Norwegian University of Life Sciences
Faculty of Environmental Science and Technology
Department of Ecology
and Natural Resource Management

Philosophiae Doctor (PhD)
Thesis 2015:12

Estimation of biomass in tropical rainforest using airborne laser scanning

Estimering av biomasse i tropisk regnskog ved bruk av flybåren laserskanning

Endre Hofstad Hansen

Estimation of biomass in tropical rainforest using airborne laser scanning

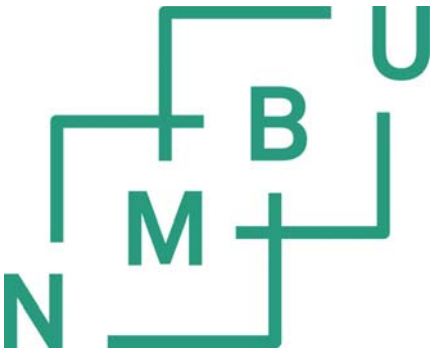
Estimering av biomasse i tropisk regnskog ved bruk av flybåren laserskanning

Philosophiae Doctor (PhD) Thesis

Endre Hofstad Hansen

Department of Ecology and Nature Management
Faculty of Environmental Sciences and Technology
Norwegian University of Life Sciences

Ås 2015



Thesis number 2015:12
ISSN 1894-6402
ISBN 978-82-575-1269-9

PhD supervisors

Professor Terje Gobakken
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO – 1432 Ås, Norway

Professor Erik Næsset
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO – 1432 Ås, Norway

Dr. Ole Martin Bollandsås
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO – 1432 Ås, Norway

Evaluation committee

Dr. Pete Watt
Indufor Asia Pacific Ltd.
P.O. Box 105 039, Auckland City 1143, New Zealand

Professor Timo Tokola
School of Forest Sciences
University of Eastern Finland
P.O. Box 111, FI – 80101 Joensuu, Finland

Associate Professor Katrine Eldegard
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO – 1432 Ås, Norway

Preface

The present thesis, together with a program of formal courses, a trial lecture and a public defence, completes my doctoral work. The work began in June 2011 with planning of the field inventory which was carried out in Amani nature reserve in Tanga region, Tanzania. After a short training of the field crew, field inventory commenced in August 2011. I stayed with the field crew until October. From then on the inventory was led by Mr. Nuru Hussein, for which I am truly grateful.

Analysis and writing of the scientific papers and synopsis was carried out at the Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences (NMBU), during the period from October 2011 to February 2015. Courses in research ethics and philosophy of science, statistics, and remote sensing were completed at NMBU, Swedish University of Agricultural Sciences and University of Eastern Finland.

I am grateful to my supervisors, Professors Terje Gobakken and Erik Næsset, and researcher Ole Martin Bollandsås for all critique, guidance, and support. I am also indebted to all members of the “Skogrover” research group for fruitful discussions and socializing. Thanks to all my co-authors outside the research group: Eliakimu Zahabu, Sokoine Agricultural University, Svein Solberg, Norwegian Forest and Landscape Institute and Annika Kangas, NMBU & Natural Resources Institute Finland. Further, I wish to thank Hugh Riley for revising the language in parts of the thesis. A special thanks to Ernest Mauya for the close collaboration and to Roar Økseter for sharing the office with me for the entire work-period. The work was made possible by the funding by the Royal Norwegian Embassy in Tanzania as part of the Norwegian International Climate and Forest Initiative. I appreciate the financial support for such an important field. Further, I would like to thank my family, Guri & Einar, Yngve & Jessica, Lars & Ann-Mari, Torgunn, Ole, and Tormod. Lastly, a loving thanks to Ingvild and Tora for brightening up my days.

Endre Hofstad Hansen
Ås, April 2015

Contents

Abstract

Forest inventories based on field sample surveys, supported by auxiliary remotely sensed data, have the potential to provide transparent and confident estimates of forest carbon stocks required in climate change mitigation schemes such as the REDD+ mechanism. Three-dimensional (3D) information about the density and height of the vegetation, obtained from remotely sensed data, is particularly useful for providing accurate estimates of forest biomass. Most of the research on biomass estimation supported by 3D remotely sensed data has been carried out in boreal and sub-boreal coniferous forests with relatively low biomass quantities and open forest structure. The studies comprising the present thesis were conducted in a dense tropical forest with challenging topography.

In the present thesis two different techniques of collecting remotely sensed 3D data were used: airborne laser scanner (ALS) and spaceborne interferometric synthetic aperture radio detection and ranging (InSAR). While the main focus was on the use of ALS, the high quality digital terrain model (DTM) derived from the ALS data also facilitated the comparison of InSAR data as auxiliary information in biomass estimation.

The analyses and results presented in Paper I of modelling aboveground biomass using ALS data resulted in root mean square errors (RMSE) of about 33% of a mean value of 462 $Mg \cdot ha^{-1}$. Use of texture variables derived from a canopy surface model constructed from ALS data did not result in improved models. Analyses showed that (1) variables derived from ALS-echoes in the lower parts of the canopy and (2) canopy density variables explained more of the aboveground biomass density than variables representing the height of the canopy.

Paper II investigated the potential of using cheaper, low-pulse density ALS data. Effects of reduced pulse density on (1) the digital terrain model (DTM), and (2) explanatory variables derived from ALS data were assessed. Random variation in DTMs and ALS variables increased with reduced pulse density. A reliability ratio, quantifying replication effects in the ALS-variables, indicated that most of the common ALS variables assessed were reliable at pulse densities >0.5 $pulses \cdot m^{-2}$, and at a plot size of 0.07 ha. The plot size of 0.07 ha corresponds to the plot size used in the national forest inventory of Tanzania.

The field plot size is of importance for the precision of carbon stock estimates, and better information of the relationship between plot size and precision can be useful in designing future inventories. The effect of plot size on the precision of biomass estimates assisted by remotely sensed data was therefore assessed in Paper III. Precision estimates of forest biomass estimates developed from 30 concentric field plots with sizes of 700, 900,…, 1900 m^2, were assessed in

a model-based inference framework. Findings indicated that larger field plots were relatively more efficient for inventories supported by ALS and InSAR data compared to a pure field-based survey. Further, a simulation showed that a pure field-based survey would have to comprise 3.5–6.0 times as many observations for the plot sizes of 700–1900 m^2 to achieve the same precision as an inventory supported by ALS data.

Sammendrag

Fjernmålte data brukt sammen med feltobservasjoner, kan potensielt gi grunnlag for troverdige estimat av karbonet som er lagret i skogens biomasse. Informasjon om lagret biomasse er nødvendig for arbeidet med å motvirke klimaendringer. REDD+ er system hvor denne kunnskapen er avgjørende. Fjernmålt 3D-informasjon om skogens høyde og tetthet er særlig nyttig fordi den gir nøyaktige estimat på skogens biomasse. Tidligere forskning på bruk av fjernmålt 3D-data til biomasseestimering har hovedsakelig blitt gjort i boreal og sub-boreal barskog med relativt lav biomasse og åpen skogstruktur. Studiene i denne avhandlingen ble utført i tett tropisk skog med utfordrende terrengforhold.

To ulike metoder for fjernmåling av 3D-data ble benyttet: flybåren laserskanning (ALS) og satellittbåren interferometrisk syntetisk apertur-radar (InSAR). Hovedfokuset for avhandlingen var på bruk av ALS. I tillegg ga en digital terrengmodell av høy kvalitet, produsert med ALS-dataene, muligheter for en sammenligning med bruk av InSAR-data til biomasseestimering. Analysene beskrevet i Paper I viste at modellering av biomasse over bakkenivå ved hjelp av ALS-data ga en standardfeil (RMSE) på ca. 33 % av et gjennomsnitt på 462 $Mg \cdot ha^{-1}$. Bruk av teksturvariabler utledet fra en modell av vegetasjonens overflate konstruert fra ALS-data ga ikke forbedret resultat. Analysene viste videre at (1) variabler utledet fra laser-ekko i lavere deler av vegetasjonen og (2) tetthetsvariabler fra vegetasjonen forklarte biomassetettheten bedre enn variabler som beskrev vegetasjonens høyde.

I Paper II ble muligheten for bruk av billigere ALS-data med lav pulstetthet undersøkt. Effekten av lav pulstetthet på (1) den digitale terrengmodellen, og (2) variabler utledet fra ALS-data ble analysert. Tilfeldig variasjon i digitale terrengmodeller og ALS-variabler økte med redusert pulstetthet. En ratio for pålitelighet, som kvantifiserer replikasjonseffekter i ALS-variablene, viste at mesteparten av de undersøkte variablene var pålitelige ved pulstettheter >0.5 $pulser \cdot m^{-2}$, ved bruk av en feltmålte flater på 0,07 ha. Denne størrelsen på feltflatene tilsvarer den som brukes i Tanzanias nasjonale landsskogtaksering.

Størrelsen på feltflatene er viktig for presisjonen i biomasseestimater. Bedre informasjon om forholdet mellom størrelse på feltflatene og presisjon er nyttig i planleggingen av framtidige skogtakseringer. Presisjonsestimater av skogens estimerte biomasse ble derfor beregnet for 30 konsentriske feltflater med størrelse på 700, 900,…, 1900 m^2. Disse estimatene ble analysert i en modell-basert statistisk metode. Resultatene indikerte at større feltflater relativt sett var mer effektive for taksering understøttet av ALS- og InSAR-data, sammenlignet med en ren feltflatetakst. Videre ble det i en simulering vist at en ren feltflatetakst ville måtte

inneholde 3,5–6,0 ganger så mange observasjoner for henholdsvis flatestørrelser fra 700–1900 m^2 for å oppnå samme presisjon som en takst understøttet av ALS-data.

List of appended papers

I. Hansen, E.H., Gobakken, T., Bollandsås, O.M., Zahabu, E. and Næsset, E. 2015. Modeling aboveground biomass in dense tropical submontane rainforest using airborne laser scanner data. *Remote Sensing*, 7(1): 788–807.

II. Hansen, E.H., Gobakken, T. and Næsset, E. 2015. Effects of pulse density on digital terrain models and canopy metrics using airborne laser scanner in a tropical rainforest. S*ubmitted*.

III. Hansen, E.H., Gobakken, T., Solberg, S., Kangas, A., Ene, L., Mauya, E. and Næsset, E. (2015). Impact of field plot size on the relative efficiency of biomass estimation in a Tanzanian rainforest using airborne laser scanner and interferometric synthetic aperture radar as auxiliary data. *Submitted*.

SYNOPSIS

1 Introduction

The climate is changing! The report from the fifth assessment of the Intergovernmental Panel on Climate Change (IPCC) is conclusive. We are experiencing increased global temperatures because of our emissions of greenhouse gases (IPCC, 2014). With the rise in temperatures we observe changes in habitats for all forms of life on earth that have not been seen in millennia. The extent of Arctic sea-ice is decreasing, glaciers are retracting, and the great ice sheets of Greenland and Antarctica are losing mass. Rainfall patterns are changing across the globe, leaving some areas wetter and some drier than normal. Extreme weather events are more frequent and intense. The salinity of the oceans is changing, some becoming saltier and some less salty. Furthermore, the oceans are becoming more acidic. All of these changes have severe implications for life forms adapted to specific climatic and environmental conditions.

1.1 Global reduction of greenhouse gases

Reduction of greenhouse gases has been a focus area in international environmental work since the late 1980s with the establishment of the IPCC in 1988 and the United Nations Framework Convention on Climate Change (UNFCCC) treaty in 1992. Since then IPCC have compiled scientific evidence about climate change, and specific treaties (called “protocols”), which set limits on greenhouse gas emissions, have been agreed upon by signatories of the UNFCCC treaty.

The main greenhouse gas in terms of emissions and global temperature change is carbon dioxide (CO_2), and CO_2 has contributed to more than 80% of the total temperature increase due to greenhouse gases in the last 15 years (Myhre et al., 2013). Further, a near-linear relationship has been found between total emissions of CO_2 and global temperature change (Matthews et al., 2009). Yearly emissions of CO_2 have increased rapidly since the beginning of the industrial revolution and are estimated to a total of 555 ± 85 petagrammes of carbon (PgC) in the period of 1750–2011 (Ciais et al., 2013). For 2013 Friedlingstein et al. (2014) reported a global estimated total of 10.75 ± 0.71 PgC, with a projected increase of 2.3% in 2014.

1.2 The role of forests in the global carbon cycle

Vegetation, especially trees, removes carbon from the atmosphere and stores it as biomass. If the forest is cleared, or becomes degraded, the carbon is released back to the atmosphere. Human- induced deforestation and forest degradation is often referred to as land use change and the total estimated emissions from land use change in the period 1750–2011 is estimated to 180 ± 80 PgC (Ciais et al., 2013). Although the proportion of emissions stemming from land use change has decreasing trend (Ciais et al., 2013), land use change is still a significant source of carbon emissions with an estimated total of 0.87 ± 0.49 PgC for 2013 (Friedlingstein et al., 2014).

Tropical forests, found on land between latitudes of 23.44°N and 23.44°S, cover around 18 million km^2 (FAO, 2011) and are estimated to store 271 ± 16 PgC (Grace et al., 2014). These forests are under great pressure for conversion to agricultural land (Houghton, 2012) and Grace et al. (2014) report a total carbon loss of 2.01 ± 1.1 PgC yr^{-1} from deforestation, harvesting and peat fires. However, the growth in forests and woodlands is reported to sequester 1.85 ± 0.09 PgC yr^{-1} resulting in a net loss of 0.16 ± 1.1 PgC yr^{-1} from tropical forests. Thus, these forests represent a substantial potential carbon sink, approaching 2 PgC yr^{-1} or up to 20% of the global carbon emissions.

With the prospect of a quick and cheap solution for mitigating carbon emissions, tropical forests have received a lot of attention and have resulted in the policy and economic incentive mechanism known as the REDD+ mechanism. REDD+ (reducing emissions from deforestation and forest degradation, conservation and enhancement of forest carbon stocks and sustainable management of forests in developing countries), described in the 16th session of the Conference of Parties to the United Nations Framework Convention on Climate Change, gives developing countries the opportunity to monetize from the reduction of emissions from deforestation and forest degradation, and enhancement of forest carbon stocks (UNFCCC, 2011).

Accessing carbon finances through REDD+ will require, among other factors, measurement of carbon stock changes in forests (UNFCCC, 2010). Furthermore, a mechanism for commercial trading of forest carbon credits earned through enhancement of forest carbon stocks, conservation of forests or sustainable forest management require trustworthy systems for verification of carbon offsets. In addition, application of the principle of conservatism, which takes into account the uncertainty of estimates to minimize the risk of overestimating emission reductions (UNFCCC, 2006; Grassi et al., 2008), and lack of accurate biomass estimates may result in loss of carbon credits for the project developer (Gibbs et al., 2007).

Establishing a robust and transparent system for measuring, reporting and verification (MRV) of biomass is therefore a requirement for the successful implementation of a REDD+ regime (Plugge et al., 2011).

1.3 Forest inventories supported by remotely sensed data

Forest inventories have the potential to provide transparent and confident estimates of aboveground biomass, hereafter simply referred to as biomass. Forest inventories are usually designed as sample surveys, with observations on the ground collected from field plots, supported by one or several sources of remotely sensed data. Remotely sensed data, in the form of aerial images, has been an important forest inventory tool since the 1940s (FAO, 1948), and the availability of optical satellite images in the 1970s has resulted in global forest cover statistics (Boyd & Danson, 2005). While high cost has prevented the use of aerial images, the use of low-cost optical satellite images has been hampered by low spatial resolution and persistent cloud cover in tropical areas. Furthermore, both aerial and satellite optical images have traditionally only provided two-dimensional information, although recent developments have resulted in three-dimensional (3D) data from aerial and satellite images with the use of digital photogrammetry and image matching (e.g. Næsset, 2002; Bohlin et al., 2012; Persson et al., 2013; Gobakken et al., 2014).

The use of LiDAR (light detection and ranging) sensors, most commonly mounted on a small aircraft and with a scanning capability, known as airborne laser scanning (ALS), has proved to be both effective and accurate for determining biomass in different forest types (Zolkos et al., 2013; Fassnacht et al., 2014). There has been a strong focus on research of ALS during the past two decades, and ALS is now used as an integral part of operational forest management inventories in several countries (McRoberts et al., 2010; Næsset, 2014). Most of the published studies on ALS to estimate biomass have been carried out in boreal and sub-boreal coniferous forests with relatively low biomass and open forest structure. However, in the last five years, use of ALS for biomass estimation has been demonstrated in tropical forests in South America (Asner et al., 2010; Clark et al., 2011; Vincent et al., 2012; Andersen et al., 2013; Asner et al., 2014), Asia (Hou et al., 2011; Jubanski et al., 2012; Ioki et al., 2014) and Africa (Asner et al., 2012; Laurin et al., 2014). The maximum biomass densities in these studies were about 500 metric tonnes of biomass per hectare ($Mg \cdot ha^{-1}$), while biomass densities in tropical rainforests can reach levels beyond 500 $Mg \cdot ha^{-1}$ (Keith et al., 2009).

