2018

Error Propagation Analysis for Remotely Sensed Aboveground Biomass

Alboabidallah, Ahmed Hussein Hamdullah

http://hdl.handle.net/10026.1/13074

---

http://dx.doi.org/10.24382/366
University of Plymouth

UNIVERSITY OF PLYMOUTH

Error Propagation Analysis for Remotely Sensed Aboveground Biomass

by

Ahmed Hussein Hamdullah Alboabidallah

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Engineering

October 2018

In the memory of my Dad Hussein Hamdullah (1954-2016)

‘The right of your father is that you know that he is your root. Without him, you would not be. Whenever you see anything in yourself, which pleases you, know that your father is the root of its blessing upon you. So praise God and thank Him in that measure.

Imam Ali Zaynul Abidin (658-713)

Treatise On Rights (Risalat al-Huquq)

Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author has been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship from the HCED (High Committee for Education Development/Office of Prime Minister/Iraq) and MTU (Middle Technical University/ Baghdad). Word count for the main body of this thesis is 34000 words. Conference papers, workshop and publications presented during the period of study are:

- ESA Sentinel-2 for Science Workshop (2014), **Accuracy Analysis for Remotely Sensed Above-Ground Biomass: Correlation between Lidar and Landsat-8 Derived Vegetation Indices.** A Poster Presentation.
- RSPSoc Wavelength (2015), **Final Accuracy Prediction Tool for Biomass Remote Sensing Projects.** Oral Presentation.
- Fifth International Conference on Remote Sensing and Geo-information of Environment: RSCy (2017) **Terrestrial laser scanning for biomass assessment and tree reconstruction: improved processing efficiency**. Oral presentation and a conference paper- SPIE.
URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10444/104440C/Terrestrial-laser-scanning-for-biomass-assessment-and-tree-reconstruction/10.1117/12.2279130.short
DOI: https://doi.org/10.1117/12.2279130
- MultiTemp (2017), **Using Landsat-8 and Sentinel-1 Data for Above-Ground Biomass Assessment in the Tamar Valley and Dartmoor**. Oral presentation and a conference paper- IEEE.
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8035247
DOI: 10.1109/Multi-Temp.2017.8035247

Word count of main body of thesis: 40000 words

Signed: Ahmed H Alboabidallah
Date: 25/10/2018

Acknowledgements

I would like to acknowledge and thank my supervisors, **Dr. Martin, Dr. Lavender**; and **Dr. Abbott.** They have offered support and encouragement throughout my PhD. I would like to acknowledge the financial support from the **HCED** and **MTU** who sponsored my scholarship and made this dream possible. I would like to thank and acknowledge the landowners of the fieldwork areas for giving me permission to do the laser scanning. I would like to thank and acknowledge the various data sources, including the **Environment Agency**, **ESA**, and **USGS**. Thanks to my big family in Iraq (my Dad, my Mum, my sisters and my brother) for their support. Thank you my little family (my wife Esraa, and my sons) for combining me and supporting me during my study in the UK.

Error Propagation Analysis for Remotely Sensed Aboveground Biomass

Ahmed H. H. Alboabidallah

Abstract

Above-Ground Biomass (AGB) assessment using remote sensing has been an active area of research since the 1970s. However, improvements in the reported accuracy of wide scale studies remain relatively small. Therefore, there is a need to improve error analysis to answer the question: Why is AGB assessment accuracy still under doubt? This project aimed to develop and implement a systematic quantitative methodology to analyse the uncertainty of remotely sensed AGB, including all perceptible error types and reducing the associated costs and computational effort required in comparison to conventional methods.

An accuracy prediction tool was designed based on previous study inputs and their outcome accuracy. The methodology used included training a neural network tool to emulate human decision making for the optimal trade-off between cost and accuracy for forest biomass surveys. The training samples were based on outputs from a number of previous biomass surveys, including 64 optical data based studies, 62 Lidar data based studies, 100 Radar data based studies, and 50 combined data studies. The tool showed promising convergent results of medium production ability. However, it might take many years until enough studies will be published to provide sufficient samples for accurate predictions.

To provide field data for the next steps, 38 plots within six sites were scanned with a Leica ScanStation P20 terrestrial laser scanner. The Terrestrial Laser Scanning (TLS) data analysis used existing techniques such as 3D voxels and applied allometric equations, alongside exploring new features such as non-plane voxel layers, parent-child relationships between layers and skeletonising tree branches to speed up the overall processing time. The results were two maps for each plot, a tree trunk map and branch map.

An error analysis tool was designed to work on three stages. Stage 1 uses a Taylor method to propagate errors from remote sensing data for the products that were used as direct inputs to the biomass assessment process. Stage 2 applies a Monte Carlo method to propagate errors from the direct remote sensing and field inputs to the mathematical model. Stage 3 includes generating an error estimation model that is trained based on the error behaviour of the training samples.

The tool was applied to four biomass assessment scenarios, and the results show that the relative error of AGB represented by the RMSE of the model fitting was high (20-35% of the AGB) in spite of the relatively high correlation coefficients. About 65% of the RMSE is due to the remote sensing and field data errors, with the remaining 35% due to the ill-defined relationship between the remote sensing data and AGB. The error component that has the largest influence was the remote sensing error (50-60% of the propagated error), with both the spatial and spectral error components having a clear influence on the total error. The influence of field data errors was close to the remote sensing data errors (40-50% of the propagated error) and its spatial and non-spatial components were also significant with ratios varying between 85-70% non-spatial and 15-30% spatial.

Overall, the study successfully traced the errors and applied certainty-scenarios using the software tool designed for this purpose. The applied novel approach allowed for a relatively fast solution when mapping errors outside the fieldwork areas.

Contents

List of Figures

List of Tables

Nomenclature

AGB	Above-Ground Biomass
ANN	Artificial Neural Network
ANOVA	the Analysis of variance
ASETR	Advanced Space-borne Thermal Emission and Reflection Radiometer
ASM	Angular Second Moment
AVHRR	the Advanced Very-High Resolution Radiometer
AVNIR-2	Advanced Visible and Near Infrared Radiometer type 2
BNG	the Ordnance Survey British National Grid
CASI II	the Compact Airborne Spectrographic Imager 2
CAVIS	Clouds, Aerosols, Vapours, Ice, and Snow
CO_2	Carbon dioxide
CPU	Central Processing Unit
DBH	Diameter at Breast Height
DCM	Digital canopy model
DOS	Dark Object Subtraction
DSM	digital surface model
DTM	Digital Terrain Model
EM	Electro-Magnetic
EO	Earth observation
EO-1 ALI	the first Earth-Observing instrument Advanced Land Imager
FCPF	the Forest Carbon Partnership Facility
GA	Genetic Algorithm

GDAL/OGR	Geospatial Data Abstraction Library / OGR simple features library
GIS	Geographic Information Systems
GLAS	Geoscience Laser Altimeter System
GLCM	Grey Level Co-occurrence Measures
GPU	Graphics Processing Unit
GUI	Graphical User Interface
H	tree height
Ha	Hectare (10000 m^2)
HCT	Hough Circle Transform
HH	Horizontal-Horizontal
HRVIR	Spot High Resolution Visible Infra-Red bands
HV	Horizontal-Vertical
ICESat	The Ice, Cloud, and Land Elevation Satellite
ID	Identification number
IDM	Inverse Difference Moment
IW	Interferometric Wide
k-NN	k-Nearest Neighbour
KP	Kyoto Protocol
Laser	Light amplification by stimulated emission of radiation
LHD	Local Heights Differences
Lidar	Light detection and ranging
MC	Monte Carlo
m.p.v.	The most probable value
MaxEnt	Maximum Entropy modelling

MERIS	Medium Resolution Imaging Spectrometer
MLC	Maximum Likelihood Classifier
MODIS	MODerate resolution Imaging Spectro-radiometer
MS	Multispectral
MTL	Landsat metadata file
NDVI	Normalized Difference Vegetation Index
NDVIR	NDVI Range
NIR	near infrared
PCs	Principal Components
PP	Profile Point
r	Correlation Coefficient
R^2	Coefficient of Determination
Radar	Radio Detection and Ranging
RAM	Random Access Memory
REDD	Reducing Emissions from Deforestation and forest degradation in Developing countries
REDD+	Reduced Emissions from Deforestation and Degradation Programme
RFE	Recursive Feature Elimination
rge	randomly generated error
RMSE	Root Mean Square Error
RS	Remote Sensing
SAR	Synthetic-Aperture Radar
SBAS	Satellite-based Augmentation System
SNR	Signal-to-Noise Ratio

SVM	Support Vector Machine
SVR	Support Vector Regressor
SWIR	Shortwave Infrared
t	Metric-tonne (1000kg)
TLS	Terrestrial Laser Scanning
TM	Landsat Thematic Mapper
TOPSAR	Terrain Observation with Progressive Scans SAR
TPU	Tensor Processing Unit
UNFCCC	United Nations Framework Convention on Climate Change
USGS	the United States Geological Survey
UTM	Universal Transverse Mercator
VH	Vertical-Horizontal
VV	Vertical-Vertical
WGS 84	World Geodetic System

Chapter 1 Introduction

1.1 Introduction

The general definition of biomass is the mass of all matter that is composed of organic compounds in a specific space and at a specific time (Achard et al., 2010). Therefore, biomass includes the mass of phytoplankton, fungi, animals and plants. However, in this study, biomass within the terrestrial environment is considered as being that within living trees. Above-ground Biomass (AGB), which is the mass in all plants' parts above the soil, is the foremost sink of terrestrial above-ground Carbon (Jandl et al., 2007), and represents the third largest source of CO_2 emissions (Keith et al., 2009).

There is an increasing interest in biomass as an environmentally friendly source of energy (Gale et al., 2005). Therefore, a number of researchers have focused on developing methods to determine vegetation biomass with applications that include Carbon sequestration, biomass as a renewable energy resource, as a pollution cause, and its impact on biodiversity as a habitat for other species e.g. Ahlström et al. (2012) and Hall et al. (2011). The importance of AGB information increased after the 1992 United Nations Framework Convention on Climate Change (UNFCCC) with many of the Carbon sequestration studies being undertaken in the context of international treaties under this framework. Good examples are the Kyoto Protocol (KP), United Nations' Reducing Emissions from Deforestation and forest degradation in Developing countries' (REDD) Programme, the Forest Carbon Partnership Facility (FCPF) of the World Bank, the Reduced Emissions from Deforestation and Degradation Programme (REDD+), and the Paris climate agreement. Despite these examples leading to strong support for scientific research on Carbon sequestration, accuracy issues have been overridden as there has been a requirement for an immediate global coverage of all relevant environmental information. Similarly, the current Paris Agreement allows only a few years before it sets in place actions in 2020 (Bodansky, 2016). There is an opportunity to make use of the accumulated experience from previous UNFCCC projects to improve the quality of the required studies and an increased need for more flexible, affordable, efficient and accurate assessment methods.

Although a considerable amount of literature has been published on AGB assessment, it remains the largest source of uncertainty in Carbon cycle computations (Houghton, 2005) causing a standard deviation of about 0.8 GtC/yr (Le Quéré et al., 2016). As a result, the estimate of terrestrial vegetation Carbon uptake has varied from about 35% of the Carbon sink (in 2010) by Le Toan et al. (2011), 27% by Le Quéré et al. (2013), and 30% by Le Quéré et al. (2015). The only way to get to the exact AGB in a specific area is by cutting down and weighing all the trees. The most accurate non-destructive alternative is to apply this destructive procedure to some of the trees and derive an allometric equation to relate non-destructive measurements, like tree trunk diameter and height, to the biomass or volume (Ketterings et al., 2001). However, this still requires high cost field measurements. Furthermore, some forest areas are remote or inaccessible. Therefore, Remote Sensing (RS) data are frequently used for biomass estimations by building the relationship between remotely sensed information and the corresponding AGB based on sample field data.

Studies have used a huge variety of remote sensing data types, bands, sensors, resolutions, and scales with no specific data combination agreed as the standard for AGB assessment. In addition, there has been a slow improvement in the AGB accuracy, even though remote sensing data have witnessed a massive improvement in spatial resolution, temporal resolution and spectral characteristics (Koch, 2010). Therefore, a full error analysis can be used to explain sources of error and might help to suggest a solution through explaining the size of the uncertainty and the share of each input and role of the processing stages in order to improve the assessment process. This way, different data combinations can be compared to each other. It is also important to focus on the weakest component i.e., improved accuracy for the remote sensing data could be degraded by low fieldwork accuracy or a bad model or vice versa.

1.2 Aim and Objectives

Aim: This study aims to develop and implement a systematic quantitative methodology to analyse the error of the remotely sensed AGB. This methodology has to include all perceptible error types and reduce the associated cost and computational effort required for conventional methods. In addition, it will investigate the accuracy of previous AGB remote sensing projects by addressing the following questions:

- What are the included errors in each component of AGB assessment models? How big is each error?
- What is the most viable technique to analyse how these errors propagate to the results?
- How big is each error's influence on the final AGB error?
- What are the weak points in the current systems, that curb the AGB assessment accuracy and need improvement, and what are the limitations of the system that are un-

improvable and so where an alternative is required?

Objectives: Four objectives are set for the study. First, to list the errors included in each input and define them quantitatively. Second, to create an accuracy prediction tool for the design stage of a standard biomass assessment; providing a rationale to breakdown time and economic costs, and to increase the cost effectiveness and resource efficiency in deriving the results. Third, to design an error propagation analysis tool that can be applied to several multi-input mathematical models (whether it is parametric based, such as regression, or non-parametric, such as an artificial neural network), and can be applied with several analysis techniques (whether it is pixel-based or object-based) with minimum computational costs. Fourth, to integrate this tool into tangible models with real data to provide a quantitative answer to the main question.

1.3 Thesis Structure

Chapter 2 situates the error analysis within the context of the academic community of spatial based subjects, and for AGB in particular. It provides a critical review of the relevant literature and points out a number of missing components within the error analysis for AGB assessments. Chapter 3 presents an error prediction tool based on the reported results from previous studies. It also provides a review to understand the data set, and the different components and the relevant techniques that are built into an AGB assessment system. Chapter 4 outlines the methodology of the terrestrial Laser scanning and the fieldwork data analysis. It also shows samples of the fieldwork results that are used for the scenarios in Chapter 5. Chapter 5 aims to focus on the error analysis tool developed within this research; including the different techniques, principals, and ideas used to develop it. It also discusses the calculation of errors for system RS and field data.

Chapter 6 develops a range of scenarios to test the applicability of the error analysis tool and provide answers to the research questions. Chapter 7 discusses the findings of and results within Chapters 2 to 6 and provides a list of recommendations for future work based on these findings. Chapter 8 concludes the findings of the research.

Chapter 2 Above-Ground Biomass Error Modelling

2.1 Introduction

AGB plays a key role in understanding the Carbon cycle. However, the accuracy of AGB assessment is still being questioned; it is agreed that AGB assessments include large errors (Le Quere et al. 2016; Réjou-Méchain et al., 2017; Njana, 2017), but it is not agreed how large these errors are. The numerous AGB assessment approaches, variety of available data sources and complexity of biomass make an accuracy assessment hard to achieve.

The effect of AGB on the Carbon cycle is consistently assessed by calculating it as the residual of all other Carbon cycle components. Therefore, it includes relative errors of more than 30% of the Carbon cycle errors. In other words, it is expected to cause a standard deviation of ±0.8 GtC/yr (Le Quéré et al., 2013; Le Quéré et al., 2014; Le Quéré et al., 2016). Also, it almost equals the increase in CO_2 emissions from human activities for the last decade (2005-2016) (Le Quéré et al., 2016).

RS based projects in general, and biomass assessment procedures specifically, usually process input data to achieve a final result. The processing is mathematically based, referred to as the mathematical model, and usually based on fieldwork data. Therefore, neither the inputs nor the mathematical models are perfect due to the fact that each RS input and each fieldwork data input has some random errors. This imperfection raises a question about how the final result is affected by the inputs and processing, with error analysis aiming to answer this.

Practically, each measurement can provide the most probable value (m.p.v.) of certain quantities with a margin of statistically evaluated random error. If this error is within the tolerance limits, it is acceptable, with these limits usually set according to how the error would affect the results. The process of the error affecting the results is called error propagation.

2.2 Techniques for error analysis

Where error tracking is needed for straightforward remote sensing applications a simple error propagation technique can be used, such as the brute force method that propagates errors in each stage by applying all possibilities without exceptions e.g., Boger et al. (2003). However, the approach makes this method very computationally expensive. In other applications, the arithmetic processes (mathematical models) are simplified mathematical relationships between real world inputs and outputs based only on linear equations. Here, errors could be derived at each stage using the chi-square method, described in Andrae (2010), which assumes the error distribution of the measured data is Gaussian. Andrae et al. (2010) states that this method is unmanageable for models with non-linear equations. Therefore, when simulating a complex reality with multivariate and high-dimensional data error outputs are harder to determine because the processing involves complicated systems of non-linear equation models. In some more complex models, error estimation via analytically driven methods could be applied with example deterministic methods being Taylor Methods and Rosenblueth's method (Heuvelink, 1998). In Taylor methods (e.g. delta method, Fisher matrix method, and second order Taylor method), any non-linear equation can be linearized using a Taylor Series. Methods such as the delta method approximate this equation by including only the first derivative terms. Some methods like the second order Taylor method only take the first and second derivative terms of the equations into consideration as the other derivatives are assumed to be small (Heuvelink, 1998). Rosenblueth's method uses a similar approach but with derivatives being computed numerically. These methods are based on the equations calculus. Therefore, they can only be applied when the equation set is well known. However, some models do not require a detailed understanding of the processes of converting inputs to outputs e.g., the use of stochastic methods. Examples of these are the Monte Carlo (MC) method, bootstrap method, and random set method.

The MC method repeatedly applies the model or system to the study. Each time, different randomly generated input errors are applied to the most probable input values; simulating both the expected range of error for each individual input. If the process is repeated a sufficient number of times, the group of outputs will be the whole distribution of expected outputs and can be statistically analysed to compute the most probable values, errors, variances and other parameters (Lee, 1993).

MC is easy to use because it is simple, direct and robust (Caflisch, 1998). The main drawback is that it is computationally expensive because it requires applying the model again and again. The other problem with this method is the random error generation process. Instead of generating full random numbers, computer algorithms can only generate pseudo-random numbers that have many of the properties but are not exactly random. Therefore, occasionally problems can occur with very long sequences (Caflisch 1998). Some studies have tried to improve the speed through improving the random sampling. For example, Caflisch (1998) used low-discrepancy quasi-random sequences. The bootstrap method is similar to MC method, in using repeatedly computed outputs.

However, it uses the original observation rather than adding errors to the m.p.v. of each input. Computationally, this method is less expensive than the MC method. It works by creating many randomly chosen samples called "bootstrap sample sets" from the training set and then applies the model to compute a pseudo-Y value based on each bootstrap sample set (Efron and Tibshirani, 1994). Instead of using new data not included in the training set (e.g. random errors generated for MC), the bootstrap world can imitate this situation by sampling randomly from the existing data to create different training sets (Tibshirani, 1996) as much as the process can afford (Efron and Tibshirani, 1994). However, sampling randomly from the existing data requires redundancy in the existing data to allow calculation to be undertaken with subset. For example, if a system is designed to be tested for 4 sample plots in each iteration. When, a 10000 iterations MC analysis to be undertaken, in each iteration, a simulated random error is to be added to each sample. Therefore, it will require a field data for four sample-plots coverage. For the same system, when 10000 iterations bootstrapping is to be applied, each iteration is required to be based on a unique set of samples. Therefore, the number of sample plots can be calculated based on combinations by applying:

$$\binom{n}{k} = I \tag{2-1}$$

Where n is the number of required samples, k is the number of samples for each iteration (in the example k=4), and I is the number of iterations (in the example I=10000). Which means that n is about 24 sample plots. Compared with the costs of the fieldwork, this increment in the required fieldwork can be significant.

Each one of these methods has its requirements and costs. Generally, the requirements can limit the applicability of the method for some application due to the mathematical analysis that does not match the requirements. In addition, the costs can limit the applicability for some models due to the analysis scale and the analysed data size. Therefore, none of these methods are perfect: some do not apply to all types of operations, others are extremely time consuming or involve large approximation errors (Heuvelink, 1999).

2.3 Error Analysis of Remote Sensing Applications

Remotely sensed environmental topics (such as geography, geology, hydrology, agriculture and ocean colour remote sensing) usually include many types of data with a range of errors. Some of the errors are systematic and can be determined and eliminated in the pre-processing. Other errors are random and cannot be eliminated and therefore their effects on the results should be analysed. Therefore, many different error analysis techniques have been applied but can be divided into two levels, results verification, and sensitivity analysis. The first level which is the most common application is results verification analysis. It involves studying the errors of the applied system by studying the results disparity (Giacco et al., 2010) or by comparing the results with fieldwork data (Handcock et al., 2006); providing a level of trust in the results. The second level is the sensitivity analysis, which is a process of error propagation or weighting input errors (Jager and King, 2004) and to study how they propagate to the results. In other words, the process seeks to produce a reliable ranking for input variables by understanding the influence of the errors on the systems' results e.g. Congalton (1991). The sensitivity analysis can be extended to be a total error budget analysis. Therefore, it can be implemented in order to determine the characteristics of a remote sensing system (Cocard et al., 1991), assess the accuracy of the system’s mathematical model (Sabia et al., 2008), choose between models (Shataee et al., 2011), improve a specific model

(Hodgson and Bresnahan, 2004), or find the optimal fusion of several models (Giacco et al., 2010). This type of analysis provides the most detailed information about errors. However, it requires a full understanding of error sources and errors propagation. Therefore, it is the most computationally costliest analysis.

As with other research fields, error analysis tends to involve systems that have very elaborate mathematical models. Methods that have already been investigated in the literature, which deal with propagating errors within spatial data, include all previously listed error analysis techniques. However, unlike in AGB remote sensing, in other RS fields there are many examples of applying successful error propagation analysis tools that are designed to adapt to these challenges by using advance computer hardware and recent software improvements. One example of successful error simulation is Ebrahimi et al. (2013) who built a Monte Carlo simulation of errors in remote sensing data using random processes theory. Another example is Goulden et al. (2016) practical experiments of error analyses for RS data error propagation within terrain surface models. An example of applying a combination of two methods is Marinelli et al. (2007), which used a first order Taylor series method, and the Monte Carlo simulation method to analyse the accuracy of agricultural monitoring data. The study shows how to use Taylor series and Monte Carlo simulation as full analytical methods for spatially varying errors. Although it did not deal with biomass, it showed how error surfaces could be achieved. The resulting accuracy map provides a comprehensive understanding of accuracy at each pixel. Another successful example is Chen et al. (2013) who applied a first order Taylor method to track error propagation within experimental sampling, using Landsat and AVHRR data to estimate foliage biomass. Similarly, Zhao (2012) and Zhao et al. (2011)

investigated the error in EO data based on random set theory. These successful examples suggest that these methods can be applied for AGB remote sensing.

2.4 Error Analysis for Aboveground Biomass

As a part of any Carbon sequestration terrestrial ecosystem service, an accuracy assessment is required for the estimation of AGB (Hill et al., 2013). There are many major research efforts focused on the capabilities of remote sensing data and techniques to help better understand AGB within different scales. However, accuracy and accuracy assessment is still among the most pressing problems that these studies are facing.

(Han et al., 2016) suggested that the low accuracy is caused by errors from inter-annual and the overall AGB dynamic changes. Namayanga (2002) and Hill et al. (2013) take it to mean the failure of the algorithms being used. Others such as Hunt Jr et al. (2002) attribute it to the high number of samples that are needed to achieve a valid inference with a large number of input parameters that are generally not available for most studies due to the large study areas compared with budget limitations. Since the 'AGB error' can be of an unknown origin and even include systematic errors (Hill et al., 2013), AGB is often poorly estimated. Therefore, almost every paper addressing AGB analyses the accuracy of the results, which can be grouped as follows: -

2.4.1 Correlation Coefficients and Coefficient of Determination

The Correlation Coefficient (r) and Coefficient of Determination (R^2) are statistical indicators of how well the field data measurements and RS based AGB relate to each other. It is possibly the most used accuracy indicator for AGB and therefore there are hundreds of examples including almost all the regression based studies in Appendix A.

Suresh et al. (2014) added confidence interval bounds to the regression coefficients. However, this assessment was still unable to understand how each error source affected the result. Furthermore, this approach is often limited to the system training samples and is not able to provide spatially based error maps.

2.4.2 Verification and Validation

Most studies use the general criteria for split sample validation and/or correlation coefficient analysis. The general criteria for split validation sample includes keeping a portion of the field samples out of the model computations. Then, the Root Mean Square Error (RMSE) can be obtained by comparing the field data to the RS based assessment results. This type of analysis is used either to justify a study result (e.g. Vaglio et al. 2014), to compare results of different types of RS data (e.g. Holopainen et al., 2010) or to compare results from different models (e.g. Xie et al., 2009). Although, this procedure is statistically valid, it has its drawbacks regarding AGB assessment. It can be completed only at the last stage of the project, i.e. when the whole fieldwork and RS data are collected and the assessment process is completed. The resulting RMSE can give an indication of the whole process accuracy, but it cannot explain where the error came from. Moreover, the accuracy variation with geographic position cannot be visualised (as an error surface or error map) because the error is a unified RMSE error for the whole area or for each vegetation species.

2.4.3 Propagation of Individual Error Type

Some studies deal with only some types of error. Sherrill et al. (2006) used two-stage Monte Carlo analysis to study the propagation of errors related to the field measurements, sampling error, error in the allometric equation (the equation for deriving biomass from field measurements like trunk diameter and tree height), and error that was associated with the regression equations for AGB estimation using Lidar data. The study found that assessing AGB for the whole landscape (instead of identical land cover areas) resulted in a dramatic drop in confidence intervals. Lo (2005) implemented the Gaussian error propagation method for allometric equation applications, and addressed the importance of ‘error budgets’ and ‘sensitivity indices’ in understanding error, and the most effective way to identify and reduce those errors. Similarly, Berger et al. (2014), Breidenbach et al. (2014) and Molto et al. (2013) used a Monte Carlo scheme in their studies of error propagation from field sampling to AGB. Chave et al. (2004) divided field errors into tree measurement error, allometric model error, sampling error and miss-representativeness of the field data samples. They pointed out that the most important source of error in the field based sampling is the model error. However, all these studies were limited to applications at the field sampling work stage only. This approach was improved by Fu et al. (2017) by introducing Taylor principles into a Monte Carlo simulation procedure to analyse quantifying the corresponding uncertainty arising from both sampling and the regression model.

In contrast, under the assumption that the allometric models are perfect, Hensley et al. (2014) studied RS data errors only; not addressing fieldwork errors or model errors that could affect the high accuracy of the RS data. They used a first-order Taylor-expansion method for assessing biomass estimation accuracy based on nominal properties of the implemented RS data. One of the study outcomes was that biomass accuracies could be highly effected by RS data characteristics. Similarly, Berra et al. (2014) studied the effects of pre-processing of RS data only on the accuracy of the results. Pacala et al. (1996) analysed the error propagation from the final results of biomass of each individual tree to the future prediction of the same tree and its neighbour. This analysis was applied on the error propagation in the post-processing stage of the AGB results rather than error propagation inside the model. Zhang et al. (2014) offered an estimation of AGB errors by comparing the study results with existing AGB estimates of the same area. The most important contribution of this study is that it provided a map of AGB error, depending on RMSE, for all pixels. Weisbin et al. (2014) used a total least squares approach for studying the propagation of errors from RS based canopy height, and vegetation indices to the assessed AGB. This study argues that this approach could be applied to a range of data and regression models, deployed to achieve a balance between costs and benefits. However, this approach can be applied only when the equations to relate AGB per unit area to measurements made with RS data are known. This means that it is not applicable with models that are not equation based.

2.4.4 Propagation of all Expected Error Types

From the review of previous studies, errors can be classified (as shown in Figure 2-1) into field data errors represented by field measurement errors, allometric equation errors, and RS data errors. In addition, these errors might propagate to the model to generate model errors and then to the model based results as shown in Figure 2-1 resulting in error in the final AGB estimation. Ahmed (2012) provides one of the most complete classifications to date of errors in AGB assessment, based on two types of RS data, with RS errors represented by the impact of measurement errors. Whereas, model errors represent the miscorrelation between field and RS datasets, and the distinct bias in RS value between the deciduous and coniferous trees contributes, he suggests that the field data errors vary considerably with the type of allometric equations used. However, the study is limited to the error description and does not study error propagation.

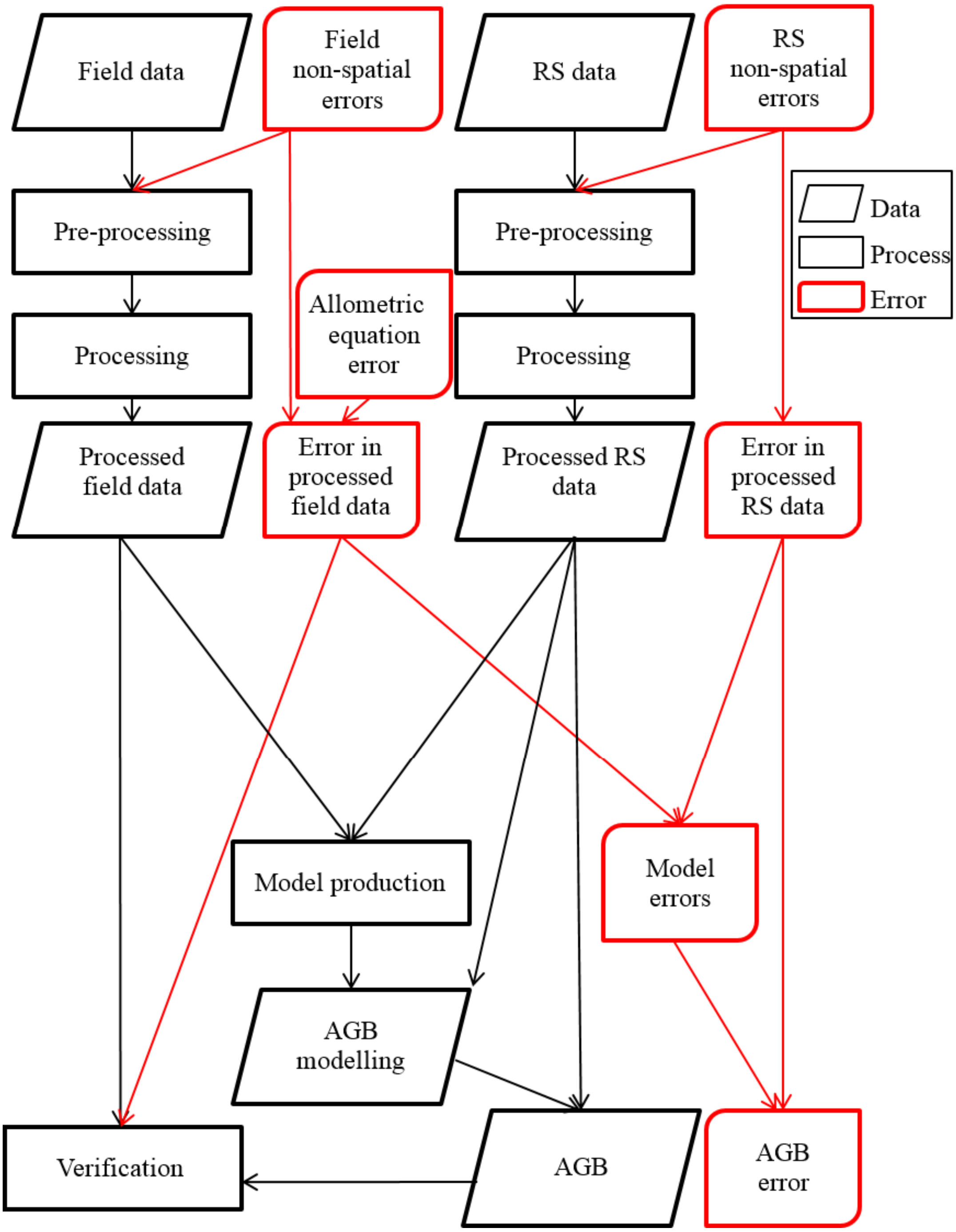


Figure 2-1: The process of AGB assessment and the outline of a comprehensive error propagation through it based on previous studies. The assessment steps are shown in black, and the error propagation steps are shown in red. The process starts with the field and remote sensing data and completes with the derived AGB.

The Colgan et al. (2013) study depends on destructive harvesting of trees to assess allometric equation model errors beside the nominal vertical error of the canopy height map derived from the implemented airborne Lidar and the applied mathematical model errors. The study results show that about one third of AGB error was related to the field based estimation, almost one half was related to the Lidar data and the remaining errors are model errors. The Weisbin et al. (2014) revolutionary study was the first study that dealt with all error types described in other studies without destructive tests; errors were propagated through a simple linear RS data fusion model. In addition, the AGB deriving model was approximated to be linear avoiding the use of non-linear models that require more computationally expensive approaches.

The most comprehensive study for AGB error analysis up to now is Chen et al. (2015). The study deals with all error types in each processing stage of a non-linear model with one RS input (airborne Lidar data) and the application of a Taylor method. This approach can be applied when the model equation has a direct well-defined relationship to the error under investigation. However, with non-parametric mathematical models, previous studies based on verification and validation analysis.

However, although some studies have used linear models, non-linear models and non-parametric models have been more widely used for AGB assessments (Chen, 2013). These models are not analysable with the listed approaches. Furthermore, these studies dealt with non-spatial errors of RS and field data based on RS data characteristics such as Signal-to-Noise Ratio (SNR) for RS data and the allometric equation error for the fieldwork data; there are number of spatial errors that have no direct relationship to the model that were not included. For example, RS data geo-location error cannot be analysed using these models and there are currently no studies within the literature that deal with this error. The geo-location error is important since it directly related to the geographic correspondence between the field data and the RS data. In other words, the geo-location is the link that pair between each pixel and its corresponding part of the fieldwork. Therefore, the error shifts each pixel to be paired with an inaccurate part of the fieldwork. In addition, all the listed studies deal with pixel-based systems; there is no available error budget analysis for object-based AGB assessment models. Besides, all studies deal with completed projects and there is no error assessment for projects during the system design stage.

2.5 Summary

This chapter discussed the available error analysis techniques for remotely sensed AGB, in order to determine the limitations that curb the inclusion of error analysis. It has identified two levels of error analysis, results verification, and sensitivity analysis. The vast majority of the literature involves a result based verification analysis with a more limited number of studies having undertaken an error analysis of the inputs, highlighting how these errors propagate to the results.

An error analysis approach requires a full understanding of error sources and error propagation, and involves computationally expensive calculations. Therefore, the error propagation analysis in the previous studies was either limited to one input error, or limited to a specific mathematical type. In general, to solve these limitations, new efforts are required to deal with these four steps:

- Error prediction for projects during the system design stage.
- Error analysis for non-parametric and non-linear parametric models.
- Error analysis that deals with spatial errors such as geo-referencing error.
- Error analysis for object-based models.

Taken together, these conclusions lead on to the following chapters that deal with these issues in order to provide a better understanding to the uncertainty of the AGB and to produce a methodology to solve these problems.

Chapter 3 Accuracy Prediction Tool for Above-Ground Biomass Assessments

3.1 Introduction

Above-Ground Biomass (AGB) studies can be classified into two main types:

- experiment studies that try to study a new data type, a new combination of data types or include a new analyses technique to extract AGB information, and
- AGB survey projects that try to design AGB assessment systems based on the outcomes and recommendations of experiment studies.

A few studies, such as (Weisbin et al., 2014), have decided the required accuracy and then tried to find the minimum cost solution to reach those requirements. Alternatively, some studies try to compare the accuracies of different techniques during the design stage; depending on previous study results. For example, Zhao et al. (2012) compared between allometric equations, Goetz et al. (2009) compared between some RS sensors, and Xie et al. (2009) compared between mathematical models. However, previous AGB studies have not employed a comprehensive analysis to assess accuracy in advance.

In addition, the highly heterogeneous combinations of fieldwork data, RS data, analysis systems, and mathematical models available for AGB remote sensing make the decision complex and requiring a high level of experience. This chapter presents the development of a tool to predict the accuracy of projects during the system design stage or half-finished projects, depending on the characteristics of the available data, study area and processing model. In addition, the chapter reviews the published literature in search of the parameters that can affect the accuracy of the assessed AGB. These parameters were used as a base to build the database to train and validate the error prediction tool. Hence, a set of data, shown in Appendix A, was built based on this literature survey.

3.2 Effective Parameters for Biomass Uncertainty

Based on the error propagation analysis of the previous studies shown in Figure 2-1, the parameters that affect the uncertainty in the results of the assessment system can be classified into the parameters that are related to the remote sensing data, fieldwork data and applied mathematical model.

3.2.1 Remote Sensing Data Parameters

Currently, no standard dataset is routinely used for AGB assessment. Hence, dataset selection often effects the set of parameters that in turn effects the accuracy as described in the following sections. Generally, remote sensing could be divided into passive and active systems with both types of systems widely used for AGB.

3.2.1.1 Passive Remote Sensing

This section summarises the parameters related to the use of data from passive sensors i.e., when the sensing system is illuminated by an external energy source (normally the Sun, but can also be the Earth). This type of remote sensing provided information to about horizontal tree land-cover structure that was correlated to AGB based on land-cover classification (Gjertsen 2007), land-cover structure analysis (Hall et al. 2011), and tree land-cover interaction with different electromagnetic energy wavelengths (Bao et al., 2009). The main parameter with this type of data is the spatial resolution, indicated by pixel size. Low and medium spatial resolution datasets are usually space-borne and have pixel sizes of 100 by 100 m or larger with a wide spatial coverage. High spatial resolution satellite data have pixel sizes of between 4 and 100 m, whereas very-high spatial resolution data would have a pixel size smaller than 4 m. Biomass studies use a wide range of spatial resolution data depending on the study scale, financial budget, required accuracy and data availability. The AGB assessment accuracy is also influenced by other vital specifications like temporal resolution (Askne et al., 2013), revisit time (Askne and Santoro, 2012), spectral coverage (Foster et al., 2002), number of bands (Daliakopoulos et al., 2009) and spectral resolution (Bao et al., 2009). Spectral resolution means the ability of the sensor to separate EM signal components according to each component's wavelength. Therefore, it can be relevant to the quality of the AGB assessment when the assessment includes computations that are based on more than one band (Bao et al., 2009). For example, vegetation indices (VIs) can be based on how green leaves interacts with different EM wavelengths (usually red and near infra-red) to indicate the health and density of vegetation cover. These indices are suitable for many different sensors, which have these broad bands, while narrowband greenness indices focus on specific aspects of plant vitality or vigor, e.g. change in pigment content during the growing season, and are restricted to sensors with specific bands.

Medium Spatial Resolution Data: Medium resolution data are used in almost all global coverage studies. Also, they are widely used for small area studies down to a country level. These studies depend on the expected correlation between AGB and the vegetation indices. Vegetation indices determining the volume of a vegetation cover through the interactions of green leaves with specific spectral wavelength (usually based on Red and Infrared bands). Therefore, they represent an integrative measure of both vegetation photosynthetic activity and canopy structural variation and consequently to the AGB. The main criteria for choosing datasets are the spatial coverage and the low cost. However, the correlation achieved by this type of data is low. For example, Lokupitiya et al. (2010) achieved an R^2 of <0.4 by using the Advanced Very-High Resolution Radiometer (AVHRR) data, Lokupitiya et al. (2010) reached an R^2 of 0.31 with Medium Resolution Imaging Spectrometer (MERIS), and Yuan et al. (2016) achieved a range of R^2 of between 0.01 and 0.75 with MODerate resolution Imaging Spectroradiometer (MODIS) data.

High Spatial Resolution Data: Over recent years, high spatial resolution satellite data have become widely used due to technical developments that have led to reduced costs, more powerful computing hardware and software and improvements in the ability of Geographic Information Systems (GIS) to handle these data. Extensive research using satellite imagery for mapping and monitoring biomass has been conducted over the last 40 years using high-resolution data. The data were provided with different sources such as the Landsat time series multispectral sensors. Examples could be Landsat Thematic Mapper (TM) e.g., Lu et al. (2002), Spot High Resolution Visible Infra-Red bands (HRVIR) e.g., Hirata et al. (2014), Modis Terra satellite hosted Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) e.g., Heiskanen (2006) and ALOS AVNIR-2 e.g., Sarker and Nichol (2011). Most high-resolution data studies depend on vegetation indices to estimate biomass. Other studies used image texture and classification as an indicator for biomass, with texture studies allowing all bands to be used rather than only (normally) two bands for vegetation indices. Nichol and Sarker (2011) concluded that combining the texture characteristics from two 10 m resolution optical sensors (Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2) and SPOT HRG) can result in a high correlation with biomass. However, Nichol and Sarker (2011) implemented an object-based analysis to generate training samples for the regression model; to deal with homogeneous forest. Most other studies found that texture indicators are weakly (Lu and Batistella, 2005, Li et al., 2008) to moderately (Sarker and Nichol, 2011) correlated with biomass. Generally, for this type of data, the relatively low spatial resolution is expected to limit the accuracy of AGB assessment.

Very-High Spatial Resolution Data: When studies require a high level of accuracy, very-high spatial resolution airborne and/or satellite sensors data can be utilized. However, such datasets require a high degree of processing and purchase costs can also be high. Hence such factors can curb the use of this data, even in small areas such as in Lu (2006) and Santi et al. (2014).

The very-high resolution satellite data was used as a source of spectral information such as in Coulibaly et al. (2008). It was used for object-extraction based approaches for example, this type of data was used for estimating tree crown size (Leboeuf et al., 2007). In addition, it was implemented as a source of detailed texture features such as in Persson (2016). Wordview-2 and Wordview-3 imagery provides eight spectral bands. However, most the other satellite sensors provide only four bands that are red, green, and blue in visible and one infrared band (e.g. IKONOS, Quick-Bird, and Pleiades). Even with only four bands, these datasets can perfectly satisfy the requirements of VIs. In addition, revisit time is around three days (Jacobsen, 2011) which is more sufficient compared to high spatial resolution sensors such as Sentinel-2A and Sentinel-2B that are together provides a revisit time of five days (ESA, 2016). The shorter revisit time can be important for AGB change monitoring such as in cases of forest fire. On the other hand, it increases the probability of providing low cloud coverage data, especially for areas that have frequent cloudy weather.

Studies such as Eckert (2012) and Migolet et al. (2007) tried to correlate the satellite data (WorldView-2 and IKONOS respectively) to the biomass by combining image classification, principal components, VIs and texture analyses. Although, the first Earth-Observing instrument Advanced Land Imager (EO-1 ALI) sensor has a spatial resolution of 30m, which is lower than IKONOS's spatial resolution of 4m, Thenkabail et al. (2004) shows that the EO-1 ALI sensor performed better (in specific cases) for biomass assessment. Airborne sensors offer a possible avenue for much denser spatial and temporal coverage. In addition, airborne imagery can provide very-high spatial and spectral resolution data alongside stereoscopic modelling. For example, Ge et al. (2007) used the Compact Airborne Spectrographic Imager 2 (CASI II) "hyper-spectral imager" (that provides up to 288 bands in the spectral range between 401 and 915 nm) for AGB assessment. The stereoscopic modelling is rarely used because it is very time consuming compared to other 3D alternative techniques like Lidar (Light Detection and Ranging) and Radar (Radio Detection and Ranging).

3.2.1.2 Active Remote Sensing

Active remote sensing involves transmitting Electro-Magnetic (EM) pulses from the sensing system itself instead of using a natural source of energy. For Radar, the pulse is at radio frequency whereas in Lidar the pulse is in optical wavelength (Müller and

Quenzel, 1985). Both of these systems can provide estimations of the elevation above the ground or vegetation/ manmade structures, and so the relative heights of plants canopies above the ground can be calculated. The plant height is closely related to the volume and therefore biomass (Edson and Wing, 2011; Rosette et al., 2012). In addition, very-high resolution data can be related to trunk diameter and canopy structure (Dalponte et al., 2009). Therefore, this type of remote sensing provides another perspective of AGB assessments that can be integrated with passive remote sensing. However, costs often make it prohibitive for wide area studies. The specification of Radar and Lidar data are dissimilar, so each has specific parameters that effect the AGB assessment accuracy.

Radar: Previous work has shown that Radar data is an efficient data type for biomass assessments. Multiband Radar is sensitive to the plant canopy because microwaves of longer wavelengths are able to reach the ground surface, by penetrating high-density vegetation cover, whereas shorter wavelengths are reflected by the vegetation surface (Becek, 2010). Radar can be used both during the day and night, and it can also penetrate cloud cover (Ghasemi et al., 2011; Becek, 2010).

Synthetic-Aperture Radar (SAR), the Radar technology that is based on the motion of the antenna to provide higher spatial resolution, has been applied to biomass assessment. The SAR techniques include polarimetric interferometric SAR (Mette et al., 2004; Persson and Fransson, 2016), multi-temporal InSAR observations with high temporal resolution (Tanase et al., 2014), tomographic SAR (Robinson et al., 2013; Minh et al., 2015), repeat-pass interferometric SAR (Khati et al., 2016), and using Tomo-SAR multi-baseline techniques (Minh et al., 2016; Li et al., 2016). Also, a number of different satellite missions have been used such as ALOS PALSAR (Lucas et al., 2010), and (Mohan et al., 2011), RADARSAT-2 (Shao and Zhang, 2016; Mohan et al., 2011), TerraSAR-X (Holopainen et al., 2010; Persson and Fransson, 2016) and COSMO-SkyMed (Deutscher et al., 2013). Generally, satellite based systems can provide a wider area and hence a more comprehensive data source compared to the available airborne sensors.

Studies show that study area characteristics have to be considered when choosing the most suitable Radar dataset. Some study area parameters relate to the target biomass such as structure, smoothness, mass distribution and orientation (Castel et al., 2002; Lucas et al., 2010; Montesano et al., 2014). Some are related to the topography of the land in the study area, such as ground smoothness slope and soil moisture content (Goering et al., 1995; Sun and Ranson, 2001; Soja et al., 2013). However, studies often pay most attention to the sensor parameters including wavelength, polarization, incidence angle, and spatial and temporal resolutions.

A crucial parameter for choosing between Radar datasets is the wavelength due to its direct impact on the Radar 's ability to penetrate vegetation (Imhoff et al., 2000). X band data (wavelength of 2.5-4.0 cm) has the least penetration power in forest areas, with the signal being scattered by leaves and canopy cover, so it primarily provides surface layer information (Ghasemi et al., 2011). The C-band (wavelength of 4-8 cm) has moderate penetration power i.e., can reach tree branches. The L band (wavelength of 15-30 cm) has the power to penetrate the tree canopy down to the trunk (Lau, 2011), and (Mercer et al., 2011), with Tanase et al. (2014) suggesting that L-band interacts with branches, leaves and needles. However, the same study concludes that even expected improvements in future L-band missions, in terms of spatial and temporal resolution, the accuracy of biomass estimation would not significantly improve. The P band (wavelength of 30-100 cm and not available with current satellite sensors) has the most penetration into the canopy cover and the major part of P band can be backscattered by trunk and ground. Therefore, the backscattering of the L and P band signal shows a relatively high correlation to tree biomass (Ghasemi et al., 2011). The VHF bands (wavelength of 1 m to 10 m, and not available with current satellite sensors either) is efficient to solve estimation problems related to the biomass saturation phenomenon (Israelsson et al., 1997), and (Imhoff et al., 2000).

Another system effective parameter is the polarization of the signal. The SAR polarization is either:

- Horizontal-Horizontal (HH): that has horizontal polarization for both transmitted and received signals.
- Vertical-Vertical (VV): that has vertical polarization for both transmitted and received signals.
- Horizontal-Vertical (HV): that has horizontal transmitted and vertical received signal.
- Vertical-Horizontal (VH): the emitted signal has vertical and the backscattered signal has horizontal polarization.

The first two types are usually called co-polarized backscatters and the other two are called the cross-polarized backscatters, and are the most sensitive to biomass (Holopainen et al., 2010), and (Tanase et al., 2010). Overall HV polarization is generally regarded as having the best potential to describe the AGB (Sun et al., 2002).

The other effective influence in choosing SAR data is the look angle of the system. The look angle relates to the incidence angle and to the reflectance of targets, the swath width and the perspective of the imagery. SAR systems used for biomass assessment can be divided into side looking and downward looking systems. Dubois-Fernandez et al. (2005) showed that the behaviour of the P band as a function of biomass is small change when the look angle varying from about 20° to about 40°. However, with an increasing look angle, the resolution would decrease and shadowing, and layover effects would increase. The shadowing occurs when the SAR system fails to illuminate parts of the terrain that hide behind other parts of the terrain from the perspective of the sensor due to the relatively low elevation of these hidden parts. While, the layover occurs when the SAR signal reaches a horizontally farther point of the terrain before it reaches a horizontally closer point due to the relatively high elevation of the farthest point that make the first one closer in the three-dimensional space. SAR layover and shadows affects one of the interferometric calculation stages, in which the topographic effect should be defined to allow individually acquired images to be co-registered to each other (Pairman and McNeill, 1997). Downward looking SAR systems overcome shadowing, and layover effects and can gather more information than side looking systems (Wencheng et al., 2010; Peng et al., 2012). Also they require less power to be operated (Imhoff et al., 2000) and so can be placed on smaller and more mobile platforms (Peng et al., 2012). Nonetheless, it still has a lower cross-track resolution and more vibration errors (Wencheng et al., 2010).

The spatial resolution of SAR data has improved with commercial satellite missions such as TerraSAR-X and COSMO SkyMed that are X-band and have a ground resolution of up to 1.5 m for TerraSAR-X and 1 m for COSMO SkyMed. Airborne SAR sensors such as OrbiSAR and IFSAR can enable surveys of very-high ground resolution of 60cm. However, the financial costs of purchasing, the data size, and the processing costs are relatively high for these very-high-resolution data types. Lower resolution space agency missions are more accessible for biomass applications, and so would be the option of choice for many future studies. For example, the Copernicus Sentinel-1 C-band SAR images have a maximum ground resolution of up to 5 m and are free to access.

Temporal resolution of SAR data is efficient because the data collection is not sensitive to cloud cover. Some satellite systems use more than one satellite to provide better temporal resolution. For example, at its full constellation of four satellites, COSMO SkyMed maximum revisit time reduced to 12 hours (Covello et al., 2010). Another example is the Sentinel-1 that has two satellites, on 12 days' orbit that result in a repeat frequency of 6 days and a revisit frequency of 3 days due to the overlap between adjacent images.

Lidar: As with other RS datasets, Lidar technology cannot directly provide three-dimensional data about AGB due to the complexity of the vegetation constructions. Instead, it provides accurate remote sensing measurements of terrain elevation and vegetation height (He et al., 2013). By processing vegetation heights, the spatial distribution and texture alongside other information such as stem counts, crown diameters (Bortolot and Wynne, 2005), and (Van Aardt, 2004), and stem diameter (Dalponte et al., 2011) can be derived.

Lidar data have their own characteristics in relation to the results uncertainty. Instead of spatial resolution, Lidar depends on the size and density of the Lidar footprint where the Lidar footprint is the illuminated area of the target by each laser pulse (Rosette et al., 2012). Footprint density is the number of footprints in a unit area. The other important parameter for Lidar data is the laser ranging systems, which could be phase-shift lasers or pulse laser. The phase-shift lasers send continuous waves with sinusoidal modulated optical power, and the reflected signal is detected and compared with the emitted signal to determine the difference in phase which can be used to find the time the signal takes to reach the target and return. Airborne and space-borne sensors use pulse lasers because it combines the high power output and the system power efficiency due to the short time required for each pulse. Pulsed laser systems can further be subdivided according to the recording mode into: double returns, multiple discrete returns and full waveform returns as shown in Figure 3-1. A dual-return Lidar (Figure 3-1A) records only two points for each pulse (usually first and last returns), and is widely used with commercial airborne Lidar systems. When the data are high density with a small footprint data, see examples in Table 3-1, the resulting height information about biomass is suited to areas of low tree density. A multiple discrete return system (Figure 3-1B) can provide more points, depending on the peaks at the power of the sampled returned signal. When this is operated with a small footprint, see Table 3-1, it is common technique for airborne Lidar surveys. However, this type of data is generally non-figurative with large footprint systems. Therefore, this form of the data has been rarely used for AGB assessments. In contrast, full waveform scanning systems (Figure 3-1C) record the whole returned signal. The examples in Table 3-1 have shown that even though the technology is still maturing, this type of system is widely used in both the small and large footprint size configurations. It showed ability to assess biomass with good accuracy for small areas.

