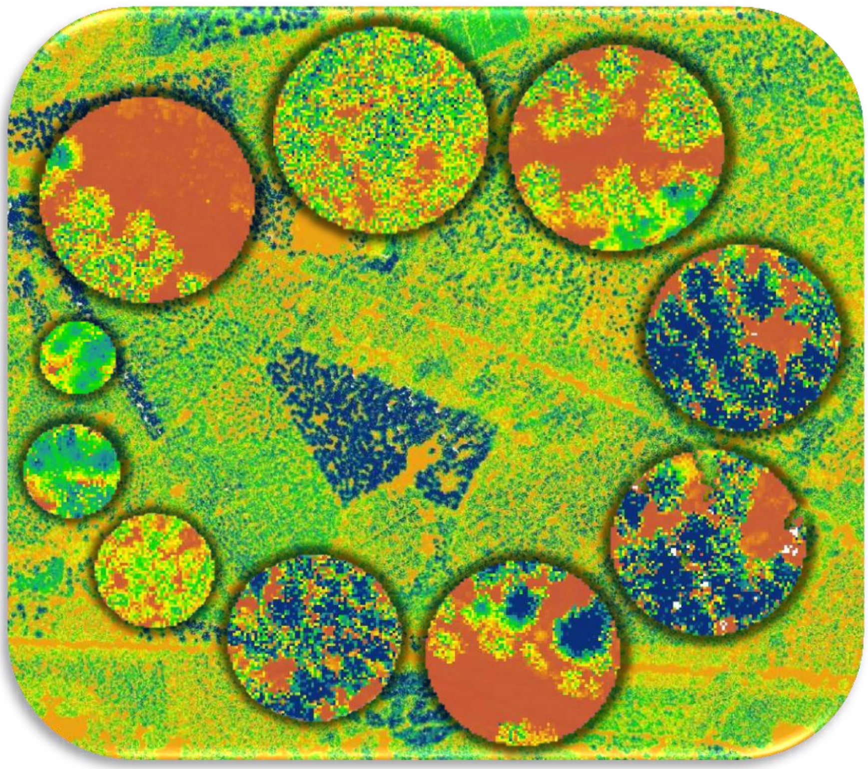
Centre for Geo-Information

Thesis Report GIRS-2011-16

Opportunities for LIDAR to characterize forest stand characteristics and biomass

Jeroen M. de Jong

August 2011



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Opportunities for LIDAR to characterize forest stand characteristics and biomass

Jeroen M. de Jong

Registration number 86 12 27 404 050

Supervisors:

Dr. ir.. Lammert Kooistra
Dr. Harm Bartholomeus

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

August 2011
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2011-16
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Table of contents

List of figures and tables

Summary

Estimations of forest carbon stock and carbon fluxes from forests have traditionally been estimated by carrying out field inventories. Another method that is increasingly being used is the measurement of forest characteristics by applying remote sensing techniques. By correlating field measured carbon densities to remotely sensed variables, estimations of carbon densities for large continuous areas can be made. Lidar remote sensing is a technique that can be applied in The Netherlands to do so. In this thesis AHN-2 lidar data as a 0.5m grid has been related to national inventory data from the Meetnet Functievervulling and a local inventory carried out for Staatsbosbeheer.

A literature review has resulted in several methods for obtaining forest characteristics for both grid data and point cloud data. These methods have been adjusted to be applicable on AHN-2 lidar data with two reference datasets, for a 3900 ha forest area on the Utrechtse Heuvelrug. The reference data has been split in a training and test dataset.

Linear regression between AHN-2 lidar data and forest inventory data has resulted in the calculation of canopy height for both the reference plots and the total forested area. Forest cover has been determined based on height differences between forest canopy and open spaces within the forest. By applying linear regression between forest cover and reference data of basal area, and timber volume linear relationships have been established. These relationships have been used to create maps of timber volume and basal area for the complete study area. Furthermore, a classification method has been has been developed to make a distinction between deciduous and coniferous forest.

Canopy height has been calculated both as maximum height and as dominant height. For the relationship between field measured and lidar derived height, R^2 values range from 0.48 to 0.66. Applying the relationships on the total forested area gives for the test data RMSE values that range from 2.6 to 4.5 meter.

The forest inventory data used as reference data, was not collected specifically for calibration and validation of lidar data. As a consequence the plot locations and the height measurements within the plots are less appropriate to use as reference data. Forest cover as estimated by lidar data did not correspond to the estimations of field measured crown cover and forest density. Forest cover however could be related to field measured basal area and timber volume, which resulted in an R^2 of 0.46 for forest cover and timber volume and an R^2 of 0.57 for crown cover and basal area. The classification in deciduous and coniferous forest has been compared with the forest classes of the Top10 topographic map and with the map of the 4th national forest inventory. For larger areas that are either coniferous or deciduous the classification based on canopy profile works reasonably well, although this has not been validated by independent reference data. The applicability outside the study area still needs to be investigated in more detail as the study area itself is relatively homogeneous when considering forest type and growing conditions. Forests on a relatively rich and moist soil are not represented by the study area.

1 Introduction

1.1 Context and background

In the global carbon balance about two third of the amount of anthropogenic emitted CO_2 is taken up by oceans or accumulates in the atmosphere (Goodale et al., 2002; Houghton et al., 2009). There is still uncertainty about 34% of all anthropogenic emitted CO_2 (Friend et al., 2007). In estimations of a global carbon balance the terrestrial ecosystem is considered a sink, which could account for most of the uncertain part of carbon uptake. Normally the terrestrial sink is quantified as the result of total emissions minus known sinks which are the atmosphere and the oceans (Houghton et al., 2009). Terrestrial ecosystems in the northern mid-latitude hemisphere and undisturbed tropical forests take up carbon, but amount, location and processes are still unclear (Friend et al., 2007). Most of the biomass in terrestrial ecosystems is stored in forests as they contain about 70 - 90% of terrestrial biomass of which 70 - 90% is contained in aboveground forest biomass (Houghton et al., 2009).

Estimations for the forest carbon sink have been made by using different approaches. Data on national forest inventories provide an estimation of biomass present in the forests (Ciais et al., 2008; Goodale et al., 2002; Luyssaert et al., 2010). However, the sample size in these inventories is very small compared to the total forested area (Houghton et al., 2009). Fluxes between atmosphere and the terrestrial ecosystems are being measured by a network of flux towers organized in for example the FLUXNET program (Friend et al., 2007). Another method that is increasingly being used is measuring of forest characteristics by remote sensing techniques. The step from measured characteristics to carbon content estimations is made by correlating field measured carbon densities, calculated with allometric equations, to the remote sensing based values (Gonzalez et al., 2010).

1.2 Problem definition

Estimating carbon fluxes and carbon stock is done on different levels of detail. For example the IPCC has made guidelines that help to develop carbon assessments on a national level (Asner, 2009). Besides on a national level, more and more information about carbon sequestration on a regional scale is required by local governments like provinces. For The Netherlands, lidar (light detection and ranging) is a remote sensing technique that might be applied to assess forest carbon stocks at the regional scale.
Lidar is an active remote sensing technique, which sends a pulse of light to an object and time of travel and the reflected energy of the pulse is measured. Combined with precise information on flying height and location, the height of the scanned terrain is calculated from the lidar signal (McRoberts et al., 2010). It can deliver information both about the height of the terrain and the objects in the terrain. In the case of a forest, terrain height and canopy height can be measured. These directly measured variables can be used to derive more complex variables like forest cover (LAI) or biomass density.

Lidar remote sensing or laser altimetry has a variety of forms. It can be collected from the ground or from airborne or spaceborne platforms. Another important characteristic in lidar remote sensing is the distinction between discrete return and full waveform sensors.

Discrete return systems measure from one pulse one or several peak reflectances whereas full waveform lidar systems record the full reflectance spectrum (McRoberts et al., 2010). One of the advantages of the full waveform, system is that it can produce more detailed information on vegetation structures in forests (Mallet & Bretar, 2009).
The discrete return system operated from an airborne platform is the most widely available system as it is used by surveyors and natural resource managers (Lefsky et al., 2002). They apply this system for surveying, photogrammetric applications and forest inventories in order to produce topographic maps, digital terrain models and obtaining information about forest structure and characteristics (Lefsky et al., 2002; McRoberts et al., 2010). The application of this type of systems is often called airborne laser scanning (ALS). The advantages of the system can be summarized as combining high spatial resolution with the possibility to cover larger areas completely (McRoberts et al., 2010). A disadvantage is that the systems designed for commercial application might not meet the requirements for scientific purposes (Lefsky et al., 2002). For example, topographic mapping gives the best result when data is collected during the leaf off season, while for research on forests, data that were collected during the leaf on season can best be used. As large scale data collection is expensive, it is profitable to use one data set more often and for more different purposes.

In The Netherlands a nationwide lidar dataset, collected by an airborne laser scanning system is available. This data is collected as part of the AHN (Actueel Hoogtebestand Nederland) program in which a nationwide detailed height map is produced. The first height map (AHN-1) that was completed in 2003, has a height accuracy of 16 cm and a point density that varies from 1 point per 16 m^2 up to 1 per m^2 (AHN, 2010). The second dataset (AHN-2) that is collected over the period 2007 - 2012 will have a mean point density of 6 - 10 points per m^2 (AHN, 2010). The AHN-2 will have a maximum systematic error and a maximum stochastic error of 5 cm. To make use of the AHN-2 dataset for retrieving forest characteristics there are some issues that need to be looked at.

For establishing a relationship between measured lidar data and forest characteristics, these have to be linked by calibration and validation with reference data often being forest inventory datasets. For the Netherlands data from the national forest inventory (Meetnet Functie Vervulling bos) might be a possible source of data for calibration and validation. This national forest inventory includes 3622 sample plots scattered over forest areas in The Netherlands which results in a density of one sample plot per 100 ha of forest (Directie Kennis, 2006). It includes information on canopy height, canopy cover, forest age, main tree species, number of trees per ha and timber volume per ha. These data were collected during the period 2001 – 2005.

As described by Van Leeuwen & Nieuwenhuis (2010) difference in scale of collected lidar data and field measurements might cause problems in the calibration and validation of tree height between datasets. This might be a problem as the data of the Meetnet Functievervulling bos was not specially collected to match the AHN data.
Lidar has been used for forest inventories in different countries and in different forest types. The characteristics of the terrain and the forest (for example tree species, canopy density and density and height of undergrowth) play a role in the accuracy of the calculations of canopy

height (Hyyppa et al., 2008). The influence of the type of terrain and type of forests is unknown for the situation in The Netherlands.

The study of Weber & Boss (2009), on an area in the Eastern United States, is of interest as it has some similarities with the situation in The Netherlands. Lidar data was collected for quite a large area in the leaf off season, not particularly for the research on forest characteristics, in an area that contained both deciduous and coniferous trees of different ages and successional stages (Weber & Boss, 2009). These circumstances are also applicable on the situation in The Netherlands. The goal of the research was to estimate forest maturity by using both lidar data and supplemental data. The lidar data was used to make an estimation of canopy height by applying regression. The supplemental information used consists of aerial pictures, land use maps and a map on hydrographic flowlines The best fitting curve for height estimation was a logistic equation with a R^2 of 0.72 (Weber & Boss, 2009).

1.3 Research objective

The objective of this thesis is to develop a method that relates discrete return AHN-2 lidar data to forest inventory data like the Meetnet Functie Vervulling (MFV) in order to retrieve forest characteristics like canopy height and crown size.

- Which (linear) regression method can be applied on AHN-2 data and MFV forest statistics and how does this method work?
- What other methods besides regression can be used for analysis of AHN-2 data?
- How does forest type influence the outcomes of the analysis?
- How well does the for the test area developed methodology work for other areas in The Netherlands?

Answering these research questions will help in the development of methods that estimate forest carbon contents, as forest carbon content is related to these forest characteristics.
Because this research field is relatively new for The Netherlands, information on methodology has to be obtained by means of studying available literature on the subject. This methodology will be applied on forest areas that have both AHN-2 and MFV data available at the moment of this thesis research.

2 Estimating forest parameters by lidar

2.1 Lidar systems

Light detection and ranging (Lidar) is an active remote sensing technique, which sends a pulse of light to an object and time of travel and the reflected energy of the pulse is measured. Combined with precise information on flying height and location of the platform, the height of the scanned terrain is calculated (McRoberts et al., 2010). It can deliver information both about the height of the terrain and the objects in the terrain. By filtering out the non-ground lidar returns a digital elevation model (DEM) is made. The non-grond lidar returns are objects in the terrain like buildings or vegetation. This data is called the digital terrain model (DTM). By subtracting the height values in the digital elevation model from the height values in the digital terrain model, the height of the objects in the terrain is calculated. This is a digital surface model (DSM). In the case of a forest, terrain height and canopy height can both be measured. These directly measured variables can be used to derive more complex variables like forest cover (LAI) or biomass density.

Lidar remote sensing or laser altimetry has a variety of forms. It can be collected from the ground or from airborne or spaceborne platforms. Besides the type of platform there are some other characteristics, which include footprint diameter, scanning mechanism, wavelength, pulse repetition frequency and type of information recorded.

The footprint diameter of a lidar system is the diameter of the pulse on the ground. Besides the divergence of the laser beam, which is constant, the footprint size is also determined by flying height. The size of the footprint varies from 20-100 m for spaceborne and from 0.1 – 20 m for airborne systems (McRoberts et al., 2010). A small footprint lidar is defined as a system which has a diameter in the range of 0.1-2 m. (Hyyppa et al., 2008; McRoberts et al., 2010).
Two scanning mechanisms can be distinguished: profiling, where information is recorded in a narrow path beneath the platform, and scanning with an oscillating mirror, which allows larger swath widths (Lefsky et al., 2002). For airborne lidar systems with a small footprint the swath width is in the range of 500 – 2000 m. (McRoberts et al., 2010).

The pulse repetition frequency is the number of pulses emitted per second. Because of technological developments the frequency has increased. In 2003, the pulse repetition frequency was around 2 kHz and in 2008 it increased to around 200 kHz. There are also systems that use a technique called multiple pulse in air. With this technique a second pulse is emitted before the reflection of the first pulse has arrived, which makes it possible to increase the frequency even further (Van der Sande et al., 2010).

A last distinction is made between discrete return and full waveform sensors. Discrete sensors can be further subdivided in single or multiple return systems. Single return systems measure only one reflectance per pulse. Multiple return systems measure from one pulse several peak reflectances, for example the reflectance of the branches of a tree and the reflectance of the earth. From each of these reflectances the return time and intensity is recorded. Full waveform lidar systems record the full reflectance spectrum (McRoberts et al., 2010).

One of the advantages of this system is that it can produce more detailed information on vegetation structures in forests (Mallet & Bretar, 2009).
The discrete return system operated from an airborne platform is the most widely available system as it is used by surveyors and natural resource managers (Lefsky et al., 2002). They apply this system for surveying, photogrammetric applications and forest inventories in order to produce topographic maps, digital terrain models and obtaining information about forest structure and characteristics (Lefsky et al., 2002; McRoberts et al., 2010). The application of this type of systems is often called airborne laser scanning (ALS). The advantages of the system can be summarized as combining high spatial resolution with the possibility to cover larger areas completely (McRoberts et al., 2010). A disadvantage is that the systems designed for commercial application might not meet the requirements for scientific purposes (Lefsky et al., 2002). For example, topographic mapping gives the best result when data is collected during the leaf off season, while for research on forests, data that were collected during the leaf on season can best be used.
Systems with larger footprints that record full waveform information are mostly used for scientific purposes. For example for research on large scale forest cover changes (Lefsky et al., 2002).

2.2 Methodological structure

Which method to use for forest inventories depends on a number of factors. These include the type of lidar data and the scale used for data analysis. The different analysis methods are summarized in table 2.1 and described in this section. They are structured according to data type, because the type of data determines the analysis possibilities in the first place.

2.2.1 Types of data

Lidar data is the result of reflectance's of the lidar pulse from a certain surface and the distance to this surface. This information can be recorded as either discrete return or full waveform. The data can be stored as either point cloud data or grid data. The main difference between the two data types is that point clouds can have several points in one location at different heights, while the grids have only one height value at a location (figure 2.1). This point cloud data furthermore can have reflectance values for the different points. Because of this, the two different file types require different analysis techniques.

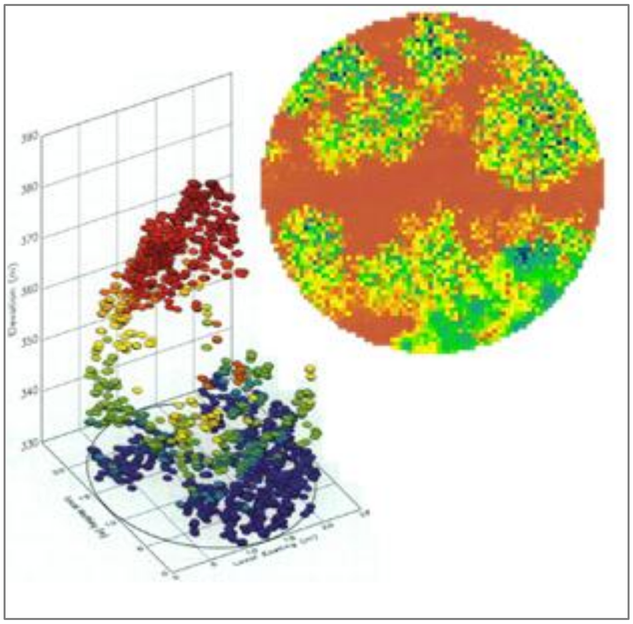


Figure 2.1 Representation of point cloud data and grid data
Point cloud data is shown on the left and grid data on the right (left image: Lefsky et al., 2002).

2.2.2 Direct and indirect measurements

A forest parameter like tree height can be directly estimated from lidar data, as it is just a height value of one of the reflectances or an average of reflectances from the tree. Another parameter that can be directly measured is forest cover. Indirect estimated forest parameters are retrieved by establishing a relationship between a directly estimated parameter and a ground measured forest parameter. Examples of these are canopy height, forest age class (Weber & Boss, 2009) and canopy volume profile which relates to basal area, crown volume and LAI (Coops et al., 2007).

2.2.3 Different scales

Forest parameters can be derived on different scales. On the level of individual trees, on plot level or on stand level. A forest stand is an area that has “homogenous tree cover and uniform site conditions”. In forest management, forest stands are the operational and planning units (Magnusson, 2006). A plot can be seen as a sampling area within a forest, used for amongst other research and inventory purposes. In The Netherlands for inventory purposes these are circular areas with a radius varying between 5 and 20 meters. This depends on tree density as a plot should include around 20 trees (Daamen & Dirkse, 2005). For analysis of individual trees and forest stands these need to be distinguished from the grid data. For analysis on plot level only the grid cells that fall within the plots have to be selected.

Table 2.1 Summary of literature search on methods to derive forest parameters from lidar data

Grid data methods			
Parameter	**Scale level**	**Method**	**Source**
	Direct estimate		
Tree height	Tree	Distinguish trees and determine maximum height	Heurich, 2008; Zhao et al., 2009; Hyyppa et al., 2008; Kwak et al., 2007; Garcia et al., 2007
Tree height	Plot	Maxima within plot or for dominant height as mean maxima van subplots	Weber & Boss, 2009; Coops et al., 2007
Tree height	Stand	Distinguish stand and determine maximum or average	Leppänen et al., 2008; Sullivan, 2008
Dominant height	Plot	Average height of maxima of different sub plots	Weber & Boss, 2009; Coops et al., 2007
Crown surface	Tree	Distinguish trees and determine surface area	Hyyppa et al., 2008
Opennes	Plot	Fraction between cells in gap and total number of cells	Coops et al., 2007
Tree density	Plot / Stand	Distinguish trees and determine number of trees per area	Hyyppa et al., 2008

Point cloud data methods			
Parameter	**Scale level**	**Method**	**Source**
	Direct estimate		
Canopy height	Plot	Regression between maximum height within forest inventory plot and field measured height in a 6 m grid	Weber & Boss, 2009
Canopy volume profile	Plot	Simulate full waveform by determining number of reflections per height for larger area	Coops et al., 2007
Lorey's mean height	Plot	Multiple regression on lidar derived parameters	Andersen et al. (2005) in Van Leeuwen en Nieuwenhuis 2010; Næsset and Økland (2002)
Stem number	Plot	Multiple regression on lidar derived parameters	Gobakken & Naesset, 2004
Basal area	Plot	Multiple regression on lidar derived parameters	Gobakken & Naesset, 2004
Total biomass	Plot	Multiple regression on lidar derived parameters	Lim & Treitz, 2004
	Indirect estimate		
Average height	Plot	Correlates to lidar derived canopy volume profile	Coops et al., 2007
Volume	Plot	Correlates to lidar derived canopy volume	Coops et al., 2007

		profile	
Crown volume	Plot	Correlates to lidar derived canopy volume profile	Coops et al., 2007
LAI	Plot	Correlates to lidar derived canopy volume profile	Coops et al., 2007
Basal area	Plot	Correlates to lidar derived canopy volume profile	Coops et al., 2007

2.3 Grid data methods

2.3.1 Distinguishing individual trees and stands

For distinguishing individual trees there are different methods available. Hyppa et al. (2008) divide methods for individual tree analysis in three categories. These are: 1) finding only the location of the tree, 2) finding the location of the tree and calculating parameters that contain information about the size of the crown and 3) fully delineating tree crowns (Hyppa et al., 2008).
For only retrieving the location of individual trees local maxima analysis can be used (Hyppa et al., 2008). This method can give good results in coniferous forests on the condition that parameter values (filter size and image smoothing) fit the forest and data characteristics (tree size and image resolution) (Hyppa et al., 2008).
For gaining more information about the tree, the tree can be further investigated by determining the edge of the crown and retrieving information about crown size. Methods for obtaining crown edges are local minima detection, edge detection and region segmentation (Hyppa et al., 2008).

One of the common approaches for fully delineating tree crowns is using the same sort of analysis techniques as for hydrological watershed analysis (Heurich, 2008; Van Leeuwen & Nieuwenhuis, 2010). In watershed analysis drainage in a landscape is modelled, where water flows to the lowest point. By reversing the canopy height information, the treetops are the lowest points and will be considered as a valley bottom (Heurich, 2008). The rim of the watershed is the edge of the tree crown (Heurich, 2008). Other methods described by Hyppa et al. (2008) are “shade-valley-following, edge curvature analysis, template matching and region growing”.
Zhao et al. (2009) use a program called TreeVaw to delineate trees. This program uses local maximum filtering to find the treetops. There is an assumption made that higher trees have wider crowns, so the distance to search for crown edges depends on the height of the tree (Zhao et al., 2009).

Heurich (2008) uses an algorithm to automatically delineate individual tree crowns. The method used in this algorithm is comparable to the automatic recognition of watershed areas in hydrological analysis. This watershed determination is done on a coarse scale and a finer scale. Comparing the two scales helps in improving the accuracy of the segmentation. When a segmented area has in both scales only one maximum the tree is correctly distinguished. Two maxima in the finer scale might be due to the inclusion of two trees in the coarser scale or because of variation within a tree (Heurich, 2008). The algorithm decides on this issue by further calculations. This method performed better on coniferous than on deciduous trees. 76.8% of the trees present in the upper canopy layer were recognized, which is better than the 45.4% for all layers together (Heurich, 2008).

Forest stands are groups of trees that have the same species composition age, structure and grow under the same site conditions (Magnusson, 2006; Leckie et al., 2003). To distinguish these other methods need to be applied then for distinguishing individual trees. Segmentation is one such a method, which groups pixels to objects by comparing them to their surroundings. An example of such a method is given by Leppänen et al. (2008). This method selects pixels to start with and compares the surrounding pixels to these start pixels. If the values are within a certain range they are added to the start pixels and in this way homogeneous areas are formed. These areas grow until they reach pixels that differ too much to be included in the area. To fill the complete area, new start pixels are chosen in the empty areas and the process starts over until all areas are filled (Leppänen et al., 2008). The software package eCognition is a program that is able to do this (Leppänen et al., 2008).

2.3.2 Height

The height of the canopy can be determined in different ways. Lidar measured tree height will show an underestimation of real canopy height (Wang & Glenn, 2008). Because lidar pulses form sample measurements of the canopy, the highest elevation points are consistently missed. The accuracy of a directly measured canopy height model depends on a number of factors, which include crown shape, sampling density, pulse diameter and peak detection-method (Wang & Glenn, 2008). By increasing the footprint of a laser pulse the chance of hitting the treetop increases, however the intensity of reflected energy from the small treetop is relatively smaller at a larger footprint and the chance of recording this reflected energy as a peak will decrease. The method of peak detection in the continuous backscatter of lidar returns determines the minimum peak size required to be registered as a peak reflectance (Wang & Glenn, 2008).

For estimating the maximum height of the canopy in a plot the maximum value within the plot can be used (Weber & Boss, 2009; Coops et al., 2007). Another method used by Coops et al. (2007) is to divide the plot in subplots, determine the maxima within these subplots and use the average of these maxima as the height value for the plot. The height value calculated this way is called dominant height (Coops et al., 2007).

Weber & Boss (2009) also related the quadratic mean (= root mean squared) height to ground measured tree height, although the result was not as good as relating maximum height to ground measured height.

2.3.3 Tree density and crown surface

After distinguishing individual trees, tree density can be calculated as the number of trees per surface area. Tree crown surface for an individual tree can be determined when tree crowns are fully delineated by calculating the surface of the crown (Hyyppa et al., 2008).

2.3.4 Fractional cover

Information about gaps can be retrieved by determining the fractional cover or gap fraction. This is the fraction between the number of grid cells that are not covered by trees down to a certain height and the total number of grid cells (Coops et al., 2007). This method is described for point cloud data, but can also be applied on grid data.

2.4 Point cloud methods

For some forest parameters the analysis method does not differ between point cloud and grid data, for example for determining grid height. However viewing and processing point cloud data requires specific software. A program that can be used to view and process these data is Fusion. It was made for viewing and analysing lidar data (McGaughey, 2010).

2.4.1 Multiple regression

For the analysis of lidar point cloud data a common approach is to use multiple regression. Regression analysis combines a field measured variable and lidar derived variables into an equation that shows the relation between these variables.
As a linear model this can be written as:

Field measured variable = $\beta_0 + \beta_1$ lidar var1 + β_2 lidar var2 + …. β_n lidar var n.

