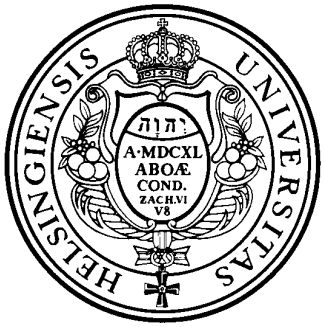

Master's thesis

Geography

Geoinformatics

MAPPING FOREST STRUCTURE WITH GINI COEFFICIENT USING DIGITAL AERIAL PHOTOGRAMMETRIC DATA FROM AN UNMANNED AERIAL SYSTEM

Johannes Nyman

2018

Supervisor:

Janne Heiskanen

UNIVERSITY OF HELSINKI
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
DIVISION OF BIOGEOSCIENCES

PL 64 (Gustaf Hällströmin katu 2)
00014 Helsingin yliopisto

Tiedekunta/Osasto Fakultet/Sektion –Faculty	Laitos/Institution– Department
Faculty of Science	Department of Geosciences and Geography

Tekijä/Författare – Author
Johannes Nyman

Työn nimi / Arbetets titel – Title
MAPPING FOREST STRUCTURE WITH GINI COEFFICIENT USING DIGITAL AERIAL PHOTOGRAMMETRIC DATA FROM AN UNMANNED AERIAL SYSTEM

Oppiaine /Läroämne – Subject
Geography (Geoinformatics)

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Master's Thesis	January 2018	60

Tiivistelmä/Referat – Abstract

Gathering information on forest structure is vital in estimating forest biodiversity, carbon stocks and temporal changes in standing forests. Currently the only viable method of collecting such information in vast areas is remote sensing (RS). Two commonly used RS methods for acquiring high resolution three dimensional information on forest structure are airborne laser scanning (ALS) and digital aerial photogrammetry (DAP). In quantifying forest structure, the distributions of tree basal areas have been used because the variation in tree sizes is closely linked to the whole concept of forest structure. Retrieving information on these distributions can be done by modelling the relationship of in situ measured distribution indices and the remotely sensed elevation information. One of these distribution indices is the Gini coefficient which has been shown to be a prominent index in describing the forest structure from ALS data.

In this study, DAP data was gathered with an unmanned aerial system (UAS) from the vicinity of the Lammi research station with the intention of investigating its suitability on modelling forest structure by using Gini coefficient (GC). Airborne laser scanning data retrieved from the National land survey was used as a comparison dataset.

The in situ measured field data consisted of tree circumference measurements from 50 circular plots (r = 5m). From these measurements, the tree basal areas were calculated and the plot level Gini coefficients determined. A comprehensive set of plot level point cloud variables were also calculated from both ALS and DAP point clouds. The most important predictor variables were chosen from the point cloud variables with an automatic exhaustive variable selection function. Then, beta regression modelling was applied to both sets of predictor variables and the best GC models determined. Finally, the models were generalized to the whole study area and GC maps were produced.

The resulting GC models for both datasets performed in a mediocre way. The best DAP model had a cross-validated RRMSE of 29.8% and the best ALS model had RRMSE of 27.2%. The coefficient of determination (R^2) was also better in the ALS model (0.49) than in the DAP model (0.39). The performance of the ALS model was slightly worse than in previous studies using ALS to predict GC. Part of this might be a repercussion of the non-optimal acquisition time of the ALS dataset. For DAP, there were no previous studies. The results of this study suggest that ALS is a more prominent method in mapping forest structure with GC. The DAP proved to be an inexpensive and flexible method of gathering three dimensional information on forests but it had poor canopy penetration abilities which affected the modelling performance negatively.

Avainsanat – Nyckelord – Keywords
Forest structure, Gini coefficient, DAP, UAS, ALS

Säilytyspaikka – Förvaringställe – Where deposited
Kumpula science library

Muita tietoja – Övriga uppgifter – Additional information

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department
Matemaattis-luonnontieteellinen tiedekunta	Geotieteiden ja maantieteen laitos

Tekijä/Författare – Author
Johannes Nyman

Työn nimi / Arbetets titel – Title
MAPPING FOREST STRUCTURE WITH GINI COEFFICIENT USING DIGITAL AERIAL PHOTOGRAMMETRIC DATA FROM AN UNMANNED AERIAL SYSTEM

Oppiaine /Läroämne – Subject
Maantiede (Geoinformatiikka)

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Pro gradu -tutkielma	Tammikuu 2018	60

Tiivistelmä/Referat – Abstract

Metsän sisäisen rakenteen tutkiminen on oleellista tietoa kartoitettaessa metsien monimuotoisuutta, hiilivarantoja tai sisäisiä muutoksia metsän rakenteessa. Tällä hetkellä kaukokartoitusmenetelmät tarjoavat parhaan menetelmän kerätä yksityiskohtaista tietoa metsän rakenteesta laajoilta alueilta. Metsien kaukokartoituksessa kaksi yleisimmin käytettyä menetelmää ovat lentolaserkeilaus ja ilmakuvaus. Niitä voidaan soveltaa myös metsien rakenteen tutkimiseen vertailemalla esimerkiksi puiden pohjapinta-alojen jakaumia, sillä puiden erikokoisuus liittyy oleellisesti metsän rakenteen käsitteeseen. Jakaumista voidaan muodostaa erilaisia jakaumaindeksejä, joista Gini indeksi on osoittautunut lentolaserkeilaustutkimuksissa parhaaksi puiden erikokoisuutta kuvaavaksi indeksiksi. Tutkimalla kaukokartoitusmenetelmien tarkan korkeustiedon yhteyttä kenttätöillä mitattuihin todellisiin jakaumaindekseihin, voidaan metsän rakennetta mallintaa laajoille alueille.

Tässä tutkimuksessa kerättiin fotogrammetrinen kolmiulotteinen pistepilviaineisto miehittämättömällä ilma-aluksella Lammin tutkimusaseman ympäristöstä. Tutkimuksen tarkoituksena oli testata fotogrammetrisen aineiston kyvykkyyttä metsän rakenteen mallintamisessa hyödyntäen Gini-indeksiä. Vertailuaineistona käytettiin Maanmittauslaitoksen lentolaserkeilausaineistoa.

Tutkimuksen kenttäaineisto koostui 50 ympyränmuotoisesta tutkimusalasta (säde 5m), joista mitattiin puiden ympärysmitat. Näistä laskettiin puiden pohjapinta-alat ja tutkimusalakohtaiset Gini-indeksiluvut. Molemmista kaukokartoitusaineistoista laskettiin koealoille korkeuteen perustuvia pistepilvimuuttujia, joista parhaat muuttujat valittiin käyttäen hyödyksi automatisoitua muuttujavalintafunktiota. Lineaarista betaregressiota hyödyntäen valituista muuttujajoukoista koostettiin molemmille aineistoille parhaat mahdolliset Gini-indeksi mallit. Lopuksi nämä mallit yleistettiin koko tutkimusalueelle ja kartat ennustetuista gini-indekseistä koostettiin.

Luodut mallit suoriutuivat Gini-indeksin ennustamisesta keskivertoisesti. Parhaan fotogrammetrisen mallin suhteellinen ristiinvalidoitu keskivirhe oli 29,8% ja parhaan laserkeilauspohjaisen mallin 27,2%. Lentolaserkeilausmallin selitysaste (0,49) osoittautui myös paremmaksi kuin fotogrammetriapohjaisen mallin (0,39). Lentolaserkeilausmallin laatuluvut olivat hiukan heikompia kuin vastaavissa aiemmissa tutkimuksissa, mikä saattoi johtua osittain ei-optimaalisesta keilausajankohdasta. Fotogrammetrinen mallin käytöstä metsän rakenteen tutkimisessa ei ole aikaisempia tutkimuksia. Tämän tutkielman tulokset puoltavat lentolaserkeilauksen käyttämistä metsän rakenteen kartoittamisessa fotogrammetrian sijaan. Fotogrammetrinen metodi osoittautui edulliseksi ja joustavaksi tavaksi kerätä kolmiulotteista tietoa metsistä, mutta sen kyvyttömyys kerätä informaatiota latvuskerroksen alta huononsi sen suoriutumista.

Avainsanat – Nyckelord – Keywords
Metsän rakenne, kaukokartoitus, Gini-indeksi, fotogrammetria, lentolaserkeilaus, miehittämätön ilma-alus

Säilytyspaikka – Förvaringställe – Where deposited
Kumpulan tiedekirjasto

Muita tietoja – Övriga uppgifter – Additional information

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
ALS	Airborne laser scanning
BIC	Bayesian information criterion
CCD	Charge-coupled device
CMOS	Complementary metal-oxide semiconductor
CO2	Carbon Dioxide
CP	Mallow's C_p
CSV	Comma separated value
CV	Cross-validation
DAP	Digital aerial photogrammetry
DBH	Diameter at breast height
dGPS	Differential global positioning system
DTM	Digital terrain model
ERS	European Remote Sensing Satellite
GC	Gini Coefficient
GCP	Ground control point
GIS	Geographical information system
GNSS	Global navigation satellite system
GPS	Global positioning system

JERS	Japanese Earth Resources Satellite
ID	Identifier
IMU	Inertia measurement unit
IUCN	International Union for Conservation of Nature
LAS	Laser file format
LASER	Light amplification by stimulated emission of radiation
LiDAR	Light detection and ranging
LOOCV	Leave-one-out cross-validation
m.a.s.l	Meters above sea level
NDVI	Normalized Difference Vegetation Index
NLS	the Finnish National Land Survey
R	Statistical program R
RADAR	Radio Detecting and Ranging
RMSE	Root mean square error
RRMSE	Relative root mean square error
RS	Remote sensing
RSS	Residual sum of squares
SAC	Special area of Conservation
SfM	Structure from motion
SIFT	Scale invariant feature transform
UAS	Unmanned aircraft system
VTOL	Vertical take-off and landing
WGS84	Coordinate system WGS84

1. INTRODUCTION

Forests cover approximately 30 percent of global land surface with an area of over 42 million km^2 (Bonan, 2008), and due to this vast quantity of biomass, they hold a significant importance globally. For instance, forests are highly influential actors on the atmosphere and climate change through their role in the global carbon cycle. Forests are also essential for the biosphere through the ecosystem services they deliver and habitats that they offer. Besides these, forests have a role in preventing erosion, securing water supplies and are an important resource of material and food for people.

The link between forests and climate change is in carbon, and especially in carbon dioxide (CO2). Forests store CO2 from the atmosphere and produce oxygen in photosynthesis. It is estimated that globally up to 45% of the terrestrial carbon is stored in forests (Bonan, 2008). On the other hand, vast amounts of carbon dioxide is also emitted to the atmosphere from deforestation. Storing carbon and not releasing it back into the carbon cycle is an essential part in mitigating climate change because deforestation is the second largest anthropogenic source of carbon dioxide to the atmosphere after fossil fuels (Van der Werf, Guido R et al., 2009). Globally forests have previously been known to act as carbon sinks that store more carbon than they emit and they have been thus slowing down the anthropogenic accumulation of CO2 in the atmosphere. The ability to store carbon has been declining due to forest degradation and deforestation. This is where measuring forest structure with different indices using remote sensing data might prove useful as changes in forest structure affect its ability to store carbon.

Besides the atmosphere, forests are also essential to the biosphere. According to the IUCN (2012) forests offer habitats to 80% of the global terrestrial biodiversity and deforestation is causing severe decline in biodiversity. To combat this, attention needs to be directed to the quantity but also to the quality and characteristics of standing forests. Different studies have shown that a forest with more heterogenetic species distribution and size distribution in trees offers more ecological niches and habitats for plants and animals (Fries, Johansson, Pettersson, & Simonsson, 1997; Lexerød & Eid, 2006; Von Gadow & Hui, 2002) and thus, increase biodiversity. Preserving forest biodiversity on the other hand has been shown to be crucial in maintaining ecosystem services that serve human well-being (Balvanera et al., 2006). Forest biodiversity also plays an important role in people's food security, nutrition and health (Arnold, Powell, Shanley, & Sunderland, 2011).

Most of the world's biodiversity resides in tropical forests but it is a topical subject in the northern latitudes too where forestry is a significant part of the economy. Nearly 25% of the total export value of Finland comes from the forest sector and the area under forest management amounts to 88% of the total forest land (Finland's forests 2017.). The intensive management of Finnish forests has succeeded in increasing the growing stock volume of Finnish forests by up to 40% from 1971 (Kortesmaa & Jokela, 2017) which also has increased the amount of carbon stored in Finnish forests making them proficient carbon sinks during this time. However, there has also been some negative repercussions following intense forest management. When forests are shaped to fit the needs of maximal production it is usually done by ensuring that the trees are homogeneous in size which changes the overall structure of a forest to a simpler form. For instance, forests that have been previously managed by clear-cutting, have trees that are mostly the same age class and thus are same sized (Lähde, Laiho, Norokorpi, & Saksa, 1991). This is not the case if forest succession would have happened with no human intervention (Hett & Loucks, 1976). This difference can be clearly seen for instance in a forest structure study by Valbuena et al. (2016) where two very differently managed forest areas were compared from the view point of forest structure. The protected National Park of Koli had significantly more heterogenic-sized trees than an intensively managed forest in the area of Kiihtelysvaara.

In Finland clear-cutting has been the dominant method in forestry but in recent years there has been growing interest in selective logging as an alternative. It would enable a richer forest structure but the method has traditionally seen as less lucrative from an economic perspective though some studies have shown it to be economically competitive with clear-cutting (Lähde, Laiho, & Pukkala, 2010). For the different methods used in forestry like selective logging or clear-cutting, indices describing inequality in tree sizes might prove useful in monitoring the current state of forests and developing the methods further.

The field of remote sensing has progressed rapidly in the last few decades with the advancements in technology and the more efficient handling of the vast quantities of data it produces. The emergence of airborne laser scanning has especially impacted the field profoundly. It offers tremendous amounts of accurate elevation data to many fields from forestry all the way to archeology. Traditional photogrammetric remote sensing methods have also progressed with the

help of more efficient algorithms and more powerful computes solving the three dimensional information from plain photographs.

The main objective of this study was to test the feasibility of digital aerial photogrammetric data in modeling and mapping Gini Coefficient in comparison to airborne laser scanning.

The more specific research questions of this study were:

1. Find out how well the digital aerial photogrammetric data performs in modeling the Gini coefficient
2. Compare the modeling performance of the digital aerial photogrammetric data to the performance of the airborne laser scanning data

2. BACKGROUND

2.1 Forest structure and Gini coefficient

2.1.1 Forest structure and boreal forest characteristics

Forest structure as a concept tries to describe the heterogeneity of a spatially defined forest area. One way to dismantle the concept of forest structure is to think of the general heterogeneity having three different aspects: spatial distribution of trees, species diversity and variations in tree dimensions. These are identified as regularity (spatial), mingling (species) and differentiation (sizes) (Albert, 1999; Pommerening, 2002). These are visualized in Figure 1. Besides these concepts forest structure can also be understood for example via tree age distribution and genetic distribution of trees (Lähde, Laiho, Norokorpi, & Saksa, 1999). For ecological purposes, the differentiation and mingling are the most relevant aspects as they have the biggest effect on biodiversity and ecological habitat creation.

In the boreal forest biome, the species diversity is relatively small compared to forests in warmer climates. According to the Natural Resources Institute of Finland (Parviainen Jari & Västilä

Sinikka, 2012) only four tree species count for 99% of trees in Finland. The Scots pine (*Pinus sylvestris*) counts for 67%, the Norway spruce (*Picea abies*) for 22% and the two birch species (*Betula pendula*, *Betula pubescens*) for 10% of the trees. The last 1% is mostly other broadleaved species. This suggests that in boreal forests mingling might not be as relevant as differentiation in describing the characteristics of a forest.