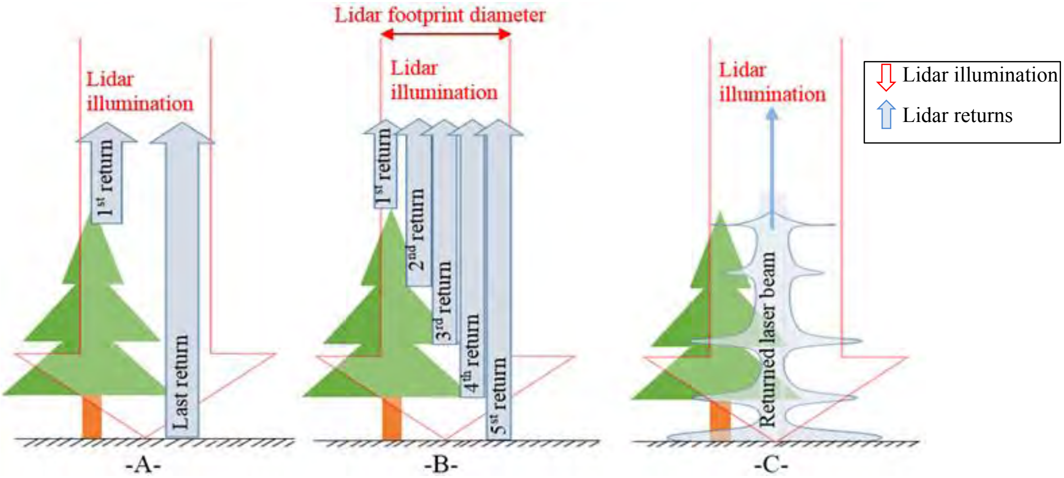


Figure 3-1: Lidar systems according to the recording mode, A- Dual-return Lidar system, B- Multiple discrete return Lidar system, and C- Full-waveform Lidar.

Table 3-1: Examples of studies that have used airborne or satellite Lidar data for AGB assessment.

Mode	Small footprint examples	Large footprint examples
Two discrete returns	(Lim and Treitz, 2004a) (Bortolot and Wynne, 2005) (He et al., 2013)	
Multiple discrete returns	(Dalponte et al., 2009) (Dalponte et al., 2011) (Vaglio et al., 2014)	(Means et al., 1999)
Full waveform scanning	(McGlinchy et al., 2014)	(Sun et al., 2007) (Baghdadi et al., 2014)

The increased ability of computer hardware can open opportunities for small footprint, full waveform scanning systems to be applied for wide areas. However, software to process such data is still limited (Rosette et al., 2012). The large footprint, full waveform scanning systems provides more ability for large scale studies. Examples show that this type is the most suitable in dense forests to ensure that laser energy reaches the ground. The focus on this type of data has increased in the period when Geoscience Laser Altimeter System (GLAS) instrument on The Ice, Cloud, and Land Elevation Satellite (ICESat) was providing it. This sensor stopped working since 2009 but a new generation of it (ICESat-2) is scheduled for launch in 2018. Terrestrial Laser Scanning (TLS) is also used in studies either for small scale studies with very-high accuracy (Yao et al., 2011), and (Dassot et al., 2010) or to provide accurate sample data for a wider scale study (Popescu and Hauglin, 2014). TLS is an accurate non-destructive estimation method used by the forestry industry, hence separate commercial software packages have been produced for this type of data. Edson and Wing (2011) tested different software packages; the results produced differed but were relative close to each other's estimations.

3.2.1.3 Combined Remote Sensing Data

Combination of more than one RS data set gives better results than using a single data set because the combination provides more information and explanatory power. Integrating the strengths of different sensors data sets can increase the confidence in forest biomass (Lu et al., 2012). Such combining can mean combining the influence of each RS data type specification on the accuracy. There are many examples of data combining systems. Attarchi and Gloaguen (2014) combined data from Landsat ETM+ and ALOS/ PALSAR data to provide topographic correction requirements for ETM+ images and to increase the number of multiple regression inputs. This combination achieved a significant enhancement in the general accuracy comparing with ETM+. However, the study does not compare results of using ALOS/POLSAR data only.

Holopainen et al. (2010) used a combination of ALS airborne Lidar and TerraSAR-X data to provide more inputs to the model. However, the general accuracy of combined data improved only slightly comparing with Lidar results. Similarly, Vaglio et al. (2014) studied the correlation between AGB and its estimation depending on laser alone and then depending on the laser coupled with hyper-spectral images which enhanced the correlation. However, the difference in the correlation indicators (and the accuracy) was slight. Lucas et al. (2008) combined Lidar and CASI hyperspectral data to achieve a relatively high accuracy. The study used a three steps procedure. First, trees were located using a CASI data. Second, tree heights were derived using Lidar. Then, height-based allometric equations were used to calculate the AGB.

3.2.2 Fieldwork data Related Parameters

3.2.2.1 Sampling Strategies

For AGB assessment fieldwork, single tree samples (which are equivalent to random individual samples) were not explicitly represented in literature due to the high cost and the spatial context required to correlate the single tree sample to the remote sensing data. Therefore, almost all studies used plots (which are equivalent to the cluster statistical samples). Those plots vary in sizes and shapes. For example, Ahmed et al. (2013) used 200 m by 50 m rectangular plots. Shao and Zhang (2016) used 30 m by 30 m square plots, Bortolot and Wynne (2005) used 30 m diameter circular plots. The plot sizes needed to be compatible with the resolution of RS data, e.g. Zheng et al. (2014) used circular plots with 300 m diameter with Landsat images, which is quite larger than plots of previous examples that work with finer spatial resolution data. However, some studies failed to achieve this condition for example Luther et al. (2006) use sample plots of 200

m^2 areas, and Zheng et al. (2014) use 667 m^2 plots. In both cases plots are smaller than the area of the pixel of Landsat TM data they used. The study by Saatchi et al. (2011) shows the importance of the size of the plot. The study provides a different accuracy result for each plot size. The choice of whole or part sample plots of some studies was to cut costs or solve logistic problems, i.e. there is no statistical logic behind choosing the sample. Similarly, some studies used an existing fieldwork data. For instance, Leboeuf et al. (2007) located his test sites near the towns whereas Yuan et al. (2016) used out-dated data observed by another study between 2000 and 2008. Other studies used statistical probability, where none of the vegetation cover types has a probability of zero to be sampled. The simplest strategy of this type is simple random clusters (plots). Dalponte et al. (2009), and Vierling et al. (2013) used simple Random Cluster samples. Gjertsen (2007), and Mitchard et al. (2013) used simple random cluster with sparse number of sample plots. This method of sampling strategy could be representative if the vegetation cover is homogeneous. However, random sampling can cause high uncertainties in heterogeneous areas if a significant vegetation cover type is not covered. The second probability strategy is systematic random sampling. In this strategy, geographical position of sample plots been chosen on a systematic pattern covers the whole area of the study. For example, Ghasemi et al. (2013) divided his study area into districts, parcels, and then plot inside each parcel. Another example is the use of Baghdadi et al. (2014), (Dalponte et al., 2011), and Heiskanen (2006) to plots of a systematic grid. This strategy could be more representative, but still can cause the same problem of the first one. The other strategy is stratified random clusters. This type can be the most statistically valid one, especially under a full-scale condition. It classified the vegetation cover into categories or classes. Then, samples could be randomly chosen for each class. The strategy has been widely employed for RS of AGB. e.g. Means et al. (1999), and Luther et al. (2006), used stratified random plots. Zheng et al. (2004) used two stages stratified random samples. The first stage results in supervised classification and the second stage results in plot sampling. Lu et al. (2002) used more complex Multi-stage probability stratified random strategy. Strata consisted of plots for trees and subplots for smaller plants. Tanase et al. (2010), and Edson and Wing (2011) used stratified clustered strategy plots that were based on the age of trees. Santos et al. (2003) used a stratified cluster sampling method where random position sample plots were chosen for each land cover class. The number of samples also varying from only two plots such as in Persson and Fransson (2016) to more than a hundred like in Dong et al. (2003) depending on the study's objective and scale.

3.2.2.2 Data Collection

The vast majority of studies use previously developed allometric equations, as it is rarely feasible to develop new ones (Ahmed et al., 2013). Studies of producing such equations depend on large scale destructive biomass sampling in order to relate the biomass of a tree to measurable parameters such as diameter at breast height (DBH) or tree age. The use of "species-specific" equations with different tree species is required due to the dependency of tree shape and wood density on the tree type (Ketterings et al., 2001). Smith and Gilbert (2003) list the tree species in the Great Britain including the study area. Almost all these species equations are listed in Zianis and Seura (2005) review book that compiles 83 equations for AGB for tree species growing in Europe. However, for other species general equations (e.g., Muukkonen 2007) can be applied. The other major concern is the time period between field data acquisition and RS data acquisition.

To study the two data sets synchronously, all biological measurements need to be processed using "tree and trait specific relative growth rates" (Lim and Treitz, 2004a). Those growth rates require sequential observations of the same sample trees for each tree category. If the data sets were taken in different seasons, the observation should be taken seasonally. Moreover, trees harvesting and deforestation should be taken in consideration. Therefore, visual comparison between fieldwork data and RS data is required. However, if synchronised RS and field data are available, they could be used for training the model and then this model could be applied on the old RS data.

3.3 Deriving Above-Ground Biomass

Extensive effort has been devoted to developing techniques for identification of the analysis techniques that can be used for deriving AGB from RS data. The techniques are used by previous studies can be divided into: parametric regression, nonparametric regression, and physical models. Those techniques can be subdivided as shown in Figure 3-2 which shows the types of regression that were used for AGB assessment on previous works.

3.3.1 Parametric Regression

Parametric regression is a statistical method that utilizes the most statistically valid functional relation between some independent variables ($X_1, X_2, \ldots, X_n$) continuous variables and a dependent variable or variables (Y) in order to predict an unobserved point in the dependent variable range (Neter et al., 1996). When the regression refers to the relation between two quantities only, one dependent variable (X_1) and one independent variable (Y), it is usually referred to as a bivariate regression. When it refers to the relation between more than two variables (X_1, …, X_n), it is called multiple regression. The selection of the parameters was highlighted by the literature as a vital part to achieve a parsimonious model. In other words, to reach the best-balanced model between using unnecessary independent variables that can cause an over-fitted model and ignoring necessary variables that can results in an under-fitted model. The applied techniques used for this purpose are stepwise regression, Analysis of variance (ANOVA), jack knife based regression, and Genetic Algorithm (GA).

The functional relation between variables is a mathematical equation is in the form of $Y=f(X_1,X_2,\ldots,X_n)$. This equation can be formed in three stages, the decision upon the equation type based on the complexity of the relationship between dependent and dependent variables, the calculation for the equation parameters based on the observed points in the dependent range, and the statistical testing of the resulting regression model. Many types of parametric regression between biomass and remote sensing products were used. For example, some studies used an exponential equation such as McRoberts et al. (2011), Schlund et al. (2015), Næsset et al. (2015) and Suresh et al. (2014) or a multiplicative equation such as Lim and Treitz (2004b). However, the majority of the literature used a form of polynomial equation. The polynomial equation was in the first-degree form (linear regression) such as Persson (2016), Mauya et al. (2015), Suresh et al. (2014) and Hansen et al. (2015), in the second-degree form (quadratic regression) such as Zaki et al. (2016) and Migolet et al. (2007), or in the third-degree form (cubic regression) such as Migolet et al. (2007). The determination of equation parameters (or the regression coefficients) is usually carried out by the method of least squares or the method of maximum likelihood (Neter et al., 1996). Both methods can ensure a minimum variance in the class of unbiased regression equation. However, for statistical reasons, non-linear regression models require a reasonably large number of samples to avoid overfitting. There is a number of statistical indicators to test the regression such as correlation coefficients (R^2 and Adjusted R^2) the F-test, root mean square errors, and the variance of the residuals. However, each one of the previous parametric regression studies yielded different values for regression parameters and regression accuracy. This disagreement between literature, even when similar data and models were used, affected the knowledge relating to the relationship between RS and AGB.

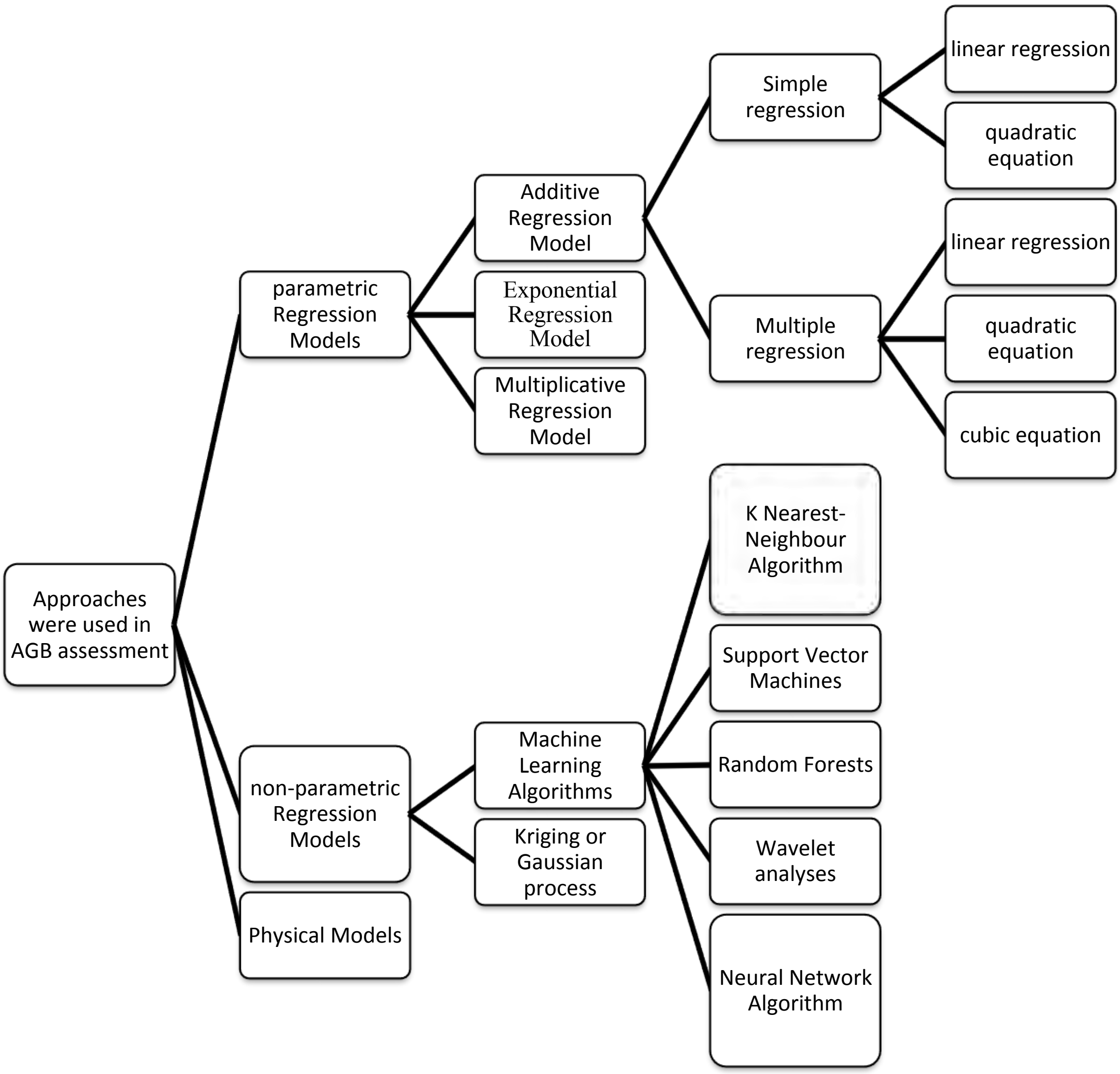


Figure 3-2: AGB deriving techniques found in the previous studies

3.3.2 Non-Parametric Regression

There are two types of non-parametric regression used for AGB assessment. These are ordinary kriging and machine learning methods. The kriging is one of the geo-statistical methods that tries to smooth the experimental variogram and produce a model variogram based on least squares. An example of this model could be found in Coulibaly et al. (2008). Machine learning approaches try to use field data and its corresponding remote sensing data to design a model and predict its parameters without previous equation form. Machine learning approaches may be grouped in several divisions: K-nearest-neighbour method (k-NN) was used by number of studies for classification (Gjertsen 2007) or for regression (Tomppo et al., 2002; Holopainen et al., 2010). Artificial Neural Network algorithm (ANN) was widely used like in Fraser and Li (2002), Zheng et al. (2014), Migolet et al. (2007), and Del Frate and Solimini (2004). Other types were also used with recent studies like Genetic Algorithm (GA) (Tuominen and Haapanen, 2013), Random Forests (Stelmaszczuk et al., 2015), Support Vector Machines (SVM) (Chen and Hay, 2011), wavelet analyses (Ghasemi et al. (2013), fuzzy logic algorithm (Walker et al., 2010) and Maximum Entropy modelling (MaxEnt) (Stelmaszczuk et al., 2015).

Englhart et al. (2012) study compared between parametric regression and nonparametric regression for AGB. It reported a higher accuracy with nonparametric regression models. However, non-parametric regression algorithms, require a larger database to be trained compared to the parametric regression models (Englhart et al. 2012). This could be a drawback when fieldwork coverage is limited. Tanase et al. (2014) combined both nonparametric regression and parametric regressions to estimate biomass depending on multi-temporal Radar and multi-temporal Lidar data. Ghasemi et al. (2013) use wavelet analyses in biomass assessment stage and ANOVA in the accuracy assessment stage. Walker et al. (2010) suggested a forest mapping with satellite Radar data using a fuzzy logic algorithm. Englhart et al. (2012) compare between SAR based AGB model with multiple linear regression, ANN and SVM; this study indicated that multiple linear regression has the lowest performance and highest error measures to retrieve AGB. The SVM model provided less saturation for higher AGB and showed a superior performance for modelling AGB. While, the ANN performance was the most accurate for low density biomass. However, other types of multiple regressions, which could perform better with nonlinear relationship, were not included in the comparison. Some models such as McRoberts et al. (2015) and McRoberts et al. (2011) combined parametric and non-parametric regression in search of optimization.

3.3.3 Physical Models

Despite the clear advantages of regression, there is a need to improve remotely sensed AGB estimates models if they are to provide accurate information on AGB. Each time a parametric regression is applied, new parameters are produced. The produced parameters for a specific study are not applicable for most other studies. The reason is that RS data are not direct measurements of the AGB (Bollandsas, et al., 2018). Moreover, non-parametric regression produces a model with an unknown equation. For example, if models employ ANN, the processing includes hidden layers where undefined processes apply on data. Therefore, transferable biomass assessment equations are essential. Some studies, such as Pouliot et al. (2005), Daliakopoulos et al. (2009), and Ferraz et al. (2016), tried to solve this problem by finding a physical model for AGB assessment based on remote sensing data. The simplest physical model is the tree count. Tree count depends on object-based classification and object detection (e.g., Daliakopoulos et al., 2009). Tree count requires very-high resolution RS data and represents a very laconic indicator of AGB in areas of mixed tree ages and/or tree species cover. Therefore its combination with tree crown delineation can give more relative information because it provides the crown diameter (Pouliot et al., 2005). Methods use spatial information extraction techniques and a productivity model. In addition, some three-dimension (Radar and Lidar) data have the power to provide an indicator of the heights. Mette et al. (2004) studied the allometric relation between trees Radar based height and forest field observed biomass. Lucas et al. (2008) and Ferraz et al. (2016) stated a procedure for extracting tree height, branches distribution, crown delineation and tree species from Lidar and hyperspectral images. However, the results of those physical models were correlated with AGB by means of regression too.

3.4 Methods and Materials

The methodology used includes training a computer software tool to emulate human decision making for the optimal trade-off between cost and accuracy for forest tree biomass surveys. It considers a variety of parameters, and their interactions, to create a flexible non-linear model; an Artificial Neural Network (ANN) fit is applied.

The input training data were based on the outputs from a number of previous biomass surveys; shown in Appendix A. The relative error (ratio of root mean square error, RMSE, to the approached value) of each sample is used as the target input. The inputs for each sample are the type of the fieldwork allometric equation that was used to derive fieldwork data (general equation or species specific equation), the processing model, allometric related properties (area, number of species and average biomass per km^2), and RS data properties. The reviewed literatures listed a number of effective characteristics for each dataset. However, to achieve a workable analysis, only the effective parameters that are reported in at least 50 samples in the literature were used.

As discussed in section 3.1, each RS data type has different specifications. Therefore, samples were divided into four groups: 100 Radar inputs, 51 Lidar, 60 optical and 50 combined RS datasets as listed in Appendix A. Only most effective characteristics were taken in consideration as listed in Table 3-2. Optical data are classified into panchromatic, multispectral or hyperspectral bands and characterised with spatial resolution. Lidar data are classified into single return and multiple return waveforms and characterised with footprint size. Radar data are defined according to wavelength, polarization, resolution, and incidence angle. Combined samples include studies that combine two of Radar, Lidar and optical data. This group is verified depending only on the data type and spatial resolution. The reason for verifying combined data, with less detailed inputs, is that the number of available studies is limited and so insufficient for a more detailed analysis. More studies can be added to the AGB assessment, increasing the size of the available samples yearly, and so more detailed accuracy production tools could be produced in the future.

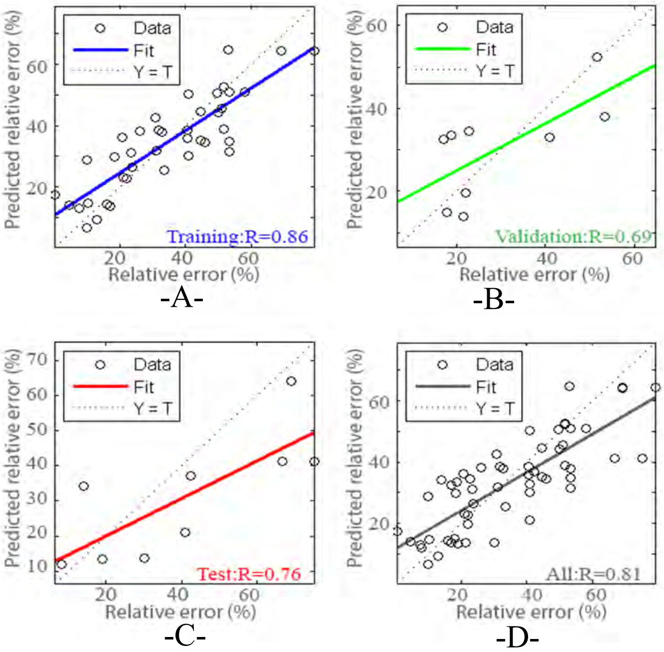
Due to the small number of the datasets available, neural network architecture design methods as described in Karsoliya (2012) resulted in over fitting. The alternative design, based on trial and error of many smaller networks, was therefore used. The best result network has one hidden layer of 10 neurons and one output layer of one neuron. In each group 60% of the data were used for training and 20% to find the epoch of best weights, with 20% used to test the trained network. This sample split was set as an alternative to the software default split that uses 70%, 15% and 15% for training, validation, and testing correspondingly. The aim was to provide more samples for testing and validation due to the low number of available samples for each ANN. Two tests were applied depending on the regression between the predicted relative errors and the relative error in the original training, validation and testing data. The tests include analysing the correlation coefficients.

3.5 Results

To apply the methodology, a separate neural network has been applied to each dataset. The effective parameters that were selected to train the ANN systems are listed in Table 3-2 alongside the characteristics of each ANN system. The correlations between neural networks results and actual studies accuracy results are shown in Figure 3-3 to Figure 3-6. The performance of neural network applied on optical data studies shows a Root Mean Square Error (RMSE) of ±9.2% in the test samples that is not included in the train process. The percentage value refers to the predicted absolute error as a percentage of the AGB. The correlation results compared to the true accuracy values is shown in Figure 3-3 with a correlation coefficient r of 0.76. Comparably, the performance of neural network applied on Lidar data studies shown in Figure 3-4 has an RMSE of ±6.5%. The correlation coefficient for this neural network is r=0.82. Similarly, Radar studies network has an RMSE of ±15.9%. The correlation coefficient of this neural network results with reality is r=0.62 (Figure 3-5). Jointly, combined data studies network has an RMSE of ±13.7% and r=0.60 (Figure 3-6).

Table 3-2: Summery of the effective parameters and number of samples for each dataset's ANN compared to the RMSE of the validation for each one.

Dataset	RS specifications used	Fieldwork specifications	Model specifications	No. of available samples	RMSE of validation samples %
Optical	Spatial resolution, and number of bands	study area coverage, and land cover heterogeneity (number of species), and field allometric equation type	Model type (parametric or nonparametric)	64	±9.2%
Lidar	Lidar type, and footprint size			62	±6.5%
Radar	Number of bands, polarization/s, resolution, and incidence angle			100	±15.9%
Combined	RS data types, and spatial resolutions			50	±13.7%



-A- -B-

-C- -D-

Figure 3-3: The regression between neural network results and corresponding study results for optical data: A-for training samples, B-for Validation samples, C- for the test samples and D- for all samples together.

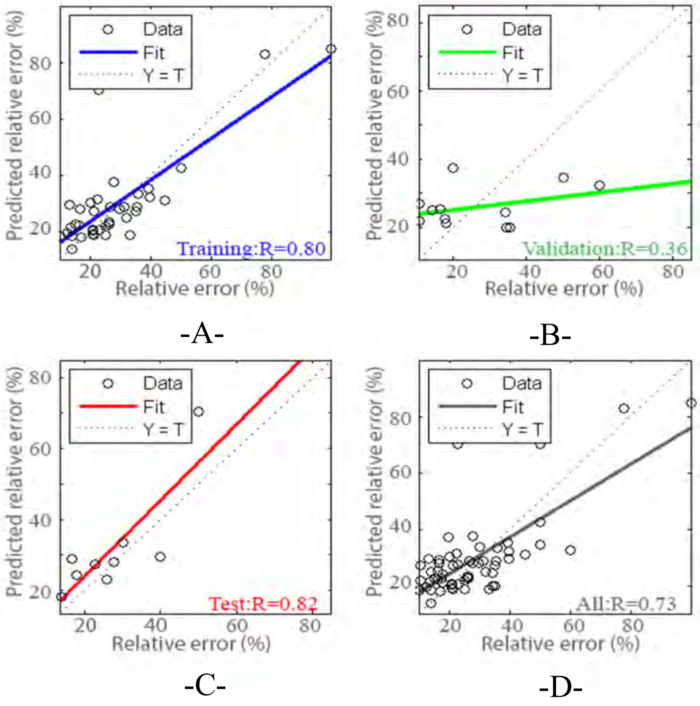


Figure 3-4: The regression between neural network results and corresponding study results for Lidar data: A-for training samples, B-for Validation samples, C- for the test samples and D- for all samples together.

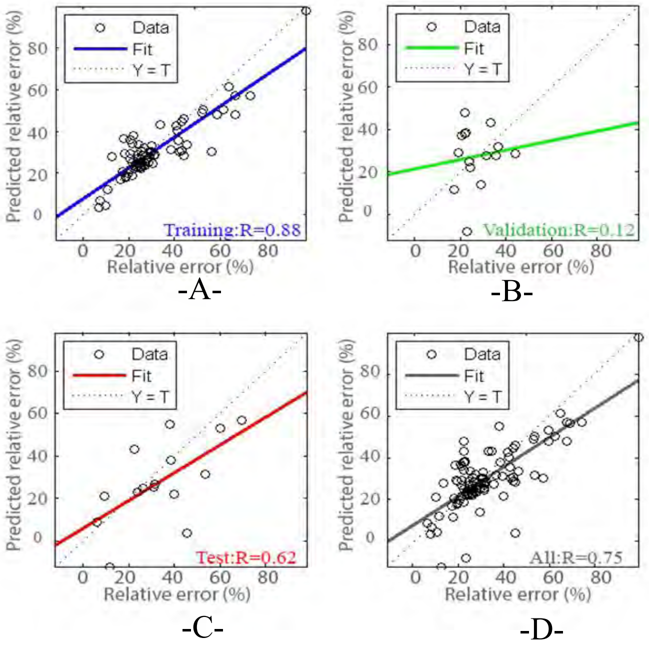
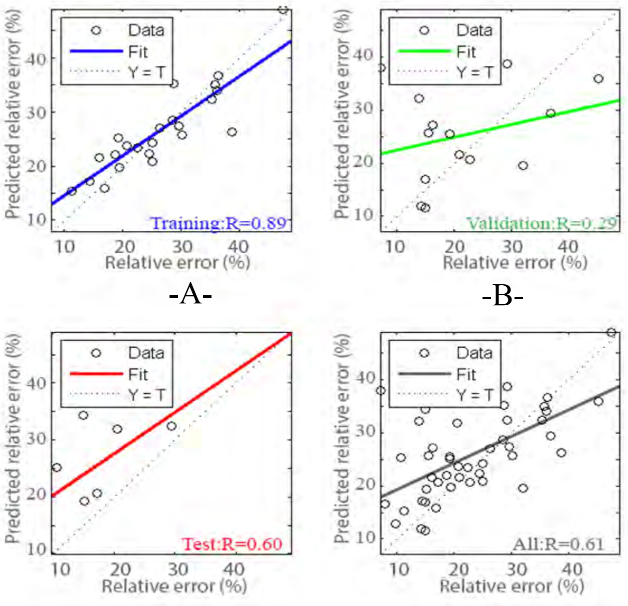


Figure 3-5: Neural network results for Radar data studies as A-for training samples, B-for Validation samples, C- for the test samples and D- for all samples together.



-A- -B-

-C- -D-

Figure 3-6:Neural network results for combined datasets, as A-for training samples, B-for Validation samples, C- for the test samples and D- for all samples together.

3.6 Summary

In this investigation, an aim was to find a methodology to predict the uncertainty of AGB assessment during the design phase. A second aim, for this chapter, was to investigate the variety of the possible AGB assessment models and inputs with error prediction based on the results of previous studies.

The studies reviewed in this experiment confirmed that a wide range of parameters can affect the uncertainty within AGB projects. The parameters include RS data types and each type's specifications, the fieldwork allometric equation type, the assessment’s coverage, the vegetation cover variety in the study area, alongside the mathematical model type and specifications. The execution of the methodology showed moderate results due to the limited number of publications compared to the required samples for sufficient training. However, the reviewed possibilities in this chapter served as a base for the design of an inclusive error analysis tool in described in Chapter 5.

Chapter 4 Fieldwork Data Collection and Processing

4.1 Introduction

Although the accuracy of the field based data is assumed to be high enough to allow the ground information data to be implicit as error free truth, it is still being questioned. With allometric equations, there is a number of error sources that can effect this accuracy such as the errors expected in the direct measurements of DBH and tree heights, the error of allometric equation as a regression equation, the error of geographic location of the trees, and the inadequate number of samples. Relative positioning and temporal errors result in the biggest share of the fieldwork AGB uncertainty (Gonçalves et al., 2017). Therefore, In recent years, there has been a growing application of Terrestrial Laser Scanning (TLS) (Liu et al., 2017) as a precise fieldwork technology to support remote sensing (RS) data, with the highly detailed field data collected within a short time giving a smaller temporal variation and higher precision to the biomass estimations. In turn, this means more information being available.

Practically, even with this detailed data, plot based sampling is still typically used to model the relationship between AGB and RS data (Næsset et al., 2015). The pixel-based approaches are not efficient due to the indirect relationship between AGB and RS data, even with very-high resolution data. Yet, high precision biomass distribution has an importance when AGB is correlated with high resolution RS data, especially when small subplots are used and plot boundaries probably intersect with some of the trees. Small subplots can provide a bigger number of samples than is required to train, calibrate and validate any AGB assessment system and so help to avoid overfitting (Fassnacht et al., 2014). In addition, the small subplots can reduce loss of detail due to averaging the biomass within each plot. Therefore, single-tree-level forest inventory is expected to play a key role in near future biomass mapping systems (Kankare et al., 2017).

The huge size of the TLS data's computer files, the wide coverage, and the complex tree structure make the processing and analysis methods to automatically retrieve the whole tree structure computationally expensive. For instance, Côté et al. (2011) stated that using their method on a single long Douglas fir tree, which has a simple structure compared with most other species, can take more than 30 hours to be processed depending on the required processing. In another example, Calders et al. (2015) stated that a sample of Eucalyptus tree took more than 18 hours to be reconstructed. The processing also requires a high density point cloud, which means more field work and larger computer files. To avoid these problems, a number of studies verified the significance of using the TLS to extract only the parameters that can be indirectly correlated with AGB (inventory parameters) like tree coordinates, tree height (h) and tree diameter at breast height (DBH). The DBH is usually measured at height of between 1.30 (e.g. see Alberti et al. 2005, Muukkonen 2007, and Forrester et al. 2017), and 1.50 m (e.g. see Pérez-Salicrup, and Barker 2000, and Evans et al. 2014). Despite these methods providing an accurate biomass assessment, they could not provide a precise biomass distribution, i.e. they provide the biomass of each tree as a point while the tree biomass is heterogeneously distributed over a few square meters. Therefore, a fast method, that combines the high precession AGB distribution provided by analysing the TLS with the high accuracy provided by inventory parameters, can overcome the problem by providing an AGB result that is fast and reduces the uncertainty of correlation with very-high resolution RS data.

Two techniques have been used for this purpose. The first technique involves converting the TLS point clouds into a volumetric pixel three-dimensional model (voxels), e.g.

Hackenberg et al. (2014), with each 3D pixel having a digital number representing the number of TLS points located inside it. The voxel reduces the data size and the processing time as a result. The second one uses one of the 3D tree modelling approaches (Pratihast 2010; Raumonen et al. 2013; Liu et al. 2017) that depends on existing knowledge of the tree shape, for instance each tree branch is expected to have a smaller diameter than its parent branch. When this knowledge is generalized, it can be used to fill the gaps in the TLS data, and predict initial values and limits of fitting parameters.

4.2 Materials and Methods

4.2.1 Study Area

The study area (Figure 4-1) is located in the south-west of the United Kingdom between Latitude 50° 15' N and 50° 90' N and between longitudes 3° 40' W and 4° 40'. This coverage was chosen due to the land cover heterogeneity that is equivalent to the variety achieved with larger scales areas; the scanned plots represent a variety of tree types, tree species and ground terrain. The area covers the river Tamar catchment and Dartmoor National Park.

The Tamar catchment includes more than 70 woodlands and forests with various tree cover densities, sizes and types distributed in rural areas of wide agricultural fields. It includes a number of towns as well. There are some water cover areas such as river Tamar, river Tavy, and Burrator reservoir. In the contrast, Dartmoor National Park has areas of tree cover surrounded by areas of moorland.

Figure has been removed due to Copyright restrictions.
Please see
Forestry_Commission (2002a) *National Inventory of Woodland and Trees- Cornwall*, United Kingdom: Forestry_Commission. Available at: https://www.forestry.gov.uk/pdf/cornwall.pdf/$FILE/cornwall.pdf. and
Forestry_Commission (2002b) *National Inventory of Woodland and Trees- Devon*, United Kingdom: Forestry_Commission. Available at: https://www.forestry.gov.uk/pdf/devon.pdf/$FILE/devon.pdf.

Figure 4-1: Tree types map and fieldwork sites in the study area.(Forestry_Commission, 2002a; Forestry_Commission, 2002b)

The main tree species in this area includes sitka spruce *(Picea sitchensis)*, larch *(Larix occidentalis)*, oak *(Quercus)*, beech *(Fagus Sylvatz'ca),* birch *(Betula),*sycamore *(Acer pseudoplatanus)*, ash *(Fraxinus excelsior)*, and elm *(Ulmus campestris)* (Forestry-Commission, 2001). These species can be categorised according to the geographical distribution with in the study area to broadleaf trees areas, conifer trees areas, and mixed trees areas, beside the non-standard tree areas like young trees, felled trees, under planting areas, and coppice trees. The distribution of these categories is shown in Figure 4-1.

Therefore, a number of sample plots were decided based on study area resonance in order to increase the representability of the field data. The selected sites for fieldwork sample plots are listed in Table 4-1.

Table 4-1: A list of the selected fieldwork sites.

Site No.	Location (ϕ,λ)	Category
1	50˚30’ 55”N, 04˚ 09’ 50”W	Broadleaf trees
2	50˚34’ 40”N, 03˚ 53’ 55”W	Conifer trees
3	50˚30’ 15”N, 04˚ 02’ 15”W	Mixed trees
4	50˚27’ 30”N, 04˚ 15’ 20”W	Mixed trees
5	50˚27’ 50”N, 04˚ 14’ 50”W	Broadleaf trees
6	50˚34’ 40”N, 03˚ 54’ 00”W	Conifer trees

4.2.2 Laser Scanning

Leica ScanStation P20 was used to scan 38 sample plots (divided on the six sites listed in Table 4-1) of about 2500 m^2 each. This scanner is a dual band laser scanner (808 nm and 658nm) with a 3D angular accuracy of ±8 second, 3D Position Accuracy of 3 mm at 50 m distance targets, and a linearity error of less than 1 mm. The scanning took place in 2014 and 2015. The scanned plots represent a variety of tree types, tree sizes and ground terrain patterns. The scanner can provide a scanning range of up to 120 m. however, the effective range of laser scanning was analysed by using a single scan in a high density forest area shown in Figure 4-2.

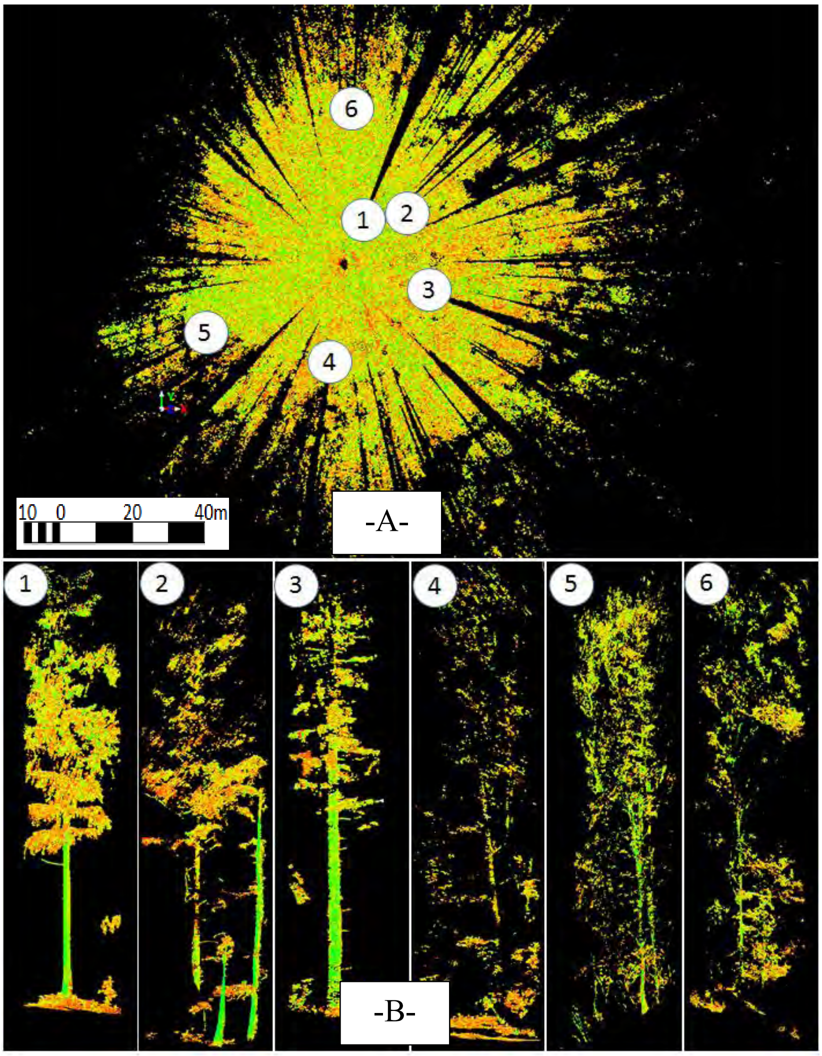


Figure 4-2: A single station scan in a relatively high density forest area. A- selected trees' location on a top view of the scan. B- the details of each tree the distances between trees and the scanner were 1- 10 m, 2- 22 m, 3- 27 m 4- 33 m, 5- 39 m and 6- 45 m correspondingly.

Six example trees were selected on different distances from the scanner on 10 m, 22 m, 27 m 33 m, 39 m and 45 m. The trees on distances of less than 30 m were presented with fine details for all above-ground tree parts. The tree scans for trees of a distance greater than 30 m were affected by laser shadows of other trees in the way between the scanner and the objected tree. These shadows on the point cloud could include tree parts that are hidden behind other objects that are closer to the scanner. The available solution is by scanning the trees from more than one station. The basic scanning plan is shown in Figure 4-3.

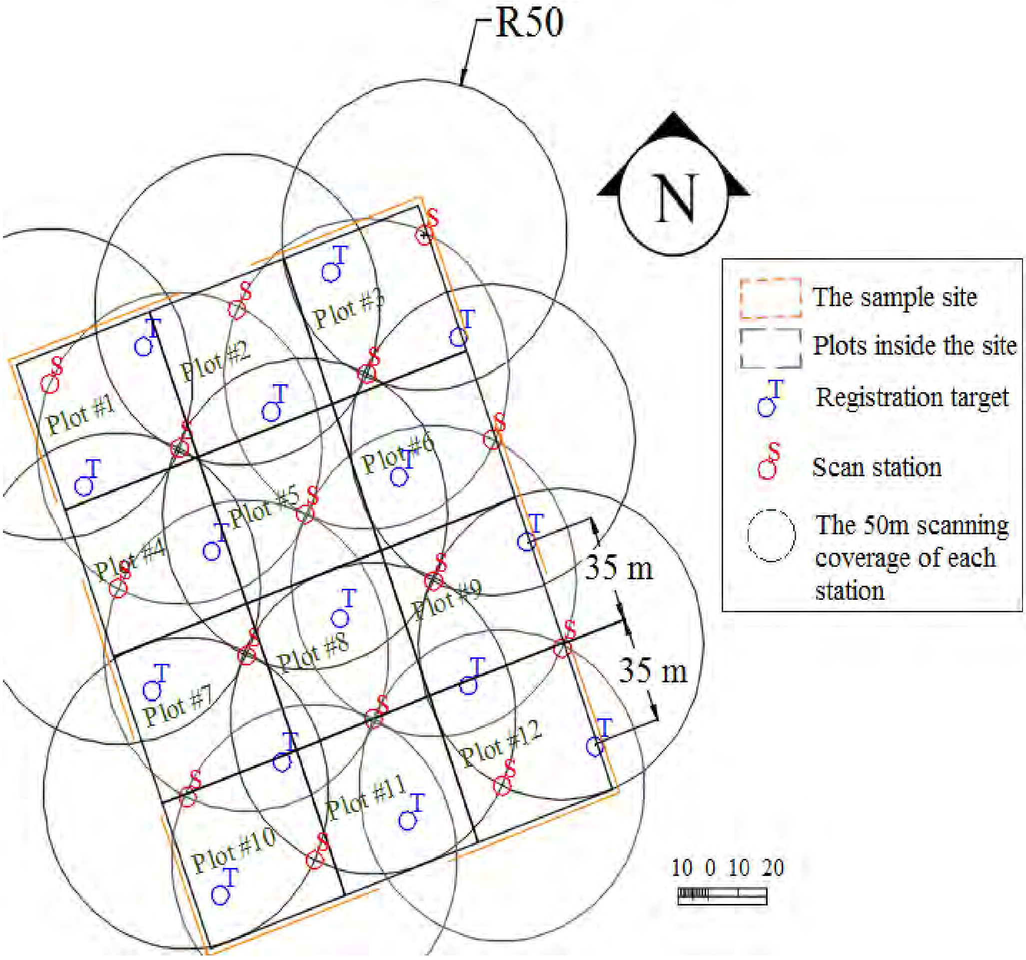


Figure 4-3: The basic scanning plan for one of the field sites.

This plan is designed to cover each tree from three to four stations with a distance of less than 50 m. The scan stations were set in rows of 35 m from each other. Each scan station provides a point cloud with a local coordinate system. Therefore, a set of connection targets were used to co-register the resultant point clouds from the scan stations to a unified coordinate system. Leica 6 inch black/white target that are compatible with the Laser scanner was used for this purpose. The scan stations are designed to be separated from the registration targets by a distance of about 35 m. However, in the field it was almost impossible to apply the exact plan because the existence of the trees that may either lie in the planned positions of the stations or targets, or be close to it and will result in a large shadow on other trees behind them. The scan stations were located to be as close to the plan as possible, but within workable positions.

4.2.3 Pre-processing Laser Scanning Data

The pre-processing includes using “Leica Cyclone 3D point cloud processing software” to co-register all the scan data to produce combined point clouds. The same software was then used to trim the point cloud to the object area boundaries, and to export the results to a standard ASCII file of the X, Y, and Z coordinates of the points.

4.2.4 Laser Scanning Data Processing

The AGB deriving process was included in the main steps in the flowchart shown in Figure 4-4. After the pre-processing, the following stages of processing were programmed using Python[1]:

- Step 1: Modelling Terrains and Excluding Ground Points.

A new technique was used to achieve a workable filtering, with minimum processing requirements to reconstruct the ground surface by processing only the set of points that lies inside far-apart parallel profiles as shown in Figure 4-5. The Profile Points (PP) are defined in terms of the minimum X coordinate of the plot (X_{min}), the distance between successive profiles (D) and the depth of the profile (d) as in Equation 4-1:

$$P \in P_{TLS}\colon \left\| \frac{X_{point} - X_{min}}{D} \right\| + \frac{d}{2} > \frac{X_{point} - X_{min}}{D} > \left\| \frac{X_{point} - X_{min}}{D} \right\| - \frac{d}{2}, \qquad (4.1)$$

where P_{TLS} is the whole TLS points data. For each plot, D was set depending on the terrain complexity and d was set depending on the point cloud intensity. An example of the filtering result of one of the fieldwork scanning areas is shown in Figure 4-5.

---

[1] The programming code was uploaded to be freely available on https://github.com/AhmedAlboabidallah/field-data-analysis

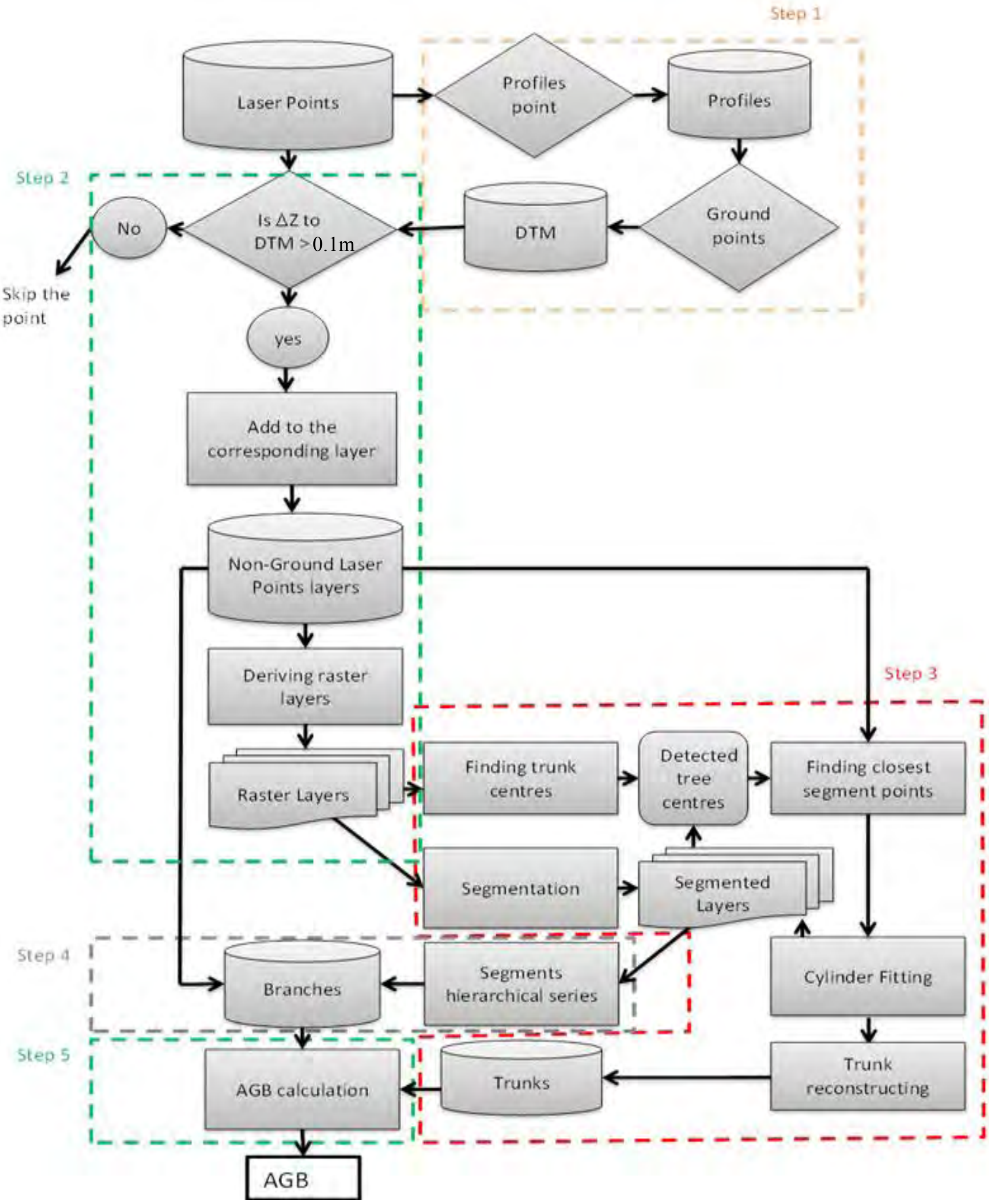


Figure 4-4: The flow chart of data processing and AGB mapping.

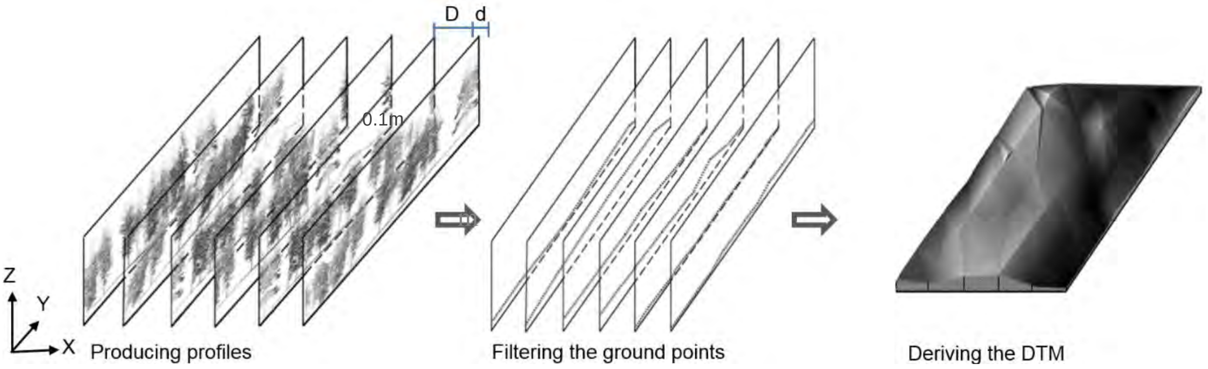


Figure 4-5: The processing steps to produce the DTM.