1.4 Sources of three-dimensional remotely sensed data

Satellite-mounted optical sensors have been used for estimation of global forest cover since the launch of the first Landsat satellite in 1972. More importantly, with data spanning over three decades, these optical sensors provide estimates of global forest cover change (Hansen et al., 2013). For biomass estimation however, the usefulness of the two-dimensional information from satellite images is limited because it lacks information about vegetation height, has limited resolution, and is often obstructed by cloud cover.

ALS systems solve these challenges by emitting a short pulse of laser light and measuring the time between the emission and the reflectance (echoes) detected by the LiDAR sensor. By emitting thousands of pulses per second and recording several echoes per pulse in a scanning motion, the ALS system effectively creates a 3D cloud of echoes. By recording the position and orientation of the sensor at the time of emitting each pulse, using a GPS (global positioning system) receiver and an inertial navigation system unit, each echo is positioned in the 3D space (x, y, and z positions). To derive information about the vegetation, a digital terrain model (DTM) is constructed by classifying echoes as ground echoes. Following the construction of the DTM, the elevation of all echoes in relation to the DTM is computed. Echoes above a certain threshold above the DTM are regarded as vegetation echoes.

Although the best results for biomass estimation have been obtained using ALS, its cost is high compared to using satellite-based sensors. While satellite based optical imagery is frequently obstructed by persistent cloud cover in the tropics, use of active synthetic aperture radio detection and ranging (SAR) sensors penetrate clouds and produce backscatter images that can be used for the prediction of forest biomass. In high biomass conditions however, radar backscatter data has so far not been able to provide data for reliable estimation and has been shown to saturate at biomass levels of between 200–250 $Mg \cdot ha^{-1}$ (Mitchard et al., 2009; Le Toan et al., 2011). Promising results have nevertheless been published for biomass values up to 450 $Mg \cdot ha^{-1}$ (Minh et al., 2014). At present, SAR technologies exist that can produce 3D data using four different techniques: clinometry, stereoscopy, interferometry and polarimetry (Toutin & Gray, 2000). In addition, optical satellite images can produce 3D data by repeat-pass-acquisition and image matching techniques. New applications are being developed continuously and a thorough overview is beyond the scope of this thesis. A shared property of these techniques is that, in order to provide information at a level similar to that of ALS, they require a high quality DTM. At present, the only technology able to provide this DTM quality is ALS, and it is therefore a prerequisite for the other sensors and techniques.

1.5 The area-based method for forest inventories using remotely sensed data

The most common method for utilizing remotely sensed auxiliary information for forest inventory purposes is known as the area-based method. This method, first outlined in Næsset (1997a; 1997b), is based on modelling the relationship between attributes of interest that have been measured or calculated from measurements on field plots, with explanatory variables derived from remotely sensed data from the corresponding field plot area. To apply the model on the area of interest, the remotely sensed data are tessellated into units, usually of the same size as the size of the field plots, and the explanatory variables are derived for each unit. The model is then applied to predict the response variable on each unit.

The alternative method to the area-based one is known as the individual-tree-crown method. As the name suggests, it is based on modelling the attributes of interest on a single tree basis. Identification of individual trees is affected by stand density and spatial pattern, which causes problems related to interlaced tree crowns and trees below the dominant canopy (Vauhkonen et al., 2014). In the tropical rainforest where tree crowns overlap, forming a closed canopy-cover, the separation of individual tree crowns is regarded as a difficult task with presently available technologies.

1.6 Improving the accuracy of biomass estimates

As described in section 1.2 accurate biomass estimates are a requirement for the REDD+ mechanism to function. Increased accuracy would also potentially lead to added carbon credits for the project developer (Gibbs et al., 2007). Accuracy is defined as the sum of trueness and precision (ISO, 2012). Accuracy of an estimation is often expressed by the mean square error (Gregoire & Valentine, 2008, p. 28), or the root mean square error (RMSE), (Equation 4) of the mean estimate as used in this thesis. Thus, accuracy incorporates both trueness, expressed herein by the mean difference (MD), and precision, expressed as standard error of estimation (SE), i.e., the square root of the estimation variance, or standard deviation of a sample (SD). Trueness can only be calculated when the true value is actually known. The simplest way of increasing the precision of biomass estimates is by increasing the sample size. In a design-based framework (see section 3.3.3), the variance of the estimation under simple random sampling is proportional to the square root of the sample size minus the number of explanatory variables minus one (Stoltzenberg, 2009, p. 181). Thus, all else being equal, doubling the number of observations would halve the variance of the estimation. Another option is to use remotely sensed auxiliary information related to the observed biomass. Depending on the correlation

between the biomass and explanatory variables derived from the remotely sensed data, the precision is improved using the auxiliary information from the data.

In boreal forests, where the correlation between ALS-derived variables and forest parameters is high, use of ALS data has been found to reduce the need for field observations by a factor of 3–9 (Næsset et al., 2011; Ene et al., 2013), without reducing the estimated precision. In high-cost countries, such as Norway, use of ALS in forest inventories thus becomes cost-efficient. In Tanzania, where the cost of field labour is low, using additional field plots would probably be the most cost-efficient way of increasing the precision of the estimated biomass. However, the remote and inaccessible nature of forest areas in tropical developing countries means that remotely sensed data can nevertheless be invaluable in providing precise biomass estimates (McRoberts et al., 2013b; McRoberts et al., 2014b).

1.7 Effects of field plot size on the accuracy of biomass estimates

The size of the field plot is a property of great importance for accuracy when estimating biomass by means of remotely sensed data. Studies of modelling the relationship between forest biomass and ALS-derived variables in tropical areas have utilized field plots sizes in the range of 0.1–1.0 ha. Larger plots inevitably increase the accuracy of the biomass estimates due to spatial averaging (Zolkos et al., 2013), as larger field plot sizes reduce the between-plot variance (cf. Gobakken & Næsset, 2009; Mascaro et al., 2011; Magnussen et al., 2012). In addition, larger plots have smaller ratios of the border zone to total plot area than do smaller plots, a zone which is subject to boundary effects (Mascaro et al., 2011; McRoberts et al., 2014a). This implies that the relative influences of the boundary effects are smaller for larger plots, regardless of plot shape. Negative consequences of GPS positioning errors are also smaller for large plots (Gobakken & Næsset, 2009). Likewise, the boundary effects will be more pronounced in forests with large tree crowns and on rectangular or quadratic plots, compared to circular plots with the smallest possible circumference-to-area ratio. Even though larger field plots, e.g. plots larger than 0.25 ha, result in models with better performance, their practical application is limited due to the difficulty of establishing them. This is especially challenging in rugged and steep terrain, and in areas with very dense vegetation. Reducing the field plot size to a more practical and manageable size will, however, reduce the precision of biomass estimates.

1.8 Research objectives

The overall objective of this thesis was to investigate the potential of using ALS as an auxiliary data-source in sample surveys of biomass in a tropical forest with a wide range in biomass densities in rugged and steep terrain. During initial work on the first study it became clear that negative boundary effects were strongly influencing the results. This led to a focus on examining the effects of plot size on the precision of biomass estimates in subsequent studies. Because the ALS data provided a high quality DTM, a comparison of ALS to interferometric SAR, in terms of sampling error, was performed. Specific objectives for the studies were:

I. To model biomass using conventional height and density variables derived from ALS data, and to explore the use of texture variables derived form an ALS canopy surface model.
II. To assess the effects of reduced pulse density on the derived DTM, and on the ALS-derived explanatory variables at spatial units ranging from 0.07 to 0.28 ha in size.
III. To assess, in a model-based inference framework, the impact of plot size on the relative efficiency of ALS and interferometric SAR data compared to models with terrain elevation as the only explanatory variable.

2 Materials

2.1 Study area

The study area, Amani nature reserve (ANR , Figures 1–3, S 5°08', E 38°37', 200–1200 m above sea level), covers around 85 km^2 of tropical sub-montane rainforest and is located in the East Usambara mountains in eastern Tanzania, part of the Eastern Arc mountains. The Eastern Arc mountains region is a global biodiversity hotspot area (Myers et al., 2000) and its unique forest ecosystem, stretching from Udzungwa in Tanzania in the south to Taita Hills in Kenya in the north, contains many endemic species of both animals and plants. Within this mountain system, the East Usambara mountains have been identified as one of three top priority areas for forest conservation (Burgess et al., 2007). Rain falls throughout the year, with two wet seasons, April–May and October–November, and the forest in ANR receives around 2000 mm rainfall per year. Daily mean temperatures vary from about 16 to 25 °C. ANR was gazetted in 1997, comprising of six former forest reserves, Amani-East, Amani-West, Amani-Sigi, Kwamsambia, Kwamkoro and Mnyusi Scarp. In addition, forest land from the neighbouring tea estate, sisal estate and local village was included in the ANR. The area also includes the Amani botanical gardens, established in 1902 under German colonial rule, which has contained over 500 indigenous and non-native tree species (Dawson et al., 2008). Very few of the non-native species have spread successfully from the area in which they were planted, but one species in particular, *Maesopsis eminii*, is found throughout the entire ANR and is the most common species in the reserve. *M. eminii* originated from the lake region in eastern Congo and is a typical light-demanding, pioneer species. It thrives in disturbed areas, but is not able to germinate under thick canopy (Newmark, 2002) and is thus not found in the less disturbed areas of the ANR. In an inventory carried out in 1986/87, about half of the ANR was classified as either logged or covered with *M. eminii* as a result of logging (Hamilton & Bensted-Smith, 1989). Logging was stopped in the late 1980s and most of the ANR is now covered by closed forest.

2.2 Field data

Two different sets of field data were used in the studies. The first field data set (FD1), used in Papers I and II, was originally established by a non-governmental conservation and development organization, Frontier Tanzania, during 1999–2000 (Frontier Tanzania, 2001). Rectangular shaped plots of 50 × 20 m were established on a 450 × 900 m grid covering the ANR (Figure 2). The horizontal area of the plots varies from 0.0639–0.1239 ha because the

plots were laid out along the terrain slope, without any slope correction. All trees with ≥10 cm diameter at breast height (DBH) were callipered, marked and species identified. During two campaigns in 2008 and 2009/2010, 143 of these plots were revisited and all trees re-measured (Mpanda et al., 2011; Mgumia, 2014). Trees that had grown larger than 10 cm in DBH since the first survey were included, and dead or missing trees excluded. All of the initial 173 plot locations were visited again between August 2011 and April 2012 and plots that were not re-measured during the 2008–2010 period were re-measured at this time. All plots were identified in the field by local personnel who had performed their establishment and previous re-measurement. Plots which were not positively identified in the field were re-established and all the trees with a DBH ≥10 cm registered. This was also done for plots with an apparent change in structure (due to landslide or human activity), or trees were added or removed in cases where there was a clear error in the earlier records. Of the 173 plots, 15 plots had one or more corners with missing coordinates after completion of the field work, due to reception of too few satellites during data recording from positioning satellites. One plot was also discarded because one of the processed corner coordinates had a reported error of >10 m (See 2.2.3 for further details about the positioning of the field plots). Furthermore, four plots were found to be outside the study area. The DBH data from the remaining 153 plots contained measurements from four different years; 2008 (19 plots), 2009/2010 (91 plots) and 2011/2012 (43 plots).

In FD1, ten trees per plot were selected for height measurement. The trees were systematically selected by choosing the closest tree to each corner; one tree in the middle of each short end of the plot; and two trees along the sides, 15 m from each corner. Tree height (H) was measured using a Vertex IV hypsometer (Figure 5) and trees with damage were noted. For plots with low stocking, in which the same tree could be selected more than once, less than ten heights were measured. A total of 1497 trees were measured during the fieldwork in 2011 and 2012.

The second field data set (FD2, Figure 3), used in Paper III, consisted of data from 30 circular field plots collected during November 2011 in pre-determined locations with the aim of capturing as much variation in biomass as possible by distributing them in different altitudinal zones. All trees with ≥5 cm diameter at breast height (DBH) were callipered, marked and species identified. The horizontal distance from the plot centre to the front of each tree was measured using a Vertex IV hypsometer. Since the distance was measured to the front of the trees, half of the tree DBH was added during data processing to get the total horizontal distance of the trees from the plot centre. The plot size was determined by the reach of the Vertex, and under the most challenging conditions in ANR, distance measurement started to fail at 25 m.

Thus, in order to contain 30 observations, the maximum plot size was 0.19 ha. The heights of three trees per plot (largest, medium and smallest) were measured using the Vertex hypsometer.

The representativeness of the plots in FD2 was evaluated in Mauya et al. (2015) by comparing the properties of FD2 to FD1. Based on this evaluation Mauya et al. (2015) concluded that, although being sampled in an opportunistic manner, the distribution in different altitudinal zones resulted in a sample which closely resembled properties of the systematic sample.

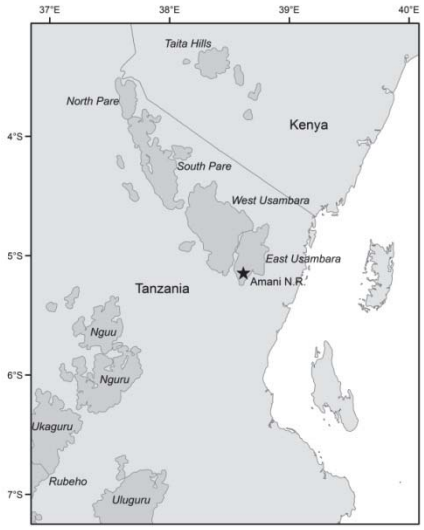


Figure 1. Study area (star) situated in the Eastern Arc Mountains (dark grey areas).

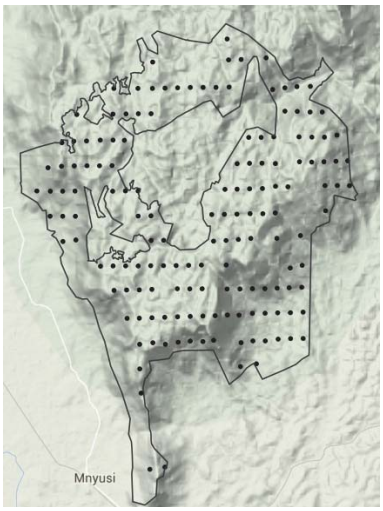


Figure 2. Plot locations for FD1 in Amani nature reserve.

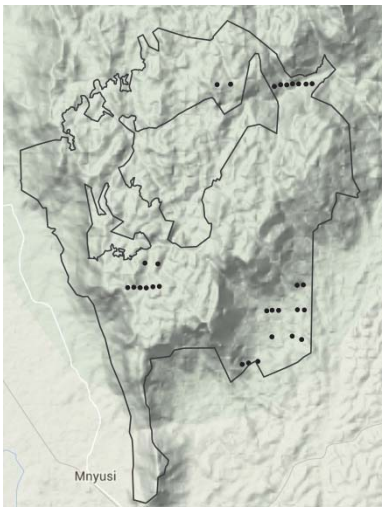


Figure 3. Plot locations for FD2 in Amani nature reserve.

2.2.1 Height-diameter models

Single tree predictions of biomass with both DBH and H as independent variables in the allometric models, give more reliable and lower biomass levels than those without height information (Henry et al., 2010; Marshall et al., 2012). Non-linear height-diameter (H-D) models were developed for both FD1 and FD2, with plot as a random effect. Using the trees measured for height, H-D models were fitted using the "fithd" function in the package "lmfor" (Mehtatalo, 2012) in R software (R Development Core Team, 2013) again with plot as random effect. The "lmfor" package contains 20 two- and three parameter model forms, and the most suited forms for our data were selected based on the Akaike information criterion. The selected model forms (Equations 1 and 2) described by Prodan (1968) and Winsor (1932) respectively,

were then re-fitted using the "nlme" function (Pinheiro et al., 2014) in R for FD1 and FD2. The selected models can be expressed as the mean (expected value) functions:

$$\mathrm{E[H]} = 1.3 + \mathrm{DBH}^2/(\mathrm{a} + \mathrm{b} * \mathrm{DBH} + \mathrm{c} * \mathrm{DBH}^2), \quad (1)$$

$$\mathrm{E[H]} = 1.3 + \mathrm{a} * \exp\big(-\mathrm{b} * \exp(-\mathrm{c} * \mathrm{DBH})\big). \quad (2)$$

This method of calibrating the H-D model is described by Lappi and Bailey (1988) and is able to include local effects. The H-D development of trees can for instance be affected by local soil conditions or by surrounding trees. To capture the local effects, field plot was specified as random effect and all three parameters of the model were allowed to describe the random effects.

2.2.2 Aboveground biomass

Aboveground biomass for individual trees ($\widehat{\mathrm{AGB}_t}$) was predicted using a locally developed allometric model (Equation 3) (Masota et al., 2015). The model is developed from 60 trees from 34 different species in the ANR and has a pseudo coefficient of determination of 0.84. The trees were felled and the green weights of stem, branches, twigs and leaves were recorded in the field, along with DBH. Wood samples from each of the three components were collected and the green-to-dry weight ratio calculated after oven drying of the wood samples. The tree biomass was then calculated by first multiplying the green weight with the green-to-dry weight ratio of each of the tree components and then summing these up for each tree. The applied model was:

$$\widehat{\mathrm{AGB}_t} = 0.402 * \mathrm{DBH}_t^{1.4365} * \mathrm{H}_t^{0.8613}, \quad (3)$$

where $\widehat{\mathrm{AGB}_t}$ is the predicted aboveground biomass in Mg for individual tree number *t*, DBH_t is the tree diameter at breast height in cm for tree number *t*, and H_t is the tree height in m for tree number *t*. For DF1 the $\widehat{\mathrm{AGB}_t}$ was then summed at field plot level and converted to per-hectare units of biomass (Table 1). For FD2 each tree was allocated to each of the concentric plot size based on the distance from the plot centre to the centre of the stem, computing per-hectare biomass values for all plots of 700, 900, …, 1900 m^2 (Table 2). Although this biomass is referred to as "observed biomass", the computed values are subject to errors related to the applied allometric model, and the subsampling and measurement of tree DBH and height.

