁻³)	(1.7 e ⁻³)
		SE _{obs}	Mg ha ⁻¹	1.7549	2.2082	5.1191	4.0812	1.3942	
		%	(5.9)	(4.9)	(8.1)	(5.3)	(2.9)		
	v1.1	SE _{an}	Mg ha ⁻¹	1.8858	2.3313	5.3684	4.1159	1.4535	
		%	(6.4)	(5.1)	(8.5)	(5.4)	(3.0)		
SRSwoR ₂	GREG	observed	Bias	Mg ha ⁻¹	-0.0248	-0.03696	-0.01079	0.0204	-0.019
				%	(-84.0e ⁻³)	(-8.2e ⁻³)	(-1.7e ⁻³)	(2.7e ⁻³)	(-3.9e ⁻³)
		SE _{obs}	Mg ha ⁻¹	4.9419	7.7351	15.0539	14.3199	2.8298	
		%	(16.8)	(17.1)	(23.8)	(18.8)	(5.9)		
	v2.1	SE _{an}	Mg ha ⁻¹	4.7463	7.5670	14.0849	14.0810	2.8009	
		%	(16.2)	(16.7)	(22.3)	(18.5)	(5.8)		
v2.2	SE _{an}	Mg ha ⁻¹	4.9626	7.7545	15.2155	14.2866	2.7928		
	%	(16.9)	(17.1)	(24.1)	(18.7)	(5.8)			
REG	observed	Bias	Mg ha ⁻¹	-0.0624	-0.0360	0.0885	0.0072	-0.0234	
			%	(-0.2)	(-0.1)	(0.1)	(9.4 e ⁻³)	(-48.6e ⁻³)	
		SE _{obs}	Mg ha ⁻¹	1.3681	2.4754	3.6363	2.2510	1.0047	
		%	(4.7)	(5.5)	(5.8)	(3.0)	(2.1)		
	v.2.3	SE _{an}	Mg ha ⁻¹	1.1027	2.3251	2.7691	1.4246	0.9287	
		%	(3.8)	(5.1)	(4.4)	(1.9)	(1.9)		
SYS ₂	GREG	observed	Bias	Mg ha ⁻¹	-0.04e ⁻⁵	4.1 e ⁻³	7.5v	17.1 e ⁻³	5.7e ⁻³
				%	(-0.1e ⁻³)	(9.0e ⁻³)	(11.9e ⁻³)	(22.4e ⁻³)	(0.0118)
		SE _{obs}	Mg ha ⁻¹	0.7610	1.0511	2.2587	1.7815	0.6079	
		%	(2.6)	(2.3)	(3.6)	(2.3)	(1.3)		
	v2.1	SE _{an}	Mg ha ⁻¹	5.2234	8.0207	15.9841	14.8125	3.0778	
		%	(17.8)	(17.7)	(25.3)	(19.4)	(6.4)		
v2.2	SE _{an}	Mg ha ⁻¹	4.9950	7.8125	15.3329	14.3956	2.8160		
	%	(17.0)	(17.2)	(24.3)	(18.9)	(5.8)			
v _{NT,DBMA}	SE _{an}	Mg ha ⁻¹	0.9222	1.0558	2.6584	1.8868	0.6641		
	%	(3.1)	(2.3)	(4.2)	(2.5)	(1.4)			
REG	observed	Bias	Mg ha ⁻¹	-0.0474	-0.0663	0.0526	-0.0149	-0.0381	
			%	(-0.2)	(-0.1)	(0.1)	(-2.0e ⁻³)	(-7.9e ⁻³)	
		SE _{obs}	Mg ha ⁻¹	1.0008	1.1681	2.5480	1.8677	0.6354	
		%	(3.4)	(2.6)	(4.0)	(2.4)	(1.3)		
	v.2.3	SE _{an}	Mg ha ⁻¹	1.0952	2.2890	2.6738	1.4216	0.9114	
		%	(3.7)	(5.1)	(4.2)	(1.9)	(1.9)		
v _{NT,MB}	SE _{an}	Mg ha ⁻¹	0.7786	0.7886	1.7889	1.0778	0.6049		
	%	(2.7)	(1.7)	(2.8)	(1.4)	(1.3)			

Table 4. Coverage and failure rates (%) of z-based CI for μ AGB following one- and two-phase designs, under design and model-based inference.

Confidence intervals		Variance estimators								
		SRSwOR			SYS					
		v2.1	v2.2	v2.3	v1.1	v2.1	v2.2	v2.3	$V_{NT,DBMA}$	$V_{NT,MB}$
z-CI										
AU ₁	below	5.08	4.17	6.33	2.80	0.00	0.00	1.64	2.30	6.08
	above	2.41	1.70	5.38	1.15	0.00	0.00	1.62	1.86	6.60
	coverage	92.51	94.13	88.29	96.05	100.00	100.00	96.75	95.85	87.32
AU ₂	below	4.92	4.31	4.27	2.59	0.00	0.00	0.03	3.57	8.42
	above	1.96	1.70	2.99	1.39	0.00	0.00	0.00	3.26	10.34
	coverage	93.12	93.99	92.74	96.02	100.00	100.00	99.97	93.17	81.23
AU ₃	below	7.13	5.22	7.17	3.38	0.00	0.00	2.48	2.91	9.62
	above	2.45	0.79	7.90	1.20	0.00	0.00	1.48	1.88	7.63
	coverage	93.42	93.99	84.93	95.42	100.00	100.00	96.04	95.21	82.76
AU ₄	below	4.72	4.35	11.10	3.22	0.00	0.00	7.18	3.55	13.54
	above	2.13	1.94	10.65	1.86	0.00	0.00	6.72	3.27	13.52
	coverage	93.15	93.71	78.25	94.92	100.00	100.00	86.10	93.18	72.94
Total	below	2.74	2.65	4.34	2.46	0.00	0.00	0.42	2.24	3.46
	above	3.93	3.21	3.04	1.71	0.00	0.00	0.12	1.96	2.96
	coverage	93.33	94.14	92.62	95.84	100.00	100.00	99.46	95.8	93.59
p-CI										
AU ₁	2.5	3.5550	4.1027	0.8549	1.6140	4.5276	4.7102	1.0139	0.5622	0.7051
	97.5	5.8611	5.7195	1.4214	2.1664	5.9426	5.2841	1.1802	1.2632	0.8606
AU ₂	2.5	5.6046	6.1154	1.7211	2.0781	7.0187	7.3558	2.0805	0.669	0.6967
	97.5	9.2443	9.0785	3.0553	2.5852	9.0281	8.2972	2.5417	1.4803	0.8781
AU ₃	2.5	8.4679	11.3577	1.7358	4.4156	13.1642	14.1785	2.4576	1.4696	1.502
	97.5	18.9280	18.2651	4.1505	6.3045	18.8933	16.4977	2.9394	3.9883	2.0769
AU ₄	2.5	10.8278	11.5955	1.1104	3.6526	13.0145	13.676	1.2002	1.0892	0.8795
	97.5	16.8371	16.3899	1.8550	4.5731	16.5894	15.1159	1.6107	2.7593	1.4044
Total	2.5	2.2301	2.3590	0.7528	1.3603	2.7017	2.6332	0.8434	0.3184	0.5251
	97.5	3.3356	3.1932	1.1529	1.5460	3.4582	2.9837	0.972	0.6398	0.6778

4.1 One-phase systematic survey with HT

The bias of the HT estimator was negligible (less than 0.01%) at both stratum and across-stratum level (Table 3). The observed standard error for AU₃ was noticeable larger (8.1%) comparing to the other AUs, which varied between 4.9 to 6.0% of the stratum average. The SE_{an} were 1-6% higher than

SE_{obs} , with the highest SE_{an} (8.5%) occurring in AU₃. The across-stratum SE_{an} (1.4535 Mg ha⁻¹) represented 3.0% of the population mean, and it slightly overestimated SE_{obs} (1.3942 Mg ha⁻¹) by 4%.

The SE_{an} and SE_{obs} at stratum and across-stratum level were not significantly different, the p-CIs covering SE_{obs} (Table 4). The actual coverage rate of the z-CIs for the stratum AGB estimates was between 95.42% and 96.06% for all strata but AU₄, where the coverage was 94.9%. The actual coverage of the z-CI for the across-strata AGB estimate was 95.84%.

The percentage failures from below systematically exceeded the failure rates from above for all strata by 1.7 to 2.8 times, the largest difference occurring in AU₃. The symmetry z-CI improved slightly for the across strata AGB estimate, the from-below failure rates exceeding the from-above failure rates by 44%.

4.2 Two-phase SRSwoR survey with regression estimator

The GREG and REG estimators attained comparable results in terms of bias (Table 3), which varied between -0.02% and 0.21% of the stratum means. Both estimators slightly underestimated the across-stratum μ_{AGB} by 0.04% and 0.05%, respectively.

The observed and analytical standard errors for stratum estimates produced by GREG matched nicely, varying by approximately 2-6%. The largest standard errors ($SE_{obs} = 15.0539$ Mg ha⁻¹, $SE_{an} = 14.0849$ Mg ha⁻¹ for $\hat{v}_{2.1}$ and $SE_{an} = 15.2155$ Mg ha⁻¹ for $\hat{v}_{2.2}$) were observed in AU₃ where they represented 23.8%, 22.3% and 24.1% of the stratum means, respectively,

while for the other strata the standard errors were approximately 16-18% from the corresponding stratum means. The stratum-wise SE_{an} produced by the $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ estimators (Table 3) were also very close, the largest difference of approximately 8% occurring in AU₃.

Comparing to GREG, the SE_{obs} obtained for REG were 65% to 75% smaller, for both stratum and across-strata level. At stratum level, the largest observed standard error (3.6363 Mg ha⁻¹) occurred in AU₃ and represented 5.7% of the stratum mean. The SE_{an} estimates produced by the $\hat{v}_{2,3}$ estimator underestimated SE_{obs} by 6% to 36%, the largest difference occurring in AU₄. The across strata SE_{an} (0.9287 Mg ha⁻¹) underestimated SE_{obs} (1.0047 Mg ha⁻¹) by approximately 8%.

For GREG, the differences between observed and analytical estimates were not found to be significant, the empirical 95% p-CIs obtained from the sampling distribution of $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ estimators covered the Monte Carlo estimate for SE_{obs} at both strata and across-strata levels. The p-CIs constructed using the $\hat{v}_{2,3}$ estimator covered SE_{obs} except for AU₄, where the SE_{obs} was significantly underestimated.

The coverage rates of the z-CIs (Table 4) obtained using the $\hat{v}_{2,1}$ estimator varied between 92.51% (AU₁) and 93.42% (AU₃) for strata, being 93.33% for across-strata level. Using $\hat{v}_{2,2}$ produced slightly wider intervals, between 93.99% (AU₂ and AU₃) and 94.14% for the across-strata estimate. The actual coverage rates constructed using REG and $\hat{v}_{2,3}$ varied between 78.25% and 92.74% at stratum level, the narrowest interval occurring for AU₄. The coverage rate attained at across-strata level was 92.62%.

The failure rates for the stratum-wise z-CI under DBMA inference were highly asymmetrical, the from-below failure rate being 2.2 to 6.6 times higher than from-below. However, at the across-stratum level the situation reversed, the from-above failure rate being 20% higher than the from-below rate. For the MB inference, the failure rates were more

balanced, yet the from-below rates being slightly higher at both the stratum and across-strata levels.

4.3 Two-phase SYS survey with regression estimator

At stratum level, the biases of the AGB estimates varied between 0.0001% and 0.16%, and the population parameter μ_{AGB} was underestimated by the GREG estimator with 0.01%, while being overestimated by the REG with 0.08%. The observed standard errors of GREG and REG were matching well although the MB estimates were systematically higher with 4% to 24% at stratum level and with 4% at across-strata level (Table 3).

The standard errors for GREG produced by the $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ estimators were approximately 6 to 8 times higher comparing to the observed estimates at stratum level, and approximately 5 times higher at the across strata level. However, the estimates produced by $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ were very close, differing by approximately 2% to 5% at stratum level and by 9% at across-strata level. The successive differencing estimates deviated by 0.5% to 21% from SE_{obs} at stratum level, the largest difference occurring in AU₁. The across-strata $\hat{v}_{NT,GREG}$ estimate was approximately 9% higher than the SE_{obs} (Table 3).

Compared to the $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ estimators, the standard errors produced by the $\hat{v}_{2,3}$ estimator were approximately 3.4 to 10.4 times smaller. The model-based SE_{obs} in AU₁, AU₂ and AU₃ were overestimated with 10%, 96%, and 4% respectively, while in AU₄ the SE_{obs} was underestimated by approximately 24%. At across-strata level, the $\hat{v}_{2,3}$ estimator gave a SE_{an} overestimated SE_{obs} by 40%. The SE_{an} produced at stratum level by the successive differencing estimator $\hat{v}_{REG,NT}$ underestimated the SE_{obs} values of REG with 22% to 42% (the largest difference occurring in AU₄), and with approximately 5% at across-strata level (Table 3).

The standard errors produced by the design-based variance estimators $\hat{v}_{2.1}$ and $\hat{v}_{2.2}$ were significantly different from the observed standard errors, the p-CIs constructed from the SE_{an} distributions missing the SE_{obs} estimates at both stratum and across-strata level. No significant differences occurred between SE_{obs} for GREG and SE_{an} estimates produced by the $\hat{v}_{GREG,NT}$ estimator. For REG, the empirical p-CIs obtained from the sampling distribution of $\hat{v}_{2.3}$ indicated that the differences between SE_{obs} and SE_{an} were significant in all situations except for AU₃. However, the p-CI constructed using the $\hat{v}_{NT,REG}$ covered the across-strata SE_{obs} (Table 4).

The coverage rates of the z-CIs constructed for GREG using $\hat{v}_{2.1}$ and $\hat{v}_{2.2}$ estimators were 100%. When using the $\hat{v}_{NT,GREG}$ estimator, the confidence interval statements matched better the nominal coverage of 95%, varying between 93.17% (AU₂) and 95.85% (AU₁).

The confidence interval for REG built with $\hat{v}_{2.3}$ came close the 95% nominal rate for AU₁ (96.75%) and AU₃ (96.04%), while for AU₂ and AU₄ the coverage rates were 99.97% and 86.10%, respectively. Using the $\hat{v}_{NT,REG}$ estimator produced z-CIs having the actual coverage varying between 72.94% (AU₄) and 87.32% (AU₁), and the coverage for the across-strata $\hat{\mu}_{AGB}$ was 93.59% (Table 4).

The failure rates of the z-CIs built using $\hat{v}_{2.3}, \hat{v}_{NT,GREG}$ and $\hat{v}_{NT,REG}$ were slightly asymmetrical, the from-below failures occurring more frequently.

4.4 Accuracy assessment for sampling strategies

From equation 14, the relative accuracy (*eff*) of DBMA and DBHT strategies varied between 0.43 (AU₁) and 0.48 (AU₂) at stratum level, and at the across-strata *eff* was 0.44. Using the MB strategy was slightly less efficient, the *eff*-values varying from 0.46 (AU₄) to 0.57 (AU₁), with an overall *eff*=0.46. In other words, the accuracy of the DBMA strategy was

approximately 52-57% better compared to the DBHT strategy, while using the MB strategy improved the accuracy with 43-54%.

The relative efficiency of the SYS_2 and $SRSwoR_2$ designs ($deff$) from equation 15 was approximately 0.02 at stratum level, with an overall design effect $deff = 0.05$. This translates into a reduction of the observed standard errors by approximately 6.6 times at stratum level, and by 4.8 times at the across-strata level.

4.5 Cost-efficiency assessment

The results of the cost-efficiency assessment are presented in Figure 4 and Figure 5. The dot labels in the plots specify the sampling intervals (km), and the curves indicate the relationships between the across-strata standard errors (% of the population mean) and the inventory costs expressed in equation 17, where c_{Field}^0 was assigned the value 1. The curves were generated using the one-term power function $SE_{obs} = aC_{Total}^b$, the parameters estimation being done using nonlinear least-squares fitting.

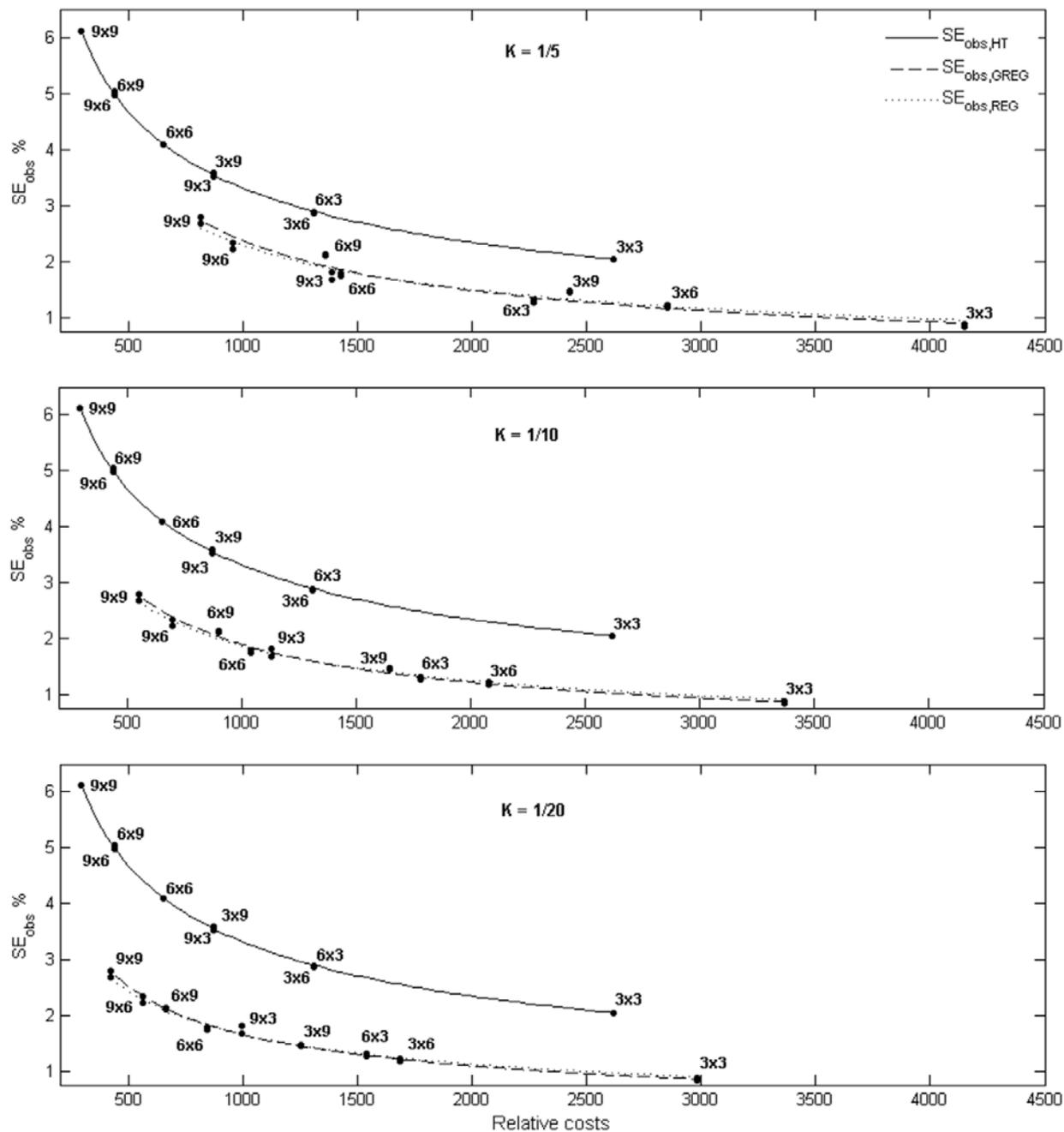


Figure 4 The relationship between observed standard errors ($SE_{obs} \%$) and inventory costs under systematic sampling. The labelled dots indicate the sampling intervals (e.g., 6x3 km), and the values on x-axes are the inventory costs relative to the average cost of a field plot, as given by the *term inside parentheses* in equation 17.

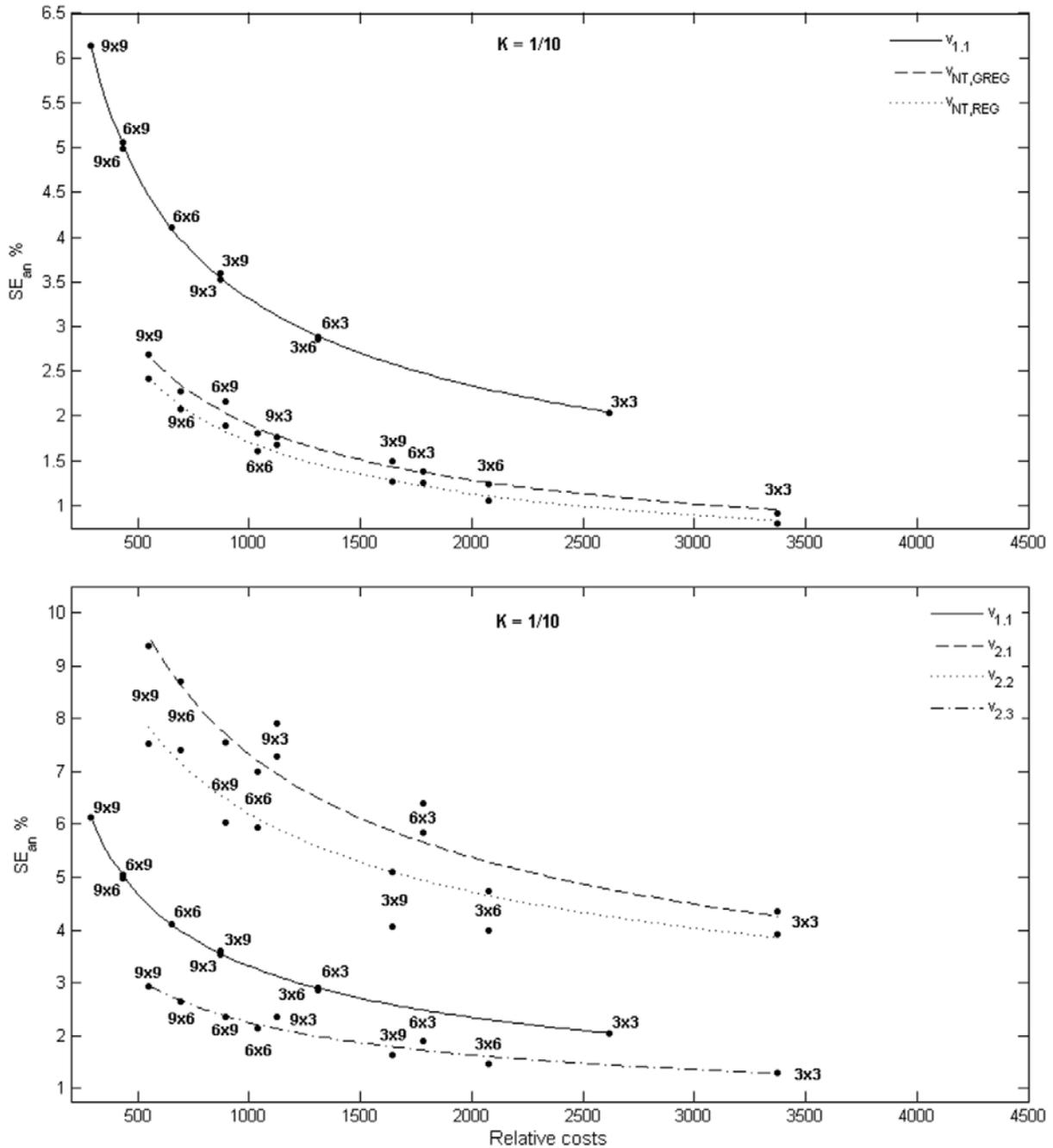


Figure 5 The relationship between analytical standard errors ($SE_{obs} \%$) and inventory costs under systematic sampling. The sampling intervals are indicated by the labelled dots (e.g., 6x3 km). To obtain the total survey costs, the relative costs on the x-axis should be multiplied by the average cost of a field plot from equation 17.