Then a simple minimum filter was applied to keep only the minimum elevation point in each 0.5 m x 0.5 m area, by applying the equation:

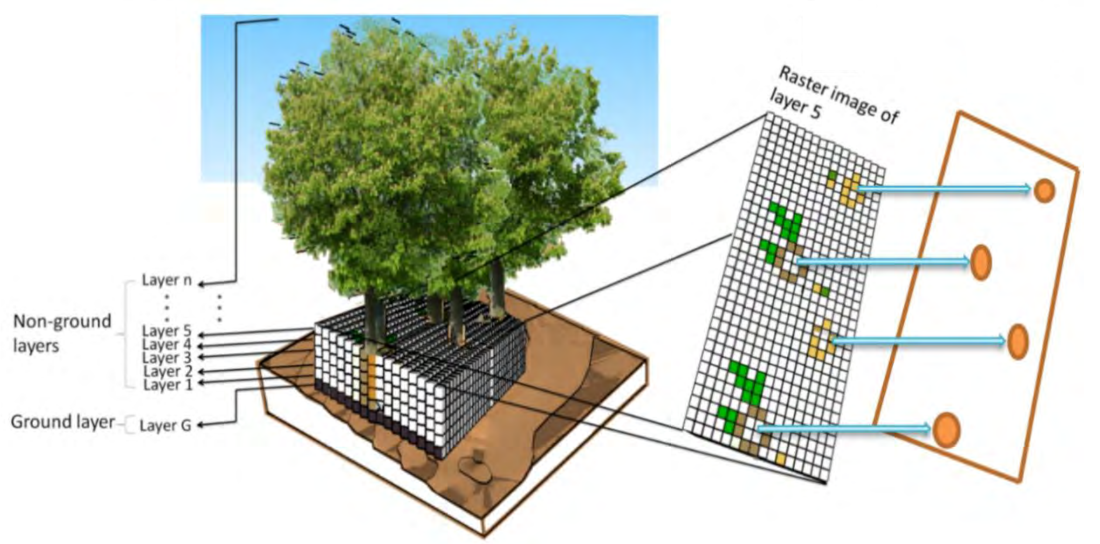
$$PG \in PP: PG = \min(C), \text{ where } C \in PP: \| \frac{X_{point} - X_{min}}{0.5} \| + 0.25 > \frac{X_{point} - X_{min}}{0.5} > \| \frac{X_{point} - X_{min}}{0.5} \| - 0.25$$

$$\wedge \| \frac{Y_{point} - Y_{min}}{0.5} \| + 0.25 > \frac{Y_{point} - Y_{min}}{0.5} > \| \frac{(Y_{point} - Y_{min})}{0.5} \| - 0.25 \quad (4.2)$$

In the next step, the DTM was built from points that passed the filter. The final step was filtering out all P_{TLS} points using a 0.2 m buffer around the DTM to exclude ground points.

- Step 2: Producing the 3D voxel

A new technique was applied to produce the voxel. Voxel layers were produced to be non-flat surfaces that are parallel to the DTM as shown in Figure 4-6 instead of being horizontal planes. The idea behind this process is to define the height of each point from the ground, allowing the trunk detection procedures to be applied on each slice evenly. Also, a small file of each voxel pixel's original TLS points was produced and linked to that pixel. This process aimed to arrange the TLS data and make it easy to be recalled for the next steps. The voxel-model is a 3d raster representation of the TLS point cloud. Therefore, the coverage of the voxel model was similar to the coverage of the TLS point cloud. The size of the 3d voxel pixel represents the resolution of the voxel. The finer the 3d voxel pixel is, the higher the accuracy is. However, the finest the 3d voxel pixel is the greater the processing computational costs is. Therefore, the pixel size was decided to 0.05 m to provide the minimum spatial resolution that can fulfil the requirements of the tree structure reconstruction steps.



-A- -B- -C- .

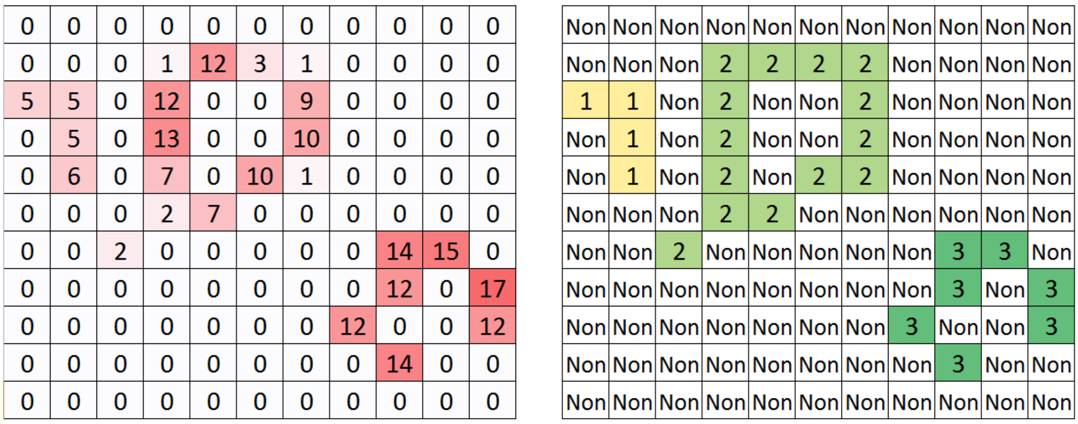
Figure 4-6: Slicing Lidar data and producing raster images. A. the concept of building a voxel by non-flat layers, B. The features of a single layer, and C. The extracted features in the layer.

- Step 3: Trunks Detection and Modelling

A new technique was applied to detect trees trunks. Trunks detection designed to work on four layers around the breast height (elevation ≈1.5m) by:

- Intersecting every two successive layers to emphasise the upstanding tree parts, like expected trunks, and to reduce other tree parts, like branches. This step is designed to reduce the effect of branches and leaves at the breast height. The intersection was used for trunk detection (by intersecting layers 1.3 m with 1.4 m and 1.5 m with 1.6 m) and for trunk tracing (by using all other layers).

Spatial segmentation of each intersection result by analysing non-zero pixel neighbourhoods' connectivity with any other non-zero pixels and giving any connected group pixel an Identification number (ID). For example, the voxel layer subset shown in Figure 4-7 has three groups of non-zero pixels each group was given an ID that is identical for the group and unique for the other groups. This step is important for image analysis and interpretation for object extraction because each segment is expected to be strongly related to the trees' trunks and branches. This segmentation transforms the voxel layer into high-level image description in terms of each extracted object and its corresponding original points in the point cloud.



0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	12	3	1	0	0	0	0
5	5	0	12	0	0	9	0	0	0	0
0	5	0	13	0	0	10	0	0	0	0
0	6	0	7	0	10	1	0	0	0	0
0	0	0	2	7	0	0	0	0	0	0
0	0	2	0	0	0	0	0	14	15	0
0	0	0	0	0	0	0	0	12	0	17
0	0	0	0	0	0	0	12	0	0	12
0	0	0	0	0	0	0	0	14	0	0
0	0	0	0	0	0	0	0	0	0	0

-A-

Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
Non	Non	Non	2	2	2	2	Non	Non	Non	Non
1	1	Non	2	Non	Non	2	Non	Non	Non	Non
Non	1	Non	2	Non	Non	2	Non	Non	Non	Non
Non	1	Non	2	Non	2	2	Non	Non	Non	Non
Non	Non	Non	2	2	Non	Non	Non	Non	Non	Non
Non	Non	2	Non	Non	Non	Non	Non	3	3	Non
Non	Non	Non	Non	Non	Non	Non	Non	3	Non	3
Non	Non	Non	Non	Non	Non	Non	3	Non	Non	3
Non	Non	Non	Non	Non	Non	Non	Non	3	Non	Non
Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non

-B-

Figure 4-7: An example of the spatial segmentation process. A. The voxel layer subset. B. The segmentation result of the subset.

- Applying the Hough Circle Transform (HCT) to the intersection results. Because the trunk sections are not perfect circles, the tolerance parameters are set to be highly tolerant. This helped to reduce the false negative results. However, this requires a next step to reduce the false positive results.

- Checking each detected circle, whether it represents a trunk section or not. This step includes visual classification at the current stage. However, all the results are kept in a separate database to be used later as a training data for a machine learning based automatic classification to true positives and false positives. This step helped to avoid false positive results.

- The centre coordinates and DBH of each tree was computed from the original TLS points of the closest segments that cover a minimum required angle around the centre detected by HCT or all segments that lies within a threshold distance which unify and combine the detected centres of processed layers.

Three fitting approaches were compared with each other in terms of accuracy and computational cost. These methods include fitting a circular cylinder with free axis orientation, an elliptical cylinder with free axis orientation and an elliptical cylinder with Z direction axis orientation.

The cylinder fitting procedure employed a non-linear least-squares fitting of cylinder that is similar to Wei and Wang (2009) fitting. The Wei and Wang (2009) algorithm includes finding the initial axis orientation angles with the three coordinate axes, and fitting a cylinder by iterating non-linear least square method. The applied algorithm makes use of the expected physical model of the trunk to reduce the cost of the cylinder fitting by setting the initial orientation to be parallel to the Z axis.

Every two successive layers were intersected and tree trunks were traced in layers of higher and lower than breast height, starting from each centre. The tracing process based on tracing the closest pixel in the next layer to the centre in the previous layer (the lower layer for heights greater than 1.3 m and the upper layer for the other layers). This type of tracing layers was designed to reduce the cost of repeating the procedure of cylinder fitting for each layer. This way the detected centre will be on the trunk surface rather than the centre which will cause a certain amount of error but it will reduce the errors at the layer where large branches intersect with trunk and the shifts tracing process to tend toward branches and get lost. If the next layer centre horizontal distance is more than the threshold distance the trunk centre the tracing process passes to the next layer to avoid the problem of the missing trunk parts in the shadows of intense branches expected at the upper part of the tree. The tree heights were determined by the highest elevation non-zero pixel in each trunk zone. To avoid possible noise points, pixels with less than three points were consider as zero-pixels for tree height determination.

- Step 4: Biomass Calculation and Mapping

This step is overlapped with trunk biomass as it requires the DBH and H to calculate the biomass, and the trunk biomass modelling requires the biomass value. Biomass were basically computed by using an allometric equation for each tree type for three reasons. The biomass calculation depending on the exact fitting of the tree is expensive in terms of the required TLS point-cloud quality and data processing. The second reason is that fitting can yield in computing the tree above-ground volume rather than its AGB and will required a destructive test to find the biomass density to convert it to AGB. The third reason is that volume-based AGB calculations involves an additional error component that is related to the wood density due to the necessity of converting volume into mass.

- Step 5: Branch Detection and Modelling

In all voxel layers, all pixels that are located with a distance of three times the radius of the trunk at that layer were set to zero. The remaining pixels were considered as branches.

To reconstruct the branches, point spatial segmentation was applied. The segmentation process included segmented pixels of each layer and segmenting the pixels' segments of all layers through their parent-child relationships. For example, layer i in Figure 4-8 is the parent layer in the branch. Because segment 1 has no parents in the lower layer (i-1), it is the first segment to be added to the set that represents the branch. Segment 1 in layer i+1 (the child layer of layer i) has a neighbourhood with a segment in the series (which is segment 1 of layer i). Therefore, this segment was added to the series too. A reverse check should be taken for other parents of segment1, layer i+1. If there is any, all parents' series should have been combined in one series. For instance, the example in Figure 4-8 segment1, layer i+5 has been added by segment1, layer i+4 and it has another parent (segment2, layer i+4) to be added to the series. The child layer is the parent of next layer and the same procedure continues until it reaches layer i+n which has no children for set segments of the previous layer. This way, the segmentation process is extended to include segments hierarchical series in between slices depending on the standard parent-child spatial. Each reconstructed branches classified to be a branch of the tree that has the closest trunk centre to it. The TLS points of each segment were sorted by their distances from the centre.

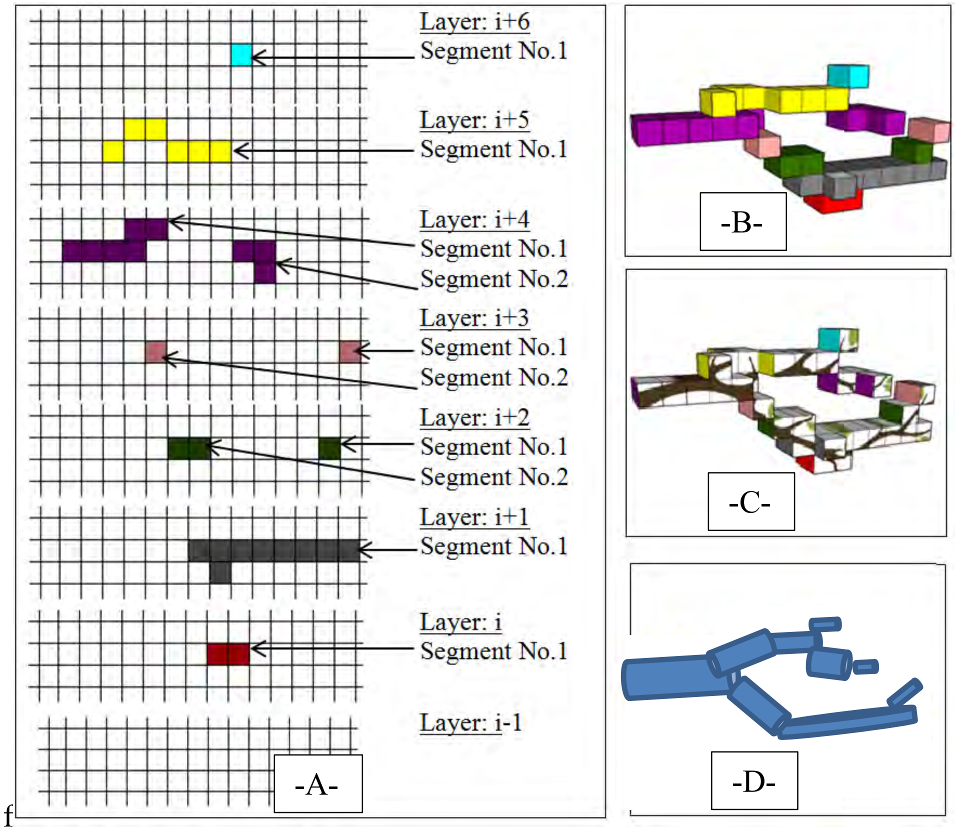


Figure 4-8: Steps of reconstructing a tree branch by using segmented raster layers. A- Finding relationships between segments of successive layers, B- Reconstructing segments, C- Cylinder fitting by Lidar points, and D- Reconstructed branch

To increase the information details and due to the fact that specific RS data types such as optical data can be better correlated to the tree branches unlike other data types such as Radar that can be more correlated to trunks. Branch maps and trunk maps were produced separately. This strategy benefits from the availability of detailed species based allometric equations for most tree species. The tree species were mapped in the field. Most scanned plots were of one tree type. Biomass of other plots were calculated by giving each try type a specific ID and call a different allometric equation accordingly.

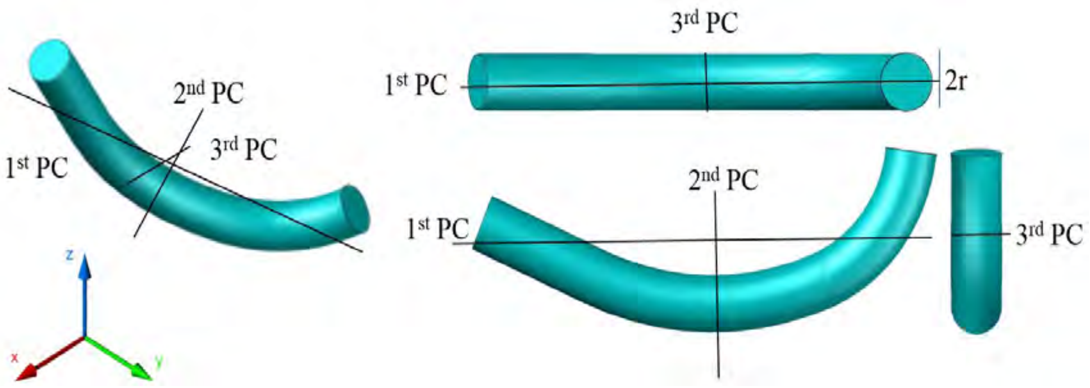
Branches were dealt with as cone and statistical weights of their biomass for each layer were calculated as the volume of the cone intersection with the layer. The calculated biomass for each layer was the statistical weight of the branch of the layer divided by the summation of the weights of the branch for all layers. Then the trunk biomass inside the layer were added as a circle of pixels. The centre of this circle is the traced centre of the trunk and the radius of it is the calculated radius of the trunk at the layer.

Branches statistical weighting is more complex and could be computationally expensive. Therefore, three techniques were tested to decide on the weight of each branch series including:

1- By fitting cylinder to the corresponding point cloud that lies on a specific distance from the centre and the branch length to build a cone. This procedure based on the fact that branches are most likely to have cylindrical shape at the parts that are close to the trunk before it takes more complex shapes that are combined by more leaves at the far parts from the trunk. The initial values setting used for trunk fitting is not expected to work for branches because it can be of any orientation. Due to the fact that the Gauss Image method used in (Wei and Wang, 2009) uses a small number of random points, the different numbers of points will affect the precision of initial vector. Therefore, this stage was replaced with simple Principal Component (PC) transformation. The analysis scale was designed to be limited to the points inside the extracted branch in the voxel. A limitation of PC is that it is sensitive to the variance due to measurement noise (Bailey, 2012). However, in this application of PC, the branch points are selected using spatial segmentation and therefore high percentage of the points are real branch points. Moreover, the branch parts close to the trunk often have less leaves than other branch parts.

The other limitation of the PC is the problem of missing data that can be cause by scanning shadows (Bailey, 2012). This problem can affect the accuracy of the cylinder fitting as well. One solution to this problem is by increasing the length of the objected part of the branch. On the other hand, this can increase the tree leaves noise and the probability of the branch to have a bow shape rather than straight shape.

2- Applying a Principal Component analysis to the same points in the cylindrical fitting in the previous method and use the length to build a cone. This analysis projects points to a new coordinate system that allows one of the projections to be vertical to the main axis of the branch. The example in Figure 4-9 shows a PC analysis of a branch and its relationship to the radius. In this example, even when the branch is not straight, the residuals of the points around the second PC axis is expected to correlate to the diameter of the branch. The range of branch points on the third PC also represented the third PC.



-A- -B-

Figure 4-9: The PC analysis of the closest part of the branch to the trunk. A. the branch part. B. the same part with respect to its PCs

3- By skeletonising the branch

The term skeletonisation refers to the thinning process of non-zero parts in the image to thin lines of one-pixel width based on the image morphology. To increase the efficiency of this method, the thinning processes were designed to avoid the 3d analysis. Alternatively, the 2d projections of the branch were used. For example, branch shown in Figure 4-8 can be projected into the top and side views shown in Figure 4-10. The branch statistical weight is calculated as the resultant of the length of the skeleton in the top-view and the vertical range of the branch. This way calculations can avoid the calculations of the branch orientation and increase the calculations speed. The scikit-image (Van der Walt et al., 2014) algorithm were used to calculate the voxels for the statistical weighting process.

These techniques were compared with each other and against manually traced branches in one of the scanning areas. PCA is used to provide initial values of each cylinder orientation and radius for the fitting process. The cylinder fitting is found to be computationally expensive. Instead the width of the branch in the second principal component of the first 30 cm of the branch was tested to be the weight.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

-A-

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

-B-

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

-C-

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	1	1	1	0	1	0	0	0	1	1	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

-D-

Figure 4-10: The skeletonisation process of the branch shown in Figure 4-8. A- the top-view of the branch's voxel, B- the skeleton of A. C. the side-view of the branch's voxel and D- the skeleton of C

The practical problem of applying this procedure is that some branches lay in a low point density of the point-cloud. The other drawback is the misleading results for branch series that include multiple tangled branches because it will count as a single branch. It is also computationally expensive because it requires using the original point cloud points that is corresponding to the first 30 cm of the branch series. Therefore, the alternative solution needs to be based on the voxel only. The branch length logically relates to the branch volume. The relationship between the length of the branch and its volume is complex and it requires a wider scale analyses to find a general equation. However, this was chosen as the approach to geographically distribute the biomass of the branches as it is a workable, relatively fast and precise technique. The other tested solution was by using the volume of the branch series in the voxel model. This technique is directly related to the branch volume. However, in the real samples the voxel volume was highly influenced by the amount of leaves on the branch.

The weights of the volumetric voxel pixels in each branch were computed by distributing the branch weight on the pixels. The pixels were given different weights based on cone volume that is parallel to the branch and has the same length, as illustrated in Figure 4-11. The cone base is at the closest part of the branch to the trunk and the top is at the farthest part from the trunk. The weight of each pixel is the volume of the corresponding part of the cone divided by the summation of volumes of all pixels along the branch.

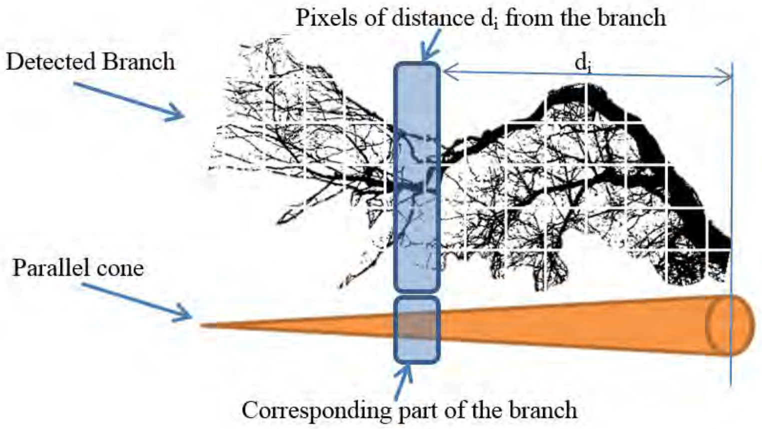


Figure 4-11: The branch's pixels statistical weighting process.

4.3 Results

The programming code was written with Python to apply the designed algorithm (Figure 4-4). The results include:

1-Modelling Terrains and Excluding Ground Points

The filter in Equation 4-1 was applied for each plot's TLS data, and the two sample results shown in Figure 4-12 represent the profile results in two different areas. The distance between each successive profiles (D) is set to 2 m for low slope terrain areas (e.g. scanning area shown in Figure 4-12a) and to 1 m for the high slope areas (e.g. scanning area shown in Figure 4-12b), with profile slice width (d) is set to 0.25m. The produced profiles are processed with filter in Equation 4-2.

The whole point cloud below the elevation of +20 cm were dealt with as ground points and were excluded. This process did not affect the tree heights because the tree height was calculated from the DTM rather than the point cloud lowest point. However, it was noticed that the elevation at the tree trunk was generally underestimated. This occurred due to the use of minimum elevation filter. The terrain at the tree trunk is generally expected to have a higher elevation than surrounding areas due to the contribution of tree root systems to slope stability and soil erosion control (Reubens, et al., 2007). The point cloud's part that is filtered with the DTM represents 20-40 % of the whole data size for the different scanned area.

4.3.1 Producing the 3D voxel

After the ground points have been removed, the remaining points in the point cloud were sliced. A corresponding raster layer of point density is produced for each voxel layer. The voxel spatial resolution was set to 0.05 by 0.05 by 0.05 m. This means that the minimum distance between tree parts has to be at least 0.05 m to be distinguished. The rational was to insure that branches that have a diameter of 5 cm can be detected. The 3D visualization of the branch is not required for following next steps. Therefore, the processing cost to plot the 3D visual representation, produced to visualize the voxel's quality, was not counted within the efficiency testing. An example of the 3D representation of a voxel is shown in Figure 4-13 that shows the voxel quality for one of the sample plots.

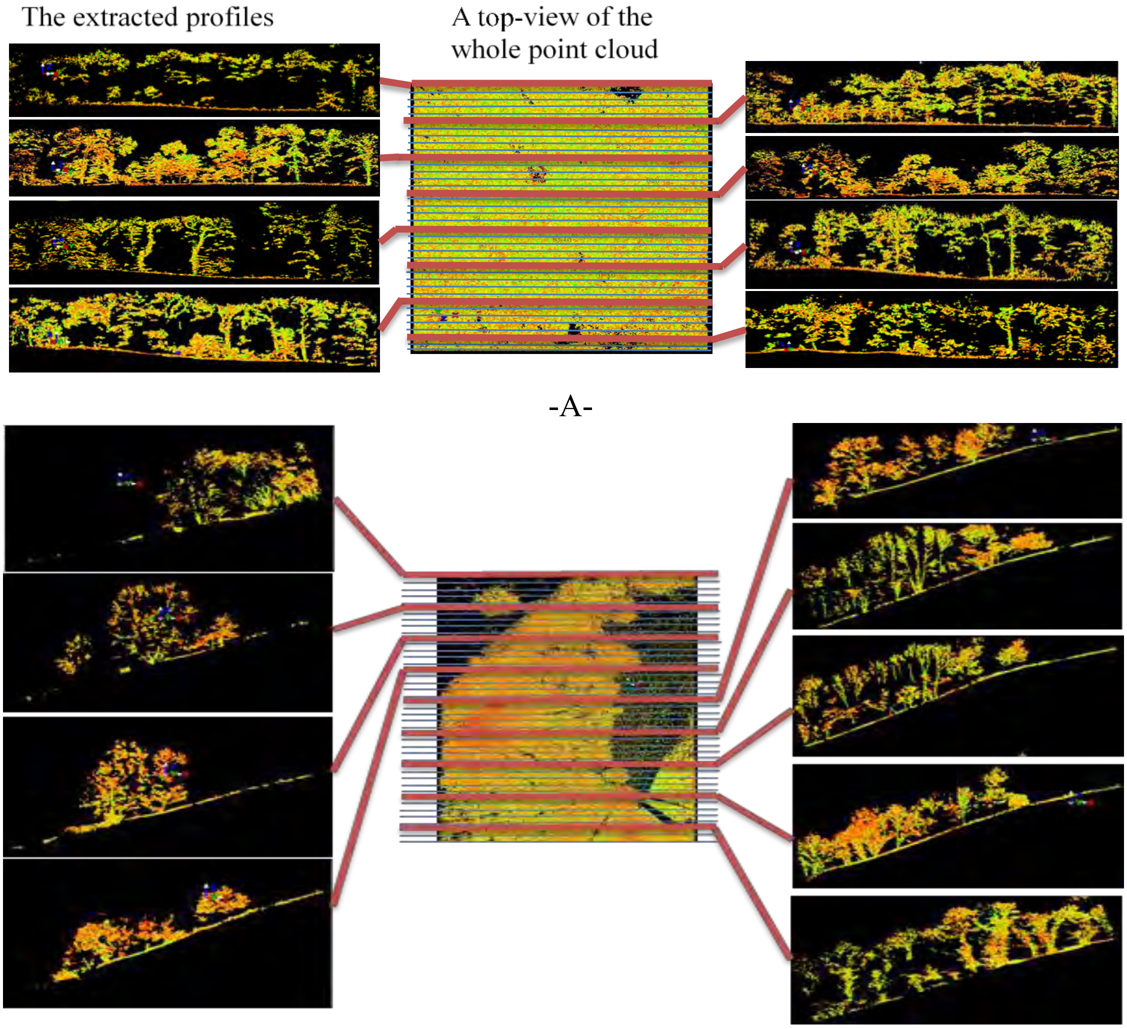


Figure 4-12: The profiles extraction results. A. sample profiles in a low slope area, B. sample profiles in a high slope area.

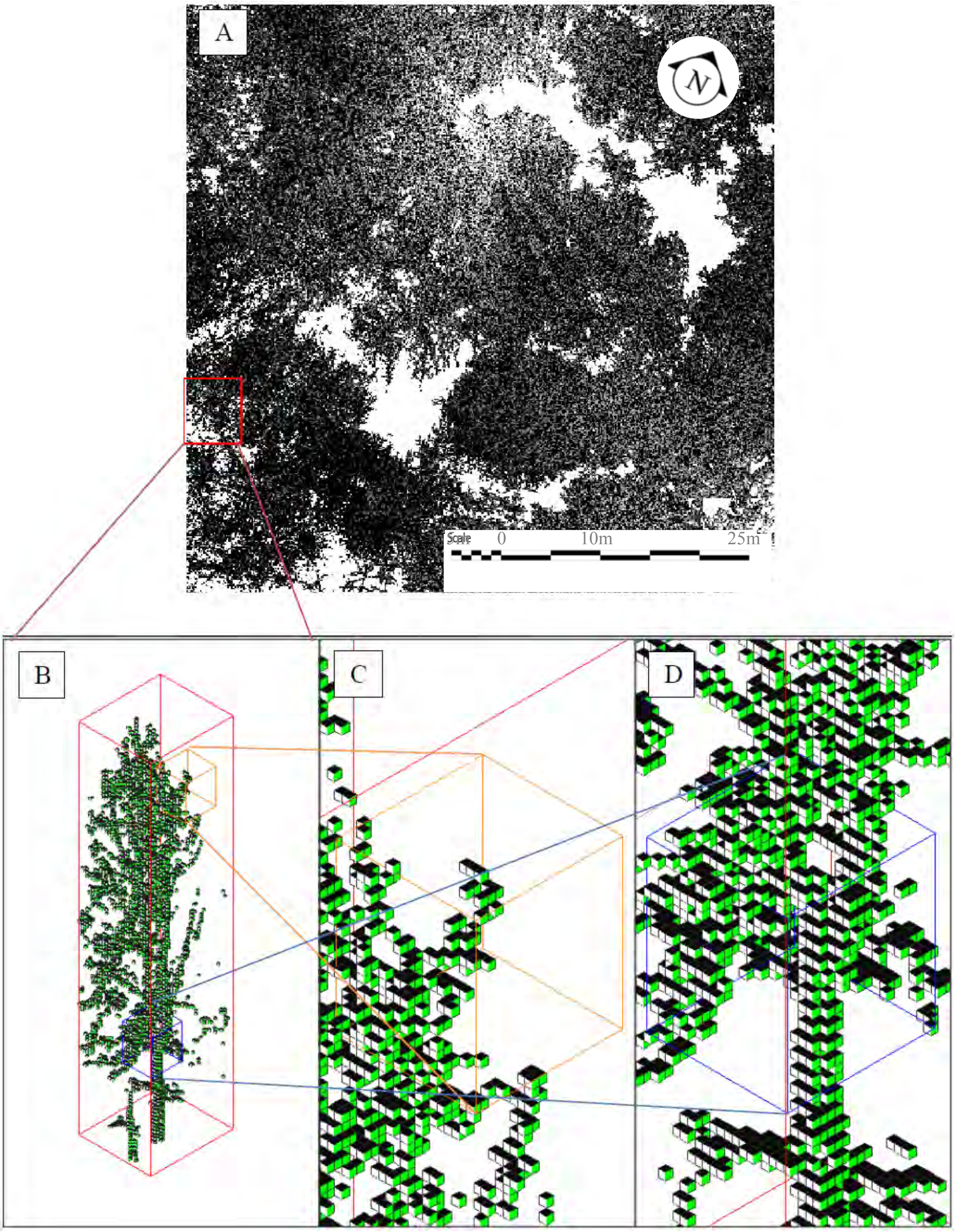


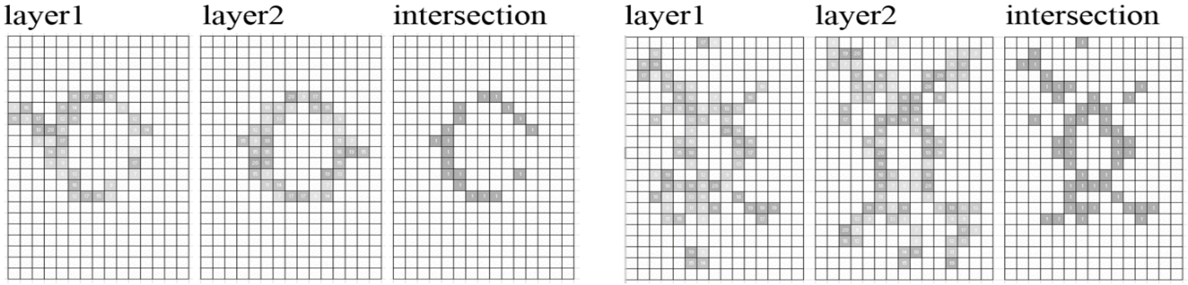
Figure 4-13: The voxel quality. A. The top view of the voxel. B. The isometric view of the voxel part inside the red colour box in A. C. The voxel quality in a branch part (the orange box part in B). D. The voxel quality in a trunk part (the blue box part in B).

4.3.2 Above-Ground Biomass Calculations

A Number of allometric equations were selected for the AGB calculations based on the scanned tree species. The selected equations needed to be geographically compatible with the study area to increase the geographic reliability of the equation. In addition, due to the objectives of this study, these equations were required to provide a sufficient error analysis. Hence, the species specific allometric equations of Alberti et al. (2005), Muukkonen (2007) and (Forrester et al., 2017) were used for biomass calculations. The calculated AGB was used to fulfil the requirements of step 4 (Trunks Detection and Modelling) and step 5 (Branch Detection and Modelling) as described in section 4.1.4.

4.3.3 Trunk Detection and Modelling

Review of the layer intersection results show an improvement in the tree trunks isolation: the majority of the tree branches attached to the trunks were separated or reduced. However, for a number of trees it does not fully exclude the branches attached to the trunk. Figure 4-14 shows an examples of a successful and a failed branch exclusion. The branch sections were either a close curve, open curve, or disconnected parts of a curve. The variety in the shape is either due to the laser shadows or due to the layer's intersection process.



-A- -B-

Figure 4-14: The effect of layer intersection on the trunk section. A- an example of a successful branches exclusion. B- an example of failed branch exclusion.

The trunk detection algorithm is applied on the raster layers at to the breast height by using OpenCV programming library (Culjak et al., 2012) under Python environment. The parameters required for trunk section recognition include the minimum distance between detected centres, the strength of the edge detection filter, the threshold score for centre detection, the minimum radius to be detected, and the maximum radius to be detected.

Beside the detected true centres (true positives) there is a possibility to detect some non-trunk features as trunks (false positives) and a probability of some true trunks that the detection algorithm fails to detect (false negatives). In order to set the trunks detection parameters to get the minimum false negatives for each plot, the minimum distance between detected centres were set to zero, the edge detection strength was set to maximum, threshold for centre detection was set to zero, minimum radius to be detected set to 5cm, and maximum radius was set to 1.5m. The visual classification of detected trunks of each plot shows that up to 45% of the detected trunks are false positives. However, the visual inspection stage reduced the ratio of the false positives to 1%. Table 4-2 shows the results of the trunk detection process before and after the visual inspection.

Table 4-2: The numbers of true positive of the trunk detection process for number of plots.

Area (m^2)	Trees	True positives		False negative trunks		False positive Trunks	
		Before inspection	After inspection	Before inspection	After inspection	Before inspection	After inspection
1600	Broadleaved	10	10	0	0	4	0
1225	Broadleaved	12	12	0	0	3	0
2500	Broadleaved	41	40	2	2	10	2
2500	Broadleaved	27	26	1	1	8	1
2500	Needle-leaved	28	28	0	0	4	0
2500	Needle-leaved	37	35	1	1	6	0
2500	Broadleaved	44	43	1	1	9	2
2500	Needle-leaved	20	20	0	0	2	1
2135	Needle-leaved	29	27	0	0	4	0
3519	Mixed	35	34	1	1	9	1
1968	Mixed	26	26	1	1	6	1
2500	Needle-leaved	17	17	0	0	4	0

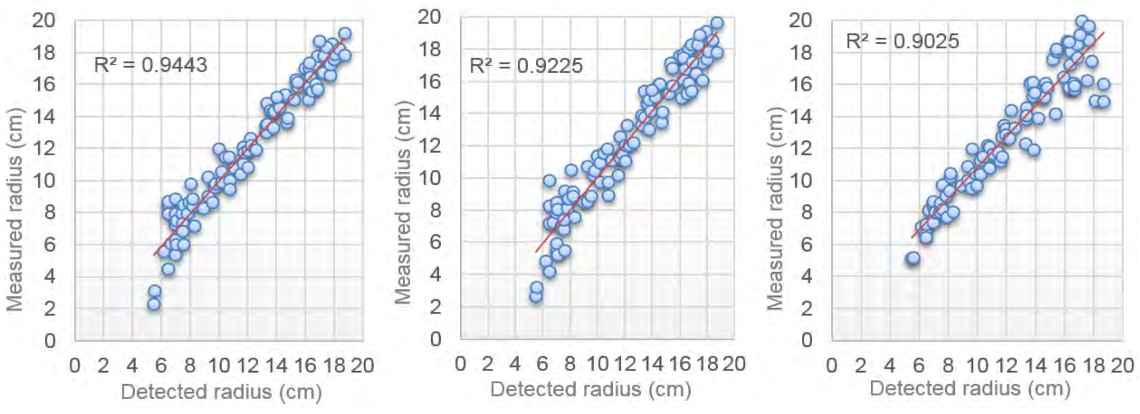
To compare the results of fitting techniques, correlations between trunk-diameter results from each fitting technique and manual-measurement was undertaken as shown in Figure 4-15, the best result was achieved with the elliptical cylinder with free axis orientation with R^2=0.94 and RMS=0.99 cm. Fitting an elliptical cylinder with Z direction axis orientation yields in an R^2 of 0.92 and an RMS of 1.20 cm. While, fitting a circular cylinder with free axis orientation shows a slightly lower correlation with R^2=0.90 and RMS=1.28 cm. However, the process of fitting an elliptical cylinder with a free axis orientation was computationally expensive compared to the process of fitting an elliptical cylinder with restricted Z direction axis orientation. This is due to the fact that the first one consists of fitting a circular cylinder, re-projecting the points to a plane that is normal to the central axis of the cylinder and then fitting an ellipse to find the radii, while the second one consists of the direct ellipse fitting only. The average time cost for free axis orientation method was 2.6 s compared with 0.2 s for the Z direction axis and 2.3 s for circular fitting with free axis orientation. Moreover, both free axis orientation algorithms have failed to converge four times out of the 97 tree used for the comparison while free axis orientation provide results for all trees. Therefore, the Z direction axis cylinder fitting was implemented.

The procedure described in Section 4.1.4, step five was applied based on the allometric equations listed in Table 4-3 The AGB map shown in Figure 4-16 represents the trunk biomass map of an example plot. These equations were mainly chosen because they have well-defined accuracy statistics that was used later in Section 6.5.3. Each pixel represents the summation of AGB of the voxel pixels located in the corresponding vertical column of the voxel. Hence, the spatial resolution of the resulted map is equal to the spatial resolution of the voxel which is 0.05 m.

The DBH was fitted around 1.5 m from the DTM that was derived in step1, section 4.2.4. The rational of choosing this height was to compensate the underestimation of the terrain elevation at tree trunk.

Table 4-3: Allometric equations implemented for AGB calculations for the fieldwork data analysis

Species		Equation	Reference
Ash	Stem	$AGB = 0.71 * 0.17 * DBH^{2.46}$	(Alberti et al., 2005; Cai et al., 2013)
	Branch	$AGB = 0.16 * 0.17 * DBH^{2.46}$	
Birch	Stem	$AGB = e^{(-2.411 + 10.21\frac{DBH}{DBH+8.291})}$	(Muukkonen, 2007)
	Branch	$AGB = e^{(-3.579+0.570\frac{DBH}{DBH+11.363})}$	
Scots pine	Stem	$AGB = e^{(-1.408 + 10.666\frac{DBH}{DBH+15.775})}$	(Muukkonen, 2007)
	Branch	$AGB = e^{(-0.928+9.889\frac{DBH}{DBH+32.338})}$	
Beech	Stem	$AGB = 0.159\ DBH^{2.346}$	(Chakraborty et al., 2016)
	Branch	$AGB = 0.233 DBH^{1.781}$	
Oak	Stem	$AGB = e^{-2.181} DBH^{2.269}$	(Forrester et al., 2017)
	Branch	$AGB = e^{-2.986} DBH^{2.309}$	
Mixed Species	Stem	$AGB = e^{-2.527} DBH^{2.414}$	(Forrester et al., 2017)
	Branch	$AGB = e^{-3.723} DBH^{2.33}$	



-A- -B- -C-

Figure 4-15: The correlation results for trunk radius algorithms. A. for elliptical cylinder with free axis orientation, B. for elliptical cylinder with Z direction axis orientation. C. for circular cylinder with free axis orientation.

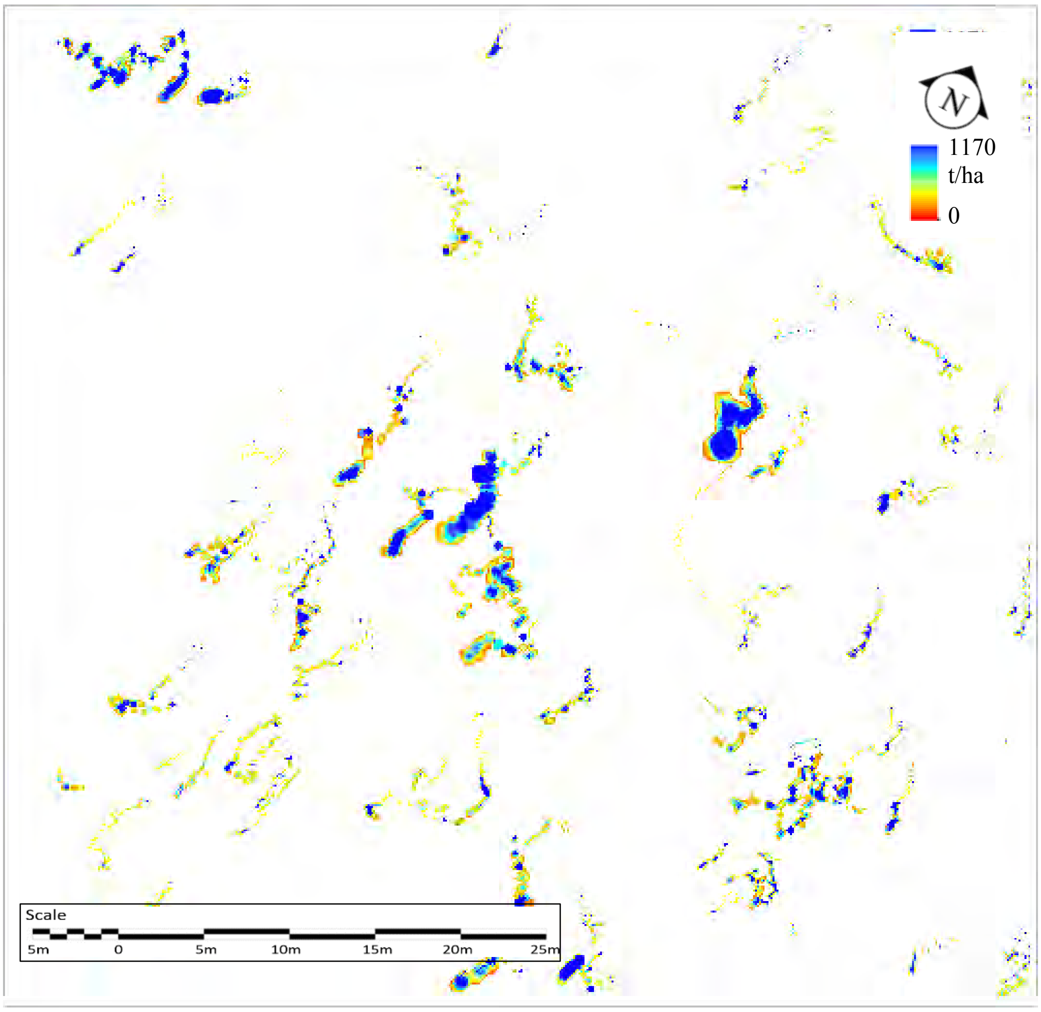


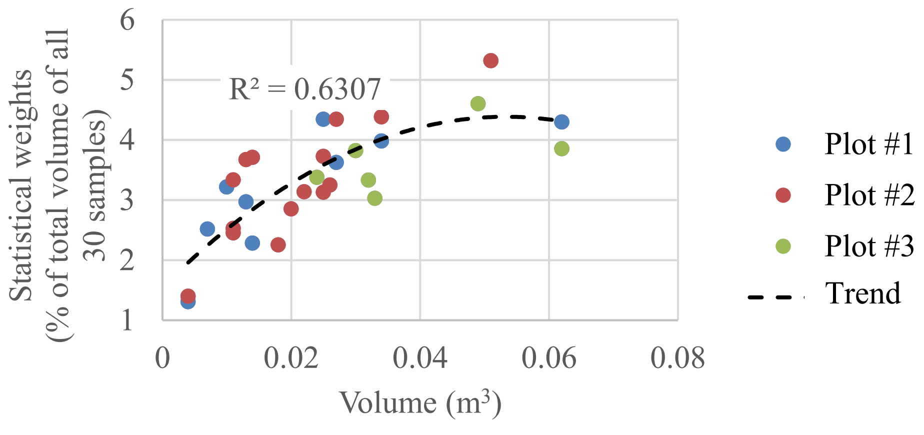
Figure 4-16: Resulting 2D tree trunk AGB map for a sample plot.

4.3.4 Branches Detection and Modelling

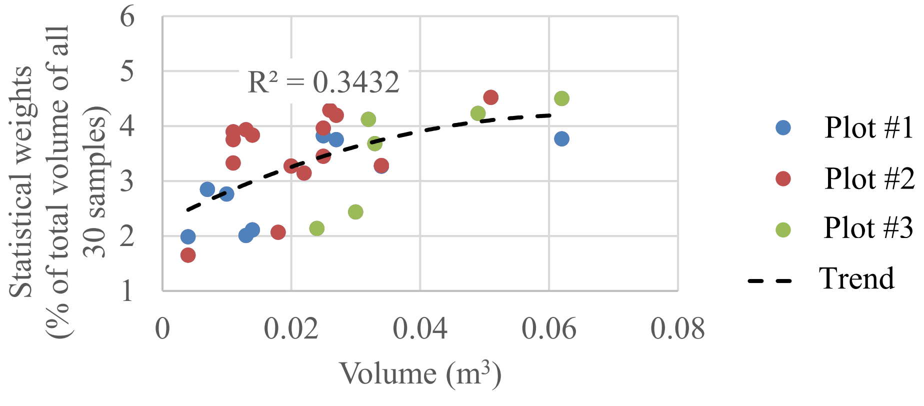
To reconstruct branches for each plot, trunk pixels were excluded from all layers and layers were segmented. Then the spatial parent-child relationship was built between successive layers' segments. The branches were assigned to the trees using nearest neighbour trunk centre, using the 3D model of trunks from the previous steps. Each branch was assigned to a statistical weight that represents the share of each branch from the total tree branch. As a comparison of branch weighting methods, 30 branches were randomly selected from three plots and their point cloud points were used to rebuild the branches by manually measuring radii along each branch CAD model (the point cloud in the AutoCAD environment). The manually reconstructed volume has a correlation of R^2=

0.63 for cylinder fitting compared to R^2=0.34 and 0.59 for the skeletonisation method as shown in Figure 4-17. The average time cost for the weighting of a single branch was 2.5s for cylinder fitting (least squares fitting (Wei and Wang, 2009)), 0.1s for the PC method (PCA Python library (Risvik, 2008)) and 0.1s for the skeletonisation method (skimage Python library (Van der Walt et al., 2014)). The branch biomass was calculated using allometric equations provided in Zianis et al. (2005) and Cai et al. (2013) and then this biomass was distributed to the tree branches based on the skeletonisation weight of each branch with the biomass of each branch redistributed to its pixel's weight.

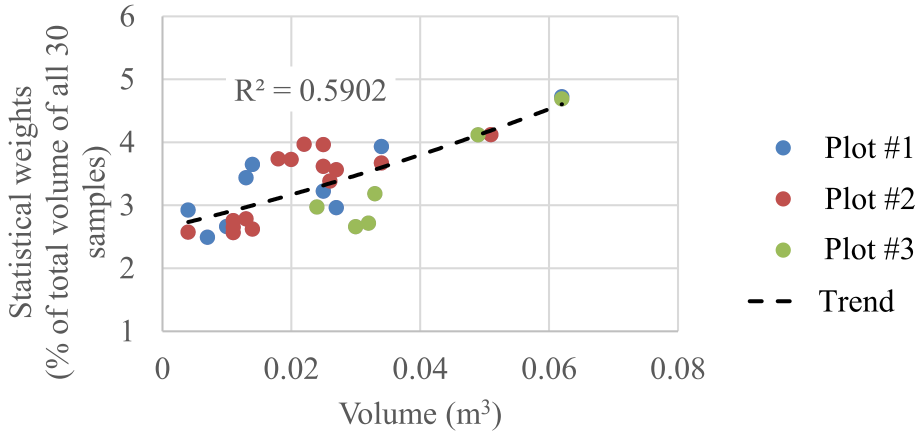
The biomass of each single pixel was added to the branch biomass layer to produce the branch biomass map. For example, Figure 4-18 shows the branches biomass of the same sample plot as shown in Figure 4-16.



-A-



-B-



-C-

Figure 4-17: Correlation between manually measured volume of branches (extracted from three sample plots) and statistical weights resulted from: A. Cylinder fitting method, B- PC method and C. Skeletonisation method.

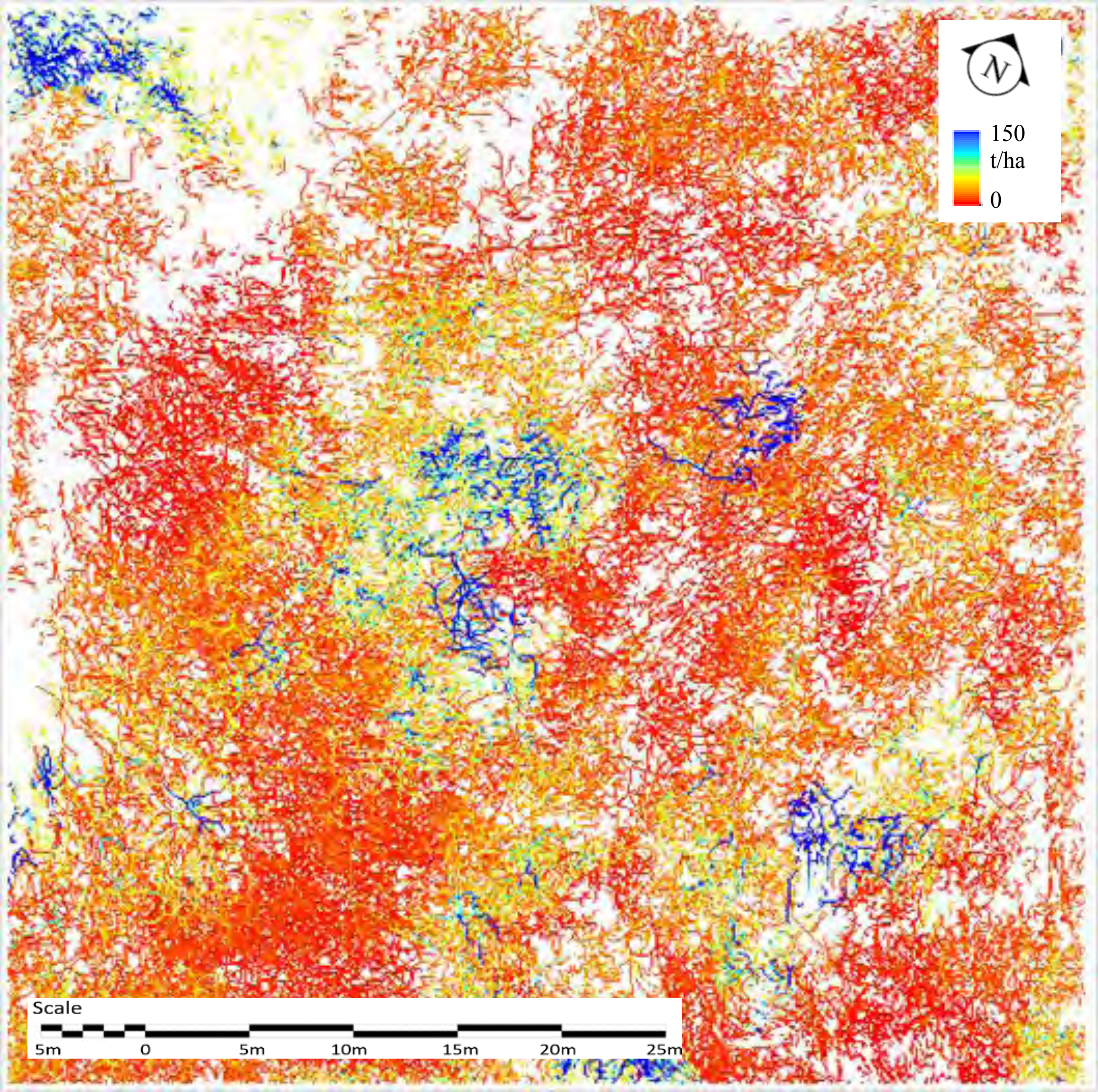


Figure 4-18: Resulting 2D tree branch AGB map for a sample plot.