Table 1. Characteristics of the 153 field plots in FD1.

Characteristic	Range	Mean	SD
Area (ha)	0.0639–0.1239	0.0914	0.011

N[a] (ha^{-1})	85.4–1085.7	471.5	161.5
DBH[b] (cm)	10.0–270.0	27.5	22.9
BA[c] ($m^2 \cdot ha^{-1}$)	5.4–144.9	47.3	22.2
Biomass ($Mg \cdot ha^{-1}$)	43.2–1147.1	461.9	214.7
H[d] (m)	8.3–51.3	19.2	8.9

[a] number of trees, [b] diameter at breast height (1.3 m), [c] basal area, [d] predicted tree height.

Table 2. Mean biomass and standard deviation (SD) in FD2 at plot sizes of 700, 900,…, 1900 m^2.

Plot size (m^2)	Mean biomass ($Mg \cdot ha^{-1}$)	SD ($Mg \cdot ha^{-1}$)
700	371.8	221.5
900	366.1	216.3
1100	365.6	203.0
1300	361.0	190.5
1500	354.2	180.4
1700	355.0	170.2
1900	351.1	159.6

2.2.3 Positioning of the field plots

During the fieldwork in August 2011–April 2012, the plot corners of the rectangular plots in FD1 and the centre point of the circular plots in FD2 were georeferenced by means of differential global navigation satellite system (dGNSS). A 40-channel dual frequency survey grade receiver (Topcon Legacy-E+) was used as a field unit (Figure 4) and a second receiver, functioning as a base station, was placed on the roof of a house at the ANR headquarters with a distance of <14 km from the plots. Before the georeferencing started, the coordinates of the base station antenna were determined with precise point positioning with GPS and global navigation satellite system data collected continuously for 24 hours according to Kouba (2009). The field unit was placed at each point on a 2.9 m rod for a minimum of 30 minutes, and a one second logging rate was used. Horizontal errors of the georeferenced points were estimated to an average of 0.57 m based on random errors reported from the post-processing using Pinnacle software (Anon., 1999) and empirical experience of the relationship between reported error and the true error documented by Næsset (2001). Mean precision in the vertical direction after post-processing in Pinnacle (Anon., 1999) was reported to 0.39 m.

Figure 4. Field unit of the differential global navigation satellite system (dGNSS) used for positioning of field plots.

Figure 5. Measuring tree height using a Vertex IV hypsometer.

2.2.4 Remotely sensed data

The ALS data used in all three papers was collected as complete coverage using a Leica ALS70 sensor mounted on a Cessna 404 twin engine, fixed wing aircraft. The acquisition was carried out in the period 19–25 January 2012 with additional flights in the period 2–18 February 2012 to fill minor gaps in the data. Average flight speed was 70 m s^{-1} at a mean altitude of 800 m above ground level and with a laser pulse repetition frequency of 339 kHz. From each pulse the sensor registered up to five echoes. A maximum scan angle of ±16° from nadir yielded an average swath width of 460 m. The beam divergence was 0.28 mrad, which produced an average footprint size on the ground of about 22 cm.

In Paper III, interferometric synthetic aperture radio detection and ranging (InSAR) was assessed as an alternative source of remotely sensed data. The InSAR data were acquired by the Tandem-X satellite mission on 6th August 2011. The incidence angle was 46°, and the polarization was horizontal transmit and horizontal receive. The normal baseline was 210 m, which corresponded to a 2π height of ambiguity of 38 m.

3 Methods

The basis for the analyses in all three papers was the conventional method for biomass modelling and estimation known as the "area-based method" described section 1.5. The method is based on modelling the relationship between attributes of interest that have been measured or calculated from measurements on field plots, and explanatory variables derived from the remotely sensed data. It is vital that the remotely sensed data is extracted from the horizontal area matching the field plot. Discrepancy in this matching is often referred to as co-registration errors, and has been identified as an important source of error in the "area-based method" (Gobakken & Næsset, 2009). To apply the model on the area of interest, the remotely sensed data are tessellated into units, usually of the same size as the size of the field plots, and the explanatory variables are derived for each unit. The model is then applied to predict the response variable on each population unit.

3.1 Construction of digital terrain models

A prerequisite for the retrieval of useful remotely sensed 3D information for biomass estimation is a high quality DTM. In Papers I and III the DTM was derived from the ALS data by the supplier of the ALS data, Terratec AS, Norway. ALS echoes reflected from the ground were identified and classified using the progressive triangulated irregular network (TIN) densification algorithm (Axelsson, 2000) in the TerraScan software (Anon., 2012). The DTM was derived from the ALS as a TIN from the planimetric coordinates and corresponding heights of the ALS echoes classified as ground echoes. In Paper II a similar classification and construction of a DTM was performed using the "GroundFilter" program in the FUSION toolkit (McGaughey, 2013). The study involved repeated reduction of the ALS pulse density, and DTMs were constructed from pulse densities of 8, 4, 2, 1, 0.5, and 0.25 pulses·m^{-2}. The algorithm presented by Kraus and Pfeifer (1998) and implemented in the "GroundFilter" program initially makes an average surface based on all ALS echoes. Further, weights are given to all echoes based on their vertical distance to the initial surface. Low weight is given to echoes above the surface, and high weight to echoes below. The weights are then used in re-fitting the surface. Two parameters in the algorithm can be adjusted to determine which echoes are weighted. Echoes located below the surface with a distance larger than parameter *g* are assigned the maximum weight value of 1.0, while echoes located above the surface with a distance larger than the parameter *w* plus the parameter g are assigned weights of 0.0 (McGaughey, 2013). To adjust for the different pulse densities the two parameters were controlled while leaving the

other parameters at the program default setting. The *g* and *w* parameter settings at different pulse densities are given in Table 4. Visual inspection of initial classifications of ground echoes showed large outliers and a smoothing filter of 3 m was applied to remove these outliers. From the echoes classified as ground, a 1 m gridded surface was created using the "TINSurfaceCreate" program in FUSION (McGaughey, 2013).

3.2 Explanatory variables derived from remotely sensed data

After creation of a DTM the elevation of the DTM was subtracted from the elevation of all ALS echoes resulting in an elevation above the ground for each echo. From the five echoes per pulse registered by the ALS sensor, each echo was classified into one of three categories: "single", "first of many" or "last of many". The "single" and "first of many" were merged into one dataset, denoted as "first echoes" while the "single" and "last of many" were merged into another dataset and denoted as "last echoes". The two classes of ALS echoes formed the basis for derivation of conventional explanatory variables from the echoes. These variables comprise two main types of variables, canopy height variables and canopy density variables, and were computed separately from the "first echoes" and "last echoes". Both variable types describe the vertical distribution of ALS echoes. Canopy height variables including maximum and mean values (E.max, E.mean), standard deviation (E.sd), coefficients of variation (E.cv), kurtosis (E.kurt), skewness (E.skewness) and percentiles at 10% intervals (E.10, E.20,…, E.90) were derived from the laser echoes above a threshold of 2 m (Paper III) or 4 m (Papers I and II) above ground. Canopy density variables were derived by dividing the height between a 95% percentile height and the threshold into 10 equally spaced vertical layers and calculating the proportion of echoes above each layer to the total number of echoes of each echo category ("first echoes", "last echoes"), including echoes below the threshold (D.0, D.1,…, D.9). To denote whether the variables were derived from the first or last echo category, a subscript L or F was used as notation, e.g. E.mean.F. This computation of ALS variables follows the procedure presented in Næsset and Gobakken (2008) and is frequently implemented in practical forest inventories (Næsset, 2014). In Paper I, additional variables describing horizontal distribution of the echoes from an ALS-derived canopy surface model were computed. Firstly, a rasterized canopy surface model with 1 m resolution was computed from the top-of-canopy ALS echoes. The raster was then converted into grey level images from the field plots, and variables originally presented by Haralick et al. (1973) were calculated using the "glcm" package (Zvoleff, 2014) in R. The texture variables were calculated using a 3 × 3 m window size and averaged in all directions (0, 45, 90 and 135°). Shifts of 3, 6, 9, 12, and 15 m were tested and variables included mean (MN),

homogeneity (HG), variance (VAR), contrast (CONT), dissimilarity (DS), entropy (ENT), angular second moment (SM) and correlation (COR) for each of the shifts. The subscript 3, 6, 9, 12, or 15 was used as notation for the shifts. These variables, originating from image analysis, were proposed as additional information to the conventional height and density variables.

As part of the analysis in Paper II ALS variables from four concentric circular plot sizes of 0.07, 0.14, 0.21, and 0.28 ha were derived. These concentric plots were centred in the centre of the 153 rectangular field plots. A set of ALS variables was derived for each concentric plot using the "CloudMetrics" program in FUSION (McGaughey, 2013). Frequently used variables were selected for analysis. These included E.mean, E.max, E.10, E.90, D.0, D.5, and the variance of the elevation above ground (E.var). The first and last echo categories were not used in the variables in Paper II.

In Paper III, the results from ALS-assisted estimation were compared to use of remotely sensed data from satellite radar imagery. Only one explanatory variable, the mean height, was derived from the InSAR data. SAR image pairs were processed, using a kind of stereo imaging known as interferometry, resulting in a digital surface model (DSM). Further, removal of phase noise and offset and ramp errors was performed using a Goldstein filter (Goldstein & Werner, 1998) and ground control points, respectively. Phase unwrapping was carried out using the minimum cost flow method, and the DSM was geocoded to a ground resolution of 10 × 10 m. Following the construction of the DSM, the DTM derived from the ALS TIN was subtracted from the DSM, resulting in obtained InSAR heights, i.e. heights of the centre of the radar echo above ground. Mean InSAR height was then derived for each field plot by weighting the height of each 10 × 10 m units of the normalized InSAR DSM by the area of the unit intersected by the field plot.

3.3 Statistical analyses

3.3.1 Modelling the relationship between biomass and remotely sensed variables

Linear least-square multiple regression analysis was used for developing biomass models in Papers I and III. To account for heteroscedasticity and non-linear relationships between the response and explanatory variables, a transformation of the response is often used. Natural-log transformations of both response and explanatory variables were performed in Papers I and III. In addition, square-root transformation of the response was tested in Paper I. These transformations introduce a bias when back-transformed to arithmetic scale and procedures for bias-adjustment, described by Goldberger (1968), Gregoire et al. (2008), and Snowdon (1991), were therefore applied. In Paper I the selection of explanatory variables to be included in the

models was performed using a best-subset regression procedure implemented in the "leaps" package (Lumley & Miller, 2009) in R. To avoid multicollinearity, variance inflation factors were controlled and models were selected based on the Bayesian information criterion (BIC). The aptness for prediction of the selected models was validated by 10-fold cross-validation. The accuracy of the models was assessed by the RMSE:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2/n}, \quad (4)$$

where n is number of plots, y_i is the observed value for plot i, $\hat{y}_i$ is the predicted value for plot i. Precision was assessed by the MD:

$$\mathrm{MD} = \sum_{i=1}^{n}(\hat{y}_i - y_i)/n. \quad (5)$$

The relative RMSE (RMSE%) and relative MD (MD%) were calculated as percentages by dividing the absolute RMSE and MD values, respectively, by the observed means.

In a study utilising FD2, Mauya et al. (2015) found that E.60.F and D.1.L were the most frequently selected variables in modelling biomass using plot sizes from 700 to 1900 m^2. E.60.F and D.1.L were therefor selected a priori for the biomass models in paper III.

To explore the relative importance of the explanatory variables, an analysis was performed in Paper I by fitting a separate simple linear model for a random sample of 1/3 of the plots. The single explanatory variable which resulted in the lowest BIC value was included in the model. Random sampling of observations, performed without replacement and model-fitting, was repeated 1000 times. The frequency with which each variable appeared in the model was used as a measure of importance for each individual variable.

3.3.2 Effects of pulse density on DTM and canopy variables

The cost of acquiring ALS data is largely governed by the flight time. By flying higher and/or faster the costs can be reduced, resulting in cheaper but lower pulse density data. The effects of reduced pulse density on the ALS-derived DTM and canopy variables were investigated in Paper II. A random thinning procedure was incorporated in a Monte Carlo simulation, in which the thinning and the subsequent analysis were repeated 50 times per pulse density level, to quantify the effects of the reduced pulse density. To study the effects of reduced pulse densities on the DTMs, the elevation z in the DTMs from reduced pulse densities (z^{DTM}) were subtracted from the elevation of each point (k) measured with dGNSS (z^{dGNSS}) to obtain the difference for each point (D_k, Equation 6):

$$D_k = z_k^{DTM} - z_k^{dGNSS}. \quad (6)$$

The mean difference ($\overline{D}$) and standard deviation (S_D) of the differences (D_k) were calculated. To compare the $\overline{D}$ at each pulse density level a t-test was performed using the Holm-Bonferroni procedure (Holm, 1979) for correction of p-values for multiple comparisons.

Conventional measures of accuracy, $\overline{D}$ and S_D, assumes no outliers and a normal distribution of errors. As pointed out by e.g. Zandbergen (2008), errors in DTMs are not normally distributed. Q-Q plots were therefore produced and checked for non-normality. In addition, robust measures of accuracy suited for characterisation of non-normal distributions suggested by Höhle and Höhle (2009) were produced. The 50% sample quantile of the errors ($P50_D$, i.e., the median value) is a robust estimator for a systematic shift of the DTM (Höhle & Höhle, 2009). The 95% quantile of the absolute value of the errors ($P95_{|D|}$) and the normalized median absolute deviation (NMAD, Equation (7)), a robust estimator for S_D, are estimators resilient to outliers (Höhle & Höhle, 2009).

$$\mathrm{NMAD} = 1.4826 * \mathrm{median}_i(|D_k - P50_D|). \quad (7)$$

The mean value ($\overline{M}$) and standard deviation of each canopy variable (S_M) on plot level were calculated from the Monte Carlo simulations across the 50 repetitions. Even though the canopy variables derived from ALS have been shown to be relatively unaffected by the density of echoes (Lim et al., 2008), reduced pulse density increase the S_M. As explained by Magnussen et al. (2010), random factors affecting the canopy variables suggests that the variables should be considered as random variables instead of fixed, as is commonly the case. These random factors can be referred to as replication effects. Replication effects weaken the fit of the biomass models with a factor termed as the reliability ratio (Fuller, 1987, p. 3). By calculating the replication variance in the variables, estimates of the reliability ratios for the variables were calculated. The method was used by Magnussen et al. (2010), in which the reliability ratio was calculated as the ratio of the variance of each variable among sample plots, to the total variance of the corresponding variable (Equation (8)):

$$\text{Reliability ratio} = (\hat{\sigma}_u^2)/(\hat{\sigma}_u^2 + \hat{\sigma}_w^2), \quad (8)$$

where $\hat{\sigma}_u^2$ is the estimated among-plot variance of the variable and $\hat{\sigma}_w^2$ is the estimated average within-plot variance. High within-plot variance in a variable compared to the variance among plots for the same variable results in a low reliability ratio, indicating that the variable is less reliable as an explanatory variable.

To assess the effect of plot size on the reliability ratio of the variables derived from concentric circles of 0.07, 0.14, 0.21, and 0.28 ha were computed.