Where lidar var1, lidar var2 represent the lidar measured variables. β0, β1 represent the regression coefficients.
By removing the non-significant variables in the equation only the relevant variables remain. The result is an equation that describes as good as possible the relationship between the lidar measured variables and the field measured variables.
For example in the study of Naesset and Okland (2002) multiple regression was used to estimate the height to the crown. 14 predictor variables were used which include quantiles, maximum values, mean values and coefficients of variation. After stepwise selection of significant predictor variables the quantile that corresponded to the 75 percentile remained as significant variable. The resulting regression equation has an R^2 of 0.61 (Naesset and Okland, 2002).
Lim & Treitz (2004) use a single regression equation for describing the relation between canopy based quantile estimators and aboveground biomass and the biomass components stem wood, stem bark, live branch and foliage. The regression equations used where in the form of:

Field measured variable = β_0 lidar var1 β_1

ln Field measured variable = ln $\beta_0 + \beta_1$ ln lidar var1

Where the measured variable represents the dependent biomass variable and var 1 represent the laser heights of a quantile. The resulting R^2 values range from 0.820 – 0.903. The root mean square error (RMSE) is more variable between the different biomass components and is highest for total biomass (50.17 – 66.65 Mg ha^{-1}) and lowest for foliage (0.87 – 2.00 Mg ha^{-1}) (Lim & Treitz, 2004).

This research was carried out in a mature hardwood forest with different forest management treatments for different forest plots. However the allometry of the trees is about the same in all plots, as the recently carried out management did not influence the allometry of the mature trees. Whether the equations that were found during this research will be usable in other areas, with different tree species and different forest structures is unknown (Lim & Treitz, 2004).

2.4.2 Canopy volume profile

A method that examines canopy structure by looking at closed and open volumes is the canopy volume profile (Coops et al., 2007). This method was developed by Lefsky et al. (1999) for full waveform data, but has been adapted for discrete return data by Coops et al. (2007). For a plot of 20 X 20 m Coops et al. (2007) used subplots with an area of 25 m^2 in which they binned all the lidar returns according to their height in 1 m height intervals. These height intervals were classified into one of the following classes: closed gap, when there is no lidar return and the height is below the highest height value within the subplot. Euphotic, when the lidar return is in the upper 65 % of the canopy, which is defined as the area where most of the photosynthesis takes place and that intercepts most light. And finally oligophotic, when the return is in the lower 35 % of the canopy, where there is less photosynthetic activity (Coops et al., 2007). All volume in a subplot that is above the highest height value in this subplot, but below the maximum height of the main plot is called open gap. A representation of this classification method is shown in figure 2.2 . The final values for the different height classes in the large plot are calculated by summing the number of occurrences from the subplots for every height value (Coops et al., 2007). These values can be compared to forest stand characteristics like crown volume, basal area and stem density (Coops et al., 2007)

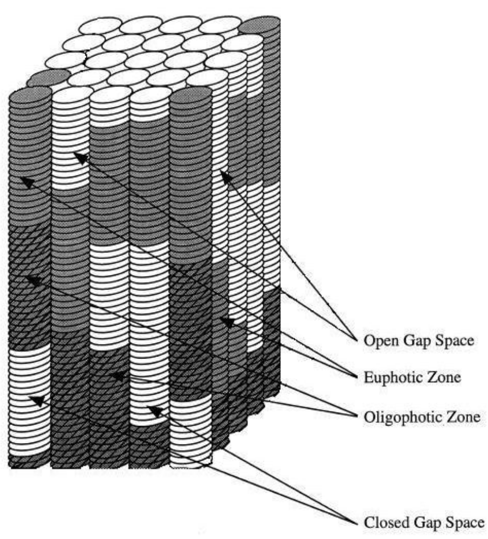


Figure 2.1 Schematic representation of the canopy volume profile method as used by Coops et al. (2007)

2.5 Accuracy assessment

McRoberts et al. (2010) describe accuracy assessment as "a location-specific comparison of a map prediction and a ground observation". To test the accuracy of a prediction, cross validation can be used. The available dataset is split in two parts, one part, the training dataset, is used to make a prediction and the other part, the test dataset, is used to test the predicted result. For example a regression formula calculated on one half of the dataset can be applied on the other half of the data and the differences between the regression and the measured values indicate the validity of the regression formula. Leave one out cross validation is mentioned as a validation method by Van Leeuwen & Nieuwenhuis (2010). It resulted in a root mean squared error (RMSE) as accuracy measurement. The difference with the method with one training and one test dataset is that every observation is used once as test set while the rest of the data is the training set. This gives for every observation an error value and the average of these error values is used as accuracy indicator (Schneider & Moore, 1997).

The coefficient of determination (r^2) and the root mean squared error are commonly used measurements to indicate the accuracy of a linear regression. The root mean squared error is a useful indicator because its value has the same unit as the data. It can for example be compared with average. The coefficient of determination is a value between zero and one and indicates the strength of a linear relationship (Ott & Longnecker, 2001).

To show the result of a categorical outcome an error matrix can be used. Rows in this matrix represent the categories in the reference data. Columns the categories in the calculated outcome (McRoberts et al., 2010). The values in the matrix show the occurrences of the different categorical combinations. On the diagonal are the cases that are both in the reference and the calculation in the same category, thus the correctly classified cases. The off-diagonal values are misclassified (McRoberts et al., 2010). Overall accuracy is derived by dividing the sum of the diagonal values by the total number of observations. Furthermore omission and commission error can be derived. Omission error or shows the percentage of the observed class that is not predicted as this class. Commission error is the percentage of a predicted class that is not observed as this class (McRoberts et al., 2010).

3 Materials and Methods

3.1 Study area

The study area for this thesis research is a forest area located at the Utrechtse Heuvelrug (figure 3.1). This area was chosen because it is part of one of the larger forested areas in The Netherlands and has both AHN-2 and forest inventory data available.

The study area can be described as a 10 by 10 kilometres square, in which about 60 % of the area is either not covered by lidar data yet or does not consist of forest. The remaining area covered by forest is about 3900 ha.

A forest inventory conducted in 2006 for 740 ha of this study area shows the species composition of the forest. According to the basal area distribution 73 % of the forest is coniferous and the remaining 27 % is deciduous. The basal area of a tree is the stem surface area at breast height.

The most abundant species is Scots pine (*Pinus sylvestris*), which covers 40 % of the total basal area (Lusink et al., 2006). Other important coniferous tree species are European black pine (Pinus nigra), douglas fir and larch (Larix decidua and Larix kaempferi). The most important decidous species are oak (Quercus robur), beech (Fagus sylvatia) and birch (Betula pendula). It is estimated that 42 % of the forest area is forest with multiple tree species mixed and the remainder is forest with only one tree species. The average age of the forest is 67 years and the oldest forest stands are about 160 years (Lusink et al., 2006). The forest is located on a poor sandy soil (Alterra, 2011), which has its influence on productivity.

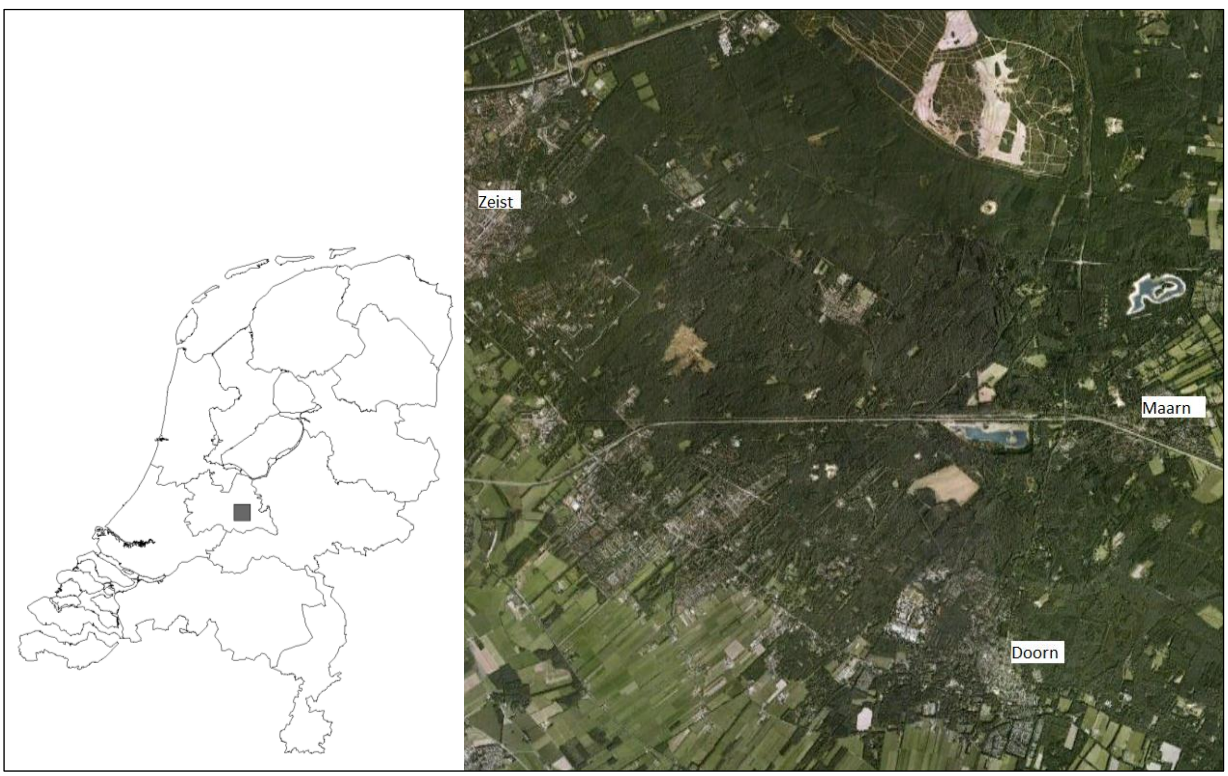


Figure 3.1 Location of the study area

3.2 AHN lidar data

3.2.1 AHN in general

In the Netherlands, lidar data is collected as part of the AHN (Actueel Hoogtebestand Nederland) program in which a nationwide detailed height map is produced. This is done on the order of the water boards (waterschappen) and the ministry of water management (Rijkswaterstaat) (Van der Sande et al., 2010). It was originally collected for water management, but it is also used for other applications like archaeological research and infrastructural planning (AHN, 2011).

The first height map (AHN-1) that was finished in 2003 has an accuracy of 16 cm and a point density that varies from 1 point per 16 m^2 up to 1 per m^2 (AHN, 2010). The second dataset (AHN-2) that is collected over the period 2007 - 2012 will have a mean point density of 6 – 10 points per m^2 (AHN, 2010). The end result will be a dataset with a maximum allowed systematic error of 5 cm and a maximum stochastic error of 5 cm. This means that at least 68.2 % of the points has an accuracy of maximum 10 cm and 95.4 % has an accuracy of maximum 15 cm (AHN, 2010). Since the AHN is meant as a terrain height model, all non-ground level points are filtered out in the end result. There are however some tolerated errors in the filtering procedure.

For objects on surface level only one object per 10.000 ha is allowed to show a mistake. For vegetation higher than 0.5 m a maximum of 1 ha per 10.000 ha is allowed to contain filtering mistakes and for vegetation below 0.5 m one ha per 5.000 ha is allowed a filtering mistake (AHN, 2010).

The data is collected in the period of the year for which the vegetation does have the least influence on the measurements. This means the data collection roughly takes place in the period from the start of December until the end of March.

There are several datasets produced out of the collected data. These include the filtered ground level data and the data that was filtered out, delivered as point data as well as an interpolated grid. There is a 0.5 m grid and a 5 m grid (AHN, 2010).

Not all data is already available as data collection continues until March 2013. Which data is available and when the rest of the data will become available is shown in figure 3.2.

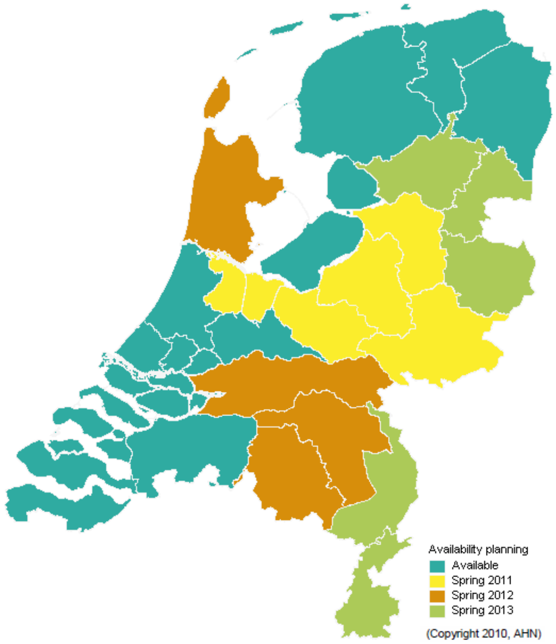


Figure 3.2 Availability planning for acquisition of AHN-2 lidar data over the Netherlands

3.2.2 AHN-2 data types

The AHN-2 data is delivered in two formats, as shown in table 3.1 (AHN, 2010). The main difference between the two file types is that point clouds can have several points in one location at different heights, while the grids have only one height value at a location. This point cloud data furthermore can have reflectance values for the different points. Because of this, the two different file types require different analysis techniques.

The 0.5 meter grid DTM and the 0.5 meter grid DEM will be used in this thesis. This data is delivered in tiles with a size of 1000 X 1250 m. The numbering of these tiles is according to the numbering of the national topographic maps, where one topographic map unit is divided in four separate sub-units, which are further subdivided in 25 tiles. The ADF-grid format can be imported in ArcGIS and can be stored as a standard grid format.

Table 3.1 Lidar data types delivered by the AHN-2 program

File description	File type
Filtered point cloud (DEM, no objects in terrain)	ASCII-XYZ
Remainder point cloud (DTM, with objects in terrain)	ASCII-XYZ
0.5 meter grid interpolated DEM (no data is filled by interpolation)	ADF-grid
0.5 meter grid not interpolated DEM (no data remains as data gap)	ADF-grid
0.5 meter grid DTM	ADF-grid
5 meter grid DTM	ADF-grid

3.3 Reference data

For assessing the results of estimating forest parameter from lidar, reference data have to be used. Three datasets that include forest information and can be used for this purpose were available for the study area. The most recent national forest inventory is the "Meetnet functievervulling" (MFV). This national inventory includes 3622 sample plots distributed over forest areas in The Netherlands, which resulted in a density of one sample plot per 100 ha of forest (Directie Kennis, 2006). Within the study area 32 plots are located as shown in figure 3.3. It includes information on canopy height, canopy cover, forest age, main tree species, number of trees per ha and timber volume per ha. This data was collected during the period 2001 – 2005. MFV plots form circles that include at least 20 trees with a DBH≥ 5 cm, unless the average diameter is less than 5 cm. When a plot circle crosses the border of a forest stand, this part is subtracted from the circle (Daamen & Dirkse, 2005).

A more specific reference dataset that falls within the area of interest is the result of a forest inventory executed for Staatsbosbeheer for one of their forests called Austerlitz. 179 plots are distributed in a regular grid over an area of 740 ha (figure 3.3). This inventory has been executed in 2006 (Lusink et al., 2006). This dataset gives detailed information on this specific forest area, so is only useful for the test area. These two datasets have 13 different tree species as main tree species. Their English, Dutch and scientific names are given in appendix 1.

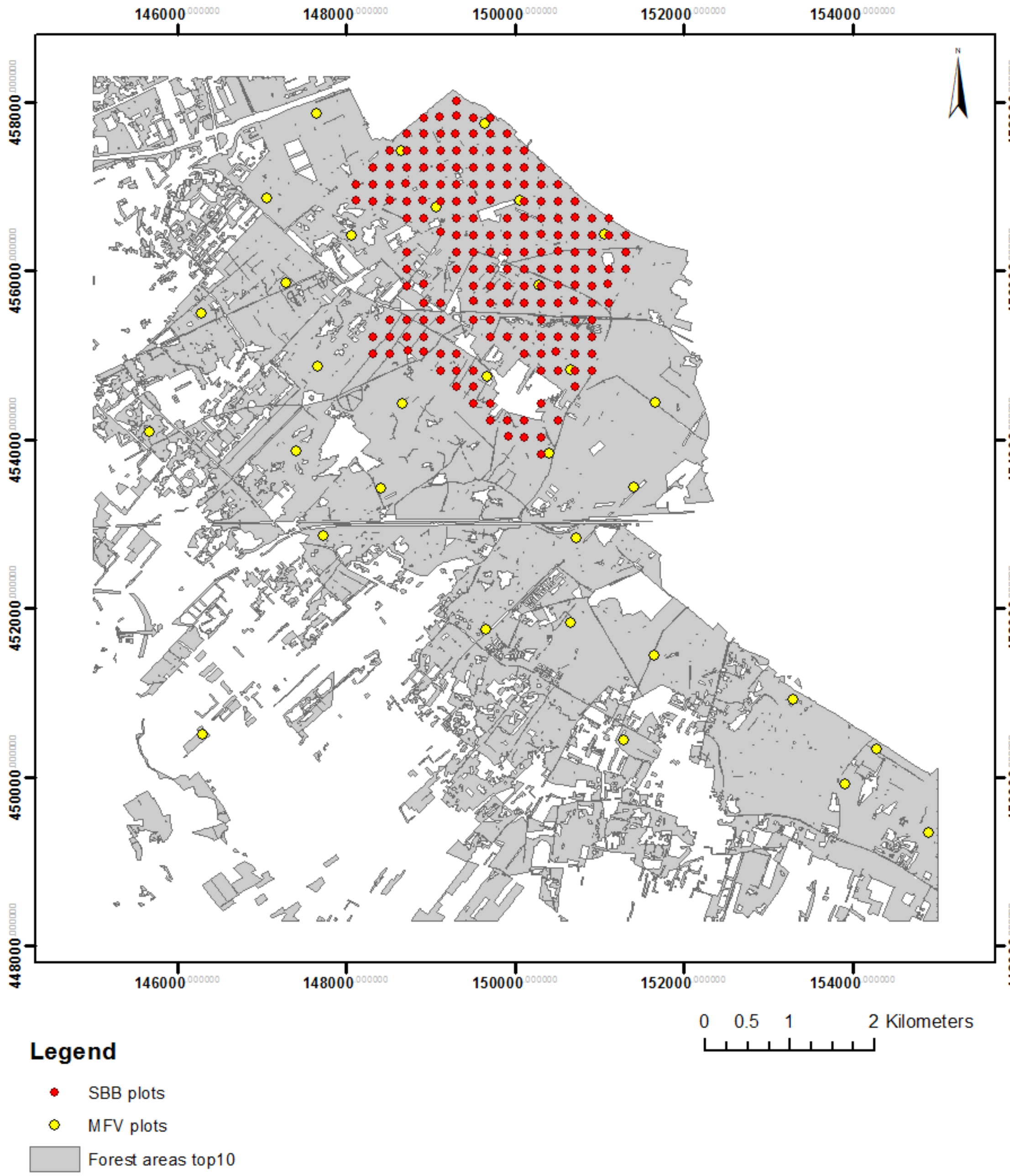


Figure 3.3 Locations of reference plots in the study area

Another national dataset is a polygon map of different forest stands, which allows to delineate the stands. This map is a result of a GIS project that digitized the maps of the fourth national forest inventory (1980 - 1983) (Clement, 2001). Amongst others, the map shows the main tree species and age of the different forest stands. When combining the MFV dataset and this map there are some dissimilarities. One of the reasons can be a mismatch between both maps. An overview of the variables that are included in the different datasets is given in table 3.2

Table 3.2 Variables in the three datasets

Characteristic	Meetnet Functievervulling	Staatsbosbeheer data (SBB data)	Fourth national forest inventory
Year of data acquisition	2001-2005	2006	1980-1983
Regeneration year	X	X	X
Main tree species	X	X	X
Dominant height (M)	X	X	X
Species composition		X	
Diameter class per species		X	
Nr trees alive (N/ha)	X	X	
Nr trees dead (N/ha)	X	X	
Nr trees dead lying (N/ha)	X	X	
Stem basal area (m^2/ha)		X	
Growing stock (m^3/ha)	X	X	X
Volume dead standing (m^3/ha)	X		
Volume dead lying (m^3/ha)	X	X	
Net annual increment (m^3/ha/y)	X	X	
Diameter middle tree (cm)	X		
DBH mean (cm)	X	X	X
Crown cover	X		
Density class		X	
Management type	X		X
Forest type	X		X
Soil type	X		

One of the factors that have to be taken into account is the collection date. The polygon dataset of which the information was collected between 1980 and 1983 can only be used for general reference and will have lost a lot of its informative value.

Within the variables that were measured in both datasets, there are some differences in the way these are presented. For both the MFV and the SBB data tree height is measured by measuring one or two trees per plot. The way these height values are presented differs. Field measured height for the MFV dataset is given as the height of the highest tree within $100m^2$ (Daamen & Dirkse, 2005). In the SBB dataset both height measurements are given. For usability in this thesis these height values have been averaged to have one height value per plot.

Also both the MFV and SBB give a measure of forest density. In the MFV dataset this is done by estimation in the field and in the SBB dataset this is derived from the basal area per hectare. The differences between these two forest density indicators is discussed in more detail in 3.5.2.

Some variables within the reference data have not been measured directly, but have been indirectly derived, for example by allometric relationships. An example of such a variable is timber volume. Within the reference data these relationships can be recognized because the directly measured variable is correlated to the indirectly derived variable. The possibility to calculate a variable indirectly from another variable can be useful in the context of this thesis. For example a forest characteristic that cannot be measured directly with lidar, can be indirectly estimated when this variable is related to a variable that can be measured with lidar.

Within the reference data the best relationships between field measured values and indirectly derived values are between stem diameter and volume (figure 3.4) for both reference datasets. Furthermore in the SBB dataset there is a relationship between basal area and growing speed (figure 3.5).

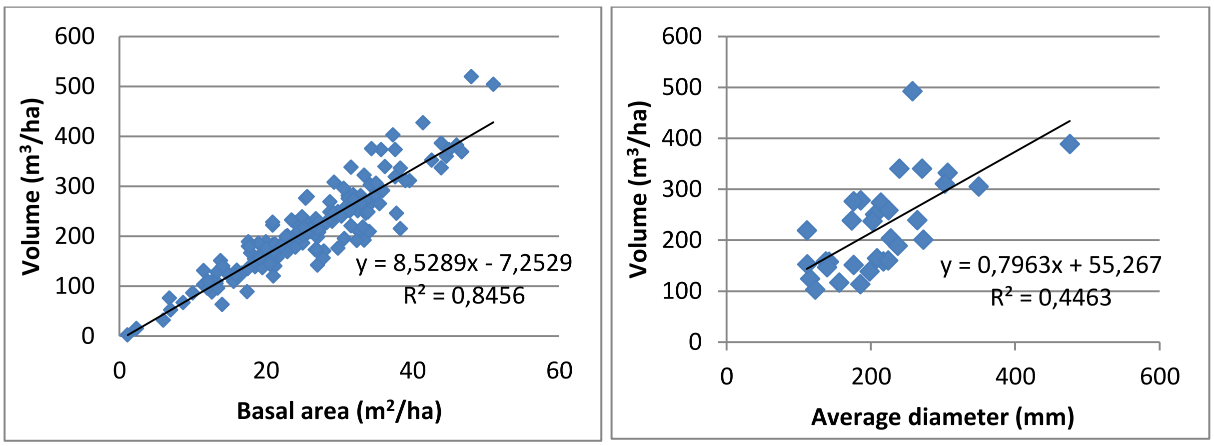


Figure 3.4 Relationship between stem diameter and volume within the SBB and MFV reference data
Relationship between basal area and timber volume for the SBB data (left)and for the MFV data between average diameter per plot and volume (right).

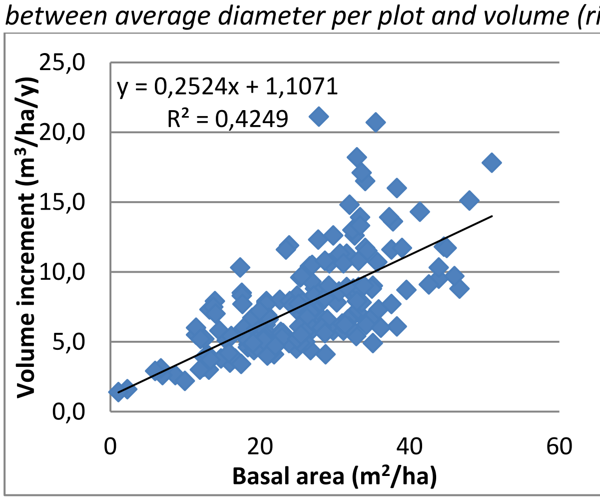


Figure 3.5 Volume increment per year plotted against basal area per ha

3.4 Data preparation

3.4.1 AHN data

In this thesis three different AHN datasets were used: 1) The AHN2 digital terrain model (DTM), 2) the AHN-2 digital elevation model (DEM) and 3) the DEM of AHN-1. The AHN-2 data were collected in 2008 and the AHN-1 data between 1998 and 2000 (AHN, 2010). The AHN-2 DTM was delivered as 77 separate tiles in the ADF format. These have been converted to an ArcGIS grid and were merged using the mosaic function. The AHN-2 DEM was delivered as a geo-referenced image file and has also been converted to an ArcGIS grid.

By a first visual inspection the AHN-2 DEM showed quite a lot of data gaps in the forested areas (figure 3.6). This dataset is thus the non-interpolated DEM as described in 3.2.2. The AHN-1 DEM model was used to fill the gaps by applying the mosaic to new raster function. This function combines the two datasets in such a way that AHN-2 information is used in the areas where both datasets are available and the AHN-1 DEM is only used at the locations where AHN-2 is not available.

For using the AHN-1 DEM it has to match with the AHN-2 DEM. The AHN-1 DEM was delivered as a geo-referenced image file. It has a grid cell size of 5 meters and the AHN-2 data of 0.5 meters. The AHN-1 file was resampled to a 0.5 meter grid, so it matches the AHN-2 files. Furthermore height in this DEM model was given in centimetres, whereas the AHN-2 files are in meters. By dividing the cell values by 100 and taking care of the right data type which is floating point, an ArcGIS grid with the same properties as the AHN-2 grids was created.