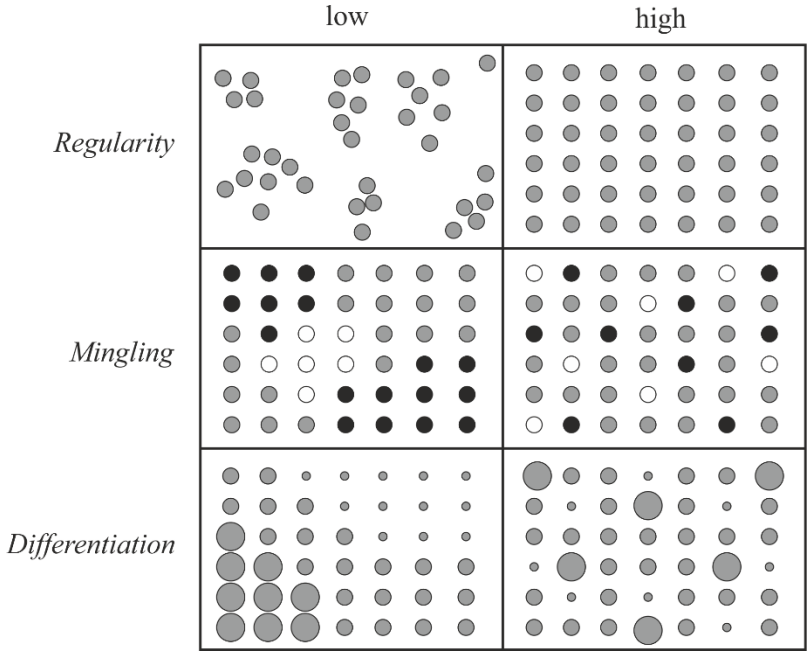

Figure 1. Description matrix of different aspects of forest structure.

From the viewpoint of differentiation, there are several methods for quantifying the concept of forest structure into a single index. A comprehensive evaluation of diameter diversity indices was made by (Lexerød & Eid, 2006) where they tested eight indices on simulated and observed diameter distribution datasets. They tested two types of diameter distribution indices: indices that are influenced by the range of diameters and indices that are influenced by the abundance of various diameter classes. The range-influenced indices were the Margalef index, the Shannon index and the Gini coefficient. The abundance-influenced were the Simpson index, the McIntosh index and the Berger-Parker index. In addition to these, the Shannon evenness and Mcintosh evenness were also tested.

Of these forest structure indices, Lexerød & Eid (2006) found that the Gini coefficient outperformed the others. It was the only index that ranked "Good" in discriminant ability and logical ranking and it also was the least sensitive to sample size.

In addition, there are various other indices suggested in the literature. These include for instance the Clark and Evans aggregation index (Clark & Evans, 1954) and the Winkelmass index (Gadow, Hui, & Albert, 1998). The downside of using these indices is that they require laborious field work as they are distance-dependent indices. This means that the location of every tree is required for the distance calculations between individuals.

2.1.2 Gini coefficient

Gini coefficient (also known as Gini index or Gini ratio) is a statistical dispersion index which was developed by an Italian statistician Corrado Gini who published it in the article "Variability and Mutability" in 1912 (Gini, 1912). Gini coefficient measures the inequality of any frequency distribution. It manifests itself as an index number in the range [0,1] where 0 represents perfect equality and 1 complete inequality. Gini himself presented the index to be used with any kind of statistical distribution dataset (Dorfman, 1979) but it is mostly known for its use in measuring income inequality at society level (Ceriani & Verme, 2012).

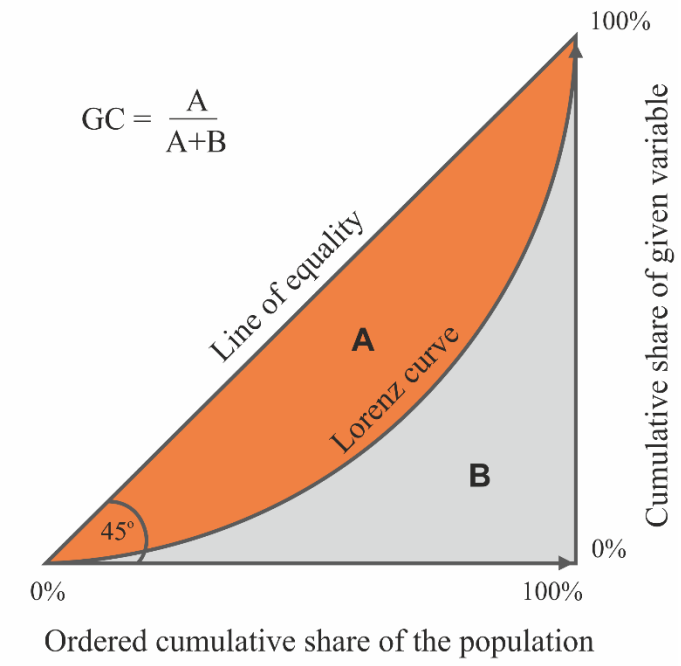

Figure 2. The relationship of the Gini coefficient and the Lorenz curve.

The Gini coefficient is closely related to the Lorenz curve which is a graphical distribution curve used to present cumulative distributions. The Lorenz curve is determined by a curve between the cumulative share of given population (x-axis) and cumulative share of the variable that is under examination (y-axis). A Lorenz curve for perfect equality would exist as a straight line in 45 degrees' angle and any inequality would be seen as a Lorenz curve below the perfect equality line. The Gini coefficient can be then defined as the area between the Lorenz curve and perfect equality line divided with the whole area (Figure 2).

Mathematically the Gini coefficient can be presented with the following equation.

$$GC = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\mu}$$

Which calculates the mean of the difference between every possible pair of individuals and divides it with the mean size μ.

Apart from societal income distributions, Gini coefficient has also been used in natural sciences to some extent. It has been used for instance in determining plant biomass distributions (Weiner, 1985) and as a measure of biodiversity (Wittebolle et al., 2009). And now, in the recent years it has been suggested as a promising index for measuring forest structure (Lexerød & Eid, 2006; Valbuena, Packalén, Martın-Fernández, & Maltamo, 2012).

In the context of forestry, Gini coefficient is used to characterize tree size inequality through the distribution of tree basal areas (areas of single tree cross-sections). In general, it determines how homogeneous or heterogeneous a given forest is in terms of tree sizes. The heterogeneity of tree sizes is related to ecological diversity as different sized trees offer habitats for a larger species abundance increasing biodiversity. Previously ecologists have been using diameter distributions in forest biodiversity estimations on several occasions and several studies suggest that diameter distributions are actually one of the key indicators when assessing forest biodiversity (Noss, 1999; Angelstam & Dönz-Breuss, 2004).

Also, as selective logging is gaining popularity in forest management for economic, ecological and recreational reasons (Lexerød & Eid, 2006), Gini coefficient might emerge as a notable tool for forest management planning and monitoring the long term changes in forest structure. Gini coefficient is not sensitive to variation in sample sizes when compared with other distribution

indices (Lexerød & Eid, 2006) which also suits forestry applications as the number of sample plots vary quite a lot in different forest datasets.

2.2 Three-dimensional remote sensing methods

2.2.1 Digital Aerial Photogrammetry history

Photogrammetry is the science of gathering information on distances and locations from the physical world through geometric analysis of photographs. The concept of photogrammetry can be traced even to the time before photographs when Leonardo Da Vinci and other painters studied the properties of perspective of paintings in the late 1400s (Doyle, 1964). After that during hundreds of years various mathematicians independently researched the area now known as projective geometry. The true breakthrough in photogrammetry did not come until Sebastian Finsterwalder published the principals of modern photogrammetry in the turn of the 19th and 20th centuries. The published papers describe key aspects of photogrammetry and perhaps the most important of them (Finsterwalder & Weinschenk, 1904) really establishes the foundation of photogrammetry by describing the principals of double-image photogrammetry and the geometry involved in it. These principals are still in use.

Aerial photogrammetry on the other hand is a remote sensing method that uses photographs taken from an aircraft to determine elevations of the ground or other objects like vegetation and buildings Aerial photogrammetry has been in operational use all over the world since the early 20th century. In Finland, aerial photogrammetry was the main method of topographic mapping from 1930s forward (Korpela, 2006). It became known as stereophotogrammetry. It was laborious work where a stereo-operator viewed aerial photograph pairs through a stereoplotter and was able to measure elevations from the change of position of a feature in two overlapping pictures. The speed of determining elevations was slow and required highly specialized hardware. Slowly topography mapping digitalized during the 1980s when the concept of digital photogrammetry emerged. (Korpela, 2004; Sarjakoski, 1981). Soon when digital photography started to be the standard image format in photography, it was possible to automatize and digitize the whole process of stereophotogrammetry and reach up to 150 measured points per second (Shears & Allan, 2004).

Due to the advances in photography, digitalization and the increase in computational power, digital aerial photogrammetry is still a viable method for elevation measuring although airborne laser

scanning has outperformed it in the topographic mapping. In 2008, the National Survey of Finland switched digital aerial photogrammetry to airborne laser scanning as the main method of national topographic mapping (Vilhomaa & Laaksonen, 2011). Digital aerial photography is still used and most of Finland is photographed in a five-year interval. The resulting ortophotographs are used for instance in classifying some of the points in the ALS point clouds (Maanmittauslaitos, 2016) and maintaining and updating geographical datasets like the national field plot registry.

As airborne laser scanning is still very expensive, the driving factor for digital aerial photogrammetry over airborne laser scanning is the cost of data acquisition. Digital aerial photogrammetry is considerably less expensive method of retrieving 3D information than airborne laser scanning (White et al., 2016) and the expenses drop even more drastically when Unmanned Aerial Systems (UAS) and modern feature recognition algorithms are used. The downside of UAS is that they are not yet a viable platform for large scale mapping projects due to operational uncertainties and the limitations of battery technology.

2.2.2 Principles of Digital Aerial Photogrammetry and Structure from Motion

Digital Aerial Photogrammetry solves the xyz-coordinates of an object by recognizing the same features from overlapping photographs and applying geometric equations to reconstruct the original geometric space (Figure 3). To achieve this, the original camera geometry (location, alignment, focal length and CCD size) has to be known so that a pinhole camera model can be created (Korpela, 2004). The pinhole camera model is a simplified model of a camera in which all light is assumed to travel through a single point. This simplifies the mathematics related to solving the xyz-coordinates of the features. It is still always a simplification - no camera is ever a true pinhole camera. In traditional digital aerial photogrammetry, the camera alignment and position are usually recorded by using expensive metric cameras designed to record constantly their position and alignment in space. It is also possible to solve the precise alignment and location of the cameras

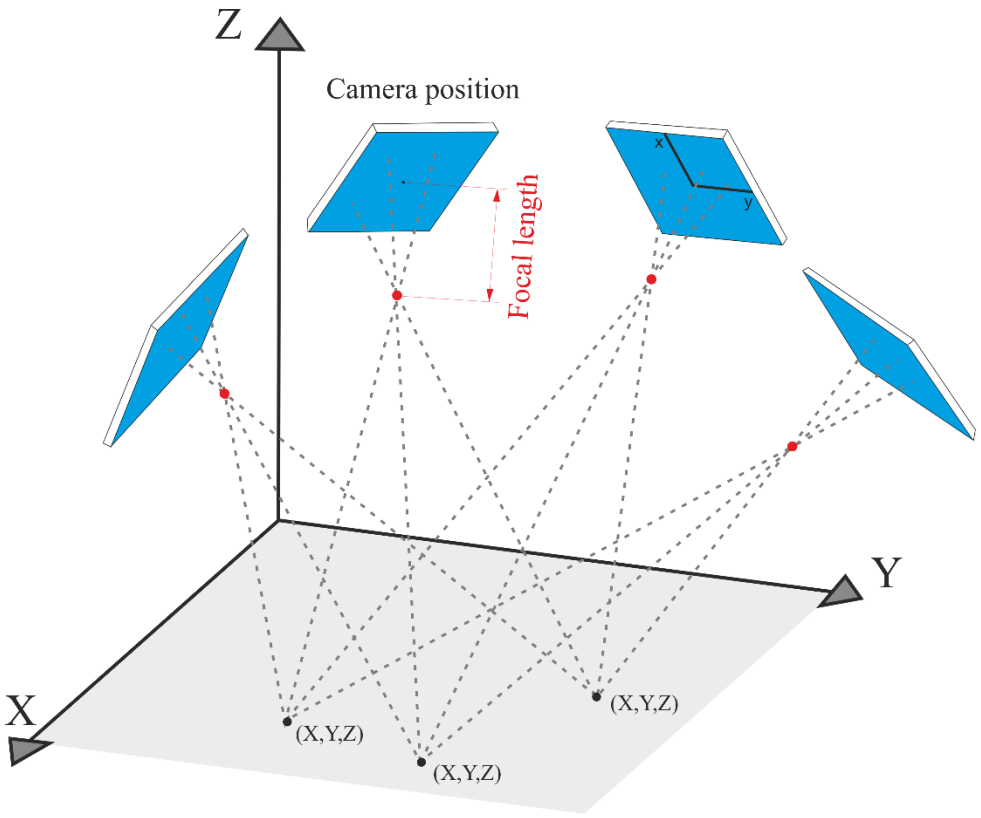

Figure 3. Pinhole camera model and the relationship of coordinates on two dimensional CCD device and the three dimensional space.

automatically through new methods like the Structure from Motion (SfM) method, which is one of the most influential recent innovations in this field of photogrammetric remote sensing.

Structure from Motion is a low-cost photogrammetric method used to derive high resolution three-dimensional information automatically from a series of overlapping photographs using some the same principals of traditional photogrammetry but automating and tuning the process through new algorithms. In the past two decades, this form of automated aerial digital photogrammetry has sustained its niche in three-dimensional topographic modelling (Westoby, Brasington, Glasser, Hambrey, & Reynolds, 2012).

The relevant difference to previous photogrammetric methods is that in SfM, the camera geometries (in essence, the camera locations and angles in space) are not needed to know beforehand but they are rather solved simultaneously with the scene geometry. This makes it cheaper and lighter as simple off-the-shelf cameras can be used rather than heavy and expensive metric cameras that know their orientation in space continuously. This also enables the use of light and unmanned aerial systems rather than utilizing proper airplanes and pilots.

The SfM method is based on iterative bundle adjustment methods (Westoby et al., 2012). Bundle adjustment itself is defined as the problem of estimating the camera locations so that the reprojection errors are minimized (Triggs, McLauchlan, Hartley, & Fitzgibbon, 1999/9/20). It is a highly computational intensive process and many algorithms have been presented, mainly originating from the computer vision community. In general, the different methods all combat the problem by iteratively applying non-linear least-square minimization for the camera position estimates (Westoby et al., 2012) which means that after the initial locations of cameras are determined, the algorithm fine tunes them iteratively by minimizing the projection errors.

The downside of SfM is that the generated point cloud is not in any geographical coordinate system as there are no real-world coordinates involved in the calculations. This can be fixed by using ground control points (GCP) which are for instance some features or markers of which's locations are precisely known in all three dimensions. When this information is included in the point cloud generation in a software like Agisoft Photoscan or Pix4D, the resulting point cloud is georeferenced according to the GCP locations.

2.2.3 Airborne Laser Scanning history

All categories of laser scanning are based on light amplification by stimulated emission of radiation (LASER) technology which was introduced roughly the same time by independent researchers from the US, Canada and Soviet Union in late 1950s (Nelson, 2013). Laser is monochromatic radiation where all waves are in phase and coherent (Shan & Toth, 2008). This gives it the directionality and narrowness it is commonly known for. Laser is produced by optical amplification and it is used in various different applications such as printing, data storage, communication and the most important regarding this thesis - measuring distances.

Laser based distance measuring gained popularity in the remote sensing community and by the 1970s the first profiling lasers were developed (Holopainen, Hyyppä, & Vastaranta, 2013). The first article where profiling laser measurement was utilized in forestry emerged in 1977 (Solodukhin, Zukov, & Mazugin, 1977) and in mid-1980s Nelson et al. (1984) demonstrated the usability of profiling laser scanners in determining forest canopy characteristics.

When positioning systems, inertia measurement units and computational power advanced, it became possible to seize larger amount of three-dimensional information and profiling lasers started to move towards laser scanning where the laser pulses are directed to different directions form the source in a systematic way. When these advanced laser scanners were attached to an aircraft the technology came known as Airborne Laser Scanning (ALS). The first commercial ALS sensor was introduced in 1994 by a Canadian company TopScan (Holopainen, Hyyppä et al. 2013). Since then, the sensors have developed significantly in accuracy and efficiency. For instance, the pulse repetition frequencies have increased by more than 3 orders of magnitude from the first sensor (Nelson 2013). This and improvements in positioning technologies have made it the standard method of estimating tree heights and metrics and it is in operational use all over the world in different forestry institutions.