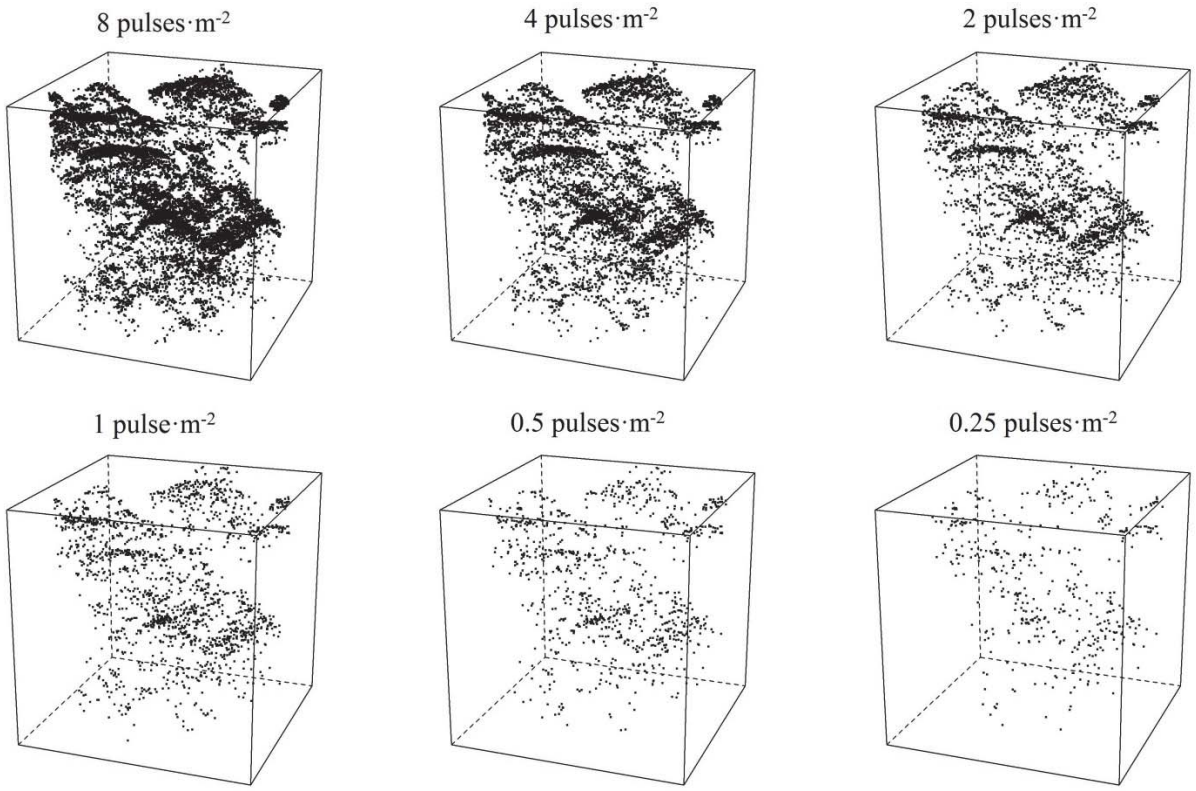


Figure 6. Illustration of thinned ALS echo cloud for plot number 49 with a plot size of 1400 m^2.

3.3.3 Variance estimation

As described in section 1.5, application of an estimated biomass model using the area-based method yields a biomass prediction for each population unit. The biomass predictions for the population elements are subsequently used to derive an estimate for the population, either as a mean or total biomass estimate. Accompanying the estimate, a variance estimate is calculated to state the precision of the estimate. Two main approaches to variance estimation have been used in forest inventories: design-based and model-based variance estimation. In the design-based approach the population, from which samples are taken, is regarded as fixed. The only source of variance is the random selection of elements included in the sample. Thus, the estimated variance is derived from the inventory sample and the probability of each population element to be included in the sample, referred to as the inclusion probability. This inclusion probability is assumed to be positive and known for all population elements. Such samples are often referred to as probability samples.

It is often the case, however, that the sample has been acquired in a non-probabilistic manner (Clark & Kellner, 2012), resulting in zero- or unknown inclusion probabilities. The

zero- or unknown inclusion probability can be the result of opportunistic sampling, i.e. sampling close to roads for economic and/or practical reasons. Similarly, purposive sampling, established to investigate a specific subject, often result in samples acquired in a non-probabilistic manner. Furthermore, the inclusion probability can be affected by the accessibility of the area (Köhl et al., 2006, p. 76). In the case where the sample data does not meet the requirements for a design-based approach to variance estimation, a model-based approach may be a viable alternative. Model-based inference does not, as opposed to design-based inference, rely on a probabilistic sample that represents the population. Instead the statistical inference relies on the model itself as a valid model of the distribution of possible observations for each population element. The population is not viewed as fixed, but rather as a result of a random process, referred to as a "superpopulation" model. This "superpopulation" model cannot be observed, but the parameters of the model can be estimated from the survey sample. The surveyed population is viewed as only one random realisation of this "superpopulation".

Because the field plots in FD2 were distributed in order to cover the height gradient in the ANR, the inclusion probabilities of the plots were unknown, and a model-based approach to estimation and inference was used in Paper III.

The approach followed the notation in Ståhl et al. (2011), where an element of the "superpopulation" was expressed as:

$$y_i = g(x_i, \alpha, \varepsilon_i), \tag{9}$$

where y is a vector of the observed plot biomass on plot i, x is a vector of variables derived from the auxiliary data, α is a vector of model parameters and ε is a vector of errors, and g is a function describing the "superpopulation". It is assumed that the errors are independent, normally distributed, with a constant variance, and without spatial auto-correlation. The parameters α were estimated with $\hat{\alpha}$ using least square regression, and used to estimate the population mean by:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} g(x_i, \hat{\alpha}), \tag{10}$$

where i indexes the population elements and N is the number of elements, i.e., i = 1, 2, …, N. Assuming that the estimated $\hat{\alpha}$ is accurate, the g function was linearized in the neighbourhood of the true function using first order Taylor series expansion:

$$\begin{aligned} g(x, \hat{\alpha}) \approx g(x, \alpha) + (\hat{\alpha}_1 - \alpha_1) * g_1'(x, \alpha) + (\hat{\alpha}_2 - \alpha_2) * g_2'(x, \alpha) \\ +\ldots+ \left(\hat{\alpha}_p - \alpha_p\right) * g_p'(x, \alpha), \end{aligned} \tag{11}$$

where $g'_j(x,\alpha) = \partial g(x,\alpha)/\partial\alpha_j$, j indexes the parameters and p is the number of parameters, i.e.,

j = 1, 2, …, k, …, p. The variance of the population mean was then estimated by:

$$\widehat{\mathrm{var}}(\hat{\mu}) = \sum_{j=1}^{p}\sum_{k=1}^{p}\widehat{\mathrm{Cov}}\left(\hat{\alpha}_j,\hat{\alpha}_k\right)\hat{\bar{g}}'_j\hat{\bar{g}}'_k, \quad (12)$$

where $\hat{\bar{g}}'_j$ and $\hat{\bar{g}}'_k$ are the estimated mean values of the first order derivatives of the g function for parameters j and k, respectively (cf. Ståhl et al., 2011).

Unlike design-based estimators, which often are unbiased or nearly unbiased, the unbiasedness of model-based estimators depends on the model being correctly specified. It was therefore paramount to assess how well the model fitted the field plot observations. Assessment of the fit of the models followed the approach used by McRoberts et al. (2013a). Scatterplots of observed *vs.* predicted biomass were produced for each plot size. Correctly specified models should result in points falling closely along a 1:1 line with intercept 0 and slope 1. Further, pairs of observations and predictions were ordered with respect to the predicted values and grouped into three classes of 10 pairs. The mean of the observed *vs.* predicted biomass was plotted for each group. A correctly specified model should again result in points falling along a 1:1 line.

3.3.4 Relative efficiency

To assess the gain in precision of using remotely sensed data to enhance the estimates, relative efficiency was calculated for both ALS ($RE_{TE:ALS}$) and InSAR ($RE_{TE:InSAR}$). Simple log-log models with the terrain elevation from the DTM as explanatory variable were developed for each plot size of 700, 900, …, 1900 m^2. These models were denoted as TE models. The relative efficiencies were calculated as ratios of the estimated variance for the mean biomass estimate ($\hat{\mu}$) for each plot size using the TE models divided by the variance estimates for each plot size using the ALS models:

$$RE_{TE:ALS_s} = \widehat{\mathrm{var}}(\hat{\mu}_{TE})_s/\widehat{\mathrm{var}}(\hat{\mu}_{ALS})_s, \quad (13)$$

where s is an indicator of the plot sizes 700, 900, …, 1900 m^2. Similarly, relative efficiency for InSAR was computed as:

$$RE_{TE:InSAR_s} = \widehat{\mathrm{var}}(\hat{\mu}_{TE})_s/\widehat{\mathrm{var}}(\hat{\mu}_{InSAR})_s. \quad (14)$$

Efficiency of ALS was also calculated relative to InSAR ($RE_{InSAR:ALS}$) in the same way by dividing the variance estimates for each plot size using the InSAR models by the variance estimates for each plot size using the ALS models:

$$\mathrm{RE}_{\mathrm{InSAR:ALS_s}} = \widehat{\mathrm{var}}(\hat{\mu}_{\mathrm{InSAR}})_s / \widehat{\mathrm{var}}(\hat{\mu}_{\mathrm{ALS}})_s. \quad (15)$$

In a design-based framework, applying simple random sampling (SRS), the relative efficiency can be used directly to calculate the additional number of field observations needed to compensate for the contribution of the remotely sensed data, which is a fundamental quantity in cost comparisons. This is because the SE of the mean estimate under SRS is proportional to the square root of the sample size minus the number of explanatory variables minus one (Stoltzenberg, 2009, p. 181). In practice, a relative efficiency of two would mean that the gain of the remotely sensed data could be compensated by twice as many field plots, assuming that the sample variance remain constant.

In the model-based framework used in Paper III the SE of the mean estimate is also assumed to reduce with increased number of observations. However, the number of observations needed to reach the same SE for the different models cannot be deduced by analytical means. Instead a basic Pólya-urn resampling scheme was applied as described in Köhl et al. (2006, pp. 195–196) to simulate the variance of the TE models. The Pólya-urn resampling scheme generates a design-consistent posterior predictive distribution of the property in interest, given that the sample is reasonably large and representative of the population (Ghosh & Meeden, 1997, p. 44–46). The field sample of u = 30 observations were considered as representative of the population, and the Pólya-urn resampling generated posterior predictive distributions of biomass for U = 60, 120, and 180 observations based on the sample. From a virtual urn, containing the 30 observations, one observation was randomly drawn, duplicated, and returned to the urn together with the duplicate. The urn thus contained u + 1 = 31 observations. The selection scheme was repeated until the desired number U of observations in the urn was reached. The simulations were repeated 200 times and the mean variance of observed biomass reported.

4 Results and discussion

Results presented in the papers constituting this thesis cover different aspects of using remotely sensed data in support of forest surveys in a dense tropical rainforest in Tanzania. With regards to the main objective in this thesis, the assessment of ALS as an auxiliary source of data, all three papers contributed to an increased knowledge in the use of ALS in dense tropical forests.

As pointed out in section 2.2.2, the observed biomass in the present thesis is subject to uncertainty not accounted for related to the allometric models and field measurements of DBH and tree height. Thus, errors related to the biomass observations are not accounted for. Overlooking these errors lead to overoptimistic precision of the variance estimates. In a study conducted in a tropical forest in Ghana, and with a plot size of 1600 m^2, Chen et al. (2015) found that the impact of allometric error contributed about 11% to the total relative prediction error. With similar forest conditions and plot sizes, it is reasonable to assume errors of similar influence in the present thesis.

4.1.1 Modelling aboveground biomass using ALS data

Paper I documented the relationship between ALS-derived variables and biomass calculated from measurements on field plots. In terms of accuracy and precision (Table 3), the results were similar to those of other recent studies in similar forests and with similar field plot sizes (Clark et al., 2011; Ioki et al., 2014; Laurin et al., 2014).

The study that was reported in Paper I also identified some modelling challenges when using the field observations in FD1. Firstly, the size of the field plots, together with their rectangular shape, possibly resulted in large negative boundary effects. These effects are due to a discrepancy between trees included in the field inventory and parts of their crown being outside the vertical boundary of the plot, and *vice versa*. Secondly, in a natural forest that has reached a climax state, old and aging trees will have reduced or even negative height and crown development. Because of the asymptotic relationship between height and diameter, canopy height variables are less suitable for discriminating between tall trees with various diameters. It is generally the large trees in a tropical forest that show this asymptotic H-D development (Poorter et al., 2006; Iida et al., 2011). Since the largest trees have great influence on the biomass in the observed biomass, this could explain the underestimation on field plots with a high biomass value, as observed in Paper I. Similar observations were made by Skowronski et al. (2007) in a temperate forest with a weak relationship between tree height and tree diameter. An analysis of the relative importance of ALS variables was therefore performed in Paper I. This analysis showed that most of the information for explaining biomass was found in

variables describing the vertical density of the full vegetation layer, and in variables from the last return echoes (Figure 8). Variables describing the height of the vegetation, which have been found to be important in boreal studies (Næsset & Gobakken, 2008), were found to be less useful in explaining the forest biomass. Similar results with regards to variable importance as those in Paper I, have also been presented by Laurin et al. (2014) for a tropical forest in Sierra Leone, West Africa.

Table 3. Summary of regression models for biomass using ALS variables.

Model	Response variable	Predictive model[a]	Model fit		10-fold cross-validation[b]			
			R^2	BIC	RMSE	RMSE%	MD	MD%
A	ln biomass	3.815 + 1.755·D.2.L+ 1.498·D.9.L + 0.016·E.90.F	0.70	98.8	149.18	32.3	−2.40	0.5
B	ln biomass	3.984 + 3.222·MN.3	0.52	160.6	173.84	37.6	−8.57	1.9
C	ln biomass	3.665 + 1.530·D.2.L + 1.231·D.9.L + 0.013·E.90.F + 0.737·MN.15	0.71	98.7	158.02	34.4	−2.85	0.6
D	sqrt(biomass)	3.796 + 11.294·D.2.L + 13.321·D.9.L+ 0.249·E.mean.L	0.62	814.4	154.44	33.4	6.12	1.3
E	sqrt(biomass)	7.563 + 0.054·MN.3 − 0.072·CONT.3	0.48	857.0	169.77	36.8	8.17	1.8
F	sqrt(biomass)	3.796 + 11.294·D.2.L + 13.321·D.9.L + 0.249·E.mean.L	0.62	814.4	156.59	33.9	5.57	1.2

[a] Explanatory variables presented in section 3.2; [b] Values after back-transformation to arithmetic scale.

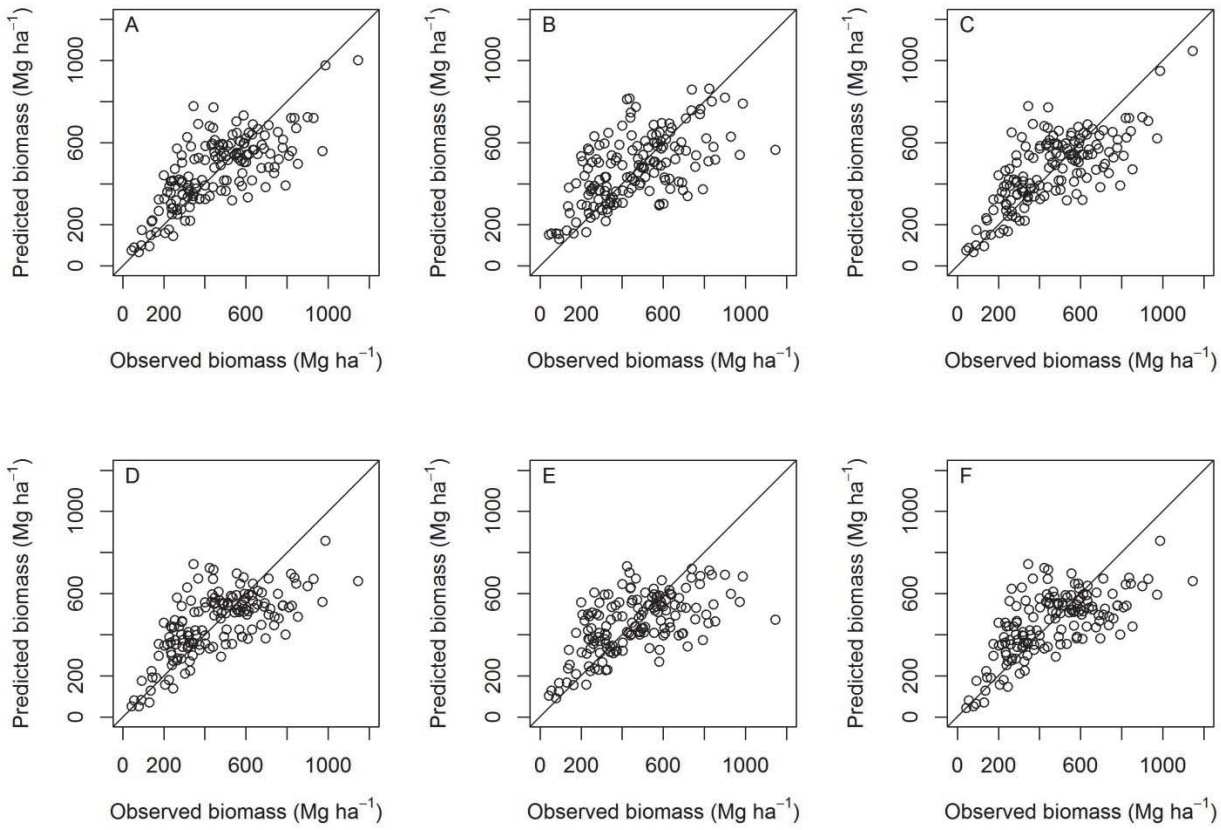


Figure 7. Scatter plots of observed *vs.* predicted biomass using logarithmic (A–C) or square root transformations of the response (D–F) in combination with vertical variables (A, D), texture variables (B, E) and both vertical and texture variables (C, F).