4.4 Summary

An algorithm was developed to map the woody biomass of all trees in a specific area, in order to combine the high accuracy of allometric equations and the high precision of biomass spatial distribution. TLS was used to both provide the input data for and evaluate the accuracy of the resulting maps. A set of techniques were used to increase the efficiency, by increasing the accuracy and reducing the computational cost. New techniques have been identified, including extracting the ground surface using a minimal proportion of the point cloud, producing a new type of voxel, supporting the trunk detection by intersecting layers, reconstructing branches through segment parenting, and distributing biomass for each tree to its pixels through statistical weighting.

Currently, the tree detection is based on a user's visual inspection to classify the detected trees into true or false detections. However, the algorithm was designed to keep the results and so can be used as training data for a future automated machine learning approach.

The computational efficiency comparison for the DBH determination methods found that fitting an elliptical cylinder, which is inclined, provided the highest accuracy but also the highest computational cost. In addition, there were a significant number of cases when the fitting process failed to converge. Therefore, the elliptical cylinder fitting was applied with a vertical Z-axis orientation. In addition, the branch weighting methods comparison shows that time efficiency can be increased by accepting slightly lower precision techniques, without affecting the total accuracy, by applying the skeletonisation method.

The resulting maps of trunk biomass and branch biomass have a high spatial resolution, similar to the voxel resolution. The approach used in the nine plots in this test stage, was generalized on six sites that included 38 plots. These plots provide the field data that were analysed for errors in Chapter 5 and implemented for the AGB assessment in Chapter 6. A summary of the field sites results is listed in Table 4-4.

Table 4-4: A summary of fieldwork data analysis results for the six field sites.

Site No.	Number of plots	Average Stem density (tree/ha)	Average DBH (cm)	Average tree height (m)	Average AGB density (t/ha)	Species percentage
1	4	310	13.0	12.3	39.1	Ash 95% Oak 5%
2	9	385	28.9	16.0	96.6	Birch 98% Scot pine 2%
3	6	106	51.1	17.3	165.7	Mixed Conifers about 70%: (spruce, and pine) Broadleaves about 30% (Oak and Beech)
4	6	86	62.5	16.2	189.6	Mixed Conifers about 55%: (fir, and pine) Broadleaves about 45% (Hazel, plumb and Sycamore)
5	3	129	47.3	18.7	115.0	Oak 65% Beech 35%
6	10	180	41.6	20.6	123.4	Scot pine 100%

Chapter 5 Error Analysis Software Tool

5.1 Introduction

As discussed in chapter two, there is a lack of theoretical approaches to spatial error analysis of biomass assessment systems. The available solutions for non-linear models are complex and computationally intensive. These solutions are further constrained, as the vast majority of available algorithms is restricted to parametric regression models and not available to non-parametrically derived models like neural network systems. In addition, the available approaches do not provide a complete solution for models that include object-based analysis. Overall, these theoretical limitations have resulted in a paucity of software tools available for AGB uncertainty analysis. In addition, they could become more pronounced as the role of more complex systems, which use non-parametric models, and object-based analysis increases in usage.

In light of this need, this chapter aims to provide solutions that bridge this gap, and use these solutions as underlying concepts to build a relatively complete software tool for AGB uncertainty analysis. Therefore, the objectives are threefold. First, to design the tool to be inclusive of the widest range of model types, fieldwork data types, and RS datasets combinations that include both pixel-based and object-based approaches. Second, to minimize the computational costs. This minimisation can be achieved by reviewing the computationally costly steps for producing error distribution maps. Third, to aim for freely available software by avoiding a dependency on any non-free commercial packages within the tool. In addition, there is the complementary design objective of providing the simplest possible user interface; helping to reduce the required training to use the tool.

One of the major hurdles is that even with minimized computation costs, the designed tool is still expected to be computationally expensive. Utilizing the maximum hardware capabilities has been taken into consideration as a solution to accelerate the processing. Therefore, the code aims to support the parallel multiprocessing technique that is suitable for multi-core computers and the graphics-processing unit computing that is suitable to the Graphics Processing Units (GPUs) whenever possible.

Another aim for this chapter is to provide the required calculations to understand the errors inherited in a number of RS and field datasets. This can help to apply the software tool for practical scenarios in the next chapter. Therefore, this chapter discusses the software tool and the random errors of field and RS data.

5.2 Underlying Concepts

As stated in chapter 3, the RS inputs could be a single dataset or a combination of more than one dataset and each dataset could be one or a wide variety of sources. Similarly, the field dataset can have a number of specifications with both the RS and field data passing through a number of data preparation steps. Currently, the data preparation steps are not included within the analysis tool. However, a number of data preparation processes will be discussed later on in this chapter.

The error analysis in the model production stage deals with both RS and field data errors, which includes both spatial and spectral errors. Previous studies, such as Ahmed (2012), Colgan et al. (2013), Chen (2013), Weisbin et al. (2014), and Chen et al. (2015) ignored spatial errors and non-parametric models, therefore, it was possible for them to use the Taylor method for error propagation. However, regardless of the model type, it is possible to use the standard equations for Monte Carlo error propagation (Equation 5.1). The aim of the MC method is to apply the model or system under study (Y) repeatedly. Each time, different randomly generated errors ($rge_{1,i}$, $rge_{2,i}$,…, $rge_{n,i}$) were applied to the most probable values of the inputs (X_1, X_2, …, X_n).

$$\mathbf{Y_i = Y\left(X_1+rge_{1,i}, X_2+rge_{2,i},\ldots, X_n+rge_{n,i}\right)} \qquad (5\text{-}1)$$

Each ***rge*** is a simulation of the expected error for each individual input. If the process is repeated a sufficient number of times, the group of outputs will be the whole distribution set of expected outputs $\mathbf{F_Y(.)}$ for **Y**. $\mathbf{F_Y}$ can be statistically analysed to compute the most probable values, errors, variances and other statistical parameters for Y.

At the model formalization stage, Y represents the model parameters and Xs represents the RS and field data therefore ***reg***s have to be spectral/ non-spatial errors for RS data and quantitative errors for field data. Spatial errors are not algebraically addable, and when geo-referencing is applied the geometric correspondence of the dataset with other datasets will differ. Therefore, there is no mathematical representation between the spatial error value and the model characteristics at this stage. However, the MC concept can still be used to simulate these errors stochastically and generate a new model training dataset for each MC iteration. This can then be applied in parallel to the algebraically addable spectral/ non-spatial errors. Equation 5.2 is the mathematical representation of this application of MC. This concept cannot be applied with deterministic methods, and therefore the workable techniques are limited to those where MC can be applied.

$$\mathbf{Y_i = Y\left([X_1]_{\text{spatial error }2,i}+rge_{1,i}, [X_2]_{\text{spatial error}_{2,i}}+rge_{2,i}, \ldots, [X_n]_{\text{spatial error }n,i}+rge_{n,i}\right)} \qquad (5\text{-}2)$$

The number of random errors required to simulate the regs is equal to the summation of pixels of all inputs because each pixel has its own noise. This high dimensionality, combined with the limited number of iterations an affordable MC can allow, limits the MC to be applied to error propagation without any attempt at correction. On the contrary, to simulate spatial errors, only three components of error are needed for each dataset. These components are the error in x coordinate, the error in y coordinate and the error in the orientation (in the x-axis direction). With this dimensionality, it might be possible to find the spatial corrections that are statistically probable because they give the highest correlation between RS and field data.

After data error propagation is used to define model errors, the model errors and RS data errors are propagated to the results. The cost of this process is expected to be huge due to the fact that it has to be applied to the whole study area compared to dealing with only the much smaller fieldwork areas as for the previous stages.

On one hand, for parametric models, the high-dimensional spectral errors require a larger MC iteration compared with the lower dimensionality spatial errors. The suggested solution is to separate this stage into two processes. One of them is to propagate the spatial errors that can propagated only with MC, and has low dimensionality, with a minimum possible number of iterations to control the computational cost. The other process is to propagate other errors that can propagated with other low cost deterministic approaches such as Taylor methods method.

On the other hand, non-parametric models cannot be partially derived for the independent variables, as required by the deterministic methods. The computational cost of applying a MC iteration to propagate errors to the whole study area can be significant enough to make this process unaffordable, especially for wide coverage studies. The suggested solution is to apply the error propagation in two stages. The first stage includes applying the error propagation to only the fieldwork areas. The second stage includes using the results of the first stage to train a machine-learning algorithm that can derive the expected uncertainty from the input values to calculate the AGB. The critical issue is being provided with enough samples for valid training. However, the chosen approach is that same number of sample plots/sample segments that is used for training the AGB assessment model will be used for the AGB error assessment model. Therefore, if the number of samples is enough to train the AGB model, it has to be enough for error mapping. Once again, the spatial errors are propagated separately because the spatial continuity of AGB cannot be represented as a feature for the machine-learning model.

One of the important options the tool has to provide is the ability to separate error sources. This can be done by giving the user the ability to turn on/off each type of error for each input. This flexibility allows the users to separate errors by error type, by input type or by a customized combination of errors. The other flexibility of the tool is allowing the user to decide the number of iterations and the model type.

5.3 Supported models

The main model types listed in Section 3.2 are supported by the software tool that includes simple regression, multiple regression, Support Vector Machine (SVM), Neural Network, Gaussian Process, Random Forest, and k Nearest Neighbours. Each one of these types is provided with a range of options that can be set by the user.

5.4 Implementation

This tool is entirely coded within the Python, which is a dynamically-typed object-oriented scripting language. It was chosen for the coding for several reasons, with the three most important being its simplicity, availability, and suitability for geospatial processing (Westra, 2016). It is relatively simple, neat, compact and elegant (Milano, 2013). In addition, it is an interpreted language that means it does not require compiling. In addition, Python is an open source language with a variety of free third-party software including statistical modules, profilers, graphical libraries, multiprocessing modules, and GIS modules.

Data importing and outcome map exporting are primarily built on the Geospatial Data Abstraction Library / OGR simple features library (GDAL/OGR) (OSGeo_Project). The

GDAL/OGR translator libraries can support 226 data formats that include all the RS raster and GIS vector standard data types. This allows the designed tool to directly import the original data.

Pandas (Pandas, 2017), Python data analysis library, is implemented to provide high-performance data structures. It can provide integrated, intuitive routines for performing statistical data manipulations and analysis on the tabular numerical data read from the original data (McKinney, 2011). This library was also chosen over other available libraries for its efficiency to reduce Random Access Memory (RAM) usage compared to the standard Python data structure (McKinney, 2011). However, other data structures are occasionally used for some processes such as random error generation with NumPy (Scipy, 2006). It is also used to export and import tabular data in the standard ASCII file formats.

Tkinter, the Python's standard library for Graphical User Interface (GUI) (Chaudhary, 2015) is used for constructing GUIs, which allow users to interact with the tool without requiring advance knowledge of the programming code.

StatsModels Python library (Seabold and Perktold, 2010) is used for simple, multiple, and stepwise regression models. This implementation of a verified package is used firstly to ensure that the statistical calculations are correct and secondly to make use of the extensive list of result statistics this package provides with each regression includes the correlation coefficients and the root mean square error.

The Scikits Learn library (Pedregosa et al., 2011) is deployed for the SVM, Gaussian process, Random Forest, and k-nearest neighbours models. The Scikits SVM variant used is the SVM for regression problems (SVR) that is based on Smola and Schölkopf (2004). The applied Gaussian process model is based on Lophaven et al. (2002), and the Random Forest model is based on Breiman and Cutler (2003).

In order to obtain the trained neural networks, TensorFlow (Abadi et al., 2016) was implemented. TensorFlow is one of the most advance systems for neural networks. It uses data flow graph technology that efficiently uses the full capabilities of the hardware system including multicore Central Processing Units CPUs, one or more Graphics Processing Units GPUs, and Google's Tensor Processing Units (TPUs). Moreover, TensorFlow supports cloud computing engines such as the Google cloud machine-learning engine, and Amazon elastic compute cloud.

5.5 Availability/Requirements

All required source code and installation URL links can be found in GitHub[1]. Because the tool can be run in Python without installation, there are no strict hardware requirements. However, there should be enough free memory to save the outputs as well as the inputs. Therefore, it is recommended to use a machine with a reasonable amount of RAM to deal with the raster data.

The parallel processing requires a multi-core CPU with a high speed; > 2.2GHz is recommended. When the Neural Network model is used, GPU operation is recommended to accelerate the processing speed. The tool is tested on a personal computer with 16GB RAM, a 3.6GHz Intel Core_i7 processor and a NVIDIA Quadro K2200 4GB GPU.

5.6 The User Interface

The user interface aims to provide a clear means to enable the user to feed the tool with all the required fields, RS data inputs, the user decisions regarding system parameters, and the considered error types. It also summarizes the inputs and provides the option to

---

[1] The tool was uploaded on https://github.com/AhmedAlboabidallah/EATool to be freely available

export them in a specific format, with the ability to import them and modify them later. Therefore, the overall approach is an interactive conversation with a series of windows.

Figure 5-1 displays a schematic diagram of the interactive components of the user interface. On the start-up, four groups of inputs are required related to the field data, RS data, Monte Carlo iteration, and processing type. The field data table properties selection has two options: the formatted table file (if there is one), and the number of the field inputs if no table file is available. The RS data table properties are similar to the field data table properties. The processing type can be pixel-based or object-based. The segments file is also required in the case of object-based processing. Spatial error of each component can be input as two values, the expected error in the coordinates and the expected error in the north direction. The non-spatial error of each component can be input as a constant when the non-spatial error is uniform all over the map. Otherwise, the user can provide the tool with the error map of the input. An example of the RS table windows is shown in Figure 5-2. This example also shows the colour codding for the RS inputs. This colour code shows the bands that are of the same database (have an identical dataset number) that are processed as if they all inherit an identical spatial error. For example, if a Landsat-8 image to be used, all bands are expected to have the similar spatial error. The similarity is due to the fact that these bands were processed using the same co-registration algorithm, geo-referencing algorithm, and with the same set of ground control points.

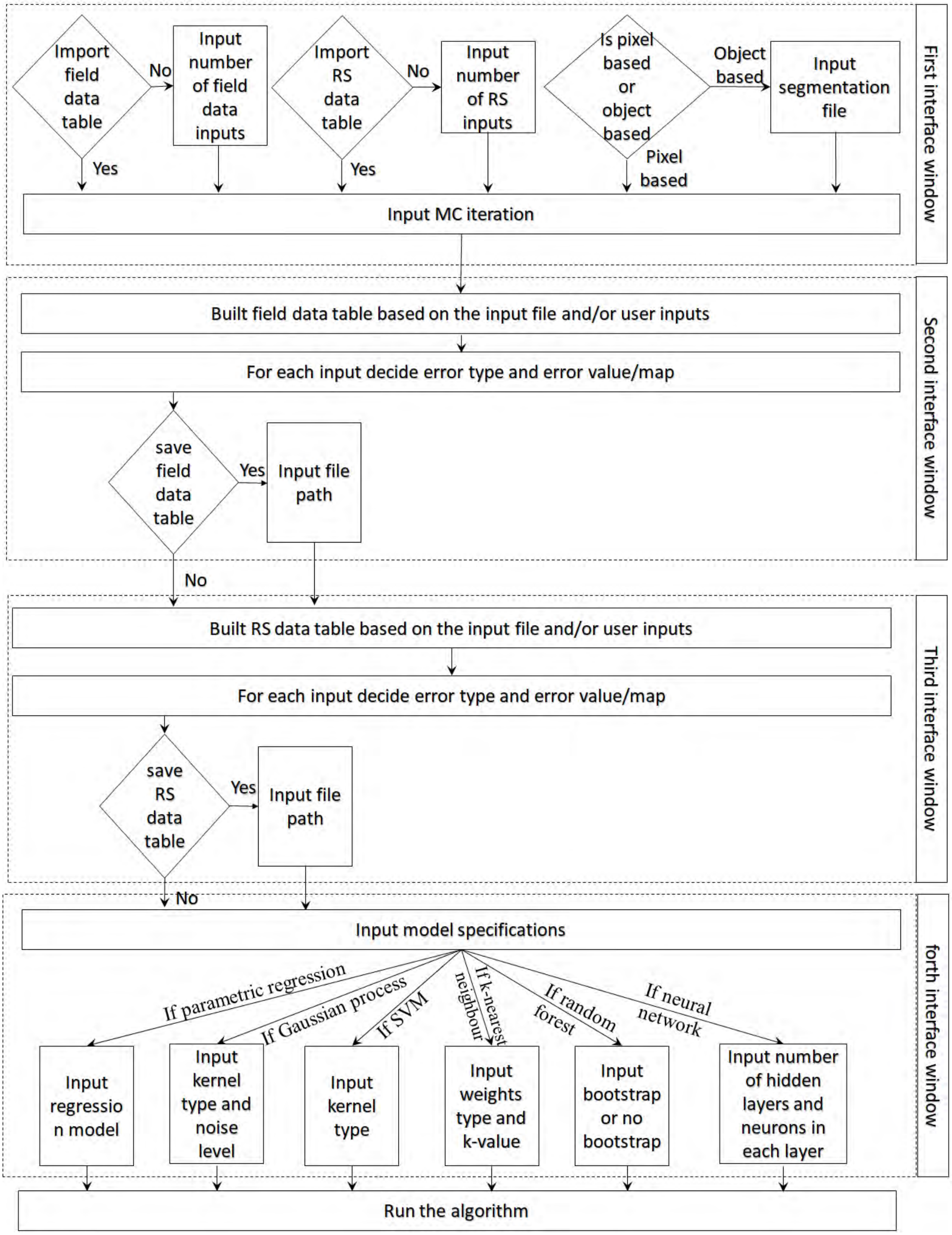
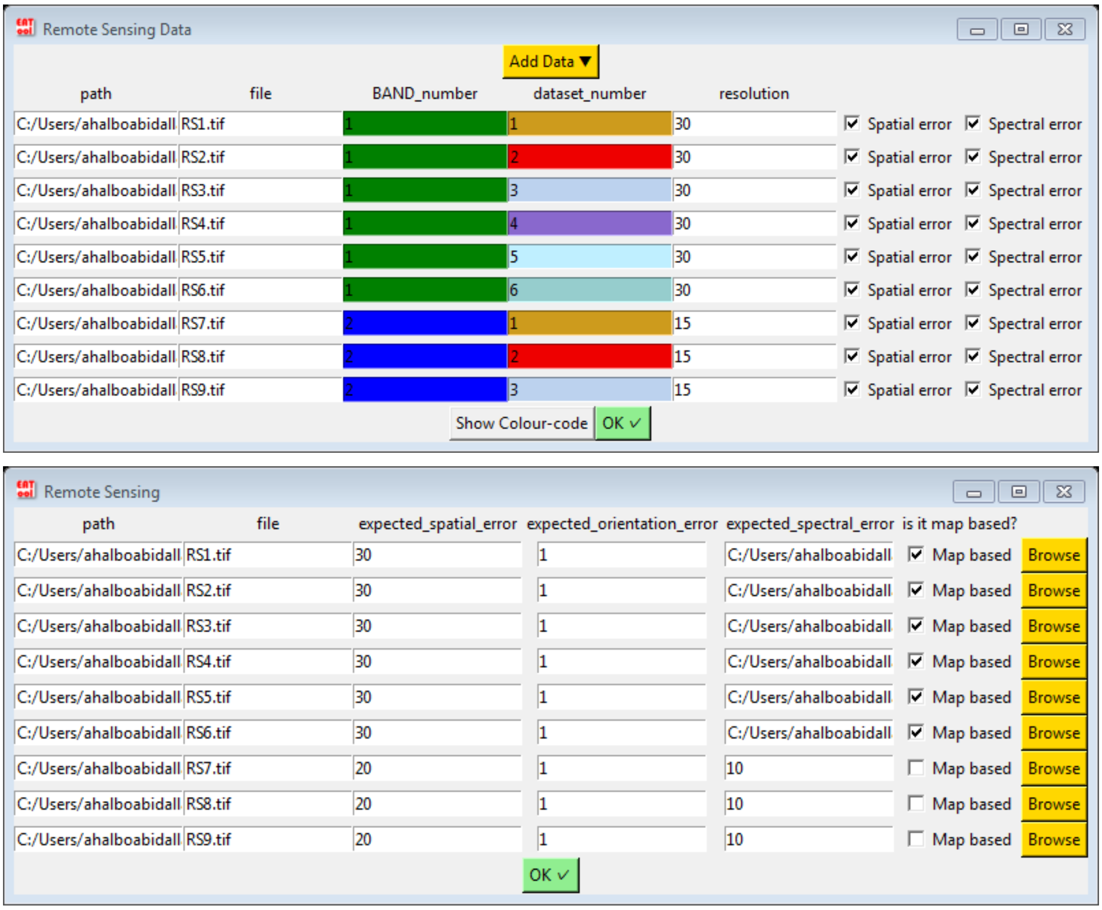


Figure 5-1: A schematic diagram that shows the components of the software tool interface windows and their relationships.



path	file	BAND_number	dataset_number	resolution		
C:/Users/ahalboabidall	RS1.tif	1	1	30	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS2.tif	1	2	30	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS3.tif	1	3	30	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS4.tif	1	4	30	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS5.tif	1	5	30	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS6.tif	1	6	30	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS7.tif	2	1	15	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS8.tif	2	2	15	☑ Spatial error	☑ Spectral error
C:/Users/ahalboabidall	RS9.tif	2	3	15	☑ Spatial error	☑ Spectral error

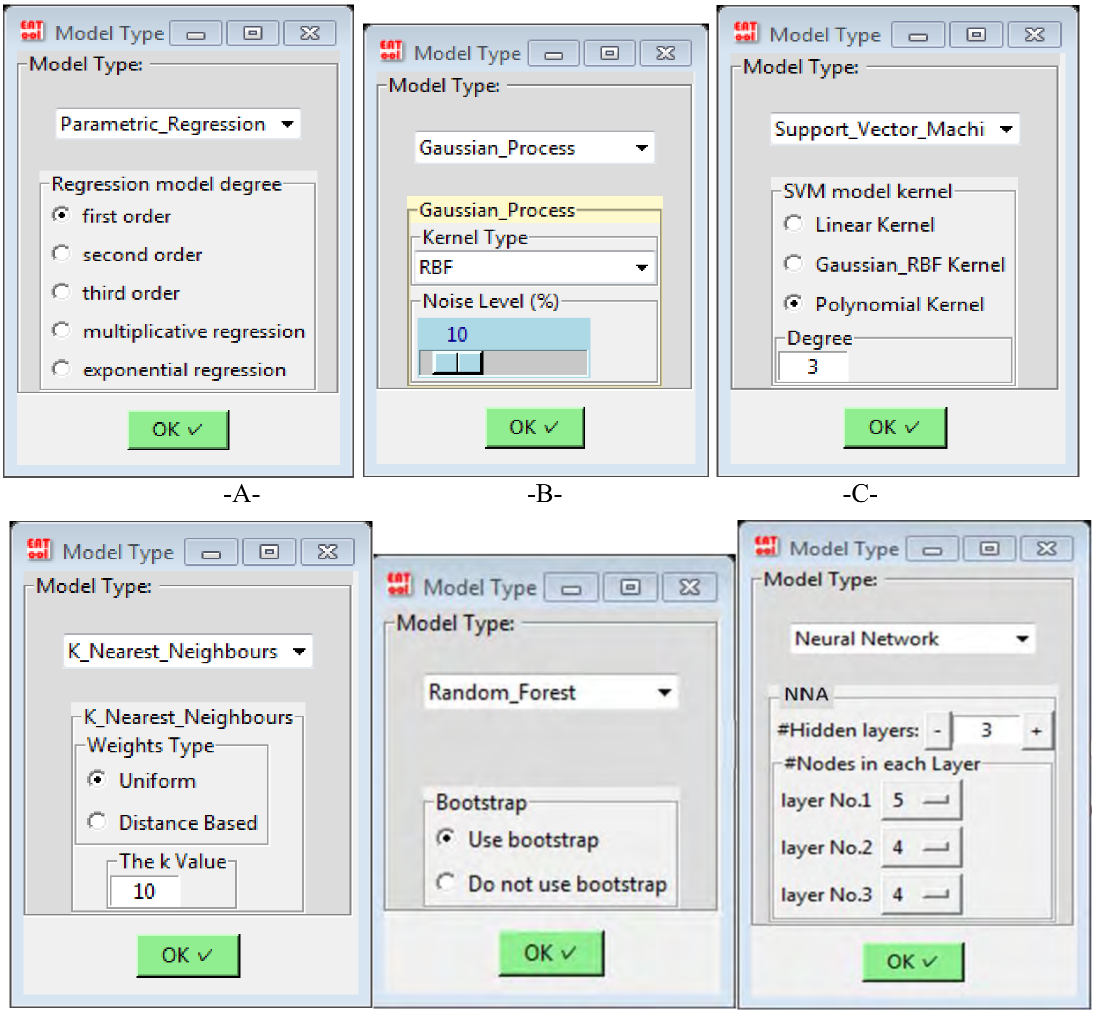
path	file	expected_spatial_error	expected_orientation_error	expected_spectral_error	is it map based?	
C:/Users/ahalboabidall	RS1.tif	30	1	C:/Users/ahalboabidall	☑ Map based	Browse
C:/Users/ahalboabidall	RS2.tif	30	1	C:/Users/ahalboabidall	☑ Map based	Browse
C:/Users/ahalboabidall	RS3.tif	30	1	C:/Users/ahalboabidall	☑ Map based	Browse
C:/Users/ahalboabidall	RS4.tif	30	1	C:/Users/ahalboabidall	☑ Map based	Browse
C:/Users/ahalboabidall	RS5.tif	30	1	C:/Users/ahalboabidall	☑ Map based	Browse
C:/Users/ahalboabidall	RS6.tif	30	1	C:/Users/ahalboabidall	☑ Map based	Browse
C:/Users/ahalboabidall	RS7.tif	20	1	10	☐ Map based	Browse
C:/Users/ahalboabidall	RS8.tif	20	1	10	☐ Map based	Browse
C:/Users/ahalboabidall	RS9.tif	20	1	10	☐ Map based	Browse

Figure 5-2: An example of the user interface window of RS data table.

The next window, Figure 5-3, is to select the properties of the mathematical model including the model type and its specifications. When a specific model is selected, its own relevant inputs will be required interactively. Both simple and multiple regressions are referred to as standard regression due to the fact that the number of inputs is open. In other words, if the model is implemented with only one independent input (single RS input) the resulted model will be a simple regression. Otherwise, if more than one RS input is used, the model will be a multiple regression model. In both cases, the user can decide upon the degrees of regression equation.

The SVM model options include the kernel to be used that can be linear, Gaussian RBF or polynomial. In cases when the polynomial kernel is to be applied, the degree of SVM can be set as well. For random forest, the available option is to use bootstrapping or not.

For the Gaussian process model, the available options are the kernel type and noise level. The available kernels are RBF, Matern, rationale quadratic, exponential sine squared, and the dot product kernels. The options for the k-nearest neighbour model are the k-value (the number of samples for each local estimation) and the weights type i.e., whether it is uniform or distance based. For the neural network model, the user can decide upon the number of hidden layers and the number of nodes in each hidden layer. The last required input is the filepath to the output workspace, where the outcomes will be kept.



Model Type: Parametric_Regression
Regression model degree
● first order
○ second order
○ third order
○ multiplicative regression
○ exponential regression

-A-

Model Type: Gaussian_Process
Gaussian_Process
Kernel Type: RBF
Noise Level (%): 10

-B-

Model Type: Support_Vector_Machi
SVM model kernel
○ Linear Kernel
○ Gaussian_RBF Kernel
● Polynomial Kernel
Degree: 3

-C-

Model Type: K_Nearest_Neighbours
K_Nearest_Neighbours
Weights Type
● Uniform
○ Distance Based
The k Value: 10

-D-

Model Type: Random_Forest
Bootstrap
● Use bootstrap
○ Do not use bootstrap

-E-

Model Type: Neural Network
NNA
#Hidden layers: 3
#Nodes in each Layer
layer No.1: 5
layer No.2: 4
layer No.3: 4

-F-

Figure 5-3: The model selection window, A - available options when parametric regression is selected, B - available options when Gaussian process is selected, C - available options when SVM is selected, D - available options when k nearest neighbour model is selected, E - available options when random forest model is selected, and F - available options when neural network is selected.

5.7 Output Files

The analysis outputs are a number of files that are saved to the user's specified directory/folder. The files are text files of the model parameters and the equivalent correlation coefficient for every MC iteration, and the binary files of the models required to calculate AGB and error maps. Specifications

5.8 Testing the Tool

The aim of the scenarios is to cover a range of AGB assessment datasets, mathematical models, and processing techniques that have been used by previous researchers, as highlighted in Section 3.2. To achieve this aim, four scenarios were designed to cover the processing types, RS data types and AGB assessment mathematical models. It was not practically applicable to test all possibilities of the variety of parameters described in chapter three. Therefore, it was considered reasonable to test a sample of each main category.

The RS data was classified according to the system type (passive or active), and according to the spatial resolution (high and very-high). The model types were classified into parametric and non-parametric models. The processing types into pixel-based and object-based. A wide variety of data processing procedures to derive new products from raw RS data are also covered such as using inter-band calculations, Principal components, Radar data interferometry, Digital Canopy Model (DCM) production, and texture features.

5.8.1 High Resolution RS Data and products

5.8.1.1 Landsat-8 Products

The two Landsat-8 scenes listed in Table 5-1 were used for model training for the first two scenarios. Landsat-8 images are sensitive to cloud cover: only a few images in the

Landsat archive have no cloud cover over the study area, resulting in the scenes having a time difference of up to 6 months from the field data acquisition. Among those images, the chosen data were selected to be the closest to canopy peak in September and the canopy minimum in June for the majority of tree species (Melaas et al., 2013). The images were downloaded from the United States Geological Survey (USGS) (USGS, 2017) as Level-1T Terrain Corrected images having a 30 m spatial resolution. The Level 1 processing means that the images are radiometrically corrected, orthorectified, and are rectified to Universal Transverse Mercator (UTM) and the World Geodetic System (WGS 84) datum and ellipsoid. The digital number was converted to top-of-atmosphere reflectance using the information in the provided Landsat metadata ancillary data file (MTL) and the Dark Object Subtraction (DOS) atmospheric correction was applied using the Semi-Automatic Classification Plugin pre-processing tool (Congedo, 2013). There are more accurate procedures that can be applied for atmospheric corrections. However, for this study's objectives, the top-of-atmosphere reflectance and DOS is practical and satisfactory as the error analysis to be applied is linked to the non-systematic errors such as noise and not directly connected to systematic errors such as atmospheric errors.

Table 5-1: High resolution Earth observation data details.

#	Mission	Position	Date	Weather	Vegetation condition
1	Landsat-8	Path:204, Row:25	30/09/2015	Dry	Leaf on
2	Landsat-8	Path:204, Row:25	20/01/2016	Dry	Leaf off
3	Sentinel-1	Absolute Orbit: 010325	11/03/2016	Dry	Leaf partially on
4	Sentinel-1	Absolute Orbit: 010150	28/02/2016	Dry	Leaf off
5	Sentinel-1*	Absolute Orbit: 009625	23/01/2016	Wet	Leaf off
6	Sentinel-1*	Absolute Orbit: 014350	12/12/ 2016	Dry	Leaf off
7	Sentinel-1*	Absolute Orbit: 014525	24/12/ 2016	Dry	Leaf off

* Used only for data temporal stability, not for model production and biomass calculations.

Landsat-8 images were subset to the study area, and Normalized Difference Vegetation Index (NDVI) has been calculated by applying the NDVI standard equation (Equation 5-3) that is based on near infrared band (NIR) and red band (R):

$$\mathrm{NDVI}=\frac{\mathrm{NIR-R}}{\mathrm{NIR+R}} \qquad (5\text{-}3)$$

A new index, the NDVI Range (NDVIR), was designed to distinct between different tree types. Needle-leaved tree land cover is likely to witness a less variation in the NDVI compared to broadleaf tree land cover over seasons. Therefore, there is a possibility that the algebraic difference in NDVI can be related to the tree land cover type. NDVIR was computed simply by subtracting the NDVIs (for 30/09/2015 and 20/01/2016) from each other's. NDVI values range between -1 and +1; the NDVI differences are expected to be between -2 and +2. Therefore, to get positive values, a constant (C = 2) was added as in Equation 5-4 to ensure positive values for all NDVIR pixels. The resulting NDVIs and NDVIR are shown in Figure 5-4.

$$\mathrm{NDVIR} = \mathrm{NDVI}_2 - \mathrm{NDVI}_1 + C \qquad (5\text{-}4)$$

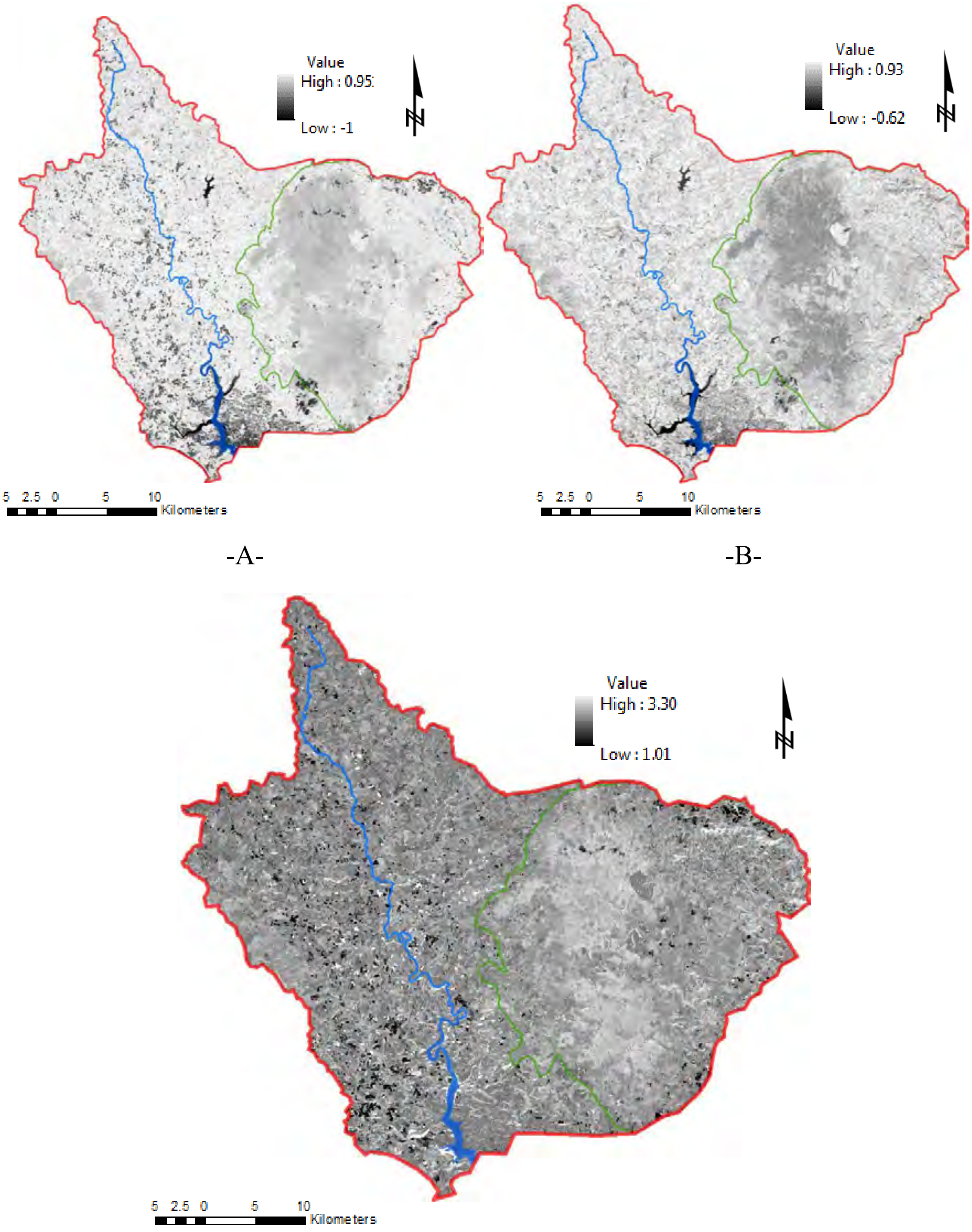
The Landsat-8 dataset number 1 (for 30/09/2015) was classified with supervised Maximum Likelihood Classifier (MLC) software (Nolè, et al., 2015). This classification was reported by previous studies such as Walker et al. (2010), Eckert et al. (2011), and Laurin et al. (2016) as a reliable algorithm for AGB classification. The land cover was classified into five classes, namely tree cover, vacant land, non-tree open area, built up area and water body. The required information for the supervised classification were collected in the field with aid of GPS observations. A pixel-based classification process was applied for the first scenario, while an object-based classification analysis was applied to the second scenario. For the pixel-based classification, shown in Figure 5-5, the percentages of user and producer accuracies, overall accuracy and Kappa coefficient are listed in Table 5-2. For the object-based classification, the percentages of user and producer accuracies, overall accuracy and Kappa coefficient are listed in Table 5-3. The classification result is shown in Figure 5-6. The Landsat-8 dataset number 2 was used alongside with Landsat-8 dataset number 1, for NDVIR production. All the resulted products were used later on as inputs for the feature selection in Section 6.2.1 and Section 6.3.1.

Table 5-2: Accuracy assessment for pixel-based land-cover classification based on Landsat-8: producer's, user's and overall accuracies, and the Kappa statistic.

	Tree cover	Waterbody	Vacant land (bare soil)	Non-tree green Space	Built-up area	Summation	User accuracy
Tree cover	109	1	0	5	0	116	94.8%
Waterbody	3	35	1	4	0	42	83.3%
Vacant land (bare soil)	2	1	18	4	3	26	69.2%
Non-tree green Space	10	0	5	19	1	32	59.4%
Built-up area	0	0	3	1	18	22	81.8%
Summation	124	37	27	33	22	238	
Producer accuracy	87.9%	94.6%	66.7%	57.6%	81.8%		
Overall accuracy=84.5% Kappa =0.79							

Table 5-3: Accuracy assessment for object-based land-cover classification based on Landsat-8: producer's, user's and overall accuracies, and the Kappa Statistic.

	Tree cover	Waterbody	Vacant land (bare soil)	Non-tree green Space	Built-up area	Summation	User accuracy
Tree cover	214	2	4	19	0	239	89.5%
Waterbody	2	42	3	4	0	51	82.4%
Vacant land (bare soil)	7	3	54	8	3	75	72.0%
Non-tree green Space	12	5	5	59	1	82	72.0%
Built-up area	0	1	3	1	35	40	87.5%
Summation	235	53	69	91	39	487	
Producer accuracy	91.1%	79.2%	78.3%	64.8%	89.7%		
Overall accuracy=83.0% Kappa =0.75							



-A-

-B-

-C-

Figure 5-4: The resulting NDVIs and NDVIR layer as: A - NDVI1 layer, B - NDVI2 layer and C - NDVIR layer

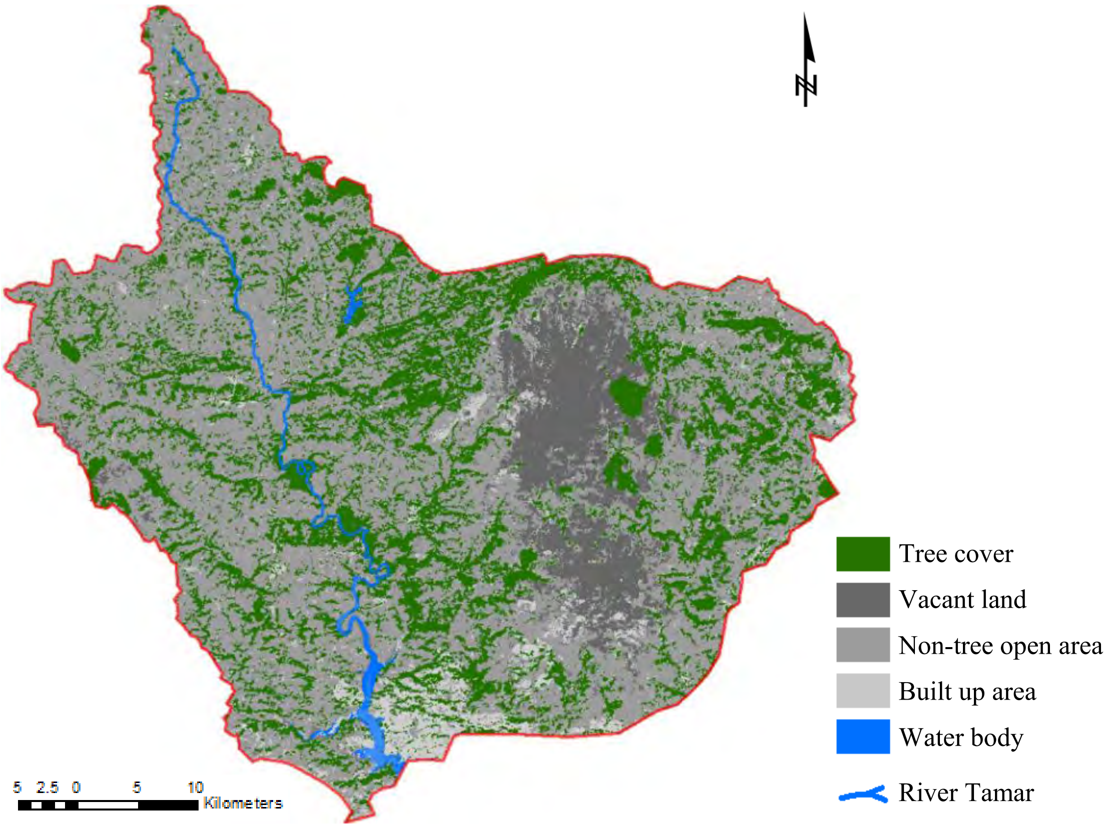


Figure 5-5: First scenario pixel-based classification result.

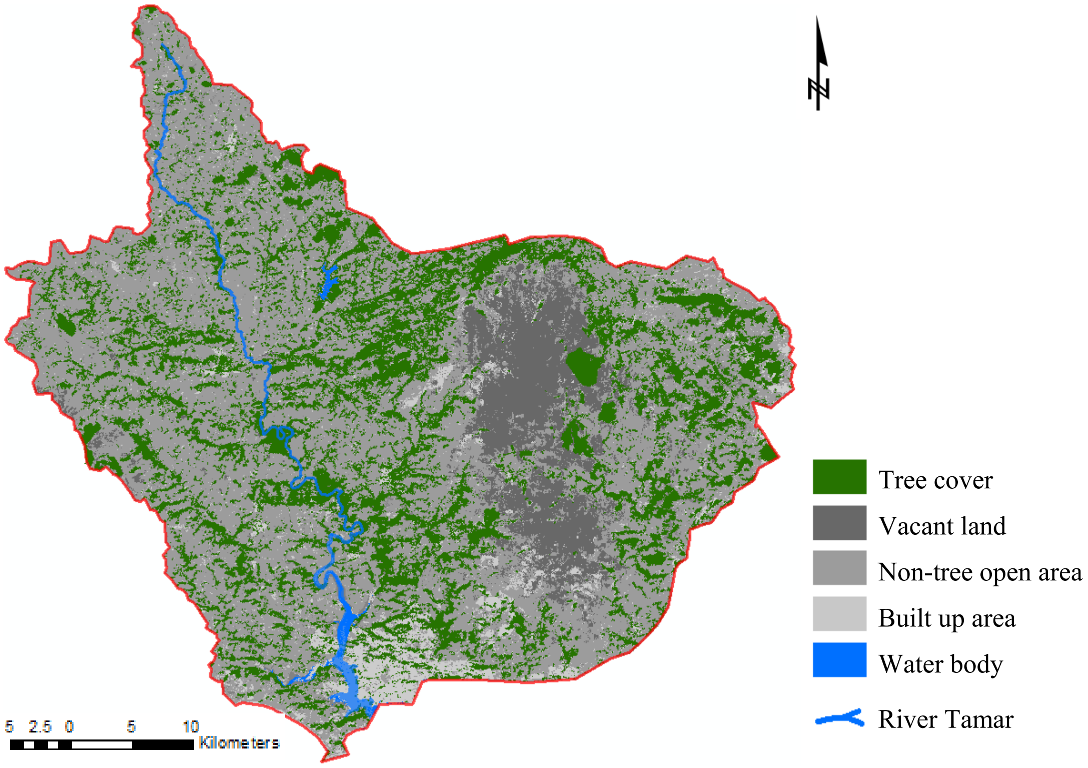


Figure 5-6: Second scenario object-based classification result.

5.8.1.2 Sentinel-1 and Interferometric Local Heights Differences

The Sentinel-1 two constellation satellites A and B, launched on 2014 and 2016 correspondingly, can provide C-band SAR images on dual polarization capacity (HH+HV and VV+VH). As part of European Space Agency (ESA)'s Copernicus, the Sentinel-1 images are distributed for free to all users. Two Sentinel-1 images, as listed in Table 5-1, were used for model training and AGB calculations. The other three Sentinel-1 images were used to study the Sentinel-1 data stability over time. The images were chosen to have an acquisition date close to the field data collection dates. They were Interferometric Wide (IW) swath, Single Look Complex (SLC) products with dual vertical-vertical (VV) and vertical horizontal (VH) polarisation. The IW mode captures three sub-swaths using Terrain Observation with Progressive Scans SAR (TOPSAR) for each image. The SLC mode images are of 5 m spatial resolution in the range direction and 20 m in the azimuth direction (Jung et al., 2013).

The datasets were delivered by the Sentinels Scientific Data Hub (Copernicus) with Level 1 processing that means the internal calibration, absolute Doppler centroid estimation (the required correction for the range and azimuth variation depends on the satellite attitude and satellite attitude as a function of time), and the single look complex focusing (converting the complex amplitude and phase information into SLC data) are already performed. Therefore, the only required pre-processing includes topographic correction. This process had been done with the final processing steps because it conflicts with the SNAPHU phase unwrapping software that requires Range increasing towards the right in the interferogram file (Chen, 2001).

To map the interferometric Local Heights Differences (LHD), a digital surface model (DSM) is produced with the terrain heights map and height differences calculated. To apply the procedure an interferometric processing chain uses both the SNAP Sentinel-1 toolbox and SNAPHU software. First, a sub-pixel co-registration was applied with the first Sentinel-1 image selected as the master and the second image as the slave. The co-registration was done with interpolation of the a priori digital-elevation-model automatically downloaded by the Sentinel-1 toolbox. Then, the interferogram is formed to produce interferometric phase and coherence. The results are then deburst (converted to continuous image in terms of azimuth time) and merged into one image. The HV polarization deburst topographic phase product is removed and the resulting layer exported to SNAPHU, with the output of phase unwrapping being the estimated unambiguous phase values derived from observed phase. The terrain heights map is then produced after the data has been imported back into the Sentinel-1 toolbox.

The heights map shows the general trends of the terrain surface and the local effect of the surface features including trees. To override the general trend, and improve the local features effect, height maximum and minimum maps were produced and subtracted from each other using a 3x3pixel convolution, and then normalized to be in the range between 0 and 1. The LHD map before normalization is shown in Figure 5-7.

5.8.1.3 Stability of Local Heights Differences

This study explored LHD as a new types of input. Based on the visual comparison between tree map in Figure 4-1 and the LHD, Figure 5-7, there is a noticeable co-existence between relatively large LHDs and tree areas. Therefore, it was reasonable to test the temporal stability for LHD over time. The first evaluation included comparing the LHD that is used for the model (derived from images number 3 and 4 in Table 5-1) with the LHD shown in Figure 5-8A (derived from images number 4 and 5 in Table 5-1).

This dataset is chosen to provide a contrast in the weather conditions compared with the original dataset. Therefore, image 5 was selected to be in a wet weather. The difference between the two images is clear, but for a systematic comparison the scatterplot of each pixel's value with the corresponding pixel's value is plotted based on the normalized LHD; shown in Figure 5-9.

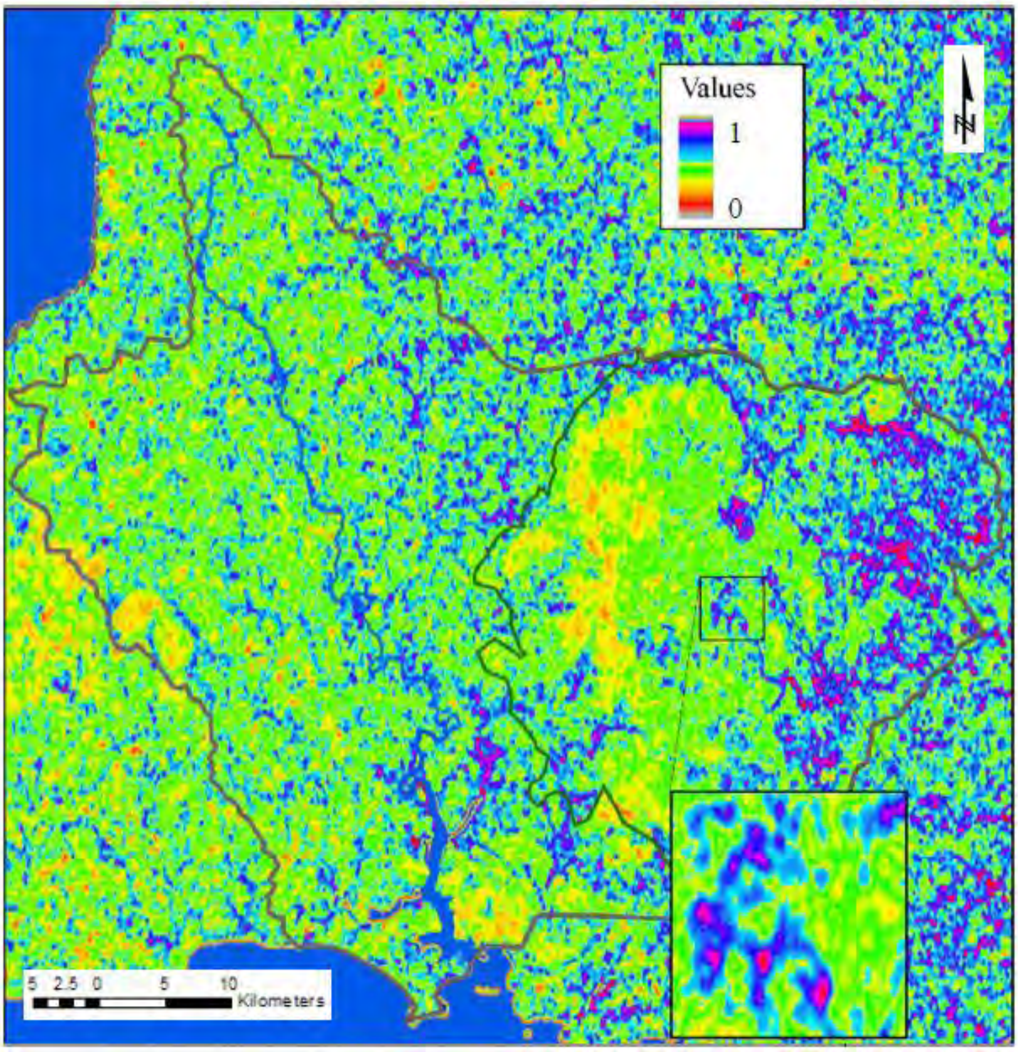


Figure 5-7: The normalised interferometric heights local differences layer. The inset shows details of an enlarged-scale sample area of the LHD map.

The second test is applied on another LHD dataset (derived from images number 6 and 7 in Table 5-1). The images were selected to be in dry weather conditions, similar to the models original dataset. The result shows a clear visual similarity, with the scatterplot shown in Figure 5-9B. The same datasets were analysed for another important factor which is the leaves on/off situation. This factor was explored based on the comparison between evergreen tree cover with deciduous tree cover.