2.2.4 Principles of Airborne Laser Scanning

Measuring distances with laser relies on the known speed of light in any given medium and the exact measurement of time. Laser ranging devices have an emitter that emits the laser pulse and a receiver that observes the reflected laser pulse. From the time difference between emission and return reflection, the distance is calculated using the following equation.

$$R = v \times t/2$$

Where R is the distance, v the speed of light in air and t the time from emission to receiving the reflection. The speed of light is known very accurately and thus, the accuracy of the distance measurement is always affected more by the measurement of time (Shan & Toth, 2008).

This simple ranging technology can be utilized to produce accurate three-dimensional information in large scales through expanding the quantity of distance measurements and changing the location of the emitter in relation to the object.

In the field of remote sensing, the laser scanner technology used to capture this kind of three-dimensional information is called light detection and ranging (LiDAR) which includes two subcategories: profiling laser ranging and laser scanning. Profiling laser ranging produces a narrow strip of measurements that describe the forests outer profile in a straight line below the aircraft whereas laser scanning produces a wide representation of the 3D structure of the forest beneath the aircraft, often up to 1000 meters in width (Holopainen et al., 2013).

Airborne laser scanning (ALS) is a technology that utilizes the accuracy of LiDAR and the mobility of an aircraft. The method is an active remote sensing method which means that it produces all the energy by itself in the form of a laser pulse and thus does not rely on reflected sunlight like satellite or aerial imagery (Holopainen et al., 2013).

The ALS system is composed of at least the following components: a laser ranging unit, optical scanning mechanism, a positioning system, electronics unit, and software (Figure 4). The laser unit includes the transmitter and receiver that operate the laser pulse emission and return signal observation. The optical scanning mechanism directs the laser pulse to the wanted direction with the help of mirrors or a prism. The positioning system is needed to get the exact location and rotation of the aircraft and laser scanner. This is achieved with a Differential Global Positioning System (dGPS) that is keeping track of the aircrafts location and Inertia Measurement Unit (IMU) that is keeping track of the different rotation parameters (yaw, roll, pitch) of the aircraft. (Shan & Toth, 2008)

The Laser ranging unit emits laser pulses in a very high frequency, ranging from 50 to 400 kHz (Holopainen et al., 2013). When the distance information of the emitted pulses is combined to the exact location of the LiDAR sensor at the time of emission, xyz coordinates can be determined to

every single laser pulse return. The data built from these measured reflection returns is essentially a point cloud where all points have a xyz coordinate and represent a point on the surface of a feature (i.e. leaf, tree trunk or ground) where the laser pulse reflected.

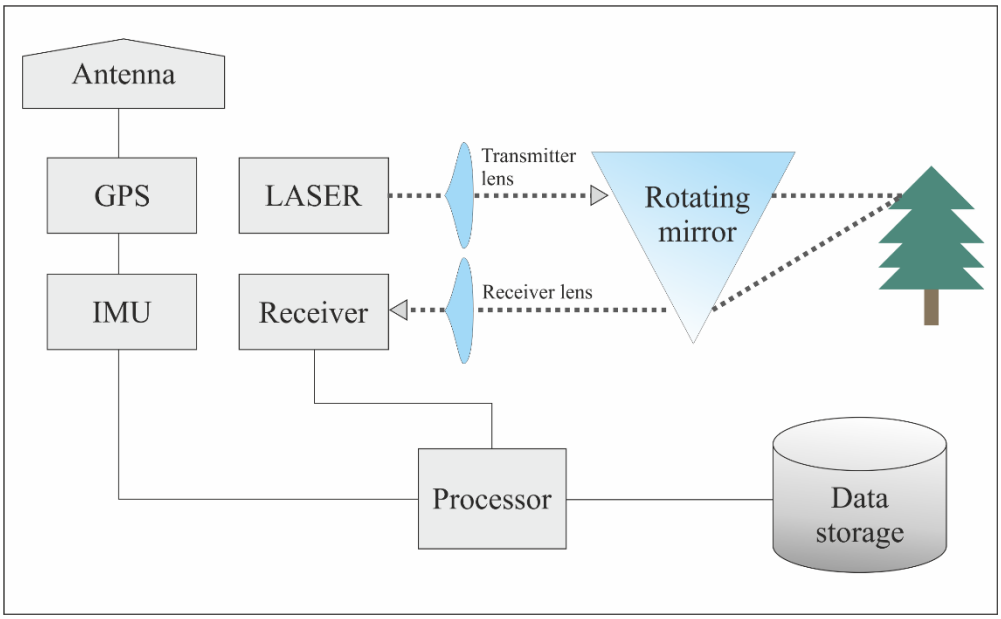

Figure 4. The different systems in a typical ALS setup.

A single laser pulse can have several recorded reflections. This is possible as the laser pulse expands in size the further it travels in space. When the pulse reaches the ground from an aircraft it is usually 0.1-1 meters in size (Holopainen et al., 2013) depending on the sensor used and the flying altitude of the aircraft. As the pulse grows and reaches an object it often hits it only partially. This sends part of the pulse back to the laser ranging unit which records it as a return and the rest of the pulse continues further away until it is reflected and recorded also by the laser ranging unit. The laser ranging unit can record often up to five different returns for a single pulse. This feature is especially useful in forests as it means the laser pulse often infiltrates deep into the forest giving information on the inner parts of the forest in addition to the information of the top of the canopy. The returns of a single pulse are branded by the sequence they return (i.e. 1st return, 2nd return etc.). The bigger the sequence number the deeper from the forest the pulse reflected. If only the last pulse returns are filtered from the data, they are presumed to represent the ground and a digital terrain model (DTM) can thus be created (Holopainen et al., 2013).

2.2.5 Unmanned Aerial Systems

The use of unmanned aerial systems (also known as drones or unmanned aerial vehicles, UAVs) originates from the defense industry where there has always been a need for cost beneficial bird eye's view of the ground. The remote sensing potential of UASs in science was understood over 35 years ago but only in the last decade or two the use of UAS has gained popularity in the remote sensing community (Colomina & Molina, 2014). This has mostly been credited to the recent progression in different technologies related to UAS but some credit also belongs to governments reducing regulation. Perhaps the most significant of which was the US government's decision to release GPS signal for the public undistorted in 2000 enabling accurate positioning data for anyone to use.

Unmanned aerial systems are usually build up from three separate systems: the actual unmanned aerial vehicle (fixed wing aircraft or multicopter), a ground control station and a communication link (Colomina & Molina, 2014). Lower sublevel components often linked with UAS are autopilots, different navigation or imaging sensors and mission planning systems.

Two general types of unmanned vehicles are used: multicopters and fixed wing aircrafts. Multicopters have the advantage of lifting heavier equipment, being easier and more reliable to operate and being capable of vertical lift off and landing. On the other hand, they use significant amounts of energy and have currently operational flight times of only 10-20 minutes. Fixed wing aircrafts use less energy and are capable of 30-60-minute operational flight times. This means they can cover larger distances and areas compared to the multicopters. It is also notable that fixed wing UAS are notoriously difficult to operate and are quite restricted in the amount of payload they can carry before the wing span becomes too large. They also need plenty of space to take off or land where multicopters can land vertically. Recently more and more fixed wing aircrafts have been developed that have the ability of vertical take-off and landing (VTOL) combining the advantages of both vehicle types.

Most sensors used in remote sensing can be attached as a payload to a multicopter and some even to fixed wing aircrafts. Although sensors have gotten smaller and there are UAS versions of hyperspectral cameras and laser scanners, they are still considered as heavy equipment for any UAS and need an expensive platform to be able to fly reasonable flight times. As bigger payload increases costs and the operational difficulty, it is necessary to minimize the payload weight.

Modern off-the-shelf cameras and digital aerial photogrammetry offer a solution to this and have become popular in scientific research as they can produce accurate spectral or photogrammetric data (Colomina & Molina, 2014).

2.3 Remote sensing of forest structure

Through most of the 20th century, gathering information on forest structure relied on technologies like satellite imagery and Radio Detection and Ranging systems (RADAR). Optical satellite sensors like the Landsat TM/ETM+ and SPOT were being used to gather spectral information on forests (Ingram, Dawson, & Whittaker, 2005). This spectral variation was then linked to structural variables like basal area by examining the relationships between structural properties and spectral data. For instance, Brockhaus & Khorram (1992) compared the correlations of the Landsat-TM sensor and the SPOT sensor on basal area and tree age class. In this study, different Landsat TM bands were shown to correlate with basal area with coefficients from -0.27 to -0.48. The group concluded that these correlation coefficients were too low to produce predictive models. Hyyppä et al. (2000) went further and used Landsat and SPOT data to model basal area. They found that the Landsat TM model did not perform well with 0.31 R^2 and 47% of relative RMSE. The SPOT sensor was slightly better with an R^2 of 0.44 and 42% standard error. The study found that spectral variables do correlate with structural features of the forest but not enough to produce accurate models.

Radar sensors have also been used in remote sensing structural features of forests. Most common radar sensors have been the European Remote Sensing satellites ERS-1 and ERS-2 and the Japanese Earth Resources Satellite (JERS). Radar pulses have the advantage of penetrating the canopy and receiving information from the inner forest but according to (Hyyppä et al., 2000), they still perform worse compared to optical sensors on predicting structural features.

In the recent decades LiDAR has outperformed all earlier techniques in the field of remote sensing forest structure with its unprecedented quality and quantity of data. It is capable of gathering three-dimensional data all the way from to the ground to the top of the canopy and everything in between. Variables derived from these three-dimensional point clouds have been shown to significantly correlate with structural features like basal area and models predicting basal area have even reached R^2 values of 0.85 - 0.95 (Holmgren, 2004). For this reason, LiDAR and ALS are nowadays in wide

operational use in estimating growing stocks in the field of forestry (Holopainen et al., 2013). As basal area only indicates the sum of tree sizes divided by area, there has been a need to understand the variation in tree sizes through a single index. This would give a more comprehensive picture of forest structure. Lexerød & Eid (2006) compared different distribution indices on measuring the variation in tree basal areas and concluded that Gini coefficient outperformed other indices. Valbuena et al. (2013) continued from this and incorporated ALS data in modelling GC achieving good results. A year later Valbuena et al. (2014) showed that variation in forest structure can be mapped accurately with ALS in a comparison study between a protected forest area and a managed forest area from the viewpoint of GC.

There has been no previous studies in modeling GC with UAS derived digital aerial photogrammetric data and only a few that model general forest structure variables like basal area. Puliti et al. (2015) mapped small forest patches with an UAS, created point clouds with structure from motion method and modeled different height and volume variables with relatively good results. In addition, Tuominen et al. (2015) gathered a large DAP dataset with a fixed wing UAS and modeled forest variables like volume and basal area. They showed that the accuracies of the UAS-DAP method were in line with a similar ALS-study (Tuominen & Haapanen, 2011) from the same area. They also compared the results of their UAS-DAP study to a DAP study by Järnstedt et al. (2012) that also used the same study area but the data was gathered with a conventional airplane rather than an UAS and from a significantly higher altitude. They found that UAS-DAP method produced significantly more accurate results in comparison to the non-UAS DAP method due to the resolution being a lot better. There are also studies that compare ALS and DAP point clouds from a more general viewpoint of forestry (Vastaranta et al., 2013; J. White et al., 2015). These studies handle similar forest metric variables as this study and they suggest that DAP is a viable and comparable method of retrieving physical forest height variables if there is an accurate digital elevation model that can be used to normalize the tree heights.

3. STUDY AREA

The forests in the vicinity of the Lammi Biological Station were chosen for the study area of this thesis. The research station is located in the municipality of Hämeenlinna near the urbanized area of Lammi (Figure 5). The area is in Southern-Finland, approximately 100 kilometers North from the city of Helsinki, 32 kilometers West from Lahti and 85 kilometers South-East from Tampere.

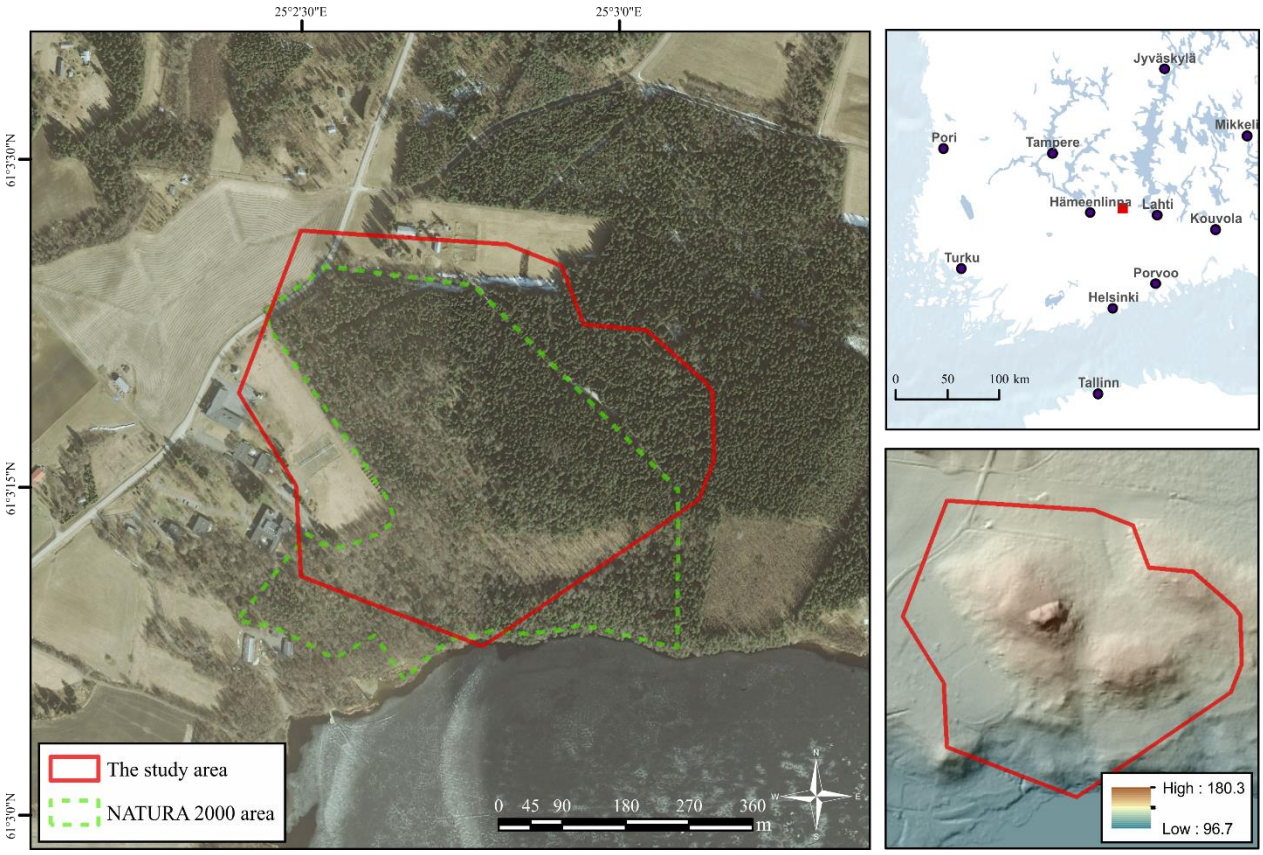

Figure 5. The study area near Lammi research station in Southern Finland and a digital terrain model (DTM) of the study area.

The Lammi area is characterized by both agricultural landscapes and forests. The biological station is situated on the shore of Lake Pääjärvi in the western part of the lake. The study area itself is located east of the Biological station's buildings, on the hill of Linnamäki and the areas between Linnamäki and Lake Pääjärvi (Figure 5). The irregular shape of the study area is a repercussion of the extent of the Digital Aerial Photography (DAP) data that was acquired.

The topography in the area ranges from the surface of the lake 96.7 meters above sea level to the top of Linnamäki which resides 180.3 meters above sea level (Figure 5. DEM map). Thus, the relative elevation difference is approximately 84 meters. In total, the study area is 27 hectares in size. The forest is mostly defined as a Special Area of Conservation (SAC) in the European Union's NATURA 2000 nature protection area network (Ympäristöministeriö, 2015).