Analyzing the regression line slopes ($\hat{\alpha}$) between estimated precision and log-transformed relative costs indicated that the most sensitive estimator to cost changes and implicitly to sample size variation was the $\hat{v}_{1,1}$ estimator ($\hat{\alpha} = -1.870$), followed by the $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ estimators ($\hat{\alpha} = -1.4282$ and $\hat{\alpha} = -0.0225$, respectively), while $\hat{v}_{2,3}$, $\hat{v}_{NT,GREG}$ and $\hat{v}_{NT,REG}$ estimators

were less affected ($\hat{\alpha} \approx -0.009$). The observed precision of the GREG was slightly more influenced by varying the sampling size ($\hat{\alpha} = -1.0620$) compared to REG ($\hat{\alpha} = -0.9679$).

The ALS-aided systematic surveys were in general more precise than the ground-based inventory except when using wide sampling intervals of 6x9, 9x6 and 9x9 km, when the precision of the ground-based survey approximately was 5% to 35% higher (Figure 3). The numerical gradients of the curves presented in Figures 4 and 5 indicated that the strongest change in precision, given a change in cost, occurred at the 6x9 km interval for $\hat{v}_{2.1}$, $\hat{v}_{2.2}$ and $\hat{v}_{2.3}$ estimators, and at the 9x9 km intervals in the rest of the cases.

5. Discussion

In this study we performed an empirical validation of the variance estimators proposed by Gregoire et al. (2011) and Ståhl et al. (2011) for post-stratified AGB estimation, following the simulation approach described by Ene et al. (in submission). In order to cope with the variance overestimation described in Ene et al. (in submission), successive differences variance estimators were also tested under two-phase systematic sampling designs. Using simulated sampling, the efficiency of the one- and two-phase systematic designs were compared, and the effect of using a systematic design instead of SRSwoR was assessed. Finally, we performed an empirical cost efficiency analysis of the ALS- and field based surveys considering systematic sampling designs with various sampling intervals.

The good performance of the $\hat{v}_{2.1}$, $\hat{v}_{2.2}$ and $\hat{v}_{2.3}$ estimators under SRSwoR₂ reported by Ene et al. (in submission) was also noticed in this study. All estimators slightly underestimated the observed standard errors, but the differences were not found to be significant using the percentile-based confidence intervals. However, the z-CIs constructed for the stratum and the population means using the design-based estimators matched more closely the 95% nominal coverage.

Under SYS_2 , the observed standard errors of GREG and REG were greatly reduced as a result of having little variability among systematic samples. The gain in observed precision was not noticed for the $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ estimators which significantly overestimate the observed standard errors by a factor of 5 to 8. This resulted into much wider confidence intervals, making difficult, for instance, to detect significant changes in biomass under shorter time intervals, which constitutes an impediment for e.g. REDD applications. Using the successive differences variance estimators improved noticeably the precision, however, only $\hat{v}_{NT,GREG}$ provided valid inference at both stratum and across-strata level.

Due to the proportional-to-size type of sample allocation scheme, the stratum size (and implicitly the sample sizes) influenced the precision of GREG and its variance estimators $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$, the estimation being more precise in larger strata (AU_1 and AU_2) and less precise in the smaller ones (AU_3 and AU_4). Under MB inference, the strata sizes did not influence the performance of the $\hat{v}_{2,3}$. There was no obvious dependency between the performances of $\hat{v}_{NT,GREG}$ and sample size. However, the results obtained with $\hat{v}_{NT,REG}$ were apparently better in the larger AUs. This result also shows that another advantage of having a sampling simulator is the possibility to assess various allocation schemes in the planning phase for design-based sampling strategies.

The poststratification was used as a tool for providing uncertainty estimates at administrative unit level. The across-stratum analytical standard errors provided by $\hat{v}_{2,1}$ and $\hat{v}_{2,2}$ were close to the ones reported by Ene et al. (*in submission*), suggesting that poststratification by administrative units does not contribute to improving the overall precision. The same is evident for the one-phase sampling, where the poststratified estimation produced only marginal gains in precision. However, poststratified estimation improved greatly the performance of the model-based estimator calculated as the ratio-of the average cluster total and average cluster size, which are positively correlated random variables.

Despite of a negligible bias, the ratio estimator is known to be more precise than the Horvitz-Thompson estimator used to estimate the total AGB under DBMA.

According to our results, the best simulation results were obtained using the GREG estimator with its variance estimated by the successive differences estimator $\hat{v}_{NT,GREG}$ under DBMA inference. The second choice would be using a model-based strategy including REG and the $\hat{v}_{2,3}$ variance estimator. Unfortunately, both choices come with certain risks due to using biased estimators ($\hat{v}_{NT,GREG}$ and REG) in a real applications. The simulation results cannot be use to correct for biases associated with the artificial population, but they can be used to minimize the risks of choosing an inappropriate estimator.

In order to reduce the inventory costs, our analysis indicated that it is more advantageous (in terms of standard error) to lower the subsampling intensity. However, reducing the size of the field sample should be done with precaution in the case of design-based inference because this can seriously affect the inference if the number of SSUs becomes too small (Särndal et al., 1992, p. 398; Gregoire et al., 2011).

In figures 4 and 5 it was exemplified how an appropriate sampling strategy can be found empirically using simulated sampling, considering either budgetary constraints (fixing an x-axis value) or precision requirements (using a fixed y-axis value). However, the results of such an exercise are valid only for the artificial population we have studied and generalization of these results to populations having different characteristics (e.g. different trends) or to different sampling strategies would not be advisable. Nevertheless, considerable empirical knowledge can be gained by performing this type of studies for various populations. According to our knowledge, the ALS-related costs are almost independent of the strip width. This suggests that the ALS sampling intensity (and thus precision) can be increased by using wider strips (e.g., flying at higher altitudes) without increasing the survey costs. The current trend in sensor development indicates that future instruments might be

operated from higher altitudes and with higher pulse repetition frequencies. Thus, it is expected that the swath width may increase significantly without increasing flight times. As an example, the HC study was conducted in 2006 with state-of-the-art instruments and had a swath width of approximately 500 m (see details in Gobakken et al. in press). A similar ALS-aided survey conducted in Tanzania in 2012 will have swaths with a width of more than 1000 m (E. Næsset, pers. comm.). Thus, a gain in sampling intensity by a factor of 2 has been achieved over a six year period mainly due to technological advances.

The systematic design boosted the efficiency of the regression estimators, but without information regarding the real precision level provided by the simulated sampling it could be wrongly concluded that the estimated precision of the ALS-aided survey would be at most comparable with the ground-based inventory, and thus not cost-efficient. For instance, Stephens et al. (2012) claimed that the maximum improvement in precision of a double-sampling ALS-aided survey relative to ground-based inventory is limited by the amount of variation explained by the regression model. This assertion is only partially true, because the variance estimation used by Stephens et al. (2012) holds for *SRSwoR* and does not account for eventual improvements due to using a systematic probability sample.

To conclude, uncertainty estimation following two-phase ALS-aided systematic AGB surveys can lead to wrong conclusions when using variance estimations tailored for *SRSwoR*. This translates into overestimated errors which would require higher sampling intensity (thus higher costs) than necessary for attaining a specific level of precision. Finding a suitable sampling strategy at the planning phase or a ‘good’ estimator for inference in a complex survey by the means of simulated sampling can provide the means for more efficient resource allocation in large area forest inventories.

Acknowledgments

This research has been funded by the Research Council of Norway (project no. 166482/i10, no. 184636/S30 and no. 180645). The authors would like to thank the Norwegian National Forest Inventory for collection of field data and Blom Geomatics, Norway, for collection and processing of the airborne laser scanner data.

References

- Andersen, H.-E. & Breidenbach, J. (2007). Statistical properties of mean stand biomass estimators in a LiDAR-based double sampling forest survey design. *International Archives of Photogrammetry and Remote Sensing*, Volume XXXVI, Part 3/W52. Espoo, Finland, September 12-14, 2007.
- Andersen, H.-E., Barrett, T., Winterberger, K., Strunk, J. & Temesgen, H. (2009). Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using airborne lidar and field plot data in a model-assisted sampling design. *Proceedings of the IUFRO Division 4 Conference: "Extending Forest Inventory and Monitoring over Space and Time,"* Quebec City, Canada, May 19-22, 2009.
- Anon. (1999). Pinnacle user's manual. San Jose, CA: Javad Positioning Systems.
- Asner G.P, Hughes R.F, Varga T.A, Knapp D.E & Kennedy-Bowdoin T. (2009). Environmental and biotic controls over aboveground biomass throughout a rainforest. *Ecosystems*, doi: 10.1007/s10021-008-9221-5.
- Asner, G. (2009). Tropical forest carbon assessment: Integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4, doi: 10.1088/1748-9326/4/3/034009.
- Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., & Kimes, D.S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, 112, 3876-3890.
- Bjørdal, I. & Bjørkelo, K. (2006). AR5 klassifikasjonssystem. Klassifikasjon av arealressurser. [*AR5 classification system. Classification of areal resources*]. Ås, Norway: Norwegian Forest and Landscape Institute. (In Norwegian with English summary).
- Cochran, W.G. (1977). *Sampling Techniques* (3rd ed). New York: Wiley.

- Corona, P., Fattorini, L. (2008), Area-based lidar-assisted estimation of forest standing volume. *Canadian Journal of Forest Research*, 38, 2911-2916.
- Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B. & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography*, 12, 147-159.
- Ene, L.T., Næsset, E., Gobakken, T., Gregoire, T.G., Ståhl, G., Holm, S. & Nelson, R, Assessing the accuracy of regional LiDAR based biomass estimation using a simulation approach. (*In submission*).
- Fahey, T.J., Woodbury, P.B., Battles, J.J., Goodale, C.L., Hamburg, S.P., Ollinger, S.V., & Woodall, C.W., (2010). Forest carbon storage: ecology, management, and policy. *Frontiers in Ecology and the Environment*, 8, 245-252.
- Gibbs, H.K., Brown, S., Niles, J.O., & Foley, J.A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, doi:10.1088/1748-9326/2/4/045023.
- Gama, F.F., dos Santos, J.R. & Mura, J.C. (2010). Eucalyptus biomass and volume estimation using Interferometric and polarimetric SAR data. *Remote Sensing*, 2, 939-956.
- Gobakken, T., Næsset, E. & Nelson, R. (2006), Developing regional forest inventory procedures based on scanning LiDAR. *Proceedings of the International Conference Silvilaser 2006* (pp. 99-104). Matsuyama, Japan: Japan Society of Forest Planning, Forestry and Forest Products Research Institute and Ehime University.
- Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., Holm, S., Ørka, H.O. & Astrup, R. (2011). Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment*, in press.

- Gonzales, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., & Palace, M. (2010). Forest carbon densities and uncertainties from LiDAR, QuicBird, and field measurements in California. *Remote Sensing of Environment*, 114, 1561-1575.
- Gregoire, T.G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28, 1429-1447.
- Gregoire, T.G. & Schabenberger, O. (1999). Sampling-skewed biological populations: behaviour of confidence intervals for the population total. *Ecology*, 80, 1056-1065.
- Gregoire, T.G. & Valentine, H., 2008, Sampling strategies for natural resources and the environment. (Boca Raton, Fla.: Chapman & Hall/CRC).
- Gregoire, T.G., Ståhl, G., Næsset, E., Gobakken, T. & Nelson, R. (2011). Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41, pp. 83-95.
- Heikkinen, J. (2006). Assessment of uncertainty in spatially systematic sampling. In Kangas, A. & Maltamo, M. (Eds.), *Forest Inventory: Methodology and Applications* (pp. 155-178). Dordrecht: Springer.
- IPCC (2000). Land-use, land-use change, and forestry: A special report of the IPCC. Cambridge and New York: Cambridge University Press.
- IPCC (2006). Good practice Guidance for Land Use, Land-Use Change and Forestry. (<http://www.ipcc-nggip.iges.or.jp/public/gpplulucf/gpplulucf.html>) (last accessed 2-Jan-2012).
- Kangas, A. (2006) Model-based inference. In Kangas, A. & Maltamo, M. (Eds.), *Forest Inventory: Methodology and Applications* (pp. 39-52). Dordrecht: Springer.
- Köhl, M., Magnussen, S., & Marchetti, M. (2006). *Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory*, Berlin: Springer.

- Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., & Brower, S.T. (2002). Lidar remote sensing of aboveground biomass in three biomes. *Global Ecology and Biogeography*, 11, 393-400.
- Marcell, W., Eriksson, M., & Popescu, S. (2009). Systematic sampling of scanning lidar swaths. In Popescu, S., Nelson, R., Kaiguang, Z. and Neusenschwander, A. (Ed.), *Proceedings of Silvilaser 2009*, 14-16 October 2009, Texas A&M University, College Station, USA, (pp. 184-192).
- Marklund, L.G. (1988). Biomass functions for pine, spruce and birch in Sweden. Umeå: Swedish University of Agricultural Sciences, Department of Forest Survey (In Swedish).
- McCulloch, C.E. & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London, UK: Chapman and Hall, 532 p.
- Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19, 164-179.
- Næsset, E. (2005). Towards a laser-scanner based biomass monitoring system. In K. Hobbelstad (Ed.), *Forest Inventory and Planning in the Nordic Countries. Proceedings of the SNS Meeting* (pp. 117-119). Sjusjøen, Norway: Report 09/05 of the Norwegian Institute of Land Inventory.
- Næsset, E., Gobakken, T. & Nelson, R. 2009. Sampling and mapping forest volume and biomass using airborne LIDARs. *Proceedings of the Eight Annual Forest Inventory and Analysis Symposium*, October 16-19, 2006, Monterey, CA, USA. General Technical Report WO-79, United States Department of Agriculture, Forest Service, Washington DC, pp. 297-301.
- Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing*, 32, 473-501.

- Næsset, E. & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112, 3079-3090.
- Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G. & Weydahl, D. (2011). Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. *Remote Sensing of Environment*, 115: 3599-3614.
- Nelsen, R.B. (2009). *An introduction to copulas* (2nd ed). New York: Springer.
- Nelson, R., Parker, G., & Hom, M. (2003). A portable airborne laser system for forest inventories. *Photogrammetric Engineering & Remote Sensing*, 69, 267-273.
- Nelson, R., Short, A., & Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LiDAR. *Scandinavian Journal of Forest Research*, 19, 500-511.
- Nelson, R., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M. L., Campbell, R., et al. (2007). Investigating RaDAR–LiDAR synergy in a North Carolina pine forest. *Remote Sensing of Environment*, 110, 98–108.
- Nelson, R., Næsset, E., Gobakken, T., Ståhl G. & Gregoire, T.G. (2008). Regional forest inventory using an airborne profiling LiDAR. *Journal of Forest Planning*, 13, 287-294.
- Nelson, R., Boudreau, J., Gregoire, T.G, Margolis, H., Næsset, E., Gobakken, T., & Ståhl, G. (2009). Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, 39, 862-881.
- Ørka, H.O., Wulder, M.A., Gobakken, T. & Næsset, E. (2012). Subalpine zone delineation using LiDAR and Landsat imagery. *Remote Sensing of Environment*, 119, 11-20.

- Parker, R.C. & Evans, D.L. (2004). An application of lidar in a double-sample forest inventory. *Southern Journal of Applied Forestry*, 29, 40-47.
- Särndal, C.E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.
- Schreuder, H.T., Gregoire, T.G. & Weyer, J.P. (2001). For what applications can probability and non-probability sampling be used? *Environmental Monitoring and Assessment*, 66, 281-291.
- Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., & Nelson, R. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41, 96-107.
- Stephens, P.R., Kimberly M.O, Beets, P.N., Paul, T.S.H., Searles, N., Bell, A., Brack, C., & Broadley, J. (2012). Airborne scanning LiDAR in a double sampling forest carbon inventory. *Remote Sensing of Environment*, 117, 348-357.
- Solberg, S., Astrup, R., Bollandsås, O.M., Næsset, E., & Weydahl, D.J. (2010). Deriving forest monitoring variables from X-band InSAR SRTM height. *Canadian Journal of Remote Sensing*, 36, 68-79.
- 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, 2353-2360. *Canadian Journal of Remote Sensing*, 36, 68-79.
- Sun, G., Ranson, K., Z. Guo, Z., Zhang Z., Montesano, P., & Kimes, D. (2011). Forest biomass mapping from lidar and radar synergies, *Remote Sensing of Environment*, **115** (11), pp. 2906-2916.
- Tomter, S.M., Hysten, G., & Nilsen J-E. (2010). National Forest Inventory Reports: Norway. In Topmppo, E., M., Gschwantner, T., Lawrence, M. and McRoberts, R.E. (Eds.),

National Forest Inventories- Pathways for Common Reporting (pp. 411-424).

Dordrecht: Springer

Wolter, K.M. (2007). Introduction to variance estimation, (2nd ed). New York: Springer.

Zhang, L., 2000, Post-stratification and calibration – a synthesis. *Journal of the American Statistical Association*, 54, 178-187.

Zheng, B. & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in medicine*, 19, 1171-1781.

Paper III

1 **Model-based inference for k -nearest neighbours predictions**
2 **using a canonical vine copula**

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

Prepared by:

Liviu Theodor Ene^{1*}, Erik Næsset¹ and Terje Gobakken¹

Address:

1. Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway,

Liviu.Ene@umb.no; Terje.Gobakken@umb.no; Erik.Naesset@umb.no.

*Corresponding author

Liviu Theodor Ene

Email: Liviu.Ene@umb.no

Phone: +47-64965766

Fax: +47-64968890

22 **Abstract**

23 The k -nearest neighbor imputation (k -NN) methods combine ground sample data provided by
24 traditional forest inventories and auxiliary information for non-parametric forest resource
25 estimation at various geographical scales. Using similarity measures calculated in the
26 auxiliary variable space, the inventory data can be imputed to locations where only the
27 auxiliary data is available. In this study, auxiliary data consisting in Landsat 5 TM satellite
28 imagery and a digital terrain model were used to perform nearest neighbor imputations of
29 plot-level above ground biomass. The study introduces a model-based inference approach
30 where a superpopulation model in the form of a canonical vine copula function is first
31 constructed from the empirical data, and then new samples are independently generated from
32 the copula model and used to perform the k -NN predictions. The method allows constructing
33 the sampling distribution for the k -NN prediction errors for assessing the statistical properties
34 of the k -NN estimator. Using a data-splitting procedure, the performance of the copula-based
35 approach was assessed against pair-bootstrap resampling. The imputations were performed
36 using k (the number of neighbours) =1 and by using optimal k -values selected according to a
37 bias-minimizing criterion in the k -NN predictions. The results indicated that the copula-based
38 approach produced confidence intervals with better coverage properties compared to pair-
39 bootstrap resampling, and it significantly reduced the root mean squared error. The
40 improvements of the copula-based approach were due to significant bias reduction, while the
41 standard errors were higher compared to the bootstrap. The best results in terms of coverage
42 properties of the confidence intervals were obtained combining the copula approach and
43 nearest-neighbor imputations with $k=1$. Using k -values produced by the bias-minimizing
44 criterion increased the accuracy of both methods, especially due to the bias reduction in the
45 case of bootstrap and by reducing the standard errors of the copula approach.

46 **1. Introduction**

47 To meet the demand for timely and accurate information about the forest ecosystems, forest
48 management and planning activities must cover a broad range of objectives which require
49 various types of information at different geographical levels (Temesgen et al., 2007).
50 Combining auxiliary information (e.g. remote sensing data, land-use and vegetation digital
51 maps, digital terrain models, etc.) with ground observations allows monitoring and
52 assessment of forest resources at different geographical scales (McRoberts & Tomppo, 2007;
53 McRoberts, 2008; Tomppo et al., 2008*ab*; Corona, 2010; McRoberts et al., 2010*ab*).

54 Allowing for univariate and multivariate predictions of continuous and categorical
55 variables, nearest neighbor methods have been deemed useful and cost-efficient solutions for
56 supporting the national forest inventory programs (Tomppo, 1991; Tomppo et al., 2008;
57 McRoberts, 2008), as well as for mapping forest attributes and for providing local estimates
58 of forest resources (LeMay & Temesgen, 2005; Chirici et al., 2008; Hudak et al., 2008;
59 Koistinen et al., 2008; LeMay et al., 2008; McRoberts, 2008; Maltamo et al., 2009;
60 McRoberts, 2011).

61 Denoting the field survey data as response variables (Y) and the auxiliary information
62 as feature data (X), the set of observations containing both X and Y variables is characterized
63 as the reference data set, and the set of observations having X variables but missing the Y will
64 be referred to as the target data set. With k near-neighbour (k -NN) imputations, the missing
65 attributes in a target data set are predicted as linear combinations of the attributes from the k -
66 nearest reference observations found in a reference data set, the distances between the target
67 and reference observations being calculated in the feature space. A thorough review of the
68 nearest neighbor methods with practical applications is given by Tomppo et al. (2008*ab*),
69 Eskelson et al. (2009) and McRoberts et al. (2010*a*).

70 The statistical properties of the nearest neighbor estimators have been addressed under
71 the design-based inferential framework (Shao, J., 2009; Baffetta et al., 2009,2011), as well as
72 under the model-based framework (Kim & Tomppo, 2006; McRoberts, 2006; McRoberts et
73 al., 2007; Koistinen et al., 2008; Magnussen et al., 2009,2010a; McRoberts, 2011; McRoberts
74 et al., 2011; Rätty & Kangas, 2012). Under the design-based inference, population elements
75 and population parameters are fixed quantities, the estimates of the population parameters
76 being functions of the sample selection probability (Gregoire & Valentine, 2008, §2). Under
77 the model-based framework, a finite population is seen as a random draw from a
78 superpopulation model (*sensu* Särndal et al., 1992, §14.5), thus the population elements and
79 population parameters are random variables. The sampling distribution of the model-based
80 estimator is defined by the estimates obtained from a long series of drawings from the
81 superpopulation model, and the inference relies on the underlying superpopulation model, not
82 on the design used for collecting the sample. Hence, large biases can be introduced if the
83 model is poorly specified, but significant gains in precision can be obtained in the presence of
84 small sample sizes compared to the design-based inference. Further insights into design- and
85 model-based inference can be found in Särndal et al. (1992), Gregoire (1998), Kangas (2006)
86 and McRoberts (2010).

87 The accuracy of k -NN imputations is influenced by factors such as gaps in the
88 reference feature space, absence of neighbours outside of reference feature space, and the
89 number of neighbours the feature used for imputations (Stage & Crookston, 2007;
90 McRoberts, 2009; Magnussen et al., 2010b). Although approximations of the prediction
91 errors have been worked out, the extrapolation bias of the k -NN predictions is difficult to
92 assess because the nearest neighbor imputations are inherently biased due to their non-
93 parametric nature (Korhonen & Kangas, 1997; Magnussen et al., 2010b). However, when a
94 strong linear relationship exists between X and Y , the bias of the k -NN imputations is

95 expected to be small (Rancourt et al., 1994), however, assumptions about the relationships in
96 the data should not be made *a priori* when using non-parametrical methods. Reducing the
97 bias has been addressed by using a small k -value or selecting variables with good coverage of
98 the feature space (Stage & Crookston, 2007), applying various weighting schemes
99 (McRoberts, 2009; Tomppo & Halme, 2004) or using model-based calibration (Magnussen et
100 al., 2010b).