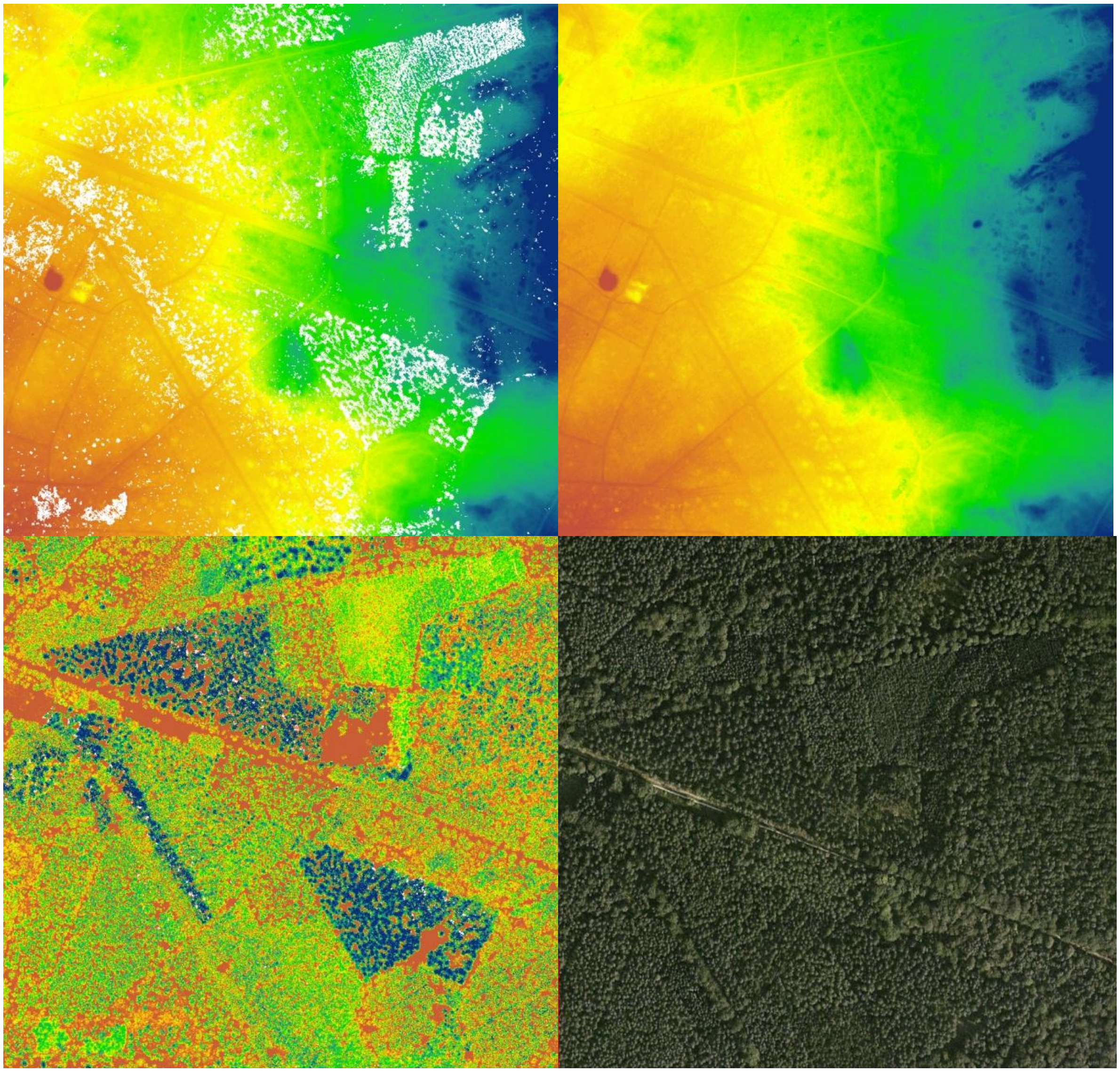


Figure 3.6 Examples of the different AHN 0.5m datasets and an aerial picture for comparison. *This area is about 750 X 700m. Top left: 0.5m AHN2 non filled DEM. Top right: Filled AHN2 DEM. The former two images show height relative to sea level. Bottom left: final 0.5m grid as used for analysis. The colours represent height in relation to ground level, where red colours represent low values and blue colours represent high values. Bottom right: aerial picture for reference.*

The AHN-2 DTM shows the heights of objects as the height above sea level (in The Netherlands NAP). To get a dataset that contains only the heights of the objects in the terrain the DEM has been subtracted from the DTM. This is the final grid that will be used for analysis. A detail of this grid is shown in figure 3.6. Different forest types are visible, by differences in colour, which represents height, and different densities of trees. Also open spaces are visible.

This grid still contains non forested areas. The Top10 topographic map can be used to clip out only the forested areas. In the Top10 dataset the categories deciduous forest, coniferous forest and mixed forest were selected. A drawback of this method is that all small paths within the forest will also be removed from the data. To get a more continuous dataset, a buffer of 3.5 meter with the option dissolve all has been used. The same buffering method has been used by Dirkse en Daamen (2000) to create a map of the forest areas in The Netherlands. The resulting shapefile will be used to clip the AHN data, so only the forested areas remain.

3.4.2 Reference data

The reference data from the MFV and SBB data has been received as DBF files. These have been imported into ArcGIS and where stored as point shapefiles. For the MFV dataset only the plots in the study area were selected out of the nationwide dataset.
The MFV and SBB reference datasets are split into two parts, a training and a test dataset. This is for both datasets done by splitting into odd and even plot numbers. The resulting datasets are well distributed over the area (appendix 2). For the SBB data, the training dataset consists of 88 plots and the test set consists of 90 plots. For the MFV dataset both the training and test dataset consist of 16 plots. The distribution of main tree species over these datasets is shown in table 3.3.

Table 3.3 Number of plots per tree species in the training and test datasets

	SBB training	SBB test	MFV training	MFV test
Red oak	5	4	1	2
Oak	3	3	2	
Birch		1	1	
Beech	1	6	1	1
Deciduous	9	14	5	3
Corsican pine	10	12	1	1
Douglas fir	8	11		2
Norway spruce	4			
Scots pine	48	45	8	8
Japanese larch	4	6	1	1
Austrian pine	4	2	1	
Western hemlock	1			
Port Orford cedar				1
Coniferous	79	76	11	13
Total	88	90	16	16

3.5 Applied methods

3.5.1 Height

The maximum and dominant height as described in 2.3.2 have been calculated. First linear regression equations will be calculated for the training dataset. These equations will be applied on the total forested area and the validity of the outcomes will be tested by using the test dataset.

To obtain the maximum value within a plot, the zonal statistics function was used. The reference plots were used as zones, and the statistics type was set to maximum. The maximum height values for the plots were linked to the field measured height values and a trend line was fitted.

The dominant height uses subplots within one plot. To achieve this, a 5 meter maximum height grid was created within the areas where plots were present. This is done by using the aggregate function. Zonal statistics was applied to calculate the mean value within a plot from this grid. As the size of the plots varies, the number of 5 meter grid cells also varies. The resulting dominant height values were treated the same as the maximum height values.
As a result, there are four regression equations, for both reference datasets an equation for maximum height and for dominant height.

To determine the height over the total forested area these equations are imposed on the digital terrain model. For the maximum height formulas the 0.5 m grid is aggregated in a maximum height grid, with a grid cell size of 25 meters. The regression equations are applied on this grid. For dominant height first a 12.5 m maximum height grid will be created. Calculating the mean of four 12.5 m grid cells results in a 25 m grid that contains dominant height. On this dominant height grid the equations for dominant height will be applied.

Comparing the resulting values with the measured height values of the test dataset gives a measure of validity of the equations. The difference between field measured values from the control dataset and the results from the regression equation is determined. The difference between these should be as low as possible.

3.5.2 Canopy openness and forest cover

Canopy openness is defined here as the percentage of cells that do not belong to the forest canopy. When the difference between DEM and DTM is below a pre-defined value, this is considered a gap. For example a height difference of 5 meters means that all cells that are below 5 m are considered gaps. Different heights are chosen for analysis, in order to take into account the effects of undergrowth in the forest. In ArcGIS this has been calculated in different ways:

- On plot scale as the number of cells within the MFV and SBB plots with a value below 1, 2, 3, 5 and 10 m respectively divided by the total number of cells within the plot.
- On the complete study area as the number of cells within a 25 m grid cell with a value below 1, 2, 3, 5 and 10 m respectively, divided by the total number of 0.5 m cells within the 25 m grid cell.

This will be calculated by combining the reclassify and zonal statistics as table tools. This results in tables that provide for a plot or a 25 m grid cell the number of cells within the chosen height class and the total number of cells within the plot or chosen grid cell. Dividing these numbers gives the gap fraction.

Both reference datasets do not include gap fraction but provide a measure of crown cover. To go from gap fraction to crown cover is done by subtracting the gap percentage from 100 %. This gives a crown cover percentage, which can be classified into the MFV classes for comparison. An error matrix is used to show the similarity between the reference and the lidar derived crown cover percentages. As the MFV dataset has only 32 reference plots within the study area and this method does not require the establishment of a regression equation, all the plots were used for creating the error matrices.

The MFV dataset includes crown cover as percentage classes. These classes are:

- 0 %
- 0 – 0.1 %
- 0.1 – 1 %
- 1 - 5 %
- 5 - 10 %
- 10 – 25 %
- 25 – 50 %
- 50 – 75 %
- 75 - 90 %
- 90 - 100 %

The SBB dataset does not have a crown cover class. Instead the degree of openness is estimated based on the basal area per hectare. The classes present are spacious, optimal, moderate dense and very dense. Different groups of tree species are distinguished, which all have their own class values (Lusink et al, 2006) (table 3.4). In this dataset not all the plots have a forest cover value. 36 of the 178 plots do not have an openness class assigned.

Table 3.4 Degree of openness related to basal area for different tree species classes
Tree species classes as used by Lusink et al. (2006)

Tree species group	Class	Basal area (m^2 / ha)
Scotch pine, larch, birch, oak, Northern red oak,	Spacious	<18
	Optimal	≥18 - < 21
	Moderate dense	≥ 21 - < 24
	Very dense	≥ 24
Douglas fir, Pinus nigra, Picea, Abies, Tsuga	Spacious	<25
	Optimal	≥25 - < 28
	Moderate dense	≥ 28 - < 33
	Very dense	≥ 33
Beech	Spacious	<25
	Optimal	≥25 - < 30
	Moderate dense	≥ 30 - < 37
	Very dense	≥ 37

By reversing this approach, it should be possible to estimate the basal area by forest cover, although according to this method tree species should be taken into account. There is a near linear relationship between canopy cover and basal area, but this works only for relatively young managed forests and does not work for mature natural forests (Jennings et al., 1999). To test this relationship, scatter plots of crown cover against basal area have been made for visual assessment. By fitting a linear regression this relationship has been quantified as a regression equation. This linear equation has been applied on the complete study area, which results in a map showing basal area for the complete study area. Canopy cover estimations can also be used to estimate forest stand volume (Jennings et al., 1999). This probably also

works for the SBB reference dataset because basal area is related to forest cover, and basal area and timber volume are related. Forest cover percentage will be used to estimate volume.
In the SBB reference data, the classes for degree of openness are not specified as crown cover percentage so an error matrix cannot be used. Instead a table is created where the degree of openness class is related to the crown cover percentage.

3.5.3 Canopy profile

The canopy volume profile method as described in 2.4.2 delivers information about canopy structure, which can be related to stem density, canopy volume and basal area. The general idea of this method will be applied on the AHN-data. The AHN data in grid format has as a drawback that there is only one data point available at one location, whereas with point cloud multiple return data, there can be more data points. This is a relevant difference for this method because now only information about the top of the canopy is available, the so called first return. For example open spaces below the top of the canopy will not be noticed. An adapted method has been developed which uses the height values within a plot and bins these into height classes. The method of Coops et al. (2007) uses 1 meter height intervals, which are plotted as canopy profile with height interval plotted against coverage value. It is expected that these canopy profiles will give information about canopy structure. A completely closed flat canopy will give as result that there are only data points in the upper part of the canopy. A more open canopy with height differences will result in a more diverse height distribution of data points. There might be a difference between deciduous trees with a rounder crown, and coniferous with a more pointed crown.

An overview of the developed method is given is figure 3.7. Dividing the 0.5m AHN grid by the maximum height is done to get height fractions instead of a height value in meters. By using a percentage instead of actual height, plots can be compared better. When profiles are plotted in a graph they begin and end at the same value when using height percentages. When using actual height in meters this is not the case, as forest height differs per plot.
In this method, each interval has a height of 10% of the maximum height within the plot. This results in 10 height intervals per plot. The number of data points within an interval are determined.
Another processing step on this dataset is dividing the values of the intervals by the total number of grid cells within the plot. The result is that every interval has a value between 0 and 1. This allows better comparison of canopy profiles. The resulting coverage values per height interval are plotted. Based on these plots a classification method was developed. The development of this method is described in the results chapter in 4.1.4 alongside the canopy profile plots.

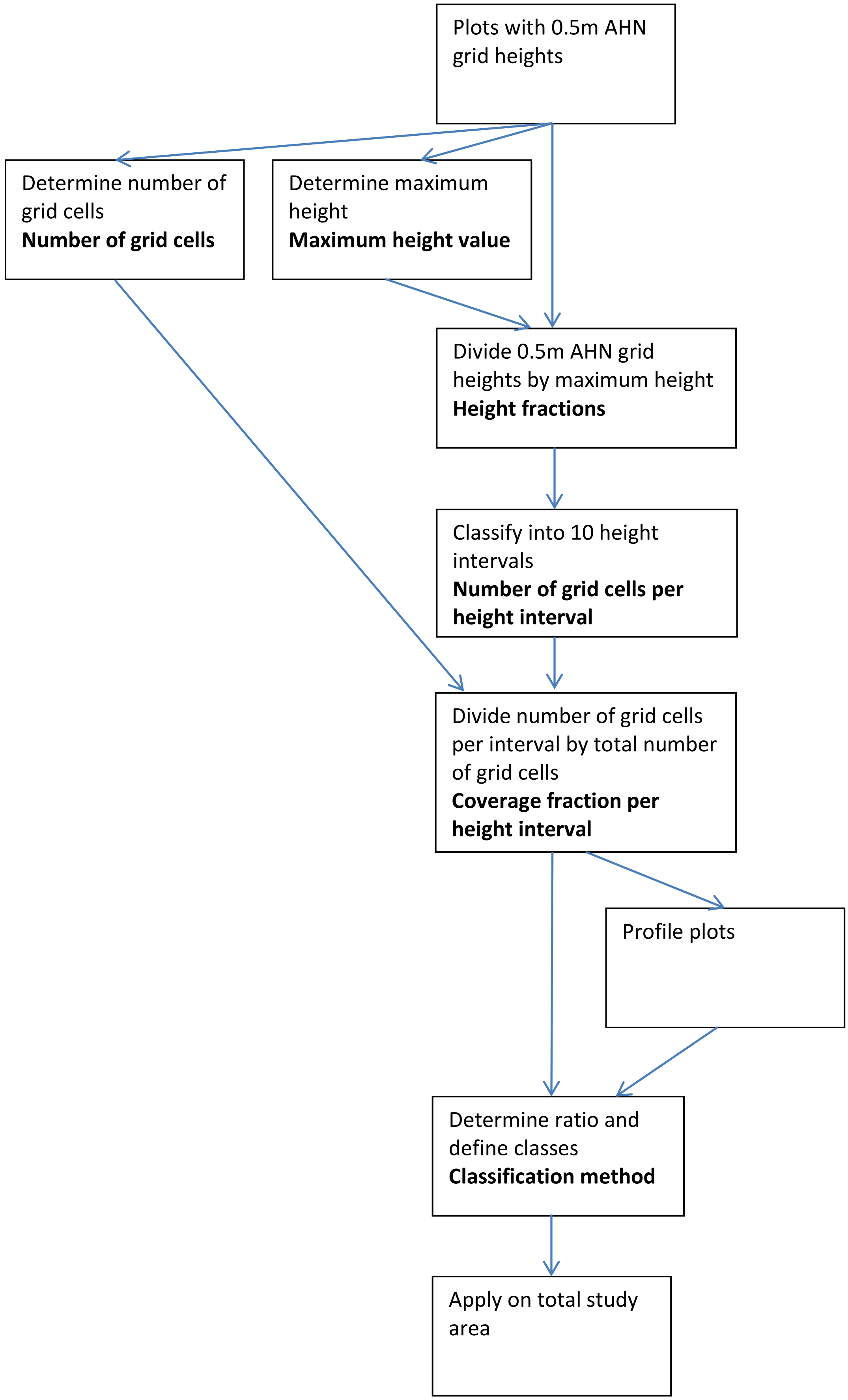


Figure 3.7 Schematic representation of the canopy profile method

4 Results

4.1 Plot scale

4.1.1 Height

Figures 4.1 and 4.2 show the relationship between maximum height and field measured height and dominant height and field measured height. A linear trend line has been fitted through the data and the equations and R-squared values are shown in the graphs.
For the MFV data one outlier was removed. This is a plot in an area, that was regenerated in 1999 and measured in 2002 as a height of 1 meter. When lidar scanning took place in 2008 the maximum height in this plot was 16 m and the dominant height 7 meters.
The gradients of the equations show that there is in general a one to one relationship between lidar measured and field measured heights, except for the maximum height from the SBB plots, where the gradient is 0.62. The intercept is larger for the SBB plots, around seven and four respectively, than for the MFV plots, where these are 2.4 and 1.3.

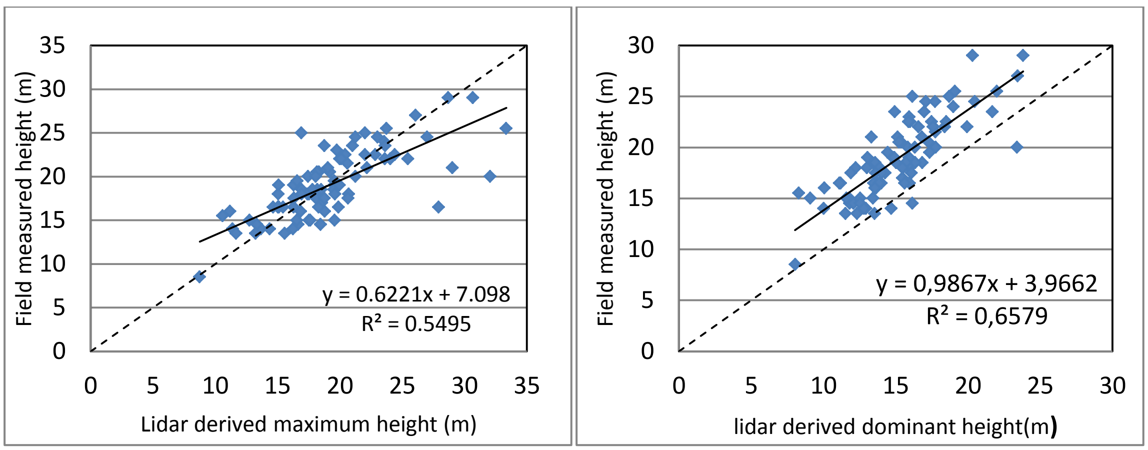


Figure 4.1 Maximum height (left) and dominant height (right) against field measured height for the SBB plots

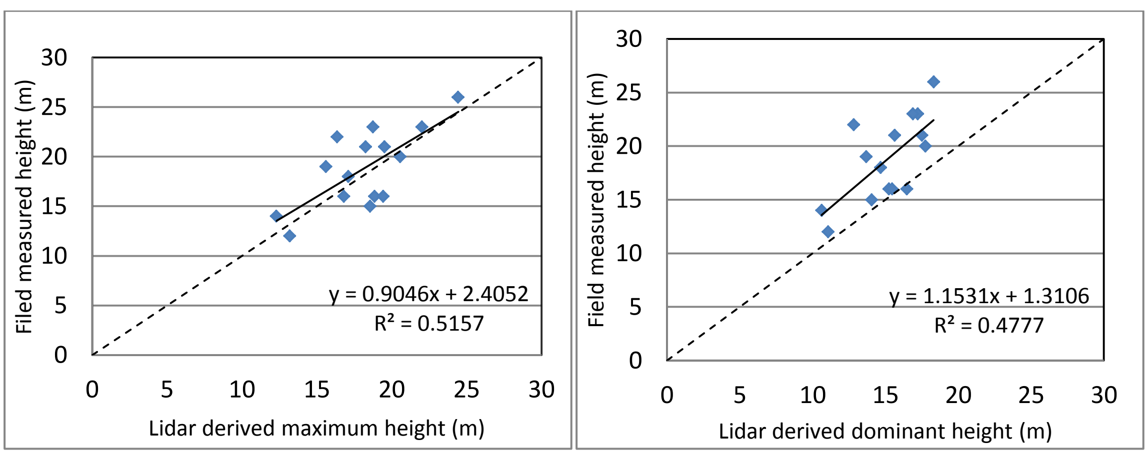


Figure 4.2 Maximum height (left) and dominant height (right) against field measured height for the MFV plots

Figures 4.3 and 4.4 show the relationship between maximum height and field measured height and dominant height and field measured height for the different tree species. The MFV dataset has so few data points that instead, the categories deciduous and coniferous have been used (figure 4.4). For the SBB dataset the coniferous and deciduous plots are shown in separate graphs for clarity. The offset to the y = x line is larger for the dominant height than for the maximum height. In the dominant height graphs the field measured values are larger than the ones calculated from the lidar data. This can be explained by the way the values were derived. Dominant height is calculated as an average value, as described in 3.5.1, which results in values lower than the maximum height of the trees and lower than the field measured height.

In the dominant height graph for the SBB deciduous plots (figure 4.3 top right), there is one plot that can be considered an outlier. This plot does not show a positive offset, but is located below the y = x line. A reasonable explanation could not be found although it was found that this plot does not consist of red oak, but mainly of douglas and scots pine. That the plot is classified as red oak might be because the surrounding forest does consist of red oak. The Japanese larch plot in the dominant height graph (figure 4.3 bottom right) that also is below the y = x line has two highly dissimilar field measured height values. A Japanese larch of 25 m height and a beech tree of 15 m height were measured. The beech tree was probably growing in the understory or in a gap. Because these two values were averaged the field measured height does not show the height of the top of the canopy.
The highest tree species are douglas and Japanese larch. In the SBB maximum height graph the field measured values are lower than the lidar derived values for these tree species. Because of these douglas and japanese larch plots, the trendline in figure 4.1 is affected. By removing these values the trendline follows the y = x line better.
The dissimilarity between field measured and lidar derived values in these plots can be attributed to the field measurements as trees were measured that did not belong to the upper part of the forest canopy or were not the main tree species.

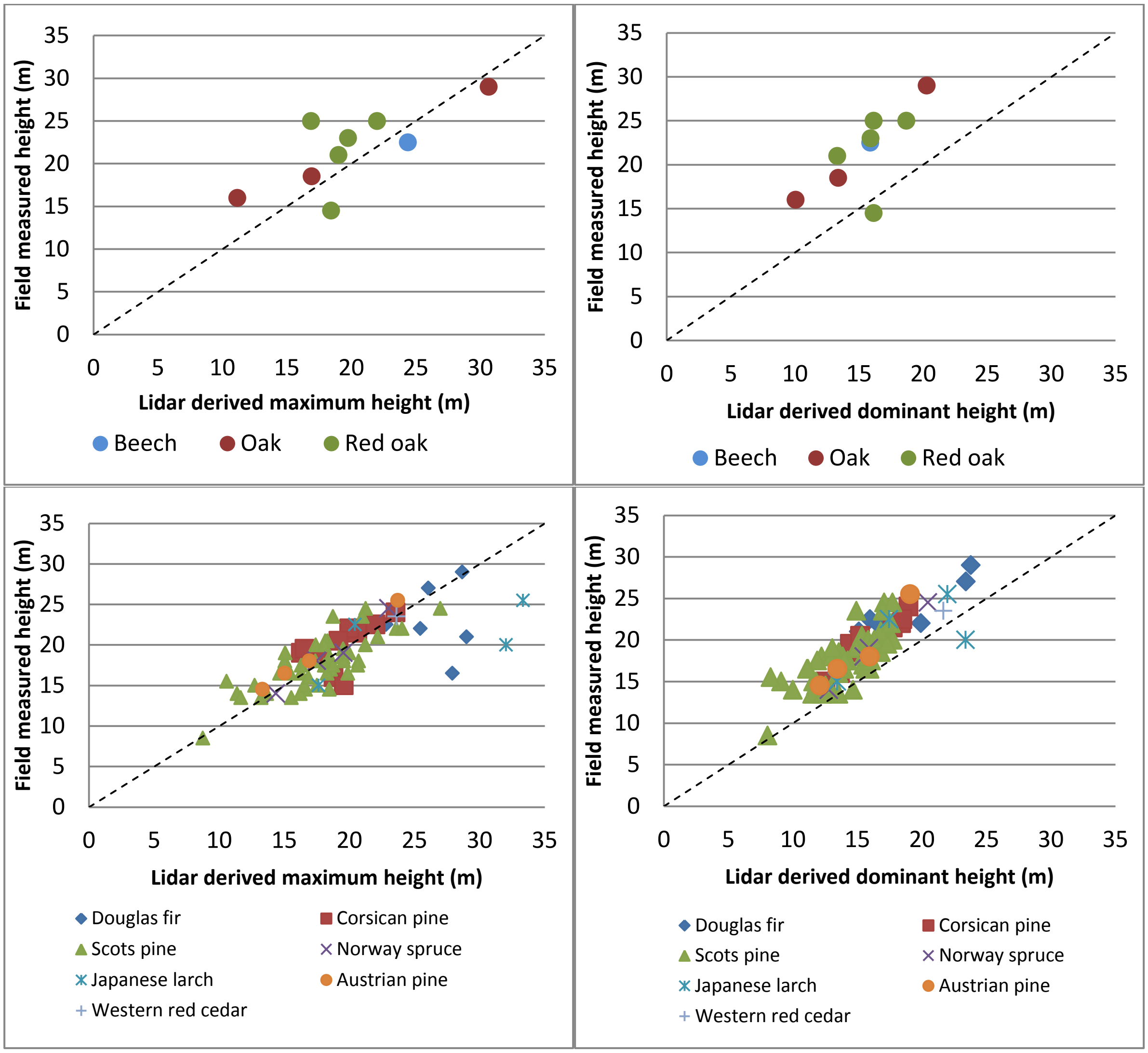


Figure 4.3 Maximum height (left) and dominant height (right) against field measured height for the SBB plots
Deciduous species are shown in the upper and coniferous species in the lower graphs.