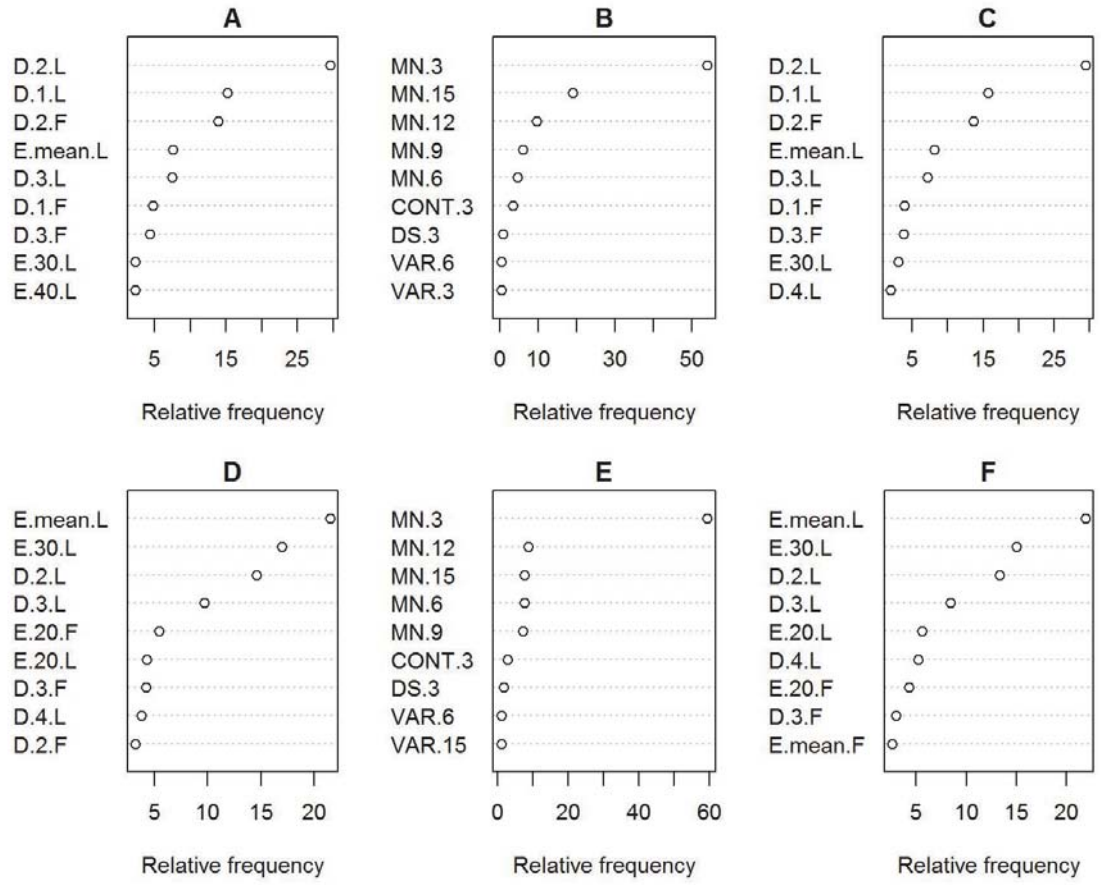


Figure 8. Dot charts of the relative frequency of the nine most frequently selected explanatory variables in a simple linear model fitting procedure with logarithmic (A–C) and square root transformation of the response (D–F) in combination with vertical variables (A, D), texture variables (B, E) and both vertical and texture variables (C, F). Variables are explained in Section 3.2.

4.1.2 Effects of pulse density on DTM quality and ALS variables

Effects of reduced pulse density on the DTMs and canopy variables derived from ALS data were assessed in Paper II. Reduced pulse density, assessed by Monte Carlo simulation, resulted in a $\overline{D}$ between the elevation of 612 point measurements recorded by the dGNSS and the elevation of the same points in the ALS-derived DTM of 1.81 m for a pulse density of 8 pulses·m^{-2} (Table 4). Thus, the elevation recorded by the dGNSS was higher than the ALS-derived DTM. Reduction of pulse density from 8 to 0.25 pulses·m^{-2} gave no significant effect on the $\overline{D}$.

Q-Q plot of the distribution of errors (Figure 9) showed non-normality and justified the presentation of robust measures of accuracy. A loss of precision was observed when the pulse density was reduced from 8 to 0.25 pulses·m^{-2}. This loss of precision was expressed both by the conventional measure of precision, S_D, and the more robust measure NMAD (Table 4).

Table 4. Summary of the difference in elevation between the dGNSS measurements and the DTM elevation for different pulse densities.

						Parameter settings[1]	
Pulses·m^{-2}	$\overline{D}$ (m)	S_D (m)	$P50_D$ (m)	$P95_{\|D\|}$ (m)	NMAD (m)	*g* (m)	*w* (m)
0.25	1.77	3.20	0.90	7.72	2.15	-1.0	1.5
0.5	1.77	3.02	0.92	7.50	1.97	-1.5	2.0
1	1.79	2.93	0.94	7.34	1.85	-2.0	2.5
2	1.81	2.90	0.96	7.28	1.80	-2.5	3.0
4	1.81	2.88	0.95	7.20	1.75	-3.0	3.5
8	1.81	2.89	0.95	7.35	1.81	-3.5	4.0

Mean difference ($\overline{D}$), standard deviation (S_D), 50% quantile of the difference ($P50_D$), 95% quantile of the absolute value of the difference ($P95_{|D|}$) and the normalized median absolute deviation (NMAD).
[1] Settings of the *g* and *w* parameters in the applied ground classification algorithm.

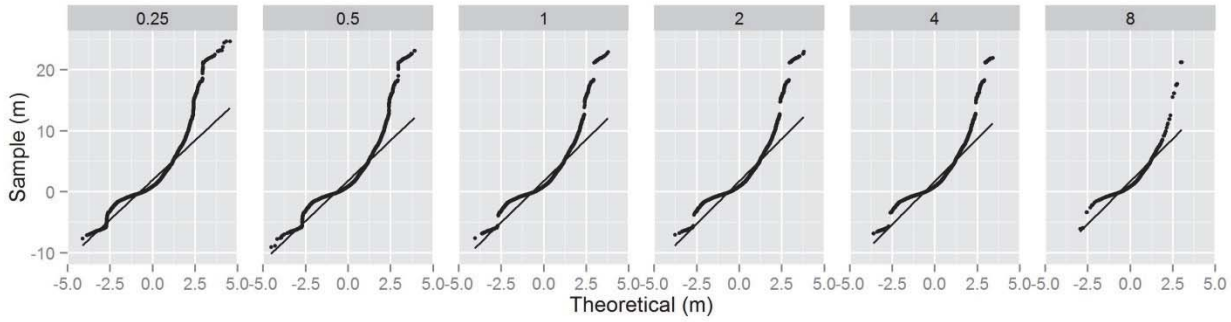


Figure 9. Normal Q-Q-plot for the distribution of the difference in elevation between the elevation recorded by the dGNSS and the elevation of the corresponding ALS-derived DTM at pulse densities of 0.25, 0.5, 1, 2, 4, and 8 pulses·m^{-2}.

Mean values from repeated simulations showed that most of the explanatory variables were unaffected by a reduced pulse density (Table 5). E.max, however, decreased with reduced pulse density, and showed a significant difference of 0.58 m at 1 pulse·m^{-2} compared to the value at 8 pulses·m^{-2} at a plot size of 0.07 ha. This effect was reduced with increased plot size, but was still significant for 1 pulse·m^{-2} at a plot size of 0.28 ha.

Reduced pulse density resulted in increased variance in the canopy variables on plot level (Figure 10). The standard deviations for the explanatory variables (S_M) showed that the variable describing the canopy elevation in the middle of the canopy (E.mean) was more stable than the elevation of the top and bottom of the canopy (E.max, E.10, E.90). Reduction of the

pulse density from 8 to 0.25 pulses·m^{-2} increased the S_M for E.mean from 0.09 to 0.90 m on a 0.07 ha plot size (Figure 10). The S_M of E.max increased from 0.16 to 1.03 m at a plots size of 0.07 ha.

Further, reduced pulse density resulted in decreased reliability ratio, the ratio of the estimated among-plot variance to the estimated total variance (Figure 11). At pulse densities ≥2 pulses·m^{-2}, and a plot size of 0.07 ha, the reliability ratio was >0.95 for all explanatory variables. At a pulse density of 0.5 pulses·m^{-2}, and a plot size of 0.07 ha, the reliability ratios of E.var and D.0 were reduced down to 0.60 and 0.90, respectively. At a pulse density of 0.25 pulses·m^{-2}, and a plot size of 0.07 ha, E.var, E.10, and D.0 had an estimated reliability ratio of <0.9, while the rest of the variables had a reliability ratio of >0.9.

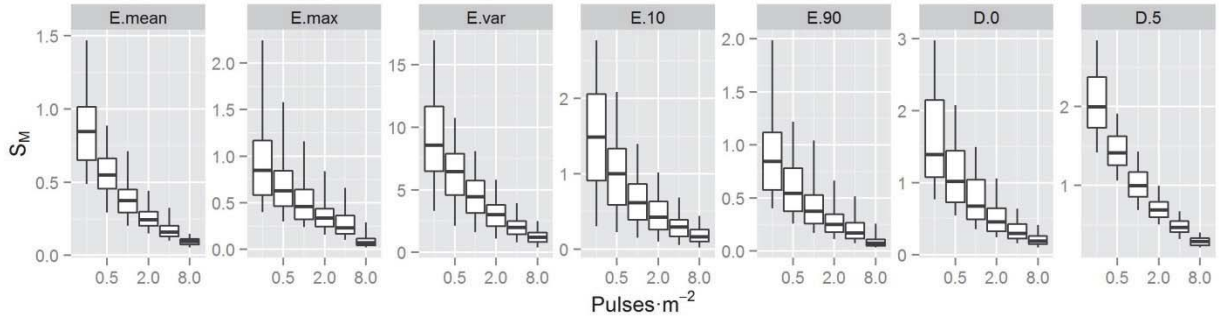


Figure 10. Box and whisker plots of standard deviations (S_M) (whiskers at 5th and 95th percentile) of ALS-derived explanatory variables for plot size of 0.07 ha and pulse densities of 0.25, 0.5, 1, 2, 4, and 8 pulses·m^{-2}. E.mean (mean elevation above ground), E.max (maximum elevation above ground), E.var (variance of the elevation above ground), E.10 and E.90 (10th and 90th height percentile of canopy points), D.0 and D.5 (the proportion of points above the ground and above the mean canopy height).

Table 5. Mean values ($\overline{M}$) of explanatory variables for plot sizes of 0.07, 0.14, 0.21, and 0.28 ha and pulse densities of 0.25, 0.5, 1, 2, 4, and 8 pulses·m^{-2}.

		$\overline{M}$						
Plot size (ha)	Pulses·m^{-2}	E.mean	E.max	E.var	E.10	E.90	D.0	D.5
0.07	0.25	24.59	41.42	107.59	10.06	36.15	89.99	55.18
0.07	0.5	24.49	41.75	108.79	9.79	36.18	90.01	55.25
0.07	1	24.47	41.98	108.93	9.72	36.16	90.09	55.39
0.07	2	24.49	42.22	109.17	9.70	36.19	90.17	55.47
0.07	4	24.59	42.36	108.72	9.82	36.21	90.29	55.68
0.07	8	24.69	42.56	108.17	9.93	36.26	90.33	55.73
0.14	0.25	24.59	43.44	112.27	9.48	36.64	90.07	54.67
0.14	0.5	24.47	43.70	113.24	9.28	36.60	90.07	54.74
0.14	1	24.45	43.92	113.60	9.21	36.60	90.15	54.80
0.14	2	24.49	44.11	113.67	9.24	36.63	90.23	54.89
0.14	4	24.60	44.23	113.15	9.40	36.68	90.38	55.05
0.14	8	24.70	44.39	112.60	9.52	36.74	90.44	55.14
0.21	0.25	24.60	44.38	114.04	9.33	36.93	90.16	54.35
0.21	0.5	24.47	44.61	114.91	9.12	36.86	90.14	54.42
0.21	1	24.43	44.81	115.22	9.04	36.84	90.18	54.50
0.21	2	24.49	45.03	115.35	9.07	36.88	90.28	54.59
0.21	4	24.59	45.17	114.83	9.22	36.93	90.40	54.72
0.21	8	24.70	45.33	114.22	9.35	37.00	90.48	54.82
0.28	0.25	24.59	45.01	115.87	9.13	37.05	90.01	54.31
0.28	0.5	24.45	45.23	116.54	8.93	36.97	89.97	54.36
0.28	1	24.44	45.47	116.95	8.87	36.98	90.04	54.45
0.28	2	24.48	45.68	116.93	8.91	37.00	90.12	54.52
0.28	4	24.59	45.85	116.40	9.05	37.05	90.24	54.66
0.28	8	24.69	45.98	115.84	9.18	37.12	90.33	54.75

Mean elevation above ground (E.mean), maximum elevation above ground (E.max), variance of the elevation above ground (E.var), 10th and 90th percentile of elevation (E.10 and E.90), and the proportion of points above the ground (D.0) and above the mean canopy height (D.5).

The statistics for the canopy variables, resulting from the Monte Carlo simulations, showed that the variances of the variables were reduced with plot size for all variables and at all pulse densities (Figure 11). Increasing the plot size means that the probability of including larger trees increase. As a result, the maximum elevation (E.max) and the elevation of the top

of the canopy (E.90) were found to increase in value with increased plot size (Table 5). Increasing the plot size also means that more of the variability in the elevation of ALS-echoes is captured by the plot, resulting in increased E.var. Variables describing the elevation of the lowest part of the canopy (E.10) and the proportion of points above the mean canopy height (D.5), however, decreased in value with increasing plot size. E.mean and D.0 did not show any trend with increased plot size. The reliability ratio increased for all variables with increasing plot size.

Dense vegetation obstructs the ALS pulses and results in fewer pulses reaching the ground and being available for DTM construction. The effect of vegetation on ALS-derived DTMs has been studied in different conditions and has resulted in both an over-prediction of terrain elevation (Bowen & Waltermire, 2002; Reutebuch et al., 2003; Töyrä et al., 2003) and under-prediction of terrain elevation (Hodgson et al., 2005; Tinkham et al., 2011). Hodgson et al. (2005) found that ALS-derived elevation was significantly under-predicted in all studied land cover classes. The under-prediction was largest in pine forest areas, by up to 0.24 m. Tinkham et al. (2011) also found an under-prediction of 0.9–0.16 m in coniferous areas, when comparing two different ground classification algorithms. In their discussion of observed under-predicted terrain elevation in heavily vegetated areas, Hodgson et al. (2005) suggested that the error was a result of echo density, and/or the accuracy of correct classification of echoes as ground.

The analysis in Paper II showed that the mean DTM elevation was unaffected by the reduction in pulse density from 8 to 0.25 pulses·m^{-2}. This was in contradiction with results from other studies on reducing ALS data density (Hyyppä et al., 2005; Anderson et al., 2006; Leitold et al., 2015). In a tropical forest with similar conditions as in the present study, Leitold et al. (2015) found an increased DTM elevation of 3.02 m at 1 pulse·m^{-2}, compared to a DTM from 20 pulses·m^{-2}. Leitold et al. (2015) attributed the increased elevation to the morphological filter algorithm (Zhang et al., 2003) used to classify ground points. Hyyppä et al. (2005), who used data collected in three separate flights, attributed an increased DTM elevation to the beam size and sensitivity of the laser.

From the analysis of the DTMs in Paper II and the results in several different studies related to construction of DTMs in different conditions, it seems clear that systematic effects of pulse thinning could be the result of the algorithm used to classify ground echoes or the parameter settings in the algorithm. Thus, great care is needed when using ALS data from different sensors and flying conditions, in diverse vegetation, and with different sensor-settings in terms of correctly classifying the ground echoes.

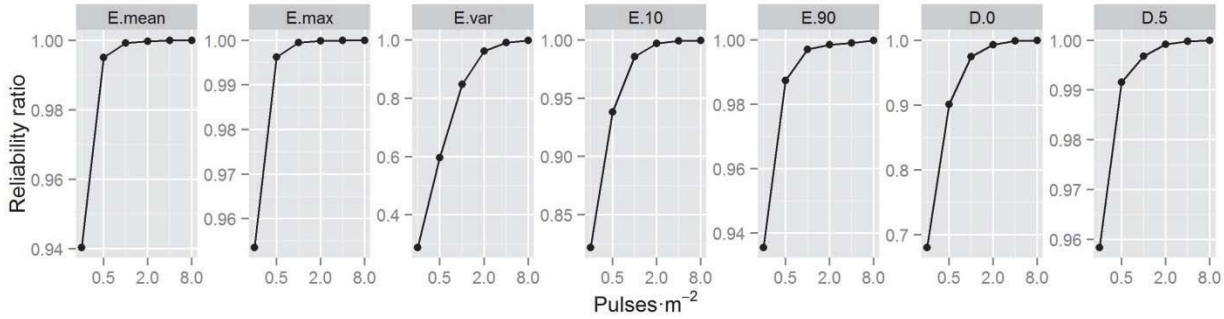


Figure 11. Plots of reliability ratios for ALS-derived variables at a plot size of 0.07 ha and pulse densities of 0.25, 0.5, 1, 2, 4, and 8 pulses·m^{-2}. E.mean (mean elevation above ground), E.max (maximum elevation above ground), E.var (variance of the elevation above ground), E.10 and E.90 (10th and 90th height percentile of canopy points), D.0 and D.5 (the proportion of points above the ground and above the mean canopy height).