101 The inference in survey sampling is required not only at the level of the entire
102 population, but also at the level of sub-population (or domains) of various sizes (e.g. regions,
103 sub-regions, counties, municipalities, individual properties or forest stands). For small-
104 domain estimation (or also small-area estimation when the domains have a geographical
105 extent (Särndal et al., 1992, p. 386-387)), the number of sample observations is likely to be
106 low or even to lack entirely, which justifies using the model-based estimators (McRoberts,
107 2011; McRoberts et al., 2011).

108 Resampling methods like bootstrap (Efron, 1979) and jackknifing (Quenouille, 1949)
109 have been considered as non-parametric alternatives to the model-based estimators for
110 assessing the prediction errors of the nearest neighbor estimates (Chen & Shao, 2001;
111 McRoberts et al., 2011). Due to the good performance and ease of implementation, pair-
112 bootstrap resampling (Freedman, 1981) has been favoured to jackknifing and has been a
113 feasible substitute for more complex parametric estimators (McRoberts, 2011; McRoberts et
114 al., 2011). However, the bootstrap approach has certain limitations due to the requirements of
115 preserving the sampling design when selecting bootstrap samples, and of having independent
116 and identically distributed (*iid*) observations (Lahiri, 2003; McRoberts et al., 2011).

117 In this study we propose a model-based approach for assessing the uncertainty of the
118 k -NN predictions for small-domains, where the model-based sampling distribution of the k -
119 NN prediction errors is derived using simulated sampling from a superpopulation model

120 represented by a copula function. Copulas are popular modelling tools in actuarial sciences,
121 and recently they were introduced in forestry applications. Among the applications of copulas
122 in forestry can be mentioned modelling of tree diameters, heights and volumes (Wang et al.,
123 2008; 2010), stochastic modelling of regeneration (Miina & Heinonen, 2008), simulation of
124 forest stand structures (Kershaw et al., 2010), or for estimating shrub cover in riparian forests
125 (Eskelson et al., 2011). Recently, Ene et al. (*in submission*) used Gaussian copulas for
126 generating ground-truth populations for simulation studies related to large-area LiDAR-based
127 biomass surveys. For an in-depth treatment of copulas see Embrechts et al. (2002), Nelsen
128 (2006), Genest & Favre (2007), Kojadinovic & Yan (2010) and Schepsmeier & Brechmann
129 (2011).

130 Copulas are mathematical functions which allow constructing multivariate
131 distributions by modelling the dependencies between univariate marginals. Although a
132 plethora of bivariate copula functions have been documented, only a few can cope with high
133 dimensional datasets (Genest et al., 2009). Recently, new approaches for building
134 multivariate copulae have been developed (Aas et al., 2009) and made available to
135 practitioners through specialized software (Schepsmeier & Brechmann, 2011). This opens the
136 possibility for constructing multivariate copula models for high-dimensional empirical data
137 sets containing forest inventory data and auxiliary information.

138 The objective of this study was to assess the validity of model-based inference
139 following nearest-neighbor predictions for small domains (*sensu* Särndal et al., 1992, p. 386-
140 387). The uncertainty estimation was performed using two methods: (1) pair-bootstrap
141 resampling (McRoberts et al., 2011; McRoberts, 2011), and (2) a copula-based approach,
142 where a canonical vine copula (Aas et al., 2009) played the role of a superpopulation model
143 generating *iid* observations used to perform the imputations. For both methods, the k -NN
144 imputations were performed using (1) $k=1$ and (2) by applying a bias-minimizing criterion for

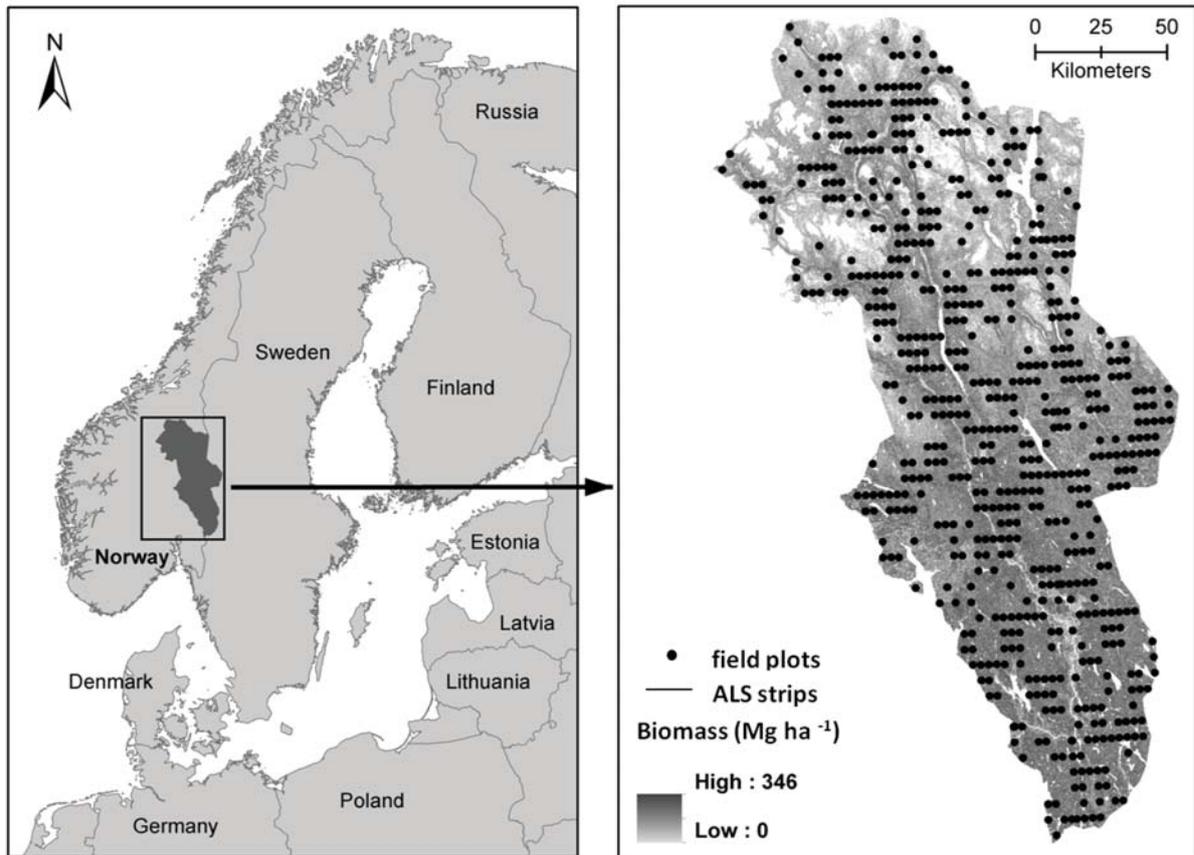
145 selecting the k -value. The domains defined in our study do not have a geographical extent.
146 They were created by splitting the empirical observations into reference and target datasets.
147 Also, it was assumed that the sample data set does not contain units from the respective
148 domains. The assessment was performed by comparing (1) the estimation precision and
149 accuracy (in terms of bias, standard errors and root mean squared errors), and (2) the
150 coverage rates of the confidence intervals produced using bootstrap replicates and copula
151 samples.

152

153 **2. Material**

154 The material was acquired across the Hedmark County (HC) located in south-eastern Norway
155 (Figure 1). The county has a land area of 27 399.72 km² and it comprises a large variety of
156 forest types and geomorphologic conditions (Gobakken et al., *in press*). The dominant tree
157 species in the study area are Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus*
158 *sylvestris* L.).

159 In this study we used three data sets representing field inventory data, satellite
160 imagery, and a digital terrain model (DTM). A more detailed presentation of the datasets is
161 given in Ene et al. (*in submission*) and Gobakken et al. (*in press*).



162
 163 **Figure 1 Geographical location of the field data**
 164

165 ***2.1 Auxiliary data***

166 The auxiliary dataset contains satellite imagery and terrain elevation. The satellite data
 167 consists of three nearly cloud-free Landsat 5 TM images acquired in June 2007 (Table 1).
 168 The georeferencing errors of the radiometrically corrected images were less than 15m for
 169 95% of the pixels. The images were converted to at-satellite reflectance and atmospherically
 170 corrected using the COST model (Chavez, 1996).

171

172 **Table 1. Acquisition parameters for Landsat 5 TM imagery**

Satellite image	Date	Path/row	Sun elevation (deg)	Sun azimuth (deg)	Spatial resolution (m)
1	03-June-2007	197/16	49.6	162.2	25.0
2	03-June-2007	197/17	50.0	162.0	30.0
3	10-June-2007	198/16	50.1	161.9	25.0

173

174 The elevation data consisted of a raster DTM produced by the Norwegian Mapping
 175 Authority at a 25 m spatial resolution, providing full coverage of the HC area. The altitudinal
 176 range of HC spans from 102 m a.s.l. to 2177 m a. s. l, the elevations being higher in the
 177 northern region which is a mountain area.

178

179 **2.2 Field data**

180 The field observations were acquired by the Norwegian National Forest Inventory (NFI)
 181 program from fixed-area circular plots of 250m² displaced in 3x3 km grid covering the entire
 182 country except the areas above the coniferous tree line where the grid spacing is 3x9 km. The
 183 plot measurements used in this study were collected during the years 2005-2007 and contains
 184 662 plots (Table 2) located on every second NFI grid line. More details regarding the field
 185 plot selection can be found in Gobakken et al. (*in press*) and Ene et al. (*in submission*).

186 On each plot, trees with breast height diameter (*dbh*) ≥ 5 cm were callipered, and
 187 approximately ten sample trees were selected proportional to stem basal area for height
 188 measurements (*h*). A detailed description regarding the NFI estimation methods is given by
 189 Tomter et al. (2010) and Gobakken et al. (*in press*). The total aboveground dry biomass
 190 (AGB) of living trees with $h \geq 1.3$ m was estimated using tree-species specific allometric
 191 equations (Marklund, 1988), and the plot-wise AGB estimates were obtaining by summing up
 192 the tree-level AGB estimates.

193

194 **Table 2. Biomass distribution by field plots and forest category**

Forest category	Plots by category		AGB (Mg ha ⁻¹)		Elevation (m a.s.l.)	
	No. of plots	%	mean	sd ⁽¹⁾	mean	sd
Productive forests	440	66	64.67	58.97	489.4	183.9
Nonproductive forest ⁽²⁾	192	29	26.05	25.59	752.6	201.1
Developed areas	30	5	27.75	34.64	211.5	99.9
Total	662	100	51.72	53.65	553.1	232.7

195 ⁽¹⁾ standard deviation (Mg ha⁻¹); ⁽²⁾ annual growth < 1m³ year⁻¹ ha⁻¹

196

197 The plots were mapped using differential Global Positioning System (GPS) and
 198 Global Navigation Satellite System (GLONASS) measurements acquired with dual-
 199 frequency Topcon LegacyE receivers. The locations of the base stations were determined
 200 such that the distances to the field plots were at most 50 km. The positioning error reported
 201 by the Pinnacle 1.0 post-processing software (Anon., 1999) varied between 0 and 2 m, with
 202 an average of 0.05 m.

203

204 **3. Methods**

205 The nearest neighbor imputations can benefit from a careful selection of the X variables
 206 (McRoberts 2002, 2008), and several feature selection procedures have been devised
 207 (Tomppo & Halme, 2004; Packalén & Maltamo, 2006; McRoberts 2008; Walter et al., 2008;
 208 Tomppo et al., 2009). For the purpose of this study, the variable selection has less relevance
 209 because the choice of the features would equally affect both inferential approaches. After
 210 preliminary analysis (not presented), the selected X variables consisted of the radiometric
 211 information from the satellite imagery (bands 1 to 5 and 7) and elevation values from the
 212 DTM raster. The Y variable was considered to be the AGB determined for each plot. The
 213 errors related to mismatching of circular NFI plots and auxiliary raster data and the
 214 estimation errors for AGB were ignored.

215 **3.1 Nearest neighbor imputations**

216 In this study, the similarity between the i th target observation and j th reference observation
217 was quantified by the means of the Euclidean distance d_{ij} calculated in the feature space as:

$$218 \quad d_{ij} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad \text{eq(1)}$$

219 where x_i and x_j are the feature vectors. Hence, the similarity between the target and reference
220 observations will increase as the d_{ij} distances decrease, and consequently the nearest neighbor
221 of the i th target observation is the reference observation located at the shortest Euclidean
222 distance in the feature space. For other types of similarity measures used with nearest
223 neighbor imputations see Chirici et al. (2008), Eskelson et al. (2009) and McRoberts (2011).

224 With the k -nearest neighbor technique (k -NN), the imputed value \hat{y}_i is expressed as a
225 weighted sum of responses taken from the nearest k reference observations:

$$226 \quad \hat{y}_i = \sum_{j=1}^k w_{ij} y_j^i \quad \text{eq(2)}$$

227 In equation 2, the k -weights associated with the reference responses were obtained

$$228 \quad \text{as } w_{ij} = d_{ij} \left(\sum_{j=1}^k d_{ij} \right)^{-1} .$$

229

230 **3.1.1 Selecting the number of neighbours (k)**

231 Although several criteria for selecting the number of neighbours have been proposed
232 (Eskelson et al., 2009), it is advisable that the choice of k should rely on careful data analysis
233 of the available data (McRoberts et al., 2002; Eskelson et al., 2009). Using $k=1$ is expected to
234 preserve the covariances in imputed responses (Franco-Lopez et al., 2001), but often using a
235 small k value may increase the variability of the imputed observations (McRoberts et al.,
236 2002; Eskelson et al. 2009). Using a large k will smooth the predictions by shifting them

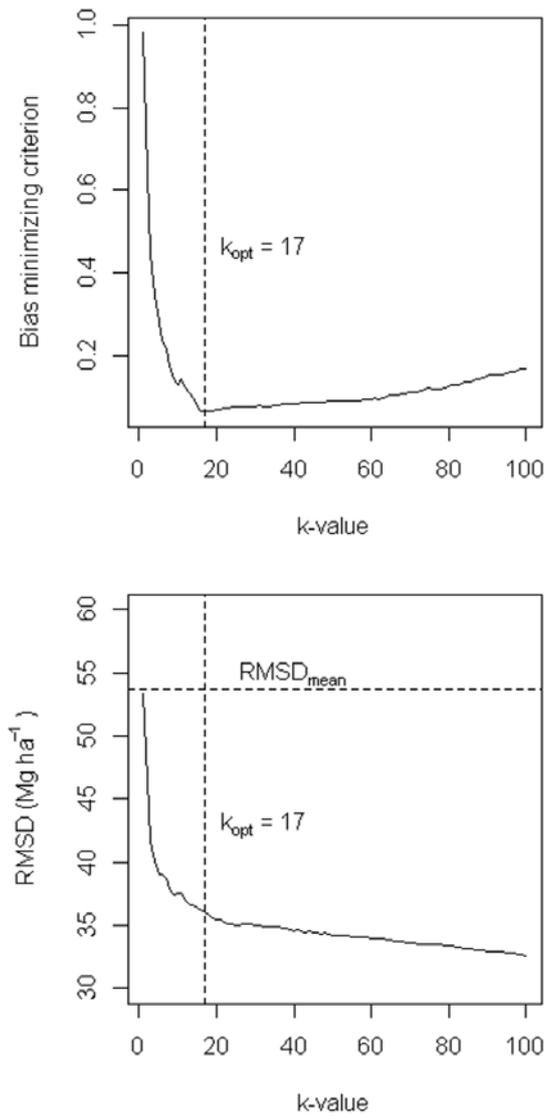
237 towards the mean response of the reference data (Nilsson, 1997) and potentially bias the
238 estimates.

239 The stochastic nature of our simulation study required devising certain criteria for
240 automatic selection of k . Since the k -NN estimates are often biased (Katila, 2006; Magnussen
241 et al., 2009; Magnussen et al., 2010b), a natural criterion for choosing the k is reducing the
242 bias. For optimizing the k -value, leave-one-out cross-validation (LOOC) is usually performed
243 (Chirici et al., 2008; McRoberts, 2009), and the k -NN predictions can be considered unbiased
244 when the 1:1 line between the imputed and the original response has intercept 0 and slope 1
245 (McRoberts, 2009). Using simple linear regression, McRoberts et al. (2011) selected the k -
246 value which minimized simultaneously the slope and intercept to (0, 1). For our data, using
247 linear regression resulted often into unreasonably large k -values due to the presence of
248 outliers. Thus, we considered a robust regression fit instead, which showed a more reasonable
249 behaviour in selecting the optimal k -values (denoted as k_{opt}) during the exploratory data
250 analysis. Fitting the robust line was performed using the ‘rlm’-package (Venables and Ripley,
251 2002) of the R Development Core Team (2011). Further in the text, we will call this approach
252 for selecting k_{opt} as the bias-minimizing criterion. Another common criterion for selecting the
253 k is minimizing the root mean squared difference RMSD (Stage and Crookston, 2007) from
254 LOOC:

$$255 \quad \text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{eq(3)}$$

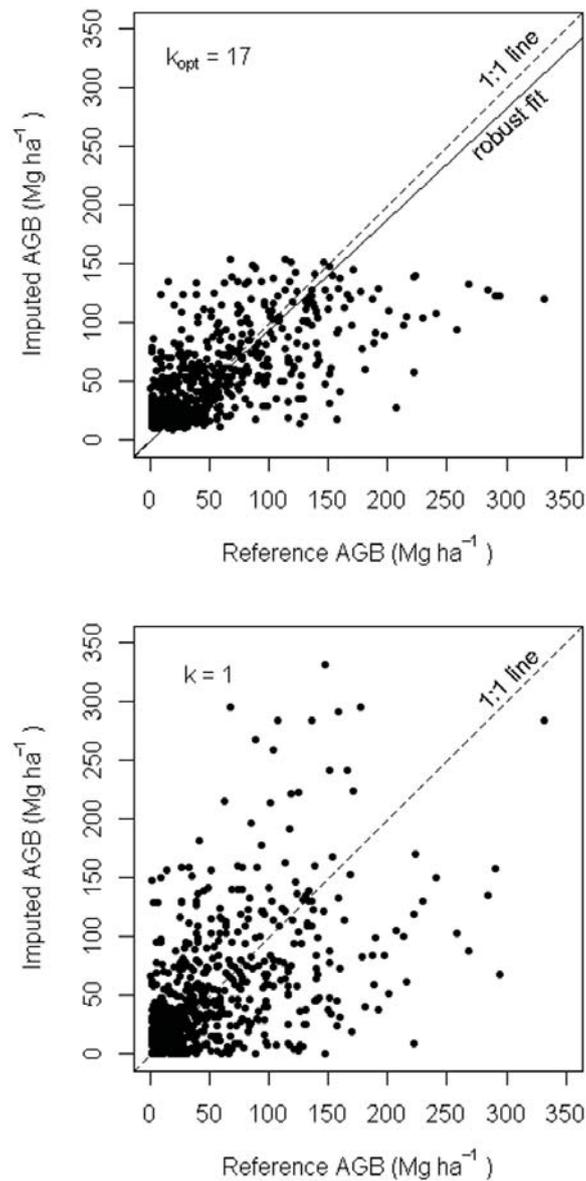
256 When the objective is minimizing RMSD, values of k between 5 and 17 were
257 commonly used (Tomppo et al., 1999, Trotter et al., 1997, Franco-Lopez et al., 2001; Reese
258 et al., 2002). The typical behaviour of the RMSD curve (McRoberts et al., 2002; McRoberts
259 et al., 2011) is to decrease with increasing k -values until it reaches a global minimum (usually
260 for a large k -value), and then to start increasing. For our data, the RMSD curve did not

261 attained a global minimum, probably due to the different set of features we used. Fortunately,
262 the RMSD decay in vicinity of the k -values produced by the bias criterion is usually
263 flattening such that increasing further the number of neighbours might not be justified
264 McRoberts et al. (2011). The use of these bias-and RMSD minimizing criteria for the entire
265 data set (662 observations) with LOOC is illustrated in Figure 2.



266
267 **Figure 2** Selecting the number of nearest-neighbours: using the bias-minimizing criterion produced a
268 distinctive optimal k -value ($k_{opt}=17$) (top), while the root means squared difference (RMSD) minimization
269 criterion did not attain a global minimum, constantly decreasing and flattening while increasing the number
270 of neighbours (bottom).
271

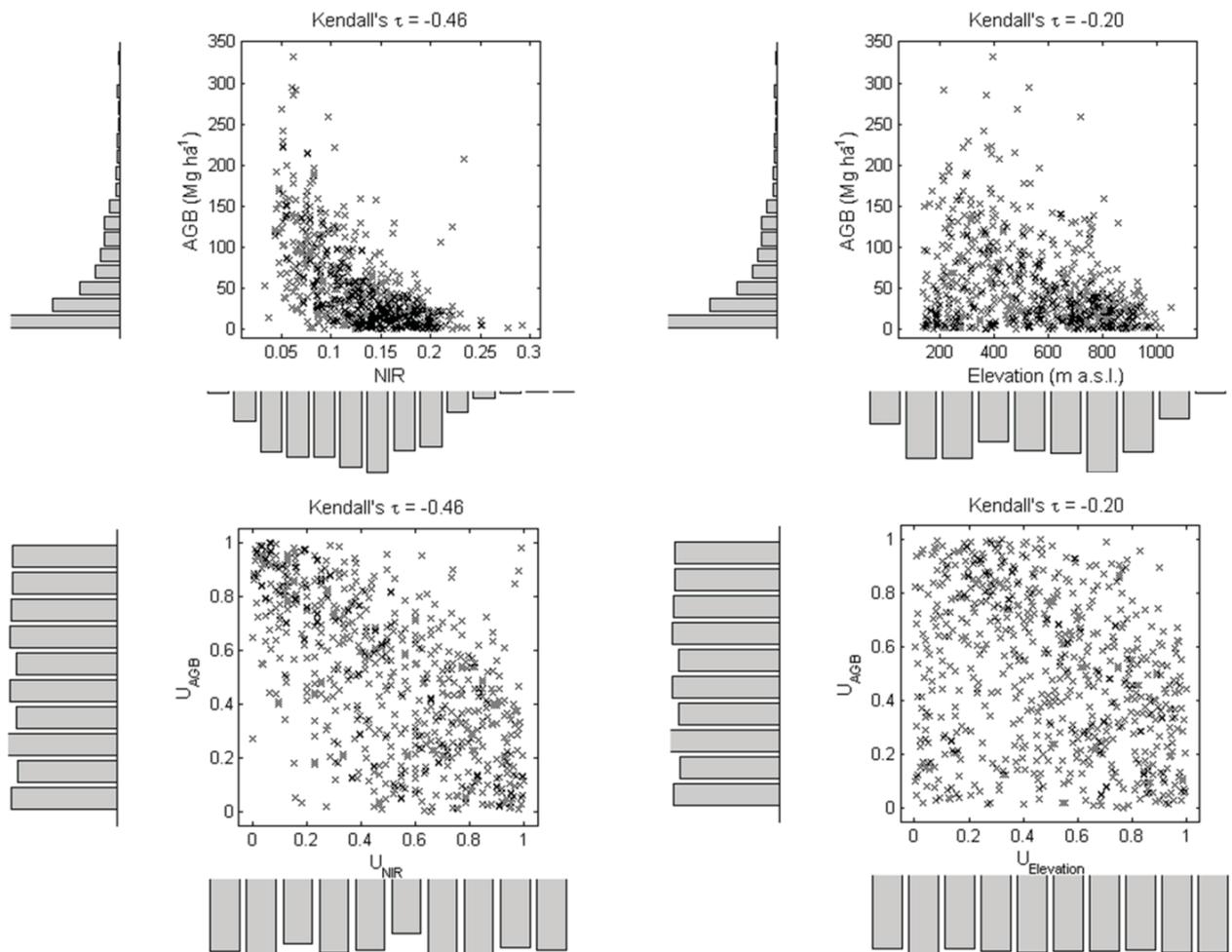
272 Beside the bias minimizing criterion for selecting the optimal number of neighbours,
273 we also performed the nearest neighbor imputations using $k=1$ for assessing the eventual gain
274 obtained when using the former procedure. The imputations obtained using LOOC with the k -
275 value produced by the bias minimizing criterion ($k_{opt} = 17$) and $k=1$ are illustrated in Figure 3.
276



277
278 **Figure 3** Original (reference) versus imputed AGB when using an optimal k -value ($k_{opt}=17$) (top) and $k=1$
279 (bottom). The 1:1 line (the dashed line) has the intercept=0 and slope=1, suggesting a perfect
280 correspondence between reference and imputed observations. The magnitude of the slope and the
281 intercept of the robust regression line (the solid line) may indicate biased predictions.
282

283 ***3.2 Inference using copula-based simulations***

284 A copula is a multivariate distribution function with uniform marginals describing the
285 dependencies among univariate random variables. Considering two random variables X_1 and
286 X_2 and their probability distributions $F(X_1)$ and $F(X_2)$, the Sklar's theorem (Nelsen 2006, p.
287 21) demonstrates the existence of a copula function C such that $F(X_1 = x_1, X_2 = x_2) = C(F_1(x_1), F_2(x_2))$.
288 The rank correlations between variables are preserved by their corresponding uniform
289 margins and captured by the copula model, making possible to capture complex relationships
290 in the data. Hence, uniform multivariate observations (called copula data) can be randomly
291 and independently sampled from the copula functions, while preserving the dependency
292 structure exhibited by the original data. The preservation of the bivariate dependencies
293 between AGB, the near-infrared band (NIR) and elevation data is illustrated in Figure 4.