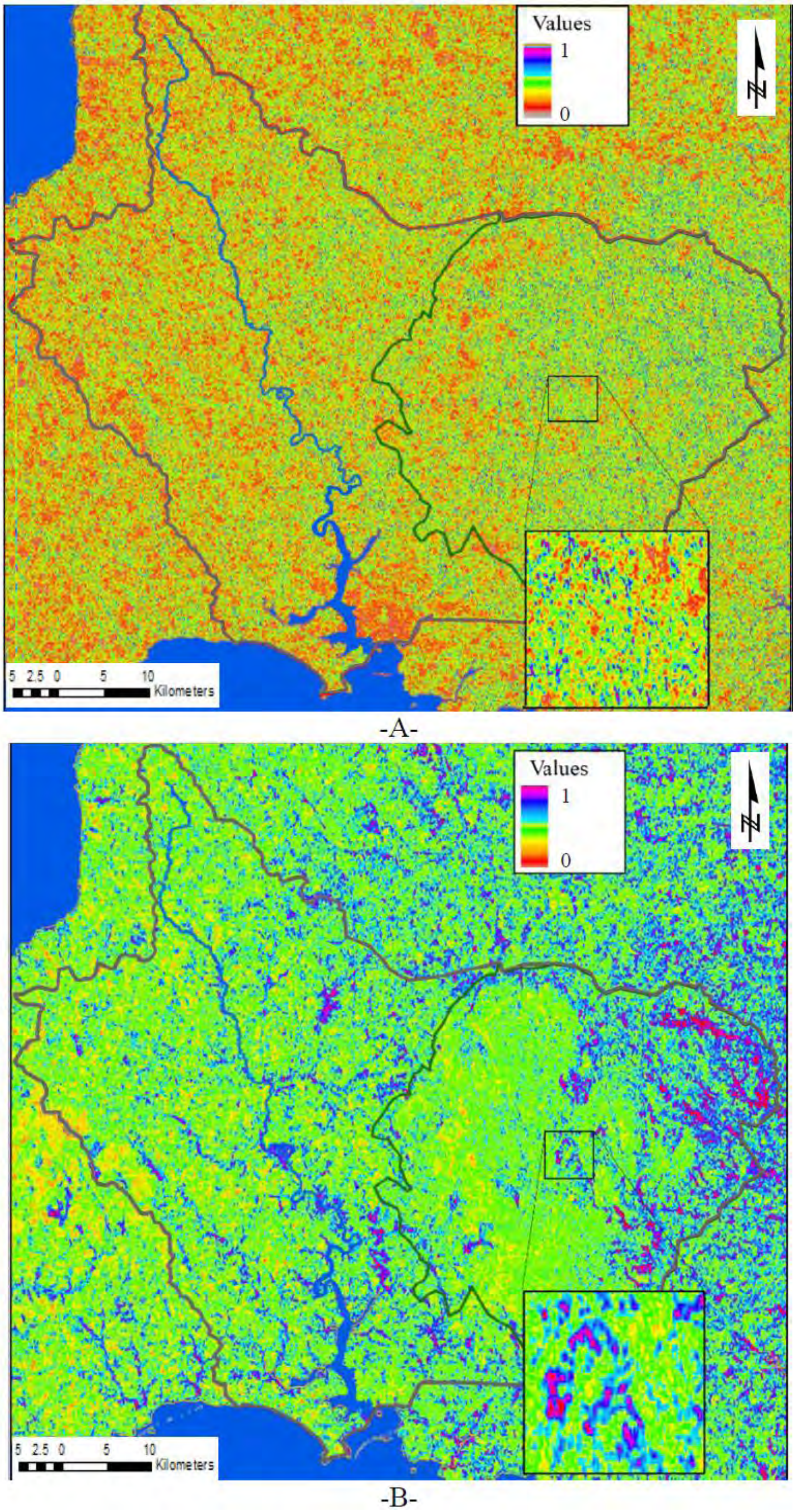
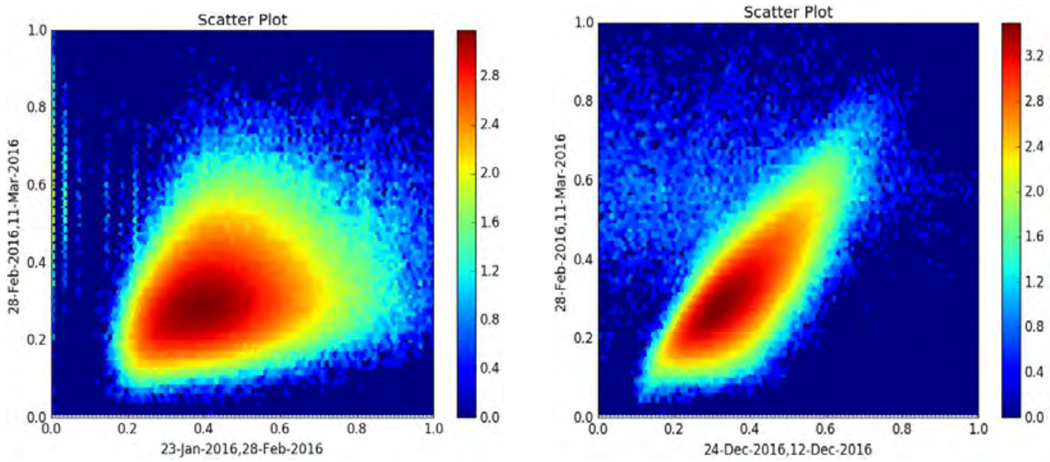


Figure 5-8: LHD layers used for comparison. A – normalised LHD derived from image 4 and 5 and B – normalised LHD derived from image 6 and 7. The insets show details of the equivalent sample to the enlarged-scale area in Figure 5-7.

The scatterplot of the second test shows a relatively high similarity between the normalized LHD comparing with the first dataset. However, the pixel value scale before normalization is not the same: the first LHD map has a mean of 115 m and a standard deviation of 116 compared to 85 m and 89 for the second LHD map. By visually comparing the LHD for the fieldwork areas, it is noticed that LHD showed a relatively similar sensitivity to both evergreen to deciduous tree covers. In addition, the same tree cover areas showed a clear dissimilarity between LHD derived from different weather condition datasets. For example, the subset in Figure 5-7, Figure 5-8A, and Figure 5-8 shows LHD of evergreen tree cover areas that is not expected to be largely affected by seasonal changes. However, it shows a clear contrast due to the weather conditions. Therefore, this instability should be taken in consideration when implementing more than one dataset for an AGB assessment.



-A- -B-

Figure 5-9: Scatterplot for the normalised model image vs A – normalised dry weather dataset and B – normalised wet weather dataset.

5.8.2 Very-High Resolution RS Data

5.8.2.1 WorldView-3 Products

The optical WorldView-3 sensor, launched by Digital-Globe in August 2014, is a multispectral commercial satellite with a very-high spatial resolution. It provides eight multispectral (MS) bands, eight Shortwave Infrared (SWIR) bands, twelve Clouds, Aerosols, Vapours, Ice, and Snow (CAVIS) bands and one panchromatic band. The utilized data were captured on 30 June 2015 in cloud free conditions, with two subsets extracted from the original image. The subsets includes three of the six fieldwork sites as shown in Figure 5-10. The 8 MS band bundle is 2 m spatial resolution and were delivered as product level LV3D (sensor corrected, radiometrically corrected, and ortho-rectified). The data were converted from digital numbers to top-of-atmosphere spectral reflectance using equations in and parameters 2015v2 provided by DigitalGlobe.

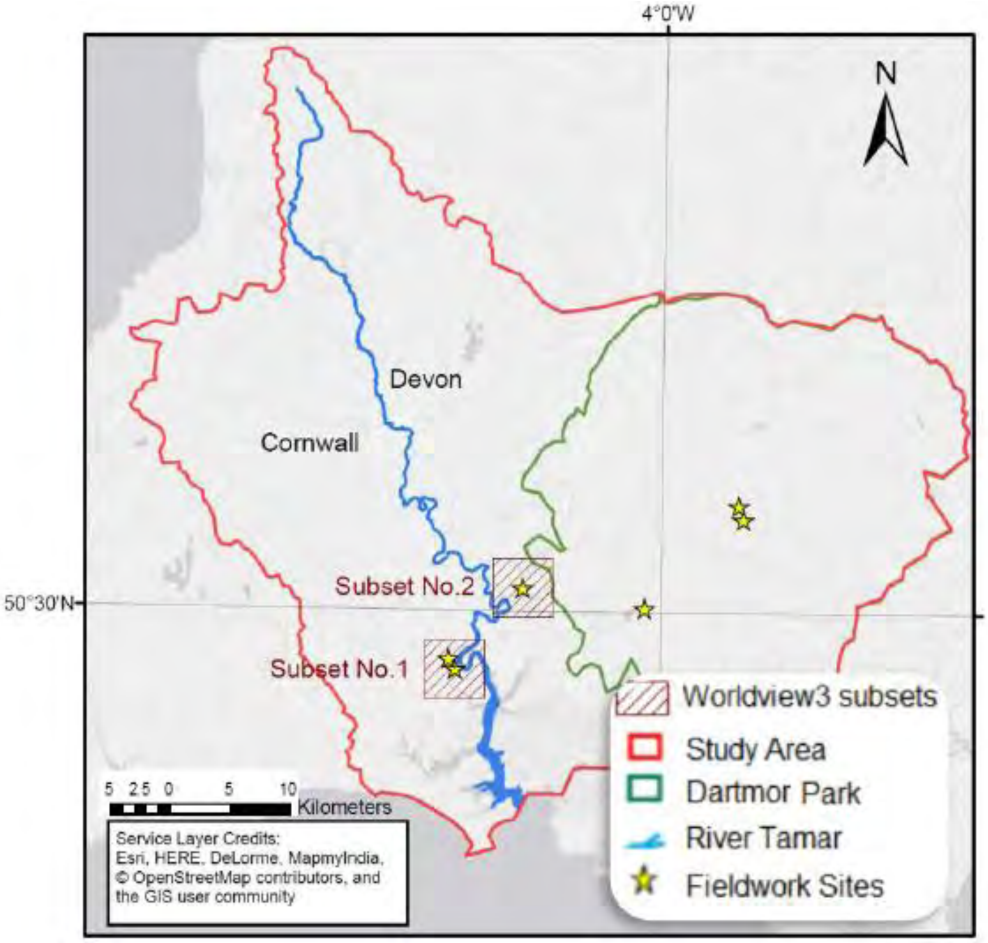


Figure 5-10: The coverage of the implemented WorldView-3 data subsets.

The atmospheric correction was then applied by using the DOS technique. As mentioned before, for this study's objectives, the DOS is satisfactory as the error analysis to be applied is linked to the non-systematic errors, while atmospheric correction deals with systematic atmospheric effect. The processing techniques were selected based on previously used optical data discussed in chapter two that includes, beside the original bands, the use of vegetation indices based on NDVI, principal component analysis and texture features. NDVI is calculated from band 5 (Red: 630-690 nm) and band 7 (Near Infrared 1: 770-895 nm), with the result shown as Figure 5-11. The first three Principal Components produced, based on the 8 MS bands, are shown in Figure 5-12.

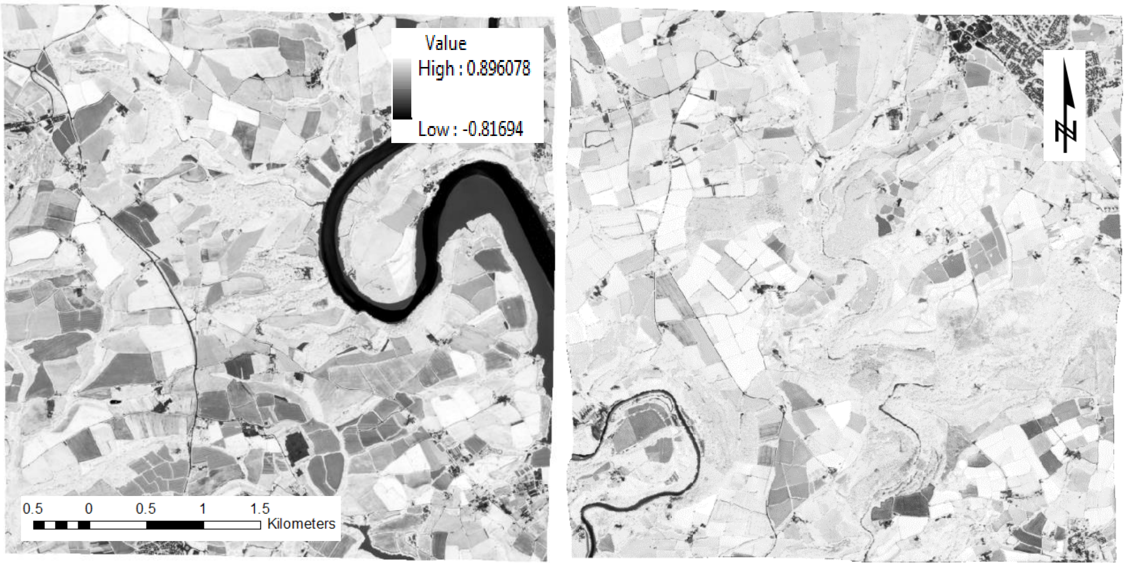


Figure 5-11: WorldView-3 based NDVI, for both subsets.

The texture Grey Level Co-occurrence Measures (GLCM) parameters were produced by using Python code that is based on GLCM texture features function provided by Python skit image library (Van der Walt et al., 2014); band 5 was used as recommended by Eckert (2012) to be the grey scale base map. Four GLCMs produced with 9x9, 11x11, 13x13 and 15x15 pixel windows, where these window sizes were chosen after initial correlation tests to be smaller than those chosen by Eckert (2012) who tests 15 × 15 to 23 × 23 pixel windows and recommended the 19x19 window due to the relatively high detailed field AGB map available for this study compared to his. Each GLCM includes six products based on the statistics of Angular Second Moment (ASM), contrast, correlation, dissimilarity, energy, and Inverse Difference Moment (IDM). Sample GLCMs are shown in Figure 5-13 for ASM, Figure 5-14 for contrast, Figure 5-15 for correlation, Figure 5-16 for dissimilarity, Figure 5-17 for energy and Figure 5-18 for IDM. These products was used later to feed the feature selection processes in Sections 6.4.1 and 6.5.1.

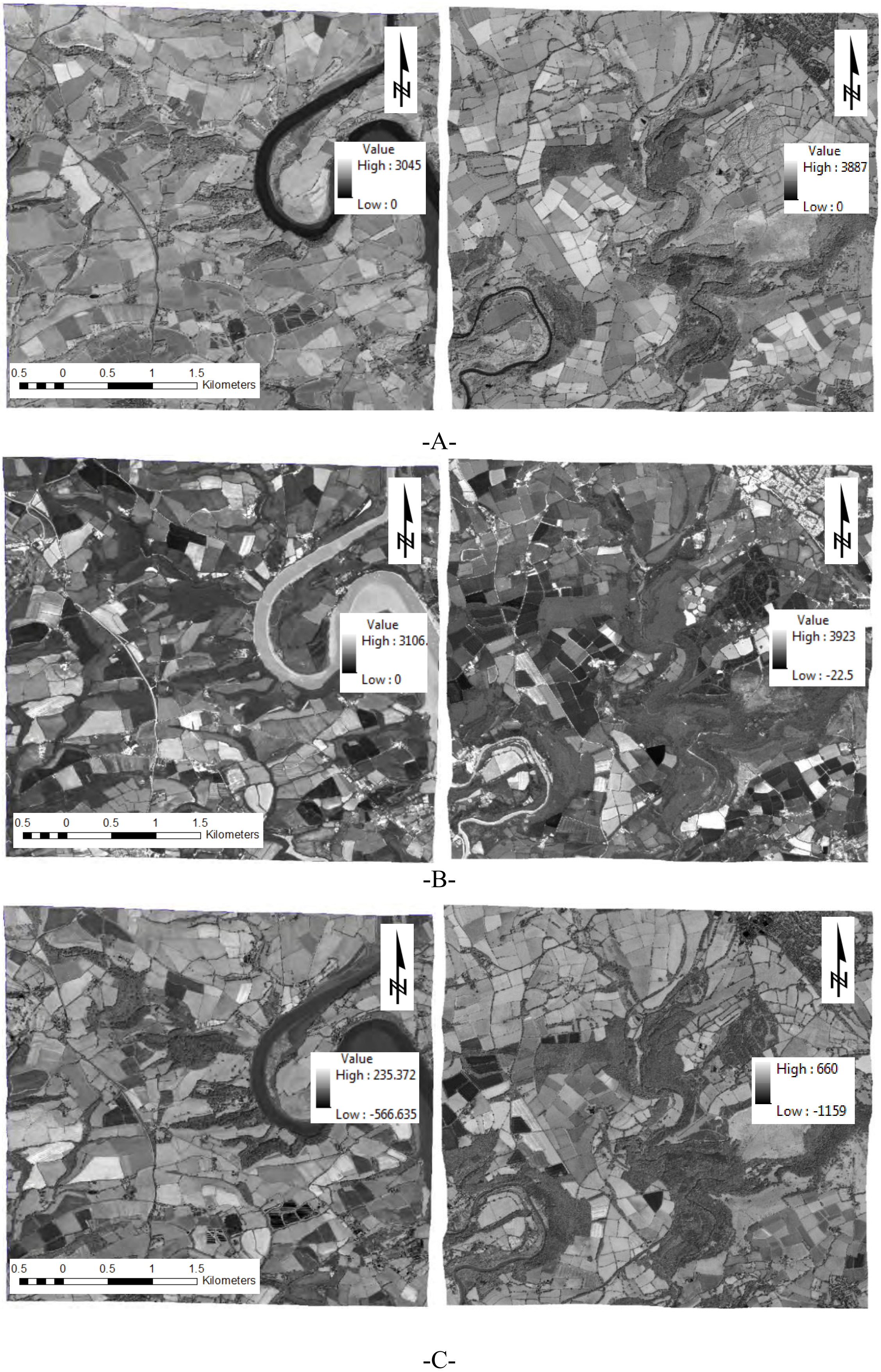
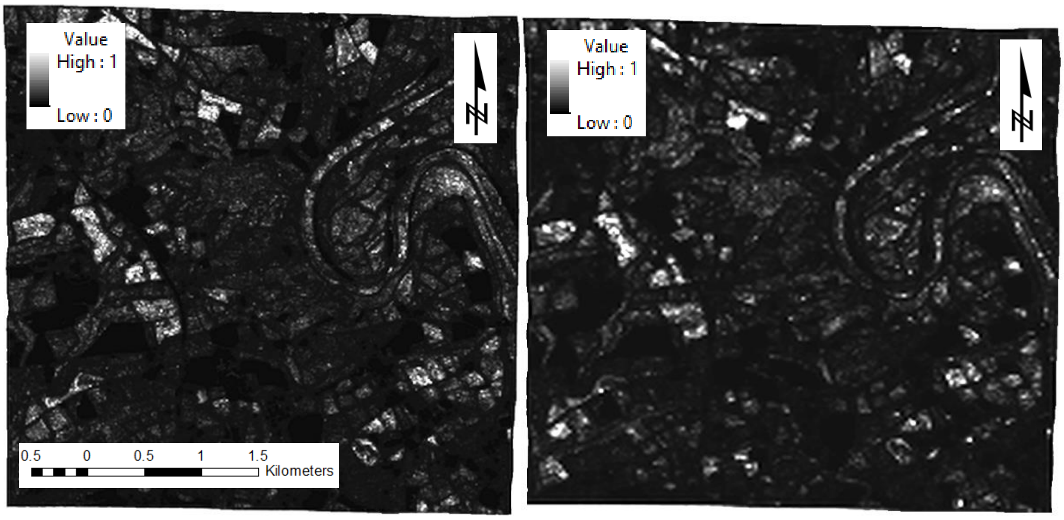
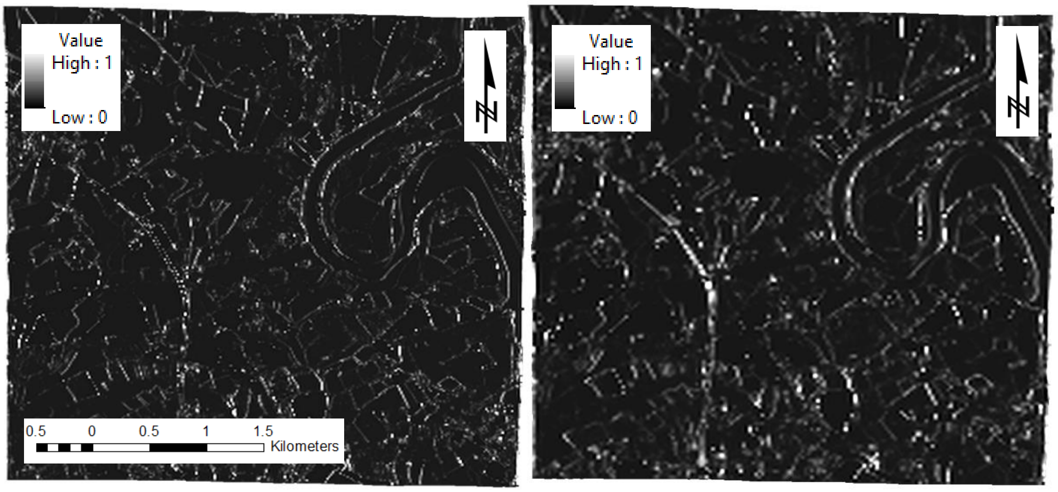


Figure 5-12: Principal components analysis for both subsets as A - PC1, B - PC2, and C - PC3.



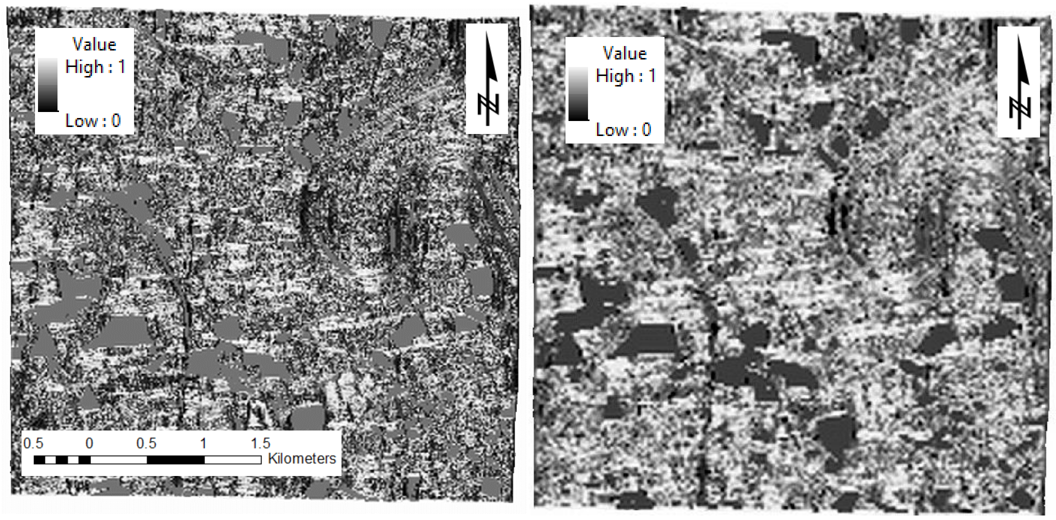
-A- -B-

Figure 5-13: Normalised ASM texture feature for subset-1 with A - 9x9 window, and B - 15x15 window.



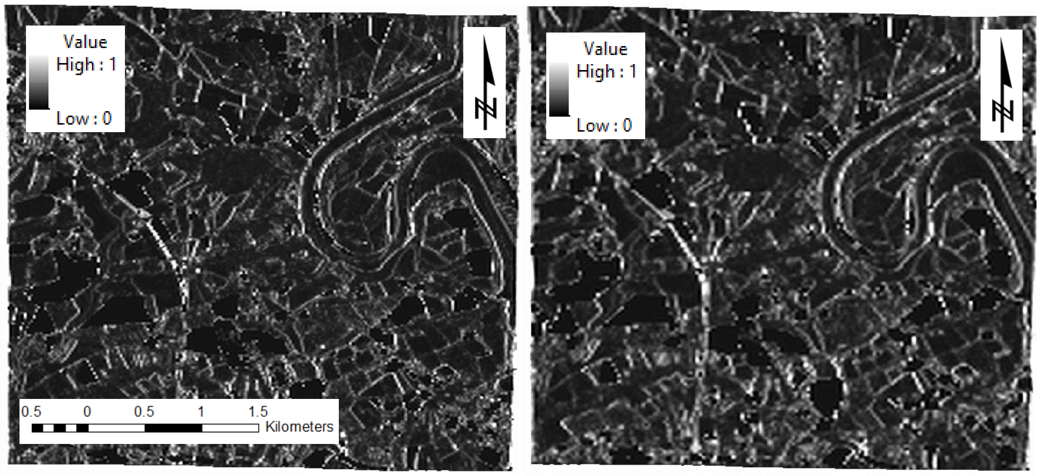
-A- -B-

Figure 5-14: Normalised contrast texture feature for subset-1 with A - 9x9 window, and B - 15x15 window.



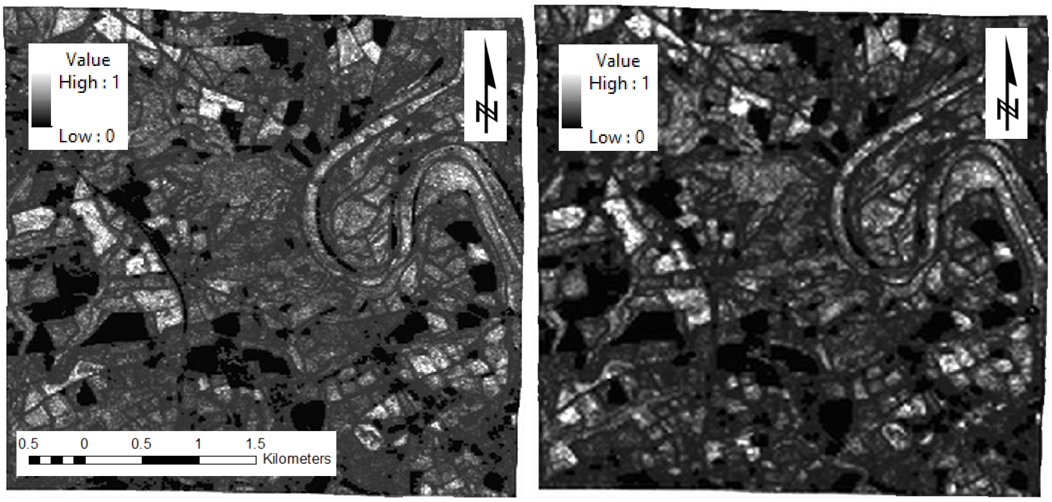
-A- -B-

Figure 5-15: Normalised correlation texture feature for subset-1 with A - 9x9 window, and B - 15x15 window.



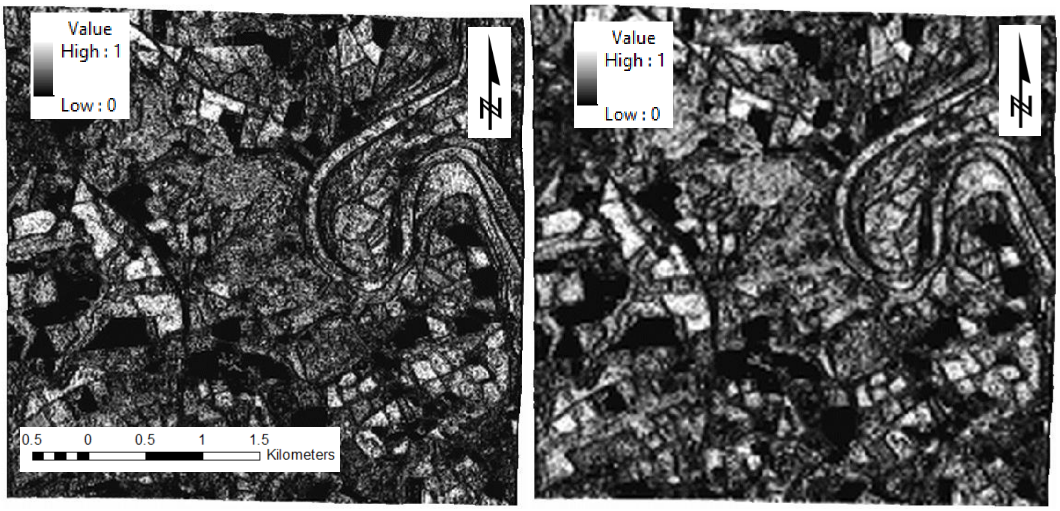
-A- -B-

Figure 5-16: Normalised dissimilarity texture feature for subset 1 with A - 9x9 window, and B - 15x15 window.



-A- -B-

Figure 5-17: Normalised energy texture feature for subset 1 with A - 9x9 window, and B - 15x15 window.



-A- -B-

Figure 5-18 Normalised IDM texture feature for subset 1 with A - 9x9 window, and B - 15x15 window.

The image classification was applied twice, the first time with a pixel-based classifier and the second time with an object-based classifier. Similar to Landsat-8 images, the data was classified into tree cover, vacant land, non-tree open area, built up area and water body. The pixel-based classification results are shown in Figure 5-19 with the percentages of user and producer accuracies for tree class, overall accuracy and Kappa coefficient being as listed in Table 5-4. The object-based classification results are shown in Figure 5-20 with the percentages of user and producer accuracies for tree class, overall accuracy and Kappa coefficient being as listed in Table 5-5.

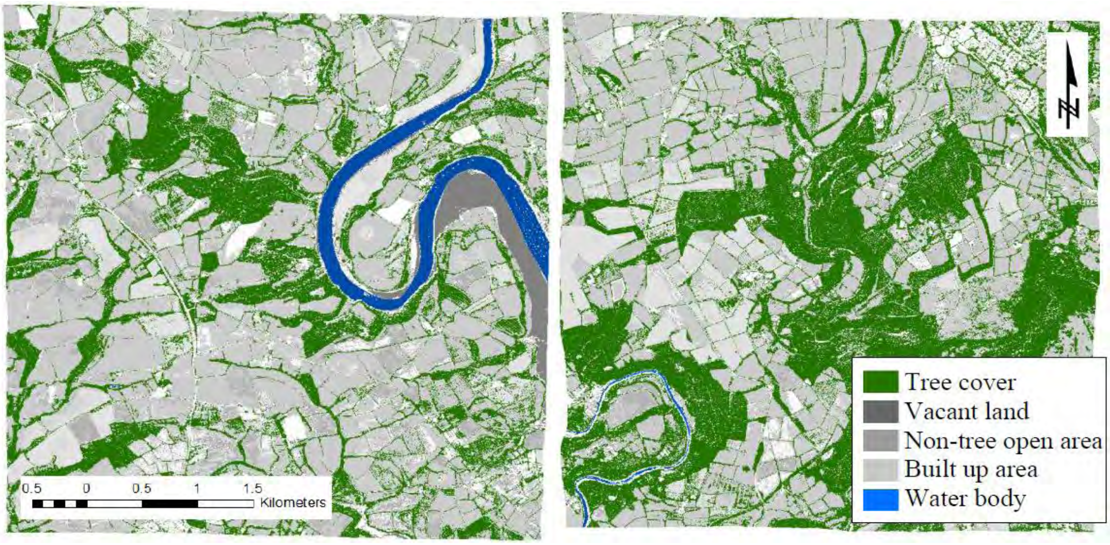


Figure 5-19: Pixel-based classification result for WorldView-3 subsets.

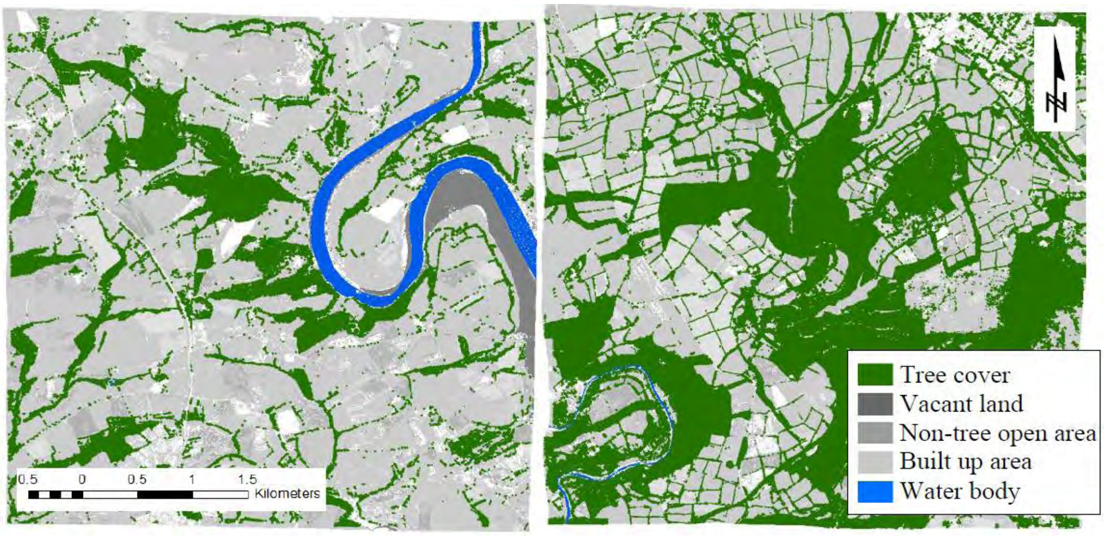


Figure 5-20: Object-based classification result for WorldView-3 subsets.

Table 5-4: Accuracy assessment for pixel-based land-cover classification based on WorldView-3 image: producers, users and overall accuracies, and the Kappa statistic.

	Tree cover	Waterbody	Vacant land (bare soil)	Non-tree green Space	Built-up area	Summation	User accuracy
Tree cover	73	0	1	3	0	77	94.8%
Waterbody	0	11	1	1	0	13	84.6%
Vacant land (bare soil)	2	1	13	2	1	19	68.4%
Non-tree green Space	6	2	2	15	0	25	60.0%
Built-up area	0	0	1	0	9	10	90.0%
Summation	81	14	18	21	10	144	
Producer accuracy	90.1%	78.6%	72.2%	71.4%	90.0%		
Overall accuracy=84.1% Kappa =0.75							

Table 5-5: Accuracy assessment for object-based land-cover classification based on WorldView-3 image: producer’s, user’s and overall accuracies, and the Kappa Statistic.

	Tree cover	Waterbody	Vacant land (bare soil)	Non-tree green Space	Built-up area	Summation	User accuracy
Tree cover	54	0	1	2	0	57	94.7%
Waterbody	0	7	1	1	0	9	77.8%
Vacant land (bare soil)	0	1	14	1	0	16	87.5%
Non-tree green Space	3	2	2	16	0	23	69.6%
Built-up area	0	0	2	0	10	12	83.3%
Summation	57	10	20	20	10	117	
Producer accuracy	94.7%	70.0%	70.0%	80.0%	100.0%		
Overall accuracy=86.3% Kappa =0.80							

5.8.2.2 Lidar Processing

The Environment Agency Geomatics Group holds a significant archive of airborne Lidar data, as a grid of small geographical blocks with different acquisition dates for each group of blocks that range from 2005 and 2016. The data used in this study were collected in the period between 2012 and 2013, and the spatial resolution is 1 m with the projection being the Ordnance Survey British National Grid (BNG). The data are available as both a Digital Terrain Model (DTM), and Digital Surface Model (DSM), so the pre-processing included the production of a Digital Canopy Model (DCM) by subtracting the provided DTM from the DSM. The DCM shown in Figure 5-21 was calculated to the same extents of the WorldView-3 data shown in Figure 5-10.

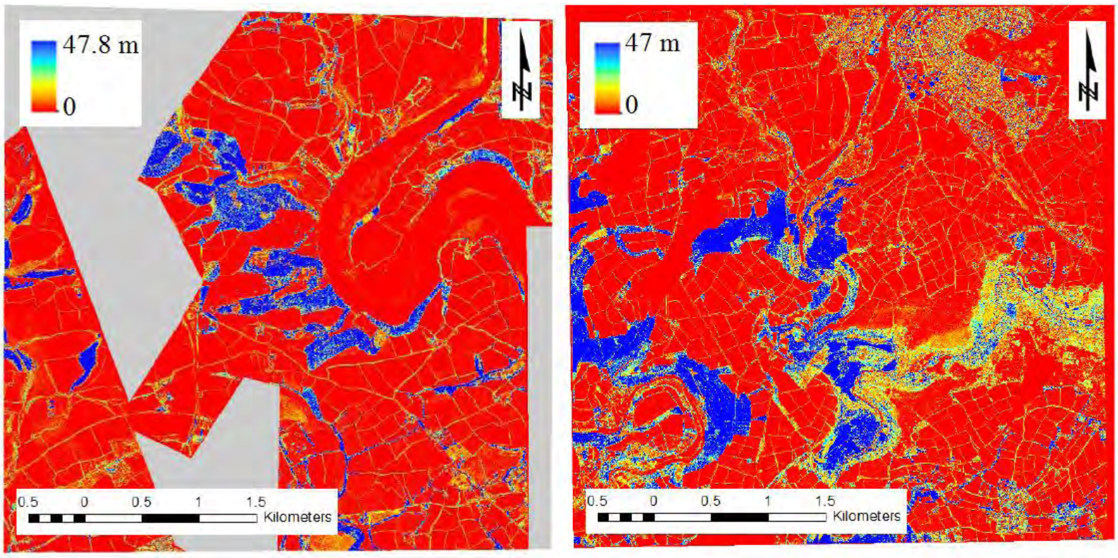


Figure 5-21: DCM for both Lidar subsets.

5.8.3 Field Data

The AGB maps derived within Chapter 4 were used as the reference AGB to produce the model. The summation of trunk and branch maps were used with high-resolution RS data scenarios, while just the branch maps were implemented to simulate the very-high-resolution data based scenarios. This decision of field maps selection was based on the best regression result achieved from the field maps for each RS data spatial resolution. The correlation between each spatial resolution type and the AGB maps will be discussed in the practical scenarios in Chapter 6.

5.8.4 Error Analysis for RS Data

5.8.4.1 Spectral Errors in Landsat-8 Products

The Landsat-8 sensor OLI has a high dynamic range of 12 bits (4096 grayscale levels) (Czapla-Myers et al., 2015) compared to 8 bits (256 grayscale levels) for Landsat-7's ETM+ (Scaramuzza et al., 2004). The increased dynamic range means a higher radiometric resolution, which improves the SNR for the OLI by an order of magnitude for typical radiance levels (Morfitt et al., 2015). The ground based tests, undertaken by Czapla-Myers et al. (2015), showed a level of noise of less than 2% for the bands used for AGB assessment.

This error can be propagated to the NDVI using the Taylor method as

$$\sigma_{NDVI} = \sqrt{(\frac{\partial NDVI}{\partial NIR})^2(\delta NIR)^2 + (\frac{\partial NDVI}{\partial R})^2(\delta R)^2} \qquad (5\text{-}4)$$

where

$\frac{\partial NDVI}{\partial NIR} = \frac{(NIR-R)+(NIR+R)}{(NIR+R)^2} = \frac{2R}{(NIR+R)^2}$, $\frac{\partial NDVI}{\partial R} = \frac{(NIR-R)-(NIR+R)}{(NIR+R)^2} = \frac{-2NIR}{(NIR+R)^2}$, $IR = \delta 0.02 * NIR$, and $\delta R = 0.02 * R$

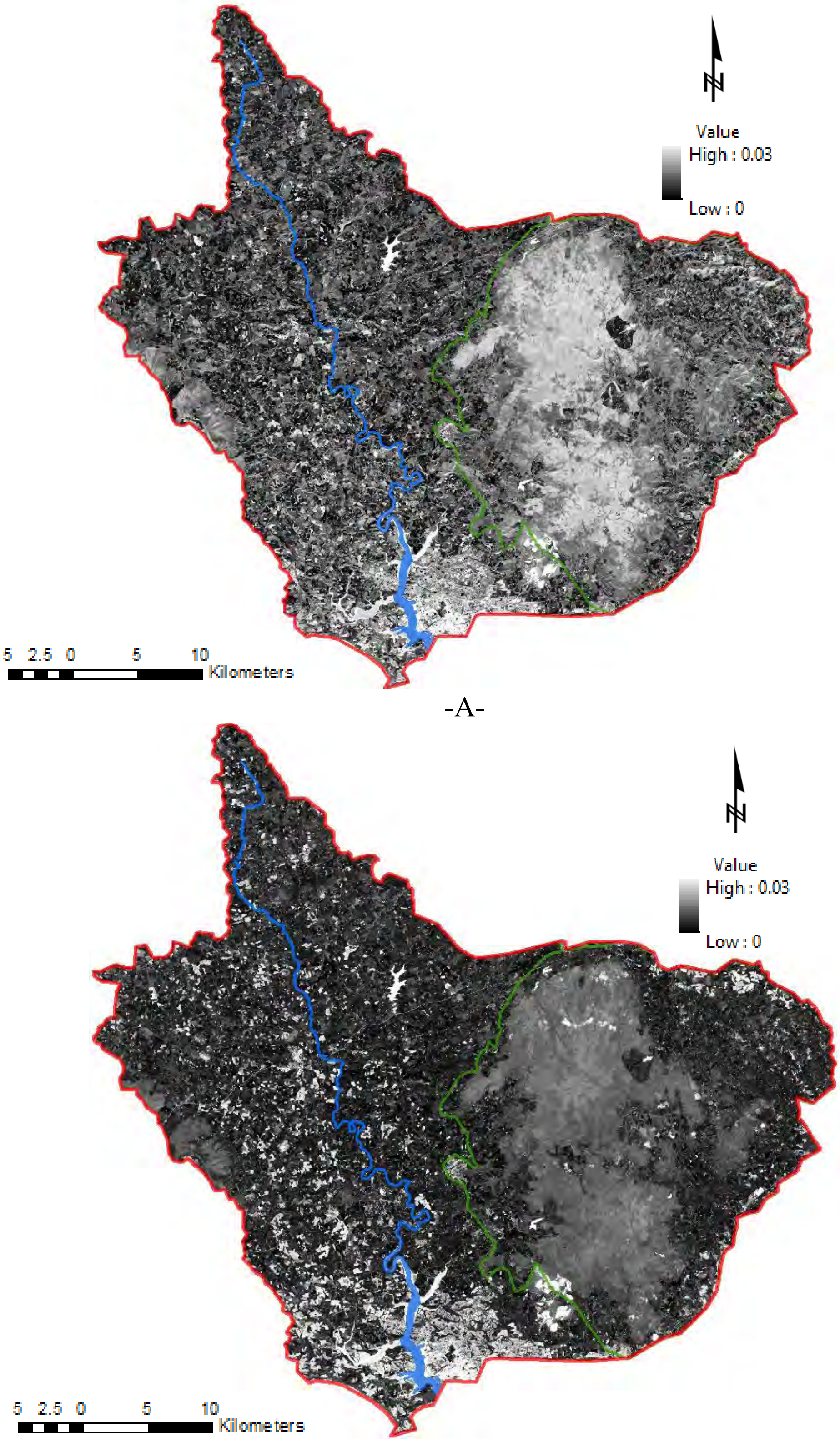
After producing the error maps for the NDVIs, the error in the NDVIs can be propagated to the NDVIR as:

$$\sigma_{NDVIR} = \sqrt{(\frac{\partial NDVIR}{\partial NDVI1})^2(\delta NDVI1)^2 + (\frac{\partial NDVI}{\partial NDVI2})^2(\delta NDVI2)^2} \qquad (5\text{-}5)$$

and therefore,

$$\sigma_{NDVIR} = \sqrt{(\delta NDVI1)^2 + (\delta NDVI2)^2} \qquad (5\text{-}6)$$

The results of spectral errors in Landsat's NDVI and NDVIR are shown in Figure 5-22 and Figure 5-23 correspondingly.



-A-

-B-

Figure 5-22: Spectral Error in Landsat-8 based NDVI Maps. A- for NDVI1, and B-for NDVI2

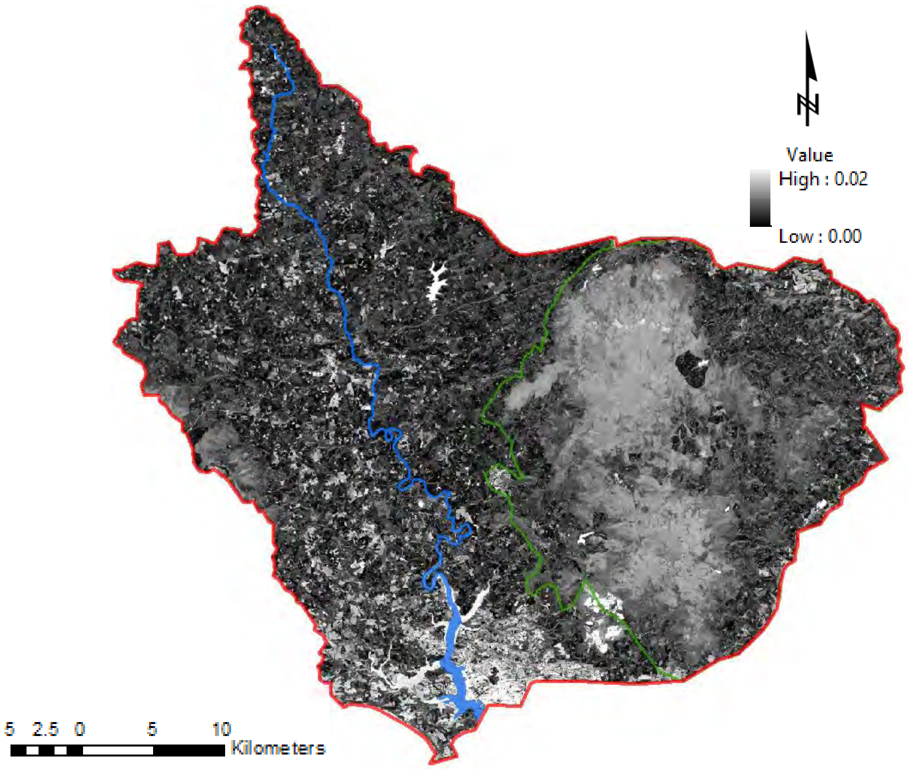
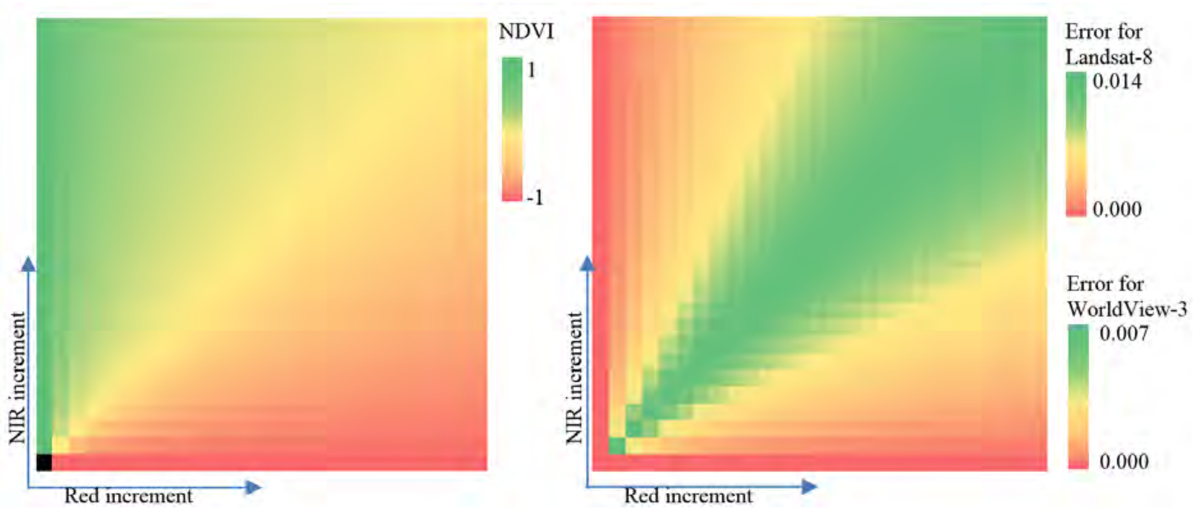


Figure 5-23: Spectral error NDVIR layer

Figure 5-22, shows high errors in the water, and in some built-up areas. In contrast, the maps show low errors for the both vegetation areas, and for the soil-covered areas where the red band has a high value compared to the NIR. Similarly, Figure 5-23, show a high errors in the water, and in some built-up areas, and low errors for vegetation and soil-covered areas.

This interesting observation was interpreted as following. The NDVI increased directly with the NIR band and inversely with the red band as shown in Figure 5-24A. While the maximum error values lie on the red=NIR diagonal axis as shown in Figure 5-24B. This non-proportional relationship between band values and band errors explains why error maps, shows high errors in the water body areas where both bands have low pixel values, and in some built-up areas where both bands have moderate pixel values. In addition, it explains why the maps also show low errors for the both vegetation areas where the NIR band is high compared to the red band, and for the soil-covered areas where the red band has a high value compared to the NIR. The NDVIR error is a resultant of two NDVI error components (equation 5-6) and therefore expected to show a similar behaviour of the NDVI non-spatial error.



-A- -B-

Figure 5-24: NDVI and NDVIs non-spatial error increment with red and near infrared bands with colour scales shows , A-the NDVI values, B- the non-spatial error values.

5.8.4.2 Height Errors in Sentinel-1

As SAR data, S1 has a number of errors such as system noise, radiometric resolution limits, orbit error, and phase error. These errors are expected to have a significant effect on the calculated heights. Unlike Landsat based products and Lidar heights, it is not practical to propagate these errors to the local height differences with error propagation techniques for two reasons. First, the height is estimated through a series of mathematically complex steps, transformations, and smoothing processes as described in Section 5-7. Second, the processing software does not provide tools for stochastic simulation of the error sources. Therefore, it was reasonable to apply approximate error estimation procedures available in the literature.

Height derivation is based on interferometric phase difference, with systematic and noise errors in the interferometric phase difference propagate to the height estimate (Richards, 2007). Systematic errors, such as baseline errors, can to be eliminated in the pre-processing stage. Therefore, the only significant error components are the phase induced random errors. These errors can be approximately calculated using Equation 5.7 provided by (Richards, 2007)

$$\delta h \approx \pm \frac{\lambda R cos\psi}{4\pi B sin(\psi+\beta)} \delta\phi \qquad (5\text{-}7)$$

where δh is the expected error in height, λ is the wavelength of the Radar system, R is the range of the SAR data, ψ is the depression angle, B is the length of the baseline, β is the baseline orientation angle with respect to horizon, and $\delta\phi$ is the phase error.

For S1, the wavelength equals 0.18m, the depression angle is 60°, and the phase error equals 5°. For the utilized dataset (image 3 and 4), the range of the SAR data equals 846662 m, the length of the baseline is 96.6 m and the baseline orientation angle equals 5.2°. Hence, the error in the height is about ±6.7 m. While for the utilized dataset (image 6 and 7), the baseline is 86.1 m, and the baseline orientation angle equals 174.8°. Hence, the error in the height is about ±7.2 m.

According to Taylor method, error in height propagates through the equation of local height differences as following

$$\sigma_{LHD} = \sqrt{(\frac{\partial LHD}{\partial H_{max}})^2 (\delta H_{max})^2 + (\frac{\partial LHD}{\partial H_{min}})^2 (\delta H_{min})^2} \qquad (5\text{-}8)$$

The errors in local maximum height δH_{max} and the error in local minimum height δH_{min} are assumed equal to the systems height error (δh). Therefore, the error in the LHD is defined as

$$\sigma_{LHD} = \sqrt{2}\,\delta h \approx \pm 10\text{m}. \qquad (5\text{-}9)$$

5.8.4.3 Spectral Error in WorldView-3

WorldView-3 images have a high dynamic range of 11 bit. However, with a higher spatial resolution and the narrower bandwidth compared to Landsat images, this type of image is expected to have a relatively higher noise level due to the trade-off between spectral resolution, spatial resolution, and Signal-to-Noise Ratio (SNR). A number of studies discussed the radiometric characteristic of WorldView-1 and WorldView-2, the previous sensors in the WorldView series, such as Krause (2008) and Poli et al. (2015). However, the radiometric characteristic studies of WorldView-3 are rare. Therefore, an image-based analysis was applied to identify the SNR, using the Homogeneous Area (HA) method discussed in Atkinson et al. (2007). Curran and Dungan (1989) showed that SNR is dependent on the land cover class and wavelength. Therefore, the geo-statistical procedure was applied to subsets of the image parts within the tree cover areas for each band. The SNR maps are visually similar, with the SNR for band 4 shown in Figure 5-25 as an example of the resulting error maps. The obtained SNR varying between 66 and 400. For the woodland areas, the SNR is around 100 for the implemented bands, which means that the spectral error is about 1%.

The NDVI error was calculated based on Equation 5.4, with the resulting NDVI error map shown in Figure 5-26. The PC error analysis was based on a 100 iterations Monte Carlo analysis by adding random noise to the image bands within the limits of the spectral errors of each band, and then calculating the PCs for each iteration. Then a statistical analysis was applied to the results of the iterations to calculate the standard deviation of each pixel. The resulting error map of the first principal component is shown in Figure 5-27.