The study area was divided to three different arbitrary categories based on tree species. This was done so that stratified random sampling could be implemented. The eastern parts are dominated strongly by the Norway spruce (*Picea abies)*, the middle parts are mostly mixed forest composed of Silver birch (*Betula Pendula),* Norway spruce (*Picea abies)* and Scots pine (*Pinus sylvestris)* and the western parts are characterized with deciduous species like the Silver birch (*Betula Pendula)* and the Common aspen (*Populus tremula).* The forest also contains small numbers of the Common juniper (*Juniperus communis),* pedunculated oak (*Quercus robur*) and Alders (*Alnus sp.).*

The station and the forests have been under the control of University of Helsinki since 1953 and before that they were managed by the vicar of Lammi who held his residence there (Lammin biologinen asema, 2016). So in essence, the study area forest has been protected and unmanaged at least 63 years and probably a lot more.

4. MATERIAL

4.1 Field measurements

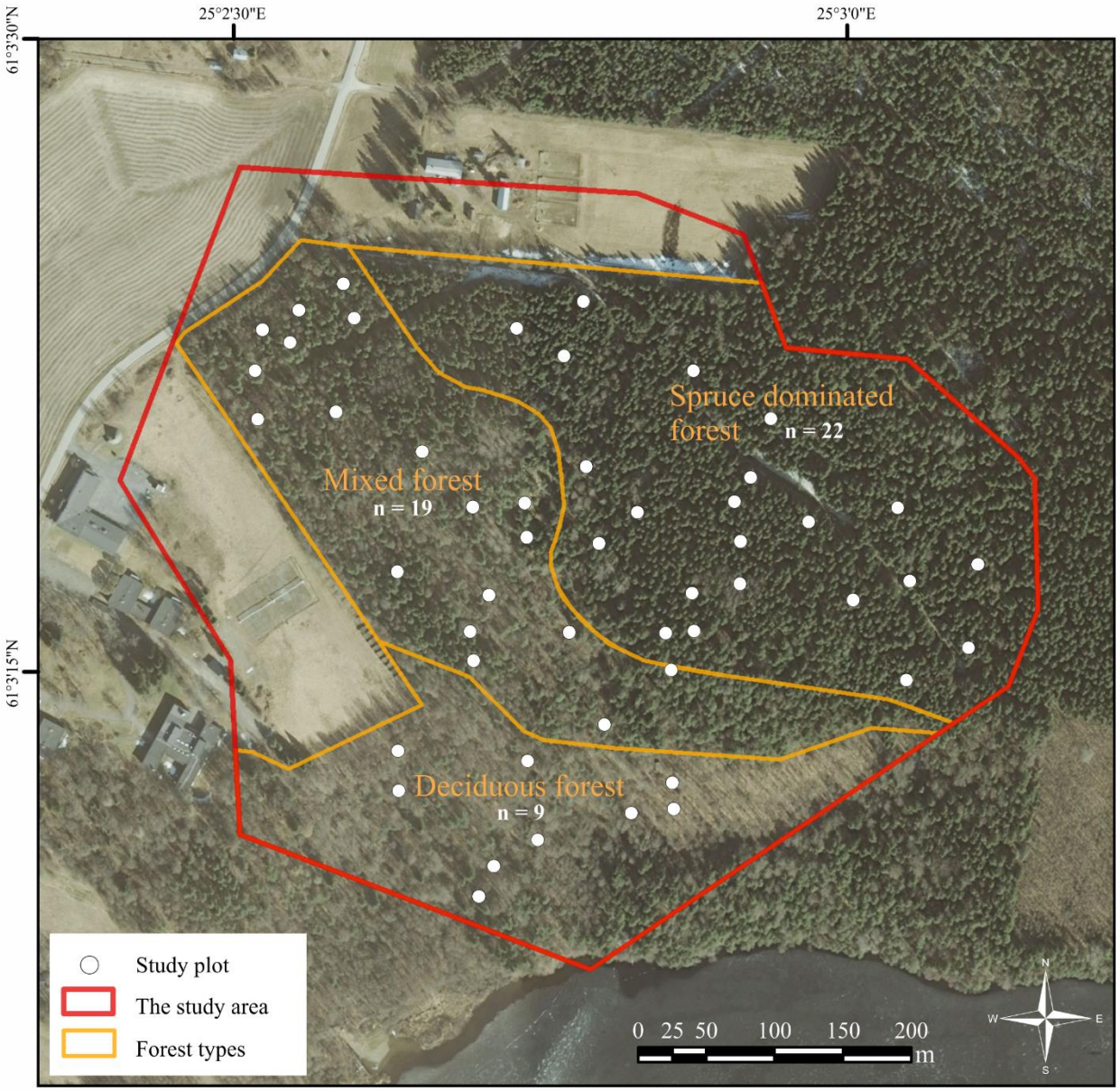

Figure 6. The locations of study plots and forest types in the study area.

The field data was gathered in June 2016. It consisted of 50 circular forest plots (Figure 6) with a radius of 5 meters. The plot locations were defined with stratified random sampling with the following method. The study area was divided into three forest types: spruce dominated forest,

mixed forest and deciduous forest. Random sampling was conducted for each of the forest types so that the amount of forest plots was in line with the forest type's surface area. The intention was to get a sample where all forest types would have representation. The original forest plot amount was 73 but after applying a 2-meter threshold for horizontal accuracy, 23 of them had to be excluded. The horizontal accuracies for the forest plot centers are given in Figure 7.

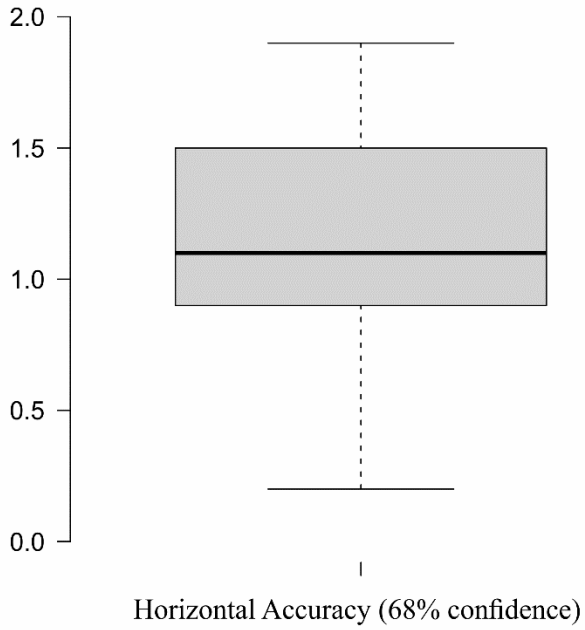

Figure 7. The positioning accuracies of the 50 study plots.

The centers of the plots were measured with Global Navigation Satellite System (GNSS) device Trimble GeoExplorer GeoXH 6000 and the measurements were averaged minimum of 30 seconds.

From these plots, circumferences of all trees in 5-meter radius from the center were measured if the circumference was over 15,7cm which indicated they had a diameter at breast height (DBH) of over five centimeters.

4.2 Flight campaign

The flight campaign was conducted on the 22th of June 2016 in a partly cloudy, partly sunny weather. Two flights were flown with a DJI 900 hexacopter which was operated by the open-source controller ArduPilot and the autopilot Pixhawk (Figure 8). The missions were planned in the software MissionPlanner and they consisted of two 8-10 minute flights. The copter was manually lifted and then switched to autopilot that maneuvered the mission and eventually returned the hexacopter to the launch site where it was manually landed. The altitude was defined as 149 meters from the site of operation and horizontal speed was set to 6 meters per second. The software calculated the optimal locations of photographs to fulfill the 90-percentage forward overlap and 75-percentage sideways overlap between pictures. During the flight, the autopilot operated the camera and took pictures on the optimal locations which software MissionPlanner had previously optimized.

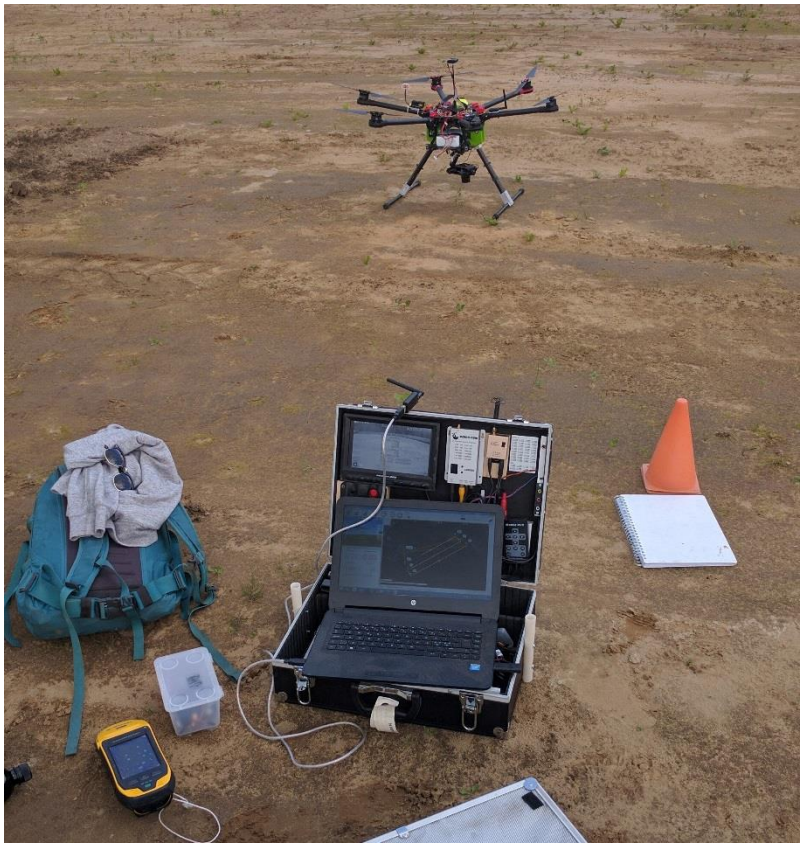

Figure 8. The DJI 900 hexacopter and the ground control station used in this study.

The camera settings for the Sony RX100 III camera were the following. The focal length was set to minimum of 8.8mm, focus to infinite, exposure time to 1/1000 second, aperture to 5.6 and ISO value to 400. The best settings for the lighting conditions were determined by sample pictures taken on ground level before the flight.

Before the flight missions a set of GCPs (Ground Control Points) were placed on the extent of the mission and their locations measured as accurately as possible with a Trimble GeoExplorer 6000 GNSS system. The GCPs were placed on open areas that were assessed being observable from the viewpoint of the copter on its flight lines. An individual GCP was formed by an orange marking cone on top of a white paper (Figure 9). Both flight missions had 6 GCPs thus 12 GCPs in total for the study area.

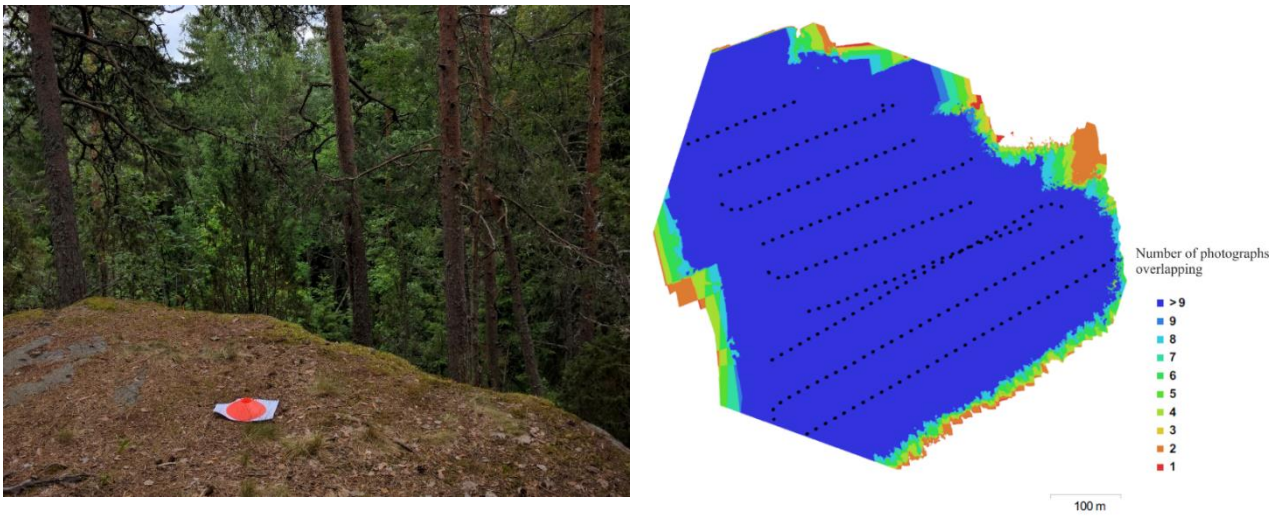

Figure 9. An individual ground control point (GCP) and amount of overlap in different photographs in the study area.

4.3 Digital Aerial Photogrammetry data

The flight campaign resulted in 447 usable images. These images were processed into three-dimensional point clouds (Figure 10) in digital photogrammetric software with methods described in the Methods chapter. The final data was a point cloud with a point count of 163 397 843 points. This amounted to a point density of 604.26 points per square meter. The data was stored in LAS

format version 1.4. The elevations ranged from 96.6 meters to 193.2 meters above sea level. Detailed parameters of the dataset are given in Table 1.

Table 1. The attributes of the digital photogrammetric dataset.

Parameter	Value
Point density (points/m^2)	604.26
Number of points	163 397 843
Acquisition day	22.6.2016
Number of images	447
Maximum flying altitude (m)	149
Average flying altitude (m)	125
Average ground resolution (cm/pix)	3.16
Camera	Sony RX-100 III
Image sensor (CMOS) size (mm)	13.2 x 8.8
Pixel size on sensor (µm)	2.41 x 2.41
Number of pixels (millions of pix)	20
Picture resolution (pix)	5472 x 3648
Focal length (mm)	8.8

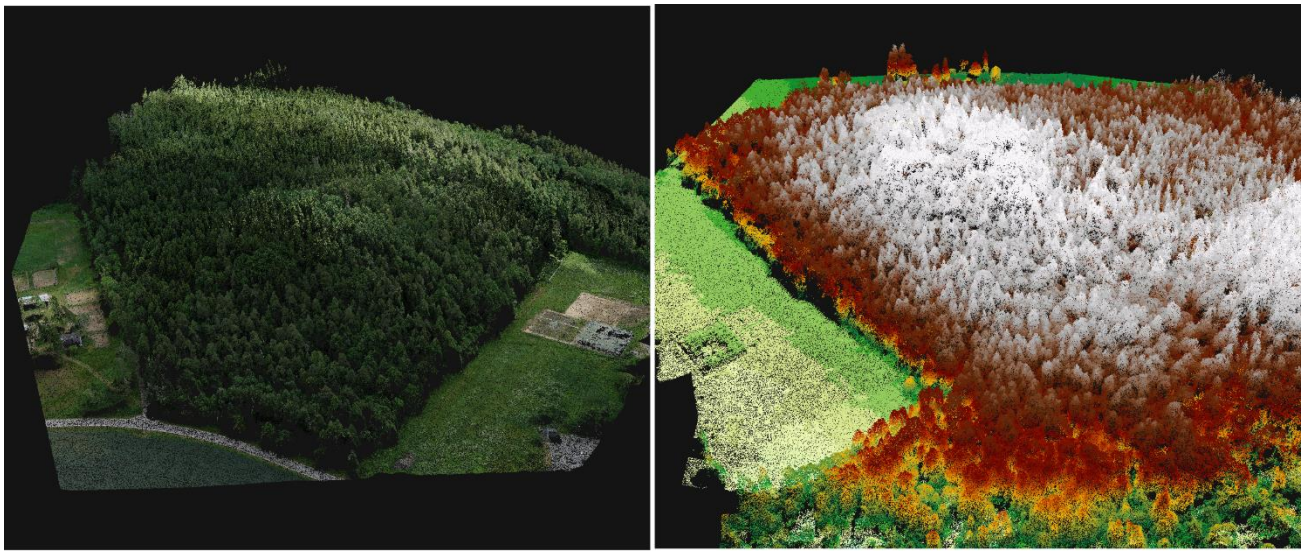

Figure 10. DAP point clouds classified by RGB values and height values.