294

295 **Figure 4 Bivariate relationships and copula data for selected variables (AGB-NIR and ABG-elevation). The**
 296 **non-linear dependencies between variables on original scales (*up*) quantified using Kendal's tau are**
 297 **preserved by their uniforms (*bottom*).**

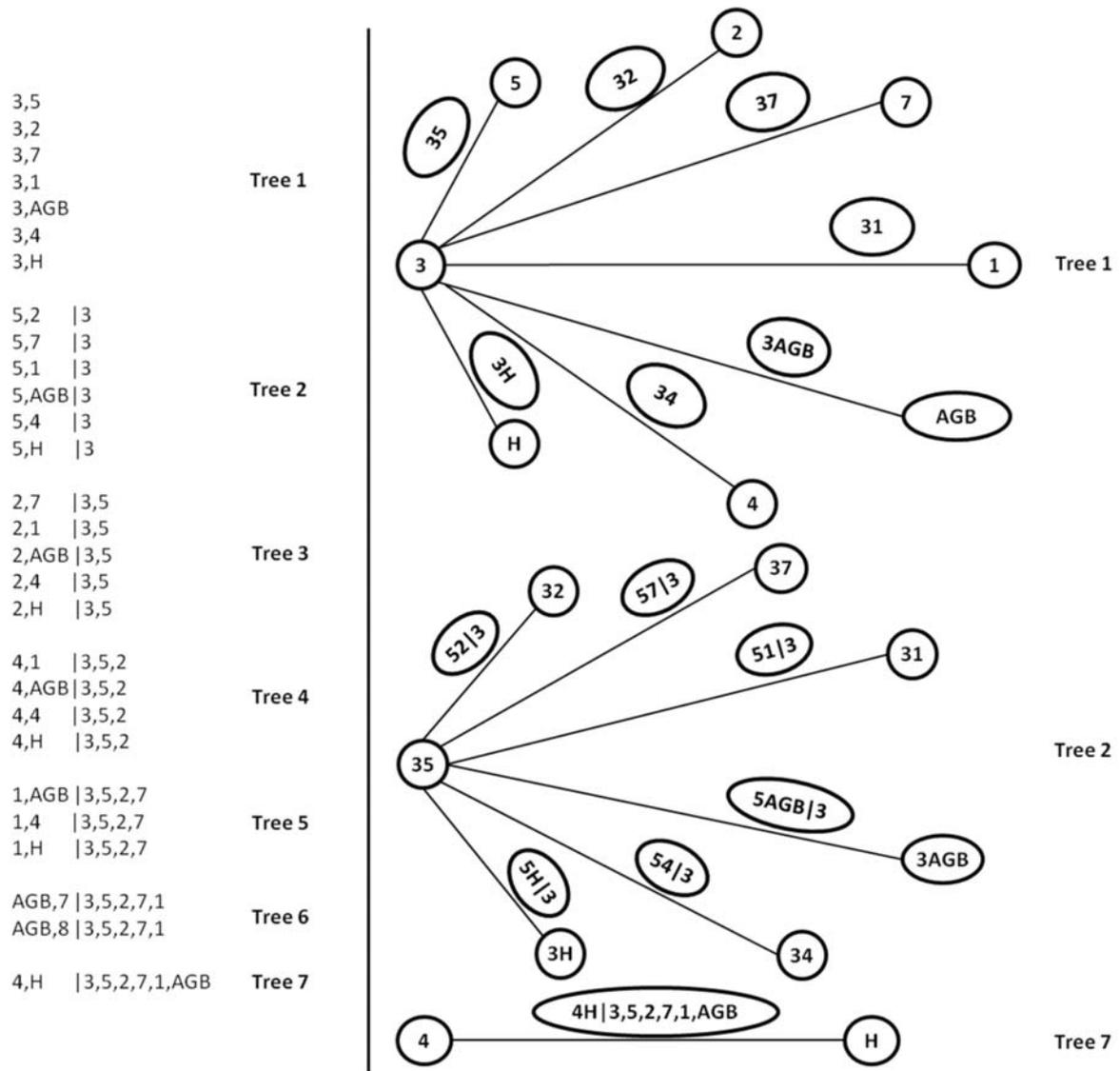
298

299 Although the advantage of using copulas is that modelling the dependencies between
 300 variables does not require distributional assumptions about the variables, the probability
 301 functions of each variable must be known for generating the copula data, and the inverse
 302 probability functions are needed for back-transforming the copula data to the original scale.
 303 Obtaining the probability distribution functions and their inverses is straightforward when
 304 using invertible parametric models for the marginals. However, this approach would assume
 305 *iid* observations, which is unlikely to be obtaining in forest applications. Alternatively,
 306 instead of using parametric distributions, the empirical cumulative distribution function (*ecdf*)
 307 of each variable can be easily obtained from each of the (X, Y) variables.

308 For coping with high-dimensional datasets, Aas et al. (2009) proposed a hierarchical
309 pair-copula decomposition of an n -dimensional distribution into $n(n-1)/2$ pair-copulas which
310 can be arranged in $n-1$ tree-like structures. The approach allows building high-dimensional
311 copula models as products of bivariate copulas and conditional marginal distributions, the
312 copula-pairs being identified by the means of graphical representation called vines (Bedford
313 & Cooke, 2001, 2002; Aas et al., 2009).

314 In this study we refer to the canonical vine (or C-vine), which allows conditioning the
315 multivariate model on selected key variables (Aas et al., 2009). In a C-vine, each tree has a
316 unique node of degree $n-i$ connected to $n-i$ edges (Bedford & Cooke, 2001; Aas et al., 2009).
317 A selected variable is placed in the root node of the first tree, and all pair-wise dependencies
318 between this variable and the other variables are modelled using bivariate copulas. Next,
319 another variable is placed in the second root node, and the pair-wise dependencies to the rest
320 of the variables are modelled conditioned to the variable placed in the first root node. The
321 process continues until all the $n-1$ trees are created, the resulting C-vine tree having a star-like
322 structure (see illustration in Figure 5).

323 Before constructing the C-vine copula model, the variables were ordered as suggested
324 by Czado et al. (2011) and Brechmann & Schepsmeier (2011), the resulting order being TM
325 bands 3-5-2-7-1, AGB, TM band 4 and elevation. That is, band 3 was set at the first root
326 node, band 5 as the second root node, etc. An overview of the tree structure of the resulted C-
327 vine structure is presented in Figure 5.

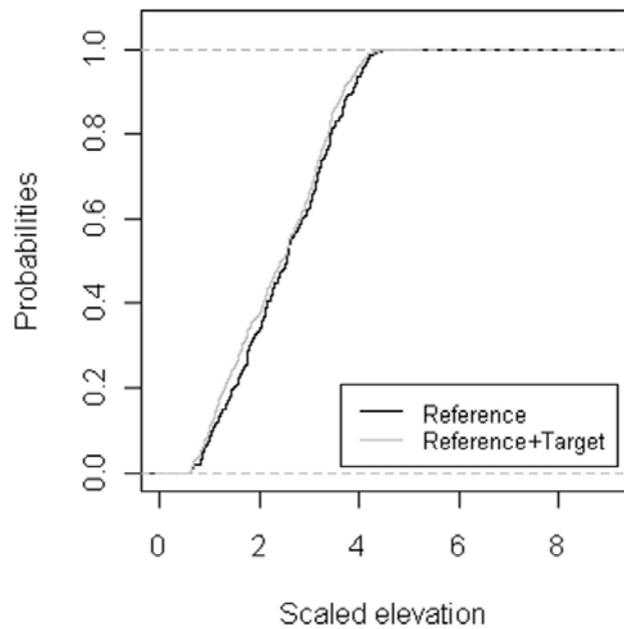
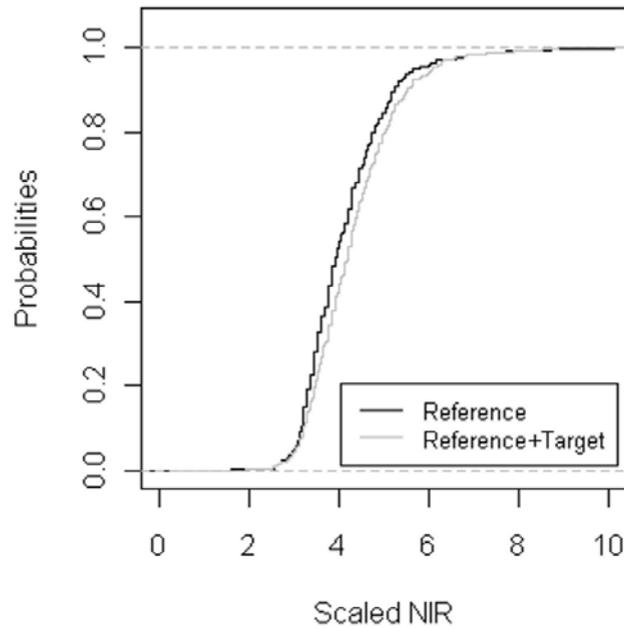


328
329 **Figure 5** Graphical representation of the C-vine copula tree. The full structure of each of the seven trees is
330 given in the left panel, while the graphical representation of selected trees (trees 1, 2 and 7) is presented in
331 the right panel.
332

333 The copula-based approach starts with building a C-vine copula model for the
334 reference data set. The copula model is supposed to capture the complex relationships
335 between X and Y , thus it can be considered a superpopulation model which can generate new
336 multivariate (X, Y) reference samples (called here copula samples). However, constructing the
337 copula model exclusively from the (X, Y) reference observations does not take into account
338 the information provided by the target features. Assuming that the reference and the target
339 data are generated by the same superpopulation model, the relationships between X and Y can
340 be considered to be approximately the same in both data sets. Hence, the reference and target

341 features were pooled into one common set of X variables for estimating the *ecdfs* of the
342 marginals, while the copula data for the Y variable was obtained exclusively from the
343 reference observations.

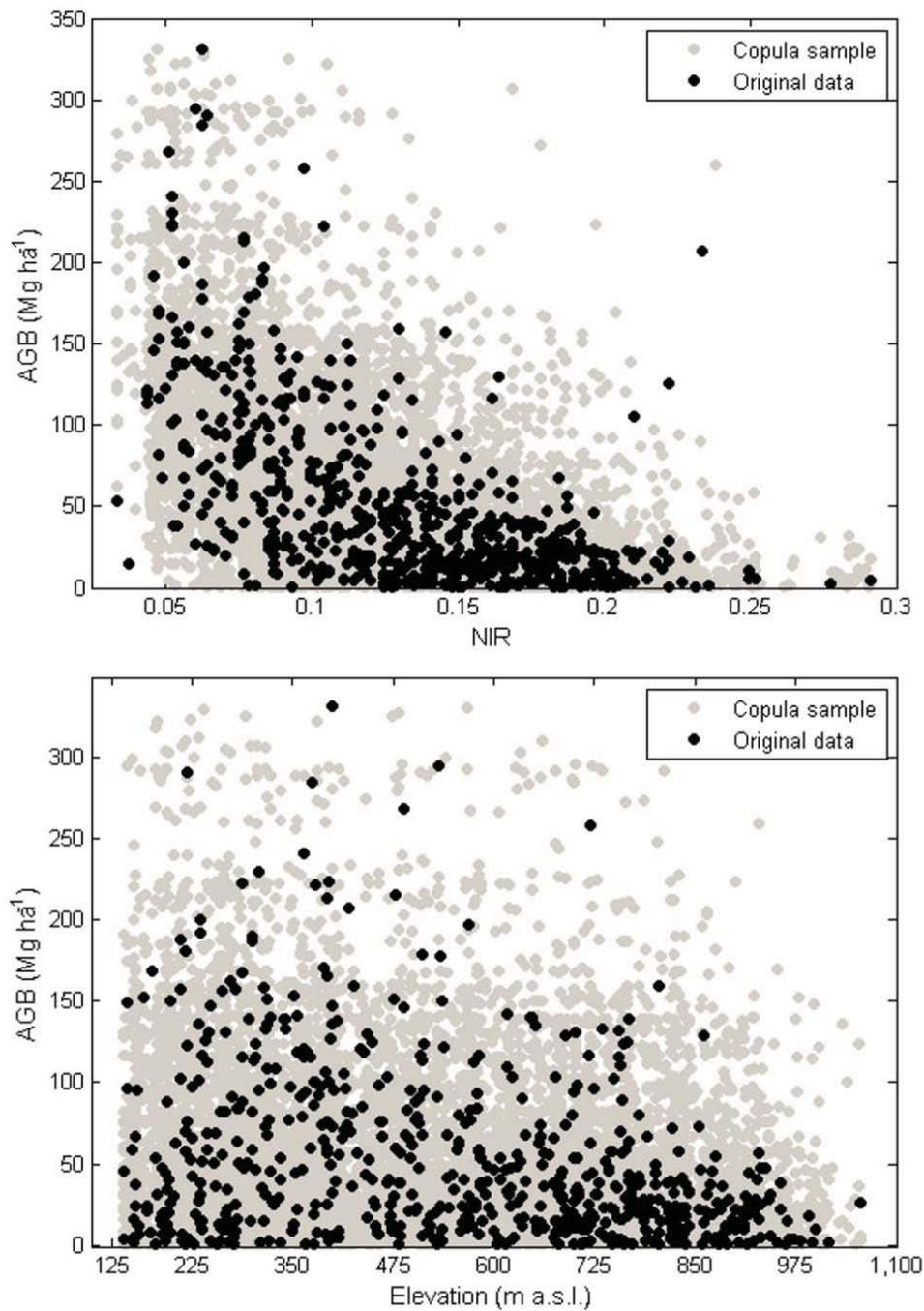
344 The *ecdf* is a step function and using it directly for deriving the uniforms will produce
345 discrete patterns in the generated data (Anon., 2012). To mitigate this effect and to obtain the
346 uniforms necessary for fitting the copula function, the reference X variables were linearly
347 interpolated between the smoothed *ecdfs*. The construction of *ecdfs* for the NIR and elevation
348 variables is exemplified in Figure 6.



349
 350 **Figure 6 Empirical cdfs for two selected features (NIR (top) and elevation (bottom)) obtained using the**
 351 **reference observations (dark line) and by combining reference and target observations (gray line).**
 352

353 In this way, the information existing in the target features could be incorporated in the
 354 C-vine copula model. Further, copula samples of the same size as the original reference
 355 sample were independently generated from the C-vine copula model and used with the
 356 nearest-neighbor imputations for predicting the AGB in the target observations. As shown in

357 Figure 7, the relationships between variables generated from the C-vine copula closely
358 resemble the relationships existing between the empirical variables.



359 **Figure 7** Scatter plots of copula samples and original observations for selected variables (AGB-NIR (top) and
360 **AGB-elevation (bottom)).**
361
362

363 The copula-based approach was implemented using the ‘CDVine’ package
364 (Schepsmeier & Brechmann, 2011) of the R Development Core Team (2011). A detailed

365 explanation regarding the construction the C-vine copula model is given by Brechmann &
366 Schepsmeier (2011).

367

368 ***3.3 Inference using bootstrap resampling***

369 Using the bootstrap resampling for model-based inference about the k -NN prediction is
370 described in detail by McRoberts et al. (2011) and McRoberts (2011). Considering the case of
371 a simple random sampling without replacement survey, the bootstrap samples are selected
372 from the reference data using simple random sampling with replacement and are used to
373 perform the k -NN predictions for the target observations. The bootstrap samples are selected
374 such that they will contain the same number of observations as the reference data. Using a
375 large number of bootstrap samples, the distribution of the k -NN prediction errors can be
376 approximated which allows estimating the bias and the error variance associated with the k -
377 NN method (Köhl et al., 2006, p. 190; McRoberts et al., 2011; McRoberts, 2011).

378 With bootstrap resampling, the parent sample is treated as a pseudo-population, and
379 the bootstrap replicates must be selected from the parent sample following the same sampling
380 design used for selecting the parent sample. Consequently, the model-based inference
381 following bootstrap resampling depends on a design component, which may be difficult to
382 replicate under complex designs.

383

384 ***3.4 Case study and assessment***

385 In this study, the model-based inference is regarded the parameter of a particular finite
386 population (descriptive inference, Kangas 2006, p. 40) and not the superpopulation itself
387 (Särndal et al., 1992, p. 514; Kangas, 2006, p. 40). Nearest neighbor estimates of the finite
388 population parameter can be derived for every realization of the finite population (that is, for
389 every random draw from the superpopulation model), and the validity of the inference can be

390 assessed using confidence/prediction intervals constructed around the point estimates
 391 (Särndal et al., p. 354). The inference is considered to be valid if the confidence intervals
 392 cover the parameter of the finite population subject to estimation. However, even for
 393 unbiased prediction error estimates, the CI can fail to cover the real parameter if the point
 394 estimators are heavily biased.

395 The validity of the inference following the bootstrap and copula-based approaches
 396 was assessed through a simulation case study following LeMay & Temesgen (2005). The
 397 empirical data set of 662 observations was iteratively separated into two sets of $S=331$
 398 observations each using simple random sampling without replacement. One of the datasets
 399 was considered to be the reference data and the other one the target dataset. The data splitting
 400 was repeated $M=1000$ times, and each time the X variables of the reference and target
 401 datasets were standardized by dividing each feature to its standard deviation. For each of the
 402 m th pair of reference-target datasets, N Monte Carlo sampling experiments were generated
 403 using the bootstrap and the copula approaches explained in sections 3.2 and 3.3. From each
 404 replication, the AGB was imputed to the target observations using nearest-neighbor
 405 imputations with $k=1$ and $k=k_{opt}$, as described in section 2.1. Exploratory analysis indicated
 406 that the copula based estimates converged faster compared to the bootstrap resampling, thus
 407 the number of replications N was set to 3000 for the copula approach and to 5000 for the
 408 bootstrap. For each of the n th replication, the average AGB was estimated as:

$$409 \quad \hat{\mu}_{m,n}^* = \frac{1}{S} \sum_{s=1}^S \hat{y}_s \quad \text{eq(4)}$$

410 where \hat{y} are the imputed AGB values and (*) symbolizes either the bootstrap (*boot*) or the
 411 copula (*cop*) approach, and S is the number of observations in the target dataset.

412 After running N replications, the m th Monte Carlo estimate of the average AGB for
 413 the target dataset was obtained as:

414
$$\hat{\mu}_m^* = \frac{1}{N^*} \sum_{n=1}^{N^*} \mu_{m,n}^* \quad \text{eq(5)}$$

415 The biases of the AGB predictions were calculated as the difference between the
 416 known AGB average of the target data set (μ_m) and the estimated AGB average $\hat{\mu}_m^*$ as:

417
$$\text{Bias}(\hat{\mu}_m^*) = \mu_m - \hat{\mu}_m^* \quad \text{eq(6)}$$

418 and the absolute biases were obtained as:

419
$$\text{ABias}(\hat{\mu}_m^*) = \left| \mu_m - \hat{\mu}_m^* \right| \quad \text{eq(7)}$$

420 The variance estimates for the m th data split were obtained as:

421
$$\hat{\text{var}}(\hat{\mu}_m^*) = \frac{1}{N^* - 1} \sum_{n=1}^{N^*} (\mu_{m,n}^* - \hat{\mu}_m^*)^2 \quad \text{eq(8)}$$

422 From equation 8, the estimated standard errors were obtained as:

423
$$\hat{SE}(\hat{\mu}_m^*) = \sqrt{\hat{\text{var}}(\hat{\mu}_m^*)} \quad \text{eq(9)}$$

424 Finally, from equations 6 and 8, the m th root mean squared error was estimated as:

425
$$\hat{RMSE}(\hat{\mu}_m^*) = \sqrt{\hat{\text{Bias}}(\hat{\mu}_m^*)^2 + \hat{\text{var}}(\hat{\mu}_m^*)} \quad \text{eq(10)}$$

426 The estimated absolute biases, SE and $RMSE$ estimates obtained after running the M
 427 data splits were averaged to obtain average absolute biases (AAB^*), average standard errors
 428 (ASE^*) and average RMSE ($ARMSE^*$).

429 For each of the m th data split, $(1 - \alpha)100\%$ confidence intervals for the estimated
 430 means were constructed at the significance levels $\alpha=0.20, 0.10, 0.05$ and 0.01 using the t -
 431 distribution, resulting in confidence intervals with nominal coverages of 80%, 90%, 95% and
 432 99%, respectively. The limits of the confidence intervals for each m th data split were
 433 calculated as:

434
$$tCI_{1-\alpha}^m = \hat{\mu}_{*,m} \pm t_{(1-\alpha/2, N-1)} \sqrt{\hat{\text{var}}_*(\hat{\mu}_m^*)} \quad \text{eq(11)}$$

435 After running the M data splits, the coverage rates of the confidence intervals were obtained
 436 as:

437
$$ACR^* = 100 \frac{T^*}{M} \tag{eq(12)}$$

438 where T^* is the number of times the tCIs covered the true means of the M target datasets.