Valbuena et al. (2012) assessed the vertical accuracy of a dGNSS receiver (Topcon Hiperpro), similar to the receiver used in the present thesis, under pine canopies in Spain. By using true coordinates obtained in a total station traverse, they found the accuracy to be 1.18 m with a standard deviation of 1.55 m. It is therefore expected that recordings under dense rainforest canopies with the reception of fewer satellite signals and more problems with multipath signals, result in lower accuracy.

Recent studies of biomass in tropical forests using ALS have been conducted using different pulse densities and plot sizes. Pulse densities from 25 pulses·m^{-2} (d'Oliveira et al., 2012) down to about 1.5 pulses·m^{-2} (Asner & Mascaro, 2014) have been used. The results from these studies are similar in terms of biomass prediction performance and show that high pulse density is not a requisite for estimation of forest biomass. In biomass studies where the key information is the vegetation height relative to the terrain elevation derived from the same ALS data, a systematic shift in the modelled surface is not a problem in itself. Of greater concern is the random error of the modelled terrain elevation. As expected, the standard deviation of $\overline{D}$ (S_D) increased with reduced pulse density. S_D increased from 2.9 m to 3.2 m when pulse density was reduced from 8 and 0.25 pulses·m^{-2}, respectively. This variation will directly translate into variation in the ALS-derived canopy variables.

Analysis of commonly applied canopy variables showed that the variables were affected differently by pulse density. As previously documented by Gobakken and Næsset (2008), E.max, which characterizes the maximum elevation of the canopy, decreases with decreasing pulse density. Mean values of the other variables assessed in the present study were found to

be stable. Reduced pulse density increased the variation in canopy variables and will result in models with increased residual variance. The estimated reliability of the explanatory variables, expressed by the reliability ratio, showed that all variables were reliable (reliability ratio >0.9) at pulse densities of down to 2 pulses·m^{-2}. Further reduction of pulse density resulted in some canopy variables becoming less reliable although most variables retain a reliability ratio of >0.9 at 0.5 pulses·m^{-2}. In low pulse density conditions (<1·m^{-2}), and with use of explanatory variables with a reliability ratio <0.9 Magnussen et al. (2010) proposed the use of a model calibration procedure.

Increasing the field plot size reduces the variation in ALS-derived variables and could counter the effects in low pulse density missions, concurring with the results of Gobakken and Næsset (2008). However, larger field plots are costly and finding the optimal balance between costs and desired accuracy has for decades been an issue of interest in designing forest inventories. Zeide (1980) presented how to optimize the plot size for systematic sampling. The optimal plot size is a function of the coefficient of variation between plots, plot measurement time, and travelling time between plots, under budgetary restraints or for a desired accuracy.

Although the study in Paper II aimed to simulate the effects caused by various flight elevations and speeds on the DTM and canopy variables derived from ALS data, effects such as increased footprint size and reduced pulse energy were not simulated and studied. Studies of the effects of footprint size on derived tree heights have shown that increased footprint size reduces the derived tree height estimates (Andersen et al., 2006). Andersen et al. (2006) found that this effect was stronger for trees with narrow crowns, and thus, we can expect that the effect is small in a tropical forest with wide tree crowns. Larger footprint sizes will, however, also have less energy per area unit and be less able to penetrate through the canopy (Hyyppä et al., 2005; Goodwin et al., 2006), resulting in a lower proportion of ground points. The influence of footprint size and pulse energy is likely to be of importance and should be investigated in future studies.

4.1.3 Effects of plot size on relative efficiency of ALS and InSAR data

In Paper III the main objective was to assess the effect of plot size on the relative efficiency of using auxiliary data from ALS and InSAR in estimation of biomass. Using FD2, separate log-log models were constructed for plot sizes of 700, 900, …, 1900 m^2 with auxiliary data from (1) the terrain elevation from a DTM (TE), (2) ALS, and (3) InSAR. TE models showed a positive correlation between biomass and elevation, and the explanatory variable was increasingly significant from $p = 0.044$ at 700 m^2 to $p = 0.002$ at 1900 m^2. Biomass was also

positively correlated to the two explanatory variables in the ALS models and the InSAR-height used in the InSAR models. All variables were significant at a 95% level except one of the ALS variables (D.1.L) at plot sizes of 1100–1700 m^2.

Inspection of the scatterplots of observed *vs.* predicted biomass (Figures 12–14) showed that the models had a lack of fit resulting in over-prediction of biomass in areas of low biomass and under-prediction in areas of high biomass. Similar lack of fit has been reported in studies from areas with high forest density (e.g. Nord-Larsen & Schumacher, 2012; Vincent et al., 2012). The plots of the grouped means of observations *vs.* predictions (Figures 15–17), however, showed small differences.

Increasing the plot size from 700 to 1900 m^2 reduced the SE of the mean estimates from 15.3 to 10.6% using TE, from 10.1 to 5.1% using ALS, and from 11.3 to 6.4% using InSAR (Figure 18). Both ALS and InSAR performed well compared to TE in terms of SE. ALS and InSAR estimates had an SE of about 5 and 4 percentage points lower than TE, respectively. Further, InSAR performed well compared to the ALS with only 0.4–1.3 percentage points higher SE depending on plot size. The differences in SE translated into relative efficiencies of 3.6–6.7 using ALS and 2.6–4.0 using InSAR, compared to TE (Figure 19). The relative efficiency of the ALS data also increased with increased plot size relative also to the InSAR data (Figure 19). At a plot size of 1900 m^2 the ALS was 6.7 times as efficient as using TE and 1.7 times as efficient as InSAR. The fact that the relative efficiency of ALS and InSAR increased with increased plot size may partly be due to reduced relative influence of boundary effects and co-registration errors. The slight increase in relative efficiency of ALS compared to InSAR may also indicate that the relative influence of boundary effects and co-registration errors is stronger for ALS than for InSAR. The relative efficiency of ALS compared to InSAR is modest compared to studies in Norway that have found the relative efficiency of ALS to be about twice to that of InSAR (Næsset et al., 2011; Rahlf et al., 2014).

As stated by Gregoire et al. (2015), information about the approach to statistical inference, design- or model-based, is essential in assessing the estimated variance. Taking the design-based approach to variance estimation d'Oliveira et al. (2012) reported a relative efficiency of 3.4 in a study utilising 50 plots of ~0.25 ha in the Brazilian Amazon. A relative efficiency of 2.1 can similarly by computed from the variance estimates in Paper I. Large negative boundary effects in the ALS variables would contribute to a lower relative efficiency for smaller plots like the plots of ~0.1 ha used in Paper I.

The relative efficiencies computed in Paper III in the model-based framework cannot be used as a factor for calculating the contribution of the remotely sensed data in terms of added

observations (see section 3.3.4). Instead a Pólya-urn resampling scheme was used to simulate the effect of additional field observations on the TE models. To reach similar levels of variance as for the ALS models with the TE models, the number of field plots would have to be increased by a factor of 3.5–6 depending on plot size (Figure 20).

The DTM used directly to derive the TE variable in the TE-models, and to derive the InSAR elevation above the terrain, was derived from the ALS data. DTMs constructed from ALS data have generally high accuracy (Meng et al., 2010). In the absence of an ALS-derived DTM, a DTM derived from other sources would have influenced the results. A DTM derived from sources like P-band SAR (e.g. Neeff et al., 2005) or the topographic map series of Tanzania, would most likely have resulted in substantially increased SE of the InSAR and TE estimates. In a study using InSAR height to estimate forest biomass in Norway Næsset et al. (2011) it was found that relative RMSE was approximately seven percentage points higher using a DTM from topographic maps with a contour interval of 20 m, compared to using an ALS-derived DTM. P-band SAR, used with good results in Neeff et al. (2005), is currently only available from airborne platforms, and was not collected in ANR.

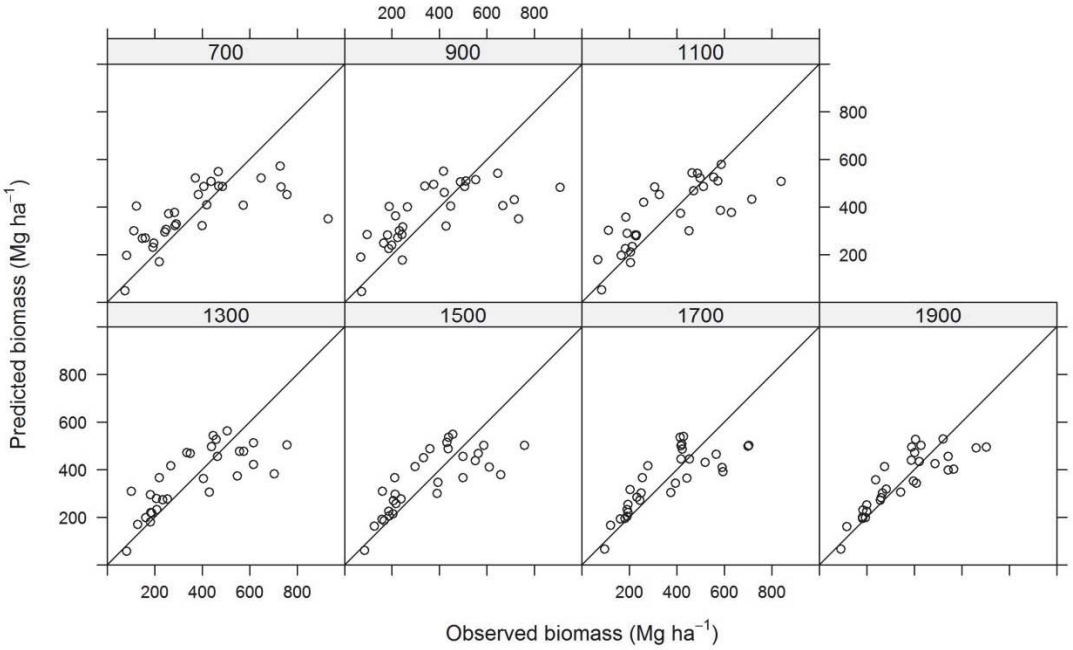


Figure 12. Observed *vs.* predicted biomass values using ALS for plot sizes of 700, 900,…, 1900 m^2.

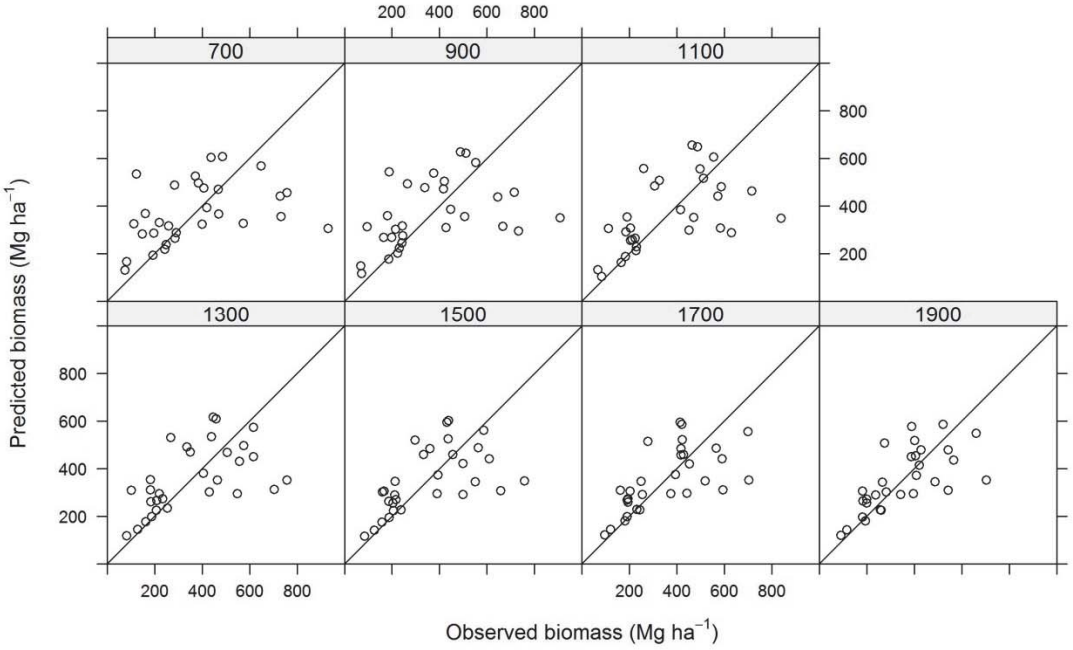


Figure 13. Observed *vs.* predicted biomass values using InSAR for plot sizes of 700, 900,…, 1900 m^2.

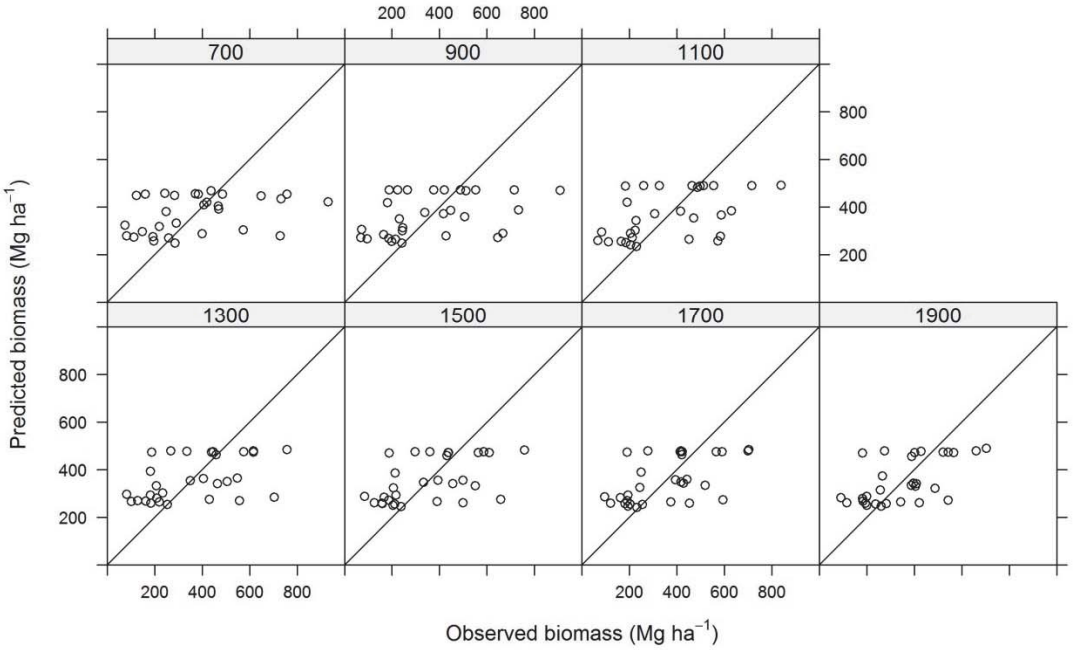


Figure 14. Observed *vs.* predicted biomass values using TE for plot sizes of 700, 900,…, 1900 m^2.

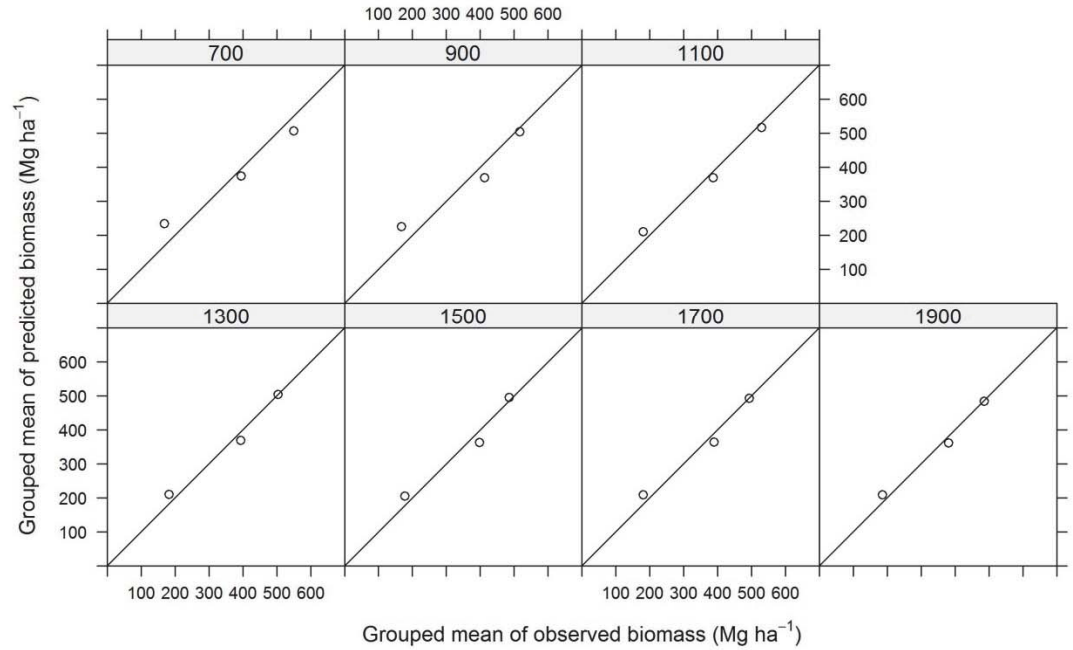


Figure 15. Grouped means of observed *vs.* predicted biomass values using ALS for plot sizes of 700, 900,…, 1900 m^2.