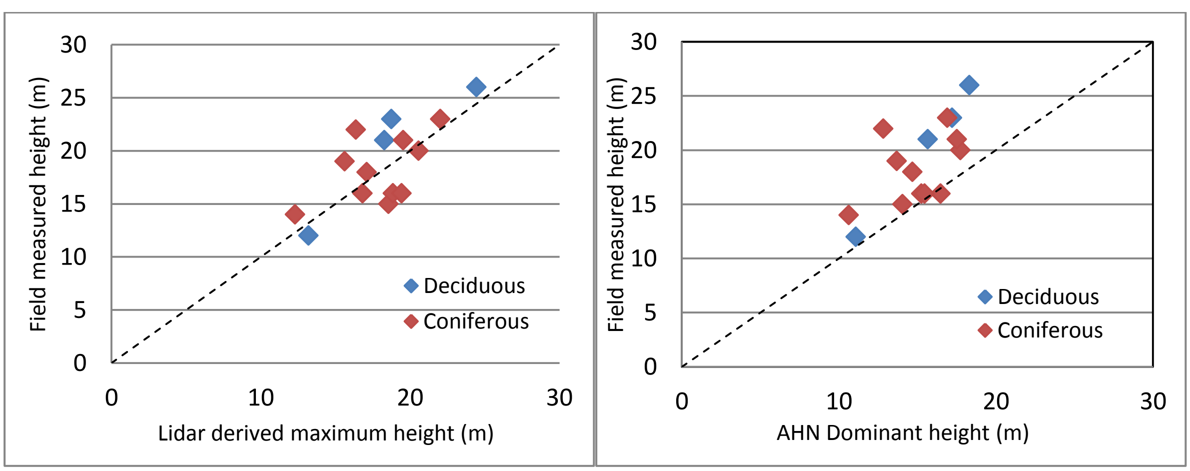


Figure 4.4 Maximum height (left) and dominant height (right) against field measured height for the MFV plots.
Deciduous and coniferous plots are shown as separate categories.

4.1.2 Gap fraction and forest cover

After calculating forest cover as described in 3.5.2, error matrices for the MFV dataset were created. These show for the MFV plots how well the classes estimated in the field (reference classification) match the lidar derived classes (map classification). Table 4.1 is the error matrix for the MFV 1m dataset. In this case for a grid cell to be considered as part of a gap, the height difference between DEM model and DTM should thus be smaller than 1 meter. Error matrices for the other height classes are included in appendix 3. Comparing the canopy openness at different heights shows that with an increasing height, the forest cover decreases, so the gap fraction is higher. For example in the MFV 1 meter error matrix with a reference classification of 90-100%, there are no plots in the classes lower than 50%. In the 3 meter height class there are 3 plots and in the 10 meter height class there are 6 plots with a lidar estimated cover lower than 50 %.

Table 4.1 Error matrix for the MFV dataset

A gap is defined as a difference smaller than 1 m between DTM and DEM.

MFV 1m	Map classification					
Reference classification	25-50%	50-75%	75-90%	90-100%	Total	Omission error
25-50%	2	2			4	50,00%
50-75%		5	5	1	11	54,55%
75-90%		1	2	2	5	60,00%
90-100%		5	5	2	12	83,33%
Total	2	13	12	5	32	
Comission error	0,00%	61,54%	83,33%	60,00%		

Overall accuracy can be calculated as the number of correct classified plots, the sum of the values in the error matrix diagonal, divided by the total number of plots. The overall accuracy decreased with increasing height as shown in table 4.2. The one and two meters height difference forest cover estimations give the best match with the MFV reference data, although the overall accuracy is only 34.38 %.

Table 4.2 Overall accuracy of forest cover classification for the MFV plots

Height	Overall accuracy
1m	34,38%
2m	34,38%
3m	31,25%
5m	28,13%
10m	9,38%

The error matrices can also be shown as a scatterplots. Figure 4.5 shows these for the height differences of 1 and 5 meters. For the other height differences the scatterplots can be found in appendix 4. The advantage of visualizing the data this way is that the degree of error is visible, as well as the degree of overlap between classes.

Also in these graphs it is visible that with an increasing height difference the lidar derived crown cover decreases.
In all MFV graphs the 90 – 100% cover class has in general lower crown cover values than the 75 – 90 % cover class.
For the SBB dataset forest cover percentage per openness class is shown in graph 4.6 for the 1 and 5 meter height differences. The graphs for the other height differences are included in appendix 5. In these graphs there is overlap between the classes. In the 1, 2 and 3 meter graphs still some increase in forest cover at the lower parts of the graph is visible. In the 10 meter graph there is hardly a relation visible between forest cover and forest cover class. Visualizing the different tree species groups shows that these do not appear to be clustered.

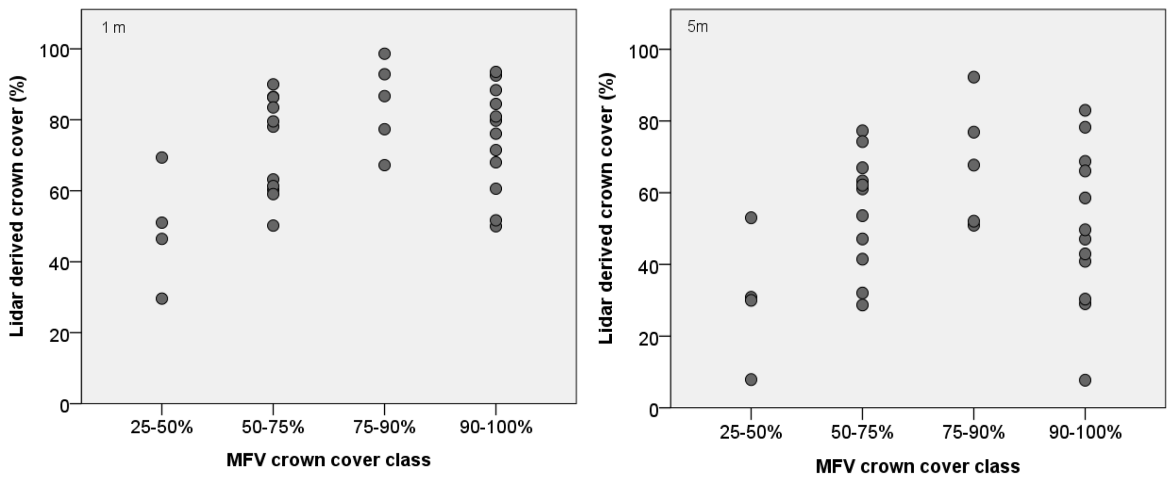


Figure 4.5 Scatterplots showing the distribution of forest cover values (y -axis) within the MFV assigned classes (x-axis) for the different height difference values.

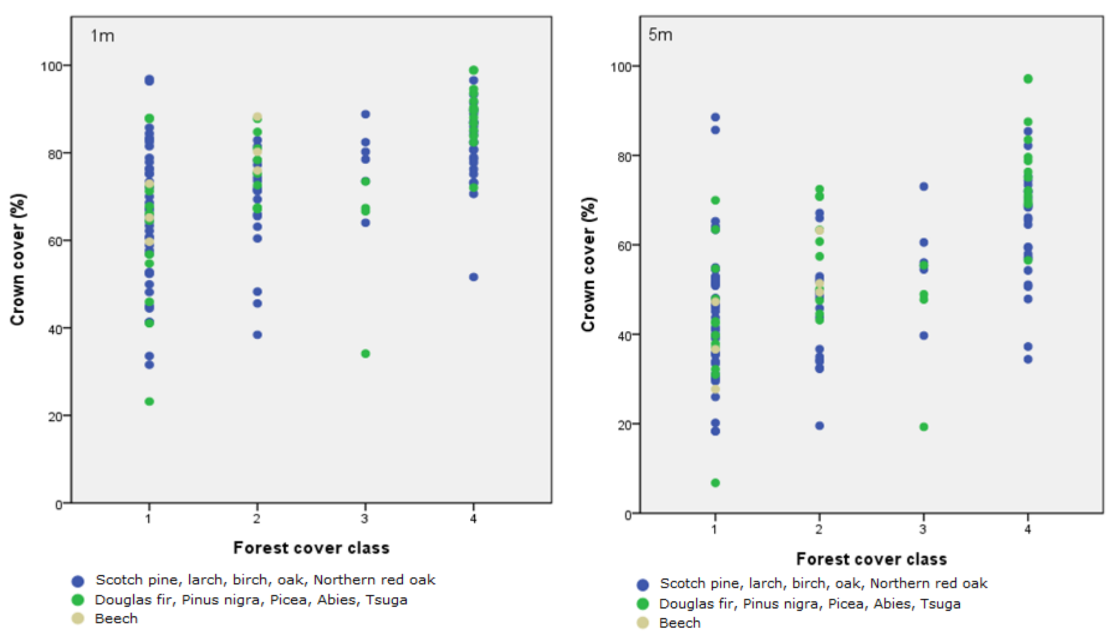
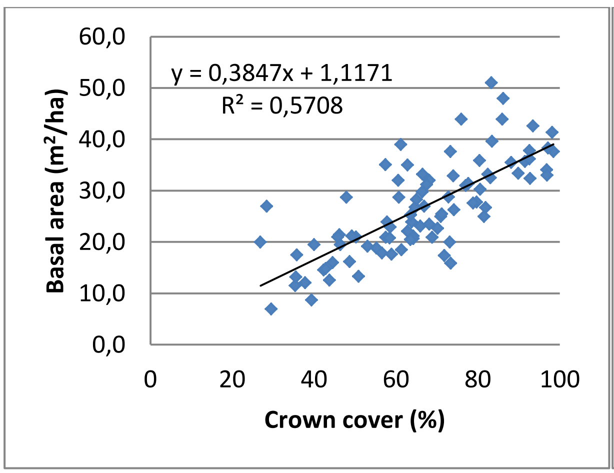


Figure 4.6 Forest cover percentage for the different forest cover classes
Lidar derived forest cover percentage (y-axis) for the different forest cover classes (x-axis) in the SBB dataset. The tree species groups are according to table 3.4.

4.1.3 Relating forest cover to basal area and timber volume

For the SBB dataset crown cover at the height difference values 1, 2, 3, 5 and 10 m has been related to timber volume and basal area. The MFV dataset does include information about volume, but does not include information on basal area. Scatterplots showing for the SBB training dataset the relationship between crown cover and timber volume and between crown cover and basal area are shown in figure 4.7 and appendix 6 This figure shows the relationships where coverage with forest is assessed at three meters above the ground. The graphs for the other heights are in the appendix.

Forest cover at a height difference value of 3m seems to give the best linear relationship between crown cover and basal area. The relationship between timber volume and forest cover has the highest R^2 value at 10 meter height (figure 4.8). A problem in using the 10 m formula for predicting volume is the intercept value of 120. This means that when there is a crown cover of zero, still a volume of 120 m^3 is predicted. Also the data points are not evenly distributed over the data range, with a concentration on the lower end of the range and few points at the high end of the range. Although the 3m graph has a lower R^2 value, the intercept is near zero. This formula will be used in 4.2.2 for predicting timber volume on a continuous surface.



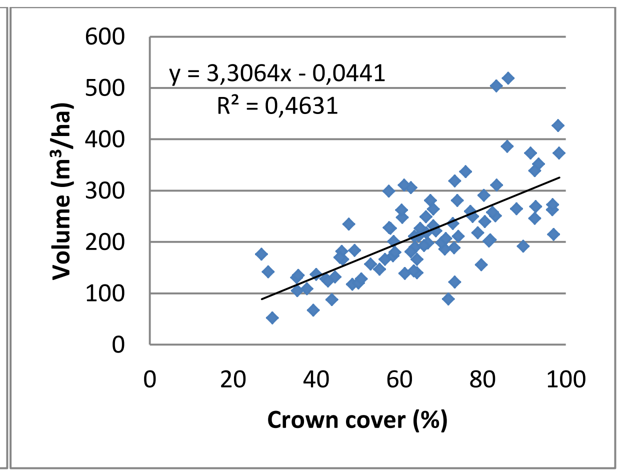


Figure 4.7 Crown cover percentage against basal area and timber volume

Scatterplots showing the relation between crown cover percentage and basal area (left), and timber volume (right), with crown cover as 3 m height difference cover percentage.

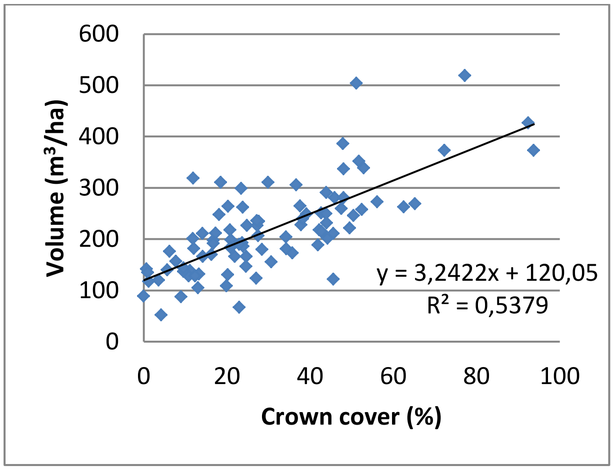


Figure 4.8 Scatterplot showing the relation between crown cover percentage and timber volume, with crown cover as 10 m height difference cover percentage

The MFV plots that have only information about timber volume do not show this relationship. When considering the complete MFV dataset R^2 values vary between 0.07 and 0.16 when forest cover is related to volume. As an example figure 4.9 shows the scatterplot for the 3m height graph which is representative for the other scatterplots. The values for the intercept at the different heights range from 122 till 170. So a realistic prediction cannot be made by using the information from the MFV volume data.

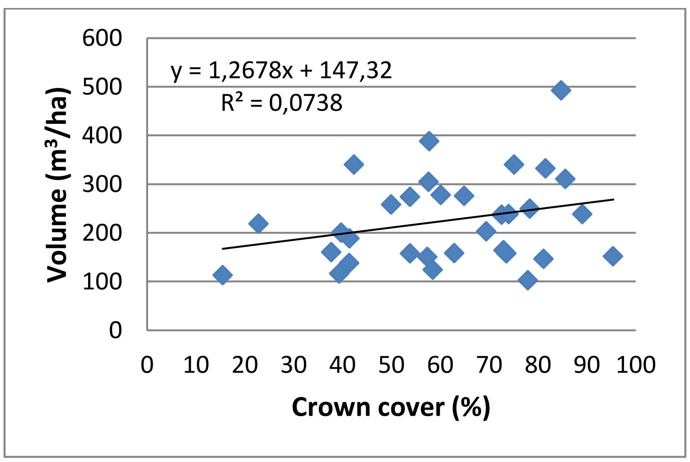


Figure 4.9 Scatterplot showing the relation between crown cover percentage and timber volume for the MFV dataset

4.1.4 Canopy profile

The canopy profiles have been studied for the different tree species. For the SBB plots information about cover percentage is available on plot scale. The plots with a high cover percentage of the main tree species have been selected for display as a sort ideal case. In the cases where only a few plots of the tree species are available also plots with a lower cover percentage were included

For display the 0-10% height class has not been included in the graphs as this class includes ground observations and has in general a higher value, which makes the pattern within the rest of the profile less visible.

Figure 4.10 shows the canopy profile for Red oak and for Douglas fir. Examples of canopy profiles for the other tree species can be found in appendix 7.

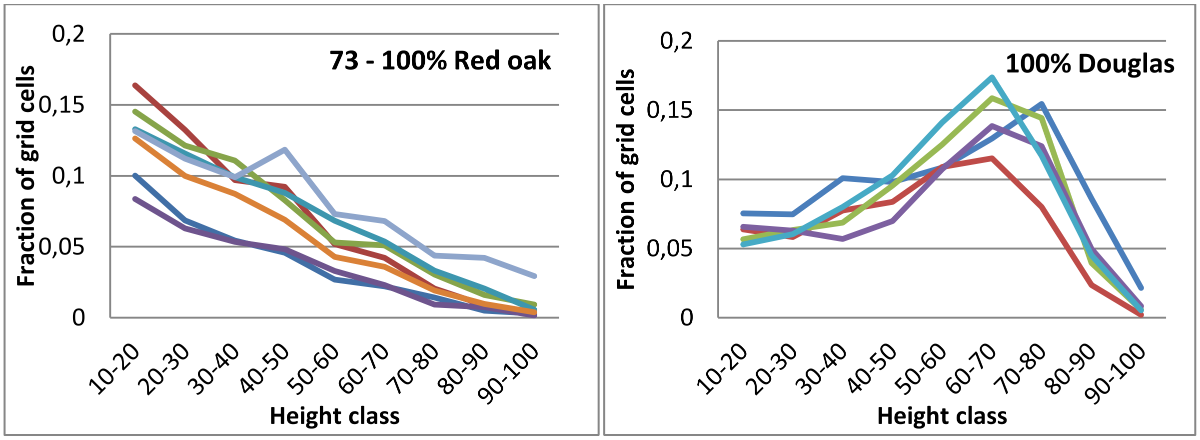


Figure 4.10 Canopy profile plots for Red oak and Douglas

In these graphs the range of cover percentages of the main tree species that are displayed is shown.

Within these graphs there is a consistent pattern for the different plots. Red oak as an example of a deciduous tree species has most grid cell values in the lower parts of the forest. Douglas fir as an example of a coniferous species has most grid cell values at 60-80% of the total canopy height. This can be explained because lidar data was collected in the leaf off season and the deciduous canopies are thus more open.

Japanese larch is an exceptional tree species. It is a coniferous tree that loses its foliage during winter. In the graph in appendix 7 it resembles more the deciduous trees than the coniferous as these curves do not have the outspoken peak around the 60 – 70% class.

For both the SBB and MFV data, the average curves for all tree species are shown in figure 4.11. Birch and western hemlock in the SBB dataset and birch, port orford cedar and Austrian pine in the MFV dataset, have only one plot, so the non-averaged curves of these plots have been used instead of an average.

In these plots the coniferous trees in general have a high lidar return fraction in the 60-70% class, and a low fraction in the lower classes. For deciduous trees this is the other way around. This information has been used to develop a classification. Dividing the 20-30% height interval class by the 60-70% height interval class results in an index with in general higher values for deciduous trees and lower values for coniferous trees. For the SBB dataset a plot was made showing the index values of the different plots, grouped per tree species, with deciduous trees placed in one group (figure 4.12 left). When including only the plots with a coverage percentage over 80%, the distinction between deciduous and coniferous becomes more clear (figure 4.12 right). Within the deciduous plots there is only one plot left, with as main tree species oak, with a value smaller than two. For the MFV dataset the same plot was made by grouping into deciduous and coniferous plots (figure 4.13). As there is no information about the dominance of the main tree species within the MFV dataset, a figure showing only plots with a high cover percentage of the main tree species could not be used.

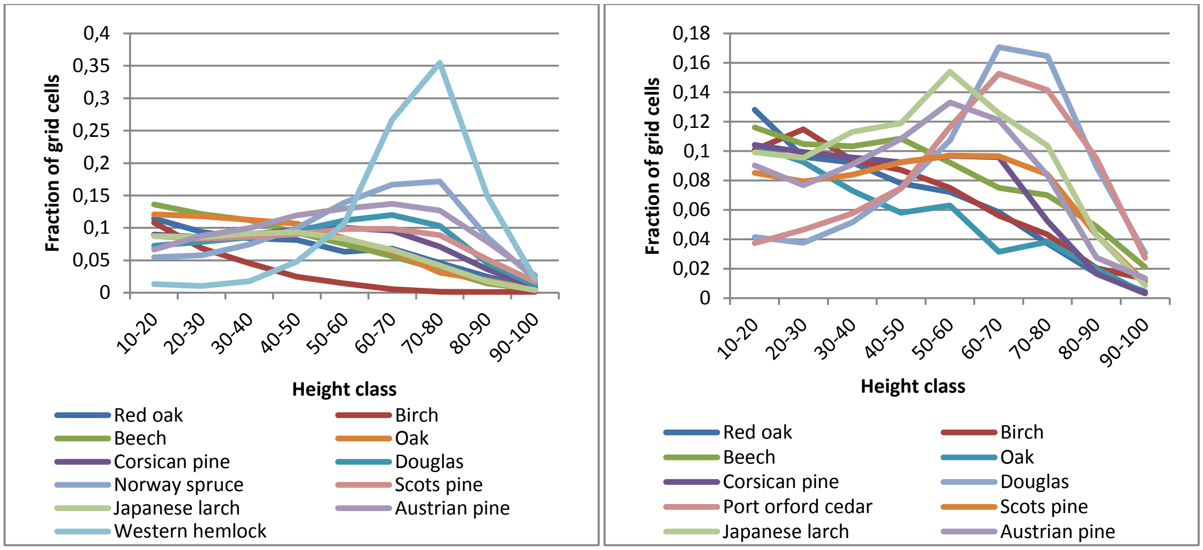


Figure 4.11 Average profiles of the SBB and MFV plots

Average profiles of the SBB plots (left) and the MFV plots (right). In the cases with only one plot available per tree species this plot is included

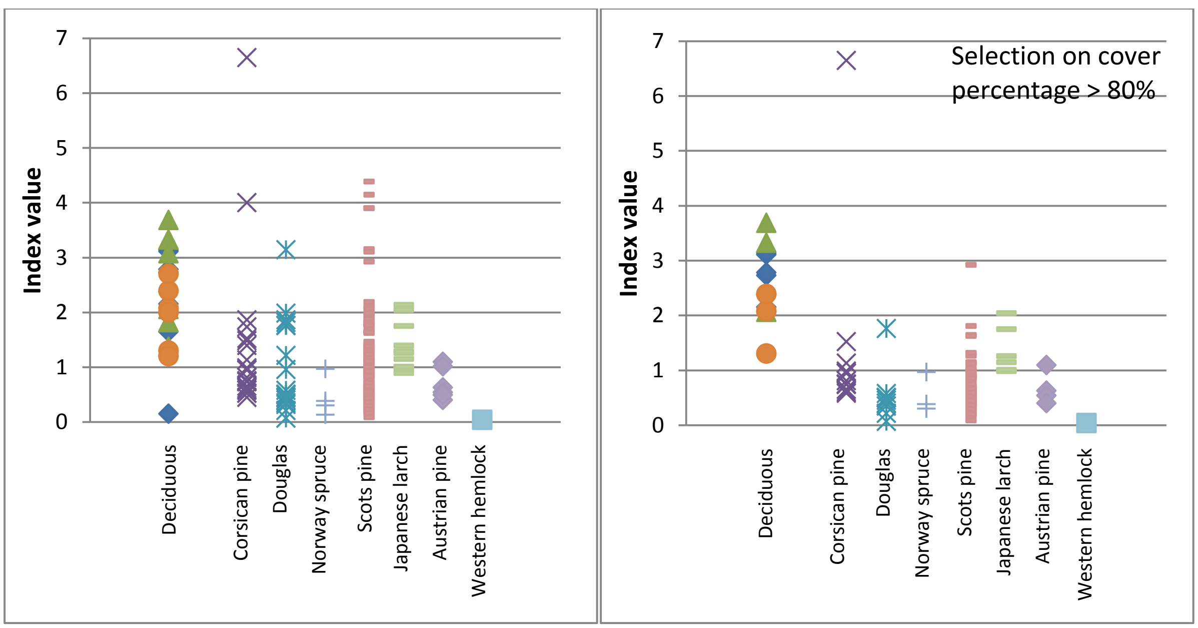


Figure 4.12 Index values grouped per tree species for the SBB dataset

In the graph on the left all plots have been included, on the right only the plots where the main tree species covers more than 80% are included

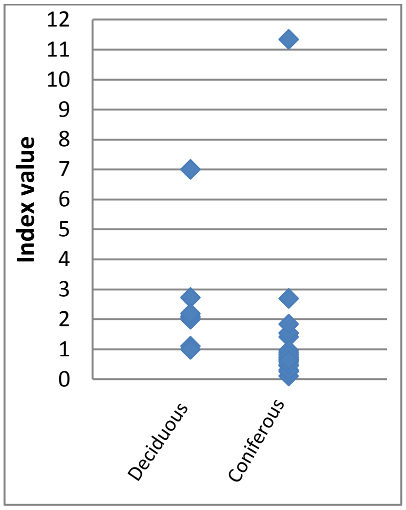


Figure 4.13 Index values for deciduous and coniferous plots for the MFV dataset

In the SBB dataset the plot with birch as main tree species is not shown, because with a value of 12.3 this outlier makes visualizing the rest of the data less clear. It is a plot with low trees and a high percentage of open space, which makes it a deviant forest plot. In the SBB dataset the Corsican pine plot with a value of 6.6 is located on the edge of a forest clear cut and this open area is also included in the plot.

Based on these plots three different groups were distinguished:
Coniferous, with an index value between 0 and 1
Deciduous, with an index value larger than 2
Intermediate, with an index value between 1 and 2

This last class is a transitional class that includes plots with both coniferous or deciduous tree species as main tree species. These are thus the mixed forest plots and also include Japanese larch. This classification has been applied on the total forested area, which is described in 4.2.3.

4.2 Continuous surface

4.2.1 Height

Applying the formulas that were derived in 4.1.2 on the total forested area results in four maps (Appendix 8). These maps have a grid cell size of 25 meters as described in 3.5.1. The two maximum height maps look the same as well as the two dominant height maps. This is because the relative differences between values did not change by applying the MFV and SBB formula on the maximum height grid. A detailed view of these maps is given in figure 4.14. The range of values between the maps however does vary. The range as shown in the legend of these maps includes unrealistic high values. For example the height map derived with the MFV maximum height formula has as maximum value 47.3 m. There are however very few of these extreme values. For example in this map there are five pixels that have a value above 40. Most of these extreme height values can be attributed to high objects present in the forest. At the locations of some of these pixels there are power pylons or a communication mast.

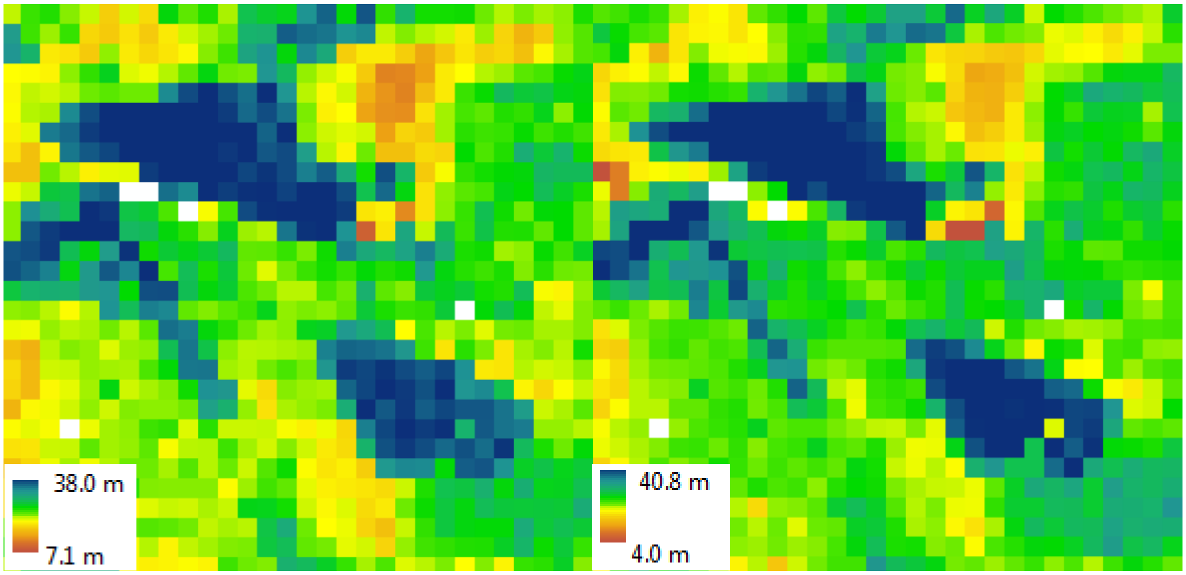


Figure 4.14 Detailed view of the maximum height map (left) and dominant height map (right). *The same area as figure 3.6 has been used. One grid cell has a size of 25 X 25 m.*

Comparing these maps with the two test datasets gives eight scatterplots (figure 4.15-4.18). The two test datasets have been compared to each of the four maps. The resulting scatterplots show the field measured values on the x-axis against the regression result on the y-axis. The dotted line is a 1:1 reference line and ideally the values are on this line, which means that the values derived with the regression formula are the same as the field measured values. The corresponding average differences and the root mean squared errors for the height regressions are given in table 4.3.