The texture-based products are not directly correlated to the original bands through a mathematical equation. Therefore, the Taylor method was not applicable to extend this analysis to include a quantitative interpretation of the effects of the SNR to the texture based products. Instead, a 100 iterations Monte Carlo method was used to simulate the effect of the spectral error and produce error maps for the texture images. The simulation result for the ASM (with a 13x13 window) is shown in Figure 5-28, and demonstrates that the spatial variation effect on the texture products is small and has a significant value only in the flat feature areas. For the woodland areas, it is less than 0.0001 and therefore it was neglected.

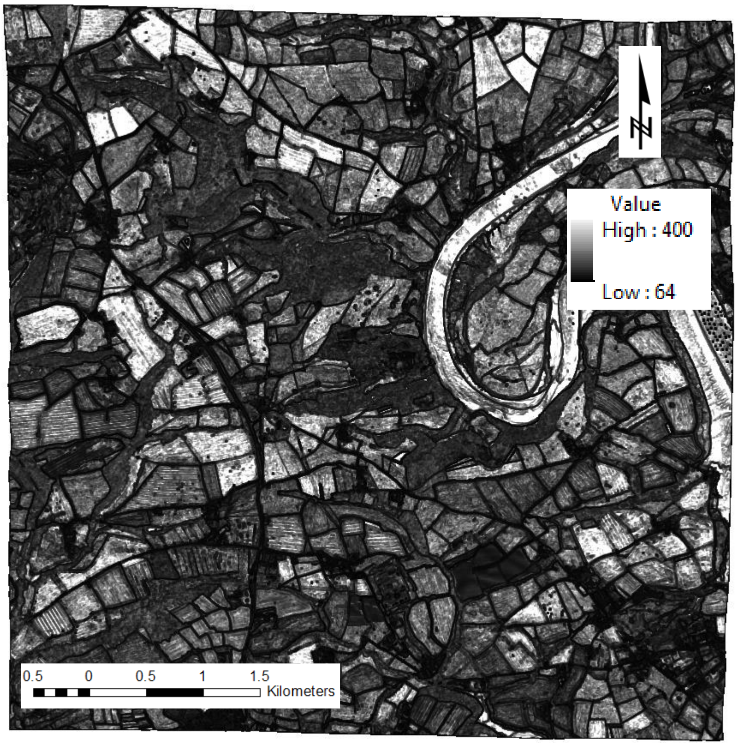


Figure 5-25: SNR of band 4 subset-1.

Figure 5-26: NDVI error.

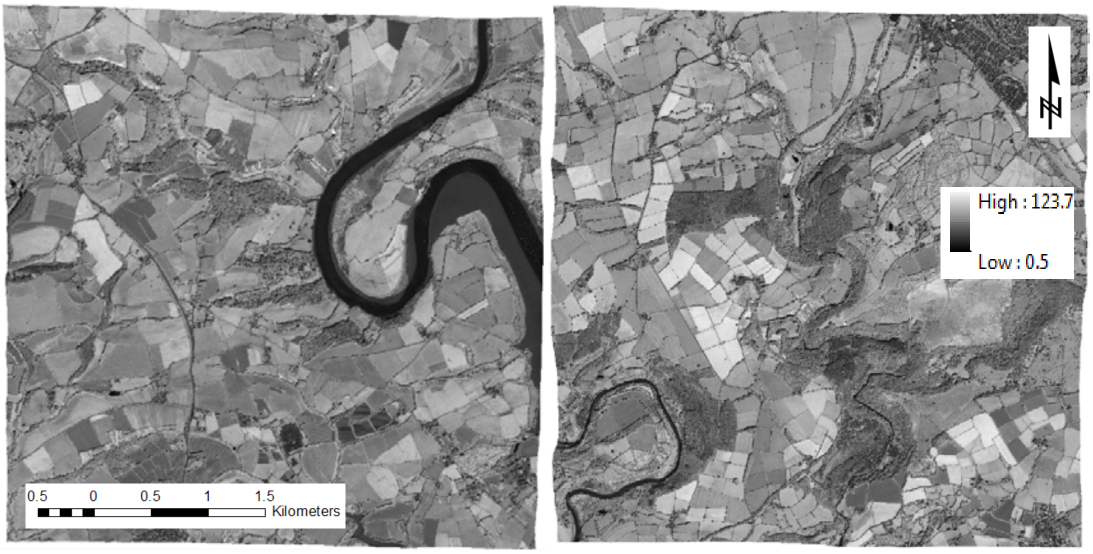


Figure 5-27: Error analysis of the first principal component.

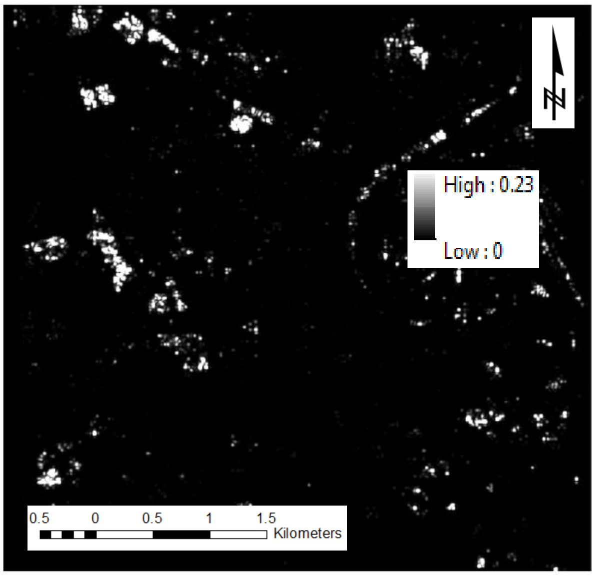


Figure 5-28: Error of texture for ASM with a 13x13 window.

5.8.4.4 Height Errors in Lidar

The elevations obtained from the Lidar DSM data ($Elev_{DSM}$) have a nominal vertical accuracy of less than ±15 cm (Agency, 2016), and the vertical accuracy of the DTM elevations ($Elev_{DTM}$) is not available. Many studies have reported a reduction in the accuracy of the DTM under tree cover. Therefore, a comparison between the TLS based DTM (derived from the field data) and the Lidar DTM was applied. The comparison was undertaken by first correcting the TLS data to ensure a constant datum; the local datum of the TLS data was converted to the global datum of the Lidar DTM by adding the algebraic difference between the averages of the DTMs to the TLS's DTM. Ten randomly selected points in each field site were then used for the comparison. The obtained RMS values were ±61 cm, ±9 cm, ±25 cm, ±28 cm, ±17 cm, and ±27 cm. The expected error in the DCM for each plot can then be calculated as:

$$\sigma_{H_{DCM}} = \sqrt{(\frac{\partial H_{DCM}}{\partial Elev_{DSM}})^2(\delta Elev_{DSM})^2 + (\frac{\partial H_{DCM}}{\partial Elev_{DTM}})^2(\delta Elev_{DTM})^2}$$

$$= \sqrt{(\delta Elev_{DSM})^2 + (\delta Elev_{DTM})^2} \qquad (5\text{-}10)$$

5.8.4.5 Spatial Errors in Landsat-8

The ground based assessment of the geolocation accuracy for Landsat-8 shows a pre-control accuracy less of than ±40 m that improves to be about ±12 m for Level 1T data (Storey et al., 2014a; Storey et al., 2014b). Tests performed for band-to-band registration accuracy show a high accuracy of better than 5 m (Storey et al., 2014a; Storey et al., 2014b).

5.8.4.6 Spatial Errors in Sentinel-1

Sentinel-1 is expected to provide high and consistent geolocation accuracy. The nominal accuracy of IW SLC data is ±7 m (Bourbigot et al., 2016). All other Sentinel-1 products shows biases, for example, Schubert et al. (2014) reports azimuth offsets of two samples from May 2014. This accuracy improved to one sample by 2015 (when the data used in this study was collected) (Schubert et al., 2015) for all products accept IW SLC that is not reported at that time. Unlike other Sentinel-1 modes, there was no estimation of the geolocation accuracy for the IW SLC product by 2015 (Schubert et al., 2015). The bias problem for IW SLC remain unsolved yet (Schubert et al., 2017). Therefore, it was reasonable to assume that the spatial accuracy is one sample (±20 m for the IW SLC data) for the datasets used in this study.

5.8.4.7 Spatial Error in WorldView-3

According to Bresnahan et al. (2016), the fieldwork data geolocation accuracy of WorldView-3 data is 3.5m, with the same study reporting a band-to-band registration error of ±3.3 m. This means that the error in each band is almost independent of the error in other bands.

5.8.4.8 Spatial Errors in Lidar

The nominal absolute spatial error for the EA Lidar data is ±40 cm (Agency, 2016), which is high compared to the resolution of 1m. The original DTM and DSM datasets were provided on the Ordnance Survey British National Grid (BNG) map projection, and so they were converted to the WGS84 datum to match the rest of the data (satellite and field data). This process is expected to reduce the horizontal accuracy of the DCM, with the conversion applied (OSGB 1936 to WGS84) having an error of up to ±2 m (EPSG, 2010).

5.8.5 Errors in the Field Data

As discussed in Chapter four, the field data were derived based on TLS. The process involved extracting the measurements of DBH and/or H, to fulfil the requirements of an allometric equation and derive the AGB. Then the calculated AGB is spatially distributed to the corresponding pixels in the AGB map according to the tree structure that is reconstructed by the algorithm described in chapter four. This process is a source of a number of errors that can be divided into three components: allometric quantity errors, spatial errors, and misdetection error.

5.8.5.1 Allometric Non-Spatial Errors

The AGB quantity error is due to errors determining DBH and H, and errors in the allometric equation parameters. It can be determined by propagating the errors of DBH, H and the allometric equation parameters through the allometric equation. This process may vary depending on the equation specifications, with the most recent allometric studies avoiding combining DBH and H in their models due to the high collinearity between them. Therefore, most studies provide equations that consist of only DBH as an independent variable, assuming that DBH can be more accurately measured and has increased correlation to AGB compared to H. However, this can be questionable when TLS is in use. Therefore, and due to the shortage in the H based allometric equations, an additional analysis was undertaken to provide allometric equations for some plots that are based on H as an independent variable.

The additional analysis involved propagating errors from the original allometric equation (AGB(DBH)) and from the conversion equation H(DBH) to the yielded equation (AGB(H)), which is key to simulating possible scenarios. The mathematical forms, shown in Table 5-6, were used to cover the tree species in the fieldwork plots, with each form requiring a slightly different calculation. However, all forms are continuously differentiable, as illustrated in Table 5-6 for DBH based equations and in Table 5-7 for H based equations. Hence, the Taylor method can be applied to propagate errors. The required information to apply the propagation are the errors inherited from the allometric equation parameters, and the errors in the tree measurements. The allometric equations are usually available with a standard error for each parameter while errors of DBH and H were derived, in chapter four, by comparing the automatically measured quantities and the corresponding manually measured ones.

Table 5-6: Allometric equation mathematical forms, and the derivatives of the parameters and measurements for DBH based equations.

		$(\partial AGB/\partial a_1)$	$(\partial AGB/\partial a_2)$	$(\partial AGB/\partial a_3)$	$(\partial AGB/\partial DBH)$
1	$AGB = a_1 DBH^{a_2}$ †	DBH^{a_2}	$a_1 DBH^{a_2} * \ln(DBH)$	-	$a_1 a_2 DBH^{a_2-1}$
2	$AGB = e^{\left(a_1+a_2\frac{DBH}{DBH+a_3}\right)}$	$e^{\left(a_1+a_2\frac{DBH}{DBH+a_3}\right)}$	$e^{\left(a_1+a_2\frac{DBH}{DBH+a_3}\right)}$ $* \frac{DBH}{DBH+a_3}$	$e^{\left(a_1+a_2\frac{DBH}{DBH+a_3}\right)} *$ $\frac{-a_2 DBH}{(DBH+a_3)^2}$	$e^{\left(a_1+a_2\frac{DBH}{DBH+a_3}\right)} *$ $\frac{-a_2 a_3}{(DBH+a_3)^2}$

† This form could be find in literature as ln (biomass) = ln (a_1) + a_2 ln (DBH).

Table 5-7: Allometric equation mathematical forms and the derivatives of the parameters and measurements for H based equations.

AGB equation	$(\partial AGB/\partial \beta_0)$	$(\partial AGB/\partial \beta_1)$	$(\partial AGB/\partial H)$
1 †	$\frac{\partial \text{AGB}}{\partial \text{DBH}} * \frac{-100H}{e^{(\beta_0+\beta_1\log(H))}}$	$\frac{\partial \text{AGB}}{\partial \text{DBH}} * \frac{-100*H*\log(H)}{e^{(\beta_0+\beta_1\log(H))}}$	$\frac{\partial \text{AGB}}{\partial \text{DBH}} * \frac{100*(1-\frac{\beta_1}{\ln(10)})}{e^{(\beta_0+\beta_1\log(H))}}$
2 ‡	$\frac{\partial \text{AGB}}{\partial \text{DBH}} * \frac{\sqrt{h}}{(1-\beta_1\sqrt{h})}$	$\frac{\partial \text{AGB}}{\partial \text{DBH}} * \frac{\beta_0 h}{(1-\beta_1\sqrt{h})^2}$	$\frac{\partial \text{AGB}}{\partial \text{DBH}} * \frac{(1-\beta_1\sqrt{h})*\frac{\beta_0}{2\sqrt{h}}+\frac{\beta_0\beta_1}{2}}{(1-\beta_1\sqrt{h})^2}$

† where $\frac{\partial AGB}{\partial a_1},\frac{\partial AGB}{\partial a_2},\frac{\partial AGB}{\partial a_3}$ *are similar to the corresponding derivatives in* Table 5-6 *with* $DBH = \frac{100*H}{e^{\beta_0+\beta_1\log(H)}}$

‡ where h=H-1.3 and $\frac{\partial AGB}{\partial a_1},\frac{\partial AGB}{\partial a_2},\frac{\partial AGB}{\partial a_3}$*, and* $\frac{\partial AGB}{\partial DBH}$ *are similar to the corresponding derivatives in* Table 5-6 *with* $DBH = \frac{\beta_0\sqrt{h}}{(1-\beta_1\sqrt{h})}$

According to (Ahmed et al., 2013) there is another type of error related to the geographic reliability of the equation when plots are sited outside the region for which the equation was originally developed for. For example, if an equation was developed to be typical for North-America, and it is to be applied in Europe, then this error can be significant. Ahmed et al. (2013) suggested that this error could be added to the error of the allometric equation as a third component, beside the parameter error and the measurement error. In the absence of an accurate non-destructive methodology for estimating the value of this component, the best solution is to have the error ranging between zero (when the equation is based on many studies) to duplicating the error of other components (when the equation based on only one site). This study applied a species-specific equation for plots where there is only one species and a general equation for the mixed plots.

The first order Taylor methods approximate the error of AGB (σ_{AGB}) calculated from the parameters ($a_i \ldots a_m$), which have errors of ($\delta a_1 \ldots \delta a_m$) correspondingly and measurements ($x_i \ldots x_n$) which have errors of ($\delta x_1 \ldots \delta x_m$) correspondingly by using the equation:

$$\sigma_{AGB} = \sqrt{\sum_{i=1}^{m} (\frac{\partial AGB}{\partial a_i})^2 (\delta a_i)^2 + \sum_{i=1}^{n} (\frac{\partial AGB}{\partial a_i})^2 (\delta a_i)^2} \qquad (5\text{-}11)$$

When this equation is extended to include the geographic reliability error ($\delta_{G.R.E.}$) as a component it becomes:

$$\sigma_{AGB} = \sqrt{\sum_{i=1}^{m} (\frac{\partial AGB}{\partial a_i})^2 (\delta a_i)^2 + \sum_{i=1}^{n} (\frac{\partial AGB}{\partial x_i})^2 (\delta a_i)^2 + (\delta_{G.R.E.})^2} \qquad (5\text{-}12)$$

The equation conversion process, from DBH based equations to H based equations, was based on Nasund's equation that is adopted by (Muukkonen, 2007) and for the Scots-

Pine, and Birch and on Hein and Spiecker (2008) for Ash. The original form of Nasund's equation defines H as a function of DBH as in Equation 5.13:

$$H=1.3+\frac{DBH^2}{(\beta_0+\beta_1.DBH)^2} \qquad (5\text{-}13)$$

Then it was converted to another form, in which DBH can be defined as a function of H to replace each DBH in the DBH based equations. The conversion involves the following process:

$$H-1.3=\frac{DBH^2}{(\beta_0+\beta_1.DBH)^2} \qquad (5\text{-}14)$$

by taking the square root of the equation and substituting h=H-1.3:

$$\sqrt{h}=\frac{DBH}{(\beta_0+\beta_1 DBH)}$$

Using cross multiplication:

$$\beta_0\sqrt{h}+\beta_1\sqrt{h}\ DBH=DBH$$

Which means that:

$$DBH-(\beta_1\sqrt{h})DBH=\beta_0\sqrt{h}$$

$$DBH(1-\beta_1\sqrt{h})=\beta_0\sqrt{h}$$

and the final equation could be:

$$DBH=\frac{\beta_0\sqrt{h}}{(1-\beta_1\sqrt{h})} \qquad (5\text{-}15)$$

While Hein and Spiecker (2008) provides the DBH as a function of H for Ash trees as in Equation 5.16.

$$\text{DBH}=\frac{100*\text{H}}{e^{\beta_0+\beta_1\log(\text{H})}} \tag{5-16}$$

The conversion parameters β_0 and β_1 are species-specific. However, the samples used for estimation were located outside the geographic region of the study area. Therefore, this might increase the effect of the geographic reliability error ($\delta_{G.R.E.}$ for the H based equations. Species-specific equation forms, parameters and parameters standard errors used for the field data analysis are listed in Table 5-8.

The error analysis for trunk AGB are shown in Figure 5-29, Figure 5-30, and Figure 5-31 for the tree species used in this study with both DBH and H based equations and Figure 5-32 for species that were explored with DBH based equations only. The error values for sample trees of each species were correlated to the corresponding AGB values of the analysed trees. The correlation equation was used to estimate errors for other trees of that species. The error component used for the analysis was based on the linear approximation of the relationship between AGB and allometric equation error. Similarly, this error analysis was applied with branch allometric equations. Examples of the analysis results for the quantity errors of the branch AGB is illustrated in Figure 5-33. The total quantity error maps were produced by the root square of the summation the squares of the error components. This root square represents the physical resultant of the error components.

Table 5-8: Species-specific equation forms, parameters and parameters standard errors used for the field data analysis.

Species		Equation	a_1	SE	a_2	SE	a_3	SE	β_0	SE	β_1	SE
Ash	stem	1	0.17 (Alberti et al., 2005) *0.71(Cai et al., 2013)	0.067*0.71	2.46	0.129	-	-	5.358 (Hein and Spiecker, 2008)	0.069	0.672	0.038
	branch	1	0.17*0.16	0.067*0.16	2.46	0.129	-	-	5.358	0.069	0.672	0.038
Birch	stem	2	-2.411 (Muukkonen, 2007)	0.204	10.210	0.182	8.291	0.736	1.460 (Muukkonen, 2007)	0.019	0.184	0.001
	branch	2	-3.579	0.299	0.570	0.350	11.363	1.728	1.460	0.019	0.184	0.001
Scots pine	stem	2	-1.408 (Muukkonen, 2007)	0.155	10.666	0.151	15.775	1.137	2.082(Muukkonen, 2007)	0.020	0.170	0.001
	branch	2	-0.928	0.141	9.889	0.523	32.338	4.556	2.082	0.020	0.170	0.001
Beech	stem	1	0.159 (Chakraborty et al., 2016)	0.021	2.346	0.062	-	-	-	-	-	-
	branch	1	0.233	0.056	1.781	0.113	-	-	-	-	-	-
Oak	stem	1	$e^{-2.181}$ (Forrester et al., 2017)	$e^{-2.181} * 0.231$	2.269	0.033	-	-	-	-	-	-
	branch	1	$e^{-2.986}$(Forrester et al., 2017)	$e^{-2.986} * 0.196$	2.309	0.061	-	-	-	-	-	-
Mixed Species	stem	1	$e^{-2.527}$(Forrester et al., 2017)	$e^{-2.527} * 0.029$	2.414	0.005	-	-	-	-	-	-
	branch	1	$e^{-3.723}$(Forrester et al., 2017)	$e^{-3.723} * 0.196$	2.330	0.008	-	-	-	-	-	-

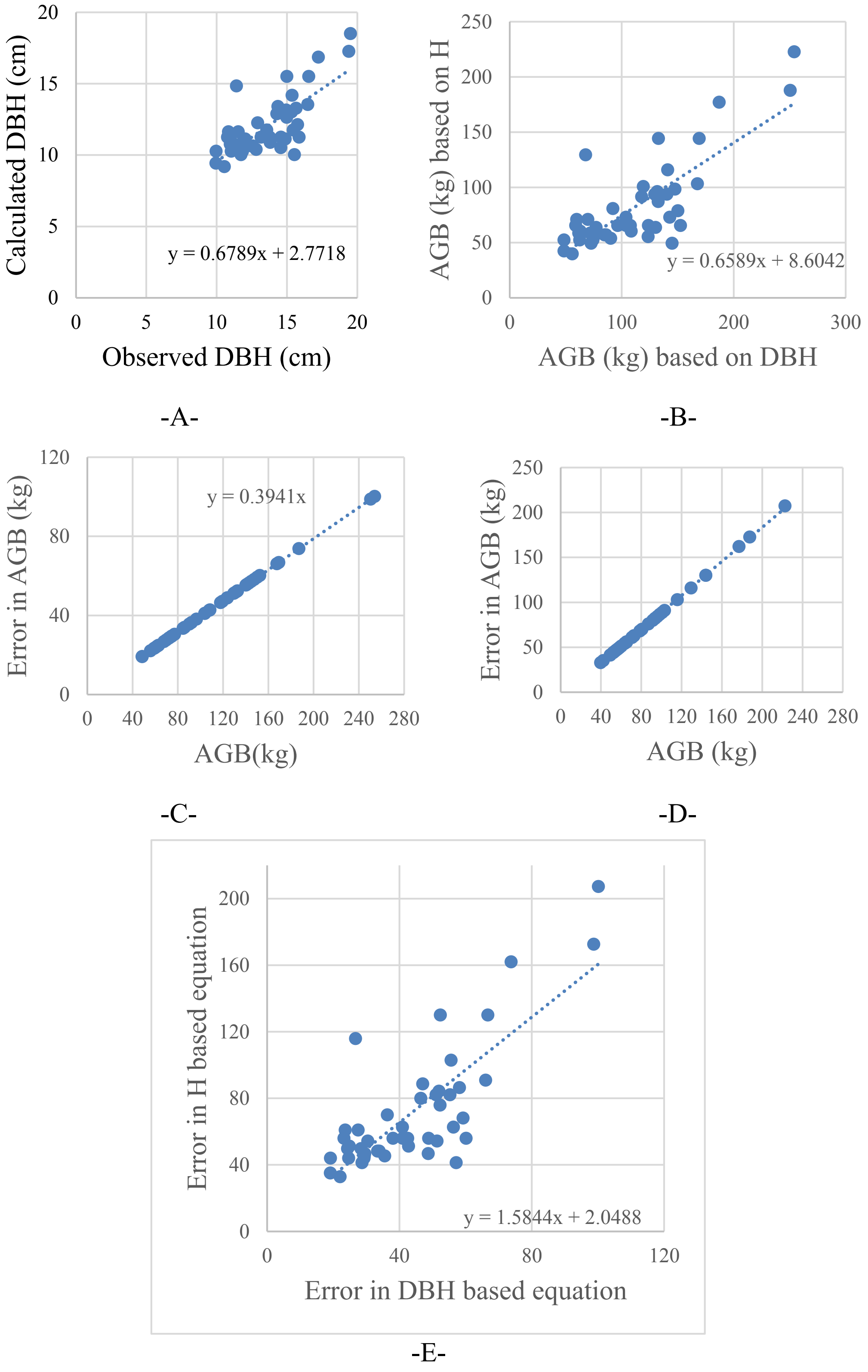
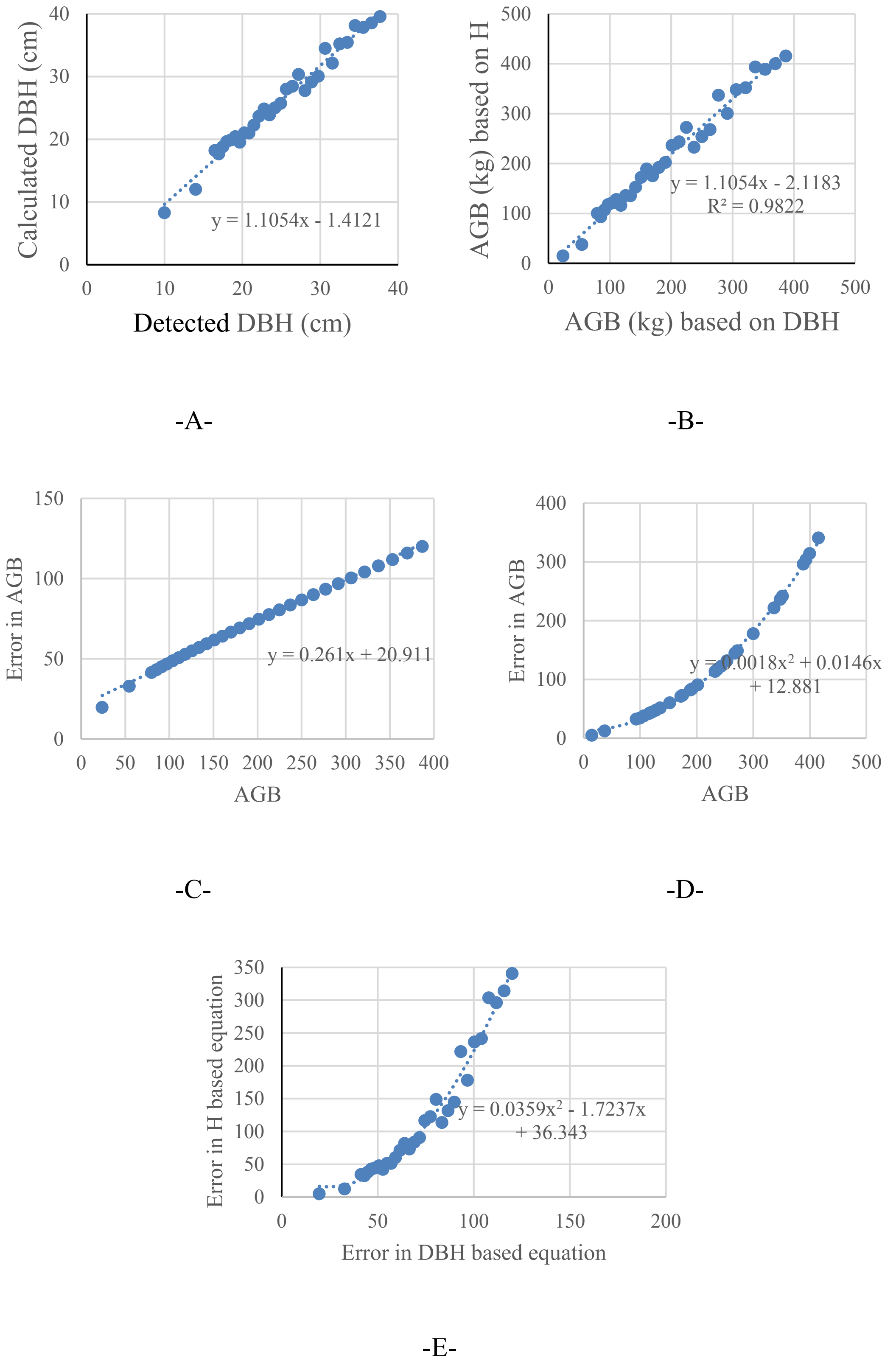


Figure 5-29: Error analysis results for Ash trees, A - DBH (Calculated vs observed DBH), B - AGB (H based vs DBH based), C - Error of the DBH based equation, D - Error of the H based equation, and E - Error of H based equation vs error of DBH based equation.



-A-

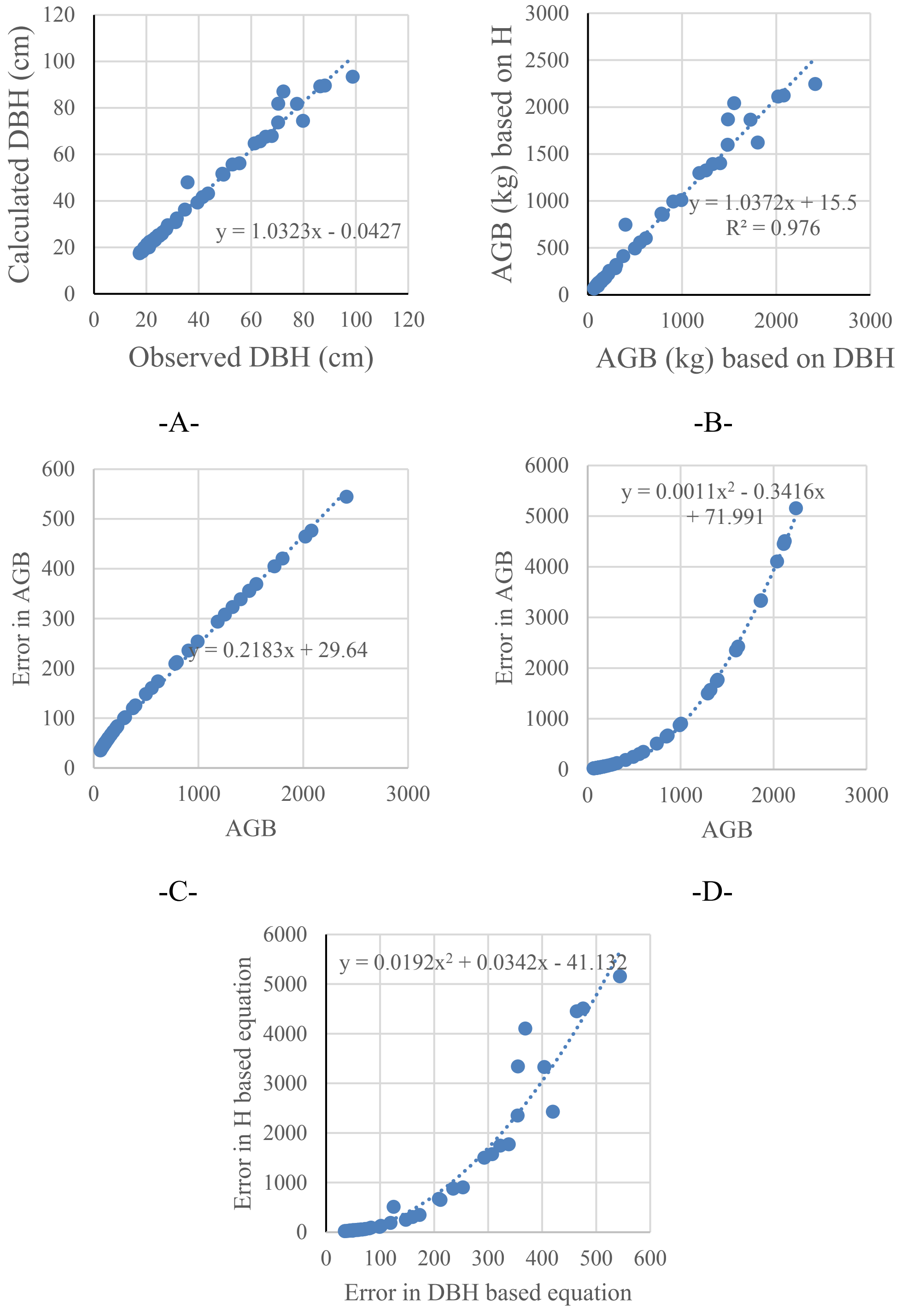
-B-

-C-

-D-

-E-

Figure 5-30: Error analysis results for Birch trees as A - DBH (Calculated vs observed DBH), B - AGB (H based vs DBH based), C - Error of the DBH based equation, D - Error of the H based equation, and E - Error of H based equation vs error of DBH based equation.



-A-

-B-

-C-

-D-

-E-

Figure 5-31: Total AGB error analysis results for Scot-Pine trees as A - DBH (Calculated vs observed DBH), B - AGB (H based vs DBH based), C - Error of the DBH based equation, D - Error of the H based equation, and E - Error of H based equation vs error of DBH based equation.

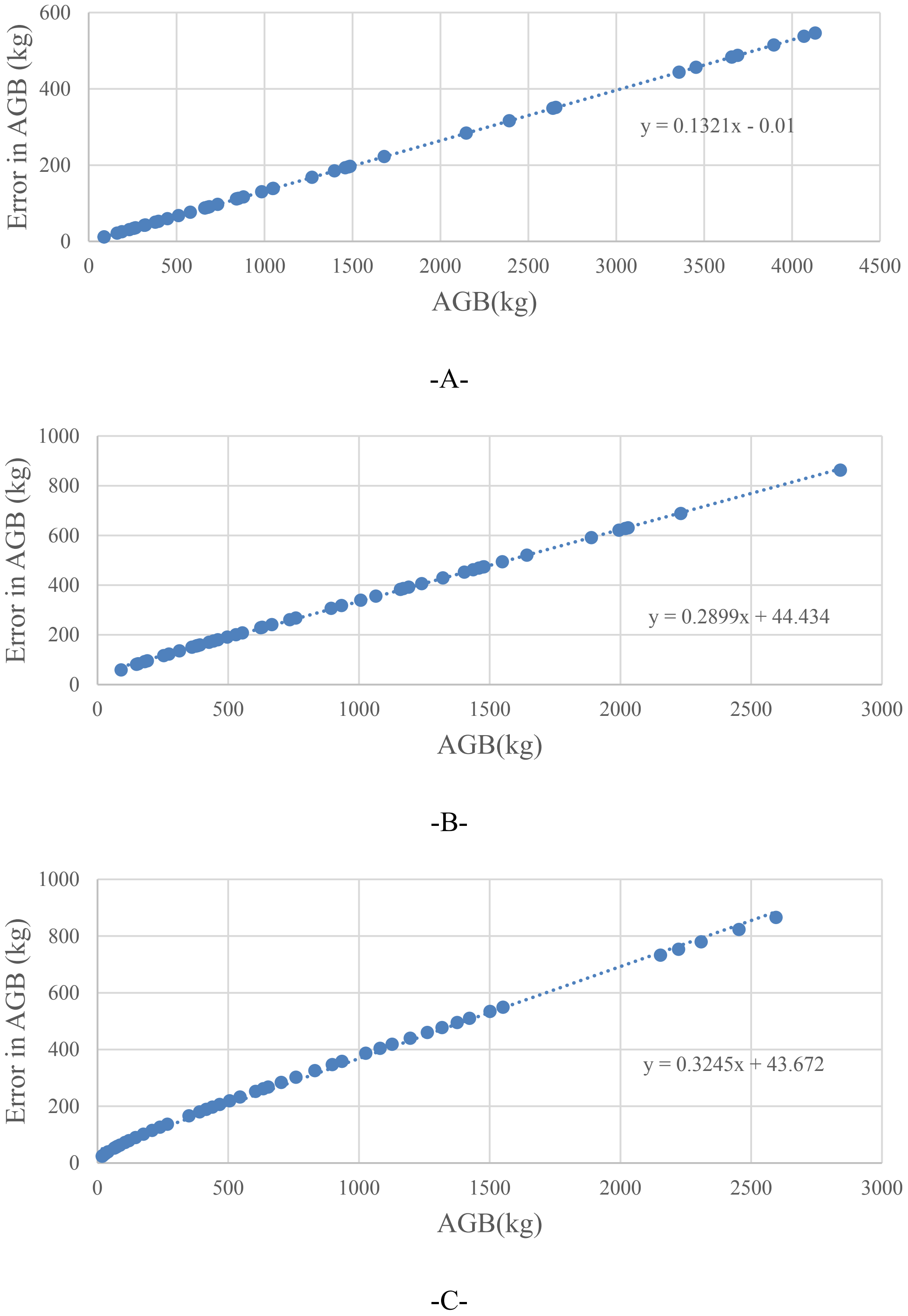
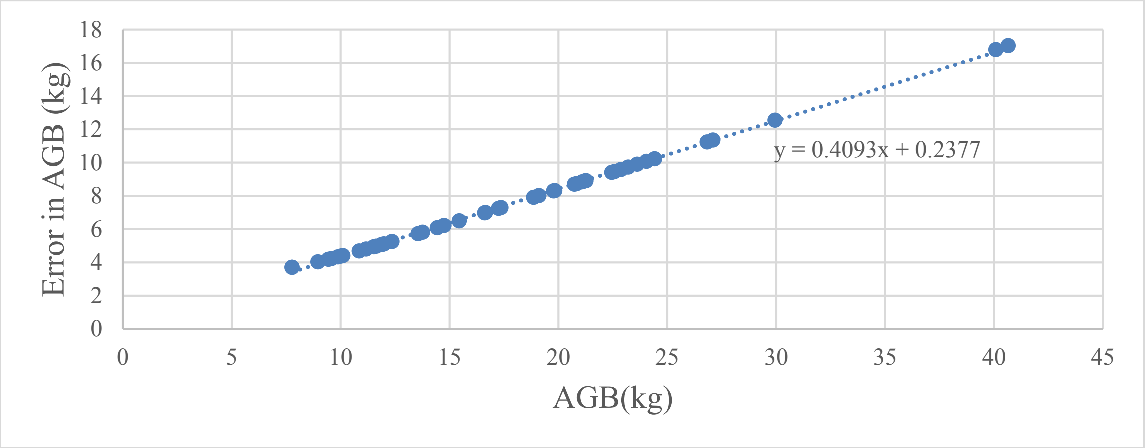
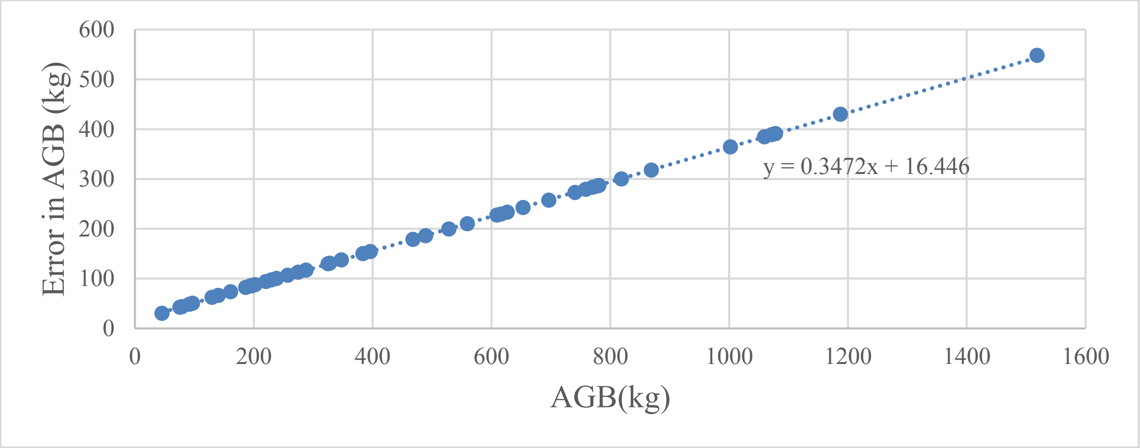


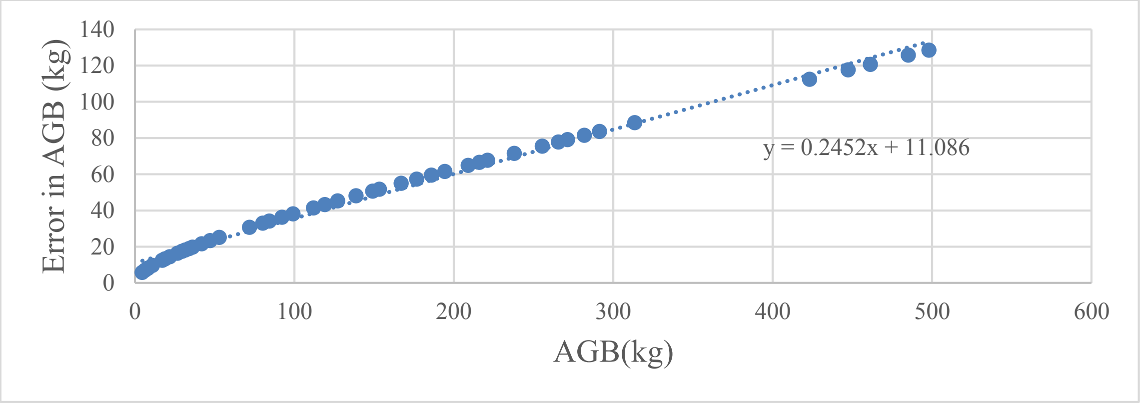
Figure 5-32: Total AGB error analysis yields from the DBH based allometric equations for A - Beech trees, B - Oak trees, and C - for mixed trees.



-A-



-B-



-C-

Figure 5-33: Branches AGB error analysis yields from the DBH based allometric equations for A - Ash trees, B- Oak trees, and C- mixed trees.

5.8.5.2 Spatial Errors

The spatial component includes errors in the geo-referencing of the AGB map, and errors in the AGB spatial distribution. The field data were geo-referenced by using GPS measurements of the position of the scanner for some scan stations in each field site. The accuracy of the hand-held Trimble Juno 3B GPS with Satellite-based Augmentation System (SBAS) correction is about ±5 m after differential correction. This accuracy can be directly provided to the error analysis software tool.

The accuracy of biomass distribution is limited by number of factors including the shadows in the TLS cloud, the tree structure approximation, and the statistical weighting of the branches. The determination of this type of error, with the spatial resolution provided within the fieldwork data analysis, is not practically possible even with destructive methods due to the fact that it is not possible to find the weight of parts of the tree parts that are located within specific spatial limits. However, the high resolution of the biomass maps makes it possible to reduce the effect of this error significantly by averaging all pixels that lie within each pixel of the remote sensing image, when pixel-based analysis is used, or for each object when object-based used. Therefore, this error is negligible for the available RS data resolutions and, as a result, the spatial errors were limited to the geo-location error.

5.8.5.3 Misdetection Errors

This error exists due to the false positives and false negatives in the tree detection stage. This error is for an approximate value of one average tree and with random spatial coordinates, meaning that this error is statistically random. The effect of false positives would be equivalent to adding an extra trunk to the trunk AGB map, and changing the classification of the close branches to be these false positive branches. This means that the biomass of a whole tree is increased and the distribution of the biomass of some trees are changed. The effect of false negatives is classifying the undetected tree as branches of the closest trees, which means missing the biomass of the whole tree and distributing it other trees instead. In both cases, this error was approximated to be the biomass of one tree per miss-detection probability with a DBH that equals the average DBH of detected trees in the objected plot.

5.8.5.4 Wood Density Variability

The conversion from above-ground volume into AGB was avoided due to the use of direct AGB allometric equations. Theoretically, this helped to avoid the need to use the wood density. Though, the indirect effect of the high spatial and temporal variability of above ground biomass is still unavoidable (Svob, et al., 2014). However, the variability of wood density affects the field samples that were used to derive the allometric equation. As a result, this type of error is an included component in the error assessment of any AGB allometric equation (Berger et al., 2014). Therefore, and due to the fact that allometric-equation error was taken into consideration in this study, the variability of wood density is expected to be implicitly included in the error propagation.

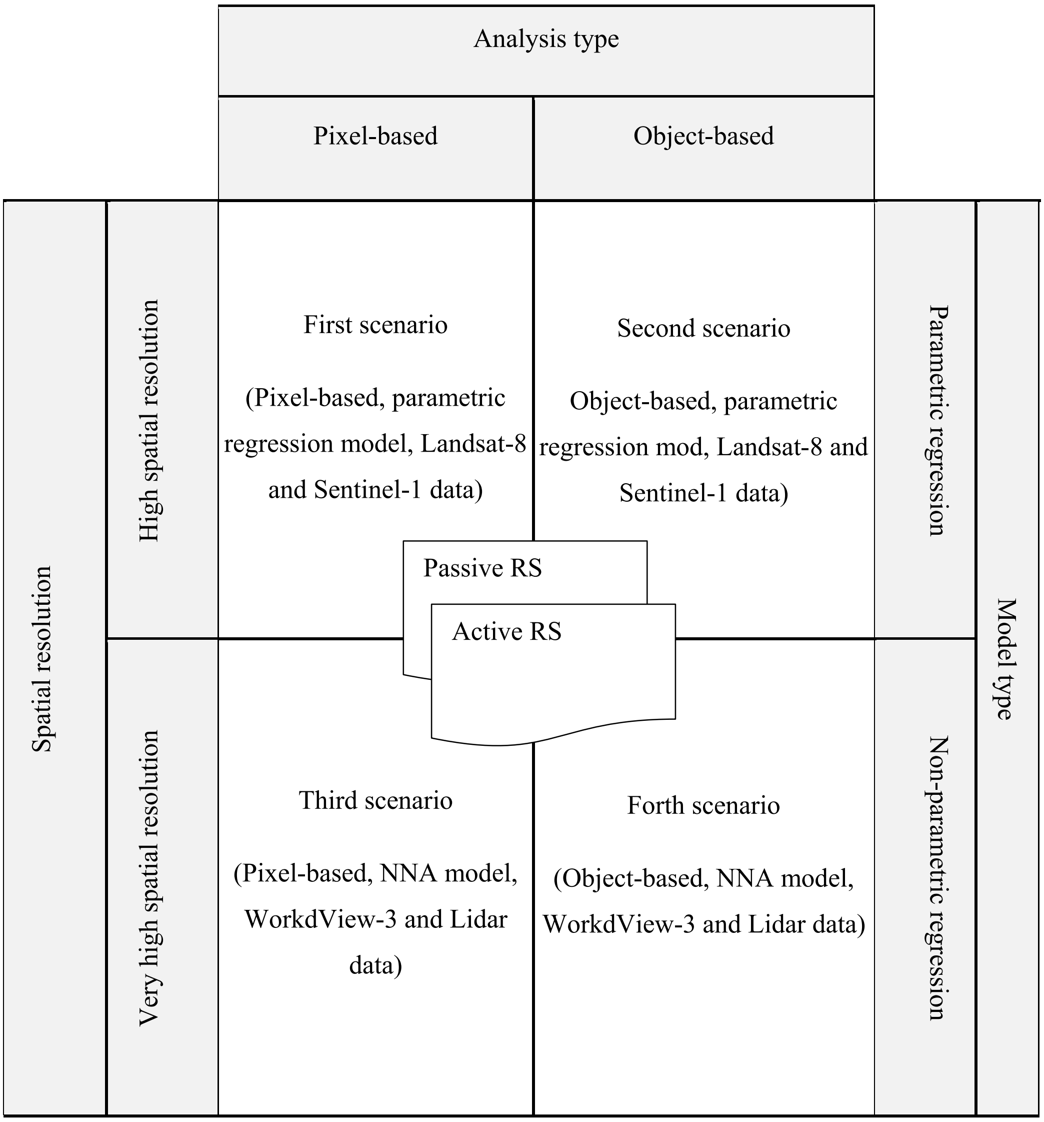
Chapter 6 Practical scenarios

6.1 Introduction

This chapter aims to analyse the propagation of remote sensing and fieldwork data errors to the AGB. This quantitative analysis was based on the error calculations discussed in Chapter 5. Four practical scenarios were designed to represent the variety of possibilities for AGB assessment systems. The rational was first to provide a better understanding of the error behaviour in a different AGB assessment system and second to prove that the software tool can manage a wide range of possible systems. As discussed in Chapter 2, the AGB assessment system possibilities can be related to the input data, the processing type, and the applied mathematical model.

Due to the wide variety of possibilities, and taking the limited time frame of the study, the scenario design was based on only main classification categories. The data related parameters are the spatial resolution (high or very high spatial resolution) and the RS system type (active or passive). The analysis type is either pixel-based or object-based. While the system type is either parametric or nonparametric. Accordingly, four scenarios were designed. The first scenario used pixel-based analysis of high spatial resolution active and passive data, with parametric regression. The second scenario used object-based analysis of high spatial resolution active and passive data, with parametric regression. The third scenario used pixel-based analysis of very high spatial resolution active and passive data, with non-parametric regression. And the fourth scenario used pixel-based analysis of very high spatial resolution active and passive data, with non-parametric regression. Therefore, the four scenarios were designed to cover the possibilities as shown in Table 6-1.

Table 6-1: The scenarios coverage to the possibilities of remote sensing data, model type, and processing type.



		Analysis type			
		Pixel-based	Object-based		
Spatial resolution	High spatial resolution	First scenario (Pixel-based, parametric regression model, Landsat-8 and Sentinel-1 data)	Second scenario Object-based, parametric regression mod, Landsat-8 and Sentinel-1 data)	Parametric regression	Model type
	Very high spatial resolution	Third scenario (Pixel-based, NNA model, WorkdView-3 and Lidar data)	Forth scenario (Object-based, NNA model, WorkdView-3 and Lidar data)	Non-parametric regression	

6.2 First Scenario

The first scenario applies a multiple linear regression as a parametric based mathematical model using the fieldwork information, derived in chapter four, combined with high-resolution active and passive RS data. A pixel-based approach was used to perform this regression, as described in the following sections.

6.2.1 Model Design

Two-stage analysis was applied to design the regression model. The first stage is a primary ANOVA that is inclusive to all inputs, and the analysis tests the relationship strength between inputs, including Landsat-8 bands 2, 3, 4, and 5, NDVI, NDVIR, and LHD with the fieldwork biomass. The Landsat bands were selected based on the results of Lu, et al. (2002); Gjertsen (2007); Melaas et al. (2013); Berra et al. (2014); and Dube and Mutanga (2015) that shows that amongst Landsat sensors bands, only these bands equivalents showed significant correlation for AGB assessments. NDVIR was derived from NDVIs on two different seasons as described in Section 5.8.1.1 and therefore is expected to relate to the tree type and AGB due to the fact that each tree type has a distinct annual cycles. LHD was used to provide active remote sensing data and hence the advantages of this type of data as discussed before in Section 3.2.1.2. All inputs are normalized to be in the range 0 to 1, and the ANOVA results are shown in Table 6-2.

Table 6-2: Model parameters and statistical P-values for the first scenario.

Primary ANOVA results	RS product	Coefficients	P-value
	Landsat B2	0.1075	0.566
	Landsat B3	1.6831	0.089
	Landsat B4	-2.8880	0.005
	Landsat B5	1.4143	0.012
	NDVI	0.4037	0.002
	NDVIR	-1.0384	0.004
	LHD	0.4112	0.111
Final ANOVA results	RS product	Coefficients	P-value
	Landsat B4	0.6667	0.001
	NDVI	0.2777	0.003
	NDVIR	0.0857	0.187

The second stage applies a final ANOVA test to the inputs that have P-values equal to or less than the P-value of band 4, and these inputs are then used to train the model. The P-value of band 4 is used as a criterion to assess the input P-values due to the fact that the red wavelength Landsat-8 band has been reported to have a moderately inverse relationship with AGB (Walker et al., 2010, Dube and Mutanga, 2015; Wu et al., 2016). The rationale of using this reduction in the number of model inputs is to avoid over-fitting, which can result when there is a small number of samples. The ANOVA results show that only three inputs (Landsat NDVI, RNDVI, and band 4) have P-values less than or equal to the Landsat-8 band-4 P-value of 0.005, and they were used to train a linear regression model. The model ANOVA results are listed in Table 6-2, with it having an R^2 of 0.62.

6.2.2 Biomass Mapping

The linear regression model was used to map the AGB. The modelled AGB for each pixel of the fieldwork areas are compared to the corresponding AGB of the fieldwork data as shown in Figure 6-1. The analysis showed that the residual level is high with an average of 56 t/ha and RMS of 66 t/ha. The resulted AGB map (density on each pixel) is shown in Figure 6-2. The map resolution is equivalent to the Landsat-8 resolution which is 30 m.