4.4 Airborne Laser Scanning data

The ALS data was acquired and downloaded from the National Land Survey (NLS) download service. Two different flight lines overlapped the study area resulting in two different downloadable files with the following IDs: M4133F1 and M4133F3. After pre-processing the ALS point cloud of the study area had a point density of 1.48 points per square meter and for last returns, 1.05 points per square meter. In total, there were 3 98 599 points in the point cloud which in elevation ranged from 103.6 meters to 187.0 meters above sea level (Figure 11). More detailed parameters are given in Table 2.

Table 2. The attributes of the airborne laser scanning dataset.

Parameter	Value
Point density (returns/m^2)	1.48
Last return density	1.05
Acquisition day	7.5.2012
Number of points	398599
Return 1 count	281994
Return 2 count	98960
Return 3 count	16659
Return 4 count	986
Altitude (m)	1830
Scan angle (degrees)	+-20
Pulse footprint diameter (cm)	60
Coordinate system	ETRS-TM35FIN

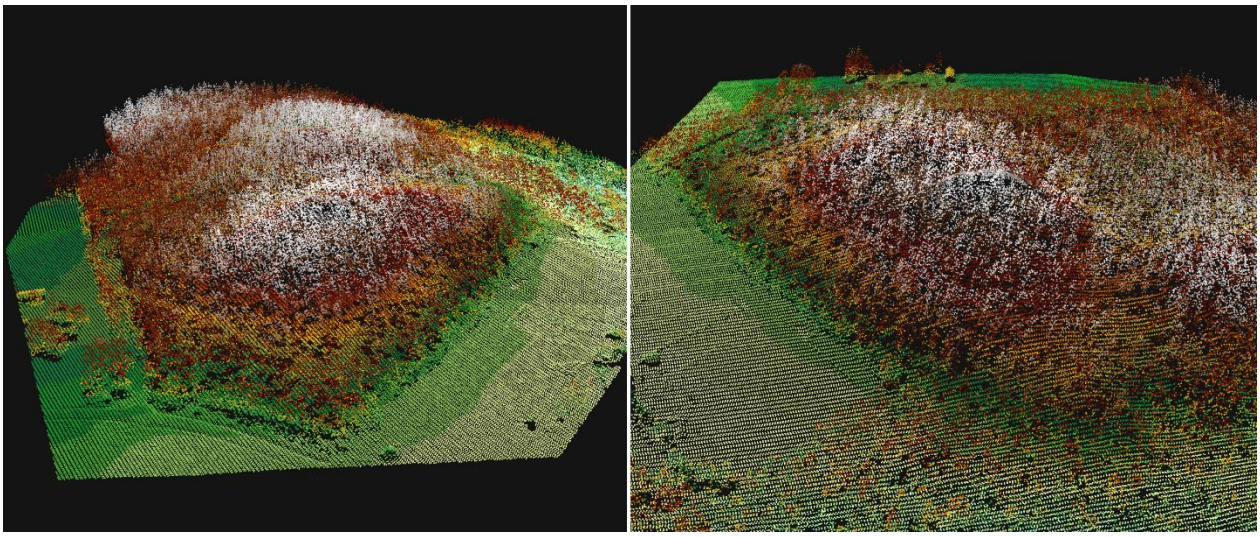

Figure 11. The airborne laser scanning point cloud visualized by height values.

5. METHODS

5.1 Overview

Methods are described comprehensively in Figure 12. There are three sources of data: DAP data, ALS data, and Field data. Different kind of pre-processing steps were conducted to both RS datasets but after producing the final point clouds, the method pipelines converged and steps were identical. Each step of the Figure 12 is portrayed in more detail in the following chapters.

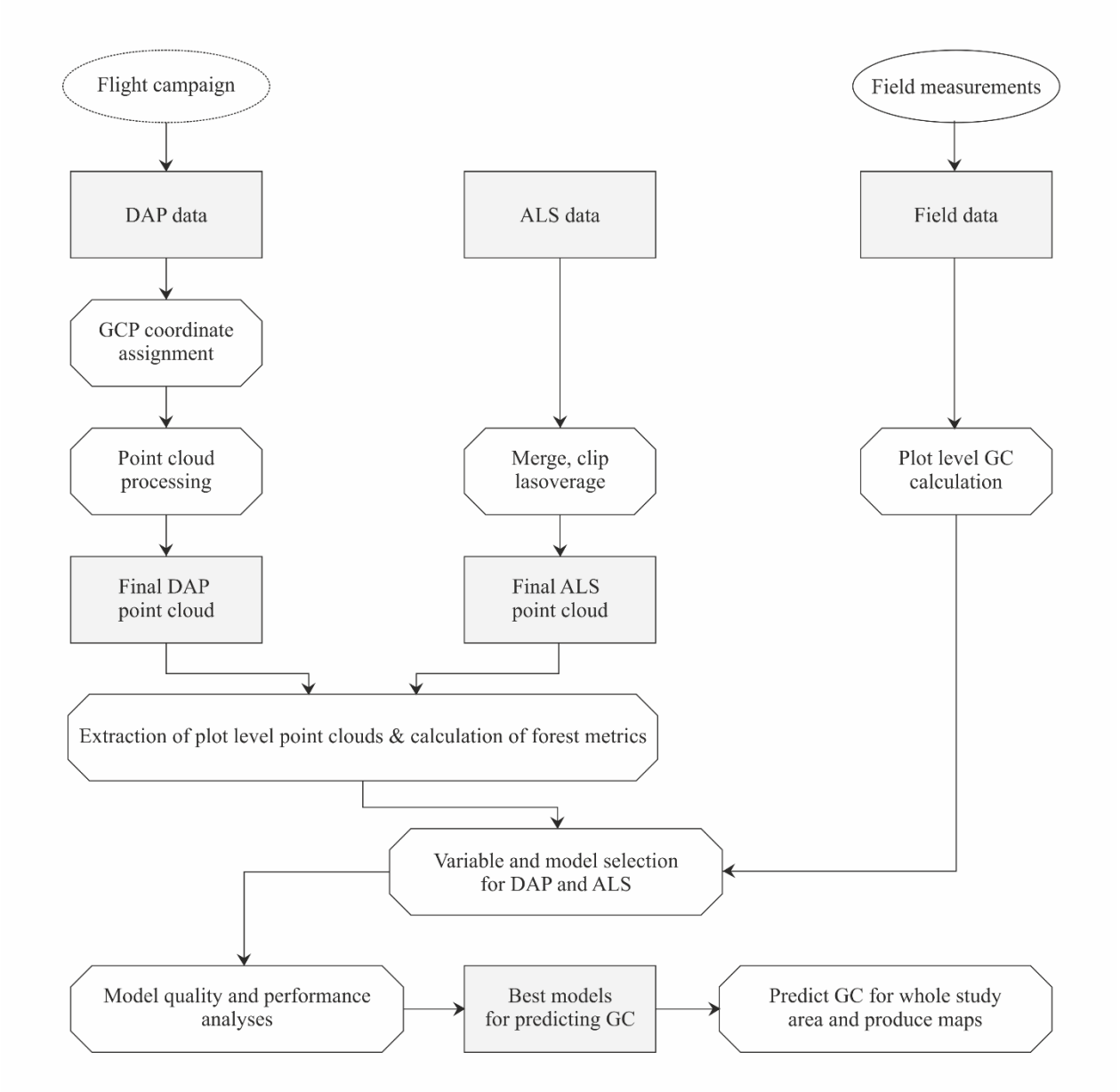

Figure 12. Overview of datasets and methods used in this study.

5.2 Gini coefficient calculation

The Gini coefficient was calculated to all the study plots by a custom R script. First the basal area of all trees was calculated from the measured circumference of the trees with the simple equation

$$BA = \pi \left(\frac{C}{2\pi}\right)^2$$

Where *C* is the circumference of the tree.

Then this list of basal areas for an individual plot was used as an input for a function in the R script that calculated a Gini coefficient value for the plot. This function was then iterated for all the plots.

As GC calculations were applied to a finite sample size *n*, finite sample correction was also implemented in the function to correct the finite population bias.

$$\frac{n}{n-1}$$

5.3 Photograph processing and point cloud generation

In the pre-processing phase, a handful of pictures were discarded on the premise of them not being perpendicular to the ground. These discarded photographs were all taken from the ends of flight lines when the copter was turning around its yaw axis. This movement together with wind produced photographs not perpendicular with the ground, in some cases even the horizon was visible in the photographs. The locations of the discarded photographs in the end of flight lines can be seen in Figure 9.

After visual inspection and filtering the photographs, the GCPs were manually located from the remaining 447 photographs in the software Agisoft Photoscan and the correct coordinates assigned to those pixels where a center of a GCP was seen. All 12 GCPs were visible in at least five photographs, most of them in 10-20 photographs with multiple GCPs visible at the same time.

Generating a point cloud from photographs in Agisoft Photoscan includes two very distinct steps. First a sparse point cloud is generated from the points in the overlapping photographs. During this step, the program tries to recognize common objects from different photographs using an algorithm similar to the scale-invariant feature transform (SIFT) algorithm (Lowe, 1999). The recognized

features can be for instance edges of objects or distinct shapes that are visible from different viewpoints. The xyz coordinates and the precise alignments of the cameras on the time of triggering are also determined.

After sparse point cloud generation, a dense point cloud is generated. The dense point cloud generation does not use the sparse point cloud, rather it utilizes the camera alignments and locations determined in the sparse point cloud step (AgiSoft, 2011) and reproduces a point cloud with significantly higher point density. The dense point cloud algorithm, like all the other Photoscan products is not open source but it is known to use pair-wise depth map calculation that can measure the depth of every pixel in a photograph if the locations of cameras are known.

The sparse point cloud generation parameters were set so that accuracy was high, pair selection generic, key point limit 120 000 and tie point limit 4000. The dense point cloud generation was processed with the following reconstruction parameters: quality was set to high and depth filtering to moderate.

5.4 Pre-processing ALS data

The ALS point clouds have been pre-processed by the National Land Survey of Finland. This includes an automatic classification procedure where the point cloud is classified to different classes and error returns are deleted (Maanmittauslaitos, 2016). The data classes are the following: unclassified, overlap, low vegetation, low point and ground.

After automatic pre-processing, the point cloud is then manually classified and inspected in a graphical working environment using stereo models from aerial imagery. This phase corrects errors made in the automatic pre-processing, creates standing water and streaming water classes and extracts bridges from water classes. (Maanmittauslaitos, 2016)

In this study, it was necessary to process the ALS point cloud further due to the area being in the middle of two adjacent flight lines. This meant the point density varied a lot and the points in the overlapping area had large scan angles. To combat this, the overlapping effect on the data was removed with the lasoverage tool in the software LAStools (Isenburg, 2016) which filters overlapping points based on minimum scan angle.

5.5 Point cloud variable extraction

In order to model the Gini coefficient, point cloud variables were derived for the study plots from the ALS and DAP point clouds. The procedure was done in FUSION software (McGaughey, 2016). All the study plots were clipped from both the ALS and DAP point clouds with FUSION's ClipData tool. The commands were dynamically created to automatize the procedure. A bounding box for the coordinates of the center point (10 meters x 10 meters) was determined and the ClipData tool configured so that it treated the study plot as a circle inside the bounding box instead of a rectangle. After clipping all the study plots, forest metric calculations were applied to them with FUSION's CloudMetrics tool. It iterated through all the clipped files with the file extension LAS and calculated forest metrics for all of them which then were merged into a single CSV file (Comma Separated Value).

From the datasets, 48 different point cloud variables that CloudMetrics calculated were taken into account on variable selection. The ALS data can have several laser pulse returns but DAP data does not due to the differences in these technologies. For this reason, the distinct return counts for the ALS data were not included in variable selection. The variables are listed in appendix 1.

5.6 Variable selection and quality assessment

Automatic variable selection was utilized in the form of regsubsets function of the leaps package in the statistical program R (Team, 2000). Regsubsets is an exhaustive variable selection function for linear regression that uses forward, backward or sequential replacement methods to find out the most important variables that explain the response variable. The criteria regsubsets uses to assess the models include for instance Akaike Information Criterion (AIC), Bayesian information criterion (BIC), residual sum of squares (RSS), and Mallows' Cp (CP). In addition, pseudo coefficient of determination (R^2) and the adjusted pseudo coefficient of determination (adjR^2) were calculated for the models.

In line with previous studies with Gini coefficient prediction from forest metrics (Valbuena et al., 2013; Valbuena et al., 2013; Valbuena et al., 2016), a maximum number of variables for the model was determined to be four to avoid overfitting the model.

A Leave-one-out cross-validation (LOOCV) method was also used to assess the models in terms of overfitting. LOOCV measures the model's performance when one of the plots is temporary left out of the data. The amount of error from the excluded data point to the model is noted and the similar procedure is applied to every plot. Then the mean of errors is calculated and models can then be compared. LOOCV was chosen as the field measurement data was limited in size and removing a separate validation data from the field data would have sliced the data set too much in size. LOOCV was the final criteria used to choose the best model alongside with adjusted pseudo coefficient of determination.

5.7 Beta regression

For the regression analysis, beta regression (Ferrari & Cribari-Neto, 2004) was chosen. It is a generalized linear model designed to be used when the response variable is a continuous proportion or an index with the range of [0,1] – like the Gini coefficient. It is described as a regression model where the dependent variable is beta-distributed (hence the name) and where the mean relates to the explanatory variables with coefficients and a link function (Cribari-Neto & Zeileis, 2009).

In practice beta regression was implemented to the Gini coefficient modeling via an R package called betareg (Cribari-Neto & Zeileis, 2009). Like in the previous studies, a logit link function was chosen as the link function to keep the predicted response variables within the theoretical limits of [0,1] (Valbuena et al., 2013).

5.8 Producing Gini coefficient maps

The Gini coefficient models were generalized to the whole study area to see how the models would predict GC on the Lammi area. The study area was divided into 10m by 10m grid and all forest variables were calculated to each grid cell from the both DAP and ALS datasets. This was done in FUSION software by applying the GridMetrics function. The DAP dataset had to be split in two and the GridMetrics function applied to both datasets due to issues related to insufficient memory on the workstation. This was done with PolyClipData tool.

The resulting tables were imported into R and the predict function of the betareg package was applied to the datasets. The best models for both DAP and ALS were used as the model input and the forest metrics for the 10m by 10m grid as the data input.

The results were converted to raster format with the help of the raster package. Finally, the results were imported to ArcGIS 10.3.1 and clipped to the forest extent. Raster statistics were calculated and the maps visualized.

6. RESULTS

6.1 Modelling results

6.1.1 Variable selection

Models chosen to predict Gini coefficient (GC) from both the DAP and ALS based data are presented in Table 3.

The equation is

$$GC = \beta_0 + \beta_1 V_1 + \beta_2 V_2 + \beta_3 V_3 + \beta_4 V_4$$

where β represents the specific model coefficients and V the different variables. The DAP model had maximum elevation, standard deviation of elevation, elevation skewness and the 10th elevation percentile as variables. The relationships of the chosen variables and the observed GC is shown on Figure 13. The ALS model had maximum elevation, variance of elevation, the 30th percentile of heights and the proportion of all returns over 1 meter as variables. The relationships are shown on Figure 14.

Table 3. The final GC models for DAP and ALS datasets.