In the two maximum height graphs derived with the regression formula that was acquired from the SBB training dataset, the field measured values are higher than the predicted values for the plots with the highest trees (figure 4.15). Overall the plots in these two graphs do not follow the y = x reference line. This results from the deviations between field measured and lidar derived height values in the SBB training set as discussed in 4.1.2. As a result the average difference that results from applying this formula is negative, whereas for the other three formulas this is a positive value. The result from the MFV maximum height formula are more accurate as can be seen in the two corresponding graphs (figure 4.16). The plot values deviate less from the y = x reference line and the root mean squared errors are lower.

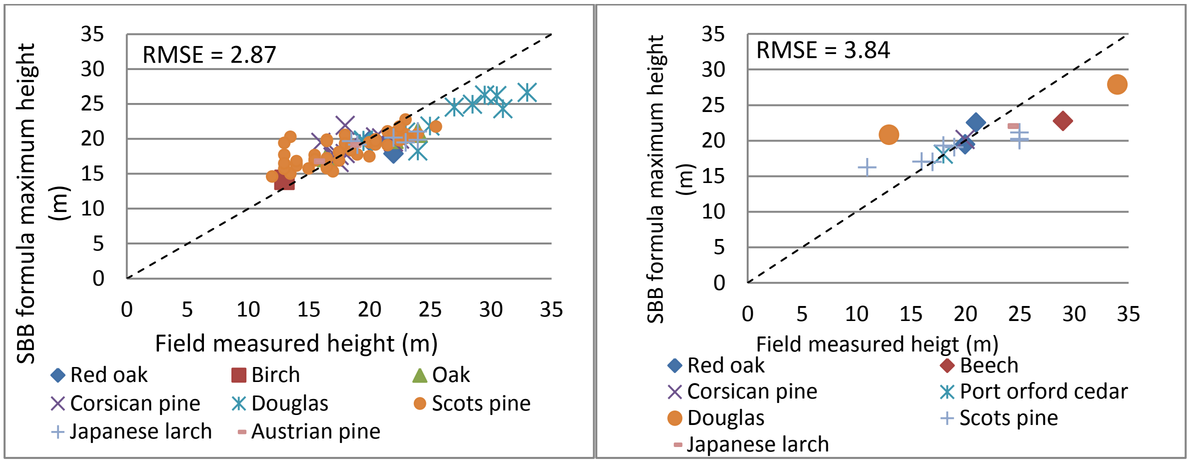


Figure 4.15 Field measured height (y-axis) against maximum height derived with the SBB formula for the SBB test dataset (left) and the MFV test dataset (right)

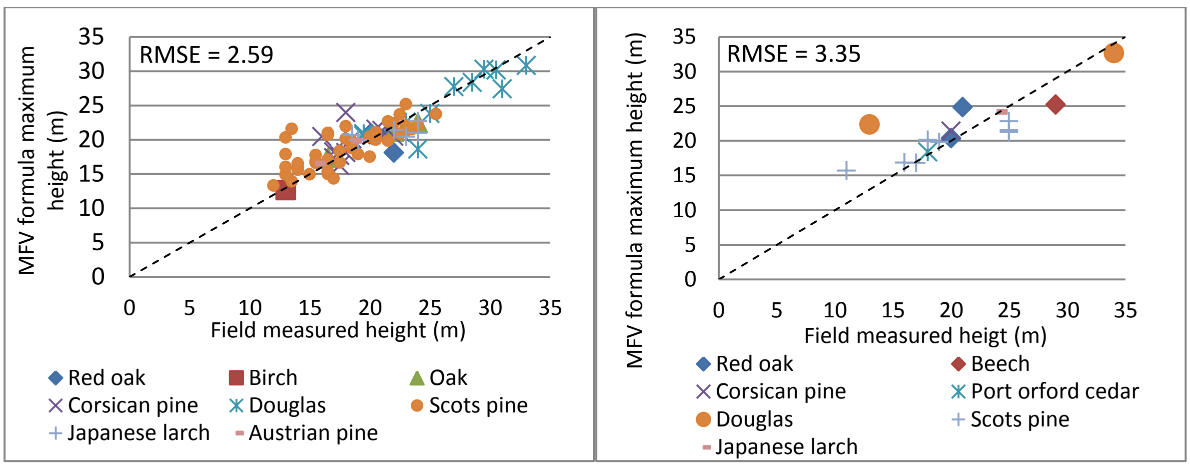


Figure 4.16 Field measured height (y-axis) against maximum height derived with the MFV formula for the SBB test dataset (left) and the MFV test dataset (right)

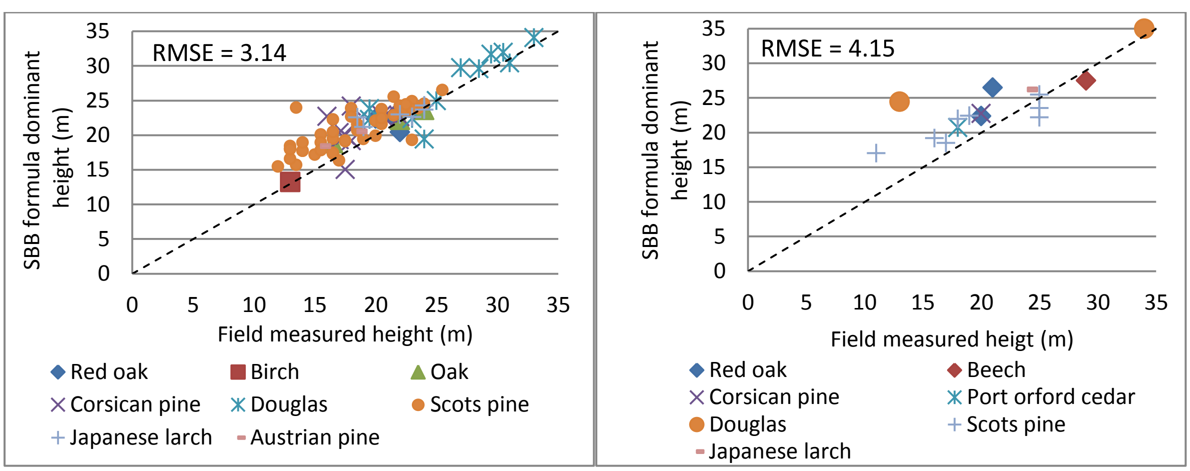


Figure 4.17 Field measured height (y-axis) against dominant height derived with the SBB formula for the SBB test dataset (left) and the MFV test dataset (right)

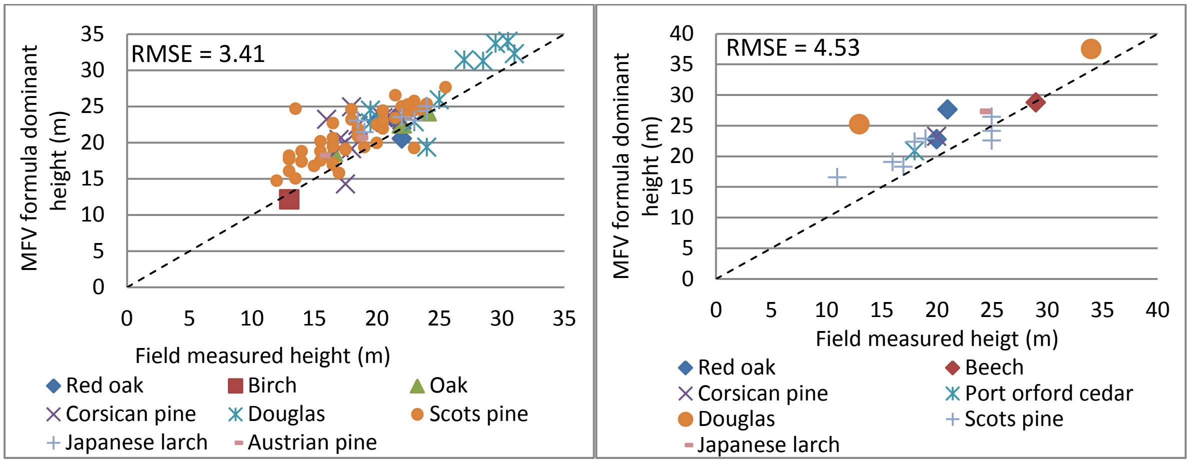


Figure 4.18 Field measured height (y-axis) against dominant height derived with the MFV formula for the SBB test dataset (left) and the MFV test dataset (right)

In the MFV dataset there is one douglas plot that has a field measured value of 13 and the predicted values range from 20.8 till 25.3 (figures 4.15-4.18 right). This douglas plot has as regeneration year 1970, was measured in the field in 2001 and laser scanning took place in 2008. This means a height increment of 7-12 m in seven years, which is an unrealistic high value (Jansen et al., 1996).

In table 4.3 the differences between the different datasets and formulas is not very outspoken. In general the RMSE and average differences are larger for the dominant height formulas and for the MFV plots. The formula that was derived with the SBB plots seems to work well on the MFV plot locations and vice versa.

Table 4.3 Average difference and Root mean squared error in meters (between brackets) for the field measured and linear regression derived values applied over the total forest area

		Maximum height (m)	Dominant height (m)
SBB plots	SBB formula	-0.44 (2.87)	2.02 (3.14)
SBB plots	MFV formula	0.37 (2.59)	2.35 (3.41)
MFV plots	SBB formula	-0.70 (3.84)	2.53 (4.15)
MFV plots	MFV formula	0.57 (3.35)	3.17 (4.53)

4.2.2 Relating forest cover to basal area and timber volume

The equations that showed the best fit to basal area and timber volume have been applied on the complete study area.

For basal area this is:

Basal area = 0.3847 x crown cover percentage at 3m + 1.1171

For timber volume this is:

Timber volume = 3.3064 x crown cover percentage at 3m – 0.0441

These equations were applied on a crown cover percentage map with a grid size of 25 m (appendix 9). The two resulting maps are visually the same, as both are linearly transformations from the same crown cover map. The values from these maps have been compared to the test datasets. The SBB dataset includes both basal area and volume, whereas the MFV data only has information on volume available. Figures 4.19 and 4.20 show these results.

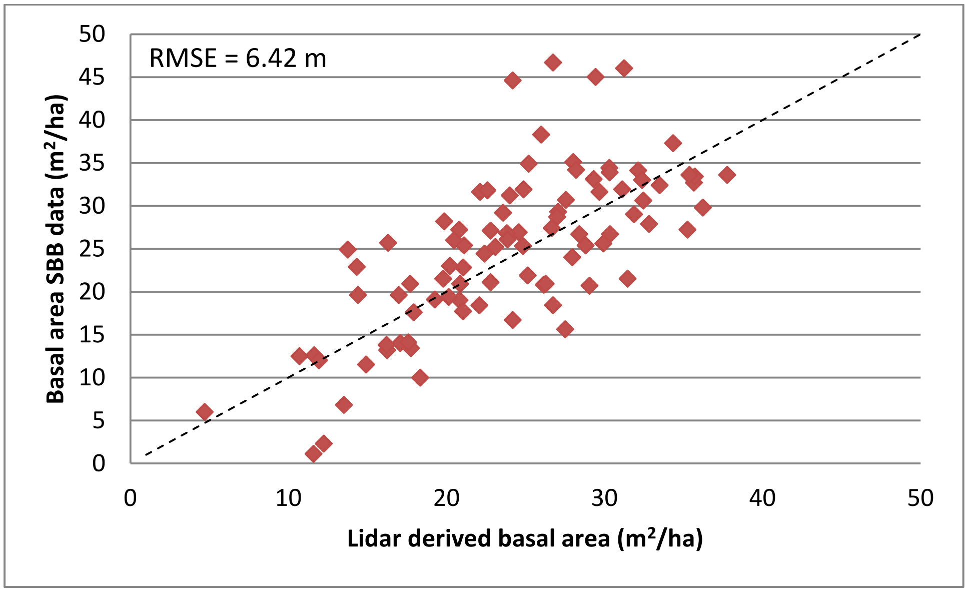


Figure 4.19 Scatterplot for the SBB test dataset relating lidar derived basal area to basal area values from the SBB data

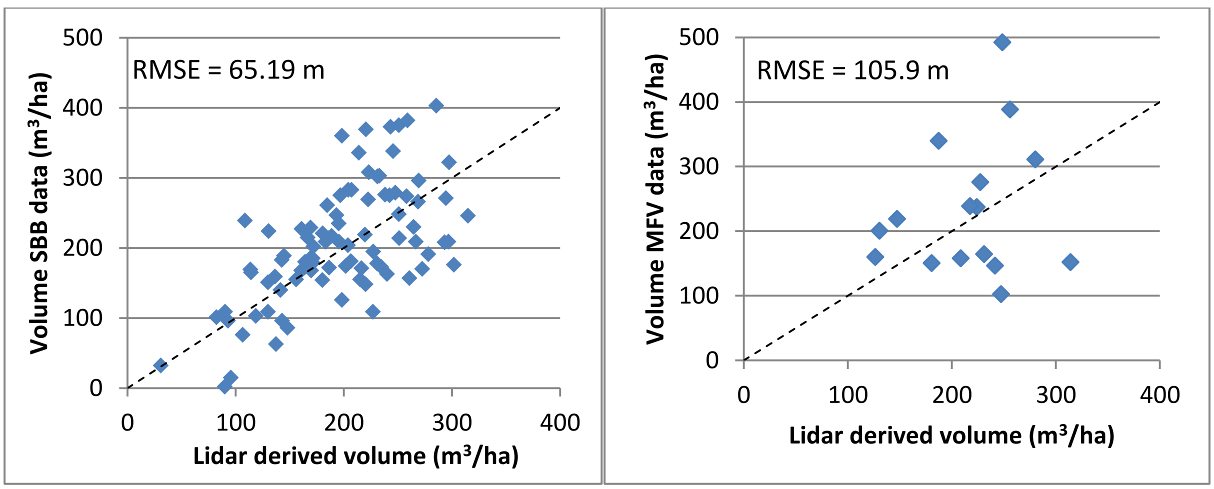


Figure 4.20 Scatterplots showing lidar derived timber volume against field measured timber volume for the SBB dataset (left) and the MFV dataset (right)

The results from the MFV test data for volume are the most scattered and have a high root mean squared error compared to the SBB results. A specific cause for the most deviating values has not been found although one douglas plot has been discussed earlier in 4.2.1 because there was a large difference between field measured and lidar derived height.

This is the data point that has a value of 314 as lidar derived value and a value of 152 in the MFV data. In figure ... there are four points that have an exceptionally high value in the SBB data, but an explanation for this could not be found.

In the basal are map the basal area values are in the range 1.1–39.6 m^2/ha. Compared to the field measured values these are realistic values as in the SBB reference data the average is 25.8 with a minimum value of 1.1 and a maximum of 51.0 m^2/ha. For timber volume the data range in the map is -0.04–330.59. Here the negative value, which occurs at some open spaces in the forest, is not a valid outcome. The average value of the field measured plots for the SBB data is 213 and 225 m^3/ha for respectively the SBB and MFV data. The average value of the lidar derived values in both test datasets is 197 and 216 m^3/ha for the SBB and MFV data. The maximum values for timber volume in both reference datasets is around 500 m^3/ha, which is higher than the maximum value in the map.

4.2.3 Canopy profile

The index that was found in 4.1.4 was applied on the total forested area. In ArcGIS the same methodology was used as before on plot scale. The grid cell size used was 25m. The height interval classes 20-30% and 60-70% were stored separately and were divided. Contrary to plot scale, dividing by the total cell size is not needed, as there are no plots that have different surface areas.

A map was created that shows the classes deciduous, coniferous and intermediate (figure 4.21).

For deriving the classification method and threshold values all the MFV and SBB plots were used. Therefore, the two test datasets cannot be used anymore as an independent test dataset, for example by making an error matrix. As a result the derived map has been compared visually with two other maps that have information about the distribution of deciduous and coniferous forest.

First, the classified map has been compared to the Top10 forest classes by plotting these on top of the classified map (appendix 10). These maps show that in general the classification works reasonably for larger areas that are either deciduous or coniferous. Only a few errors in small areas are visible where a coniferous area is classified as deciduous or the other way around. The intermediate class contains some areas that are considered coniferous or deciduous by the Top10 data. Areas that are considered mixed forest by the Top10 classification have a more diverse classification. The coniferous and deciduous classes have larger continuous areas, whereas the intermediate class occurs more scattered.

As a second reference map, the map of the 4th national forest inventory has been used. When comparing the classification result with this map, appendix 11, the different classes are also quite similar to each other. The larger coniferous and deciduous areas match between both maps. The 4th national forest inventory map does not have a mixed forest type and the amount of coniferous forest is higher than in the Top10 dataset. The higher amount of coniferous forest in this old reference map can be explained because the amount of coniferous forest in The Netherlands has decreased since this inventory (Directie kennis, 2006).

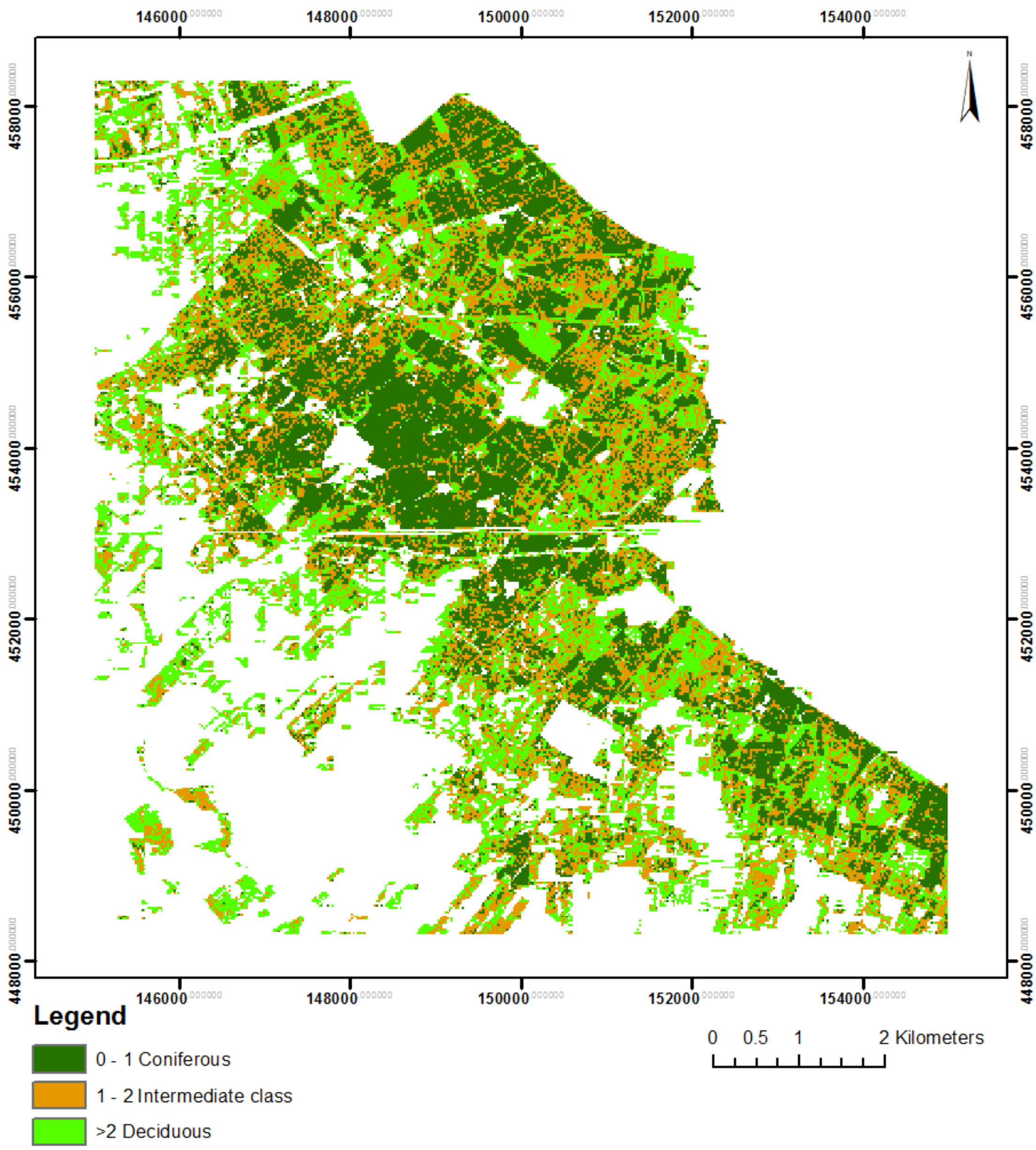


Figure 4.21 Map showing the classification result with the classes coniferous, deciduous and intermediate

5 Discussion

5.1 Reference data

The forest inventory data used as reference data was not specifically collected to be used as validation dataset for the parameters retrieved from the AHN-2 lidar data. As a consequence there are some traits in the reference data that make it less suitable as reference data.
The locations of the forest plots has been chosen by using a systematic sampling method. As a consequence, part of the plots is located at the edge of a forest stand. So part of a forest path or of another forest type would be included in the plot. In the field this is dealt with by using not the complete circular plot, but by measuring only part of the circular plot and calculating this surface. In the methodology used in this thesis this has not been taken into account. When the methodology used is applied for example on a large region or on a continuous surface these cases will also be included, unless very detailed forest maps are available.
If a tree species is considered main tree species in a mixed forest, this does not mean it has to be the majority of the trees. For example the main tree species can occupy 40% of the surface, the rest of the surface area is covered by other tree species with a lower cover percentage. This makes interpreting the data more difficult.
In the results section (4.1.1) it was found that some of the differences in height between field measured and lidar derived height values could be attributed to the reference data. These differences are not caused by incorrect measurements, but by the way they were collected. The goal of the height measurements was not to determine the maximum canopy height, so measurements were not specifically done on the highest trees in the plot. When multiple trees within one plot are measured their height is averaged to get one plot height value. This has also been applied on the SBB data.
Forest cover estimation by lidar and by field measurements as applied in the MFV data do not correspond. It is known that visual assessment of forest cover is subjective and not very accurate (Paletto & Tosi, 2009). The SBB coverage estimation is a measure of canopy closure from one point of view. By using the AHN lidar data forest cover as a vertically projected ground surface is calculated. These are two different methods, which can explain the differences in cover estimations. These issues have their influence on the accuracy of the obtained results.

5.2 Accuracy of results

This research mainly focused on finding and testing methods that can be applied on already available data. Therefore the focus was not on retrieving a high accuracy but on testing multiple methods.
The results obtained by the methods applied in this thesis can partly be compared to results described in literature. Tree height estimations in literature have in general a R^2 value between 0.8 and 0.98 (based on 6 studies), and a RMSE around 1.5 m (Van Leeuwen & Nieuwenhuis, 2010). Only one article reviewed by Van Leeuwen & Nieuwenhuis (2010) used the RMSE as accuracy indicator for retrieving tree height on plot level. The value of the RMSE can only be assessed accurately when information on canopy height itself is also given. The R^2 values found in this thesis are between 0.46 and 0.66. These are low values compared to the results found in literature, but comparing only on values has its shortcomings. The study area

in this thesis is diverse with different tree species and different forest ages. The studies mentioned by Van Leeuwen & Nieuwenhuis (2010) are performed on less diverse forests with only one or two tree species present in the canopy. Also suitability of reference data will play a role as discussed in 5.1.

When considering the results of the timber volume estimation the article of Breidenbach et al. (2010), is of interest as it reports a RMSE of 34.56 m^3/ha for a mainly coniferous forest in Norway with an average volume of 202.38 m^3/ha within the sample plots. This study used point cloud data and has collected field data specifically for this research. For the SBB dataset the average timber volume within the plots is 213 m^3/ha and for the MFV data this is 225 m^3/ha. A RMSE of 65 (SBB test data) and 106 m^3/ha (MFV test data) was found, so these RMSE values are two and three times higher as the value obtained by Breidenbach et al. (2010).

The accuracy of the forest cover estimations were tested by using the error matrix. The accuracy of this method cannot be determined adequately because the reference datasets are not appropriate as discussed in 5.1. The method itself is quite straightforward and does not rely on this reference data for establishing a relationship. The classification into the classes deciduous, coniferous and intermediate could not be referenced by field observations from the MFV or SBB data because all plot information was used for establishing the methodology. Therefore, an appropriate validation of this result did not take place and only a visual assessment was made (section 4.2.3 and appendix 10).

Overall, the derived results agree with the field measured data. The derived methods and relationships work better for the SBB than for the MFV data, which has much less data points within the study area.

5.3 General applicability

During this thesis it was not tested how well the found relationships work for other areas that have different forest types. The study area is a relatively homogeneous area when looking at forest type and growing conditions. However, this forest type is one of the most common forest types in The Netherlands (Directie Kennis, 2006). Forests on a relatively rich and moist soil are not represented by the study area. For example tree species as ash, alder, willow and poplar do not occur in the study area.

The method for determining canopy cover will probably work for other forest types, as only height information is important and a linear regression does not have to be established. This is different for retrieving canopy height and calculating basal area and timber volume. Also the classification method will probably have to be adjusted or does not work, as this relies on tree species. The applicability of these methods in other areas thus has to be tested.