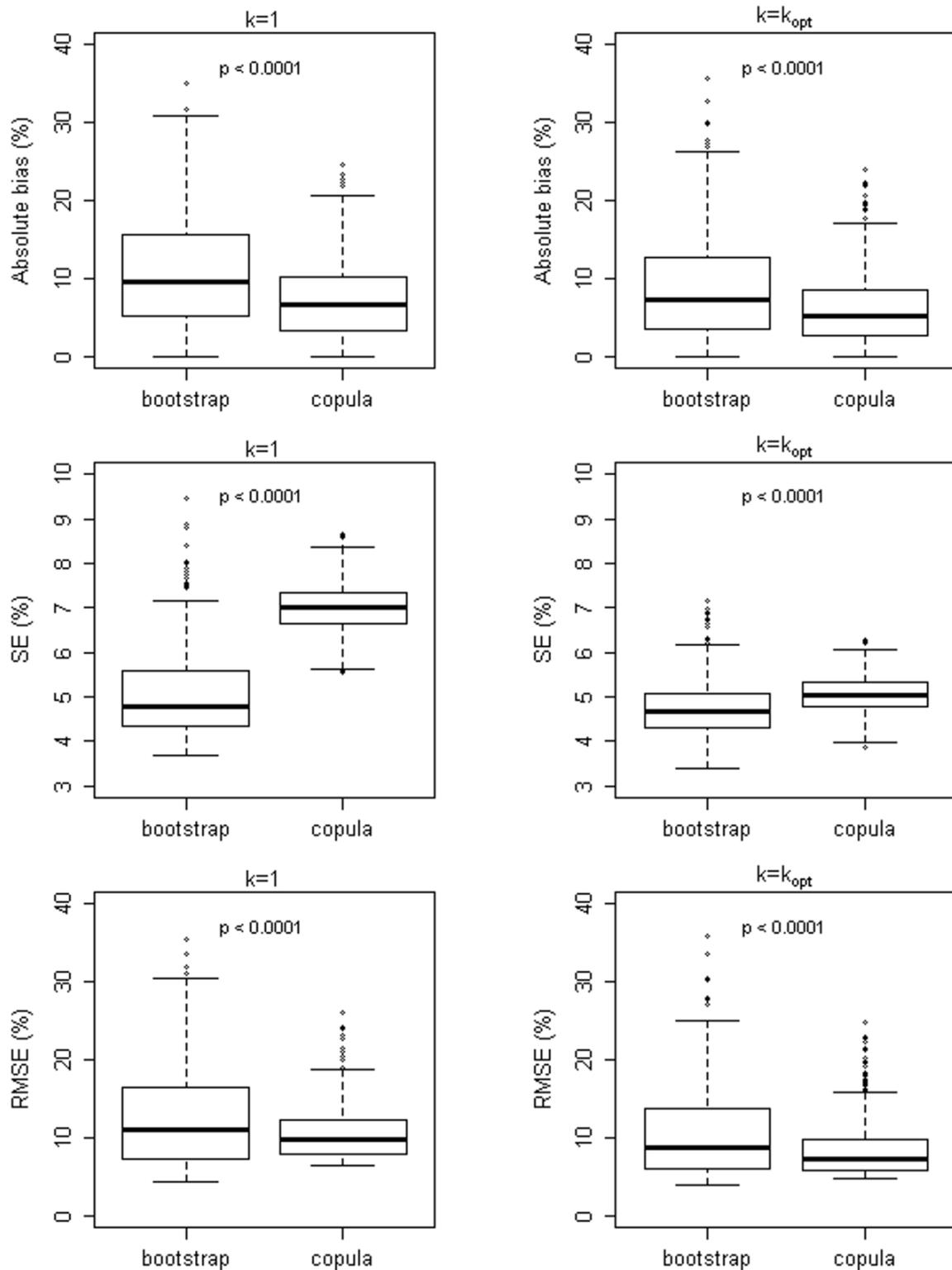
439 Finally, we tested whether the results produced by the bootstrap and copula-based
440 approaches were significantly different ($\alpha=0.025$) using one-sided paired t-tests for the bias,
441 SE and RMSE, and one-sided binomial tests for ACR^* .

442 For our study, the sampling design used for selecting the parent sample will not
443 influence the analysis, because the estimation is focused on the target data sets created by the
444 randomly splitting the original data, and not on making inference about the parent population.

445

446 **4. Results**

447 The number of nearest neighbours found by the bias-minimizing criterion varied between $k=7$
448 and $k=42$, with an average of 17. The simulation results for nearest neighbor imputations with
449 $k=1$ and $k=k_{opt}$ are shown in Figure 8.



450
 451 **Figure 8** Box and whisker plots of absolute bias (top), standard errors (middle) and root mean squared errors
 452 (RMSE) (bottom) resulted from simulations using $k=1$ (left) and $k=k_{opt}$ (right). The p-values indicate the
 453 statistical significance of the paired t-tests.
 454

455 For nearest neighbor imputations with $k=1$, the AAB^{boot} (10.73%) was 52% higher
 456 compared with AAB^{cop} (7.08%). The ASE^{boot} (5.05%) was 28% lower compared to ASE^{cop}
 457 (7.01%), and the $ARMSE^{cop}$ (10.51%) was 15% lower compared to $ARMSE^{boot}$ (12.36%).

458 When using $k=k_{opt}$, AAB^{boot} (8.72%) was reduced with 19%, and the AAB^{cop} (6.06%)
 459 with 14%, the difference between the AAB^{boot} and AAB^{cop} decreased from 52% (for $k=1$) to
 460 44%. Also, both the ASE^{boot} (4.75%) and ASE^{cop} (5.04%) decreased by 6% and 28%,
 461 respectively, the ASE^{boot} being 6% lower than ASE^{cop} .

462 On average, the root mean square error decreased for both approaches ($ARMSE^{boot}$
 463 =10.46% and $ARMSE^{cop}$ = 8.40%) by approximately 20 and 15% respectively, the $ARMSE^{cop}$
 464 being 20% lower than $ARMSE^{boot}$.

465 Compared to using $k=1$, choosing $k=k_{opt}$ resulted in a statistically significant
 466 reduction ($p<0.0001$) of the absolute bias, standard error and root mean squared errors for
 467 both approaches. Moreover, the bootstrap and copula results were also significantly different
 468 when using the same k -values (Figure 8).

469 The confidence intervals estimated using the copula and bootstrap methods for $k=1$
 470 and $k=k_{opt}$ are presented in Table 3.

471

472 **Table 3. Confidence interval coverage (%) obtained using bootstrap and copula methods for $k=1$ and $k=k_{opt}$**

k -value	Method	tCI coverage			
		80 %	90 %	95 %	99 %
$k=1$	Bootstrap	34.4	45.2	49.8	64.0
	Copula	66.8	82.2	92.0	97.0
$k=k_{opt}$	Bootstrap	40.2	53.0	59.8	72.6
	Copula	60.2	73.6	81.6	91.6

473

474 For both approaches, the CI coverages improved as the confidence statements were
 475 relaxed. However, the CIs produced by the copula-based approach were systematically closer
 476 to their nominal coverage rates, and the binomial test indicated significant differences

477 between the coverage rates produced by copula and bootstrap ($p < 0.0001$). The best overall
478 coverage rates were obtained by the copula approach for $k=1$, but the improvement was not
479 found significant ($p > 0.05$) compared to using the copula approach with $k=k_{opt}$. Also, when
480 using $k=k_{opt}$ the bootstrap approach produced better CIs compared with using $k=1$, although
481 the differences were found to be significant only for the 90, 95 and 99% CIs.

482 For the copula-based approach, there was a more pronounced reduction in the
483 standard error estimates than in the bias when using k_{opt} . Hence, the estimated CIs narrowed,
484 failing to cover the real AGB mean more frequent compared to using $k=1$. For the bootstrap
485 approach, the reduction of standard error when using k_{opt} was minor (although statistically
486 significant) compared to the bias reduction, and the CI coverage rates increased compared to
487 using $k=1$.

488

489 **5. Discussion**

490 This study introduced a new model-based inferential approach for nearest-neighbor
491 predictions, which can be useful especially for small-area estimation, when sampled
492 observations within the area of interest are not available. Combining reference observations
493 with the features from the target data, a superpopulation model was created using a C-vine
494 copula function. From this model, simulated samples containing *iid* observations were
495 iteratively generated and predictions for the target data were made via nearest-neighbor
496 imputations. Predictions based on a number of neighbours determined by minimizing the
497 prediction bias were compared to predictions using the nearest neighbor only.

498 The inference following the copula approach was found superior to the bootstrap
499 resampling for all cases. The copula observations are ‘filling’ the gaps among observations in
500 the reference feature space, reducing the biases but also increasing the variability of the
501 nearest neighbor estimates. The most conservative results (larger standard errors, but valid

502 CIs) were produced by the copula-based approach with $k=1$. Despite a lower precision, this
503 might actually be a desirable solution especially when there are chances for severe
504 extrapolation outside the reference feature space. In turn, the bootstrap approach is prone to
505 larger biases due to the limited variability of the observations available in the original sample
506 dataset. Due to the with-replacement resampling, the bootstrap samples will contain common
507 observations which and can be repeatedly selected as the nearest neighbours to the same
508 targets, thus reducing the variability of the nearest neighbor estimates. Furthermore, using
509 larger k -values had a more pronounced effect on reducing the between-sample variability of
510 the k -NN estimates based on copula data than reducing the bias.

511 The results suggested that the main source of uncertainty producing the failure of the
512 CI coverage is the bias of the nearest neighbor estimator. Magnussen et al. (2010*ab*) and
513 McRoberts et al. (2011) reported rather negligible biases in small-area estimation studies,
514 while Katila (2006) noticed the occurrence of important biases at various geographical scales.
515 However, the material used in the present study spans a large geographical area
516 (approximately 300 km north-south and 80 km east-west), containing various trends and
517 representing many forest types (Ene et al., in submission; Gobakken et al., *in press*), and due
518 to the data splitting procedure we have used, it is reasonable to assume that extrapolations
519 occurred frequently. Screening of the reference observations and subsequently imposing
520 restrictions regarding nearest neighbor selection and thus avoiding reference observations
521 from being imputed to unreasonable target locations might help reducing the bias (Katila
522 & Tomppo, 2001; Tomppo & Halme, 2004). However, the criteria for applying such
523 restrictions in Norwegian conditions require a thorough investigation which was not
524 considered in this study. Furthermore, there is nothing that would preclude combining the
525 copula-based approach with existing error reduction strategies developed for k -NN methods
526 (e.g. Magnussen et al., 2010*b*; Malinen, 2003; Tomppo & Halme, 2004).

527 Usually, a large number of variables are collected during large-area forest inventories,
528 and it is not clear yet for which of these variables the methods used in our study are
529 applicable. Our study considers exclusively the case of having continuous variables, but other
530 field variables might be registered either as nominal data (e.g. vegetation type, site index, tree
531 species, cluster membership, etc.) or as ordinal data (stem number, abundance of plant
532 species, etc.). While bootstrap resampling may be applicable in the presence of discrete
533 variables, it is not yet obvious how the copula approach would accommodate to such data
534 types. Nevertheless, theoretical aspects regarding the possibilities and the limitations of
535 modelling copulas with ordinal margins have been investigated (Genest & Nešlehová, 2007;
536 Nikoloulopoulos & Karlis, 2009) and recent progress in developing discrete vine-copula
537 constructions (Smith, 2011; Panagiotelis et al., *in submission*) is reported. Arguably, a naive
538 approach for dealing with the presence of ordinal variables would be to use classification
539 schemes derived from the reference data for labeling the target and the copula observations
540 prior to k -NN imputations. However, there is a question whether the prediction of all
541 variables acquired during large-area inventories is actually needed when focusing on small-
542 domain estimation. Very detailed information might not be required for very small forest
543 areas, thus the copula-based approach might be applied only for a subset of the forest
544 variables.

545 To conclude, our approach is distribution-free and can provide uncertainty estimates
546 at different level of spatial aggregation. Moreover, using copulas to join the *ecdfs* of the
547 marginals relaxes the *iid* assumptions required by the bootstrap. This may constitute an
548 advantage in forestry applications where data collection schemes and population
549 characteristics may induce various types of dependencies among observations. Still, using the
550 copula-approach with data collected using cluster and/or multi-phase sampling requires
551 further investigations.

552 **Acknowledgments**

553 This research has been funded by the Research Council of Norway (project no. 166482/i10,
554 no. 184636/S30 and no. 180645). The authors would like to thank the Norwegian National
555 Forest Inventory for collection of the field.

556 **References**

- 557 Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple
558 dependence. *Insurance: Mathematics and Economics*, 442, 182-198.
- 559 Anon. (2012). MATLAB Online Help (<http://www.mathworks.com/help/toolbox/stats/>)(*last*
560 *accessed 07-Feb-2012*).
- 561 Baffetta, F., Fattorini, L., Francheschi, S. & Corona, P. (2009). Design-based approach to k -
562 nearest neighbours techniques for coupling field and remotely sensed data in forest
563 surveys. *Remote Sensing of Environment*, 113, 463-475.
- 564 Baffetta, F., Corona, P., Fattorini, L. (2011). Design-based diagnostics for k -NN estimators of
565 forest resources. *Canadian Journal of Forest Research*, 40, 59-72.
- 566 Bedford, T. & Cooke, R. M. (2001). Probability density decomposition for conditionally
567 dependent random variables modelled by vines. *Annals of Mathematics and Artificial*
568 *Intelligence*, 32, 245–268.
- 569 Bedford, T. & Cooke, R. M. (2002). Vines - a new graphical model for dependent random
570 variables. *Annals of Statistics* 30, 1031–1068.
- 571 Brechmann, E.C. & Schepsmeier, U. (2011). Modelling dependence with C- and D-vine
572 copulas: The R-package CDVine. *R vignette of the R-package CDVine (version 1.1-*
573 *5)*, URL [http://cran.r-project.org/web/packages/CDVine/vignettes/CDVine-](http://cran.r-project.org/web/packages/CDVine/vignettes/CDVine-package.pdf)
574 [package.pdf](http://cran.r-project.org/web/packages/CDVine/vignettes/CDVine-package.pdf) (*last accessed 01-Feb-2012*).
- 575 Chavez, P.S. (1996). Image-based atmospheric corrections- revisited and improved.
576 *Photogrammetric Engineering and Remote Sensing*, 62, 1025-1036.
- 577 Chen, J. & Shao, J. (2001). Jackknife variance estimation for nearest neighbor imputations.
578 *Journal of the American Statistical Association*, 96, 260-269.

579 Chirici, G., Barbati, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F & Bertini, R.
580 (2008). Non-parametric and parametric methods using satellite images for estimating
581 growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sensing*
582 *of Environment*, 112, 2686-2700.

583 Corona, P. (2010). Integration of forest mapping and inventory to support forest management.
584 *iForest*, 3, 59-64 (online: 2010-05-17). URL: <http://sisef.it/iforest/show.php?id=531>
585 (last accessed 10-Feb-2012).

586 Czado, C., Schepsmeier, U. & Min, A. (2011). Maximum likelihood estimation of mixed C-
587 vines with application to exchange rates. Accepted for publication in *Statistical*
588 *Modelling*.

589 de Leon, A. R., Wu, B. & Withanage, N. (2011). Gaussian copula-generated distributions for
590 correlated discrete and continuous variables. (in submission). URL:
591 <http://math.ucalgary.ca/~adeleon/robit.pdf> (last accessed 14-Feb-2012)

592 Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*,
593 7, 1-26.

594 Embrechts, P., McNeil A. J. & Straumann, D. (2002). Correlation and Dependency in Risk
595 Management: Properties and Pitfalls, in: M. Dempster, ed., *Risk Management: Value*
596 *at Risk and Beyond* (Cambridge: Cambridge University Press), 176-223.

597 Ene, L.T., Næsset, E., Gobakken, T., Gregoire, T.G., Ståhl, G. & Nelson, R. Assessing the
598 accuracy of regional LiDAR-based biomass estimation using a simulation approach
599 (in submission).

600 Eskelson, B.N.I., Temesgen, H., LeMay, V., Barret, T.M., Crookston, N.L & Hudak, A.T.
601 (2009). The roles of nearest neighbor methods in imputing missing data in forest
602 inventory and monitoring databases. *Scandinavian Journal of Forest Research*, 24,
603 235-246.

604 Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics*, 9, 1218–
605 1228.

606 Genest, C & Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, 37,
607 475-515.

608 Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula
609 modelling but were afraid to ask. *Journal of Hydrologic Engineering*, 12, 347-368.

610 Genest, C., Gerber, H.U., Goovaerts M.J., & Laeven, R.J.A. (2009). Editorial to the special
611 issue on modelling and measurement of multivariate risk in insurance and finance.
612 *Insurance: Mathematics and Economics* 44, 143-145.

613 Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., Holm, S.,
614 Ørka, H.O., & Astrup, R. Estimating biomass in Hedmark County, Norway, using
615 national forest inventory field plots and airborne laser scanning (*in press*).

616 Gregoire, T.G. (1998). Design-based and model-based inference in survey sampling:
617 appreciating the difference. *Canadian Journal of Forest Research*, 28, 1429-1447.

618 Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, O.E. & Falkowski, M.J. (2008). Nearest
619 neighbor imputation of species-level, plot-scale forest structure attributes from
620 LiDAR data. *Remote Sensing of Environment*, 112, 2232-2245.

621 Kangas, A. (2006). Model-based inference. In A. Kangas, & M. Maltamo, (Eds.), *Forest*
622 *Inventory: Methodology and Applications* (p. 39-52). Dordrecht: Springer.

623 Katila, M. (2006). Empirical errors of small area estimates from the multisource national
624 forest inventory in eastern Finland. *Silva Fennica*, 40, 729-742.

625 Katila, M. & Tomppo, E. (2001). Selecting estimation parameters for the Finnish multi-
626 source national forest inventory. *Remote Sensing of Environment*, 76, 16-32.

- 627 Kershaw, J.A. (Jr.), Richards, E.W., McCarter, J.B. & Oborn, S. (2010). Spatially correlated
628 forest structures: a simulation approach using copulas. *Computers and Electronics in*
629 *Agriculture*, 74, 120-128.
- 630 Kim, H.-J. & Tomppo, E. (2006). Model-based prediction error uncertainty estimation for *k*-
631 nn method. *Remote Sensing of Environment*, 104, 257-263.
- 632 Koistinen, P., Holström, L. & Tomppo, E. (2008). Smoothing methodology for predicting
633 regional averages in multi-source forest inventory. *Remote Sensing of Environment*,
634 112, 862-871.
- 635 Korhonen, K.T & Kangas, A. (1997): Application of nearest-neighbour regression for
636 generalizing sample tree information. *Scandinavian Journal of Forest Research*, 12,
637 97-101.
- 638 Köhl, M., Magnussen, S. & Marchetti, M. (2006). *Sampling methods, remote sensing and*
639 *GIS multiresource forest inventory* (Springer: Berlin Heidelberg), 392 pp.
- 640 Lahiri, P. (2003). On the impact of bootstrap in survey sampling and small area estimation.
641 *Statistical Science*, 18, 199–210.
- 642 LeMay, V. & Temesgen, H. (2005). Comparison of nearest neighbor methods for estimating
643 basal area and stems per hectare using aerial auxiliary variables. *Forest Science*, 51,
644 109-119.
- 645 LeMay, V., Maedel, J. & Coops, N.C. (2008). Estimating stand structural details using
646 nearest neighbor analysis to link ground data, forest cover maps, and Landsat
647 imagery. *Remote Sensing of Environment*, 112, 2578-2591.
- 648 Malinen, J. (2003). Locally adaptable non-parametric methods for estimating stand
649 characteristics for wood procurement planning. *Silva Fennica*, 37, 109-120.

650 Magnussen, S., McRoberts, R.E. & Tomppo, E. (2009). Model-based mean-squared error
651 estimators for k -nearest neighbour predictions and applications using remotely sensed
652 data for forest inventories. *Remote Sensing of Environment*, 113, 473-488.

653 Magnussen, S., McRoberts, R.E. & Tomppo, E. (2010a). A resampling variance estimator
654 for the k nearest neighbours technique. *Canadian Journal of Forest Research*, 40,648-
655 658.

656 Magnussen, S., Tomppo, E. & McRoberts, R.E. (2010b). A model-assisted k -nearest
657 neighbor approach to remove extrapolation bias. *Scandinavian Journal of Forest
658 Research*, 25, 174-184.

659 McRoberts, R.E., Nelson, M.D., and Wendt, D.G. (2002). Stratified estimation of forest area
660 using satellite imagery, inventory data, and the k -nearest neighbors technique. *Remote
661 Sensing of Environment*, 812, 457-468.

662 McRoberts, R.E. (2006). A model-based approach to estimating forest area. *Remote Sensing
663 of Environment*, 103, 56-66.

664 McRoberts, E.R. & Tomppo, E. (2007). Remote sensing support for national forest
665 inventories. *Remote Sensing of Environment*, 110, 412-419.

666 McRoberts, E.R., Tomppo, E.O., Finley, A.O. & Heikkinen, J.H. (2007). Estimating areal
667 means and variances of forest attributes using the k -nearest neighbours technique and
668 satellite imagery. *Remote Sensing of Environment*, 111, 466-480.

669 McRoberts, R. (2008). Using satellite imagery and the k -nearest neighbours technique as a
670 bridge between strategic and management forest inventories. *Remote Sensing of
671 Environment*, 112, 2212-2221.

672 McRoberts, R.E. (2009). Diagnostic tools for nearest neighbor techniques when used with
673 satellite imagery. *Remote Sensing of Environment*, 113, 489-499.

674 McRoberts, R.E. (2010). Probability- and model-based approaches to inference for proportion
675 forest using satellite imagery as ancillary data. *Remote Sensing of Environment*, 114,
676 1017-1025.

677 McRoberts, R.E., Cohen, W.B., Næsset, E., Stehman, S.V. & Tomppo, E.O. (2010a). Using
678 remotely sensed data to construct and assess forest attribute maps and related spatial
679 products. . *Scandinavian Journal of Forest Research*, 25, 340-367.

680 McRoberts, R.E., Tomppo, E.O. & Næsset, E. (2010b). Advances and emerging issues in
681 national forest inventories. *Scandinavian Journal of Forest Research*, 25, 368-381.

682 McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbor
683 techniques. *Forest Ecology and Management* (2011),
684 doi:10.1016/j.foreco.2011.06.039.

685 McRoberts, R.E., Magnussen, S., Tomppo E.O. & Chirici, G. (2011). Parametric, bootstrap,
686 and jackknife variance estimators for the k -nearest neighbours technique with
687 illustrations using forest inventory and satellite image data. *Remote Sensing of*
688 *Environment*, 115, 3165-3174.

689 Nelsen, R.B. (2009). An introduction to copulas, (2nd ed.). New York: Springer.

690 Nikoloulopoulos, A.K. & Karlis, D. (2009). Modeling Multivariate Count Data Using
691 Copulas. *Communications in Statistics - Simulation and Computation*, 39,172-187.

692 Nilsson, M. (1997). Simultaneous estimation of forest variables using Landsat TM data. *Acta*
693 *Universitatis Agriculturae Suecia. Silvestria 17*. Estimation of forest variables using
694 satellite image data and airborne Lidar. Doctoral thesis Paper IV. Swedish University
695 of Agricultural Sciences.