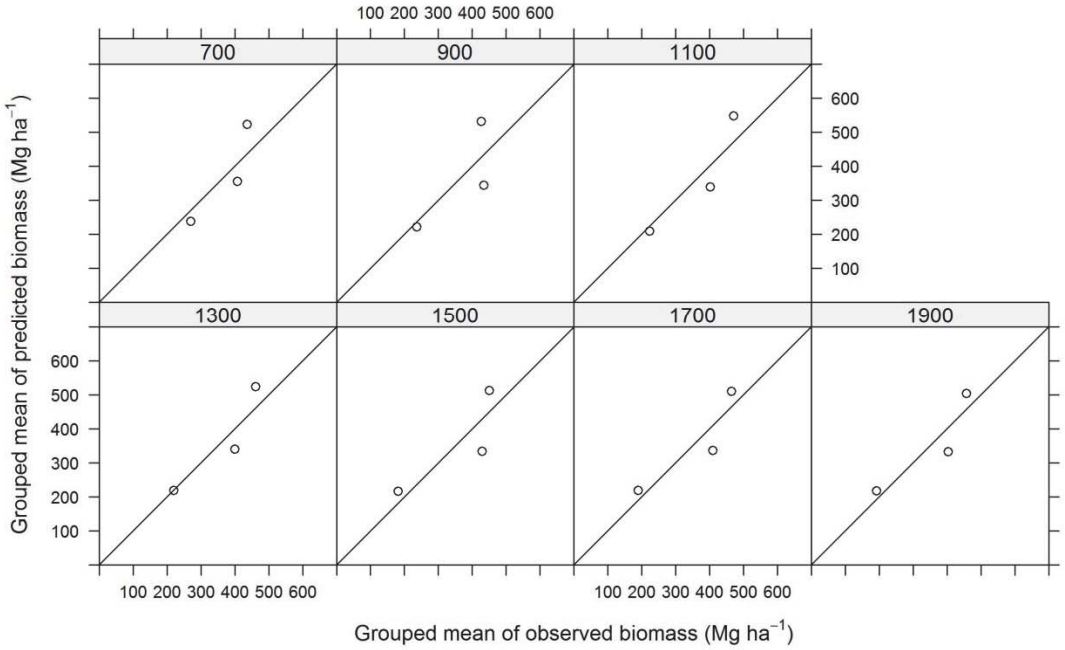


Figure 16. Grouped means of observed *vs.* predicted biomass values using InSAR for plot sizes of 700, 900,…, 1900 m^2.

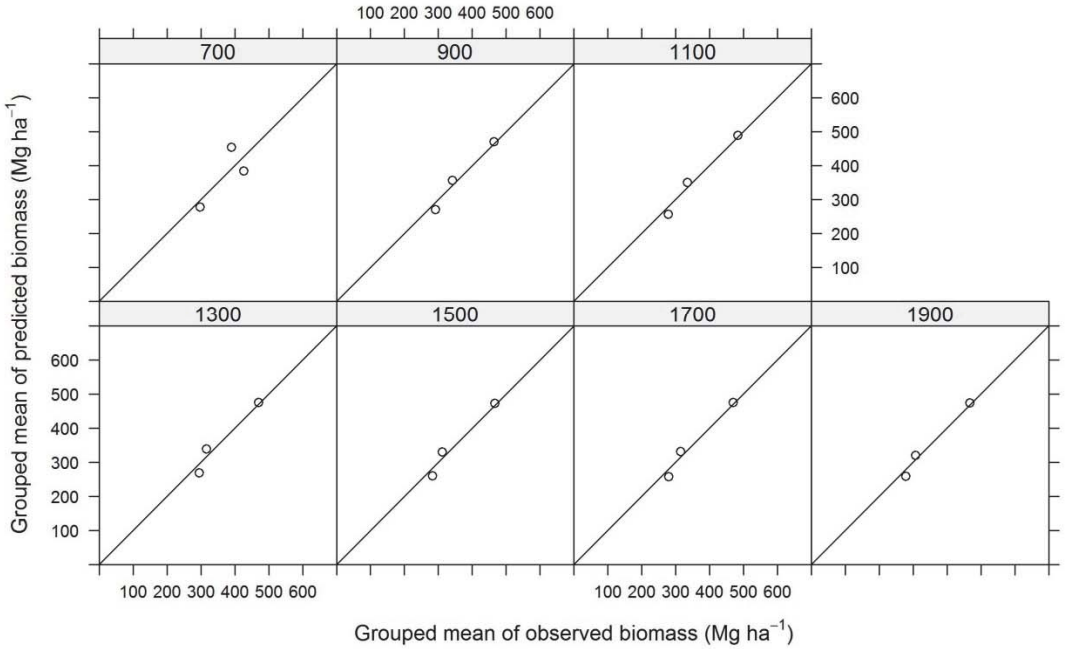


Figure 17. Grouped means of observed *vs.* predicted biomass values using TE for plot sizes of 700, 900,…, 1900 m^2.

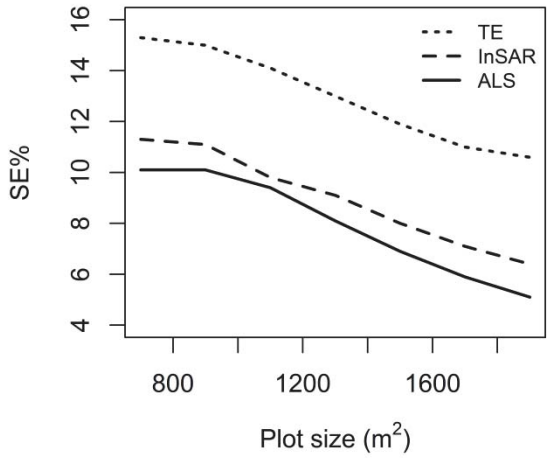


Figure 18. Relative standard error of biomass estimates (SE%) using models with auxiliary data of terrain elevation (TE) derived from a digital terrain model (dotted line), InSAR (dashed line), and ALS (solid line).

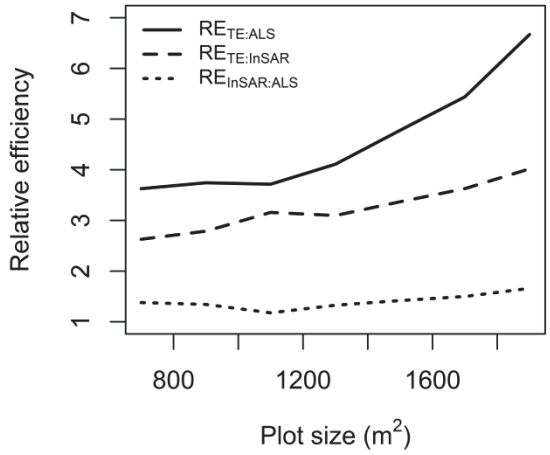


Figure 19. Relative efficiency of using InSAR ($RE_{TE:InSAR}$, dashed line) and ALS ($RE_{TE:ALS}$, solid line) relative to TE for biomass estimation, and ALS relative to InSAR ($RE_{InSAR:ALS}$, dotted line).

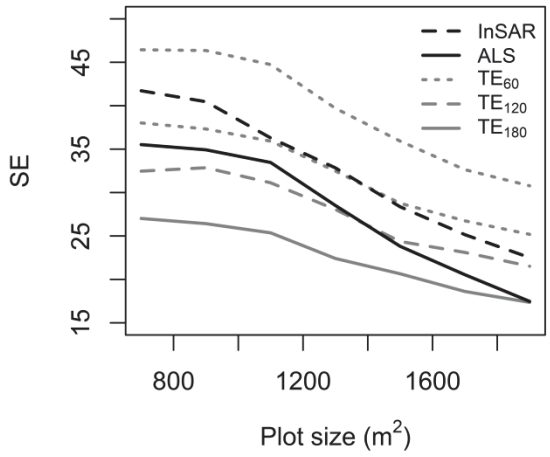


Figure 20. Standard error of biomass estimates (SE) using models with auxiliary data of InSAR (dashed line), ALS (solid line), and TE. TE model SE is derived from 60 (dotted grey line), 120 (dashed grey line), and 180 (solid grey line) simulated observations.

5 Final comments and future prospects

The analyses undertaken in the three papers which constitute this thesis have increased the understanding and basis for forest sample surveys in high-biomass tropical rainforests supported by remotely sensed auxiliary ALS- and InSAR-data. The studies shed light on important aspects of the relationship between biomass observed in the field and the remotely sensed data. A particular focus was directed at how the size of field plots affects this relationship.

Initially, Paper I established that aboveground biomass could be estimated using ALS and field plots of ca. 0.1 ha, at an accuracy of approximately 33% RMSE. Furthermore, the analyses in Paper II demonstrated that similar accuracies could be achieved using potentially cheaper ALS data, with lower pulse densities. Lastly, comparisons of estimated precision in estimation of biomass using both ALS and InSAR data documented that ALS gave the highest precision. The study reported in Paper III further documented that larger plots were relatively more efficient in improving the precision of estimates supported by remotely sensed data, compared to estimates from field observations supported by terrain elevation.

Although the use of ALS data as auxiliary information can reduce the number of field observations needed to obtain a desired accuracy level, the cost of field labour in Tanzania is low. With a cost of establishing an additional inventory plot of only 100–150 USD per plot (E. Mauya 2015, pers. comm. 19 Jan.) increasing the number of field observations is the cheapest way of increasing the accuracy of biomass estimates. However, there are still good reasons for investing in ALS as an auxiliary source of data. Firstly, the ALS DTM can be utilized with other, cheaper sources of remotely sensed data, such as SAR or optical satellite images. Secondly, stability in measurement of InSAR height over time as documented in Solberg et al. (2015) presents a potential of repeatedly estimating biomass without the need of repeated field surveys. This would permit biomass stock estimates at short intervals, annually or bi-annual for instance. Finally, the ALS-assisted inventory can, in addition to the total biomass estimate, provide a map of desired forest parameters. Maps of other forest traits, such as structure, are increasingly being used for improving our understanding of a range of ecological subjects (Maltamo et al., 2014; Pettorelli et al., 2014; Potts et al., 2014).

Collection of remotely sensed data using aircraft is inevitably relatively expensive. With an increasing demand for cheaper and more available information about forest ecosystems, satellite based sensors such as the ICESat-2 and Global Ecosystem Dynamics Investigation (GEDI) LiDARs of National Aeronautics and Space Administration and the P-band SAR

BIOMASS mission of the European Space Agency, are planned for launching in 2017, 2018 and 2020 respectively. In addition, new ALS missions are continuously being carried out and calls have been made for greater coverage of ALS (Mascaro et al., 2014) financed under a REDD+ framework. This gives us great possibilities of understanding further aspects of these important ecosystems.

Knowledge of how the remote sensing technologies are related to ground observations at different scales can be utilized to optimize both plot- and sampling design, in order to minimize the total inventory cost whilst still reaching the required level of precision.

References

Andersen, H.-E., Reutebuch, S. E. & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing*, 32 (5): 355-366.

Andersen, H.-E., Reutebuch, S. E., McGaughey, R. J., d'Oliveira, M. V. N. & Keller, M. (2013). Monitoring selective logging in western Amazonia with repeat lidar flights. *Remote Sensing of Environment*, 151: 157-165.

Anderson, E. S., Thompson, J. A., Crouse, D. A. & Austin, R. E. (2006). Horizontal resolution and data density effects on remotely sensed LIDAR-based DEM. *Geoderma*, 132 (3–4): 406-415.

Anon. (1999). *Pinnacle user's manual*. San Jose, CA: Javad Positioning Systems.

Anon. (2012). *TerraScan user's guide*. Jyvaskyla, Finland: Terrasolid Oy. p. 311.

Asner, G. P., Powell, G. V. N., Mascaro, J., Knapp, D. E., Clark, J. K., Jacobson, J., Kennedy-Bowdoin, T., Balaji, A., Paez-Acosta, G., Victoria, E., et al. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (38): 16738-16742.

Asner, G. P., Clark, J., Mascaro, J., Vaudry, R., Chadwick, K. D., Vieilledent, G., Rasamoelina, M., Balaji, A., Kennedy-Bowdoin, T., Maatoug, L., et al. (2012). Human and environmental controls over aboveground carbon storage in Madagascar. *Carbon Balance and Management*, 7 (2).

Asner, G. P., Knapp, D. E., Martin, R. E., Tupayachi, R., Anderson, C. B., Mascaro, J., Sinca, F., Chadwick, K. D., Higgins, M., Farfan, W., et al. (2014). Targeted carbon conservation at national scales with high-resolution monitoring. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (47): E5016-E5022.

Asner, G. P. & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, 140: 614-624.

Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *International Archives of the Photogrammetry and Remote Sensing*, 33: 110-117.

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

Bowen, Z. H. & Waltermire, R. G. (2002). Evaluation of light detection and ranging (LIDAR) for measuring river corridor topography. *Journal of the American Water Resources Association*, 38 (1): 33-41.

Boyd, D. S. & Danson, F. M. (2005). Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography*, 29 (1): 1-26.

Burgess, N. D., Butynski, T. M., Cordeiro, N. J., Doggart, N. H., Fjeldsa, J., Howell, K. M., Kilahama, F. B., Loader, S. P., Lovett, J. C., Mbilinyi, B., et al. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, 134 (2): 209-231.

Chen, Q., Vaglio Laurin, G. & Valentini, R. (2015). Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. *Remote Sensing of Environment*, 160 (0): 134-143.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., et al. (2013). Carbon and Other Biogeochemical Cycles. In Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P. M. (eds) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 465-570. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Clark, D. B. & Kellner, J. R. (2012). Tropical forest biomass estimation and the fallacy of misplaced concreteness. *Journal of Vegetation Science*, 23 (6): 1191-1196.

Clark, M. L., Roberts, D. A., Ewel, J. J. & Clark, D. B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*, 115 (11): 2931-2942.

d'Oliveira, M. V. N., Reutebuch, S. E., McGaughey, R. J. & Andersen, H. E. (2012). Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sensing of Environment*, 124: 479-491.

Dawson, W., Mndolwa, A. S., Burslem, D. & Hulme, P. E. (2008). Assessing the risks of plant invasions arising from collections in tropical botanical gardens. *Biodiversity and Conservation*, 17 (8): 1979-1995.

Ene, L. T., Næsset, E., Gobakken, T., Gregoire, T. G., Stahl, G. & Holm, S. (2013). A simulation approach for accuracy assessment of two-phase post-stratified estimation in large-area LiDAR biomass surveys. *Remote Sensing of Environment*, 133: 210-224.

FAO. (1948). *Forest resources of the world*. Unasylva, vol. 2, no. 4. Available at: http://www.fao.org/docrep/x5345e/x5345e00.htm (accessed: 16.01.2015).

FAO. (2011). *State of the World's Forests 2011*: Food and Agriculture Organization of the UN, Rome. 164 pp.

Fassnacht, F. E., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P. & Koch, B. (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment*, 154: 102-114.

Friedlingstein, P., Andrew, R. M., Rogelj, J., Peters, G. P., Canadell, J. G., Knutti, R., Luderer, G., Raupach, M. R., Schaeffer, M., van Vuuren, D. P., et al. (2014). Persistent growth of CO2 emissions and implications for reaching climate targets. *Nature Geosci*, 7 (10): 709-715.

Frontier Tanzania. (2001). Amani Nature Reserve: A biodiversity survey. In Doody, K. Z., Howell, K. M. & Fanning, E. (eds). *East Usambara Conservation Area Management Programme Technical Paper*. Dar es Salaam, Tanzania and Vantaa, Finland: Forestry and Beekeeping Division and Metsähallitus Consulting.

Fuller, W. A. (1987). *Measurement error models*. New York: John Wiley and Sons. 440 pp.

Ghosh, M. & Meeden, G. (1997). *Bayesian methods for finite population sampling*. London, UK: Chapman & Hall.

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*, 2 (4).

Gobakken, T. & 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 (5): 1095-1109.

Gobakken, T. & Næsset, E. (2009). Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, 39 (5): 1036-1052.

Gobakken, T., Bollandsås, O. M. & Næsset, E. (2014). Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data. *Scandinavian Journal of Forest Research*: 1-14.

Goldberger, A. (1968). Interpretation and estimation of Cobb-Douglas functions. *Econometrica*, 36 (3-4): 464-472.

Goldstein, R. M. & Werner, C. L. (1998). Radar interferogram filtering for geophysical applications. *Geophysical Research Letters*, 25 (21): 4035-4038.

Goodwin, N. R., Coops, N. C. & Culvenor, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, 103 (2): 140-152.

Grace, J., Mitchard, E. & Gloor, E. (2014). Perturbations in the carbon budget of the tropics. *Global Change Biology*, 20 (10): 3238-3255.

Grassi, G., Monni, S., Federici, S., Achard, F. & Mollicone, D. (2008). Applying the conservativeness principle to REDD to deal with the uncertainties of the estimates. *Environmental Research Letters*, 3 (3).

Gregoire, T. G., Lin, Q. F., Boudreau, J. & Nelson, R. (2008). Regression Estimation Following the Square-Root Transformation of the Response. *Forest Science*, 54 (6): 597-606.