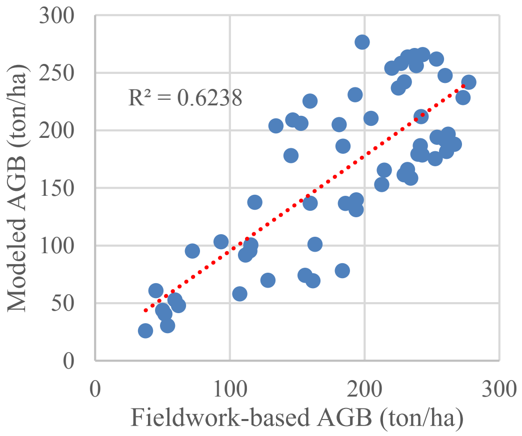


Figure 6-1: Modelled AGB vs fieldwork based AGB for the first scenario.

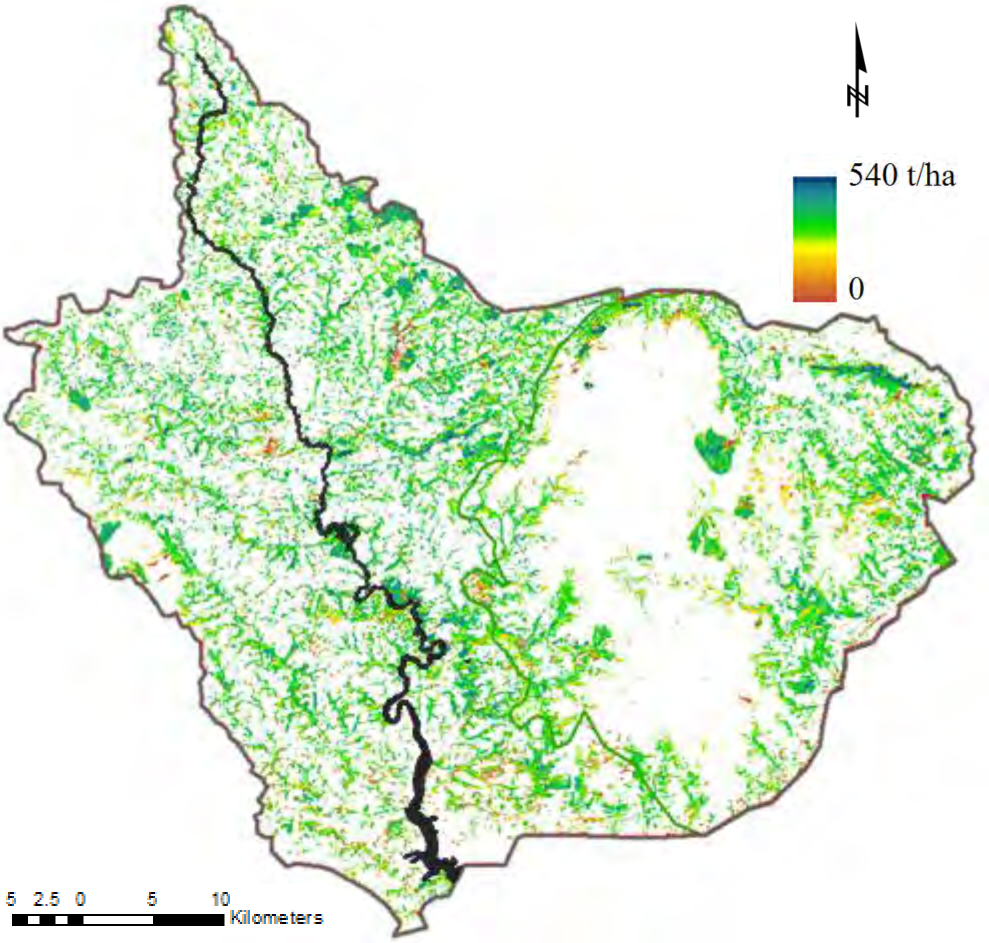


Figure 6-2: AGB map resulting from the first scenario.

6.2.3 Error Analysis

The error analysis tool was implemented with seven settings, with the:

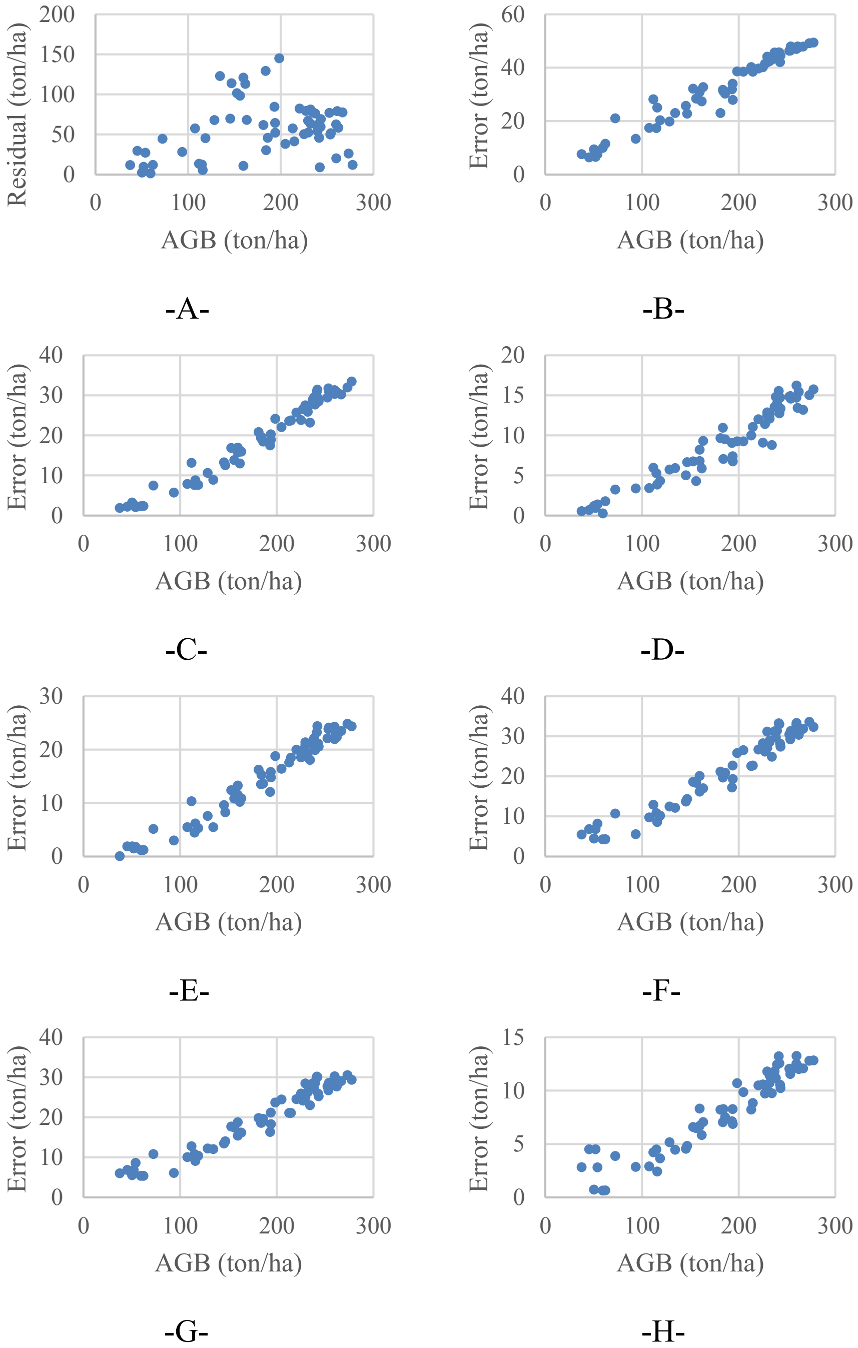
- First setting including all the spatial and spectral field data errors and the RS data spatial and spectral errors.
- Second setting including the spatial and spectral errors for the field data only.
- Third and fourth settings separating the field data spatial errors from spectral errors.
- Fifth setting including the spatial and spectral errors of RS data only.
- Sixth and seventh settings separating the RS data spatial errors from spectral errors.

Each setting was applied with 1000 iterations. The execution time consumed to apply the software tool ranged between 5 to 7 minutes for each setting. The regression parameters standard errors are listed in Table 6-3, with the uncertainty components of the field plot pixels of each analysis is shown in Figure 6-3.

Table 6-3: The influence of errors on the model coefficients.

	coefficient	SE regression (overall coefficient error)	SE all errors	SE due to field data errors influence			SE due to RS data errors influence		
				All	Spatial	Spectral	All	Spatial	Spectral
Landsat B4	- 0.1667	0.065	0.047	0.029	0.010	0.019	0.032	0.018	0.026
NDVI	0.2777	0.092	0.069	0.033	0.026	0.002	0.045	0.041	0.033
NDVIR	0.0857	0.058	0.032	0.024	0.008	0.022	0.019	0.017	0.012

The highest absolute residuals (Figure 6-3 A) lie in the areas of AGB around the 180 metric-tonne/Hectare (t/ha), with the areas of low AGB showing a lower average residual density within the study area, with the AGB map (density on each pixel) shown in Figure 6-2. The error propagation results show that the errors propagated from the RS and field data have an obvious trend that is proportional to the AGB. The combined RS and field data error (Figure 6-3 B) starts low for low AGB pixels and increases to reach a level similar to the residuals for the high AGB pixels. The average of these errors is about 33 t/ha that is significantly lower than the average of the residuals. The RS errors (Figure 6-3 C, D and E) shows that the average RS data error is about 19 t/ha. The spectral and spatial components of the RS data have averages of about 14 and 9 t/ha, respectively, with the average field error being around 21 t/ha. The non-spatial and spatial error components have averages of 19 and 9 t/ha, respectively. The software tool used these error analyses to model the errors all over the study area.



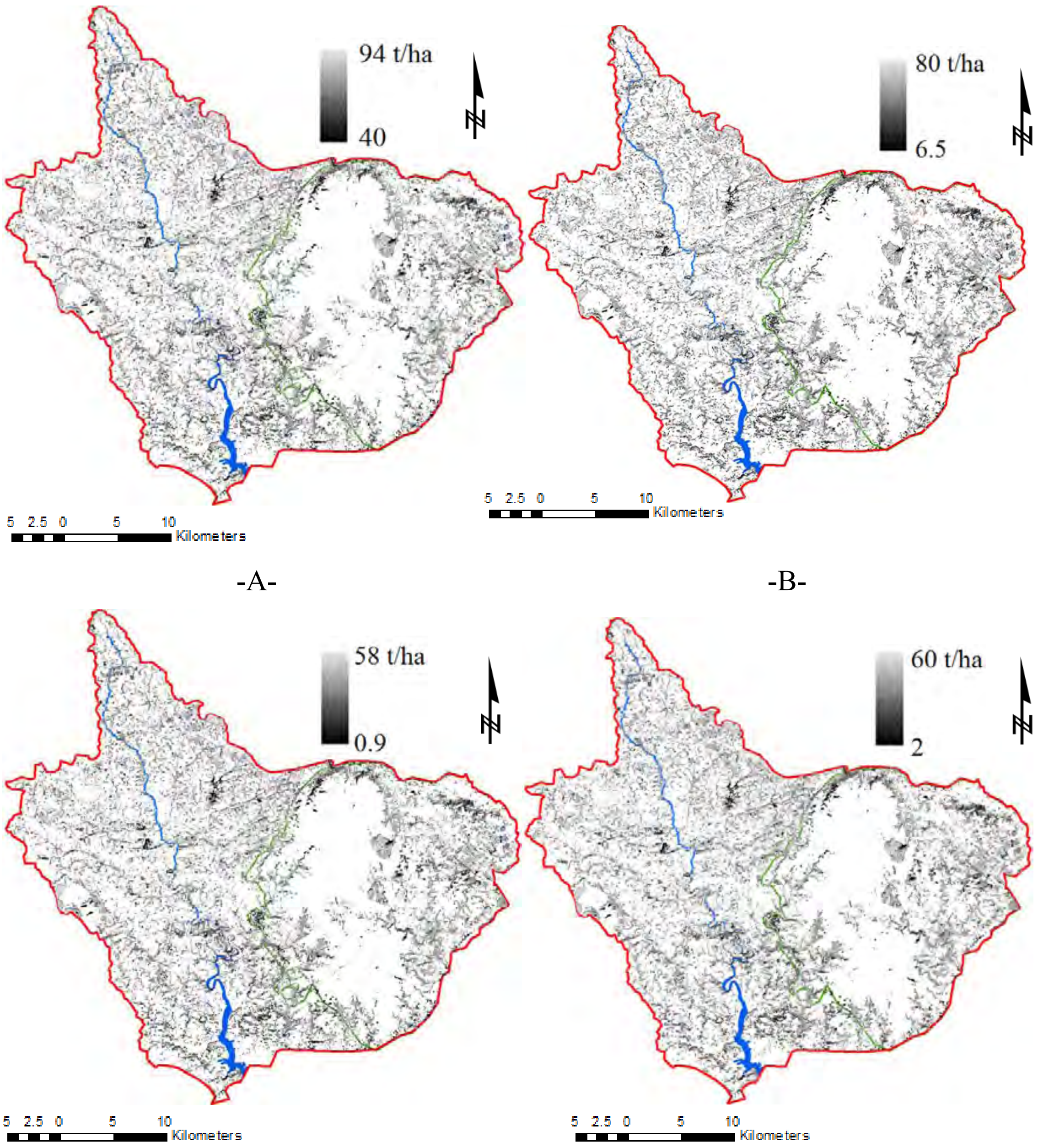
-A- -B-

-C- -D-

-E- -F-

-G- -H-

Figure 6-3: The uncertainty components for the first scenario as: A - residuals of AGB fitting vs AGB, B - standard deviation yields from applying all RS and field errors vs AGB, C - standard deviation yields from applying all RS errors vs AGB, D - standard deviation yields from applying spectral RS vs AGB, E - standard deviation yields from applying spatial RS vs AGB, F - standard deviation yields from applying all field errors vs AGB, G- standard deviation yields from applying spectral field vs AGB, and H - standard deviation yields from applying spatial field vs AGB.



-A- -B-

-C- -D-

Figure 6-4: Error maps for the first scenario as the A – fitting error, B - uncertainty component that is related to RS and field errors, C - uncertainty component that is related to RS errors only, and D - uncertainty component that is related to field errors only.

The correlation between residual error and the modelled error was on an R^2 of 0.33. The correlation between simulated RS and field errors, and the errors mapping model results, was high with R^2s between 0.82 and 0.94. The error model result was the error maps shown in Figure 6-4. The maps represent the error map for the main settings including fitting residual errors, the uncertainty component that is related to the combined RS and field errors, the uncertainty component that is related to RS errors only, and the uncertainty component that is related to field errors only.

6.3 Second Scenario

The same RS data that were explored for the first scenario were explored for the second scenario that included Landsat-8 images, NDVIs and RNDVI alongside Sentinel-1 LHD. The same field data were used to provide the reference AGB data and, in addition, the fieldwork data used for classification in the first scenario was also used in this scenario.

6.3.1 Model Design

That same procedure used in the first scenario was applied in this scenario. However, an object-based correspondence was used instead of the pixel-based correspondence between field and RS data. The average of each layer's pixels for each objects polygon was applied. For example, if an object polygon covers 10 pixels of some layer the average of the 10 pixels was used as a value for the object. Therefore, the number of resulted fieldwork samples was fewer compared to the first scenario; each sample segment consists of a number of pixels instead of one pixel samples in the first scenario. Primary ANOVA yields in results that are shown in Table 6-4. Accordingly, the final model ANOVA test was applied on Landsat-8's B4, NDVI, and Sentinel-1's LHD. The model ANOVA results are listed in Table 6-4, which has an overall R^2 of 0.84.

Table 6-4: Model parameters and statistical P-values.

Primary ANOVA results	***RS product***	***Coefficients***	***P-value***
	Landsat B2	-0.2946	0.359
	Landsat B3	-0.3338	0.605
	Landsat B4	-0.7655	0.107
	Landsat B5	0.6791	0.181
	NDVI	0.5918	0.003
	NDVIR	-0.1131	0.605
	LHD	0.3283	0.0670
Final ANOVA results	***RS product***	***Coefficients***	***P-value***
	Landsat B4	-0.2816	0.160
	NDVI	0.3997	0.001
	LHD	0.1452	0.094

6.3.2 Biomass Mapping

The resulting linear regression model was used to map the AGB density within the study area. The modelled AGB for each object of the fieldwork areas are compared to the corresponding AGB of the fieldwork data as shown in Figure 6-5. The analysis showed that the average of the absolute residual is with an average of 21 t/ha and RMS of 36 t/ha. The resulted AGB map (density on each pixel) is shown in Figure 6-6. The map is object-based. Hence, each polygon has an AGB average.

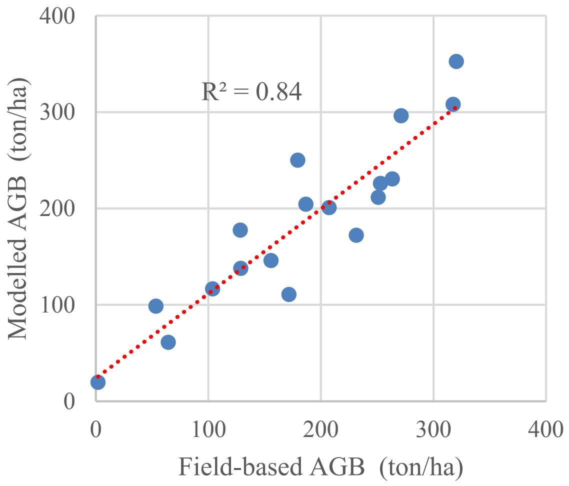


Figure 6-5: Modelled AGB vs fieldwork based AGB for the second scenario.

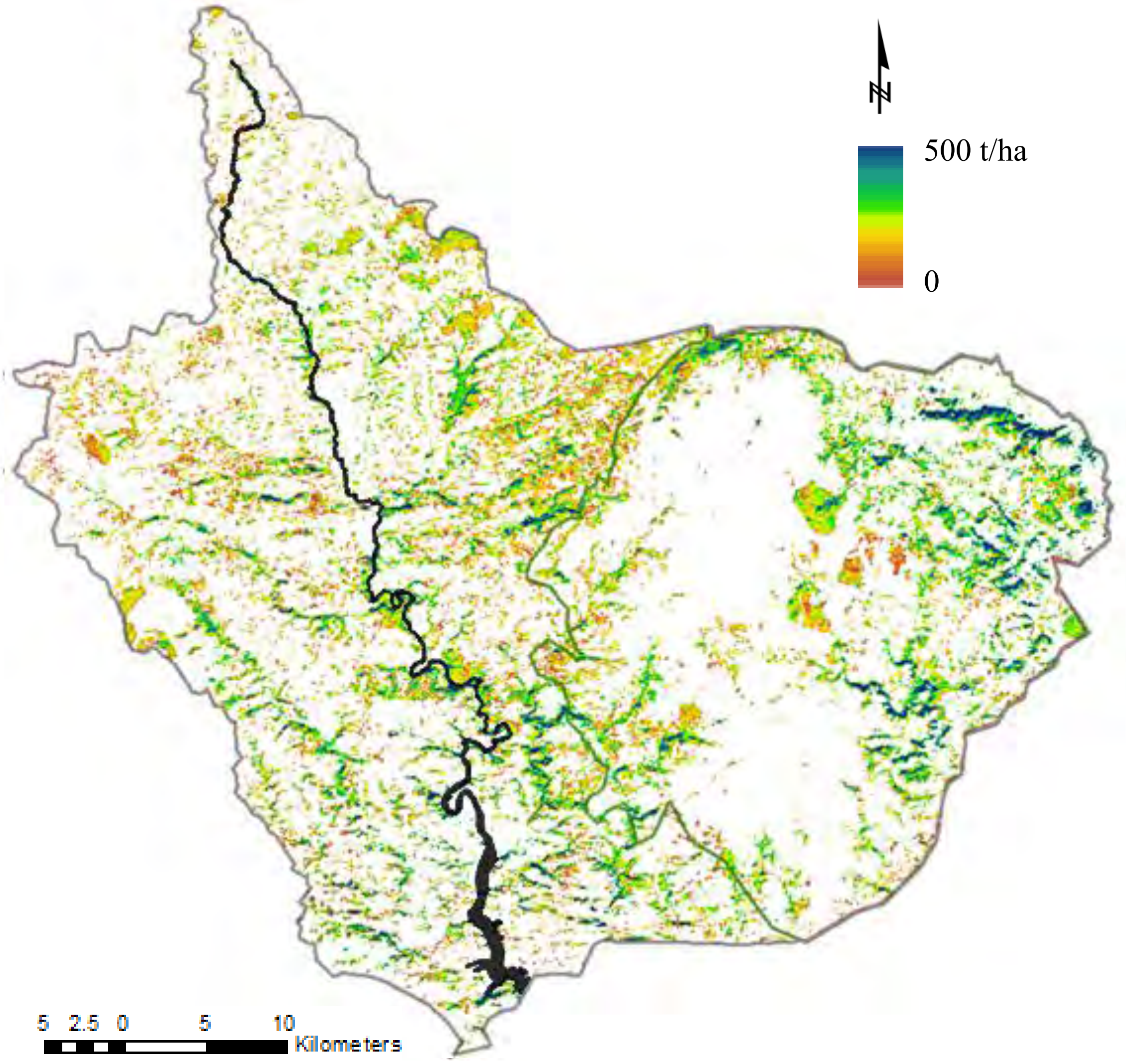


Figure 6-6: AGB map resulting from the second scenario.

6.3.3 Error Analysis

The error analysis tool was implemented with the seven settings used in the first scenario. Each setting was applied with a 1000 iterations. The execution time consumed to apply the software tool ranged between 3 to 6 minutes for each setting. The parameters standard errors are as listed in Table 6-5, and the uncertainty components of the field plot polygons for each setting are shown in Figure 6-7. The highest residuals lie in the areas of AGB around the 200 t/ha. The areas of low AGB showed a lower average residual density within the study area, with the AGB map (density on each polygon). The correlation between residual error and the modelled error was on an R^2 of 0.14. The correlation between simulated RS and field errors, and the errors mapping model results, was high with R^2s between 0.61 and 0.64. The result of applying the error model was the AGB error maps shown in Figure 6-8. The maps represent the error map for the main settings.

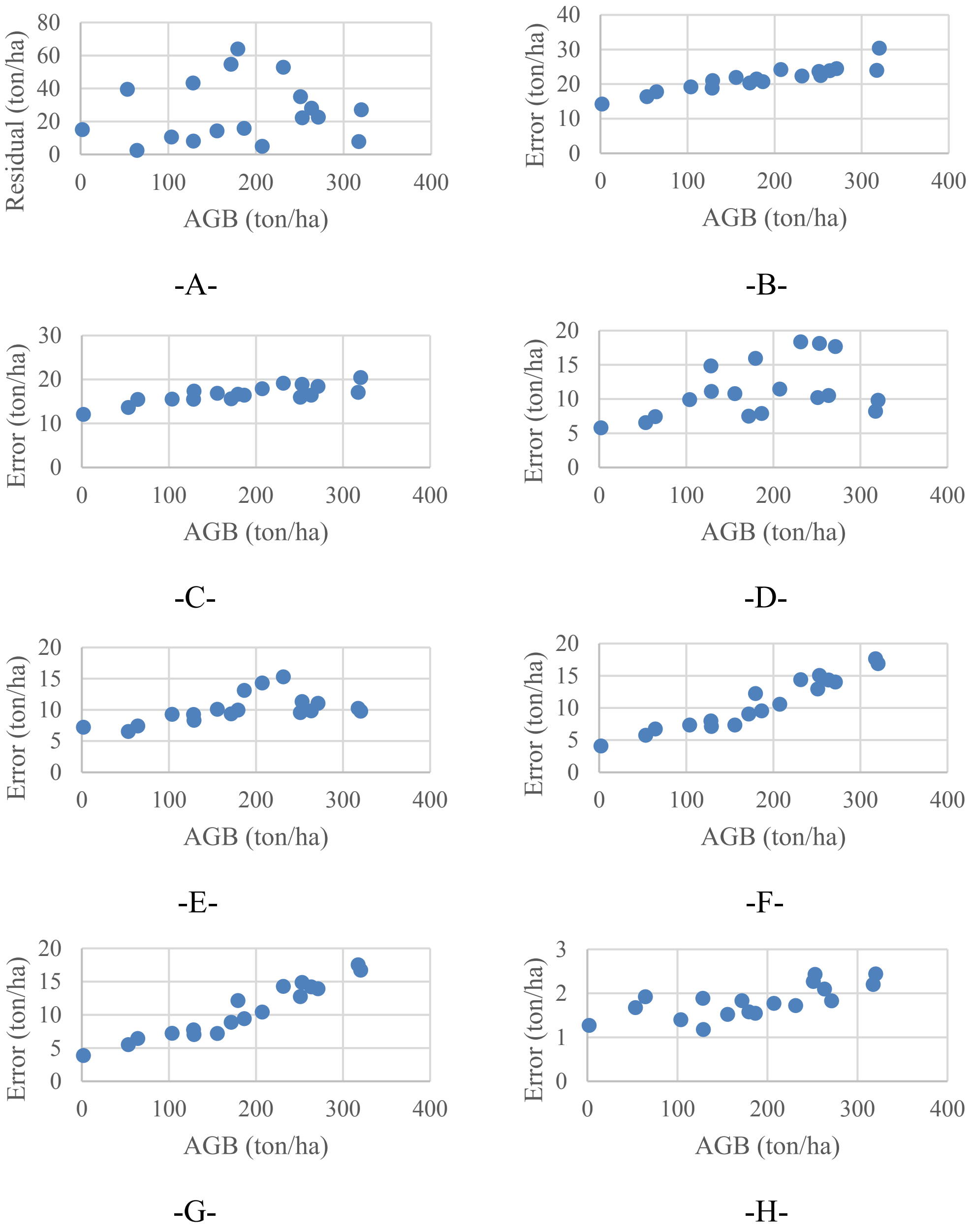
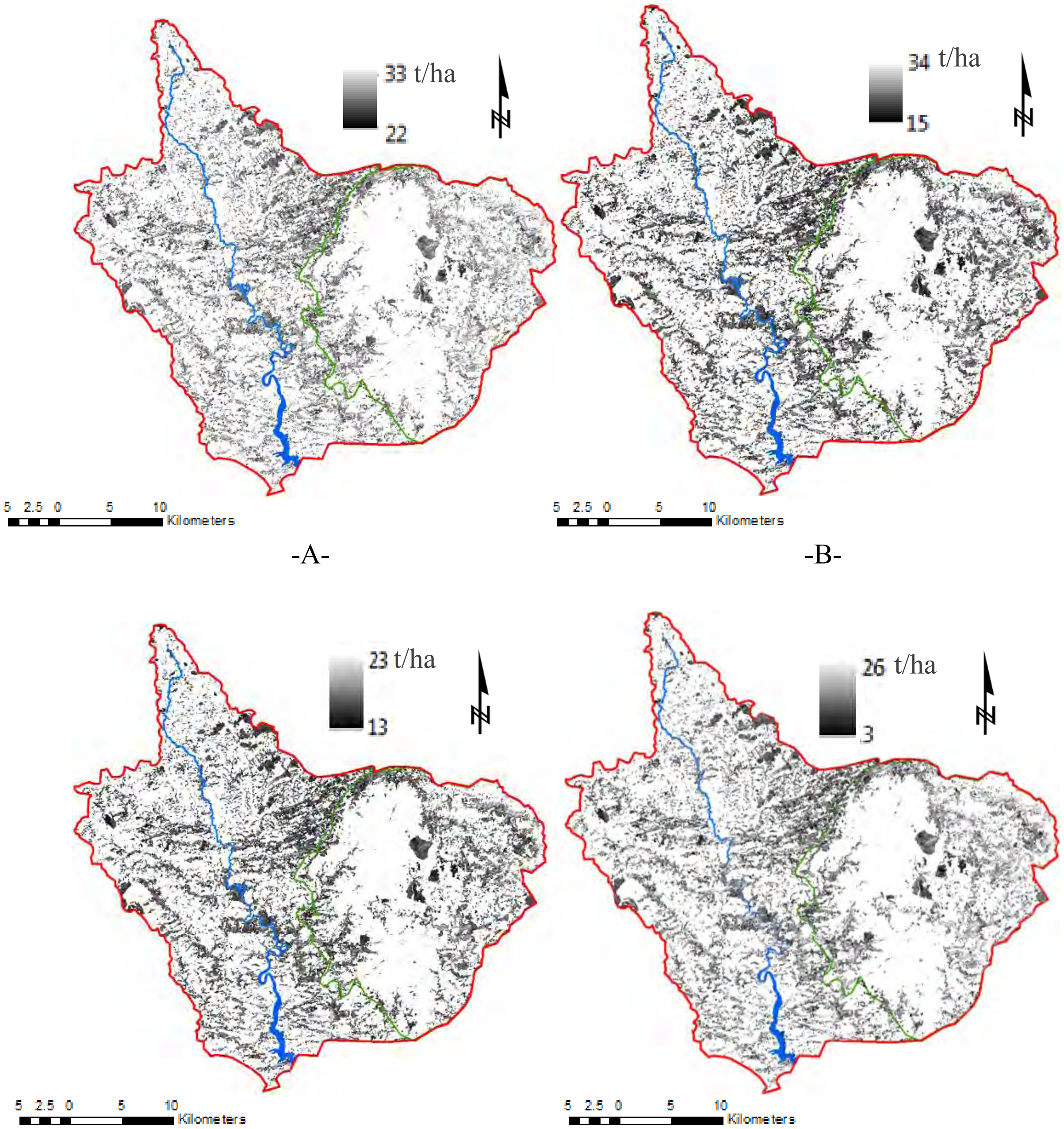


Figure 6-7: The uncertainty components of the second scenario, A- residuals of AGB fitting vs AGB, B- standard deviation yields from applying all RS and field errors vs AGB, C- standard deviation yields from applying all RS errors vs AGB, D- standard deviation yields from applying spectral RS errors vs AGB, E- standard deviation yields from applying spatial RS vs AGB, F- standard deviation yields from applying all field errors vs AGB,. G - standard deviation yields from applying spectral field vs AGB, and H- standard deviation yields from applying spatial field errors vs AGB.

As in the first scenario, the error propagation results showed that the errors propagated from the RS and field data have an obvious trend that is proportional to the AGB. The combined RS and field data error (Figure 6-7B) starts relatively high for low AGB pixels compared to the first scenario. The average of these errors is about 22 t/ha that is significantly lower than the average of the residuals which equals 29 t/ha. The RS errors (Figure 6-7C, D and E) shows that the average RS data error is about 17 t/ha. The spectral and spatial components of the RS data have averages of about 11 and 10 t/ha, respectively, with the average field error being around 11 t/ha. The non-spatial and spatial error components have averages of 10 and 2 t/ha, respectively.

Table 6-5: The influence of errors on the model coefficients for the second scenario.

	coefficient	SE regression (overall coefficient error)	SE all errors	SE due to field data errors influence			SE due to RS data errors influence		
				All	Spatial	Spectral	All	Spatial	Spectral
Landsat B4	-0.2816	0.159	0.081	0.022	0.003	0.020	0.089	0.027	0.064
NDVI	0.3997	0.107	0.120	0.014	0.005	0.012	0.077	0.041	0.032
LHD	0.1452	0.091	0.048	0.005	0.001	0.005	0.040	0.019	0.032



-A- -B-

-C- -D-

Figure 6-8: Error maps for the second scenario as A - fitting error, B - the uncertainty component that is related to RS and field errors, C - the uncertainty component that is related to RS errors only, and D - the uncertainty component that is related to field errors only.

6.4 Third Scenario

The third scenario applies a neural network regression as a non-parametric based mathematical model using the fieldwork information and very-high resolution active and passive RS data. The passive data includes the eight bands of WorldView-3, the NDVI, the first three principal components of WorldView-3 bands, and the six texture features with four window sizes for each feature. The active data includes the Lidar based DCM. A pixel-based approach was used to perform this scenario as described in the following sections.

6.4.1 Model Design

Some products are expected to have a high multi-collinearity (i.e., strongly correlated to each other) especially when they were based on the same band/bands for their calculations. Some other products can have low correlation to the AGB. Therefore, it was important to apply a feature selection process i.e., selecting a subset of only the relevant products to reduce the risk of overfitting, cut the noise of unnecessary products and increase the processing speed. Moreover, the flexibility of the neural network model can increase the probability of overfitting. Therefore, a simple ANOVA based feature selection that was used for the first two scenarios, was not applicable as it is based on the correlation between dependent and independent features. Therefore, the applied The Recursive Feature Elimination (RFE) algorithm for feature selection was adopted, as it was able to provide a robust solution.

RFE algorithm described in Liu et al. (2011) was used for this study. As a wrapper algorithm, it can analyse the correlation amongst features (Phuong et al., 2005) and hence be combined with many machine learning techniques for optimization based on the kernel or cost function used (Liu et al., 2011). The RFE used here is a two-step algorithm. The backward sample ranking runs an iteration based optimization for finding the best performing subset of features by repeatedly training a model and selecting the top performing feature based on the model based optimization criteria, excluding the selected feature and then repeating the process until all features are excluded. The results are then used to rank features according to model performance when they were excluded. Then the forward selection includes adding features to the model based on the ranking of the backward step, starting with the top feature and adding the next ranked feature in a recursive fashion according to the reported performance of the model at each iteration along with the features utilized in that step. The forward step is important for deciding which features do not improve the performance, and so can be eliminated. The skit-learn Python (Pedregosa et al., 2011) library was used to perform the feature selection. The model was tested with two fieldwork data types. First, the system was tested by implementing the total AGB map that was obtained by adding branch AGB map (an example plot is shown in Figure 4-18) to the trunk AGB map (an example plot is shown in Figure 4-16). Second the system was tested by implementing the branch AGB map only. The maximum R^2 obtained when the total AGB fieldwork data was 0.28 compared to 0.62 when branch AGB map was implemented instead of the total AGB map. Six out of the thirty-six products were selected according to the obtained correlation coefficient as shown in Figure 6-9; the selected features are DCM, NDVI, band 3, first principal component, the energy texture feature, and the ASM. The neural network applied used a single hidden layer with four neurons. The small number of neurons was decided to avoid overfitting due to the limited number of fieldwork samples as suggested by Del Frate and Solimini (2004) and Migolet et al. (2007) for similar applications with similar number of inputs.

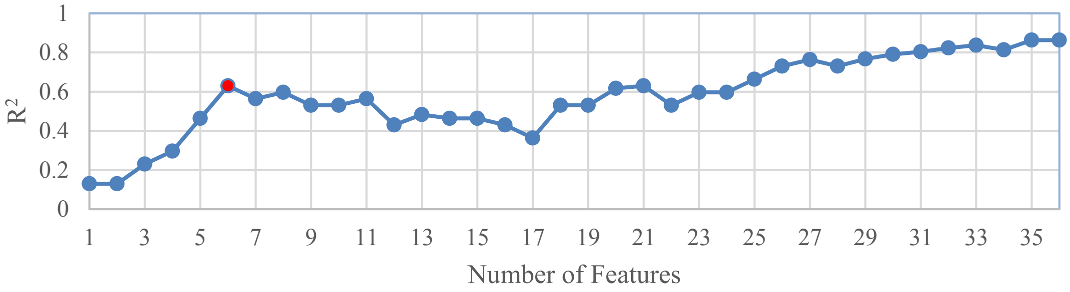


Figure 6-9: The obtained correlation coefficient for each number of features.

6.4.2 Biomass Mapping

The model was used to map the branch AGB. The R^2 of the correlation between the fieldwork data and calculated AGB based on the model was 0.62 as shown in Figure 6-10. The resulting branch AGB map is shown in Figure 6-11. Due to the pixel-based analysis, the map is of a spatial resolution that is equivalent to the spatial resolution of WorldView-3 image.

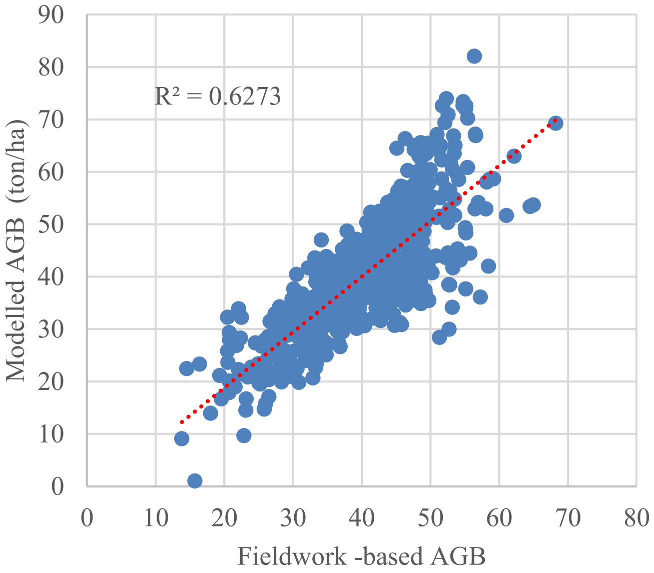
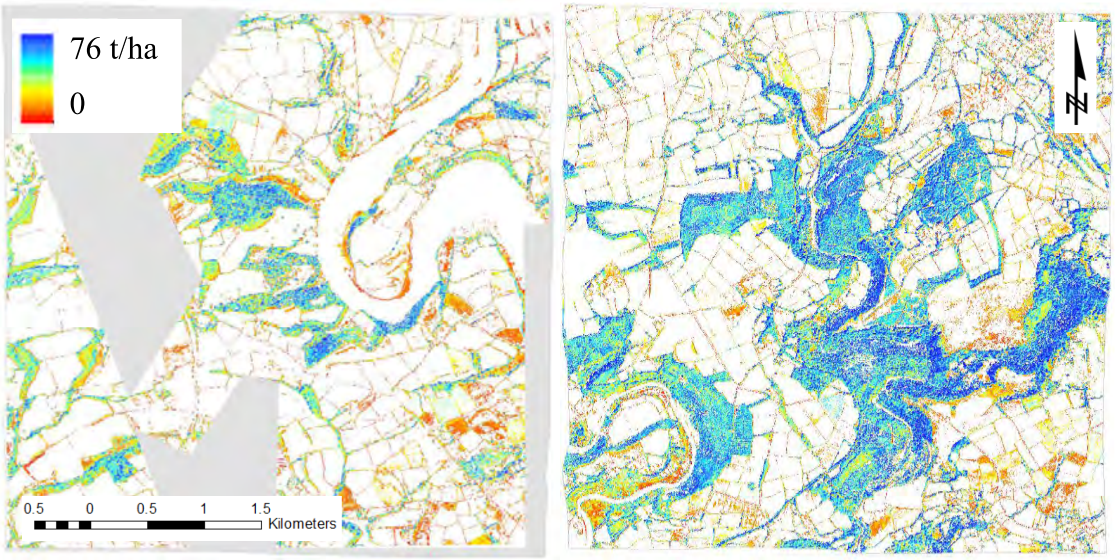


Figure 6-10: Modelled AGB vs fieldwork based AGB for the third scenario.

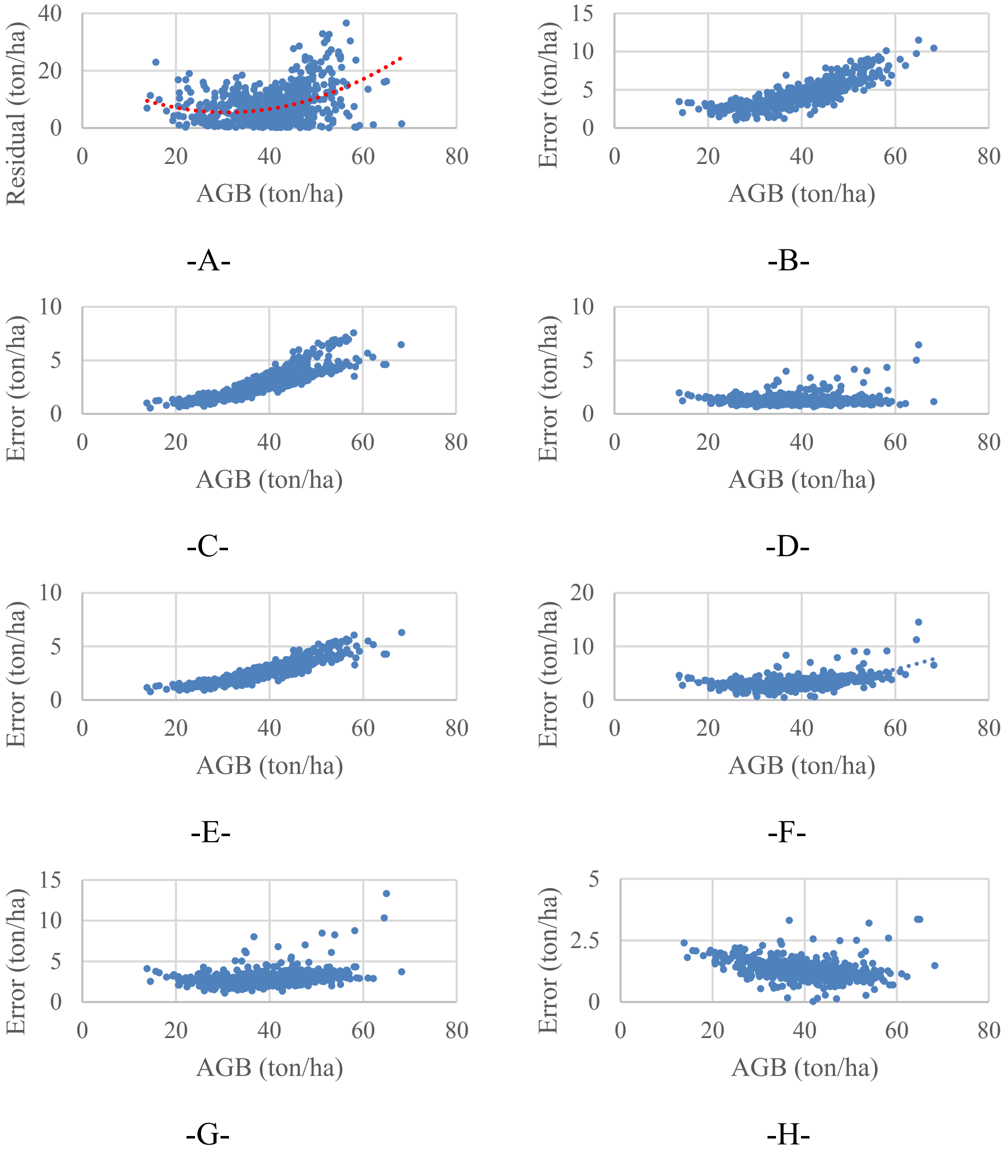


-A- -B-

Figure 6-11: The branch AGB map resulting from the third scenario.

6.4.3 Error Analysis

The error analysis tool was implemented with the seven settings, as used in the first scenario. Each setting was applied with 1000 iterations. The execution time consumed to apply the software tool ranged between 6 to 11 minutes for each setting. The uncertainty components of the field plot polygons for each setting analysis are shown in Figure 6-12. It shows that the average of model fitting residuals is about 8 t/ha and an RMSE of 10 t/ha compared to an average of branch AGB of about 40 t/ha. These residuals start with an average of about 10 t/ha for the low AGB, decreases to about 4 t/ha at the median value AGB areas and then increases again to about 20 t/ha with the high AGB pixels. The RS error has an average of about 3 t/ha the spatial error caused an average variation of 2.8 t/ha, whereas, the non-spatial error caused an average variation of 1.3 t/ha. The field error was 3.3 t/ha. Non-spatial field error component resulted in an average variation of about 3 t/ha, whereas, spatial errors results in about 1.3 t/ha. The general trend of all the error components to the AGB seems to be of a second order. Therefore, a quadratic regression model was used to generalised the error relationship to the AGB in order to map the error all over the study area. The correlation between residual error and the modelled error was on an R^2 of 0.12. The correlation between simulated RS and field errors, and the errors mapping model results, was high with R^2s between 0.51 and 0.84. The resulted error maps of the main setting are shown in Figure 6-13.



-A-

-B-

-C-

-D-

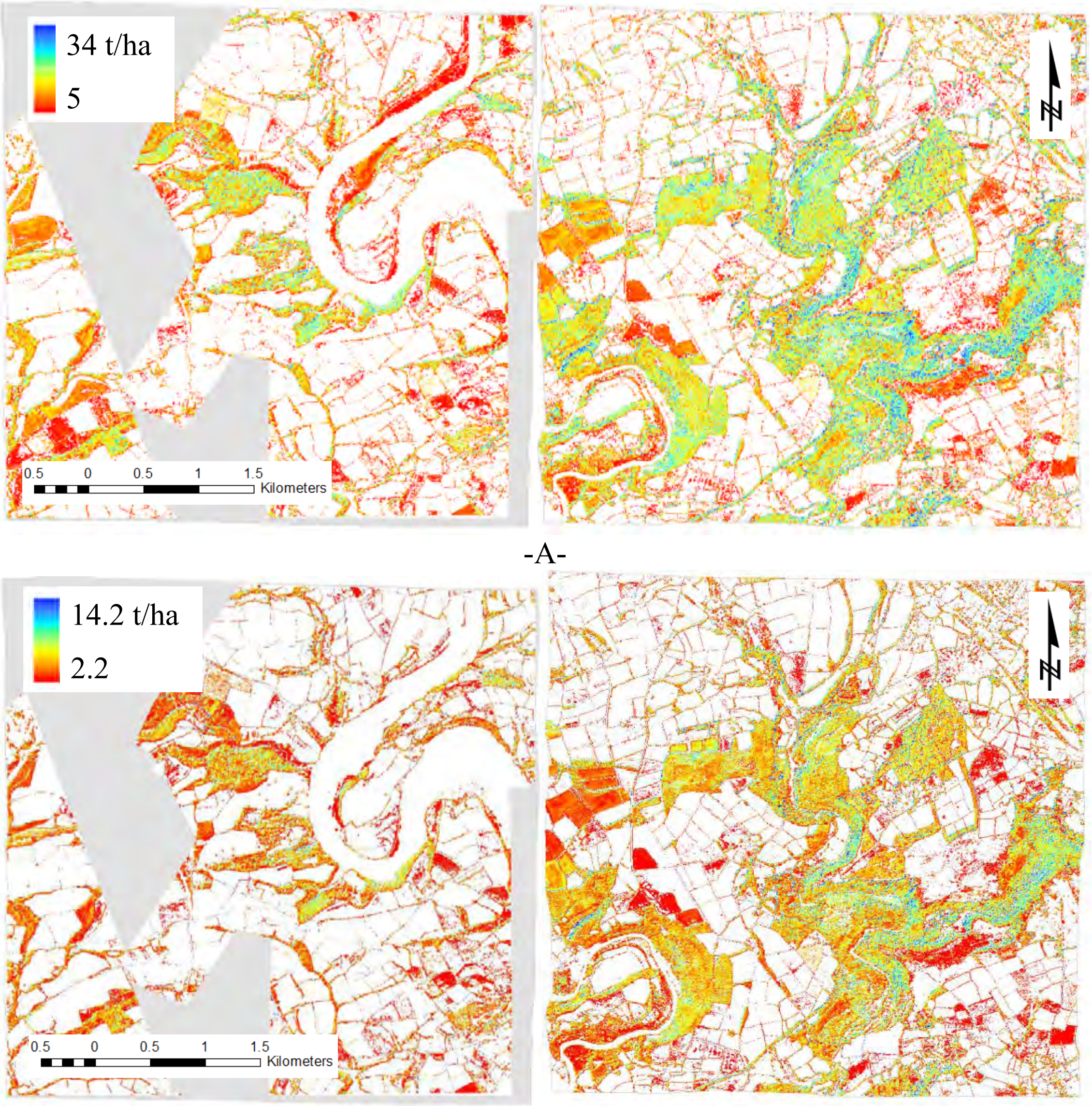
-E-

-F-

-G-

-H-

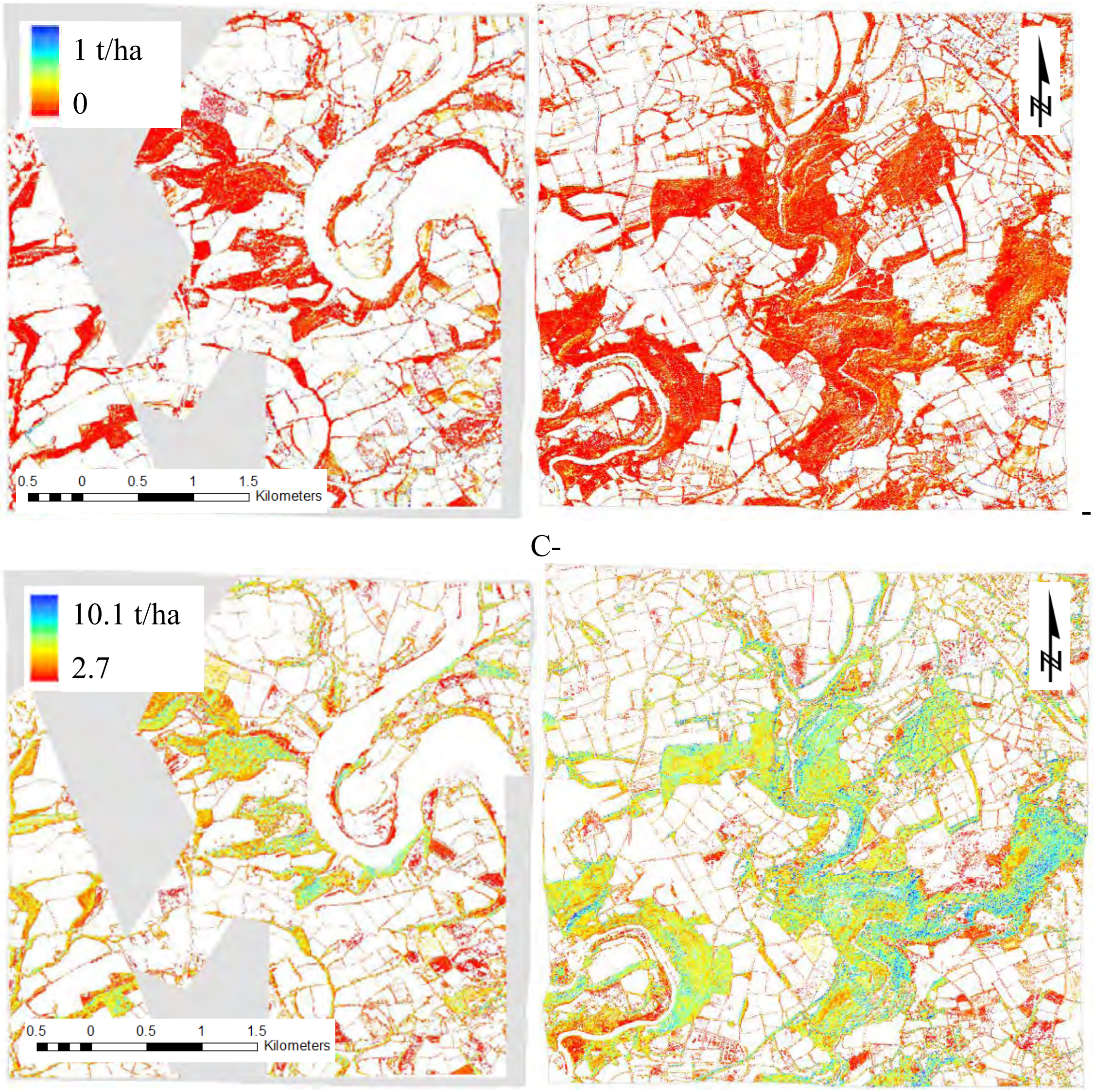
Figure 6-12: The uncertainty components of the third scenario as the A- residuals of AGB fitting vs AGB, B - standard deviation yields from applying all RS and field errors vs AGB, C - standard deviation yields from applying all RS errors vs AGB, D - standard deviation yields from applying spectral RS vs AGB, E - standard deviation yields from applying spatial RS vs AGB, F- standard deviation yields from applying all field errors vs AGB, G - standard deviation yields from applying spectral field vs AGB, and H - standard deviation yields from applying spatial field vs AGB.



-A-

-B-

Figure 6-13: Error maps for the third scenario, A- fitting error for both subsets, B - the uncertainty component that is related to RS and field errors for both subsets, C - the uncertainty component that is related to RS errors only for both subsets, and D - the uncertainty component that is related to field errors only for both subsets (continued).



-D-

Figure 6-13: (continued) Error maps for the third scenario as the, A - fitting error for both subsets, B - uncertainty component that is related to RS and field errors for both subsets, C - uncertainty component that is related to RS errors only for both subsets, and D - uncertainty component that is related to only field errors for both subsets.