696 Packalén, P. & Maltamo, M. (2006). Predicting the plot volume by tree species using
697 airborne laser scanning and aerial photographs. *Forest Science*, 52, 611-622.

698 Panagiotelis, A., Czado, C. & Joe, H. Pair copula constructions for discrete data
699 (submitted).

700 URL: [http://www.asb.unsw.edu.au/schools/economics/Documents/A.%20Panagiotelis%20-](http://www.asb.unsw.edu.au/schools/economics/Documents/A.%20Panagiotelis%20-%20Pair%20Copula%20Constructions%20for%20Discrete%20Data.pdf)
701 [%20Pair%20Copula%20Constructions%20for%20Discrete%20Data.pdf](http://www.asb.unsw.edu.au/schools/economics/Documents/A.%20Panagiotelis%20-%20Pair%20Copula%20Constructions%20for%20Discrete%20Data.pdf) (last
702 accessed 08-Feb-2012).

703 R Development Core Team (2011). R: A language and environment for statistical computing.
704 R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL
705 <http://www.R-project.org/>

706 Rancourt, E., Sarndal, C & Lee, H. (1997). Estimation of the variance in the presence of
707 nearest neighbor imputation. *Proceedings of the Section on Survey Research Methods,*
708 *American Statistical Association*, 888-893.

709 Rätty, M. & Kangas, A. Reprint of: Comparison of k -MSN and kriging in local prediction.
710 *Forest Ecology and Management* (2012), doi: 10.1016/j.foreco.2011.12.046.

711 Reese, H., Nilsson, M., Pahlén, T.G., hagner, O., Joyce, S., Tingelöf, U, Egberth, M. &
712 Olsson, H. (2002).

713 Schepsmeier, U. & Brechmann, E.C. (2011). CDVine: Statistical inference of C- and D-
714 vine copulas. R package version 1.1-5, URL [http://CRAN.R-](http://CRAN.R-project.org/package=CDVine)
715 [project.org/package=CDVine](http://CRAN.R-project.org/package=CDVine) (last accessed 03-Feb-2012).

716 Shao, J. (2009). Nonparametric variance estimation for nearest neighbor imputation. *Journal*
717 *of Official Statistics*, 25, 55-62.

718 Smith, M. S. (2011). Bayesian Approaches to Copula Modelling, *Working Paper*. URL:
719 http://works.bepress.com/michael_smith/30 (last accessed 14-Feb-2012).

720 Stage, A.R. & Crookston N.L. (2007). Partitioning error components for accuracy-assessment
721 of near-neighbor methods of imputation. *Forest Science*, 53, 62-72.

722 Särndal, C.E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New
723 York: Springer.

724 Quenouille, M. (1949). Approximate tests of correlations in time series. *Journal of the Royal*
725 *Statistical Society, Series B*, 11, 18-44.

726 Temesgen, H., M.E. Goerndt, G. P. Johnson, D.M. Adams, and R.A. Monserud. 2007. Forest
727 measurement and biometrics in forest management: status and future needs of the
728 Pacific Northwest USA. *Journal of Forestry*. 105: 233-238.

729 Tomppo, E. & Halme, M. (2004). Using coarse scale forest variables as ancillary information
730 and weighting of variables in k -NN estimation: a genetic algorithm approach. *Remote*
731 *Sensing of Environment*, 92, 1-20.

732 Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O. & Katila, M. (2008a). Combining
733 national forest inventory plots and remote sensing data for forest databases. *Remote*
734 *Sensing of Environment*, 112, 1982-1999.

735 Tomppo, E., Haakana, M., Katila, M. & Peräsaari, J. (2008b). *Multi-Source National Forest*
736 *Inventory Methods and Applications*, Springer Series: Managing Forest Ecosystems,
737 374 p.

738 Trotter, C.M., Dymond, J.R. & Goulding, C.J. (1997). Estimation of timber volume in a
739 coniferous plantation forest using Landsat TM. *International Journal of Remote*
740 *Sensing*, 18, 2209-2223.

741 Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S*. Fourth edition.
742 Springer, New York.

743 Walter, B.F., Finley, A.O. & McRoberts, R.E. (2008). Estimating optimal model parameter
744 values for k -nearest neighbor prediction. *FIA Symposium*, October 2008, Park City,
745 UT.

746 Wu, B. & de Leon, A. R. (2012). Flexible random effects copula models for clustered mixed
747 bivariate outcomes. (*in submission*). URL:
748 http://math.ucalgary.ca/~adeleon/re_copula_paper.pdf (*last accessed 14-Feb-2012*).

Paper IV

Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates

LIVIU ENE*, ERIK NÆSSET and TERJE GOBAKKEN

Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, NO-1432 Ås, Norway

(Received 25 April 2010; in final form 21 November 2011)

Adaptive single tree detection methods using airborne laser scanning (ALS) data were investigated and validated on 40 large plots sampled from a structurally heterogeneous boreal forest dominated by Norway spruce and Scots pine. Under the working assumption of having uniformly distributed tree locations, area-based stem number estimates were used to guide tree crown delineation from rasterized laser data in two ways: (1) by controlling the amount of smoothing of the canopy height model and (2) by obtaining an appropriate spatial resolution for representing the forest canopy. Single tree crowns were delineated from the canopy height models (CHMs) using a marker-based watershed algorithm, and the delineation results were assessed using a simple tree crown delineation algorithm as a reference method ('RefMeth'). Using the proposed methods, approximately 46–50% of the total number of trees were detected, while approximately 5–6% false positives were found. The detection rate was, in general, higher for Scots pine than for Norway spruce. The accuracy of individual tree variables (total height and crown width) extracted from the laser data was compared with field-measured data. The individual tree heights were better estimated for deciduous tree species than for the coniferous species Norway spruce and Scots pine. The estimation of crown diameters for Scots pine and deciduous species achieved comparable accuracy, being better than for Norway spruce. The proposed methodology has the potential for easy integration with operational laser scanner-based stand inventories.

1. Introduction

Small footprint airborne laser scanning (ALS) has become one of the most common remotely sensed data sources for analysing the tree canopy structure at the scale of operational forest management. Area-based ALS forest inventory is nowadays used commercially in the Nordic countries (Næset 2004a,b, 2007), relying on sound statistical principles and commonly accepted field inventory practices (Hyypä *et al.* 2008). However, area-based ALS forest inventory is probably not suitable for the entire Nordic forest area. Especially for heterogeneous forests, the stand-wise inventories do not provide the detailed information required by the sustainable forest management and planning programmes. For harvesting operations, biomass and carbon stock estimation, forest damage assessment and forest monitoring single tree data might be required, and it has been shown that small footprint ALS can provide such detailed

*Corresponding author. Email: liviue@umb.no

information (Yu *et al.* 2003, Chen *et al.* 2006, Koch *et al.* 2006, Solberg *et al.* 2006, Hyyppä *et al.* 2008).

Since the pioneering work of Hyyppä and Inkinen (1999) and Hyyppä *et al.* (2001a,b), various algorithms for the extraction of single tree information using ALS data have been developed and reported as being appropriate for various forest conditions. Methods for single tree delineation based on ALS data have traditionally employed tree height–crown diameter relationships (Pitkänen *et al.* 2004, Popescu and Wyne 2004, Chen *et al.* 2006, Koch *et al.* 2006), multiscale techniques (Persson *et al.* 2002, Brandtberg *et al.* 2003, Pitkänen *et al.* 2004, Falkowski *et al.* 2006, Zhao and Popescu 2007) or information extracted from the full three-dimensional structure of the laser cloud (Wang *et al.* 2008, Rahman and Gorte 2009).

In Norway, there is lack of research regarding the development of single tree detection methods adapted to the particularities of the Norwegian forest conditions, although ALS data are widely used for area-based forest inventories (Næsset 2007). A singular but interesting contribution is the algorithm for tree crown delineation with enhanced control of the crown segment proposed by Solberg *et al.* (2006), which requires mapped stems for fine-tuning.

The accuracy of ALS-based single tree detection algorithms is mainly influenced by forest conditions (Falkowski *et al.* 2008, Kaartinen *et al.* 2008) and the spatial resolution of the remote-sensing data (Wulder *et al.* 2000, 2002, Tesfamichael *et al.* 2009). Most of the applications that are focused on single tree delineation are based on a rasterized representation of the canopy surface in the form of a canopy height model (CHM), which is a very convenient framework for integrating common image-processing methods. The CHM is obtained by interpolating the laser data into a grid with a predefined spatial resolution – i.e. the size of the grid cell or pixel. Using fine spatial resolutions can produce unnecessary details (or noise) at the scale of single crowns, while at coarse resolutions the tree crowns are merged. Appropriate low-pass filtering strategies can eliminate a large amount of noise, but little can be done to improve the single tree detection if the CHM's spatial resolution is too coarse. The rule of thumb recommended by Hyyppä *et al.* (2001a,b) is to use a grid size of 0.5 m to represent the tree crowns in boreal forests, while Heinzl *et al.* (2008) suggested using lower-resolution CHMs for delineating tall trees and finer resolutions for smaller trees. In some studies the pixel size was directly related to laser echo density (Persson *et al.* 2002, Chen *et al.* 2006) or indirectly to laser echo spacing (Leckie *et al.* 2003), but these criteria can lead to unreasonable spatial resolutions when the echo density is either too high or too low.

Accurate estimates of single tree characteristics obtained using ALS data can potentially increase the value of area-based inventories, providing a detailed description of dominant and co-dominant trees that can be transformed into merchantable assortments of wood products or can be used for the retrieval of various silvicultural and ecological indicators. Performing area-based forest inventories and single tree delineation simultaneously would be feasible and useful if some of the stand-wise estimates provided by the former method could be used to improve the latter, such that the single tree information is obtained without extra costs (except for the acquisition costs for higher-density ALS data). One of the stand characteristics that is relatively easy to measure during the field inventory is the stem number, and this biophysical attribute might be useful for single tree delineation – e.g. by providing prior information for the choice of an appropriate spatial resolution of the CHM and the amount of smoothing to be applied. However, guiding the tree delineation process using this particular

variable may pose a challenge because the expected accuracy of the stem number prediction in area-based ALS forest inventory for typical Norwegian conditions is rather low, varying between 10% and 50% (Næsset 2007).

The two main objectives of this study are

1. to assess whether the area-based stem number estimates resulting from operational ALS-based forest inventories can guide the CHM creation and single tree delineation process;
2. to assess the accuracy of the single tree variables obtained from ALS data by applying the methodology developed in (1).

2. Material

2.1 Study area and field data

The study area of 960 km² is located in the municipality of Aurskog-Høland, situated in south-eastern Norway (59° 80' N, 11° 55' E, 172–388 m a.s.l.). A sample of 40 circular plots was established and field data were collected between October 2007 and April 2008. The plot locations were distributed along five systematically located and separate strips of ALS data. The locations were selected to cover a broad range of forest conditions with respect to tree species composition, site quality and stage of stand development. Hence 10 plots were established in young forests, 13 plots in mature forests with poor site quality and 17 plots in mature forests with good site quality. Four plots located in dense young forests were established with an area of 500 m² (12.62 m radius) and the remaining 36 plots had an area of 1000 m² (17.84 m radius). In terms of basal area (BA) per hectare, 17 plots were dominated by Norway spruce and 23 plots by Scots pine. The terrain conditions across the study area can be considered gentle compared to average terrain conditions in the productive forests in Norway, but the local topography still varies significantly among plots.

The plot centre coordinates (x , y) were determined using dual-frequency Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) measurements, acquired by two Topcon dual-frequency receivers (Topcon Positioning Systems Inc., Livermore, CA, USA). Differential post-processing was performed using Pinnacle 1.00 software (Anon 1999), using the corrections provided by temporary base stations installed on National Geodetic Grid ('Stamnett') points. The average *a priori* planimetric accuracy of the plot centre position after post-processing was about 0.12 m. According to Næsset (2001), the real positional error can be considered twice as large as the standard error resulting from the differential post-processing.

On each plot, tree species and diameter at breast height (dbh) were recorded for all trees with dbh \geq 5 cm, and their stem centre locations were mapped using a Sokkia SET5F total station (Sokkia B.V., LJ Capelle a/d IJssel, The Netherlands). In total, 4299 trees were recorded (52% spruce, 34% pine and 14% deciduous species). The dbh range for the trees of dominant species was divided into five diameter classes, and two trees were selected as sample trees for height and crown measurements in each class. Where applicable, up to five trees from each of the secondary species were also sampled proportionally with the tree's basal area using a relascope with a BA factor of 1 (Gregoire and Valentine 2008). For all sample trees, the heights (h) were measured using a Vertex hypsometer (Haglöf Sweden AB, Långsele, Sweden), and the crown projections along the eight cardinal directions (N, NE, E, SE, S, SW, W and NW) from the stems were determined using a metallic tape. Crown diameter (cd) was

Table 1. Field data summary: field plots.

Characteristics	Range	Mean
No. of sample trees	11–20	17
h_L (m)	10.3–25.4	16.27
d_g (m)	0.10–0.26	0.171
N (ha ⁻¹)	460–2440	1171
BA (m ² ha ⁻¹)	13.6–42.0	25.19
Tree species distribution by percentage of basal area (%)		
Spruce	0–98	42
Pine	0–100	49
Birch	0–34	8

Note: h_L , Lorey's height; d_g , mean basal area diameter; N , stem number; BA, plot basal area.

Table 2. Field data summary: single tree variables.

Single tree Variable*	Spruce			Pine			Birch		
	Range	Mean	St. dev.	Range	Mean	St. dev.	Range	Mean	St. dev.
dbh (cm)	5.0–49.5	14.4	7.56	5.0–48.0	16.7	8.04	5.0–64.0	11.6	6.96
h (m)	3.5–29.4	14.6	6.12	5.1–27.1	14.6	4.39	3.7–25.9	13.8	5.23
cd (m)	1.2–7.9	3.2	1.05	0.9–6.8	3.3	1.26	1.3–8.5	3.4	1.31

Notes: *only for sample trees; dbh, breast height diameter; h , total tree height; cd, crown diameter; st. dev., standard deviation.

calculated as the diameter of a circle with the same area as the crown projection. In total, 669 sample trees (273 Norway spruce, 266 Scots pine and 130 deciduous trees) were selected across the plots, with an average of 17 trees per plot (1171 stems/ha). The most represented deciduous tree species in the data set was birch. Tree-specific (Norway spruce, Scots pine and birch) height–diameter regression models were developed using the sample trees as training data (Magnussen *et al.* 2010). These models were used to predict the heights of the trees not included in the sample. Mean height (h_m) and Lorey's mean height (h_L) were calculated for each plot using field-measured and predicted tree heights, and plot BA was computed from the dbh measurements. A summary of the field data is presented in tables 1 and 2.

2.2 Laser scanner data

ALS data were acquired at a mean flying altitude above ground of 800 m. Data were collected under leaf-on canopy conditions on 6 June 2006 using a PA31 Piper Navajo fixed-wing airplane. The strips were E–W oriented, with an N–S spacing of approximately 8.7 km. Data were acquired with an Optech ALTM 3100 system (Optech Inc., Vaughan, ON, Canada) operating with a pulse repetition frequency of 100 kHz, a scan frequency of 70 Hz and a half-angle of 5°. Up to four echoes per pulse were recorded (approximately 1.4 echoes per emitted pulse), resulting in an echo density of 10.4 echoes m⁻². Laser data processing was accomplished by the contractor (Blom Geomatics AS, Oslo, Norway). The GPS/INS (Inertial Navigation System) processing was performed using Applanix POSPac software (Applanix Corp., Richmond Hill, ON, Canada), and then planimetric coordinates (x and y) and ellipsoidal heights

were obtained by processing the navigation and laser data using the Optech REALM tools (Optech Inc.). The three-dimensional adjustment and classification (canopy and ground) was performed using TerraScan and TerraModel software from TerraSolid Ltd, Helsinki, Finland. A triangulated irregular network (TIN) model representing the terrain surface was built using the TerraScan software, and the heights of all laser returns relative to the interpolated ground surface were obtained by subtracting the TIN heights from the laser echo heights. According to the contractor's technical report, the expected accuracy of the relative heights was ± 10 cm. Finally, the 'first' and 'single' echoes were extracted within buffers of 10 m extending around the plot borders and retained for further analysis, resulting in a density of approximately 7.4 echoes m^{-2} , which is considered to be relatively high (Hyypä and Inkinen 1999, Persson *et al.* 2002).

3. Methods

The main steps of the standard workflow for single tree crown delineation and single tree variable extraction from CHMs can be summarized as follows (Hyypä *et al.* 2001a, 2008): (1) creation of a CHM, (2) smoothing the CHM using a low-pass spatial filter for removing spurious details, (3) regional maxima detection and segmentation of the CHM for delineating single tree crowns and (4) extraction of single tree variables using various metrics derived from the CHM regions or laser echoes contained in each segment.

In this study, two adaptive approaches for guiding the tree crown delineation are proposed: Poisson 1 and Poisson 2. The crown delineation method described by Hyypä and Inkinen (1999) and Hyypä *et al.* (2001a) was used as a reference method (labelled 'RefMeth') for assessing the performance of the Poisson methods. In the Poisson 1 method, adaptive low-pass filtering was applied to CHMs created at pre-defined spatial resolutions. This approach considers that the CHMs for all 40 plots are created at the same spatial resolution, while the size of the low-pass filter varies for each plot. The Poisson 2 method allows for the CHM spatial resolution and the size of the low-pass filter to vary by plot.

Both Poisson methods rely on a mild assumption about the process describing the spatial stem distribution at the plot level. More exactly, it was assumed that tree locations are generated by a homogeneous spatial Poisson process, thus making possible the calculation of the expected nearest neighbour distance (D_{NN}) between the closest pairs of trees, as follows:

$$D_{\text{NN}} = 0.5\sqrt{\frac{A}{N}}, \quad (1)$$

where A is the plot area (m^2) and N is the number of stems on the plot (Clark and Evans 1954).

Exploratory analysis was performed to assess the spatial distribution of trees on the plots, following the procedure described in detail by Salas *et al.* (2006). The analysis was conducted using the 'spatstat' package (Baddeley and Turner 2005), version 1.17–5, of the open-source statistical environment R (R Development Core Team 2009). The null hypothesis of a random process describing the tree spatial distribution at the plot level was tested using 95% confidence envelopes of the Ripley's $\hat{L}(t)$ -estimator obtained through Monte Carlo simulations, using isotropic edge correction

(Ripley 1988). The hypothesis of a random spatial process held for 24 plots (9 dominated by 6 spruce and 15 by pine), clustering was identified on 12 plots (6 dominated by spruce and 6 by pine) and regularity was found on 4 plots (2 dominated by spruce and 2 by pine). However, the spatial distribution is also influenced by the silvicultural treatments applied within each stand. Since the stand history was not available for most of the stands, the results of the spatial analysis had to be interpreted with caution.

In practice, using equation (1) requires that the stem number (N) and the forest areas (A) are known. While the forest areas are usually available, the stem numbers can be estimated from the ALS data using a predictive regression model, following the standard area-based approach used in operational forest inventories (Næsset 2002). With plots of various sizes (500 and 1000 m²), the regression equation (root mean squared error (RMSE) = 24.5%) fitted for stem number prediction was

$$\ln\left(\frac{N}{A}\right) = 0.03091 - 0.0809\bar{h} - 2.3133(CV), \quad (2)$$

where \bar{h} is the mean echo height (m) and CV is the coefficient of variation of echo heights. In a practical application, the regression equation would be used to predict the stem number for forest areas not included in the training data set. Thus to mimic the stem number prediction on new plots, a leave-one-out data-splitting procedure was employed. Each plot was therefore in turn withheld from the training data set, and equation (2) was re-fitted for the remaining 39 plots. The RMSE increased to 28.9%, and the new estimates of stem number obtained by applying the leave-one-out procedure (labelled \hat{N}_{Loo}) were plugged into equation (1) to calculate the mean nearest neighbour distances. The resulting mean nearest neighbour distances varied between 1.0 and 2.4 m, with an average of 1.5 m and a standard deviation of 0.3 m. Because spruce-dominated plots tend to have higher stem densities compared with pine-dominated plots, the average D_{NN} was slightly smaller in the former (1.35 and 1.67 m, respectively).

An overview of the methodology is presented in figure 1. The applied methods are detailed in the following sections.

3.1 CHM processing

The CHMs were obtained using the natural neighbour interpolation algorithm known to perform very well in interpolating dense, irregularly spaced data (Falkowski *et al.* 2008, Bater and Coops 2009). Before applying low-pass filtering, a thresholding procedure was applied to the CHM of each of the 40 plots to separate the canopy regions from the background. Visual assessment of the vertical distribution of laser echo heights corroborated by field observations did not indicate a clear presence of multilayer canopies for our data set; thus the three-dimensional laser echo clouds were split into two groups – canopy and background – using the minimum error histogram thresholding method (Kittler and Illingworth 1986). The method sets the threshold by minimizing the classification error between background and canopy echoes assuming that the echoes in the two groups are normally distributed. The CHM heights below the threshold value were considered to represent the forest floor and understory vegetation and were set to 0. The laser echo height threshold varied from 1.00 to 3.25 m,

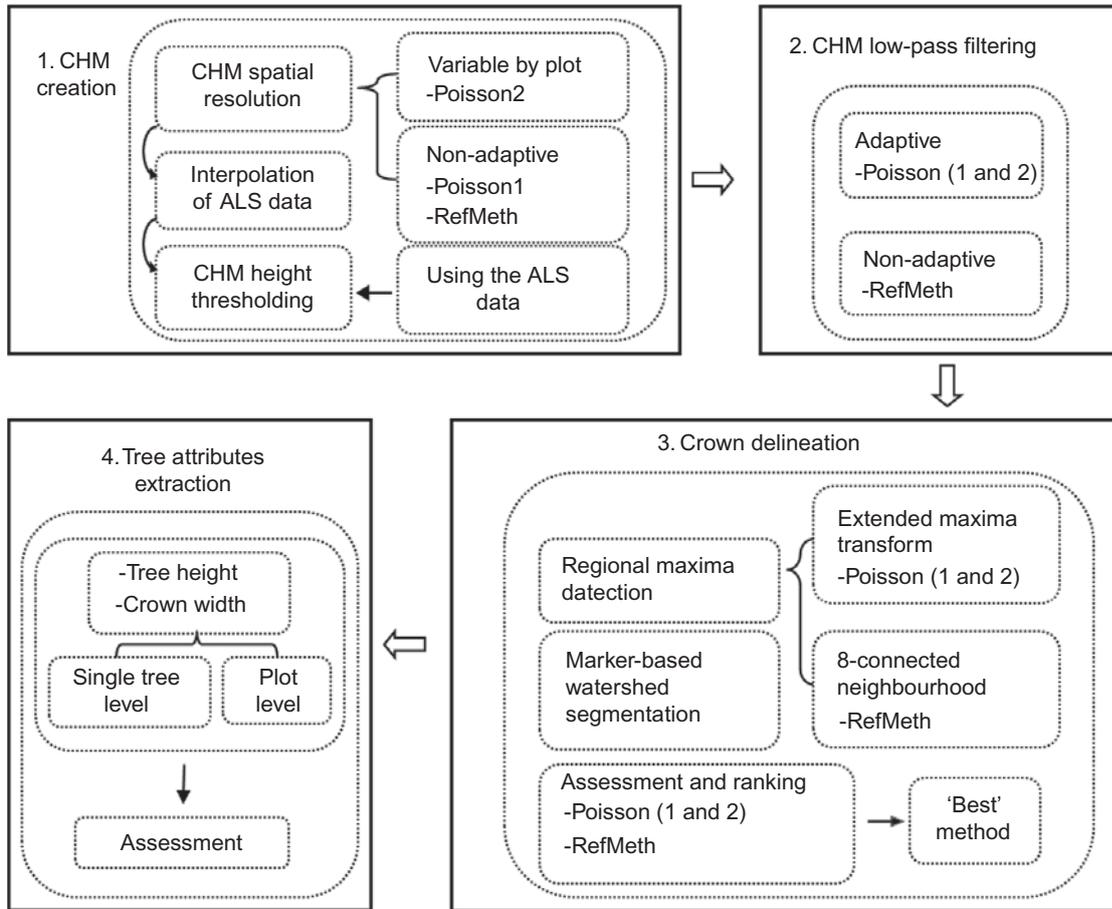


Figure 1. Methodological flow chart.

with an average of 1.2 m. On all plots, the threshold values were consistently smaller than the heights of the smallest sample trees, and it was considered that the chances of identifying small trees were not affected by removing height information from the CHM.