	β_0	β_1	V_1	β_2	V_2	β_3	V_3	β_4	V_4
GC_{DAP}	2.669***	-0.299	Elev.maximum***	0.691	Elev.stddev**	0.498	Elev.skewness**	0.234	Elev.P10***
GC_{ALS}	-3.646**	-0.118	Elev.maximum**	0.025	Elev.variance***	-0.122	Elev.P30***	0.075	Cover***

Codes for statistical significance: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

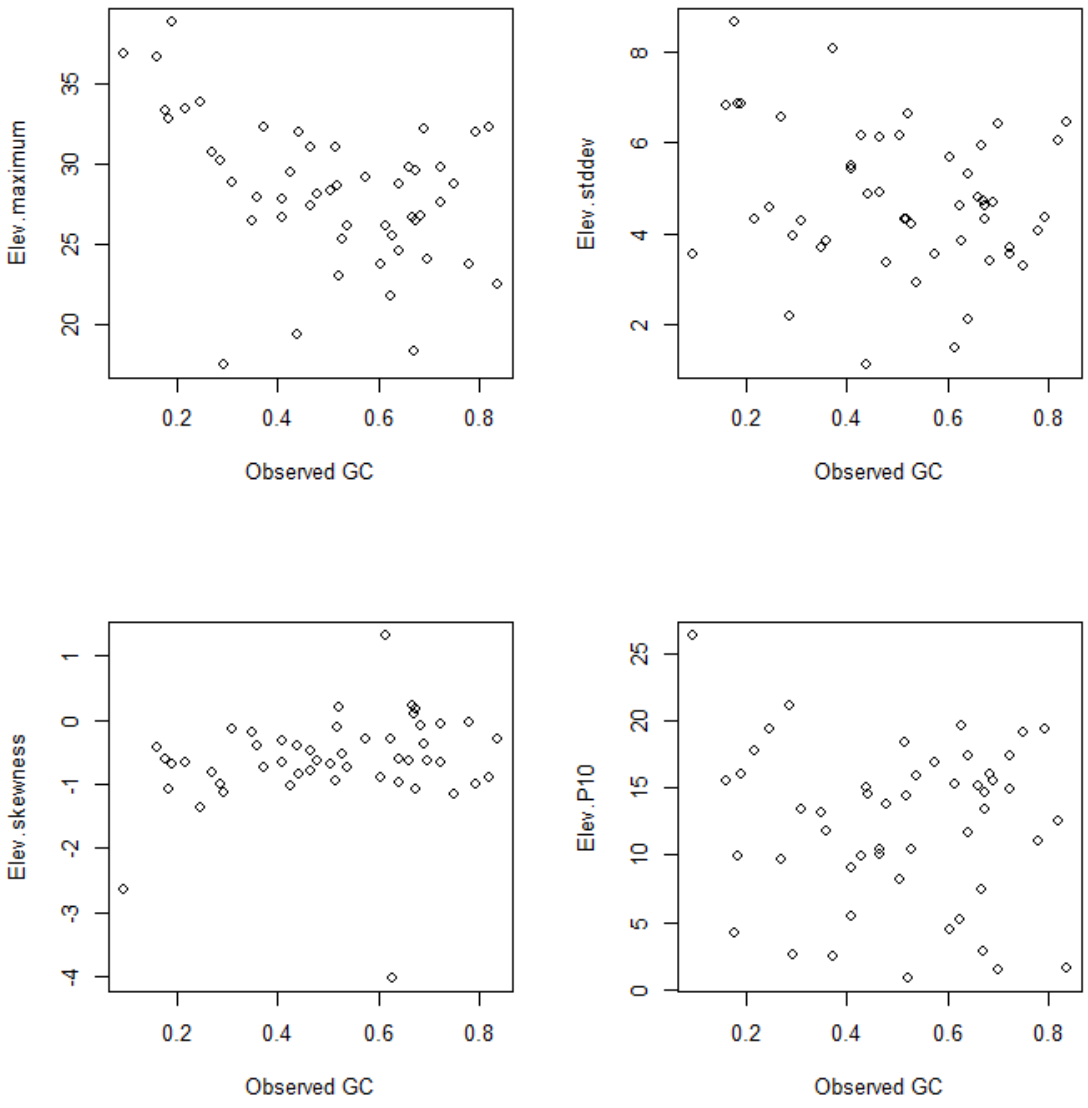

Figure 13. The DAP model variable relationships with observed GC.

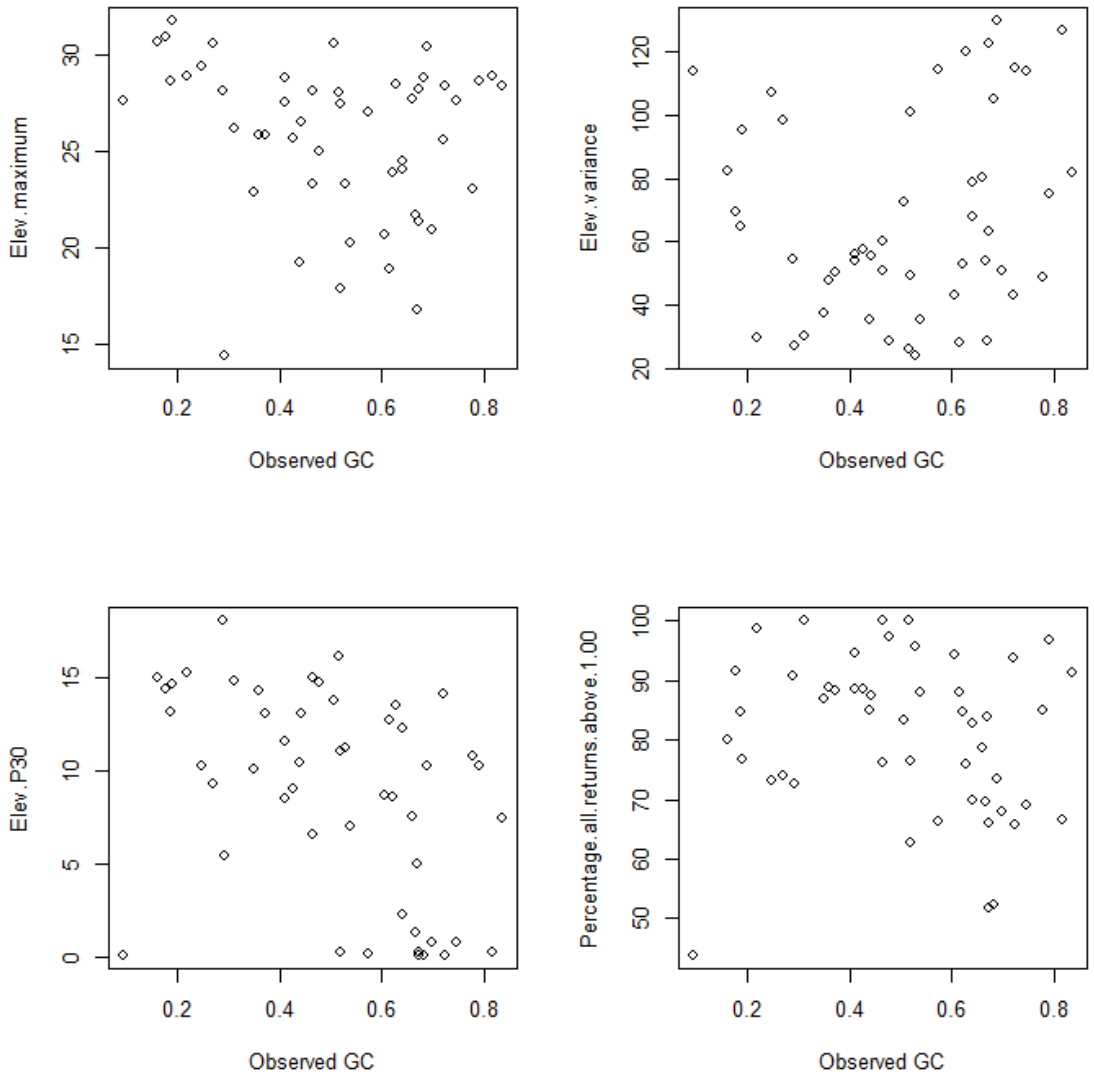

Figure 14. The ALS model variable relationships with observed GC.

6.1.2 Model performance

In general, both models performed only in a mediocre manner. The DAP based model did not perform as well as the best ALS based model in the accuracy indicators (Table 4). The DAP based model had a smaller coefficient of determination (0.39) than the ALS based model (0.49). Also, the residual sum of squares was larger in the DAP model (1.14) than in the ALS model (0.95). The

cross validated RRMSE was similar in both, capping just under 30%. LOOCV bias was slightly higher in the ALS model. AIC and BIC values were better for the ALS model.

Table 4. The model performance measures for ALS and DAP models.

	R^2	RSS	adj R^2	BIC	AIC	CV BIAS	CV BIAS R	CV RMSE	CV RMSE R
GC_{ALS}	0.494	0.951	0.448	-14.452	-48.092	-0.005	-1.020	0.138	27.200
GC_{DAP}	0.392	1.142	0.338	-5.295	-39.029	-0.003	-0.606	0.151	29.800

The difference in performance between GC_{DAP} and GC_{ALS} can also be seen in the Figure 15 and 16 where predicted Gini coefficient values are put against observed values. ALS model predictions are in tighter formation.

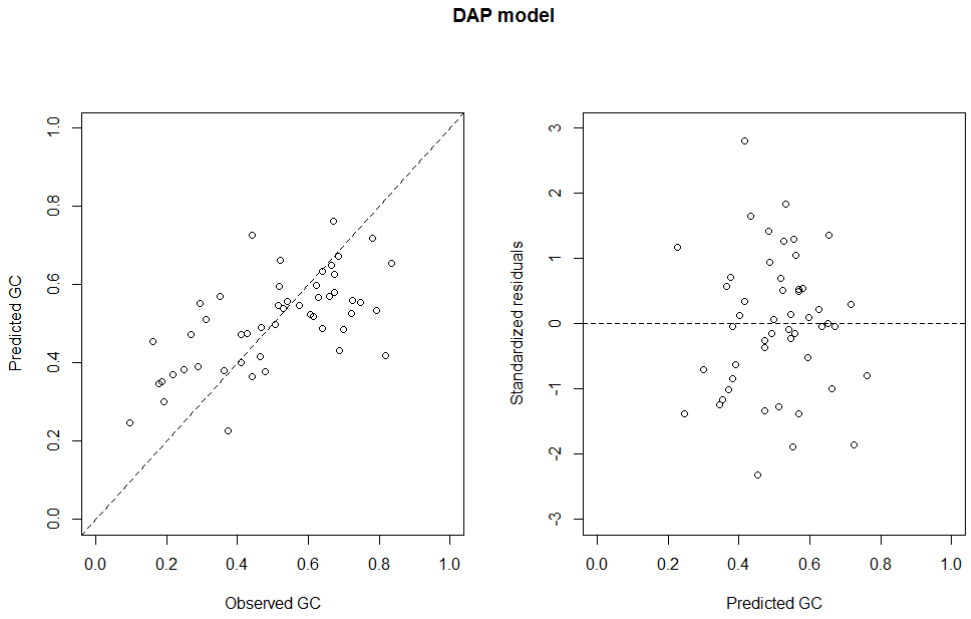

Figure 15. The predicted and observed GC values for the DAP model and the variation in standardized residuals.

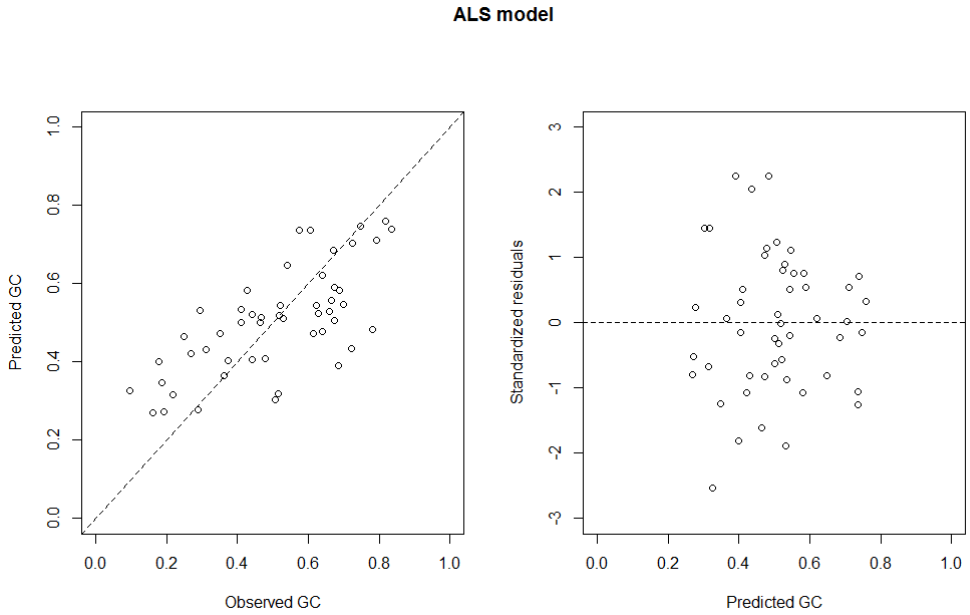

Figure 16. The predicted and observed GC values for the ALS model and the variation in standardized residuals.

6.2 Gini coefficient maps

The ALS and DAP models were used to predict Gini coefficient to the whole study area. This produced two maps (Figure 18, A and B) corresponding the two models. The DAP dataset based GC raster has a range of 0 to 0.98, median 0.50, a mean of 0.49 and standard deviation of 0.15 (Figure 17). The ALS dataset based GC raster has a range of 0 to 0.79, median of 0.40, a mean of 0.37 and standard deviation of 0.18 (Figure 17).

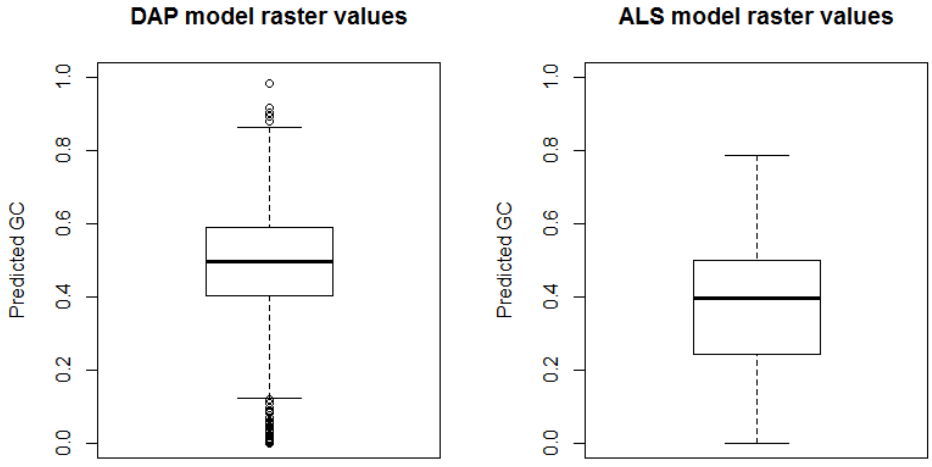

Figure 17. The variation in the raster cell GC values.