Gregoire, T. G. & Valentine, H. T. (2008). *Sampling Strategies for Natural Resources and the Environment*. Chapman & Hall/CRC Applied Environmental Statistics. Boca Raton, Florida, USA: Chapman and Hall/CRC

Gregoire, T. G., Næsset, E., Ståhl, G., Andersen, H.-E., Gobakken, T., Ene, L., Nelson, R. F. & McRoberts, R. E. (2015). Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sensing of Environment*, under review.

Hamilton, A. C. & Bensted-Smith, R. (1989). *Forest conservation in the East Usambara Mountains, Tanzania*. Gland, Switzerland and Cambridge, UK: IUCN-The World Conservation Union ; Dar es Salaam, Tanzania : Forest Division, Ministry of Lands, Natural Resources, and Tourism, United Republic of Tanzania. 392 pp.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., et al. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342 (6160): 850-853.

Haralick, R. M., Shanmugam, K. & Dinsteein, J. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern*, 3: 610-621.

Henry, M., Besnard, A., Asante, W. A., Eshun, J., Adu-Bredu, S., Valentini, R., Bernoux, M. & Saint-Andre, L. (2010). Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. *Forest Ecology and Management*, 260 (8): 1375-1388.

Hodgson, M. E., Jensen, J., Raber, G., Tullis, J., Davis, B. A., Thompson, G. & Schuckman, K. (2005). An evaluation of lidar-derived elevation and terrain slope in leaf-off conditions. *Photogrammetric Engineering and Remote Sensing*, 71 (7): 817-823.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6 (2): 65-70.

Hou, Z., Xu, Q. & Tokola, T. (2011). Use of ALS, Airborne CIR and ALOS AVNIR-2 data for estimating tropical forest attributes in Lao PDR. *Isprs Journal of Photogrammetry and Remote Sensing*, 66 (6): 776-786.

Houghton, R. A. (2012). Carbon emissions and the drivers of deforestation and forest degradation in the tropics. *Current Opinion in Environmental Sustainability*, 4 (6): 597-603.

Hyyppä, J., Yu, X., Hyyppä, H., Kaartinen, H., Honkavara, E. & Rönnholm, P. (2005). *Factors affecting the quality of DTM generation in forested areas*. Proceedings of ISPRS Workshop on Laser Scanning 2005, 12–14 September, Enschede, Netherlands. 85-90 pp.

Höhle, J. & Höhle, M. (2009). Accuracy assessment of digital elevation models by means of robust statistical methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64 (4): 398-406.

Iida, Y., Kohyama, T. S., Kubo, T., Kassim, A., Poorter, L., Sterck, F. & Potts, M. D. (2011). Tree architecture and life-history strategies across 200 co-occurring tropical tree species. *Functional Ecology*, 25 (6): 1260-1268.

Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M.-H., Wong, W. V. C., Ling, Z.-Y., Saito, H. & Takao, G. (2014). Estimating above-ground biomass of tropical rainforest of different degradation levels in Northern Borneo using airborne LiDAR. *Forest Ecology and Management*, 328: 335-341.

IPCC. (2014). Fifth Assessment Synthesis Report. 116 pp.

ISO. (2012). *ISO 5725-1. Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*: International Organization for Standardization,

Geneva, Switzerland. Available at: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en (accessed: 15.01.2015).

Jubanski, J., Ballhorn, U., Kronseder, K., Franke, J. & Siegert, F. (2012). Detection of large above ground biomass variability in lowland forest ecosystems by airborne LiDAR. *Biogeosciences Discuss.*, 9 (8): 11815-11842.

Keith, H., Mackey, B. G. & Lindenmayer, D. B. (2009). Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *Proceedings of the National Academy of Sciences*, 106 (28): 11635-11640.

Kouba, J. (2009). *A guide to using international GNSS service (IGS) products*. IGS Central Bureau, Jet Propulsion Laboratory, Pasadena, CA. p. 34.

Kraus, K. & Pfeifer, N. (1998). Determination of terrain models in wooded areas with airborne laser scanner data. *Isprs Journal of Photogrammetry and Remote Sensing*, 53 (4): 193-203.

Köhl, M., Magnussen, S. & Marchetti, M. (2006). *Sampling methods, remote sensing and GIS multiresource forest inventory*. Berlin: Springer. 365 pp.

Laurin, G. V., Chen, Q., Lindsell, J. A., Coomes, D. A., Frate, F. D., Guerriero, L., Pirotti, F. & Valentini, R. (2014). Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 89: 49-58.

Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., et al. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115 (11): 2850-2860.

Leitold, V., Keller, M., Morton, D. C., Cook, B. D. & Shimabukuro, Y. E. (2015). Airborne lidar-based estimates of tropical forest structure in complex terrain: opportunities and trade-offs for REDD+. *Carbon Balance Manag*, 10 (1): 3.

Lim, K., Hopkinson, C. & Treitz, P. (2008). Examining the effects of sampling point densities on laser canopy height and density metrics. *The Forestry Chronicle*, 84 (6): 876-885.

Lumley, T. & Miller, A. (2009). *leaps: regression subset selection*. R package version 2.9.

Magnussen, S., Næsset, E. & Gobakken, T. (2010). Reliability of LiDAR derived predictors of forest inventory attributes: A case study with Norway spruce. *Remote Sensing of Environment*, 114 (4): 700-712.

Magnussen, S., Næsset, E., Gobakken, T. & Frazer, G. (2012). A fine-scale model for area-based predictions of tree-size-related attributes derived from LiDAR canopy heights. *Scandinavian Journal of Forest Research*, 27 (3): 312-322.

Maltamo, M., Næsset, E. & Vauhkonen, J. (2014). *Forestry Applications of Airborne Laser Scanning - Concepts and Case Studies. Part III Ecological Applications*. Managing Forest Ecosystems, vol. 27: Springer Netherlands. 335-462 pp.

Marshall, A. R., Willcock, S., Platts, P. J., Lovett, J. C., Balmford, A., Burgess, N. D., Latham, J. E., Munishi, P. K. T., Salter, R., Shirima, D. D., et al. (2012). Measuring and modelling above-ground carbon and tree allometry along a tropical elevation gradient. *Biological Conservation*, 154: 20-33.

Mascaro, J., Detto, M., Asner, G. P. & Muller-Landau, H. C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of Environment*, 115 (12): 3770-3774.

Mascaro, J., Asner, G., Davies, S., Dehgan, A. & Saatchi, S. (2014). These are the days of lasers in the jungle. *Carbon Balance and Management*, 9 (1): 7.

Masota, A. M., Zahabu, E., Malimbwi, R., Bollandsås, O. M. & Eid, T. (2015). Tree Allometric Models for Above- and Belowground Biomass of Tropical Rainforests in Tanzania. *Submitted to: Southern Forests: a Journal of Forest Science*.

Matthews, H. D., Gillett, N. P., Stott, P. A. & Zickfeld, K. (2009). The proportionality of global warming to cumulative carbon emissions. *Nature*, 459 (7248): 829-832.

Mauya, E., Hansen, E. H., Gobakken, T., Bollandsås, O. M., Malimbwi, R. E. & Næsset, E. (2015). Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. *Submitted to: Carbon Balance Manage.*

McGaughey, R. J. (2013). *FUSION/LDV: Software for LIDAR Data Analysis and Visualization*. 3.30 ed. Available at: http://forsys.cfr.washington.edu/fusion (accessed: 09.01.2015).

McRoberts, R. E., Tomppo, E. O. & Næsset, E. (2010). Advances and emerging issues in national forest inventories. *Scandinavian Journal of Forest Research*, 25 (4): 368-381.

McRoberts, R. E., Næsset, E. & Gobakken, T. (2013a). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sensing of Environment*, 128: 268-275.

McRoberts, R. E., Tomppo, E. O., Vibrans, A. C. & de Freitas, J. V. (2013b). Design considerations for tropical forest inventories. *Brazilian journal of forestry research*, 33 (74): 189-202.

McRoberts, R. E., Andersen, H. E. & Næsset, E. (2014a). Using Airborne Laser Scanning Data to Support Forest Sample Surveys. In Maltamo, M., Næsset, E. & Vauhkonen, J. (eds) Managing Forest Ecosystems, vol. 27 *Forestry Applications of Airborne Laser Scanning*, pp. 269-292: Springer Netherlands.

McRoberts, R. E., Næsset, E. & Gobakken, T. (2014b). Estimation for inaccessible and non-sampled forest areas using model-based inference and remotely sensed auxiliary information. *Remote Sensing of Environment*, 154: 226-233.

Mehtatalo, L. (2012). *Forest biometrics functions of Lauri Mehtatalo*. R package version 1.1 ed. Available at: http://cs.uef.fi/~lamehtat/rcodes (accessed: 09.01.2015).

Meng, X. L., Currit, N. & Zhao, K. G. (2010). Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sensing*, 2 (3): 833-860.

Mgumia, F. (2014). Implications of forestland tenure and status changes on resource base, forest govenance, and community livelihood sorrounding Amani Nature Reserve. *Unpublished draft*.

Minh, D. H. T., Le Toan, T., Rocca, F., Tebaldini, S., d'Alessandro, M. M. & Villard, L. (2014). Relating P-Band Synthetic Aperture Radar Tomography to Tropical Forest Biomass. *Geoscience and Remote Sensing, IEEE Transactions on*, 52 (2): 967-979.

Mitchard, E. T. A., Saatchi, S. S., Woodhouse, I. H., Nangendo, G., Ribeiro, N. S., Williams, M., Ryan, C. M., Lewis, S. L., Feldpausch, T. R. & Meir, P. (2009). Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, 36 (L23401).

Mpanda, M. M., Luoga, E. J., Kajembe, G. C. & Eid, T. (2011). Impact of forestland tenure changes on forest cover, stocking and tree species diversity in Amani Nature Reserve, Tanzania. *Forests, Trees and Livelihoods*, 20 (4): 215-229.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403 (6772): 853-858.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., et al. (2013). Anthropogenic and natural radiative forcing. In Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P. M. (eds) *Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change*, pp. 659-740. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Neeff, T., Dutra, L. V., dos Santos, J. R., Freitas, C. d. C. & Araujo, L. S. (2005). Tropical Forest Measurement by Interferometric Height Modeling and P-Band Radar Backscatter. *Forest Science*, 51 (6): 585-594.

Newmark, W. D. (2002). *Conserving biodiversity in East African forests: a study of the eastern arc mountains*. Berlin: Springer. 197 pp.

Nord-Larsen, T. & Schumacher, J. (2012). Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. *Remote Sensing of Environment*, 119: 148-157.

Næsset, E. (1997a). Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52 (2): 49-56.

Næsset, E. (1997b). Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 61 (2): 246-253.

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 (9): 1021-1026.

Næsset, E. (2002). Determination of Mean Tree Height of Forest Stands by Digital Photogrammetry. *Scandinavian Journal of Forest Research*, 17 (5): 446-459.

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 (6): 3079-3090.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Stahl, 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 (12): 3599-3614.

Næsset, E. (2014). Area-Based Inventory in Norway – From Innovation to an Operational Reality. In Maltamo, M., Næsset, E. & Vauhkonen, J. (eds) Managing Forest Ecosystems, vol. 27 *Forestry Applications of Airborne Laser Scanning*, pp. 215-240: Springer Netherlands.

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

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H. & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology*, 51 (4): 839-848.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Development Core Team. (2014). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-115.

Plugge, D., Baldauf, T. & Köhl, M. (2011). Reduced Emissions from Deforestation and Forest Degradation (REDD): Why a Robust and Transparent Monitoring, Reporting and Verification (MRV) System is Mandatory. In Blanco, J. (ed.) *Climate Change - Research and Technology for Adaptation and Mitigation*, pp. 155-170: InTech.

Poorter, L., Bongers, L. & Bongers, F. (2006). Architecture of 54 moist-forest tree species: Traits, trade-offs, and functional groups. *Ecology*, 87 (5): 1289-1301.

Potts, J. R., Mokross, K., Stouffer, P. C. & Lewis, M. A. (2014). Step selection techniques uncover the environmental predictors of space use patterns in flocks of Amazonian birds. *Ecology and Evolution*, 4 (24): 4578-4588.

Prodan, M. (1968). *Forest biometrics*. Oxford, UK: Pergamon Press. 447 pp.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org/ (accessed: 10.01.2015).

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E. & Astrup, R. (2014). Comparison of four types of 3D data for timber volume estimation. *Remote Sensing of Environment*, 155: 325-333.

Reutebuch, S. E., McGaughey, R. J., Andersen, H. E. & Carson, W. W. (2003). Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing*, 29 (5): 527-535.

Skowronski, N., Clark, K., Nelson, R., Hom, J. & Patterson, M. (2007). Remotely sensed measurements of forest structure and fuel loads in the Pinelands of New Jersey. *Remote Sensing of Environment*, 108 (2): 123-129.

Snowdon, P. (1991). A ratio estimator for bias correction in logarithmic regressions. *Canadian Journal of Forest Research*, 21 (5): 720-724.

Solberg, S., Lohne, T.-P. & Karyanto, O. (2015). Temporal stability of InSAR height in a tropical rainforest. *Remote Sensing Letters*, 6 (3): 209-217.

Stoltzenberg, R. M. (2009). Multiple Regression Analysis. In Hardy, M. A. & Bryman, A. (eds) *Handbook of Data Analysis*, pp. 165-207. London: SAGE Publications Ltd.

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 (1): 96-107.

Tinkham, W. T., Huang, H. Y., Smith, A. M. S., Shrestha, R., Falkowski, M. J., Hudak, A. T., Link, T. E., Glenn, N. F. & Marks, D. G. (2011). A Comparison of Two Open Source LiDAR Surface Classification Algorithms. *Remote Sensing*, 3 (3): 638-649.

Toutin, T. & Gray, L. (2000). State-of-the-art of elevation extraction from satellite SAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55 (1): 13-33.

Töyrä, J., Pietroniro, A., Hopkinson, C. & Kalbfleisch, W. (2003). Assessment of airborne scanning laser altimetry (lidar) in a deltaic wetland environment. *Canadian Journal of Remote Sensing*, 29 (6): 718-728.

UNFCCC. (2006). *Good practice guidance and adjustments under Article 5, paragraph 2, of the Kyoto Protocol. FCCC/KP/CMP/2005/8/Add.3 Decision 20/CMP.1*.

UNFCCC. (2010). *Report of the Conference of the Parties on its fifteenth session, held in Copenhagen from 7 to 19 December 2009. Addendum. Part Two: Action taken by the Conference of the Parties at its fifteenth session*: United Nations Office, Geneva, Switzerland. p. 43.

UNFCCC. (2011). *Report of the Conference of the Parties on its sixteenth session, held in Cancun from 29 November to 10 December 2010. Addendum. Part two: Action taken by the Conference of the Parties at its sixteenth session*: United Nations Office, Geneva, Switzerland. p. 31.

Valbuena, R., Mauro, F., Rodriguez-Solano, R. & Antonio Manzanera, J. (2012). Partial Least Squares for Discriminating Variance Components in Global Navigation Satellite Systems Accuracy Obtained Under Scots Pine Canopies. *Forest Science*, 58 (2): 139-153.

Vauhkonen, J., Maltamo, M., McRoberts, R. & Næsset, E. (2014). Introduction to Forestry Applications of Airborne Laser Scanning. In Maltamo, M., Næsset, E. & Vauhkonen, J. (eds) Managing Forest Ecosystems, vol. 27 *Forestry Applications of Airborne Laser Scanning*, pp. 1-16: Springer Netherlands.

Vincent, G., Sabatier, D., Blanc, L., Chave, J., Weissenbacher, E., Pélissier, R., Fonty, E., Molino, J. F. & Couteron, P. (2012). Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. *Remote Sensing of Environment*, 125: 23-33.

Winsor, C. P. (1932). The Gompertz Curve as a Growth Curve. *Proceedings of the National Academy of Sciences of the United States of America*, 18 (1): 1-8.

Zandbergen, P. A. (2008). Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy. *Transactions in GIS*, 12 (1): 103-130.

Zeide, B. (1980). Plot size optimization. *Forest Science*, 26 (2): 251-257.

Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J. & Zhang, C. (2003). A progressive morphological filter for removing nonground measurements from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41 (4): 872-882.

Zolkos, S. G., Goetz, S. J. & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, 128: 289-298.

Zvoleff, A. (2014). *Calculate textures from grey-level co-occurrence matrices (GLCMs) in R*. R package version 0.3.1 ed. Available at: http://cran.r-project.org/web/packages/glcm/.

Paper I
Hansen, E.H., Gobakken, T., Bollandsås, O.M., Zahabu, E. & Næsset, E. 2015. Modeling aboveground biomass in dense tropical submontane rainforest using airborne laser scanner data. - Remote Sensing 7: 788-807.
DOI: 10.3390/rs70100788

Paper II
Hansen, E.H., Gobakken, T. & Næsset, E. Effects of pulse density on digital terrain models and canopy metrics using airborne laser scanner in a tropical rainforest.
(Submitted)

Paper III
Hansen, E.H., Gobakken, T., Solberg, S., Kangas, A., Ene, L., Mauya, E. & Næsset, E. Impact of field plot size on the relative efficiency of biomass estimation in a Tanzanian rainforest using airborne laser scanner and interferometric synthetic aperture radar as auxiliary data.
(Submitted)

ISBN: 978-82-575-1269-9
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no