Visual inspection of the CHMs of the plots did not reveal any speckle noise, which eventually would have required non-linear filters (e.g. a median filter) for noise reduction. Low-pass filtering was applied in the spatial domain using binomial filters. The binomial coefficients representing the filter weights were obtained from Pascal's triangle. Following Popescu and Wyne (2004), circular kernels were used to mimic circular tree crown shapes.

3.2 Adaptive tree crown delineation methods

The CHMs created at fine spatial resolutions depict the canopy surface in greater detail, and in such situations larger spatial filters (larger f_{size}) are required to remove spurious local maxima to keep the number of false positives low. At coarse resolutions the small-scale spatial variation in the CHMs is reduced, and using large spatial filters will lower the detection rates owing to regional maxima attenuation. The challenge is then twofold: (1) to represent the tree crowns using an appropriate CHM resolution and (2) to use a filter size such that a reasonable number of trees are detected while the proportion of false detected trees is kept low.

3.2.1 The Poisson 1 method. Here, it was assumed that the tree stems are randomly located at the plot level and that the stem number on each plot is obtained using a regression model (equation (2)).

Assuming randomly located tree stems at the plot level and having the stem number estimated by a regression model (equation (2)), the mean nearest neighbour distance can be predicted for each plot by substituting the real stem number, N , in equation (1) with the estimated stem number, \hat{N}_{Loo} . For each plot the filter size (f_{size}) was then set equal to the predicted mean nearest neighbour distance ($D_{\text{NN},p}$):

$$f_{\text{size}} = D_{\text{NN},p} = 0.5 \sqrt{\frac{A}{\hat{N}_{\text{Loo}}}}. \quad (3)$$

The choice of grid resolution for analysing the spatial data has been much debated and it is accepted that an ideal grid resolution generally does not exist (Hengl 2006). When the stem positions are available from field measurements, an optimal CHM resolution can be obtained by applying a trial-and-error procedure, where the performance of the method is assessed using several CHM resolutions and the resolution giving the best results is selected. Feasible candidates for the CHM resolution can be selected considering that objects smaller than the low-pass filter will be affected more than the larger objects (Gonzales and Woods 2002). On the assumption of having randomly located trees, a lower bound of the CHM resolution can be determined considering that a filter having the minimum meaningful size of 3 pixels \times 3 pixels should not simultaneously cover the tops of two neighbouring trees situated at the mean neighbour distance from each other (figure 2).

Following this reasoning, the maximum allowable pixel size, $p_{s,\text{max}}$, can be calculated for each plot as

$$p_{s,\text{max}} = \frac{f_{\text{size}}}{3}, \quad (4)$$

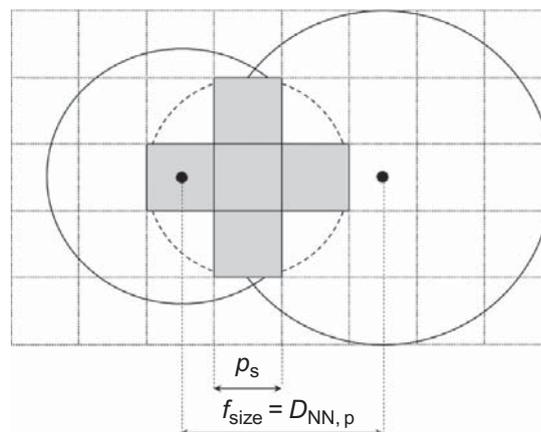


Figure 2. Calculation of pixel size (p_s): the maximum pixel size is chosen such that a 3 pixel \times 3 pixel filter does not cover the tree tops of two neighbouring trees situated at $D_{\text{NN},p}$ from each other. The solid line circles illustrate the outline of the crowns of two neighbouring trees found within the range of $D_{\text{NN},p}$ and the dotted line circle outlines the 3 \times 3 circular low-pass filter shown in grey. The two black dots indicate the tree tops.

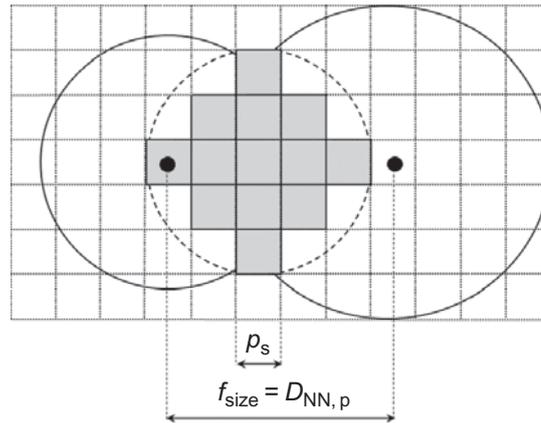


Figure 3. Representation of a circular spatial filter of size f_{size} , given the pixel size (p_s) and the expected nearest neighbour distance between trees.

resulting in $p_{s,\text{max}}$ values varying between approximately 0.3 and 0.8 m, with an average of 0.5 m and a standard deviation of 0.10 m. The minimum pixel size was subjectively set to 0.1 m, and a series of CHMs were created at resolutions from 0.1 to 0.8 m in steps of 0.1 m. The f_{size} values obtained in metres were then scaled to pixels based on the CHM resolution, while constraining the f_{size} to be an odd number (figure 3). The performance of the Poisson 1 method was judged using the f_{size} values (in pixels) obtained for each series of the CHM resolutions.

3.2.2 The Poisson 2 method. The main difference between the Poisson 1 and Poisson 2 methods is that the latter allows one to obtain the CHM resolution for each plot. As in the case of the Poisson 1 method, the filter size was obtained from equation (3). Since stem positioning is usually not performed during operational forest inventories, we developed a strategy for obtaining reasonable CHM resolutions at the plot level, using the number of local maxima obtained from the unsmoothed CHM and the predicted stem number per plot, (\hat{N}_{Loo}).

The number of local maxima found after running low-pass filtering was previously used to optimize the smoothing (Culvenor 2000, Pouliot and King 2005, Pouliot *et al.* 2005). In our study, we used the number of local maxima found in the unsmoothed CHM to obtain a reasonable CHM resolution. Given that the low-pass filtering can drastically reduce the number of local maxima and assuming an ideal situation when all tree tops are ‘visible’ in the CHM, the number of local maxima found before applying the low-pass filtering should be at least equal to the number of trees. The approach is illustrated in figure 4, where the number of local maxima (values on the y -axis) is represented as a function of several candidate pixel size (p_s) values (on the x -axis) obtained using equation (4). The optimal p_s value, at which the number of local maxima equals the predicted stem number (\hat{N}_{Loo}), is then selected for CHM creation. It was considered that using finer resolutions (that is, resolutions producing more local maxima than the number of trees) is not necessary because, in reality, many of the smaller trees would be fully covered by the dominant ones, and the stem numbers used by the method usually deviate from the real values, being predicted by a regression model.

Our exploratory results indicated that 90% of the p_s values (for 36 plots) obtained using N and \hat{N}_{Loo} differed by less than 0.1 m, and in only 10% of the cases (four plots) was the difference as high as 0.2 m.

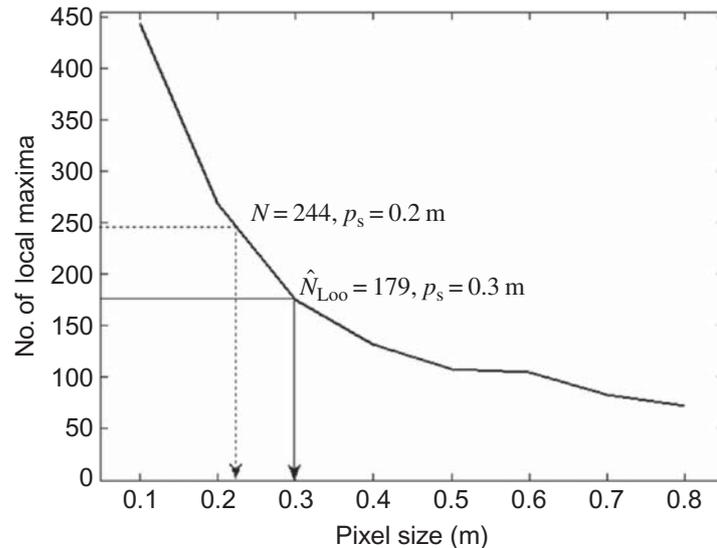


Figure 4. Determining the CHM resolution using the number of local maxima extracted from the unsmoothed CHM and the predicted number of stems per plot ($\hat{N}_{Loo} = 179$). The optimal pixel size (0.3 m) is the p_s value at which the number of local maxima (the values on the y-axis) equals \hat{N}_{Loo} . When the field-measured stem number ($N = 244$) is used instead of \hat{N}_{Loo} , the optimal p_s value is 0.2 m. By convention, the pixel size was allowed to vary in steps of 0.1 m.

Given the higher stem density in spruce-dominated forests, the average CHM resolution for plots dominated by spruce was also higher (0.3 m) than that for pine-dominated plots (0.4 m).

3.3 CHM segmentation

Various segmentation approaches have been explored across different forest conditions – e.g. watershed algorithms (Chen *et al.* 2006, Kwak *et al.* 2007, Zhao and Popescu 2007, Heurich 2008), region growing (Solberg *et al.* 2006), pouring algorithm (Koch *et al.* 2006) and slope-based segmentation (Hyypä *et al.* 2001a, Persson *et al.* 2002). In the present study, we used the marker-based watershed algorithm (Soille 2003) for CHM segmentation because it is less prone to produce over-segmentation than the typical watershed segmentation algorithm. The segmentation algorithm ensures a 1:1 relationship between object markers and delineated CHM segments, so the segmentation success depends on how accurately the object markers represent the objects. Having already established the approaches for controlling the filter settings to be used with the Poisson methods, we decided to use the extended maxima transform to detect CHM regional maxima after performing the adaptive low-pass filtering.

The extended maxima transform (Soille 2003) is a morphological operation that allows for regional maxima detection in two steps: (1) regional maxima are found from the intensity image by morphological operations and (2) all regional maxima with values below a certain threshold, h_{max} , from the background are removed. The scalar value, h_{max} , can be considered as the average intensity value separating regional maxima across the image or, in our case, the average height difference between the tree tops and their surroundings. The number of regional maxima produced by this method is inversely proportional to the value of h_{max} . Marker-based watershed segmentation using an extended maxima transform for marker selection has been previously used

by Chen *et al.* (2006) and Kwak *et al.* (2007). In their approach, the optimal threshold, h_{\max} , was obtained empirically using training data, but we aimed to obtain the h_{\max} value directly from the CHM. Hence, the local means were obtained from the smoothed CHM using averaging moving windows of size f_{size} , and local residuals were calculated by subtracting the local means from the smoothed CHM. The threshold value h_{\max} was then set to a percentile obtained from the distribution of the residuals. During the calculations, the CHM regions containing background pixels were discarded. Choosing a low percentile will increase the detection rate, while a higher percentile will decrease the number of regional maxima detected. For our data, little change occurred when the h_{\max} value varied between the first quartile and the median due to the CHM smoothness obtained after low-pass filtering. For this reason the h_{\max} value was set as the 25th percentile of the absolute local residuals estimated at the plot level.

The result of the marker-based watershed segmentation consists of a labelled matrix in which positive elements correspond to the watershed regions (delineated segments). The zero labels (also called watershed pixels) are areas that do not belong to a unique watershed, and they create contiguous separation lines (watershed ridge lines) between watershed regions (Anon 2010). The laser echoes and tree stem positions were linked to the watershed regions using their coordinates, so we could identify the laser echoes and the trees located inside each of the segments. Morphological cleaning was performed to eliminate spurious pixels produced after height thresholding, and dilation using a disc-shaped structural element of one-pixel radius was applied to each segment before linking it to the trees and laser echoes in order to extend the segment borders into the watershed ridges.

3.4 The reference method

The method described by Hyypä and Inkinen (1999) and Hyypä *et al.* (2001a) was used as a ‘RefMeth’ to assess the results obtained with the Poisson methods. The algorithm runs a CHM low-pass filter using a 3 pixel \times 3 pixel discrete kernel with the weights obtained from the binomial approximation of the Gaussian function, and the regional maxima are found at the locations where the values in the smoothed CHMs are larger than their eight neighbours. For assessment, the Poisson 1 and ‘RefMeth’ methods were run on identical CHMs. The ‘RefMeth’ and ‘Poisson 2’ methods were assessed on CHMs created at resolutions related to the number of local maxima found by each method.

3.5 Extraction of single tree variables

The single tree variables estimated in this study – i.e. tree height and crown diameter – were obtained from the three-dimensional coordinates of laser echoes located inside the CHM segments resulting from the segmentation process.

The (x, y, z) coordinates of the highest laser echo within each segment were considered to represent the tree top position and were assigned to the highest tree located inside the same segment.

Crown diameters were estimated in two steps. First, the laser echoes in each segment were separated into two groups – background and canopy. Preliminary results indicated that the minimum error histogram thresholding did not perform well at the single tree level, when the number of echoes varies substantially among segments. Instead we used the Otsu histogram thresholding method (Otsu 1979), which sets

the threshold by minimizing the intra-class variance of the background and canopy echo heights. After thresholding, crown areas were derived by dividing the number of echoes above the thresholds obtained for each of the segments by the average echo density of the plot (Morsdorf *et al.* 2004). The crown diameters were subsequently derived, considering that the tree crown projections are circular.

3.6 Accuracy assessment

One problem that affects the accuracy assessment is that some of the segmented regions cross the plot boundaries, covering areas where field measurements are missing. To avoid such problems, the results were assessed using only a subset of the trees initially mapped on each plot, which consisted of trees from the segments entirely located inside the plots. The trees located inside the segments crossing the plot boundaries were then discarded, and the initial stem numbers obtained from the field inventory were modified accordingly, obtaining plot-wise adjusted stem numbers, which were further considered to represent the ground truth.

3.6.1 Delineation errors. Following Pouliot *et al.* (2005) and Pouliot and King (2005), an accuracy index (AI; a percentage) quantifying the trade-off between omission and commission errors can be defined as

$$AI(\%) = 100 \left(1 - \frac{N_{OM} + N_{COM}}{N_{adj}} \right), \quad (5)$$

where N_{adj} is the adjusted stem number, and N_{OM} and N_{COM} represent the omission and commission errors, respectively. By omission errors (N_{OM}), we mean the number of undetected trees, while commission errors (N_{COM}) refer to the number of local maxima that do not correspond to the tree tops. If the omission errors (OM) and commission errors (COM) are expressed as percentages, then

$$OM = 100 \frac{N_{OM}}{N_{adj}} \quad (6)$$

and

$$COM = 100 \frac{N_{COM}}{N_{adj}}. \quad (7)$$

The AI can then be written as

$$AI = 100 - ((OM) + (COM)). \quad (8)$$

It should be noticed that equations (5) and (8) allow for negative AI values in the case of very poor performance, when the sum of delineation errors is larger than the actual number of trees. Given the omission errors, the detection rate can subsequently be calculated as $100 - (OM)$. Finally, the Poisson and RefMeth methods were ranked in order of increasing AI.

3.6.2 Single tree variables. The delineation results produced by the method with the highest AI were further used to assess the accuracy of the estimated single tree variables.

In a practical application, each delineated segment would be treated as representing a single tree, regardless of the number of trees it actually contains. Thus, the effect of tree grouping (more than one tree located inside the same segment) would affect the estimation of single tree variables compared with the situation where the segments contain only one tree. For this analysis, we used only those segments where the laser tops were linked to sample trees. Further, the effect of the tree species on the height and crown width estimation was also assessed, and in this case we used only identified reference trees standing alone inside their corresponding segments.

At the plot level, the mean height and mean crown diameter obtained from single tree estimates obtained from ALS data were compared with the mean height, Lorey's height and mean crown diameter derived from field data. For this analysis, the ALS-based single tree estimates for height and crown diameter also included the values corresponding to false positives (commission errors), which in a practical application would be taken into account as real trees.

The strength of the relationship between the single tree variables estimated from the ALS data and their field-based counterparts was assessed using the Pearson linear correlation coefficient, r . The estimation accuracy was assessed by using the errors e_i calculated as the differences between field- and ALS-based single tree estimates. The mean errors and the root mean squared deviations (RMSD) calculated for the n identified trees as

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} \quad (9)$$

were also reported.

The null hypothesis of having the group mean errors greater or less than zero was assessed using paired one-tailed t -tests (significance level $\alpha = 0.025$), and one-way ANOVA ($\alpha = 0.05$) was used to test for significant differences between group means. Where significant differences were found, multiple comparison tests based on Tukey's least significant difference procedure were used for assessing pairs of group mean errors that were significantly different.

4. Results

4.1 Delineation accuracy

The overall delineation accuracy results are presented in table 3. The highest accuracy (43.6%) was obtained by the Poisson 1 method for a CHM resolution of 0.2 m. The worst results (-95.4%) were obtained when using the RefMeth for a 0.1 m CHM resolution.

The performance of the Poisson 1 method increased from 41.2% to a maximum of 43.6% when the CHM resolution decreased from 0.1 to 0.2 m, and it decreased steadily as the CHM resolution became coarser, attaining the minimum (31.8%) for a pixel size of 0.8 m. The AI of the RefMeth also increased with finer pixel size having a peak of 41.8% at 0.3 m, and then steadily decreasing to 32.7% at 0.8 m resolution.

Table 3. Overall delineation accuracy.

Method	CHM resolution (m)	No. of trees (adjusted)	*OM (%)				†COM (%)	‡AI (%)
			Spruce	Pine	Birch	Total		
Poisson 1	0.1	4126	51.8	32.4	54.1	45.5	13.3	41.2
RefMeth		4161	35.2	20.1	34.8	30.5	164.8	-95.4
Poisson 1	0.2	4105	55.8	36.5	59.2	50.2	6.2	43.6
RefMeth		4137	47.6	27.6	47.7	41.6	33.8	24.6
Poisson 1	0.3	4113	57.1	39.3	60.3	52.3	5.3	42.4
RefMeth		4125	52.8	32.6	52.5	47.2	11.1	41.8
Poisson 1	0.4	4091	59.9	42.8	63.4	55.5	3.7	40.8
RefMeth		4103	57.4	38.6	59.5	52.4	5.9	41.7
Poisson 1	0.5	4084	63.4	47.7	66.4	59.2	3.2	37.6
RefMeth		4088	61.5	44.4	63.7	57.0	3.9	39.0
Poisson 1	0.6	4066	67.1	50.7	69.5	62.8	2.2	35.0
RefMeth		4076	66.2	49.5	67.6	61.8	2.6	35.6
Poisson 1	0.7	4074	66.3	48.9	69.1	62.0	2.4	35.6
RefMeth		4076	65.6	48.1	68.6	61.4	2.7	35.9
Poisson 1	0.8	4037	70.1	54.1	74.3	66.5	1.7	31.8
RefMeth		4045	68.7	52.8	73.5	65.3	2.0	32.7
Poisson 2	0.2–0.5	4104	57.6	42.2	62.5	53.7	4.3	42.0
RefMeth	0.2–0.6	4113	52.3	36.1	54.5	48.3	9.8	41.8

Note: *omission error, †commission error (%), ‡accuracy index.

The largest number of commission errors was produced by the RefMeth at 0.1 m resolution (164.8%), while the Poisson 1 method produced only 13.3% false positives for the same pixel size. At CHM resolutions of 0.3 and 0.4 m, the Poisson 1 and benchmark methods performed nearly identically, achieving an AI of approximately 41–42%.

The Poisson 2 and the RefMeth achieved nearly equal AIs (42.0% and 41.8%, respectively) when a variable CHM resolution was used, the RefMeth having a higher detection rate (51.7% against 43.3%) at the expense of a larger number of false positives (9.8% against 4.3%).

All methods produced a higher percentage of omission errors for spruce (approximately 35–70%) and deciduous trees (approximately 35–74%) compared with pine (approximately 20–54%).

Because the Poisson 1 method reached the highest overall AI, the results of this method were further analysed per plot (table 4). The Poisson 1 method could delineate an average 51.8% of the adjusted number of trees on each plot, producing 6.6% false positives. The method could delineate 54.6% of the trees on pine-dominated plots (which is about 7% more than that for spruce-dominated plots), but the average commission error level was not affected much by the species composition, being around 6–7%.

The delineation accuracy (AI) was not found to be significantly influenced by the number of trees (figure 5) and by their spatial distribution (figure 6), either for the spruce-dominated or for the pine-dominated plots.

Table 4. Plot-wise average delineation errors and accuracy indices for the dominant tree species.

	Spruce			Pine			Total		
	OM (%)	COM (%)	AI (%)	OM (%)	COM (%)	AI (%)	OM (%)	COM (%)	AI (%)
Min	13.5	2.4	29.0	15.9	2.0	30.0	13.5	2.0	29.0
Max	66.2	19.3	80.2	60.4	15.5	81.0	66.2	19.3	81.0
Mean	52.0	7.0	41.0	45.4	6.3	48.4	48.2	6.6	45.2
Std. dev.	12.79	4.56	13.30	11.50	3.39	12.78	12.36	3.89	13.35

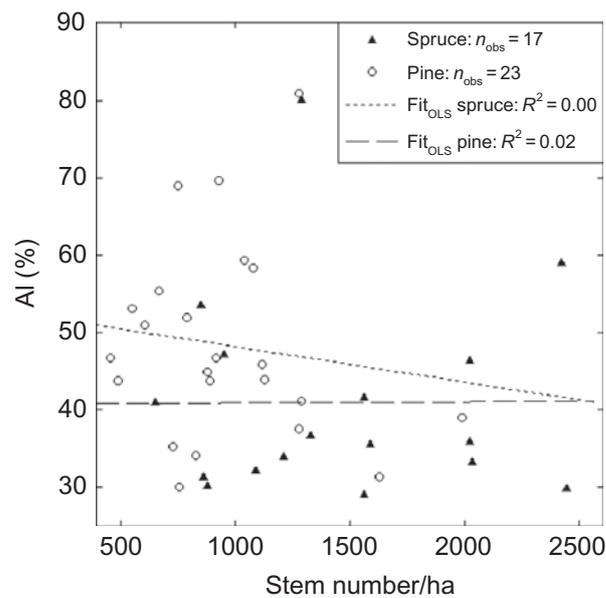


Figure 5. Relationship between the delineation accuracy (AI) and the stem number, by tree species.