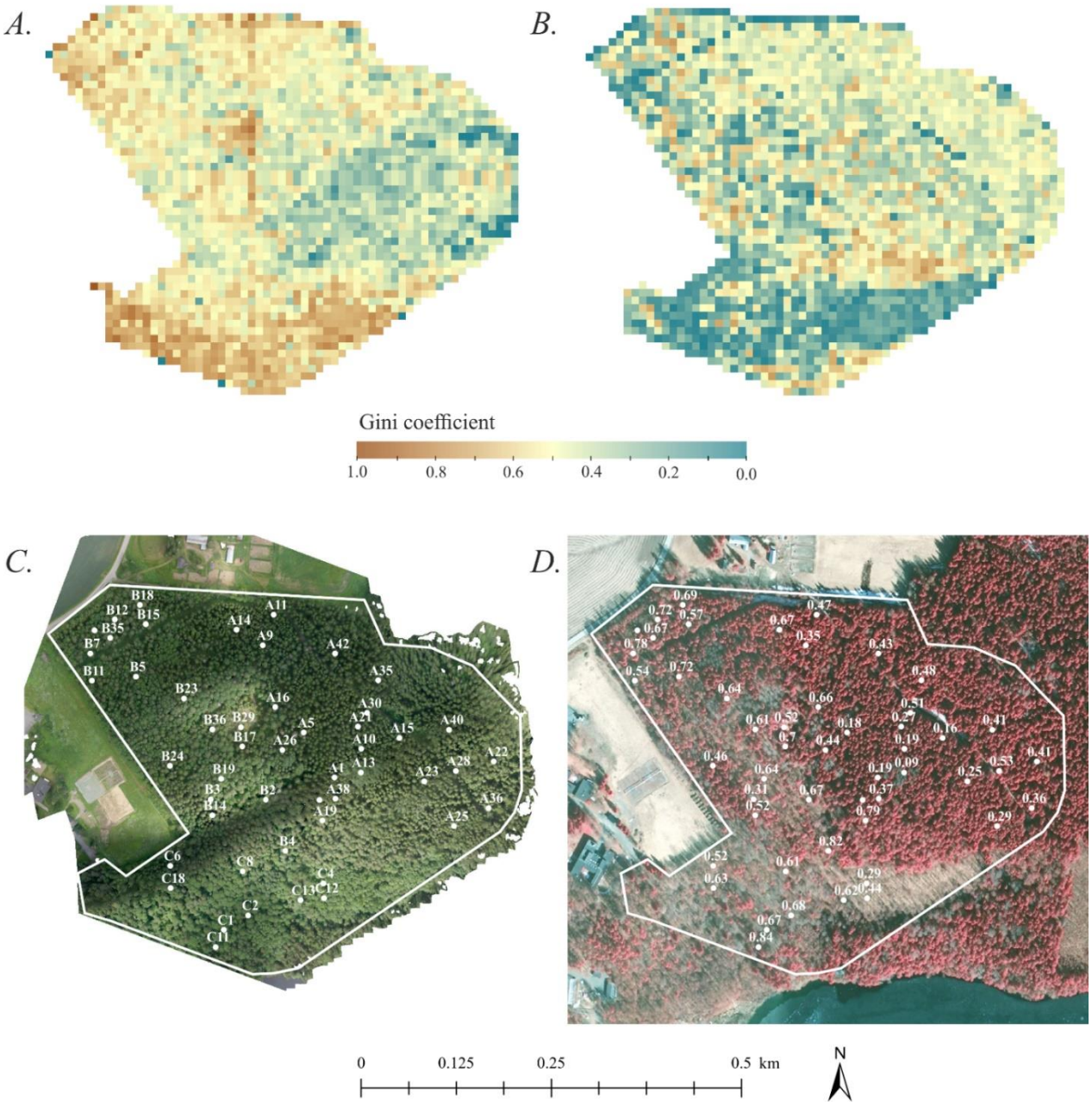

Figure 18. The results of GC generalization to the whole study area. The DAP GC map (panel A), the ALS GC map (panel B), the orthoimage and locations of study plots (panel C) and a infrared aerial image and the measured GC values (panel D).

Spatially, low GC areas on the DAP map (Figure 18A) are situated especially in the eastern and middle parts of the forest, mainly the spruce dominated part of the forest. On the ALS map (Figure

18B) low GC areas are located in the southern parts of the area, mainly in the deciduous forest. Also, the Oak forest pasture in south-east parts is clearly a low GC area.

In the DAP map, high GC areas are spatially spread to all the parts of the study area. The Southern broadleaved part of the map is clearly a high GC area. Besides that, the highest part of the area – Linnamäki – is also distinguishable as a high GC area. Other smaller high GC clusters exist in the northern parts and various locations all around the map. In the ALS map, high GC areas are not as abundant. They are located in small patches all around the study area with a bigger cluster north and south from the oak forest pasture.

The electric wiring (which can be seen in auxiliary maps C and D, Figure 18) is clearly visible in the ALS map (Figure 18B) in the north-western part of the area. Also, pieces of the dirt road are visible in the eastern part of the map. Both instances can be identified as blue low GC stripes.

7. DISCUSSION

7.1 Accuracy of GC modelling using DAP and ALS point cloud variables

The first objective of this study was to find out how well the digital aerial photogrammetric data performs in modeling Gini coefficient. There are no previous studies on the subject and only a few that handle forest variable modelling in general from UAS-DAP data. In one of these studies, Tuominen et al. (2015) modelled height, volume and basal area variables from DAP data. Their RRMSE values were in line with previous ALS studies and basal area RRMSE (20-30%) in line or slightly better than the DAP RRMSE values of this study (30%). In another study, Puliti et al. (2015) implemented a very similar study setting than used in this study for estimating general forest variables. They achieved RRMSE value of 15.38% for basal area predictions.

Other studies like Kachamba et al. (2016) have estimated above ground biomass in tropical woodlands using UAS-DAP point clouds as a data source and have also found that the capabilities of UAS-DAP as a method in biomass prediction were in line with ALS predictions. The study also notes that even DAP was able to record accurate information on height and volume, it was not able

to convey as much information about the inner structure of forests as ALS which is suggested also by the results of this study.

The second objective of this study was to compare the DAP GC modelling performance to ALS GC modelling from the same area. In the context of previous ALS studies predicting Gini coefficient (Valbuena et al., 2013; Valbuena et al., 2014; Valbuena et al., 2013; Valbuena et al., 2016), the ALS model of this study performed in a slightly poorer manner. These studies have achieved for instance R^2 values of 0.52 – 0.91 compared to the 0.49 R^2 of this study. In the previous studies proportional cross-validated RMSEs for the best models have been around 17-20% in contrast to 27% RMSE of this study. In a comparison study of different Gini coefficient modeling methods, Valbuena et al. (2014) recorded relative RMSEs between 16-47% for different modeling methods. In that context, the ALS model of this study performed in an average manner.

There are some possible reasons for the modeling uncertainties of this study found in ALS related literature. The pulse density of the ALS data used in this study (1.4 m^{-2}) was lower than on the previous GC studies (~3m^{-2}) which might suggest that it might be too low to gather enough information from the inner forest structure. This is also backed up by a recent study by Adnan et al. (2017) that concluded that the nationwide NLS datasets used in this study are not very suitable for GC predictions as their return density is below their recommendation of 3 points m^{-2}. Also, one relevant detail regarding the ALS data is that the data used in this study had relatively high flight altitude and larger scan angles in comparison to the previous studies relating to GC prediction. This accompanied by the notation that the study area was situated between two flight lines which means there are substantially more returns with larger scan angles, might have some effect on the results. Large scan angles are known to cause for instance underestimation of low height percentiles (Holmgren, Nilsson, & Olsson, 2003) which - in this study - are represented in both of the models in the form of 10^{th} and 30^{th} height percentiles. In addition, the differences in the resulting GC maps are strongest in the parts of the forest where deciduous tree species are dominant. This suggests also that phenological reasons might influence the ALS modelling which is also backed by the notation that the ALS dataset was created in early May when trees are only in the early phase of leaf production. This means that a significantly higher proportion of laser pulses reached the ground (Ørka, Næsset, & Bollandsås, 2010) and this distorts the modelling because this does not happen in coniferous parts of the forest.

Positioning inaccuracies in plot center measurements are also known to weaken the model performance and the effect grows the smaller the plot size is (Gobakken & Næsset, 2009). Even a small positioning error in the center of the plot can produce proportionally a larger effect in a small circle plot than a bigger one. This study had a plot radius of 5 meters and a median positioning error of 1.1 meter which means that a median plot in the field dataset shared a common area of 86% with the actual in-situ plot. For comparison, if the radius was 10 meters and the positioning error the same, the corresponding overlapping area would have been 93%. Sensitivity analyses have been recently applied to find the optimal settings for mapping GC with ALS data (Adnan et al., 2017) and they suggest that the minimum plot radius for reliable GC estimations is 6 meters in contrast to 5 meters used in this study.

This study's significantly higher DAP point cloud density also did not noticeably increase the modeling accuracy. The vast quantity of information did not improve the physical variable retrieval and using such massive point clouds in forest variable estimation is questionable. A lighter point cloud would have made the entire process computationally less demanding and it would have been possible perhaps to expand the study area without losing too much details.

7.2 Variable selection

The variables selected in this study (Table 3) were in line with previous GC related studies. For instance, in a study by Valbuena et al. (2013) the variables were maximum, standard deviation, the 10th and 25th height percentiles, skewness of L-curve and percentage of returns above mean. In another study by Valbuena et al. (2016) the variables included also height variables like 95th percentile, and cover variables like canopy cover and percentage of returns above mode. These are similar or related to the ALS model variables in this study which were maximum elevation, variance of elevation, 30th height percentile and canopy cover.

The DAP model variables (Table 3) also shares similarities with the variables in previous ALS related GC studies. Maximum elevation, 10th percentile and standard elevation are all present in previous studies.

The variables that are not found in any previous GC studies are the elevation variance in the ALS model and elevation skewness in the DAP model. They both describe the vertical profile of the forest under the canopy which is also what the variable L-CV does in the previous studies.

It can be deduced from these findings that Gini coefficient modeling in general tends to perform better when it is accompanying at least one variable correlating with the highest point elevation (e.g. maximum elevation, 95^{th} percentile), one variable related to the vertical structure of the inner forest (e.g. standard error/variance of elevation) and one proportional variable that describes the horizontal canopy cover (e.g the canopy cover or percentage of returns above mean/median/mode).

These three categories can be seen as good measures of forest structure as they characterize vertical structure, horizontal structure and inner structure. With this view point in mind, it is easy to understand for instance that if a method is unable to produce information on some of these categories, it is quite clear it will not be good for Gini coefficient modeling. This might be the case on DAP data.

7.3 Feasibility of DAP data in studying forest structure

The main objective of this study was to assess the feasibility of DAP data in modeling Gini coefficient. As the performance of the DAP model in this study was not up to par with its ALS counterpart, it can be presumed that it might not be the best of methods for Gini coefficient modeling as the availability of the ALS data is significantly better at least in the Nordic countries. That said, it is also true that the performance of the ALS model was not as expected so it is arguable that the study settings might have influenced both methods.

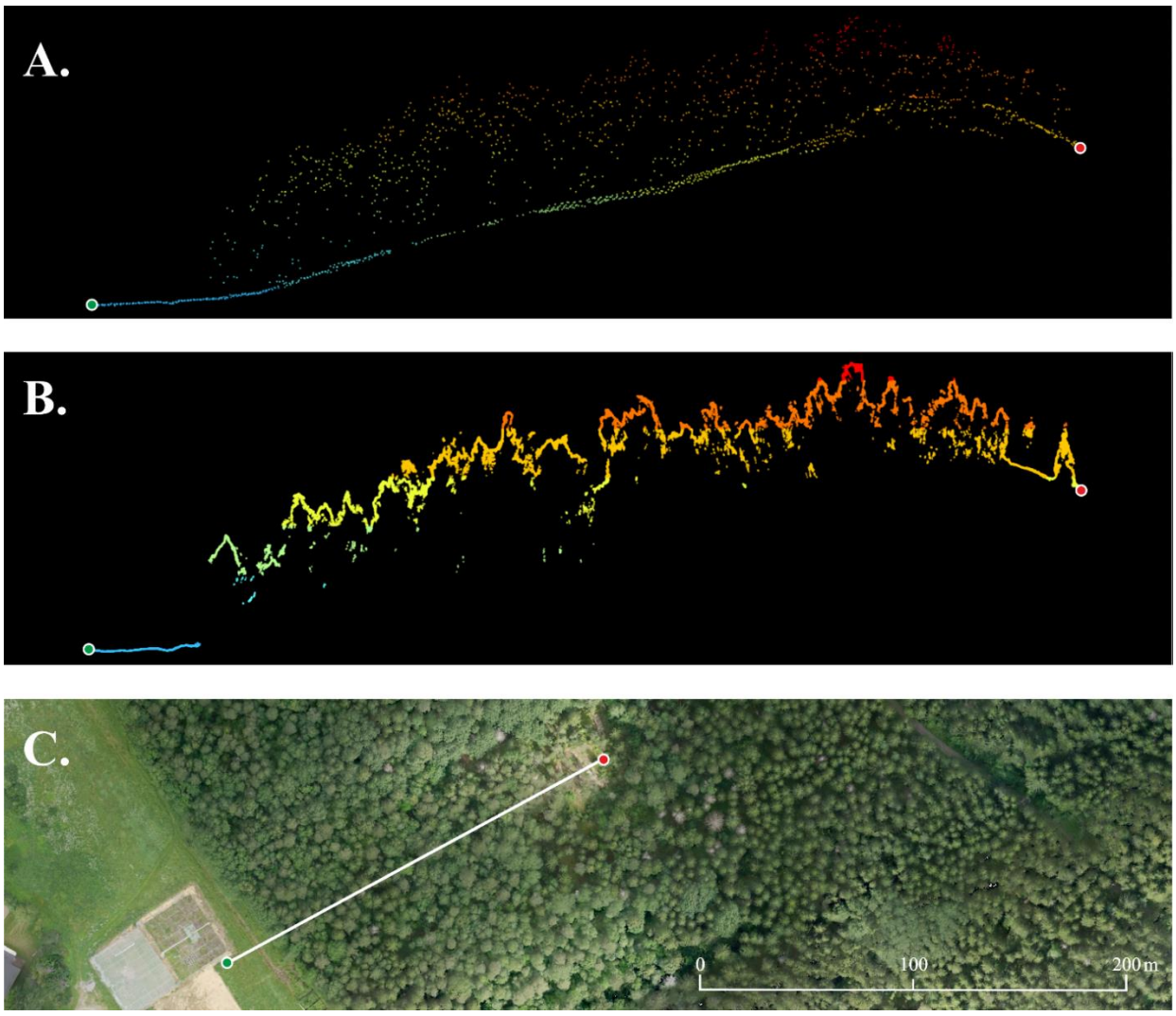

Figure 19. A cross cut of ALS (panel A) and DAP (panel B) point clouds and the location of the cross cut (panel C).

The DAP method lacks certain fundamental characteristics that might be prominent for GC modeling. Because the xyz information is photogrammetrically derived from photographs, its ability to penetrate the outer canopy layer is limited even though the point density might be superior. This is demonstrated in Figure 19. The first two represent cross cuts of the point clouds from the same line segment visible in the third panel. The differences in these datasets are clear. As the laser pulses in ALS rarely stop to the first echo, it is able to record more observations about the middle and bottom part of the forest and especially the ground. The DAP method on the other hand is able to produce superior amounts of information on the shape of the outer canopy but lacks observation capabilities on the middle and lower parts of the forest which can be seen in Figure 20

where both point cloud datasets for one plot are visualized. The difference in density and vertical characteristics between the datasets is noticeable.

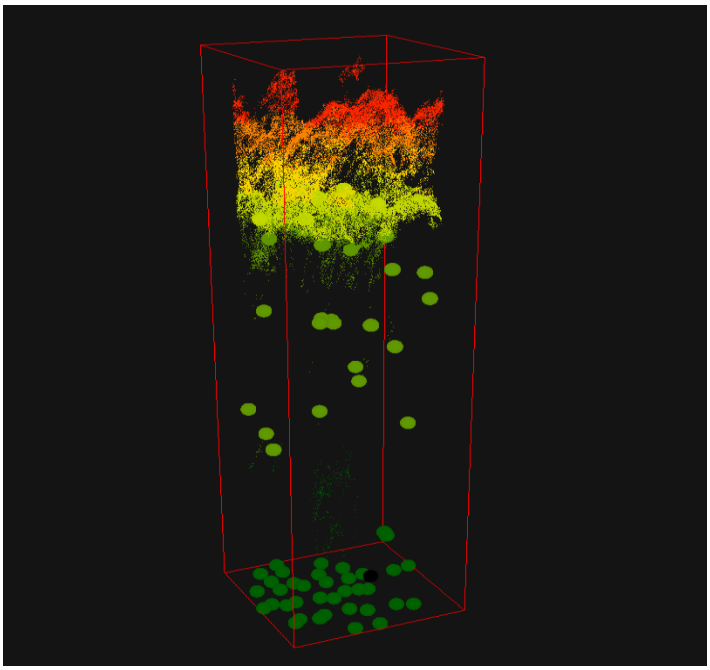

Figure 20. Two clipped point clouds of a single study plot. DAP points visualized with small points, ALS with large points.

As DAP and UAS technologies advance, the strengths and weaknesses of this methodology should be acknowledged and the capabilities and applications mapped. One characteristic of DAP data has not received much attention yet. It originates from the point cloud generation process where every point also receives the spectral value of the pixel it was derived from. For instance, in this study every point had a RGB value which was not utilized in modeling. This means that in addition to the three-dimensional position, every single point has information of the spectral properties of the feature it was derived from. This topic was touched by Puliti et al. (2015) in a UAS-DAP study where they briefly tested if spectral properties would increase the model performance. The study concluded that it only increased performance slightly but also that the lighting conditions were not optimal and varied between flights and that this topic needs more research.