6 Conclusion

Forest inventory data can be obtained by a variety of methods. During this thesis research the applicability of AHN-2 lidar data for retrieving forest parameters available from forest inventory data has been studied. By examining current literature on this topic, different methods have been developed that relate AHN-2 lidar data to already existing forest inventory data. These methods have been tested on a forested area located on the Utrechtse Heuvelrug.
Linear regression between AHN-2 lidar data and forest inventory data has resulted in the calculation of canopy height, basal area, and timber volume. Forest cover has been determined based on differences in height between forest canopy and open spaces within the forest. Furthermore, a classification method has been developed to make a distinction between deciduous and coniferous forest.
The forest inventory datasets, in this study used as reference datasets for calibration and validation, were not specifically collected to do so. The methodology to obtain canopy height information in the field made these data less appropriate to use as reference data. Forest density and forest cover estimation in the reference data did not match with the lidar derived values. Nonetheless cover values derived from the AHN-2 data could be related to basal area and timber volume.
Despite shortcomings in both reference datasets, the developed methods work reasonably well. The applicability outside the study area is something that still needs to be investigated in more detail.

Further research can focus on two topics. Testing the methodology and the found relationships for other forest types than the one studied in this research by applying the same methods on other areas in the Netherlands. Improving the methodology and the relationships by being more strict in selecting reference data so that only valid forest measurements are used or by collecting new reference data especially for calibrating and validating AHN-2 data within forests.

References

AHN. (2010). *Kwaliteitsdocument AHN-2*. Amersfoort: AHN.

AHN. (2011). *Wat is het AHN?* Retrieved July 1, 2011, from http://www.ahn.nl/wat_is_het_ahn.

Alterra. (2011). *Bodemdata.nl.* Retrieved July 3, 2011, from http://www.bodemdata.nl.

Breidenbach, J. Naesset, E. Lien, V. Gobakken, T. Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, 114, 911-924.

Ciais, P. Schelhaas, M.J. Zaehle, S. Piao, S.L. Cescatti, A. Liski, J. Luyssaert, S. Le-Maire, G. Schulze, E.-D. Bouriaud, O. Freibauer, A. Valentini, R. Nabuurs, G.J. (2008). Carbon accumulation in European forests. *Nature geoscience*, 1, 425-429.

Clement, J. (2001). *GISBOS Vierde Bosstatistiek; Documentatie van bestanden.* Wageningen: Alterra.

Coops, N.C. Hilker, H. Wulder, M.A. St-Onge, B. Newnham, G. Siggins, A. Trofymow, J.A. (2007). Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees*, 21, 295-310.

Daamen, W.P. Dirkse, G.M. (2005). *Veldinstructie Meetnet Functie Vervulling bos 2005*. Wageningen: Alterra.

Directie Kennis. (2006). *Meetnet Functievervuling bos 2001-2005. Vijfde Nederlandse bosstatistiek.* Ede: Directie Kennis, Ministerie van Landbouw, Natuur en Voedselkwaliteit.

Dirkse, G.M. Daamen, W.P. (2000). *Pilot Meetnet Functievervulling bos, natuur en landschap.* Wageningen, Alterra. Alterra-rapport 097.

García, R. Suárez, J.C. Patenaude, G. (2007). Delineation of individual tree crowns for LiDAR tree and stand parameter estimation in Scottish woodlands. *The European Information Society*, 55-85.

Gobakken, T. Naesset, E. (2004). Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research*, 19(6), 529-542.

Heurich, M. (2008). Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management*, 255, 2416-2433.

Houghton, R.A. Hall, F. Goetz, S.J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research*, 114.

Hyyppa, J. Hyyppa, H. Leckie, D. Gougeon, F. Yu, X. Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29(5), 1339-1366.

Jansen, J.J. Sevenster, J. Faber, P.J. (1996). *Opbrengsttabellen voor belangrijke boomsoorten in Nederland.* Wageningen: IBN. IBN rapport nr. 221.

Jennings, S.B. Brown, N.D. Sheil, D. (1999). Assessing forest canopies and understory illumination: canopy closure, canopy cover and other measures. *Forestry*, 72(1), 59-73.

Kwak, D. Lee. W. Lee, J. Biging, G. Gong, P. (2007).Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research*, 12, 425-434.

Leckie, D. Gougeon, F. Walsworth, N. Paradine, D. (2003). Stand delineation and composition estimation using semi-automated individual tree crown analysis. *Remote Sensing of Environment*, 85, 355-369.

Lefsky M.A. Cohen, W.B. Acker, S.A. Parker, G.G. Spies, T.A. Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70, 339-361.

Lefsky, M.A. Cohen, W.B. Parker, G.G. Harding, D.J. (2002). Lidar Remote Sensing for Ecosystem Studies. *BioScience*, 52(1), 19-30.

Leppänen, V.J. Tokola, T. Maltamo, M. Mehtätalo, L. Pusa, T. Mustonen, J. (2008). Automatic delineation of forest stands from lidar data. *Proceedings of GEOBIA 2008 - Pixels, Objects, Intelligence: GEOgraphic Object-Based Image Analysis for the 21St Century*. Calgary, Canada.

Lim, K. S. Treitz, P.M. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research*, 19(6), 558-570.

Lusink, M. Schoonderwoerd, H. Wondergem, H.E. (2006). *Ontwikkelingen in het bos van Austerlitz: resultaten SyHI 1994-2006*. Wageningen: Silve.

Magnusson, M. (2006). *Evaluation of Remote Sensing Techniques for Estimation of Forest Variables at Stand Level.* Doctoal thesis. Swedish University of Agricultural Sciences, Umeå, Sweden.

Mallet, C. Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 1-16.

McGaughey, (2010). *FUSION/LDV: Software for lidar data analysis and visualization*. Washington: U.S. Department of Agriculture, Forest Service.

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

Næsset, E. Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79, 105-115.

Ott, R.L. Longnecker, M. (2001). *An introduction to statistical methods and data analysis. 5th edition*.Pacific Grove, Ca, USA: Duxbury.

Paletto, A. Tosi, V. (2009). Forest canopy cover and canopy closure: comparison of assessment techniques. *European Journal of Forest Research*,128, 265-272.

Schneider, J. Moore, A.W. (1997). *A locally weighted learning tutorial using vizier 1.0. .* Retrieved August 15, 2011, from http://www.cs.cmu.edu/~schneide/tut5/node42.html.

Sullivan, A.A. (2008). *LIDAR Based Delineation in Forest Stands.* MSc thesis. University of Washington, Washington, USA.

Van der Sande, C. Soudarissanane, S. Khoshelham, K. (2010). Assessment of Relative Accuracy of AHN-2 Laser Scanning Data Using Planar Features. *Sensors*, 8198-8214.

Van Leeuwen, M. Nieuwenhuis, M. (2010). Retrieval of forest structural parameters using LiDAR remote sensing. *European Journal of Forest Research,* 129, 749-770.

Wang, C. Glenn N. F. (2008). A linear regression method for tree canopy height estimation using airborne lidar data. *Canadian Journal of Remote Sensing*, 34(2), 217-227.

Weber, T.C. Boss, D.E. (2009). Use of LiDAR and supplemental data to estimate forest maturity in Charles County, MD, USA. *Forest Ecology and Management,* 258, 2068-2075.

Zhao, K. Popescu, S. Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, 113, 182-196.

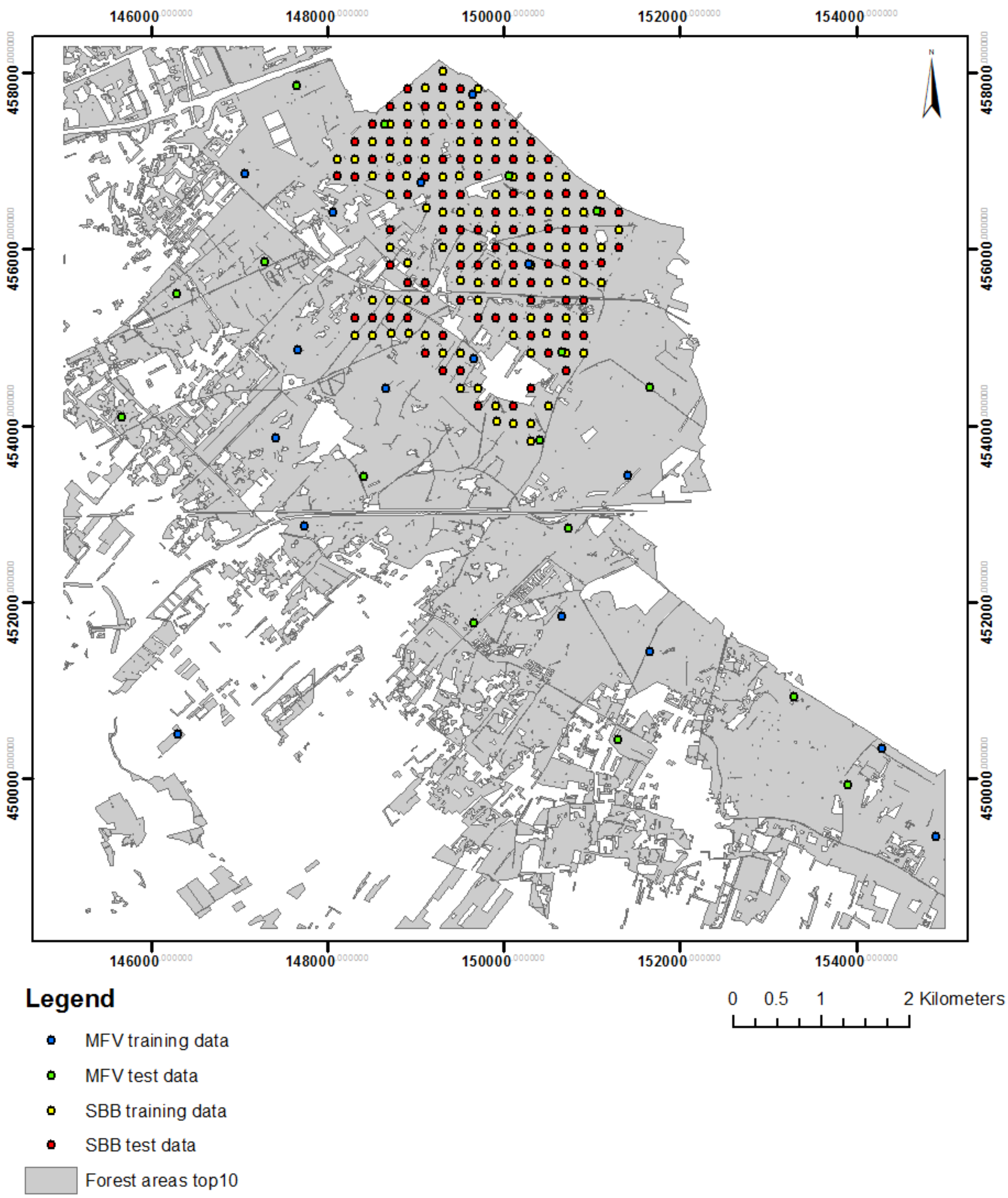
Appendices

Appendix 1: Tree species present as main tree species
Appendix 2: Distribution of training and test data plots
Appendix 3: Forest cover error matrices for the MFV dataset
Appendix 4: Scatterplots showing the distribution of lidar derived crown cover for the MFV crown cover classes
Appendix 5: Scatterplots showing the distribution of lidar derived crown cover for the SBBcrown cover classes
Appendix 6: relating crown cover percentage to basal area and timber volume
Appendix 7: Canopy profiles per main tree species for SBB plots
Appendix 8: maps showing dominant and maximum height
Appendix 9: Maps showing basal area and timber volume
Appendix 10: Top10 forest classes displayed on lidar derived classification
Appendix 11: Map of 4th national forest inventory

Appendix 1: Tree species present as main tree species

English name	Dutch name	Latin name
Deciduous		
Red oak	Amerikaanse eik	*Quercus rubra*
Oak	Eik	*Quercus robur and Quercus petraea*
Birch	Berk	*Betula pendula and Betula pubescens*
Beech	Beuk	*Fagus sylvatica*
Coniferous		
Corsican pine	Corsicaanse den	*Pinus nigra ssp. laricio*
Douglas fir	Douglas	*Pseudotsuga menziesii*
Norway spruce	Fijnspar	*Picea abies*
Scots pine	Grove den	*Pinus sylvestris*
Japanese larch	Japanse lariks	*Larix kaempferi*
Austrian pine	Oostenrijkse den	*Pinus nigra ssp. nigra*
Western red cedar	Reuzenlevensboom, Thuja	*Thuja plicata*
Western hemlock	Westelijke hemlockspar	*Tsuga heterophylla*
Port Orford cedar	Californische cypres	*Chamaecyparis lawsoniana*

Appendix 2: Distribution of training and test data plots



Appendix 3: Forest cover error matrices for the MFV dataset

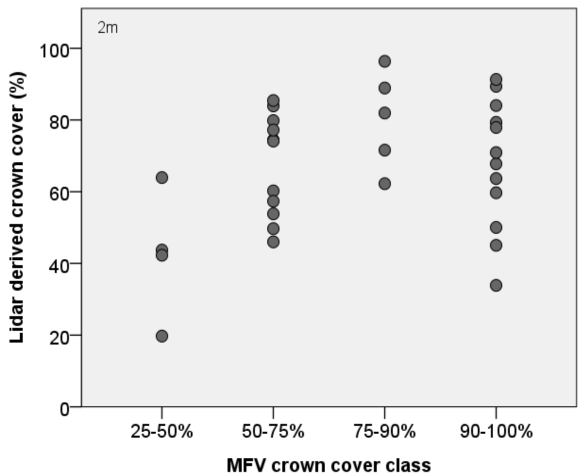
MFV 2m	Map classification						
Reference classification	< 25%	25-50%	50-75%	75-90%	90-100%	Total	Omission error
25-50%	1	2	1			4	50.00%
50-75%		1	6	4		11	45.45%
75-90%			2	2	1	5	60.00%
90-100%		2	5	4	1	12	91.67%
Total		5	14	10	2	32	
Commission error		60.00%	57.14%	80.00%	50.00%		

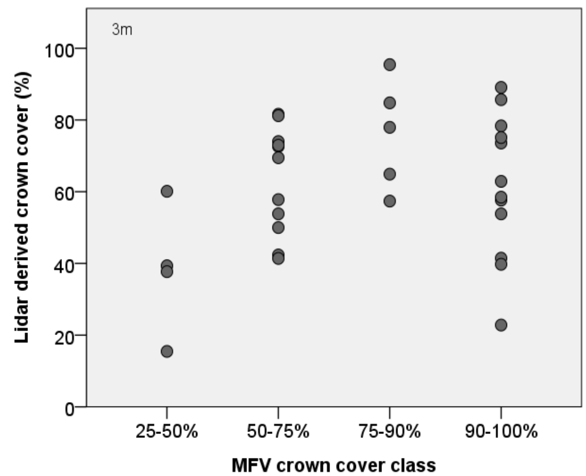
MFV 3m	Map classification						
Reference classification	< 25%	25-50%	50-75%	75-90%	90-100%	Total	Omission error
25-50%	1	2	1			4	50.00%
50-75%		3	6	2		11	45.45%
75-90%			2	2	1	5	60.00%
90-100%	1	2	5	4		12	100.00%
Total		7	14	8	1	32	
Commission error		71.43%	57.14%	75.00%	100.00%		

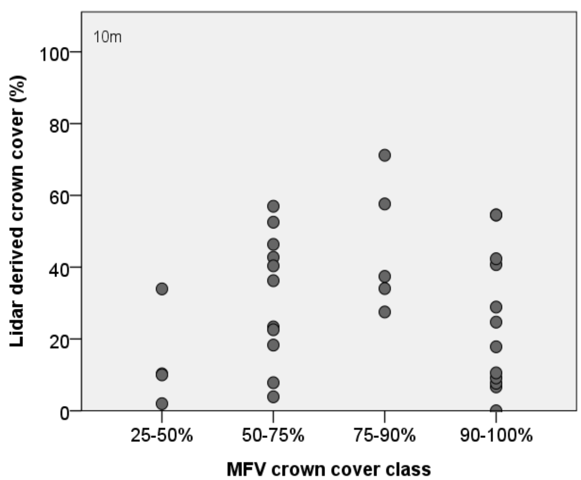
MFV 5m	Map classification						
Reference classification	< 25%	25-50%	50-75%	75-90%	90-100%	Total	Omission error
25-50%	1	2	1			4	50.00%
50-75%		4	6	1		11	45.45%
75-90%			3	1	1	5	80.00%
90-100%	1	5	4	2		12	100.00%
Total		11	14	4	1	32	
Commission error		81.82%	57.14%	75.00%	100.00%		

MFV 10m	Map classification						
Reference classification	< 25%	25-50%	50-75%	75-90%	90-100%	Total	Omission error
25-50%	3	1				4	75.00%
50-75%	5	4	2			11	81.82%
75-90%		3	2			5	100.00%
90-100%	6	4	2			12	100.00%
Total		12	6	0	0	32	
Commission error		91.67%	66.67%	100.00%	100.00%		

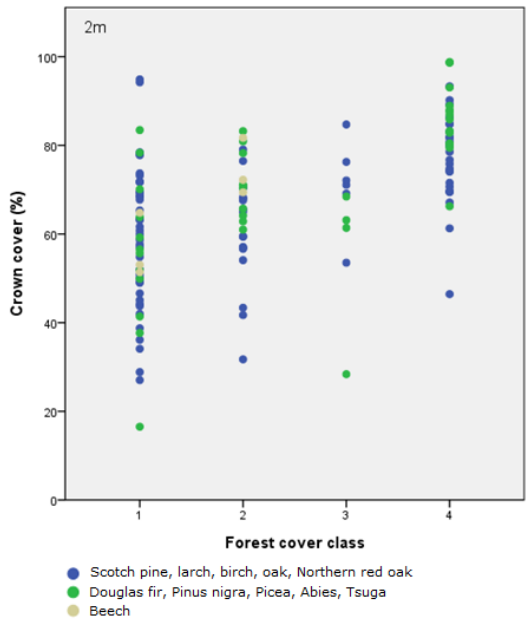
Appendix 4: Scatterplots showing the distribution of lidar derived crown cover for the MFV crown cover classes

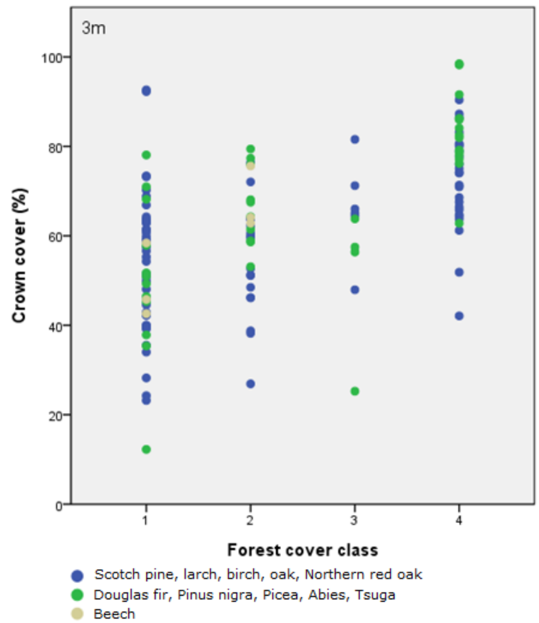


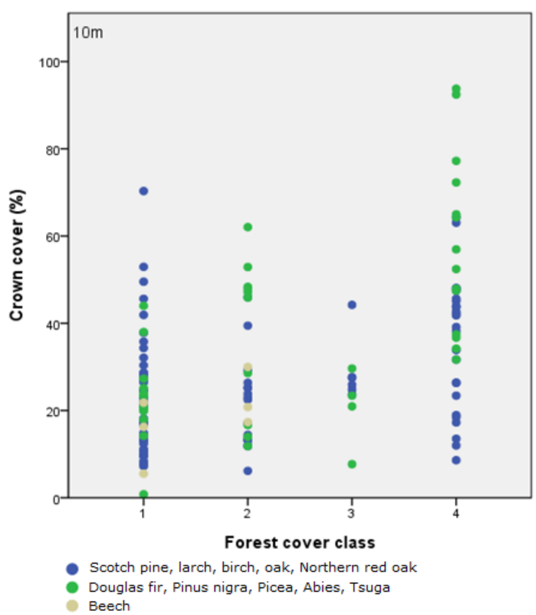




Appendix 5: Scatterplots showing the distribution of lidar derived crown cover for the SBB crown cover classes







Appendix 6: relating crown cover percentage to basal area and timber volume

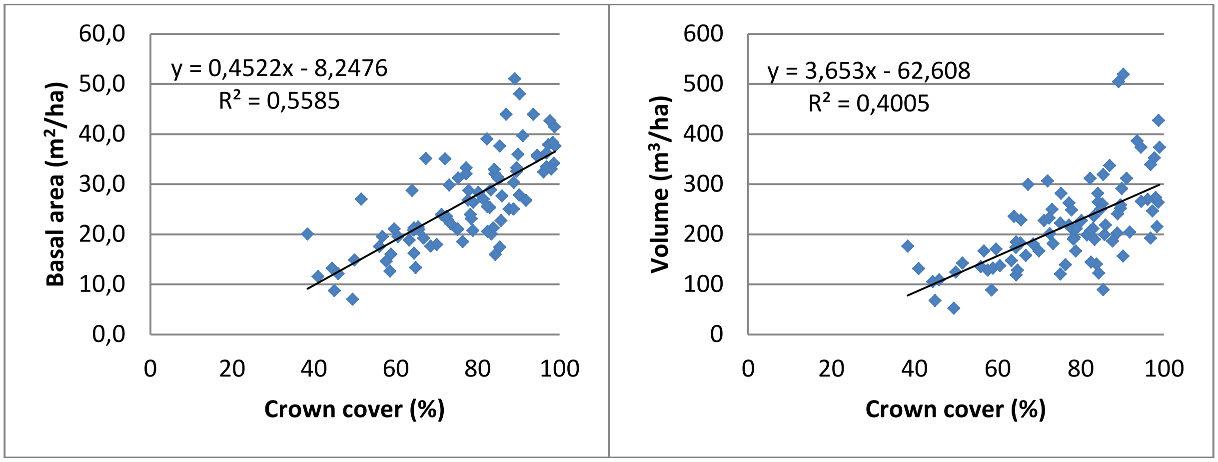


Figure ... Scatterplots showing the relation between crown cover percentage and basal area (left), and timber volume (right), with crown cover as 1m height difference cover percentage

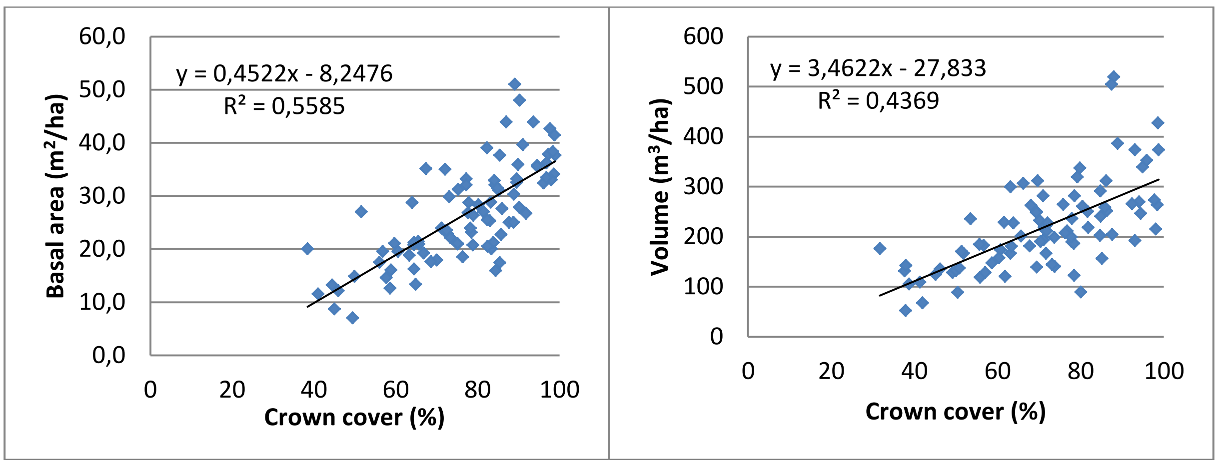


Figure ... Scatterplots showing the relation between crown cover percentage and basal area (left), and timber volume (right), with crown cover as 2m height difference cover percentage

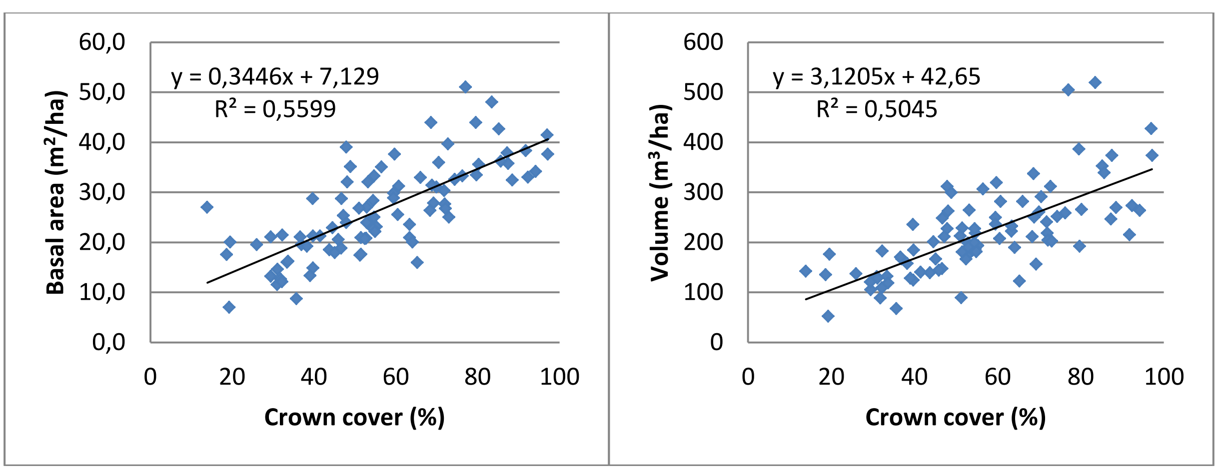


Figure ... Scatterplots showing the relation between crown cover percentage and basal area (left), and timber volume (right), with crown cover as 5m height difference cover percentage