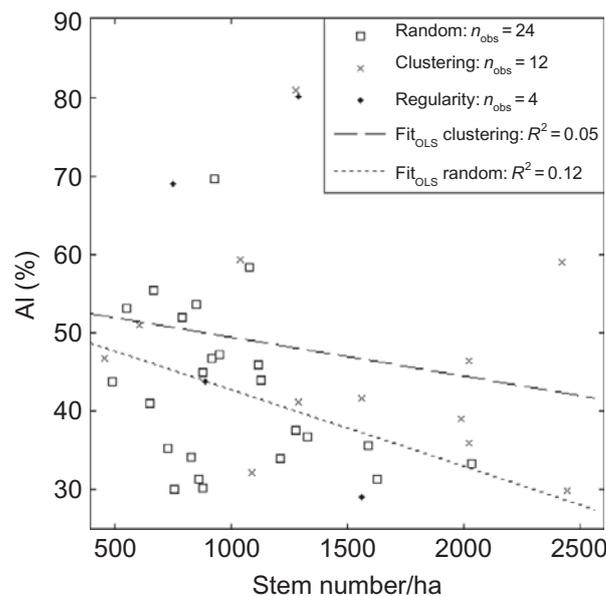


Figure 6. Relationship between the delineation accuracy (AI) and the stem number, by spatial distribution of the stems.

4.2 Single tree variable extraction

4.2.1 Accuracy of single tree variables. The Poisson 1 method identified 423 sample trees from a total of 669, and these trees were further used to compare the field-measured attributes with the corresponding laser-based estimates.

The influence of the number of trees located inside the watershed segments (1, 2 and >2 trees) on the ALS-based single tree estimates is shown in table 5.

Tree heights were systematically underestimated (table 5), and the ANOVA test ($p > 0.5$) indicated that the magnitude of the bias was not dependent on the number of trees per segment. Although the estimation slightly improved with an increase in the number of trees in a segment (table 5), the underestimation was found to be statistically significant by the paired t -tests ($p < 0.001$). Paired t -tests indicated that crown diameters were unbiasedly estimated for segments containing only one tree ($p > 0.1$), but significant overestimations occurred when at least two trees were contained in the segments ($p < 0.001$ and $p < 0.0001$, respectively). Moreover, multiple comparison tests indicated that the biases for segments containing two and more than two trees were also significantly different, the bias increasing with the number of trees in the segments.

From the initially matched trees, 189 sample trees (78, 87 and 24 Norway spruce, Scots pine and deciduous trees, respectively) were found standing alone inside their corresponding segments. The differences between the laser- and the field-based single tree estimates for sample trees grouped by tree species are presented in table 6.

For all species the single tree heights were systematically underestimated (paired t -test: $p < 0.001$), but the ANOVA tests ($p > 0.5$) did not indicate significant differences between tree species. The t -tests showed that the estimation of crown diameters was unbiased for birch ($p > 0.5$), underestimated for spruce ($p < 0.001$) and overestimated for pine ($p < 0.0001$).

Table 5. Accuracy of ALS-based single tree estimates grouped by the number of trees located in each segment.

No. of trees in each segment	Height			Crown diameter		
	1	2	>2	1	2	>2
No. of observations	189	104	130	189	104	130
Pearson's correlation coefficient, r	0.96	0.96	0.97	0.85	0.75	0.73
Mean error (m)	0.59	0.52	0.46	0.03	-0.28	-0.63
RMSD (m)	1.35	1.22	1.05	0.59	0.91	0.81

Table 6. Accuracy of ALS-based single tree estimates for sample trees grouped by tree species.

Species	Height			Crown diameter		
	Spruce	Pine	Birch	Spruce	Pine	Birch
No. of observations	78	87	24	78	87	24
Pearson's correlation coefficient, r	0.97	0.95	0.99	0.76	0.92	0.90
Mean error (m)	0.60	0.52	0.79	0.26	-0.18	0.02
RMSD (m)	1.58	1.24	0.89	0.67	0.46	0.49

Table 7. Accuracy of ALS-based plot estimates.

Species	Mean height		Lorey's height		Crown diameter	
	Spruce	Pine	Spruce	Pine	Spruce	Pine
No. of observations	17	23	17	23	17	23
Pearson's correlation coefficient, r	0.93	0.89	0.91	0.82	0.89	0.87
Mean error (m)	0.50	0.96	2.98	2.67	-0.43	-0.56
RMSD (m)	1.69	1.03	1.88	1.24	0.34	0.25

4.2.2 Accuracy of plot-level estimates. Single tree heights and crown diameters derived using all the segments delineated from the ALS data at the plot level (including commission errors) were aggregated to obtain plot averages. The accuracy of the plot-level estimates is presented in table 7.

The mean plot height was underestimated for both of the main species, the underestimation being greater in pine-dominated plots (0.96 m) compared with spruce-dominated plots (0.50 m), but paired t -tests indicated that the underestimation was significant only in the former case ($p < 0.0001$). A strong linear correlation was found between the ALS- and the field-based mean plot heights ($r = 0.93$ for spruce-dominated plots and $r = 0.89$ for pine-dominated plots). The ALS-derived mean plot height significantly underestimated the Lorey's height ($p < 0.0001$), the underestimation being slightly higher in spruce-dominated plots (2.98 m) compared with pine-dominated plots (2.67 m). ANOVA tests ($p > 0.5$) indicated that the underestimation of Lorey's height was not significantly influenced by the main tree species. The mean crown diameters were significantly overestimated (paired t -tests: $p < 0.0001$) but the overestimations did not significantly differ between the two main tree species (ANOVA tests: $p > 0.1$). A strong linear correlation between the ALS- and field-based mean crown diameters was also found ($r = 0.89$ for spruce-dominated plots and $r = 0.87$ for pine-dominated plots).

5. Discussion and conclusions

In this study, two adaptive approaches for tree crown delineation – Poisson 1 and Poisson 2 – were developed and assessed. The adaptive behaviour of these methods is guided using stem number estimates to obtain the spatial resolution and the amount of smoothing of the CHMs. Although these approaches consider the trees being randomly positioned across the study area, departure from this hypothesis did not seem to affect the delineation results, at least not for our data set.

In general, the adaptive approaches seemed to be quite versatile, and they suit best in cases where they can be used in addition to area-based ALS forest inventories, where stem number estimates at the stand or stratum level are available. In situations lacking reference data, stem number estimates can possibly be roughly approximated using data provided by previous inventories, forest operations logs (e.g. harvesting intensity, planting distance) and national forest inventory statistics.

Although the Poisson 1 method achieved the highest accuracy according to the AI criterion, the Poisson 2 came very close and it has the advantage of not requiring trial-and-error attempts for fine-tuning. The Poisson 2 method uses only area-based

estimates as input and it does not require tree-level data for training (e.g. stem positions and crown measurements), which may be an asset in a practical application. The slightly poorer performance of the Poisson 2 method can also be seen as a consequence of using less information compared with the Poisson 1 method.

The non-adaptive RefMeth was the most sensitive to changes in CHM resolution, the number of false positives being reduced to a fifth (from 164.8% to 33.8%) by increasing the pixel size from 0.1 to 0.2 m and approximately a third by decreasing the resolution to 0.3 m (from 33.8% to 11.1%). For CHMs with 0.3–0.4 m spatial resolution, the performances of the adaptive and non-adaptive approaches were nearly equal.

The delineation results seem to be less influenced by the choice of the regional maxima detection method than by the spatial size of the low-pass filter and CHM resolution. The non-adaptive method had a slightly higher detection rate at coarser resolutions also because the regional maxima detection method was more sensitive, searching for maxima in a 3 pixel \times 3 pixel window, but reducing the threshold of the extended maxima transform would also increase the detection rate. However, the extended maxima transform with the CHM-based calibration procedure described in this article is recommended because it is more flexible and offers the possibility of refining the results independently of other processing steps.

Overall, the accuracy of our approaches is comparable to that of other published results obtained for similar forest conditions, but a direct comparison might not be appropriate because different data sets were used. Another obstacle to comparing our results with other findings is that we did not classify the trees according to their social status (e.g. following Kraft's social tree classes, Pretzsch (2010), while many previous studies have reported results following that classification, which might be relevant to forest managers. Moreover, the assessment methods vary among studies and the reported accuracy results are thus inconsistent. For boreal forest conditions, Persson *et al.* (2002) reported a 71% detection rate and very low commission errors, but the methodology used to link the field-measured trees with the crown segments served analytical purposes and it is unlikely to be used in a practical application. Using a subjective assignment of trees in social classes, Solberg *et al.* (2006) reported a 93% detection rate for dominant trees and 19% for suppressed trees, but the number of false positives was not reported. Morsdorf *et al.* (2004) reported a tree detection rate of approximately 60% in boreal-type coniferous forest conditions, yet the number of commission errors was not reported. However, it may be worth mentioning that our results were in agreement with those reported by Hyyppä *et al.* (2001b), Pitkänen *et al.* (2004) and Maltamo *et al.* (2004) for Nordic forest conditions (40–50% of the trees identified, $\leq 10\%$ false positives).

The higher detection rate in pine-dominated plots compared with spruce can be explained by the forest management practices of introducing Scots pine in low productive sites. Thus on such locations the stem number is low and the tree crowns are well separated from each other, improving the delineation results. The results deteriorated in structurally rich forest stands located on high productive sites, which are typically dominated by Norway spruce, often mixed with deciduous species.

Also, an AI that accounts only for the proportion of omission and commission errors might be difficult to interpret, and the practical significance of both types of errors should be quantified using methodologies that target the subsequent use of the inventory data and the consequences of different types of errors – for example, on management decisions and their economic implications based on the provided data

(Wulder *et al.* 2000, Eid *et al.* 2004). It should also be mentioned that performing the assessment using only delineated segments entirely located inside the plots may underestimate the omission errors, if the number of trees contained in the segments and the probability of crossing the plot boundary are correlated with the size of the segments.

The assessment of the accuracy of single tree variable estimation did not account for the errors introduced by field measurements (Lim *et al.* 2001, Persson *et al.* 2002, Maltamo *et al.* 2004) and by incorrect linking between sample trees and ALS data. Also, it is likely that a higher laser pulse density would increase the accuracy of the estimation of single tree variables. For this reason, comparing our results with the findings of other studies might not be very informative. The one growth season offset between ALS and field data acquisition is expected to have a low impact on the delineation results because no harvesting took place and no significant natural damage occurred within this period. However, we believe that the height increment can influence the accuracy of height estimates by increasing the bias resulting from the ALS measurements. Hence the reported accuracy of height estimation is rather conservative, but unfortunately we did not have the possibility of properly correcting our results. Also, it is not clear how this time offset has influenced the crown width estimation, but we do not believe that significant changes in the crown structure occurred.

Overlapping tree crowns within segments caused the overestimation of single crown areas, but the results obtained for the segments containing only one tree seem promising. Delineating segments containing more than one tree had little influence on the accuracy of the tree height estimates. The linear relationship between field- and ALS-based single tree estimates was stronger for height than for crown diameter. Among species, the height estimates for deciduous trees seemed to be more accurate, and crown diameter estimation performed better for Scots pine and deciduous trees than for Norway spruce. The underestimation of tree heights led to underestimated mean heights at the plot level. As a result of having many segments containing more than one tree, the average crown width obtained at the plot level was overestimated.

The results of statistical significance testing of the differences found between ALS- and field-based estimates have to be interpreted with caution. Lacking the evidence to reject the null hypothesis of non-significant differences between field measurements and ALS estimates does not mean that the differences are not important in a practical setting, and does not eliminate the need for calibrating the ALS estimates.

To obtain reasonable delineation results, the ALS sampling density also has to be suitable for the size of the target objects (e.g. small tree crowns). Nowadays, state-of-the-art ALS sensors can achieve very high sampling densities, but this does not constitute a prerequisite for area-based ALS inventories, where very low echo density (~ 0.5 echo m^{-2}) often suffices (Gobakken and Næsset 2008). Using low echo densities has a cost-saving effect on area-based operational forest inventories, but in turn it may become prohibitive for single tree detection and single tree variable extraction. Thus, finding a suitable echo density to simultaneously satisfy the requirements of area-based forest inventories and of single tree detection requires further investigation.

To conclude, this study has shown that plot-wise stem number estimates obtained from area-based ALS forest inventories can be used to guide the extraction of single tree information. The key factors controlled through the estimated stem number are the spatial resolution of the CHM and the size of the low-pass filter. At least for our data set, the results obtained at the plot level look promising, and further research will be aimed at assessing the applicability of our approach at the stand level in an operational forest inventory framework.

Acknowledgements

We are grateful to Mr Marius Hauglin, Mr Vegard Lien and Mr John Gunnar Dokk for designing and performing the field inventory, and Dr Ole Martin Bollandsås for developing the regression models for tree height prediction. We also thank the two anonymous reviewers who helped us to improve our manuscript.

References

- ANON, 1999, *Pinnacle User's Manual*, pp. 1–123 (San Jose, CA: Javad Positioning Systems).
- ANON, 2010, Matlab® Online Help, Mathworks Inc., USA. Available online at: <http://www.mathworks.com/help/toolbox/images/ref/watershed.html> accessed 21 June 2011.
- BADDELEY, A. and TURNER, R., 2005, Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**, pp. 1–42.
- BATER, C.W. and COOPS, N.C., 2009, Evaluating error associated with lidar-derived DEM interpolation. *Computers and Geosciences*, **35**, pp. 289–300.
- BRANDTBERG, T., WARNER, A.T., LANDENBERGER, R.E. and MCGRAW, J.B., 2003, Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, **85**, pp. 290–303.
- CHEN, Q., BALDOCCHI, D., GONG, P. and KELLY, M., 2006, Delineating individual trees in a savanna woodland using small footprint LiDAR data. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 923–932.
- CLARK, P. and EVANS, E., 1954, Distance to the nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35**, pp. 445–453.
- CULVENOR, D.S., 2000, Development of a tree delineation algorithm for application to high spatial resolution digital imagery of Australian native forest. PhD thesis, Land and Food Resources, Forest and Ecosystem Science, The University of Melbourne. Available online at: <http://repository.unimelb.edu.au/10187/5557> (accessed 21 June 2011).
- EID, T., GOBAKKEN, T. and NÆSSET, E., 2004, Comparing stand inventories based on photo interpretation and laser scanning by means of cost-plus-loss analyses. *Scandinavian Journal of Forest Research*, **19**, pp. 512–523.
- FALKOWSKI, M.J., SMITH, A.M.S., GESSLER, P.E., HUDAK, A.T., VIERLING, L.A. and CROOKSTON, N.L., 2006, Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of LiDAR data. *Canadian Journal of Remote Sensing*, **32**, pp. 153–161.
- FALKOWSKI, M.J., SMITH, A.M.S., GESSLER, P.E., HUDAK, A.T., VIERLING, L.A. and EVANS, J.S., 2008, The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data. *Canadian Journal of Remote Sensing*, **34**, pp. 338–350.
- GOBAKKEN, T. and NÆSSET, E., 2008, Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, **38**, pp. 1095–1109.
- GONZALES, R.C. and WOODS, R.E., 2002, *Digital Image Processing*, 2nd ed., p. 120 (Upper Saddle River, NJ: Prentice Hall).
- GREGOIRE, T.G. and VALENTINE, H.T., 2008, *Sampling Strategies for Natural Resources and the Environment*, pp. 247–262 (Boca Raton, FL: Chapman & Hall/CRC).
- HEINZEL, J.N., WEINACKER, H. and KOCH, B., 2008, Full automatic detection of tree species based on delineated single tree crowns – a data fusion approach for airborne laser scanning data and aerial photographs. In *Proceedings of SILviLASER 2008: 8th International Conference on LiDAR Applications in Forest Assessment and Inventory*, 17–19 September 2008, R.A. Hill, J. Rosette and J. Suárez (Eds.), Edinburgh, UK (Bournemouth: SilviLaser 2008 Organizing Committee), pp. 76–86.

- HENGL, T., 2006, Finding the right pixel size. *Computers and Geosciences*, **32**, pp. 1283–1298.
- HEURICH, M., 2008, Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest national park. *Forest Ecology and Management*, **255**, pp. 2416–2433.
- HYYPÄ, J., HYYPÄ, H., LECKIE, D., GOUGEON, F., YU, X. and MALTAMO, M., 2008, Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, **5**, pp. 1339–1366.
- HYYPÄ, J. and INKINEN, M., 1999, Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, **16**, pp. 27–42.
- HYYPÄ, J., KELLE, O., LEHIKONEN, M. and INKINEN, M., 2001a, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 969–975.
- HYYPÄ, J., SCHARDT, M., HAGGRÉN, H., KOCH, B., LOHR, H.U., SCHERRER, R., PAANANEN, R., LUUKKONEN, H., ZIEGLER, M., HYYPÄ, H., PYYSALO, U., FRIEDLÄNDER, H., UUTTERA, J., WAGNER, S., INKINEN, M., WIMMER, A., KUKKO, A., AHOKAS, E. and KARJALAINEN, M., 2001b, HIGH-SCAN: the first European-wide attempt to derive single-tree information from laserscanner data. *The Photogrammetric Journal of Finland*, **17**, pp. 58–68.
- KAARTINEN, H., HYYPÄ, J., LIANG, X., LITKEY, P., KUKKO, A., YU, X., HYYPÄ, H. and HOLOPAINEN, M., 2008, Accuracy of automatic tree extraction using airborne laser scanner data. In *Proceedings of SILviLASER 2008: 8th International Conference on LiDAR Applications in Forest Assessment and Inventory*, 17–19 September 2008, R.A. Hill, J. Rosette and J. Suárez (Eds.), Edinburgh, UK (Bournemouth: SilviLaser 2008 Organizing Committee), pp. 467–476.
- KITTLER, J. and ILLINGWORTH, J., 1986, Minimum error thresholding. *Pattern Recognition*, **19**, pp. 41–47.
- KOCH, B., HEYDER, U. and WEINACKER, H., 2006, Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 357–363.
- KWAK, D.A., LEE, W.K. and LEE, J.H., 2007, Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research*, **12**, pp. 425–434.
- LECKIE, D., GOUGEON, F., HILL, D., QUINN, R., AMSTRONG, L. and SHREENAN, R., 2003, Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing*, **29**, pp. 633–649.
- LIM, K., TREITZ, P., GROOT, A. and ST-ONGE, B., 2001, Estimation of individual tree heights using LIDAR remote sensing. In *Proceedings of the 23rd Annual Canadian Symposium on Remote Sensing*, 20–24 August 2001, Québec (Ottawa, ON: CASI), pp. 243–250.
- MAGNUSSEN, S., NÆSSET, E. and GOBAKKEN, T., 2010, Reliability of LiDAR derived predictors of forest inventory attributes: a case study with Norway spruce. *Remote Sensing of Environment*, **114**, pp. 700–712.
- MALTAMO, M., MUSTONEN, K., HYYPÄ, J., PITKÄNEN, J. and YU, X., 2004, The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. *Canadian Journal of Forest Research*, **34**, pp. 1791–1801.
- MORS DORF, F., MEIER, E., KÖTZ, B., ITTEN, K.I., DOBBERTINN, M. and ALLGÖWER, B., 2004, LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, **92**, pp. 353–362.
- NÆSSET, E., 2001, Effects of differential single- and dual frequency GPS and GLONASS observations on point accuracy under forest canopies. *Photogrammetric Engineering and Remote Sensing*, **67**, pp. 1021–1026.
- 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**, pp. 88–99.

- NÆSSET, E., 2004a, Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, **19**, pp. 164–179.
- NÆSSET, E., 2004b, Accuracy of forest inventory using airborne laser-scanning: evaluating the first Nordic full-scale operational project. *Scandinavian Journal of Forest Research*, **19**, pp. 554–557.
- NÆSSET, E., 2007, Airborne laser scanning as a method in operational forest inventory: status of accuracy assessment accomplished in Scandinavia. *Scandinavian Journal of Forest Research*, **22**, pp. 433–442.
- OTSU, O., 1979, A threshold selection method from gray level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, **9**, pp. 62–66.
- PERSSON, Å., HOLMGREN, J. and SÖDERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, **68**, pp. 925–932.
- PITKÄNEN, J., MALTAMO, M., HYYPPÄ, J. and YU, X., 2004, Adaptive methods for individual tree detection on airborne laser scanner based canopy height model. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-8/W2, 3–6 October 2004, Freiburg, Germany.
- POPESCU, S.C. and WYNE, H.R., 2004, Seeing the trees in the forest: using lidar and multi-spectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 589–604.
- POULIOT, D.A. and KING, D.J., 2005, Approaches for optimal automated individual tree crown detection in regenerating coniferous forests. *Canadian Journal of Remote Sensing*, **31**, pp. 255–267.
- POULIOT, D.A., KING, D.J. and PITT, D.G., 2005, Development and evaluation of an automated tree detection-delineation algorithm for monitoring regenerating coniferous forests. *Canadian Journal of Remote Sensing*, **35**, pp. 2332–2345.
- PRETZSCH, H., 2010, *Forest Dynamics, Growth and Yield: From Measurement to Model*, p. 154 (Berlin: Springer).
- R DEVELOPMENT CORE TEAM, 2009, R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Available online at: <http://www.R-project.org> (accessed 20 January 2012).
- RAHMAN, M.Z.A. and GORTE, B.G.H., 2009, Tree crown delineation from high resolution airborne lidar based on densities of high points. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-3/W8, 1–2 September 2009, Paris, France.
- RIPLEY, B.D., 1988, *Statistical Inference for Spatial Processes*, 148 pp. (Cambridge: Cambridge University Press).
- SALAS, C., LEMAY, V., NÚÑEZ, P., PATRICIO, P. and ESPINOSA, A., 2006, Spatial patterns in an old-growth *Nothofagus obliqua* forest in south-central Chile. *Forest Ecology and Management*, **231**, pp. 38–46.
- SOILLE, P., 2003, *Morphological Image Analysis*, p. 392 (Berlin: Springer).
- SOLBERG, S., NÆSSET, E. and BOLLANDSÅS, O.M., 2006, Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing*, **12**, pp. 1369–1378.
- TESFAMICHAEL, S.G., AHMED, F., VAN AARDT, J.A.N. and BLAKEWAY, F., 2009, A semivariogram approach for estimating stems per hectare in Eucalyptus grandis plantations using discrete-return lidar height data. *Forest Ecology and Management*, **258**, pp. 1188–1199.
- WANG, Y., WEINACKER, H. and KOCH, B., 2008, A Lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest. *Sensors*, **8**, pp. 3938–3951.
- WULDER, M., NIEMANN, K.O. and GOODENOUGH, D., 2000, Error reduction methods for local maximum filtering of high spatial resolution imagery for locating trees. *Canadian Journal of Remote Sensing*, **28**, pp. 621–628.

- WULDER, M., NIEMANN, K.O. and GOODENOUGH, D., 2002, Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, **73**, pp. 103–114.
- YU, X., HYYPPÄ, J., KAARTINEN, H. and MALTAMO, M., 2003, Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sensing of Environment*, **90**, pp. 451–462.
- ZHAO, K. and POPESCU, S., 2007, Hierarchical watershed segmentation of canopy height model for multi-scale forest inventory. In *Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'*, 12–14 September 2007, P. Rönholm, H. Hyypä and J. Hyypä (Eds.), Espoo, Finland (Espoo: ISPRS), pp. 436–441.

ISBN 978-82-575-1056-5
ISSN 1503-1667



NORWEGIAN UNIVERSITY OF LIFE SCIENCES
NO-1432 Ås, NORWAY
PHONE +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no