Other suitable applications for DAP data could be for instance mapping gaps in the canopy cover. A gap in the canopy cover usually means a tree has died and fell or a tree has been cut down.

Finding these might be relevant in assessing storm damages or mapping illegal logging in the tropics. Gaps are visible because the point density in the DAP data is high and otherwise the observations from the bottom of the forest are very limited in number. An example of this can be seen in Figure 19B, where at the end of the line segment a distinct clearing in the forest is visible.

8. CONCLUSIONS

The main objective of this thesis was to test the feasibility of DAP data in modeling and mapping Gini coefficient.

The study found that the DAP data did not perform ideally in modelling forest structure with Gini coefficient. The RRMSE and R^2 values of the DAP model were mediocre in general and poorer than in the ALS model of this study. Also, when comparing the DAP results to previous GC studies, the model performance was unsatisfactory. The reasons behind this are likely related to the inability of the DAP data to gather information on the inner parts of the forest which affects the important proportional variables that are associated with GC modeling.

The flight campaign and the processing of the digital aerial photogrammetric data succeeded without any problems. A high-density point cloud of the area was successfully created and georeferenced.

During the analysis phase some shortfalls emerged. For instance, substantial amount of measured field plots had to be excluded due to the quality of their positioning. Secondly phenological differences due to the different time of year the two datasets were gathered (ALS in early May, DAP in June), likely influenced especially the ALS modeling.

The DAP data might not be optimal for mapping forest structure but its capability of providing cheap and accurate elevation data is recognized in the remote sensing community. These capabilities combined with increasing battery capacity, more intelligent autopilots and even more powerful algorithms lay down a bright future for unmanned aerial systems and digital aerial photogrammetry.

9. ACKNOWLEDGEMENTS

I wish to thank The Global Change and Conservation research group (GCC) and especially its team leader Dr. Mar Cabeza for enabling me to use their UAS and encouraging me to pursue whatever thesis topic I felt inspiring. I also want to thank Tero Vuorenmaa for using his own time to fly and operate the UAS and thus making the data acquisition of this study possible. The flight campaign was very successful and also great fun.

A great thank you belongs also to Dr. Janne Heiskanen for supervising this thesis and helping me in shaping the idea for the subject of this thesis. Thank you for giving always relevant comments and guidance in this project. Also, thank you Dr. John Loehr from the Lammi research station for giving tips and helping us to complete our flight campaigns.

Finally, for all of the numerous people who have supported me during this process, thank you!

10. REFERENCES

Adnan, S., Maltamo, M., Coomes, D. A., & Valbuena, R. (2017). Effects of plot size, stand density, and scan density on the relationship between airborne laser scanning metrics and the gini coefficient of tree size inequality. *Canadian Journal of Forest Research, 47*(999), 1590-1602.

AgiSoft, L. (2011). AgiSoft PhotoScan user manual: Professional edition, version 1.1.

Albert, M. (1999). *Analyse der eingriffsbedingten strukturvernderung und durchforstungsmodellierung in mischbestnden* Hainholz.

Angelstam, P., & Dönz-Breuss, M. (2004). Measuring forest biodiversity at the stand scale: An evaluation of indicators in european forest history gradients. *Ecological Bulletins*, (51), 305-332.

Arnold, M., Powell, B., Shanley, P., & Sunderland, T. (2011). Forests, biodiversity and food security. *International Forestry Review, 13*(3), 259-264.

Balvanera, P., Pfisterer, A. B., Buchmann, N., He, J., Nakashizuka, T., Raffaelli, D., & Schmid, B. (2006). Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecology Letters, 9*(10), 1146-1156.

Bonan, G. B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science, 320*(5882), 1444-1449.

Brockhaus, J. A., & Khorram, S. (1992). A comparison of SPOT and landsat-TM data for use in conducting inventories of forest resources. *International Journal of Remote Sensing, 13*(16), 3035-3043.

Ceriani, L., & Verme, P. (2012). The origins of the gini index: Extracts from variabilità e mutabilità (1912) by corrado gini. *The Journal of Economic Inequality, 10*(3), 421-443.

Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology, 35*(4), 445-453.

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 92*(Supplement C), 79-97.

Cribari-Neto, F., & Zeileis, A. (2009). Beta regression in R. *Research Report Series / Department of Statistics and Mathematics,* (98)

Dorfman, R. (1979). A formula for the gini coefficient. *The Review of Economics and Statistics,* , 146-149.

Doyle, F. (1964). The historical development of analytical photogrammetry. *Photogrammetric Engineering, 30*(2), 259-265.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*(7), 799-815.

Finland's forests 2017. (2017). Retrieved from https://www.luke.fi/wp-content/uploads/2017/06/finlands-forests-facts-2017-www.pdf

Finsterwalder, S., & Weinschenk, E. (1904). *Eine grundaufgabe der photogrammetrie und ihre anwendung auf ballonaufnahmen* G. Franz Komission.

Fries, C., Johansson, O., Pettersson, B., & Simonsson, P. (1997). Silvicultural models to maintain and restore natural stand structures in swedish boreal forests. *Forest Ecology and Management, 94*(1-3), 89-103.

Gadow, K. v., Hui, G. Y., & Albert, M. (1998). Das winkelmaß–ein strukturparameter zur beschreibung der individualverteilung in waldbeständen. *Centralblatt Für Das Gesamte Forstwesen, 115*(1), 1-9.

Gini, C. (1912). Variabilità e mutabilità. contributi allo studio dele relazioni e delle distribuzioni statistiche. *Studi Economico-Giuridici Della Università Di Cagliari.,*

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.

Hett, J. M., & Loucks, O. L. (1976). Age structure models of balsam fir and eastern hemlock. *The Journal of Ecology*, , 1029-1044.

Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, *19*(6), 543-553.

Holmgren, J., Nilsson, M., & Olsson, H. (2003). Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing*, *29*(5), 623-632.

Holopainen, M., Hyyppä, J., & Vastaranta, M. (2013). *Laserkeilaus metsävarojen hallinnassa*

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S., & Zhu, Y. (2000). Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*, *128*(1), 109-120.

Ingram, J. C., Dawson, T. P., & Whittaker, R. J. (2005). Mapping tropical forest structure in southeastern madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment*, *94*(4), 491-507.

Isenburg, M. (2016). *LAStools-efficient tools for LiDAR processing*

IUCN. (2012). Facts and figures on forests. Retrieved from https://www.iucn.org/content/facts-and-figures-forests

Järnstedt, J., Pekkarinen, A., Tuominen, S., Ginzler, C., Holopainen, M., & Viitala, R. (2012). Forest variable estimation using a high-resolution digital surface model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *74*, 78-84.

Kachamba, D. J., Ørka, H. O., Gobakken, T., Eid, T., & Mwase, W. (2016). Biomass estimation using 3D data from unmanned aerial vehicle imagery in a tropical woodland. *Remote Sensing, 8*(11), 968.

Korpela, I. (2004). *Individual tree measurements by means of digital aerial photogrammetry* Finnish Society of Forest Science.

Korpela, I. (2006). Geometrically accurate time series of archived aerial images and airborne lidar data in a forest environment. *Silva Fennica, 40*(1), 109.

Kortesmaa, T., & Jokela, A. (2017). E-yearbook of food and natural resource statistics for 2016: Statistical facts on agriculture, forestry, fisheries and hunting in finland.

Lähde, E., Laiho, O., Norokorpi, Y., & Saksa, T. (1991). The structure of advanced virgin forests in finland. *Scandinavian Journal of Forest Research, 6*(1-4), 527-537.

Lähde, E., Laiho, O., Norokorpi, Y., & Saksa, T. (1999). Stand structure as the basis of diversity index. *Forest Ecology and Management, 115*(2), 213-220.

Lähde, E., Laiho, O., & Pukkala, T. (2010). Eri-ja tasarakenteiskasvatuksen vertailua pohjoismaissa. *Metlan Työraportteja / Working Papers of the Finnish Forest Research Institute, 176*

Lammin biologinen asema. (2016). Aseman historia. Retrieved from http://www.helsinki.fi/lammi/esittely/historia.html

Lexerød, N. L., & Eid, T. (2006). An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management, 222*(1), 17-28.

Lowe, D. G.Object recognition from local scale-invariant features. *Computer Vision, 1999. the Proceedings of the Seventh IEEE International Conference On, , 2* 1150-1157.

Maanmittauslaitos. (2016). Laserkeilausaineisto. Retrieved from http://www.maanmittauslaitos.fi/digituotteet/laserkeilausaineisto

McGaughey, R. J. (2016). FUSION/LDV: Software for LIDAR data analysis and visualization. *US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA, 123*(2)

Nelson, R. (2013). How did we get here? an early history of forestry lidar. *Canadian Journal of Remote Sensing, 39*(sup1), S17.

Noss, R. F. (1999). Assessing and monitoring forest biodiversity: A suggested framework and indicators. *Forest Ecology and Management, 115*(2–3), 135-146.

Ørka, H. O., Næsset, E., & Bollandsås, O. M. (2010). Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. *Remote Sensing of Environment, 114*(7), 1445-1461.

Parviainen Jari, & Västilä Sinikka. (2012). *State of finland's forests 2012 based on the criteria and indicators of sustainable forest management* Ministry of Agriculture and Forestry & Finnish Forest Research Institute.

Pommerening, A. (2002). Approaches to quantifying forest structures. *Forestry: An International Journal of Forest Research, 75*(3), 305-324.

Puliti, S., Ørka, H. O., Gobakken, T., & Næsset, E. (2015). Inventory of small forest areas using an unmanned aerial system. *Remote Sensing, 7*(8), 9632-9654.

Sarjakoski, T. (1981). Concept of a completely digital stereo plotter. *The Photogrammetric Journal of Finland, 8*(2), 95-100.

Shan, J., & Toth, C. K. (2008). *Topographic laser ranging and scanning: Principles and processing* CRC press.

Shears, J. C., & Allan, J. W. (2004). Softcopy photogrammetry and its uses in GIS. *The AGI Source Book for GIS,* , 41.

Solodukhin, V., Zukov, A., & Mazugin, I. (1977). Possibilities of laser aerial photography for forest profiling. *Lesnoe Khozyaisto (Forest Management), 10*, 53-58.

Team, R. C. (2000). R language definition. *Vienna, Austria: R Foundation for Statistical Computing,*

Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W.Bundle adjustment — A modern synthesis. *International Workshop on Vision Algorithms,* 298-372.

Tuominen, S., Balazs, A., Saari, H., Pölönen, I., Sarkeala, J., & Viitala, R. (2015). Unmanned aerial system imagery and photogrammetric canopy height data in area-based estimation of forest variables. *Silva Fennica Vol.49 Nr.5, Article Id 1348,*

Tuominen, S., & Haapanen, R. (2011). Comparison of grid-based and segment-based estimation of forest attributes using airborne laser scanning and digital aerial imagery. *Remote Sensing, 3*(5), 945-961.

Valbuena, R., Eerikäinen, K., Packalen, P., & Maltamo, M. (2016). Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. *Ecological Indicators, 60*, 574-585.

Valbuena, R., Maltamo, M., Martín-Fernández, S., Packalen, P., Pascual, C., & Nabuurs, G. (2013). Patterns of covariance between airborne laser scanning metrics and lorenz curve descriptors of tree size inequality. *Canadian Journal of Remote Sensing, 39*(Suppl. 1), S31.

Valbuena, R., Packalén, P., Martın-Fernández, S., & Maltamo, M. (2012). Diversity and equitability ordering profiles applied to study forest structure. *Forest Ecology and Management, 276*, 185-195.

Valbuena, R., Packalen, P., Mehtatalo, L., Garcia-Abril, A., & Maltamo, M. (2013). Characterizing forest structural types and shelterwood dynamics from lorenz-based

indicators predicted by airborne laser scanning. *Canadian Journal of Forest Research, 43*(11), 1063-1074.

Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., & Maltamo, M. (2014). Comparison of airborne laser scanning methods for estimating foreststructure indicators based on lorenz curves. *ISPRS Journal of Photogrammetry and Remote Sensing, 95*, 33.

Van der Werf, Guido R, Morton, D. C., DeFries, R. S., Olivier, J. G., Kasibhatla, P. S., Jackson, R. B., . . . Randerson, J. T. (2009). CO2 emissions from forest loss. *Nature Geoscience, 2*(11), 737-738.

Vilhomaa, J., & Laaksonen, H. (2011). Valtakunnallinen laserkeilaus–testityöstä tuotantoon. *The Photogrammetric Journal of Finland, 22*(3), 82-91.

Von Gadow, K., & Hui, G. Y. (2002). Characterizing forest spatial structure and diversity. *W: Bjoerk L.[Red.].Sustainable Forestry in Temperate Regions.Materiały Konferencyjne IUFRO, Lund, ,* 20-30.

Weiner, J. (1985). Size hierarchies in experimental populations of annual plants. *Ecology, 66*(3), 743-752.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). 'Structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology, 179*, 300-314.

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing, ,* 1-23.

Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., . . . Boon, N. (2009). Initial community evenness favours functionality under selective stress. *Nature, 458*(7238), 623-626.

Ympäristöministeriön asetus natura 2000 –verkostoon kuuluvien alueiden luettelosta, (2015).

11. APPENDICES

Appendix 1. Predictor variables

Abbreviation	Description	Present also in DAP variable selection
Elev.minimum	Minimum elevation	X
Elev.maximum	Maximum elevation	X
Elev.mean	Mean elevation	X
Elev.mode	Mode of elevation	X
Elev.stddev	Standard deviation of elevation	X
Elev.variance	Variance of elevation	X
Elev.CV	Coefficient of variation of elevation	X
Elev.IQ	Interquartile distance of elevation	X
Elev.skewness	Skewness of elevation	X
Elev.kurtosis	Kurtosis of elevation	X
Elev.AAD	Average absolute deviation of elevation	X
Elev.MAD.median	Median of the absolute deviations from the overall median	X
Elev.MAD.mode	Mode of the absolute deviations from the overall mode	X
Elev.L1	1st elevation L-moment	X
Elev.L2	2nd elevation L-moment	X
Elev.L3	3rd elevation L-moment	X
Elev.L4	4th elevation L-moment	X
Elev.L.CV	Coefficient of variation of L-moment	X
Elev.L.skewness	Skewness of L-moment	X
Elev.L.kurtosis	Kurtosis of L-moment	X
Elev.P01	1st percentile of elevation	X
Elev.P05	5th percentile of elevation	X
Elev.P10	10th percentile of elevation	X
Elev.P20	20th percentile of elevation	X
Elev.P25	25th percentile of elevation	X
Elev.P30	30th percentile of elevation	X
Elev.P40	40th percentile of elevation	X
Elev.P50	50th percentile of elevation	X
Elev.P60	60th percentile of elevation	X
Elev.P70	70th percentile of elevation	X
Elev.P75	75th percentile of elevation	X
Elev.P80	80th percentile of elevation	X
Elev.P90	90th percentile of elevation	X
Elev.P95	95th percentile of elevation	X

Elev.P99	99th percentile of elevation	X
Canopy.relief.ratio	Canopy relief ratio	X
Elev.SQRT.mean.SQ	Generalized means for the 2nd power	X
Elev.CURT.mean.CUBE	Generalized means for the 3rd power	X
Percentage.first.returns.above.1.00	Percentage of first returns above 1 meter	
First.returns.above.1.00	First returns above 1 meter	
Percentage.first.returns.above.mean	Percentage of first returns above mean	
Percentage.first.returns.above.mode	Percentage of first returns above mode	
First.returns.above.mean	First returns above mean	
First.returns.above.mode	First returns above mode	
Cover	Percentage of all returns above 1 meter. Also known as canopy cover	X
Cover.mean	Percentage of all returns above mean	X
Cover.mode	Percentage of all returns above mode	X
ARATFR	Number of returns above 1 meter / total first returns * 100	X
ARAMETFR	Number of returns above the mean height / total first returns * 100	X
ARAMOTFR	Number of returns above the mode height / total first returns * 100	X
All.returns.above.1.00	All returns above 1 meter	X
All.returns.above.mean	All returns above 1 mean	X
All.returns.above.mode	All returns above 1 mode	X
Total.first.returns	Total number of first returns	
Total.all.returns	Total number of all returns	X