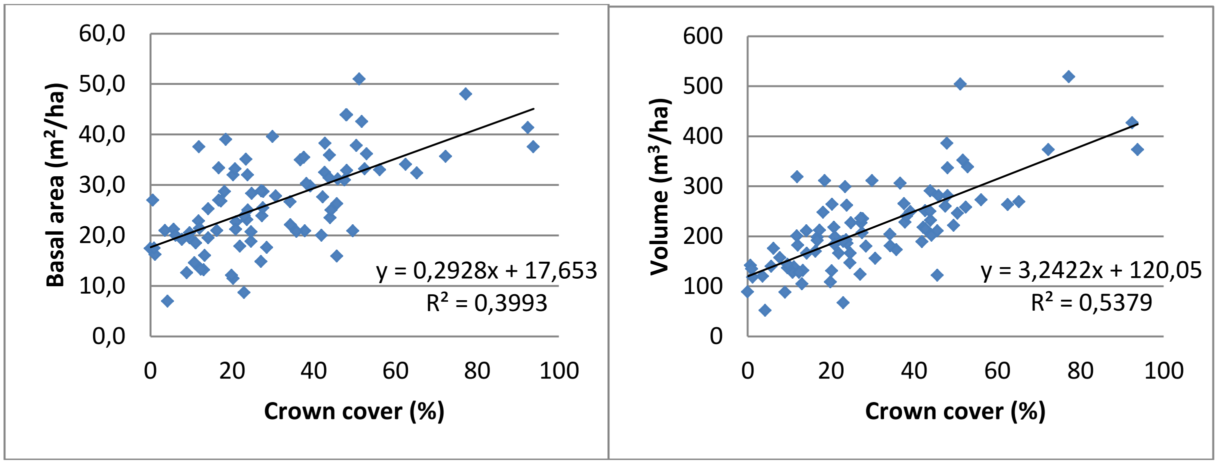
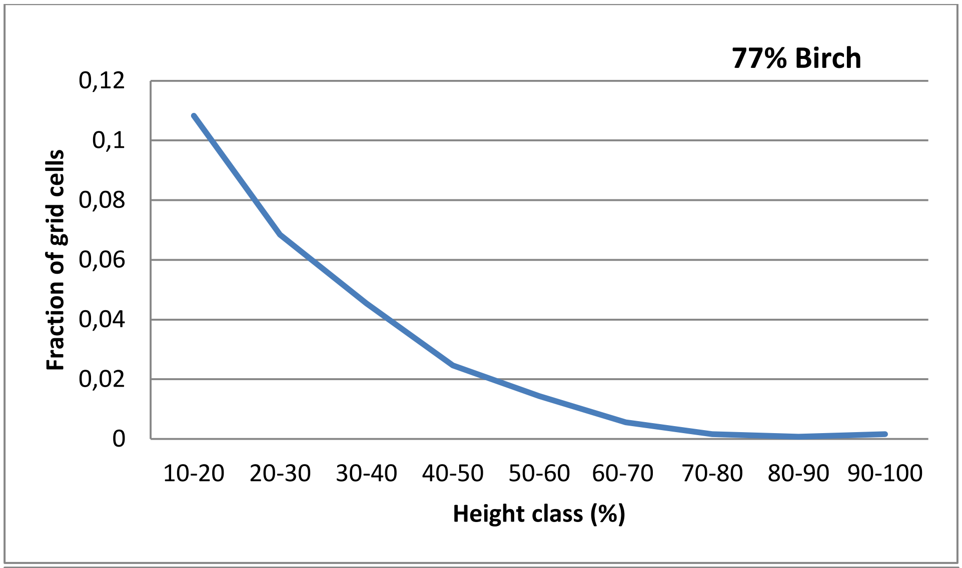
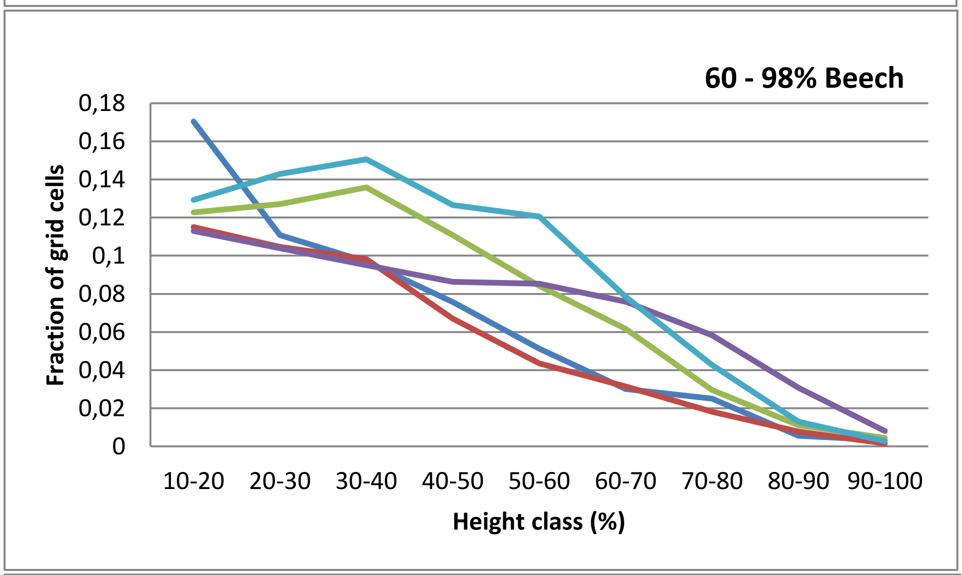
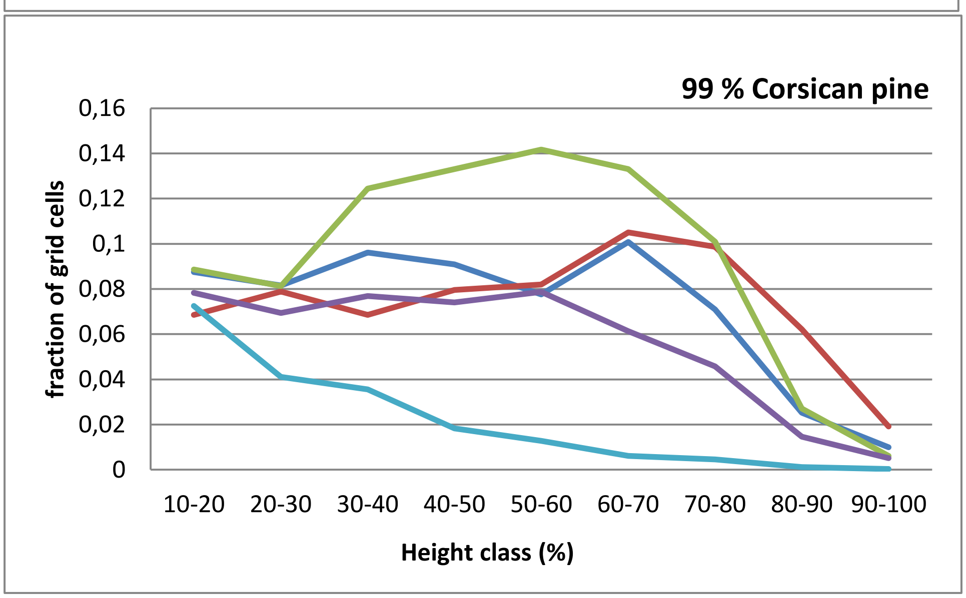


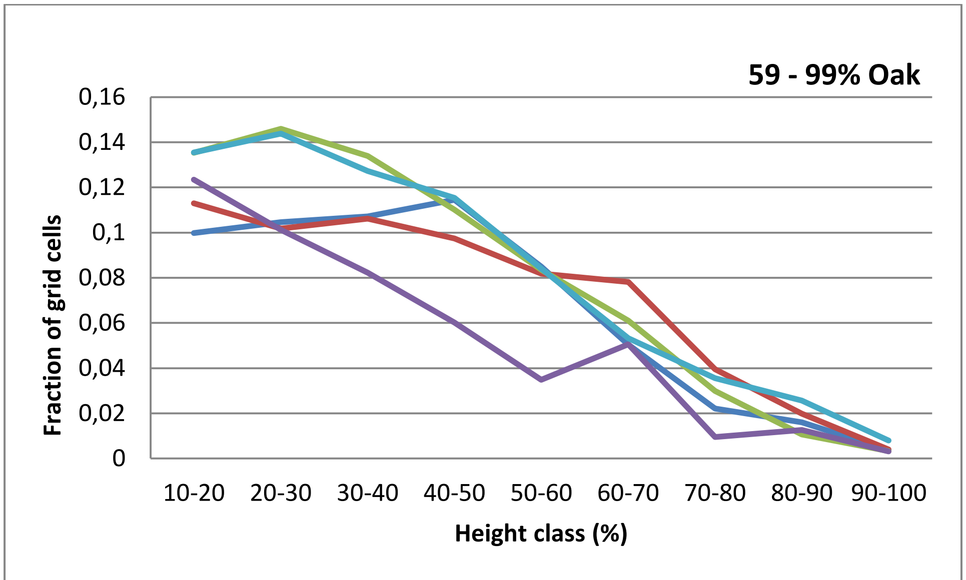
Figure ... Scatterplots showing the relation between crown cover percentage and basal area (left), and timber volume (right), with crown cover as 10 m height difference cover percentage

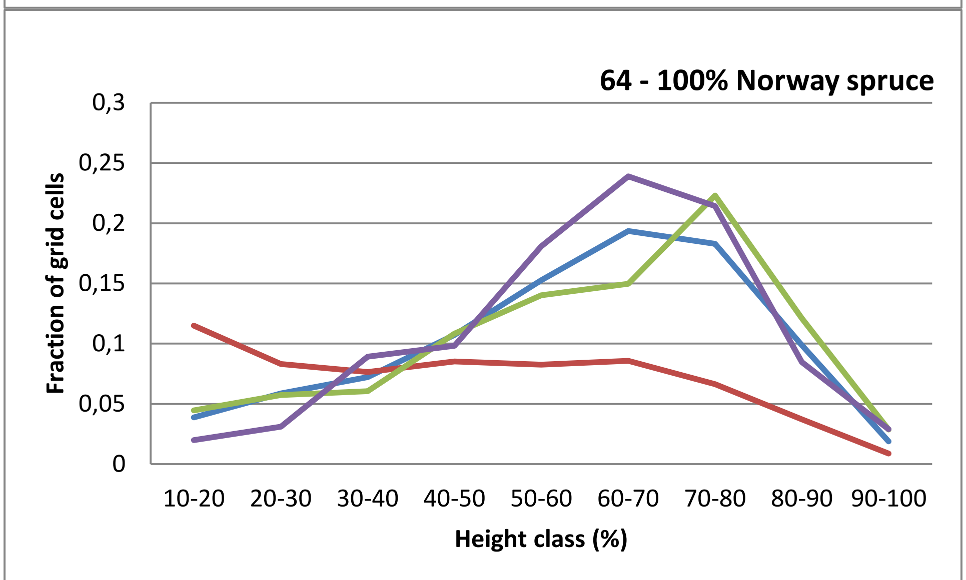
Appendix 7: Canopy profiles per main tree species for SBB plots

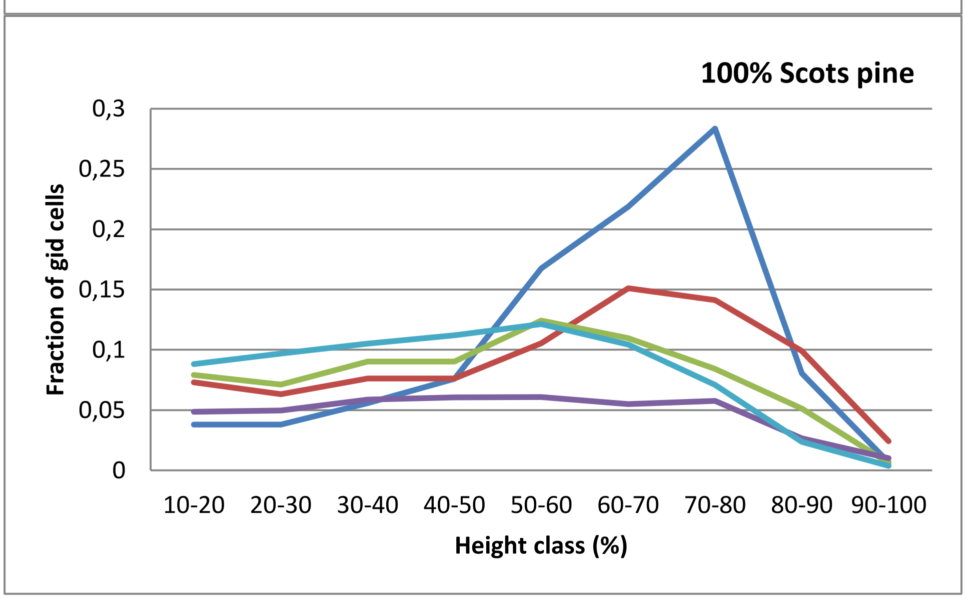


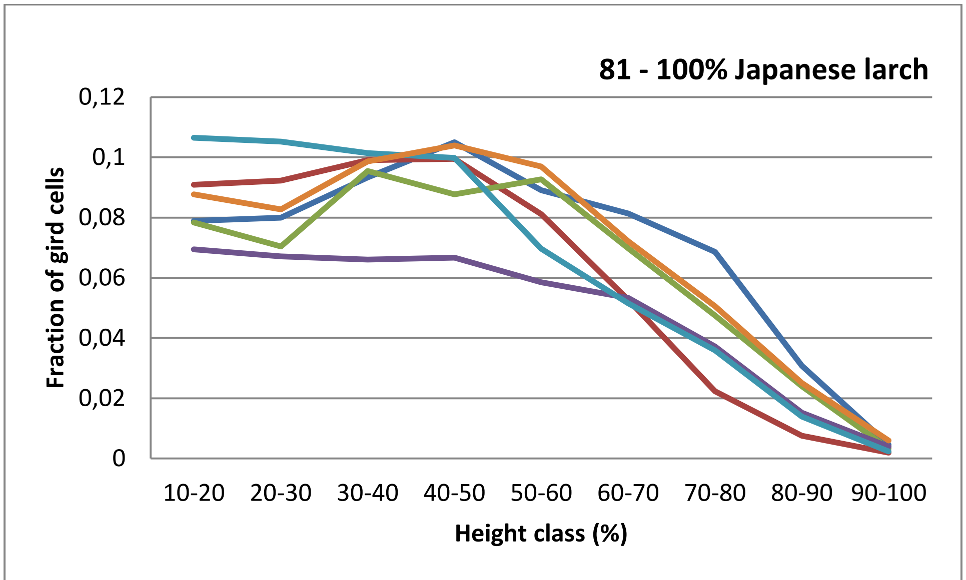


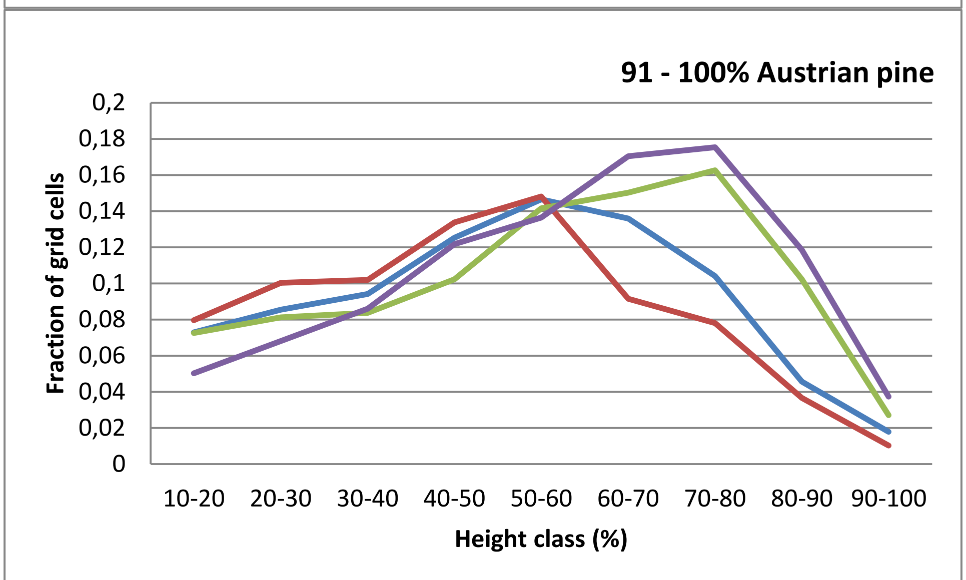


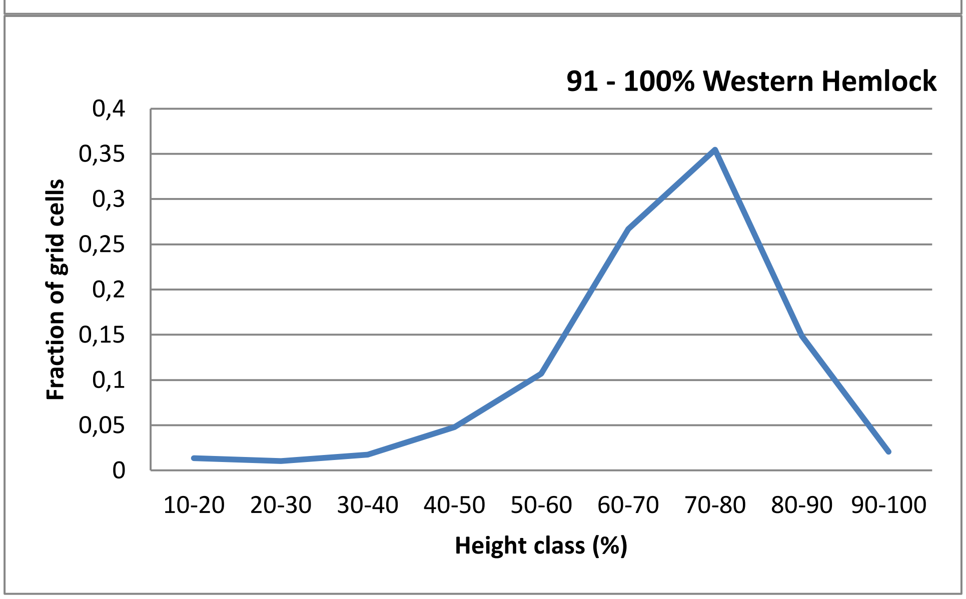

59 - 99% Oak
Fraction of grid cells
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Height class (%)


64 - 100% Norway spruce
Fraction of grid cells
0,3
0,25
0,2
0,15
0,1
0,05
0
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Height class (%)

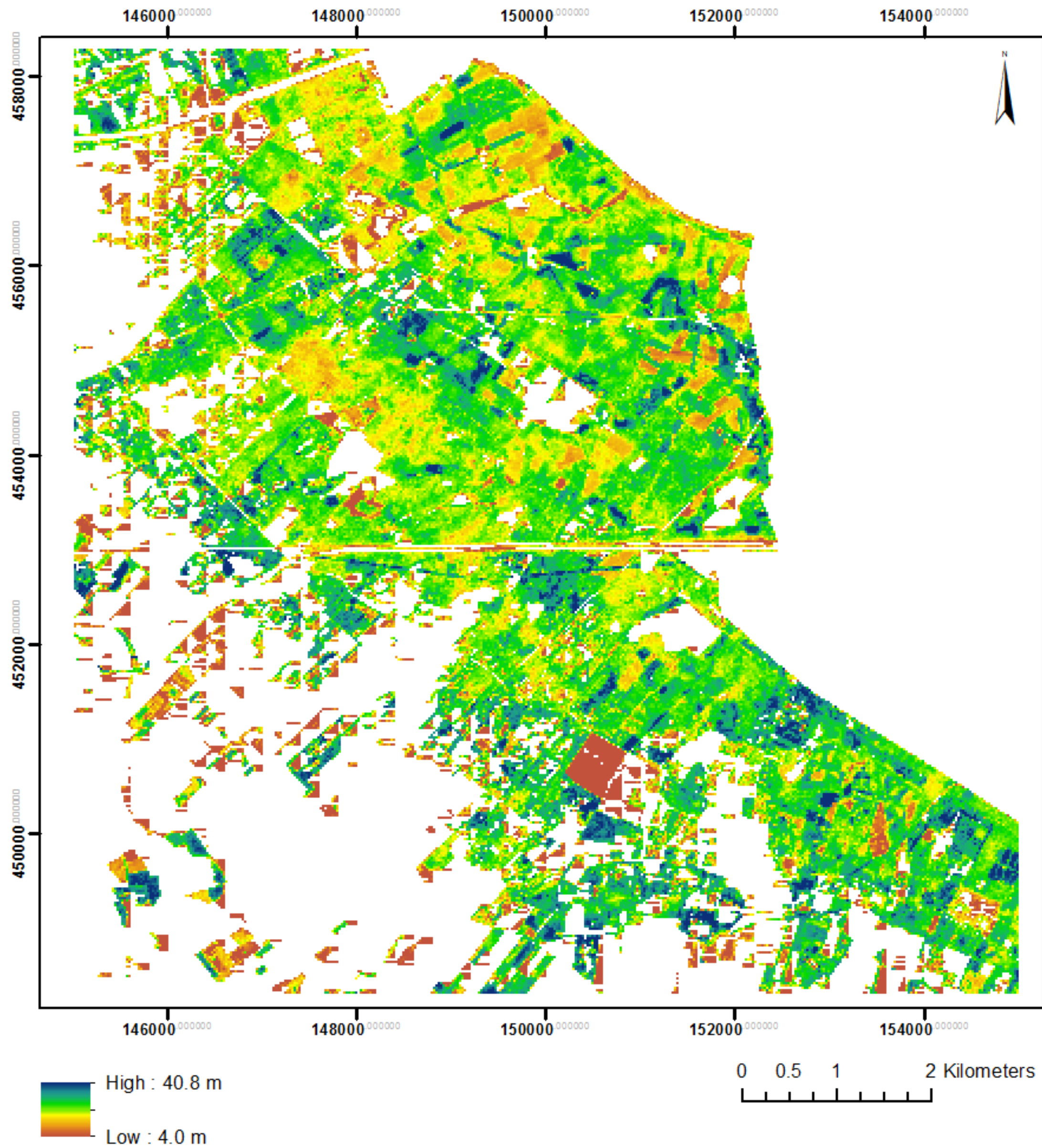

100% Scots pine
Fraction of gid cells
0,3
0,25
0,2
0,15
0,1
0,05
0
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Height class (%)


81 - 100% Japanese larch
Fraction of gird cells
0,12
0,1
0,08
0,06
0,04
0,02
0
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Height class (%)

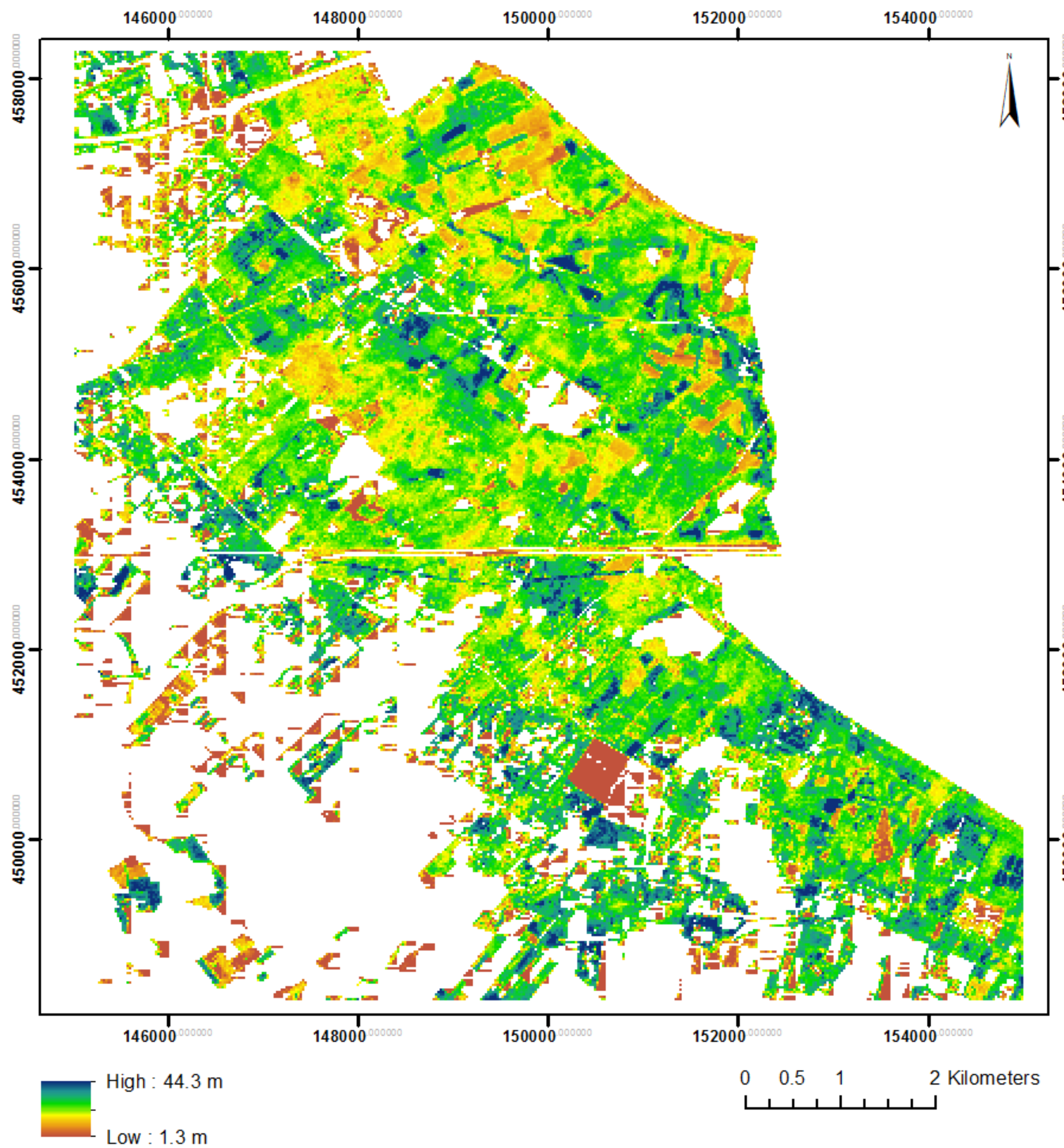

91 - 100% Austrian pine
Fraction of grid cells
0,2
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Height class (%)


91 - 100% Western Hemlock
Fraction of grid cells
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Height class (%)

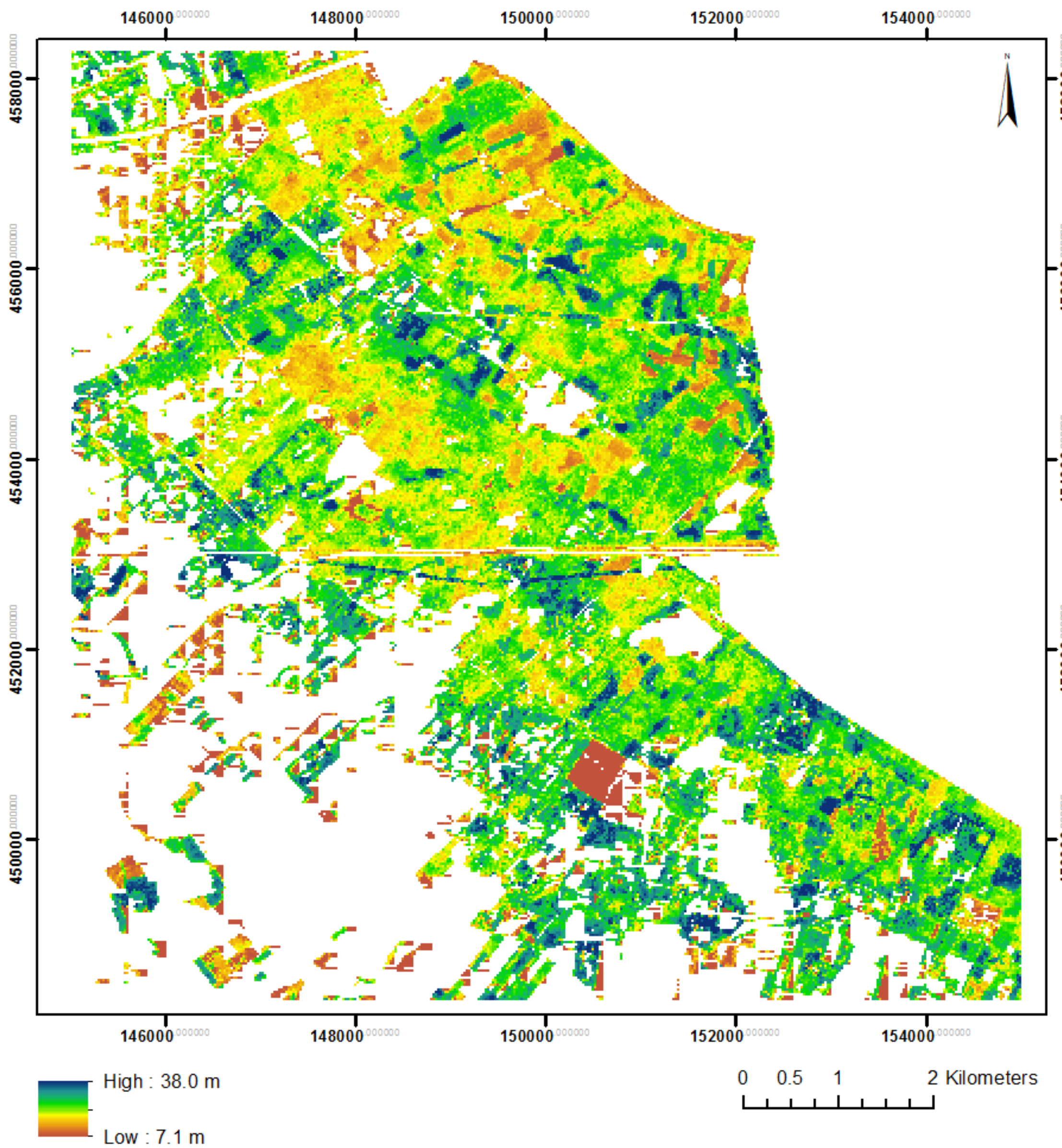
Appendix 8: maps showing dominant and maximum height



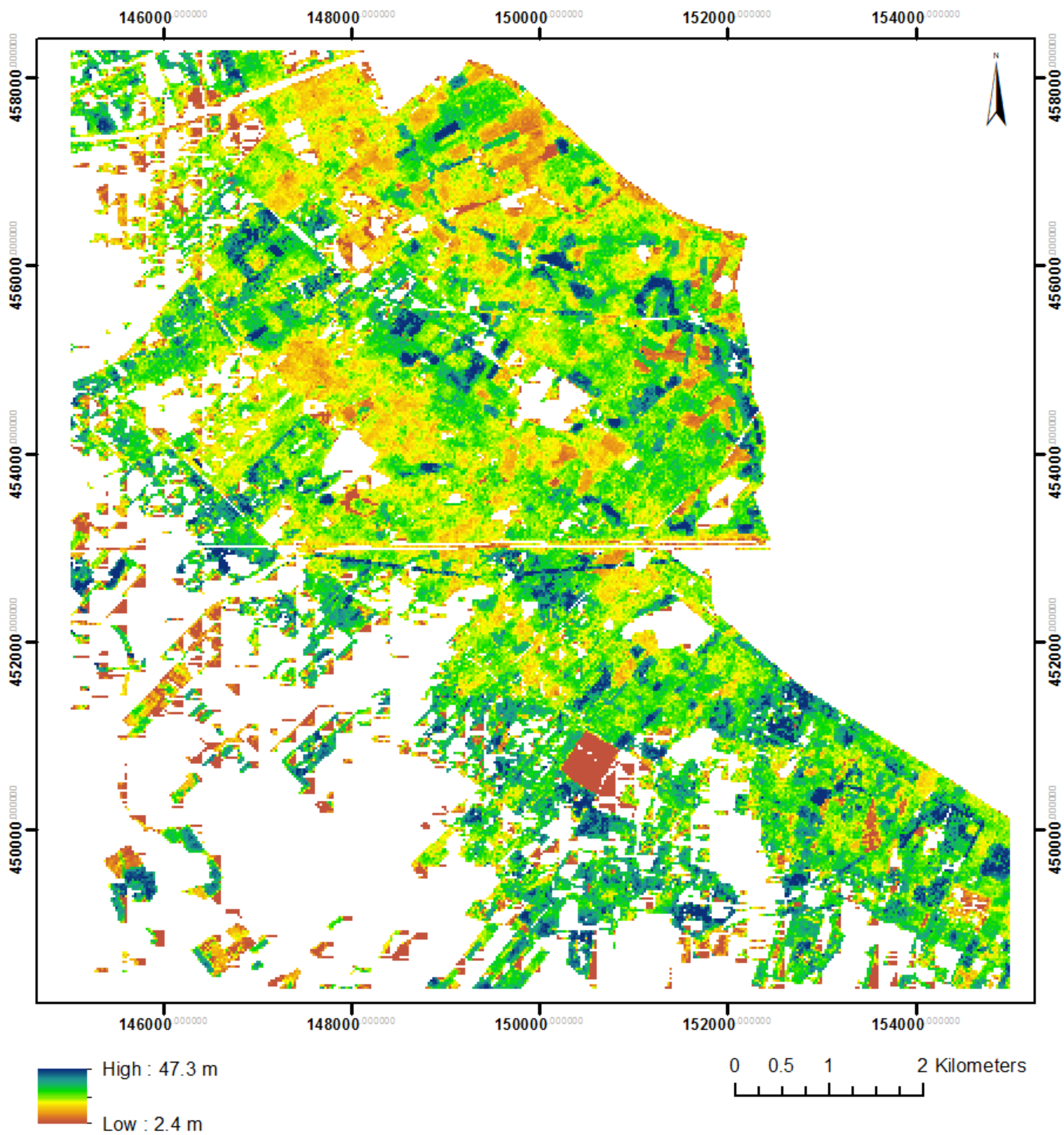
Dominant height SBB formula



Dominant height MFV formula

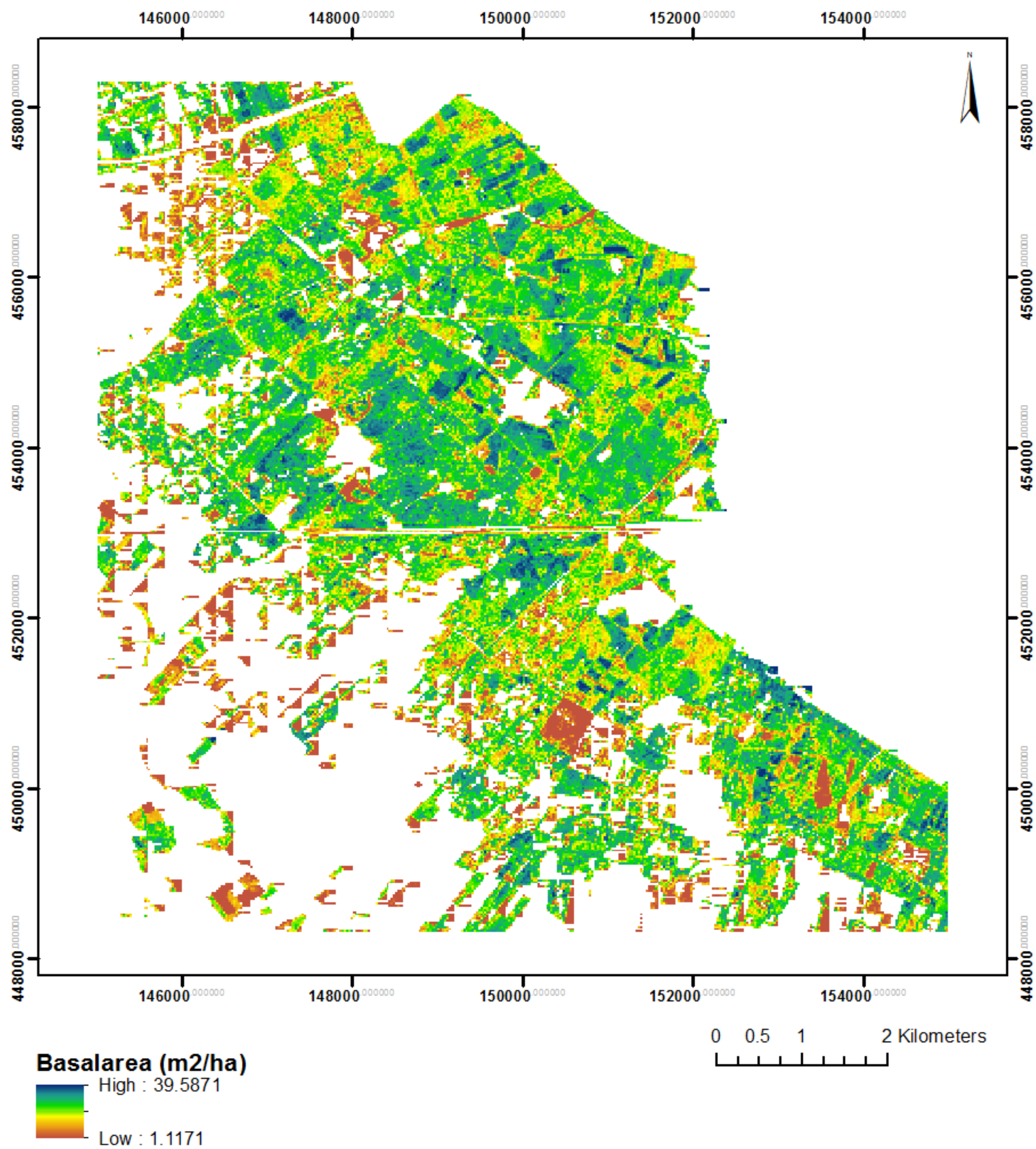


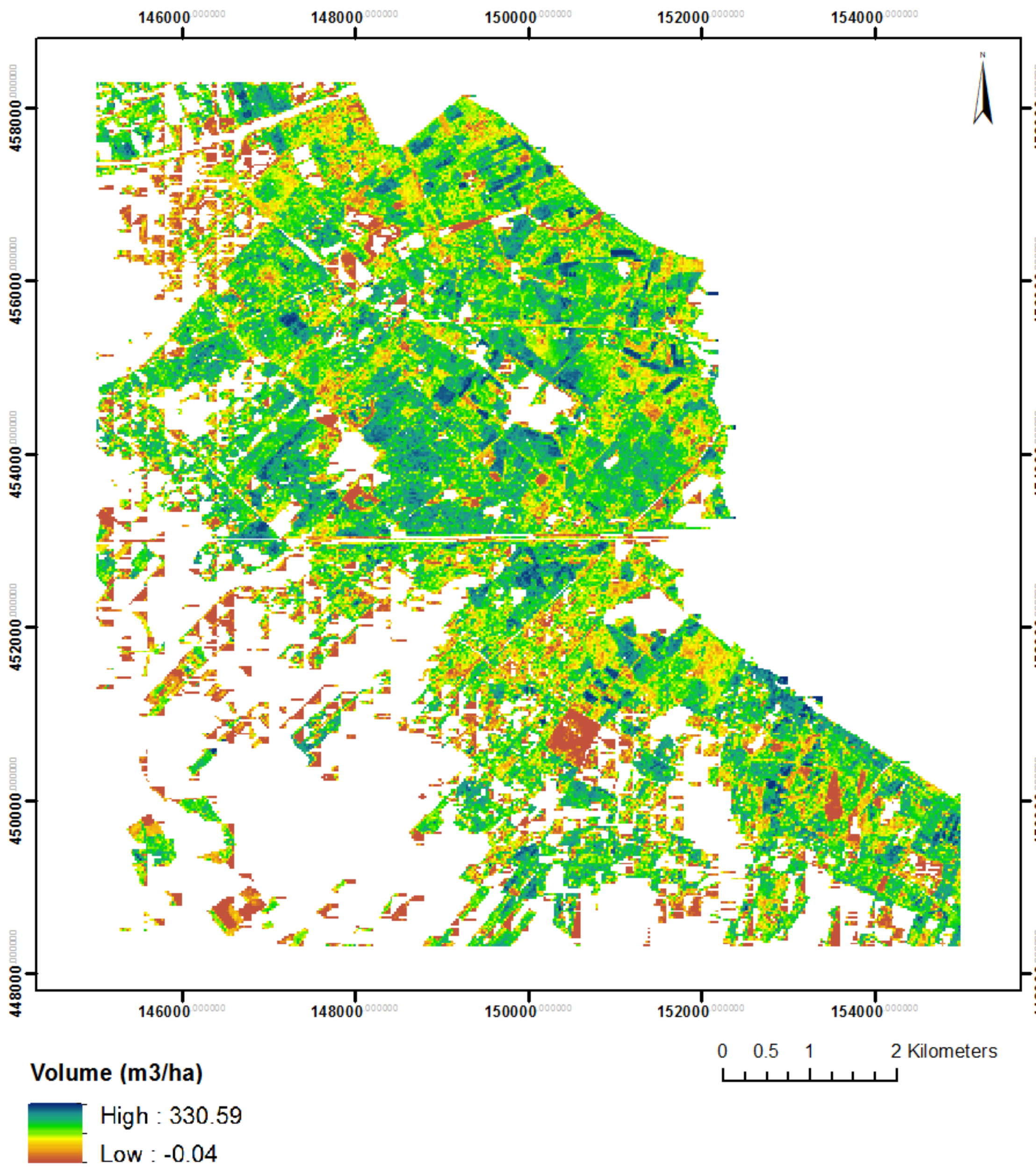
Maximum height SBB formula



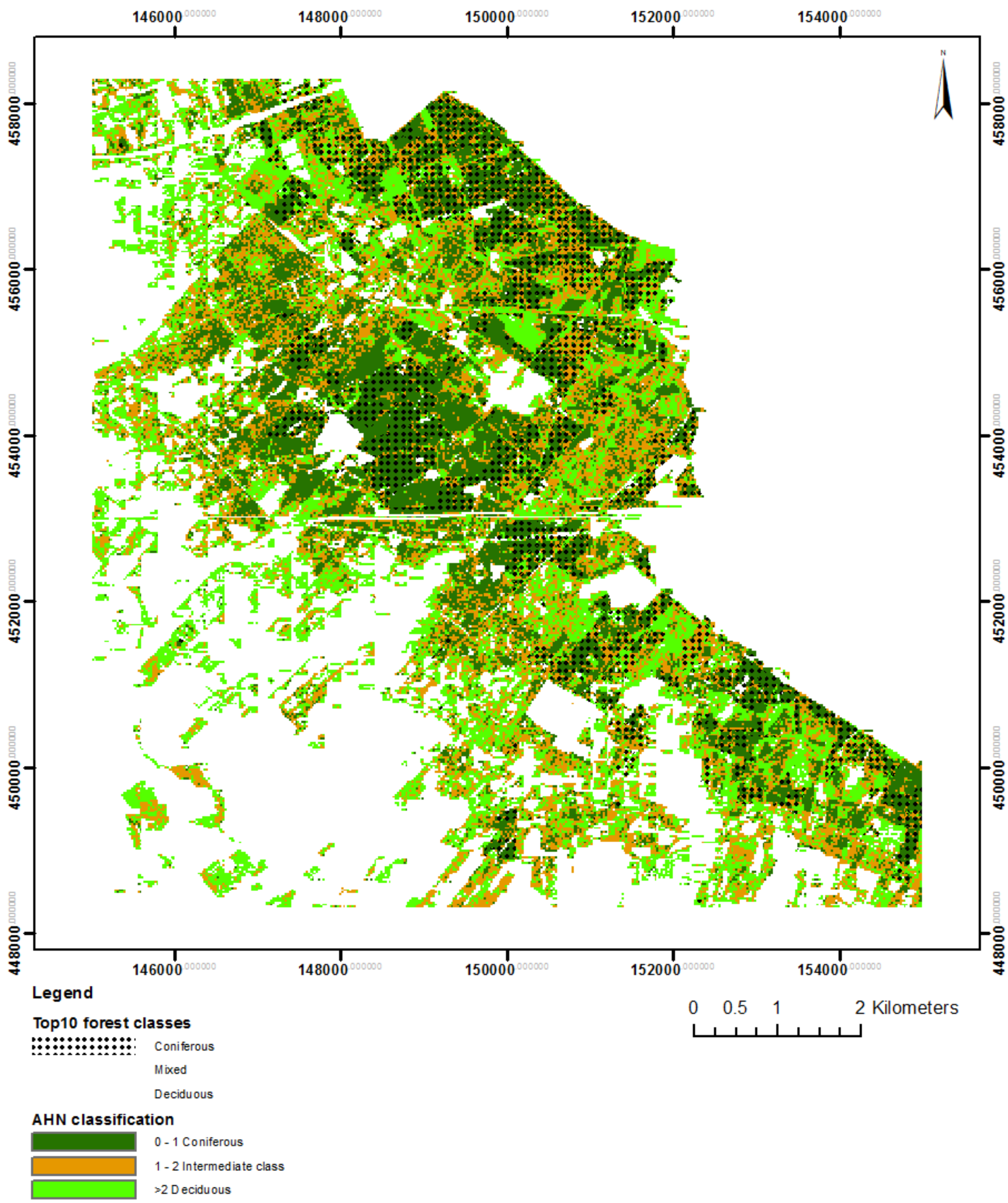
Maximum height MFV formula

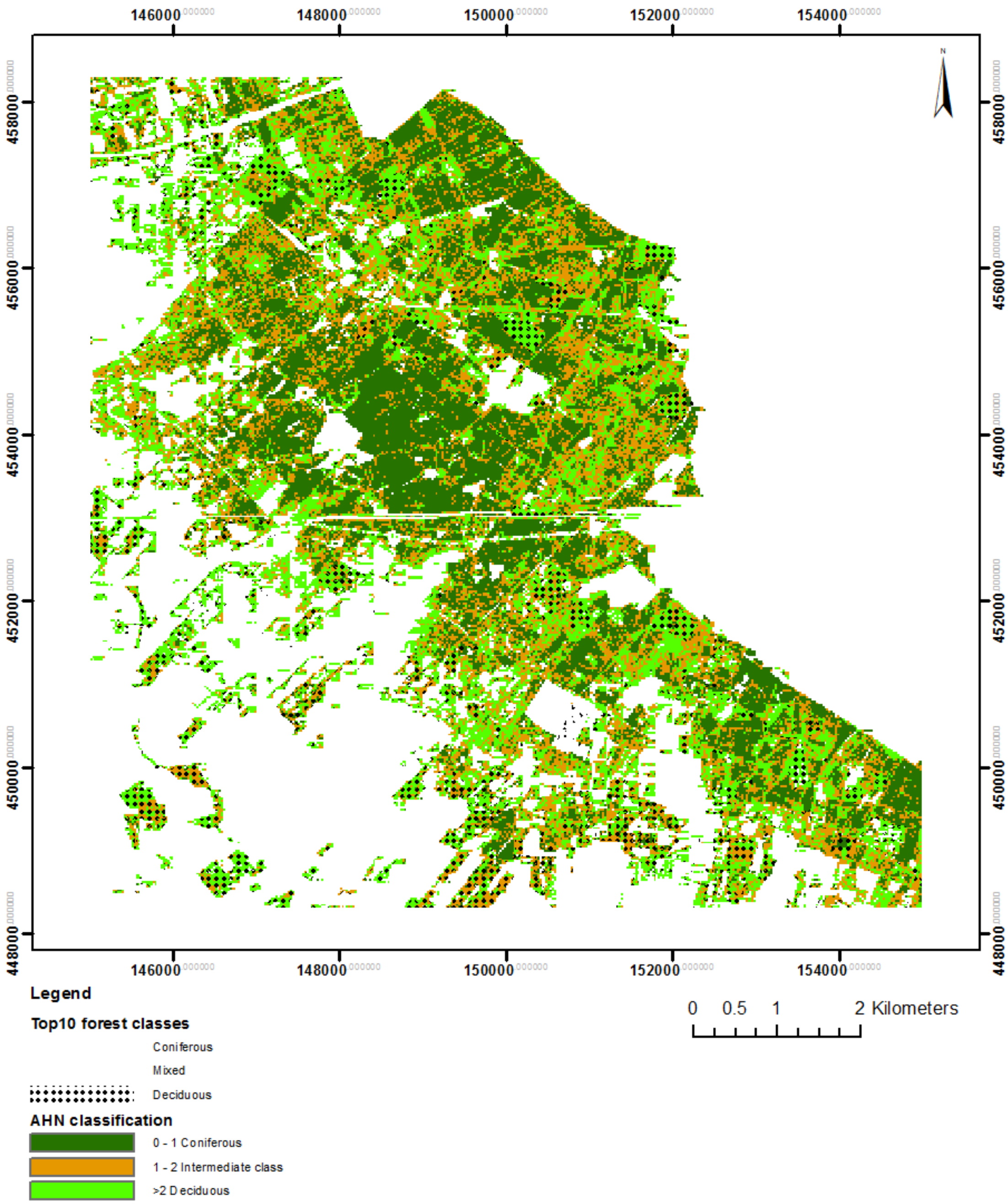
Appendix 9: Maps showing basal area and timber volume

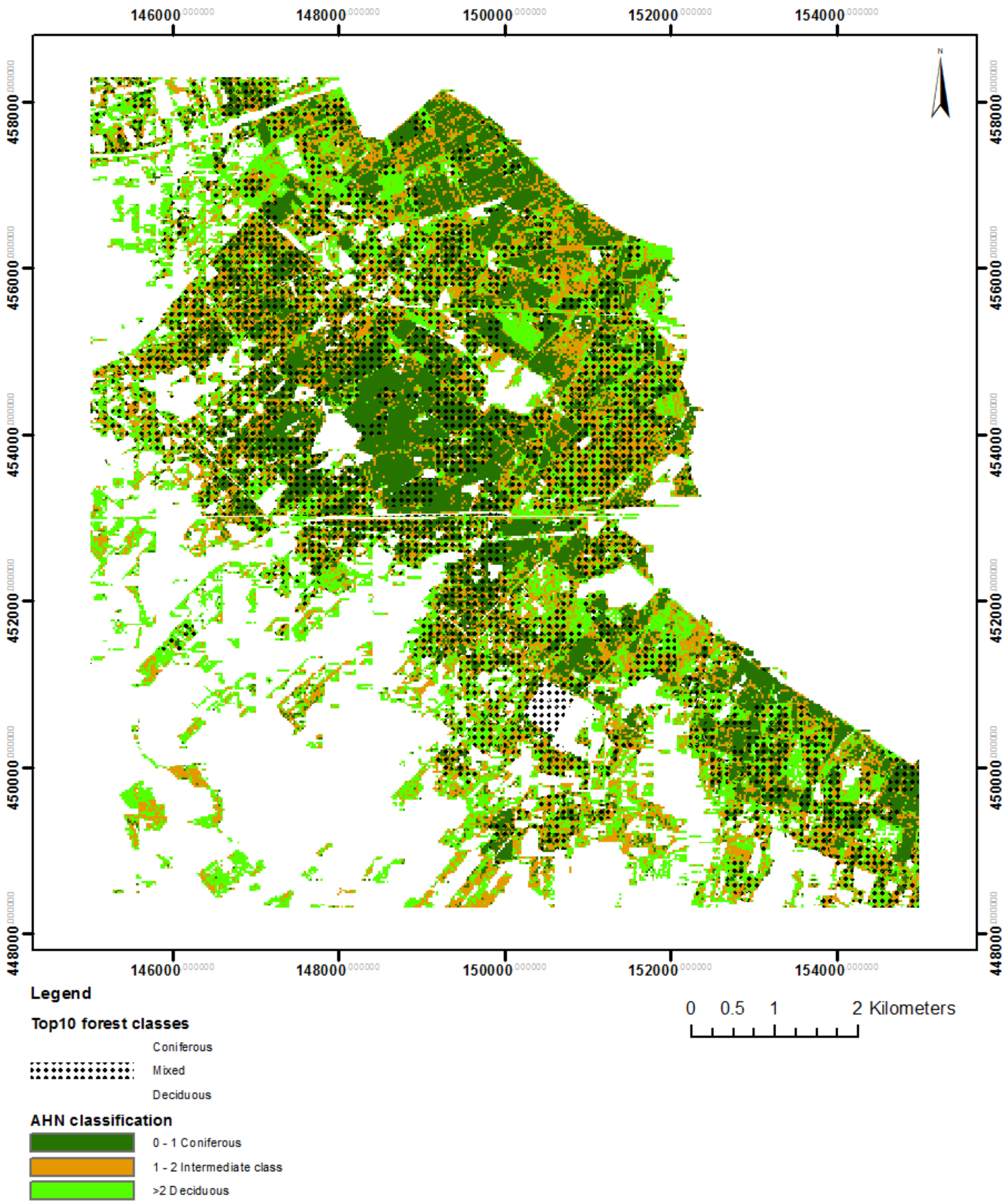



146000
148000
150000
152000
154000
458000
456000
454000
452000
450000
448000
0 0.5 1 2 Kilometers
Volume (m3/ha)
High : 330.59
Low : -0.04

Appendix 10: Top10 forest classes displayed on lidar derived classification




146000
148000
150000
152000
154000
458000
456000
454000
452000
450000
448000
Legend
Top10 forest classes
Coniferous
Mixed
Deciduous
AHN classification
0 - 1 Coniferous
1 - 2 Intermediate class
>2 Deciduous
0 0.5 1 2 Kilometers


146000
148000
150000
152000
154000
458000
456000
454000
452000
450000
448000
Legend
Top10 forest classes
Coniferous
Mixed
Deciduous
AHN classification
0 - 1 Coniferous
1 - 2 Intermediate class
>2 Deciduous
0 0.5 1 2 Kilometers

Appendix 11: Map of 4th national forest inventory