6.5 Fourth Scenario

The fourth scenario applies an object-based analysis to satisfy a non-parametric based mathematical model using the same data as for the third scenario. However, due to the use of object-based analysis, the number of samples dropped compared to the third scenario. Therefore, and due to the fact that neural networks required large quantities data to be trained properly, a support vector machine regression was applied instead of a neural network approach.

6.5.1 Model Design

Five out of the thirty six products were selected according to the obtained correlation coefficient for each number of features as shown in Figure 6-14; the selected features were DCM, NDVI, band 3, first principal component, and the ASM. The support vector machine model was implemented with an RBF kernel as suggested by (Chen and Hay, 2011) for a similar data combination.

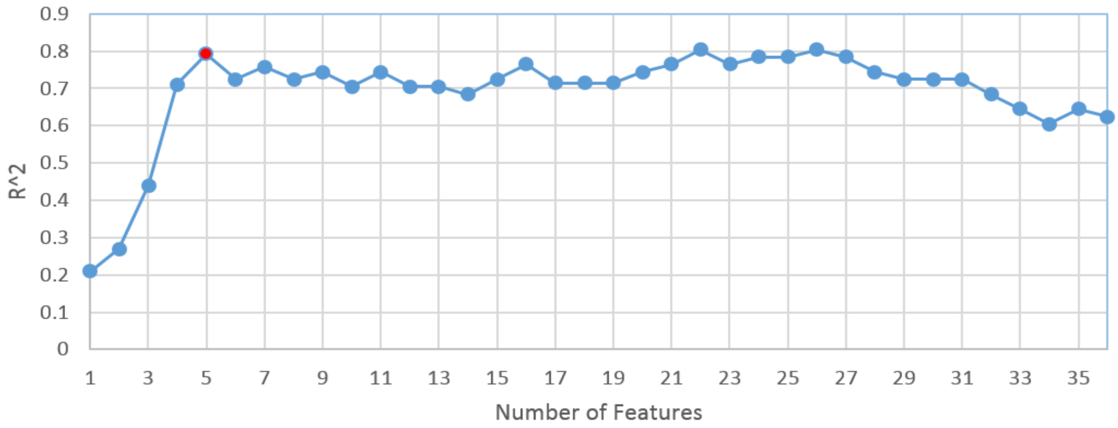


Figure 6-14: The obtained correlation coefficient for each number of features.

6.5.2 Biomass Mapping

The trained SVM model was generalised to the study area to produce a branch AGB map. The resulted branch AGB map, shown in Figure 6-15, is an object-based map. In other words, each object polygon was assigned to an AGB value. Figure 6-16 shows the modelled AGB versus the fieldwork based AGB for the object polygons inside the fieldwork sites. The R^2 of the correlation between fieldwork data and the calculated AGB based on the model was 0.78. The polygon AGB values ranged between 0 and 74 t/ha.

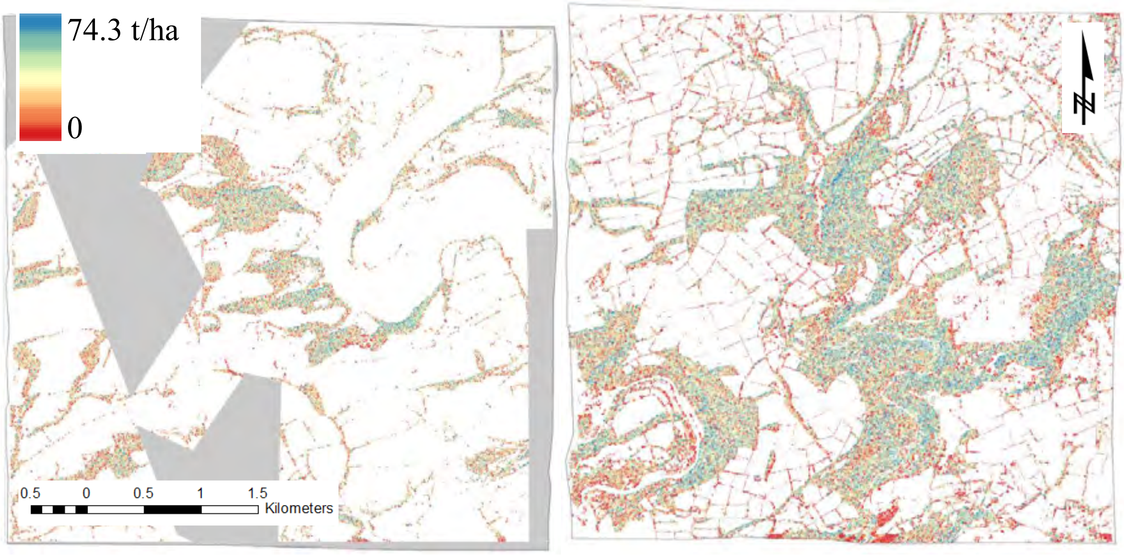


Figure 6-15: Branch AGB map resulted of the fourth scenario.

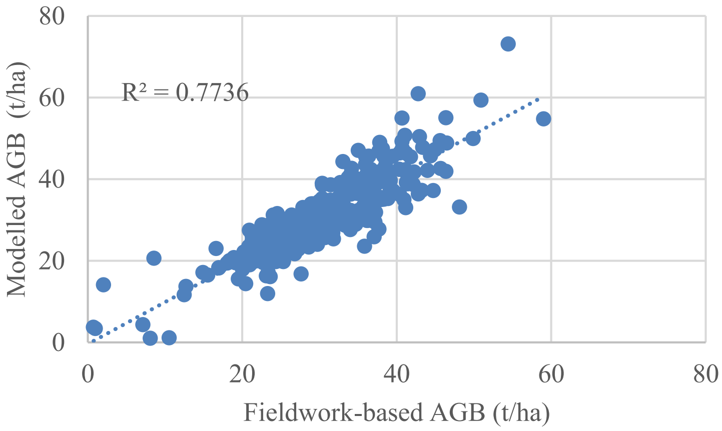
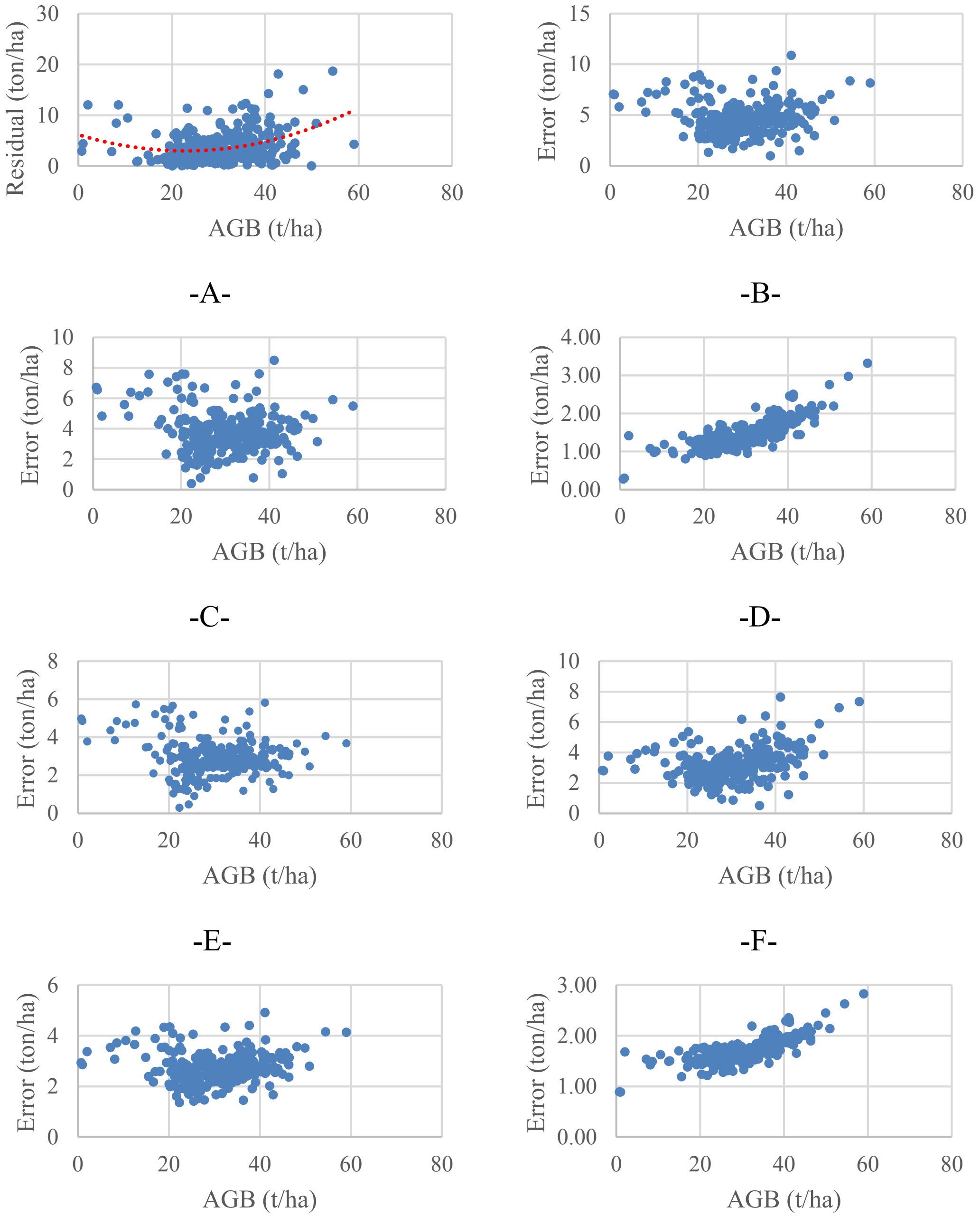


Figure 6-16: Modelled AGB vs fieldwork based AGB for the forth scenario.

6.5.3 Error Analysis

The error analysis tool was implemented with the seven settings used in the previous scenarios. Each setting was applied with 1000 iterations. The execution time consumed to apply the software tool ranged between 5 to 9 minutes for each setting. The uncertainty components of the field plot polygons of each setting are shown in Figure 6-17.



-A- -B-

-C- -D-

-E- -F-

-G- -H-

Figure 6-17: The uncertainty components of the fourth scenario, as: A - residuals of AGB fitting vs AGB, B - standard deviation yields from applying all RS and field errors vs AGB, C - standard deviation yields from applying all RS errors vs AGB, D - standard deviation yields from applying spectral RS vs AGB, E - standard deviation yields from applying spatial RS vs AGB, F - standard deviation yields from applying all field errors vs AGB, G - standard deviation yields from applying spectral field vs AGB, and H - standard deviation yields from applying spatial field vs AGB.

The error analysis shows that the average of the absolute residuals is 6 t/ha and the RMSE is 8 t/ha, with a distribution that is similar to the residuals of the third scenario. The average influence of the combined RS and field errors propagated to the AGB is about 5 t/ha. The influence of the RS data error was on average of about 4 t/ha. The spatial error component of RS inputs shows an influence in average of about 4 t/ha, whereas, non-spatial error component was on average of about 1.5 t/ha. The field data error was lower than the RS data error, with an average of about 3 t/ha and most of it being from the non-spatial error component. All the propagated error components show a quadratic relationship to the AGB. As in previous scenarios, the software tool modelled the error all over the study area based on the error behaviour of the fieldwork samples. The correlation between residual error and the modelled error was on an R^2 of 0.13. The correlation between simulated RS and field errors, and the errors mapping model results, was high with R^2s between 0.45 and 0.67. The error maps for main setting are shown in Figure 6-18.

6.6 Evaluation of the Scenarios Outcomes

To summarize the results of all scenarios, the averages over the error components were converted into percentage ratios as illustrated in Figure 6-19. For all four scenarios, it is noticeable that the error values are high although a variety of datasets, model types and processing techniques were included. The share of the errors between the RS and field error components are close, but the relative proportion is slightly different for different scenarios. For example, the spatial error propagation was computationally efficient; especially for the third and fourth scenarios. These errors have tended to be ignored in previous studies, and the only studies that included a similar results to this study were the studies based on destructive tests such as Colgan et al. (2013). All the scenarios showed that fitting residuals exceeded the error propagation results, with significant values that can indicate a hidden error component that was not included in the analysis.

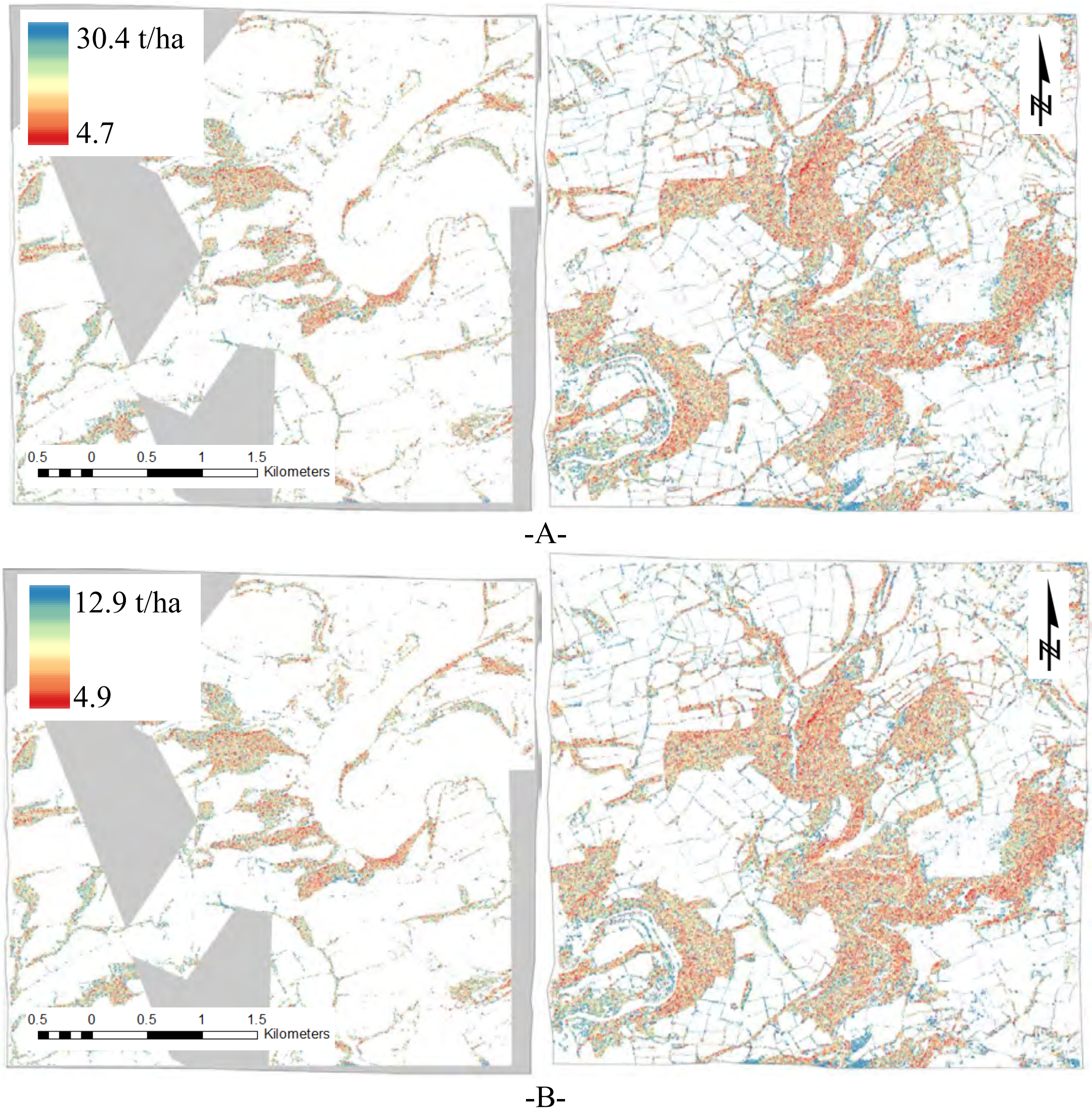


Figure 6-18: Error maps for the fourth scenario, as the A - fitting error for both subsets, B - uncertainty component that is related to RS and field errors for both subsets, C - uncertainty component that is related to RS errors only for both subsets, and D - uncertainty component that is related to field errors only for both subsets. (continued)

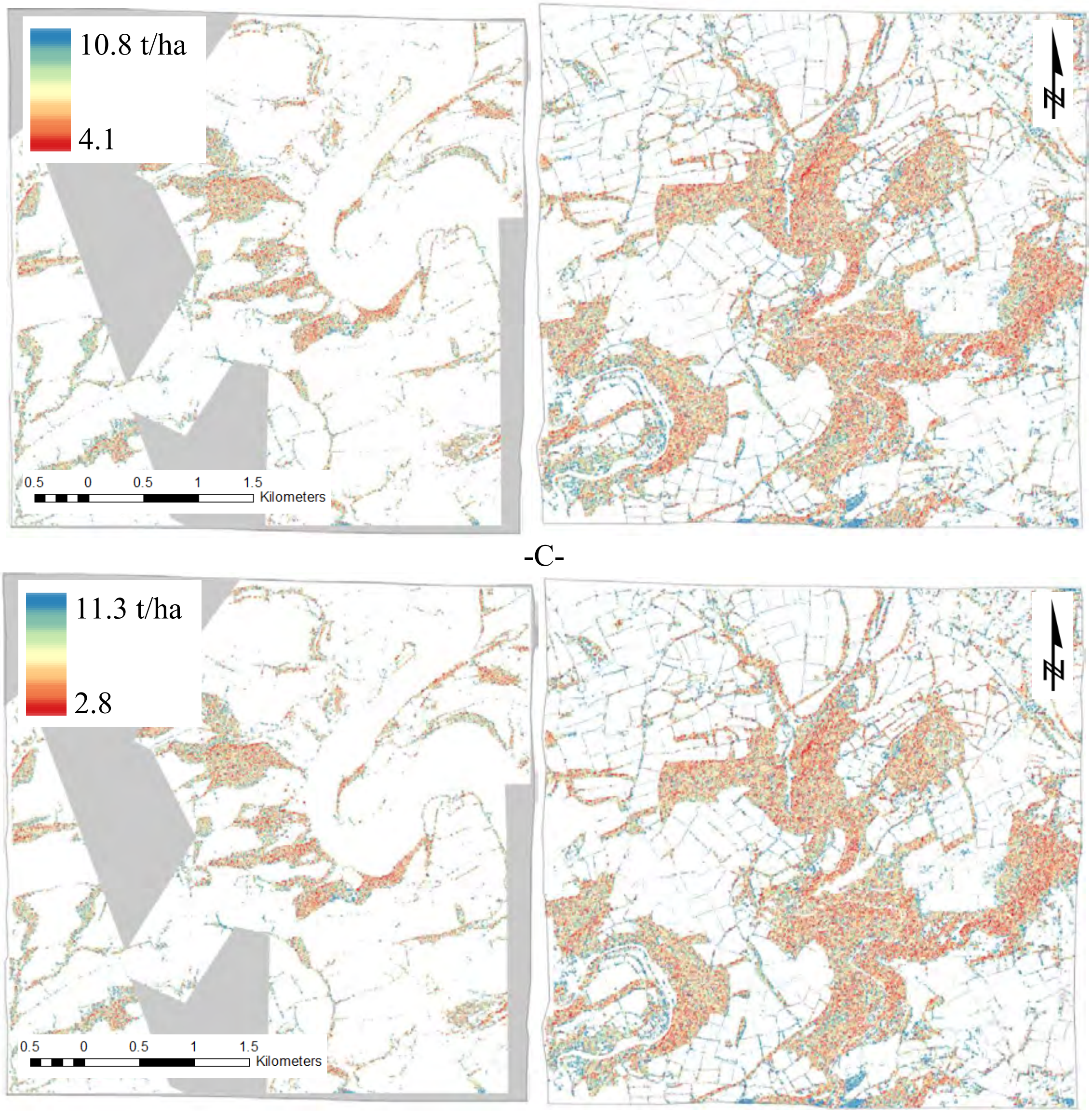


Figure 6-18: (continued) Error maps for the fourth scenario, as the A - fitting error for both subsets, B - uncertainty component that is related to RS and field errors for both subsets, C - uncertainty component that is related to RS errors only for both subsets, and D - uncertainty component that is related to field errors only for both subsets.

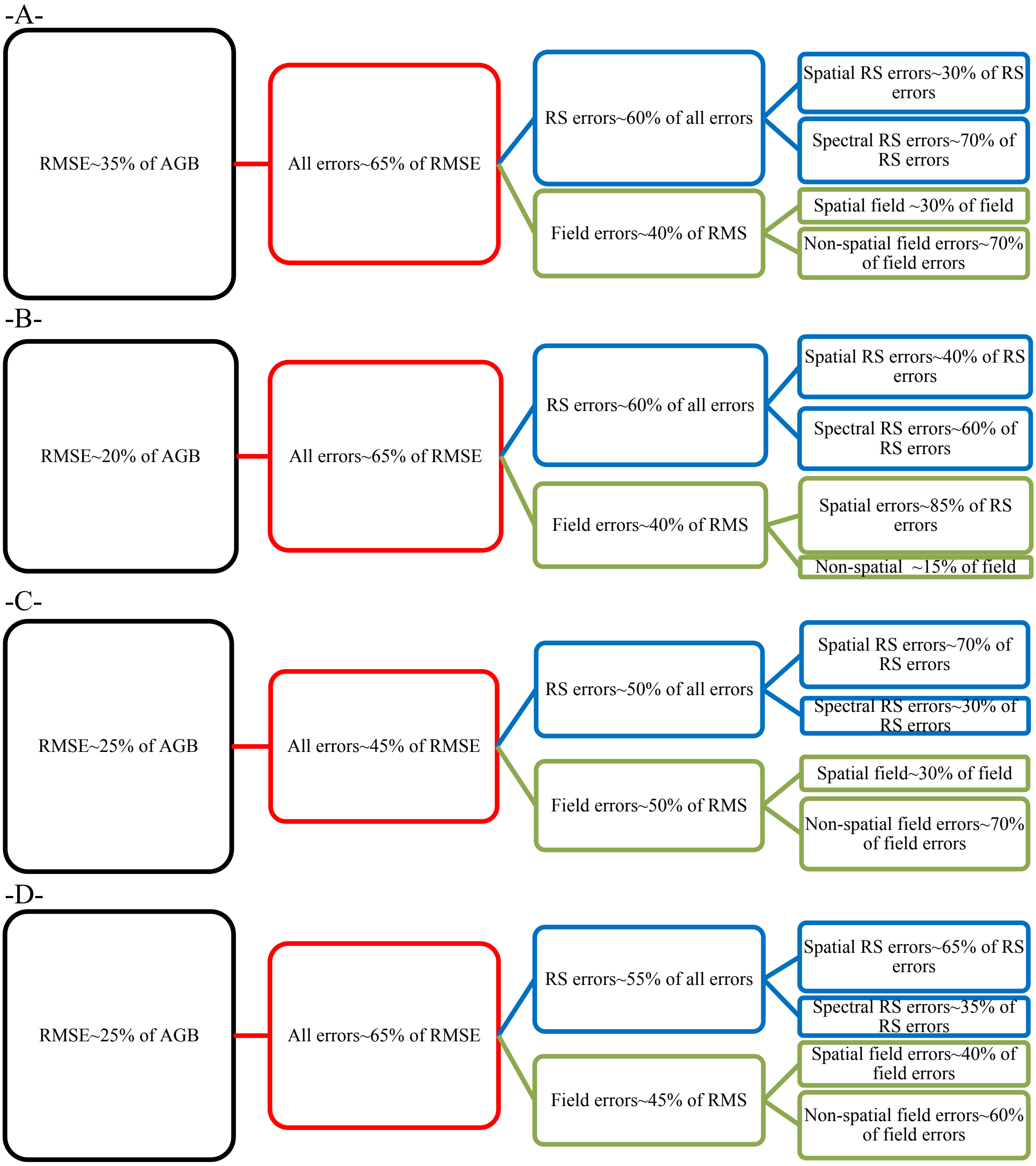


Figure 6-19: Approximate proportion of error component for: A- 1st scenario. B- 2nd scenario. C- 3rd scenario. D- 4th scenario.

6.7 Summary

The purpose of this chapter was to design, execute, and test a software tool for error propagation in the AGB assessments. The aim was to be applicable to the widest possible range of AGB assessment systems and inclusive to all definable input error types. Therefore, a Monte Carlo based error propagation tool was designed to deal with quantified errors of each input.

To reduce the computational costs, the designed system was based on a novel technique that can derive general error equations that can be applied to the whole study area, from analysing datasets that only exist within the fieldwork areas. In addition, the software was supported by a number of techniques that can accelerate the computation process, such as parallel computing technique and GPU processing capabilities. A user interface was designed to make the tool accessible for wide range of users, and it has been uploaded to an online sharing platform so it is available to the community.

The input error types were considered as RS data errors and field errors. Each has two types of errors, non-spatial and spatial. The errors in the RS inputs were explored using nominal error indicators provided with the raw data or previous experimental tests whenever these were available. The errors in the field data were based on the nominal uncertainty indicators of the allometric equations and the experimentally identified errors in the tree DBH and H.

The tool was successfully tested on four scenarios that cover a wide variety of model RS inputs (Sentinel-1, Landsat-8, Lidar and WorldView-3 data), model types (parametric and non-parametric), and processing methods (pixel-based and object-based). The results showed a relatively high RMSE for all scenarios, with error propagation analysis showing a significant effect for all defined error components. However, the propagated error result was generally smaller than the Root Mean Square Error (RMSE) as shown in Figure 6-19.

Chapter 7 Discussion

The research has dealt with different types of error involved in AGB assessment, which included assessing those coming from the field and RS data and the influence on the results. Section 7.1 discusses the findings of the literature review (Chapter 2), with the design criteria and accuracy prediction tool (Chapter 3). The fieldwork data collection, field data processing, and the resulting fieldwork based information (Chapter 4) are discussed in Section 7.2. The uncertainty analysis tool (Chapter 5) and its application scenarios (Chapter 6) are discussed in Section 7.3, and then finally future work is discussed in Section 7.4.

7.1 Literature review and resulting accuracy prediction tool

Several reports, such as (Le Quéré et al., 2016, Réjou-Méchain et al. 2017, and Njana 2017), have shown there is often large uncertainties in global estimations of AGB and there is no agreed standard system for AGB assessment, with many of the published studies using either new techniques or new data types. As mentioned in the literature review, there was a lack of analytical approaches that could support non-parametric models, non-linear parametric models, and object-based models within error propagation techniques. Spatial errors were often excluded, and there was limited use of accuracy prediction schemes for projects that are still in the system design stage.

The reviewed data types and model types (Chapter 3) has limited the error propagation methods to the Monte Carlo method. Deterministic methods, such as Taylor's and Rosenblueth’s methods, are not applicable for the non-differentiable non-parametric models such as ANN. Methods based on iterating data subsets in their original forms (such as bootstrap methods) are not applicable for some error types, due to the dependency between the errors of different subsets. The Monte Carlo method used independent random error values and error orientations for each iteration making the variance between iterations more representative of the effect of the simulated error. This finding is contrary to previous studies which have suggested that deterministic methods should be used to minimise the computational costs, e.g. Chen et al. (2015).

The accuracy prediction tool presented in Chapter3 raises the possibility that a machine-learning tool (ANN) using previous studies can solve the complex task to assess accuracy during a system’s design stage and the rationale of using the ANN was that it could deal with inputs that are class-based as well as quantity inputs. For example, the RS data type could be Lidar, Radar, optical etc. with each represented by a class number. The drawback is that it requires a large number of training samples compared to the limited number of studies available so far, which affected the accuracy of the results obtained. Therefore, the solution was to apply a separate model for each RS data type while the fieldwork specification and the model type are less complex and so could be represented by one variable each. In general, the obtained correlation further supports the idea of Ahmed (2012) and Weisbin et al. (2014), that the accuracy of AGB is largely based on the system design and the data specifications.

7.2 Fieldwork Data

The results showed that the accuracy of the inputs to the tree biomass allometric equation was relatively high compared to the other tree's parts, possibly because of the provided knowledge of the tree trunk physical model as an upstanding object. Emphasising the trunk gave more confidence in tree diameter at breast height (DBH) fitting and in tree height (h) extraction. Even though the accuracy of the total biomass of a single tree is based on the accuracy of the DBH and H, the accuracy of a wider scale process of biomass assessment can be influenced by the tree detection efficiency. Therefore, any weakness in the tree detection can affect the accuracy of both biomass and distribution.

The proposed algorithm succeeded to reduce the false negatives automatically, by applying tolerance parameters. However, it failed to reduce false positives automatically, and so a manual step was unavoidable. Nevertheless, this stage can be considered as a learning algorithm because it is designed to retain all the user answers to be able to use them as training data for a future machine learning classifier; the relative accuracy of this algorithm's results could be accepted or rejected based on the required application of the biomass. In all cases, it cannot be dealt with as error-free information. This issue has not previously been described as previous studies either dealt with individual trees such as Côté et al. (2011), or assumed that tree detection is an error free process, e.g. Calders et al. (2015).

The 3D voxels that are parallel to the terrain can be applied regardless of the terrain smoothness. This technique improved the detection by providing fewer layers with all the trunks at the same height. It also allowed the algorithm to focus on a particular layer, and so increase the computational efficiency. Another benefit was that it could provide the tree height with minimum required calculations because it provides relative heights from the ground rather than absolute elevations.

The practical applicability and time efficiency were considered when fitting techniques were chosen on condition that they did not affect the archived accuracy. For example, the vertical elliptical cylinder fitting was used because the inclined cylinder fitting failed to converge in number of cases and therefore is expected to affect the system consistency. Similarly, the branch weighting methods showed that all methods had a moderate correlation with the branch biomass. Even though it has a slightly lower correlation than the cylinder fitting method, the skeletonisation method was adopted since it required lower computation resources. Unlike the trunk DBH fitting, the branch weighting was applied to a relatively large number of branches in each plot, but will not affect the total biomass because it is not an input to the allometric equations. For example, for branch reconstruction using skeletonisation can increase the time efficiency by about 42 mins for each 1000 tree branches given that some plots can have more than 10000 branches.

The biomass maps are at a high spatial resolution (0.05 m). This resolution is significant in at least two respects; the plot could be subdivided to give more spatially accurate subplots; and the subplot shape is flexible enough to fit any object when object-based analysis is used. For example, this type of output can be correlated to 1 m resolution RS data on a one-to-one pixel basis. At the same time the subplot can be correlated to random shape segment objects.

7.3 Error Propagation Analysis

Spatial errors were considered within the error analysis for AGB; maybe for the first time. The spatial errors can be input as a shift error and orientation error with other spatial errors such as geometric distortions, relief displacement, and scale errors, not included in the analysis due to the complexity of these types of errors that will affect the time efficiency of the error analysis. It is argued that these errors should be small compared to the shift and orientation error for the geometrically corrected images.

The error analysis tool was used in four scenarios, intended to cover the main categories of AGB assessments by exploring active and passive RS data with high and very-high resolutions, and with both parametric and non-parametric models. Landsat-8 radiometric data and NDVI have been widely used, as discussed in chapter 3. The newly introduced product was NDVIR, which is expected to have a relationship with the tree type and AGB because each tree species can have its distinctive annual cycles.

SAR data are also widely utilised for AGB assessments, with Sentinel-1 providing a unique combination of global coverage, C-band SAR data and free to access data. Therefore, implementing this dataset illustrated a highly challenging, but not uncommon, conditions for data processing (Argamosa, et al. 2018). However, the LHD based on Sentinel-1 data had not been explored in previous studies. Therefore, the aim was to explore its capabilities for AGB assessment and to simulate the error propagation through an interferometric based input. The stability analysis (Figure 5-8 and Figure 5-9) showed that this type of data is unstable and therefore it cannot be generalised, meaning that the LHD map used for the model training should also be used for applying the model. This result agrees with Kyriou and Nikolakopoulos (2018), which stated that stable interferometric borders are not achievable with Sentinel-1 data.

The RS input error analysis showed three types of spectral/non-spatial errors: constant error i.e., when the error value is constant all over the map; error value has a defined linear relationship to the RS product, e.g. error of 2% of the pixel’s value; and non-linear relationship between the error and the RS product/s. e.g. the error of NDVI explained in Section 5.8.4.1. This finding supports Ahmed (2012) and Chen et al. (2015)'s suggestions to use a theoretical framework to estimate errors.

Previous studies, including Ahmed (2012), Colgan et al. (2013), Chen (2013), Weisbin et al. (2014), and Chen et al. (2015), have tended to limit the field error analysis to the non-spatial errors, while this study extended the analysis to include the spatial errors and the misdetection error. The non-spatial error analysis was based on the determined DBH and H errors and the standard error for each allometric equation parameter.

Figure 7-1 combined the behaviour of non-spatial error of observed tree species in Figure 5-29C, Figure 5-30C, Figure 5-31C, and Figure 5-32. Comparing these errors shows that the tree species and the selected allometric equation error can affect the accuracy. The figures also illustrate that the error has a direct relationship to the AGB, with all species showing non-linear behaviour for small AGB values that changes to an almost linear behaviour for the vast majority of the AGB range. Therefore, this relationship was approximated to a first order linear equation for each tree species. This approximation shows a close similarity with (Chen et al., 2015)'s field error equation (σ_{AGB}=0.38xAGB_{tree}) and (Chave et al., 2004)'s field error equation (σ_{AGB}=0.31xAGB_{tree}), i.e. in both the error equation form and the equation parameters.

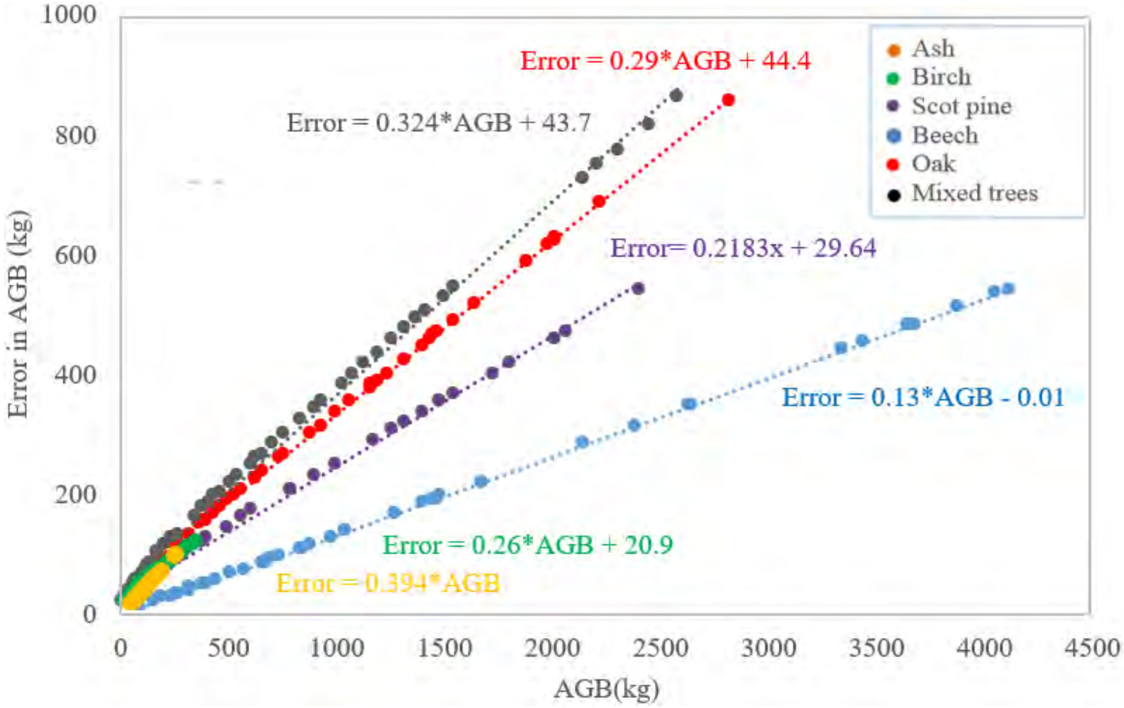


Figure 7-1: Non-spatial errors per tree in the field AGB of explored species compared to the tree's AGB.

The non-spatial error analysis for field data demonstrated:

- an agreement with statements in previous studies about the influence of the allometric equation specifications on the error of field data, highlighting the importance of the availability of allometric equations with a minimum uncertainty level;
- that the error in the field data is proportional to the AGB that means the field error is a map based error rather than a constant number;
- that converting an allometric equation from the DBH based form to the H based form will reduce the high relative accuracy of H that can be obtained from the TLS.

The most important error source is the allometric equation, with the error in DBH being negligible (Molto et al., 2013). Therefore, improving an H-based equation was not expected to improve the accuracy. However, the relative error in DBH that is based on

TLS is high compared to the relative error in H that is small compared to conventional methods. This result agrees with the conclusion of Calders et al. (2015) that a TLS provided H that is closer to the destructive test results than other field inventory approaches

Error analysis for the field data was tree rather than pixel-based and therefore the ideal procedure for each iteration should include separating each tree's AGB, adding the random error for the tree AGB, and then reassembling the AGB map. However, this solution can slow the analysis down and so the solution was to use the plots as a cluster of trees (samples). Each cluster has an error in one direction (either positive or negative for all pixels) with random quantities for each pixel. This field data averaging process of field data for each RS pixel/object can be bias. The unavoidable disadvantage of this solution is a higher probability of getting an imbalanced number of positive to negative error trees for each fieldwork site due to the limited number of plots per site.

The model errors consisted of two components: the error that was propagated from the inputs to the model and then from the model to the final AGB; the error from the mathematical model's inappropriateness for the relationship between the RS data and AGB. The second component should be eliminated in the model generation stage, by providing a sufficient number of field samples, applying feature selection to avoid overfitting and designing the model and the model characteristics based on statistical tests to avoid under-fitting. Therefore, the model error was not considered as an external source of error. This finding is contrary to previous studies, such as Sabia et al. (2008) and Ahmed (2012), which have suggested that model error is an independent external error.

The scenarios were designed to cover the widest possible range of RS based AGB assessment. For the first two scenarios, the feature selection process was limited to implementing an ANOVA procedure. Both scenarios used the same inputs but yielded two different sets of products due to the processing technique (either pixel or object-based). Kellndorfer et al. (2004) established that SAR based heights are not stable unless averaging a minimum of 20 pixels. Therefore, it is expected that a higher LHD performance would be achieved with the object-based processing that averages the pixels with each polygon. The interpretation for the failure of NDVIR, to pass the feature selection in the second scenario, could be that the NDVIR did not perform as good as it performed in the first scenario due to the averaging process. The other interpretation could be that the performance of band 4 data has improved with object-based analysis and therefore its P-value increased and exceeded the P-value of the NDVIR. For both scenarios, the selected Landsat-8 products somewhat disagree with previous studies such as Hall, et al. (2006) which chose Landsat-7 bands 3, 4, and 5, and Dube and Mutanga (2015), which selected Landsat-8 bands 3,4,5, and 7; and vegetation indices. This disagreement could be due to the differences in the fieldwork data sampling that uses pixel/object-based samples instead of plot based sampling.

There are similarities between the ratios of remote sensing and fieldwork error splits to the total error in this study and those described by Colgan et al. (2013); Weisbin et al. (2014); and Chen et al. (2015); the error component averages for all scenarios showed that the highest share of errors came from RS data followed by the field data. However, the total error seems to be underestimated in the non-destructive studies, namely Weisbin et al. (2014); and Chen et al. (2015). This difference was interpreted to be due to the neglected error component in these studies, as they did not analyse the spatial error component.

An important implication of the error analysis is the possibility to determine the link between the local AGB and its error for each pixel/object. For all scenarios, the error propagation shows that the combined error from RS data and field data has a general trend that is directly proportional to the AGB with a relatively high average.

Although there was a significant increase in the R^2 of the second and fourth scenario model compared to the first and third scenarios, the relative absolute residuals were still high. This finding, while preliminary, suggests that reporting high correlation without validating the results, as many previous AGB assessment did, does not mean low uncertainty. Instead it can indicate a risk of overfitting.

In the first and second scenarios, the highest residuals corresponded to the medium AGB, which might indicate that there are two or more areas that have tree covers of distinctive behaviours linked to the RS inputs that were covered by fieldwork validation samples. However, in the second scenario, the error value for the low AGB polygons was high compared to the corresponding errors for the low AGB pixels in the first scenario. This increase could be because of the constant spectral error of the LHD; the average error from the error propagation is close to the value of the average residuals.

For the third scenario, the smallest possible sample size, which is one pixel, was used in order to make use of the detailed fieldwork-based AGB maps to provide the largest number of samples for the model training as a small sample size can reduce the precision by averaging the data over the plot area (Colgan et al. 2013; Hensley et al., 2014). However, the resulting direct correlation between the RS data and the total AGB was unsatisfactory with a weak correlation due to the high density of the AGB of the tree trunk over a small area that is not completely visible by the optical RS data. In other words, the tree trunk that has the largest share of the tree's AGB is hidden behind tree branches and leaves. Therefore, the branch AGB map was used to provide a better correlation. These findings may help to understand why finer resolution data can possibly fail to improve the accuracy of AGB assessments in general.

Applying an averaging filter to both the RS data and branch AGB data was key to reducing the effect of the pixel to pixel registration errors between the RS datasets and field data, and the effect of shadow areas in the WorldView-3 data. One of the disadvantages of using the branch biomass was that the relationship between total biomass and branches is very dependent on the tree species. For example, based on equation parameters in Table 43, an Oak tree with a branch AGB is about 400 kg should have a trunk AGB is about 2500 kg compared to a trunk AGB is about 860 kg for an Ash tree of the same branch AGB. Therefore, the conversion from branch AGB to the total AGB requires accurate tree species mapping.

When compared to the Eckert (2012) feature selection result for WorldView-3 data (Figure 6-9 and Figure 6-14), both results agree on selecting a vegetation index and more than one texture feature. However, Eckert (2012)'s feature selection excluded the PCs and the original bands, which could be linked to the difference in the sample sizes as Eckert (2012) used plot samples.

The error analysis, (Figure 6-12), showed that the model fitting residuals are high compared to the average branch AGB. These residuals start high with the low AGB, decreases for the middle values and then increase again for the high AGB pixels. A possible interpretation is that the high number of samples which are concentrated in the low and high AGB ranges and hence force the model to fit these samples. An implication of this is the possibility that error analysis can be also used to indicate the weaknesses in the fieldwork sampling.

The total error yield from propagating RS and field data to the AGB shows a direct correlation with the square of AGB which means that a second order model is reasonable to model the error outside the field areas. The quadratic relationship between error and AGB occurs as a result of the error fitting model matching the AGB model regarding complexity and flexibility because it was designed to be applied to the same model specifications.

The fourth scenario all the selected features were common with the third scenario. Although it uses a smaller number of features, the trained model shows a higher R^2 than the third scenario model (R^2=0.78). This could be because the object-based analysis provided a smaller number of samples compared to the pixel-based analysis; the smaller number of samples can train a lower complexity model and therefore requires a smaller number of input features to avoid overfitting (Gu et al., 2016; Lever et al., 2016).

The variances between residuals and error propagation results of the scenarios could be due to the existence of one or more error components that were not included in the analysis. In this study, all possible spatial and non-spatial errors from the field and RS data were taken in consideration. Still, the gap between residuals and simulated errors is large. Therefore, a possible hypothesis is that this component is related to a weakness in the relationship between AGB and the RS data.

One of the issues that emerges from the scenarios is that the portions of errors can vary based on the processing methodology. Figure 6-19 shows that the ratios of spatial and non-spatial error component varies for each scenario. For example, for the first two scenarios, the non-spatial RS errors have more influence than the spatial errors.

In contrast, for other scenarios, most of the RS error was caused by the influence of the spatial error.

The results summarised in Figure 6-19 also shows that all error component were significant. According to the findings of this study, the error sources considered by previous studies (Figure 2-1) could be extended to include the spatial error components of both field and RS inputs. Therefore, an alternative illustration is shown in Figure 7-2. This lack of error analysis for these two components (the spatial errors and the weakness in the relationship between RS data and AGB) can explain why previous simulation models were optimistic compared with the practical AGB models in this research. An example of an optimistic simulation study is Weisbin et al. (2014) that expected the Lidar based systems to achieve an uncertainty level of 5%. Also, Feng et al. (2017) showed that after a certain level of accuracy, adding a new type of RS data cannot improve the accuracy.

The error analysis was applied with the minimum number of settings to satisfy the objectives of this study. Therefore, only seven settings were used for each scenario. However, the error analysis tool is flexible enough to provide further specific settings to separate the error components related to each single data set, which can satisfy a more sophisticated error analysis. Therefore, the outcomes of the undertaken quantitative research describe the state of concerns within the applied scenarios.

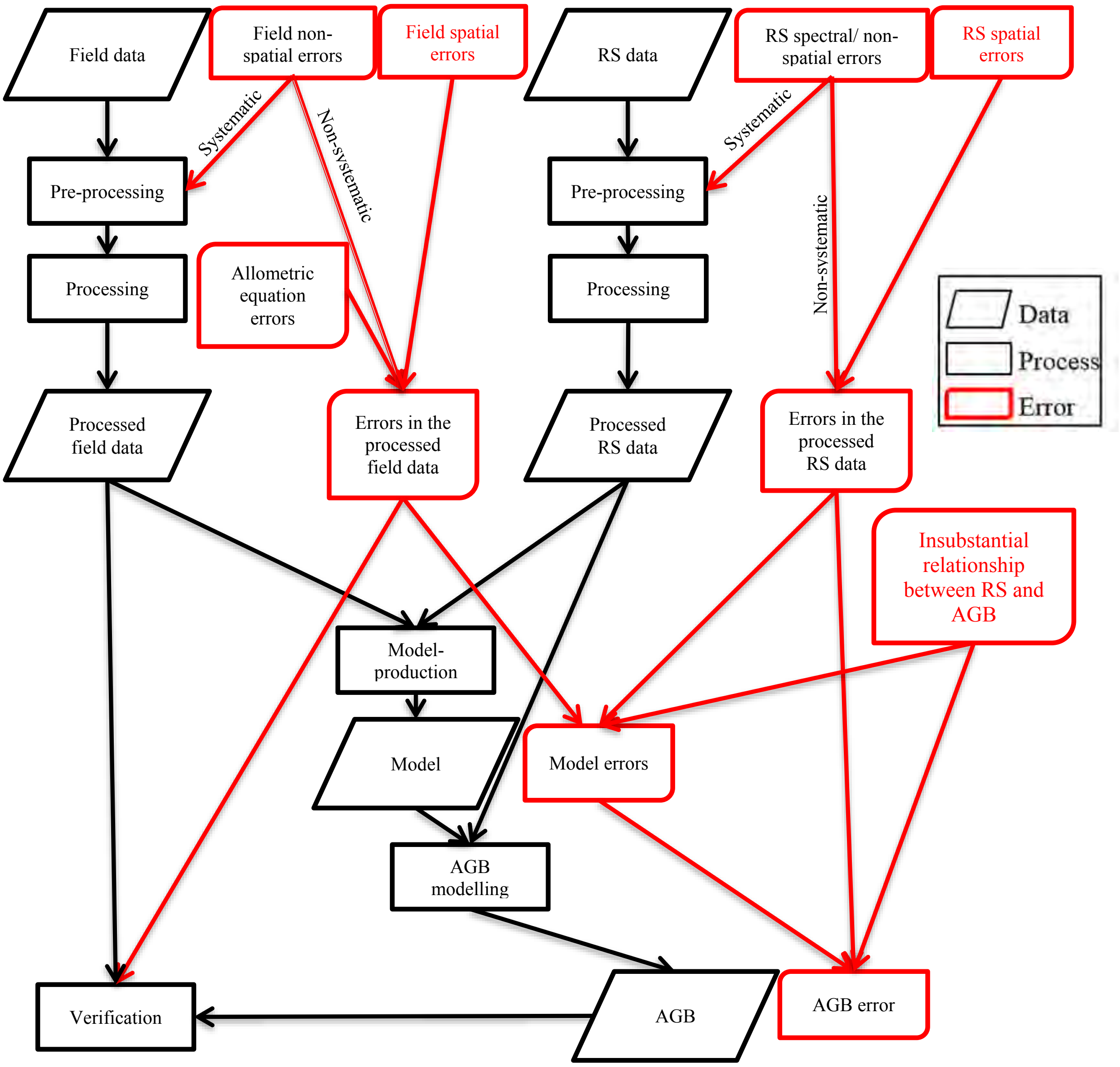


Figure 7-2: The process of AGB assessment and the comprehensive error propagation through it according to this study results. The assessment steps are shown in black, and the error propagation steps are shown in red.

In addition, the study tried to accurately represent the sophisticated possibilities of AGB assessment systems in general with representative practical scenarios. The error propagation tool was applied to several varied scenarios and showcased the efficiency of applying a Monte Carlo method. A new technique was designed and implemented in order to overcome the computational cost limitation of previous studies. The technique was based on dividing the analysis into two stages. The first stage is applying the analysis to the fieldwork areas, and the second stage is modelling the errors for the first stage and generalizing the resulting error model to estimate errors over the whole area of interest. Therefore, it can be claimed that this study is externally valid for other AGB assessment models. In addition, due to the new technique design, this methodology is applicable to the wider coverage AGB studies whenever the AGB modelling applicable. Moreover, this methodology is not limited to the AGB applications. It can be applied to any RS system that uses RS correlation to fieldwork-based data.

7.4 Future Work

With TLS based fieldwork improvements, such as faster and more spatially detailed information compared to the direct tree DBH measurements, there is a need to develop H-based allometric equations. The error analysis for the allometric equation error shows that the relative error of tree height measurements extracted from the TLS point cloud can be less than the relative error of the DBH extracted from the same point cloud. However, the accuracy can be reduced when the DBH allometric equation is converted to an H based equation.

The plot based fieldwork had a limited spatial resolution information. The idea of distributing each plot to a number of smaller sized field samples can be extended in the future, providing a higher number of samples for the model training compared to the number of plots over the same fieldwork area and with higher spatial resolution. Also, distributing each plot to a number of samples, can provide maps that are more detailed; especially when very-high resolution RS data are utilized. However, the sampling strategy that was based on a large number of small area samples requires a detailed spatial distribution of the field AGB maps. Therefore, improving this technique will require improving the spatial accuracy of the fieldwork methods.

The relationship between branch AGB maps and total AGB needs to be explored further. The third and fourth scenarios showed that when small-area field samples were used, the branch maps showed a higher correlation to the very-high resolution RS data. However, the conversion from the branch maps to the total AGB maps can be a large source of uncertainty unless a well-established conversion is provided.

This work can be expanded to be generalized for AGB and other remote sensing applications especially that the accuracy analysis applications are not limited to the AGB applications. The accuracy prediction tool (Chapter 3) can be applied on any application when enough literature is available. A better precision result is expected if the application bases on less variables or when more published detailed studies are available. The error sources classification (Chapter 6) is applicable on wide range of remote sensing applications. Therefore, the error propagation analysis tool (Chapter 5) can be applied on many uses that correlate field information and remote sensing data.

Chapter 8 Conclusions

The overall research questions focused on what are the significant error components in AGB assessment systems, how to propagate errors to the results, and what influence each error component has on the assessed AGB. In answering these, the strengths of this research, compared to previous studies, are that:

- The research was based on a careful classification of the expected error sources. A preliminary list of errors was formed, based on the literature review, and then extended to include the spatial errors within both the field AGB maps and remote sensing data that were shown to have a significant influence on the total error.

- A new, fast, fieldwork approach was developed that provides very high spatial resolution (5 cm) outputs. The approach combines the high precession (3D accuracy of a few millimetres) AGB distribution assessment provided by terrestrial laser scanning data (with 3d accuracy of few millimetres for the P20) with the high accuracy existing allometric equations that are based on destructive tests for a large number of sampled trees. The AGB maps were separated into trunk and branch AGB maps with a higher correlation achieved when branch AGB maps were correlated against very-high resolution remote sensing data, compared to the total AGB.

- The analysis tool was sufficiently flexible that it could propagate both non-spatial (e.g. spectral) and spatial errors, for all field and remote sensing data, to the model and hence results. It also dealt with both pixel-based and object-based analysis techniques and could analyse a single error component or combination of all components, providing error models that gave an error analysis for each pixel or object.

- The new error analysis technique, developed in this study, can derive general error equations (based on the fieldwork sites) that can be applied to the whole study area. This reduced the computational cost of stochastic simulation and allowed the processing to be accelerated within a parallel computing framework with GPU capabilities. Therefore, it was possible to make use of the widely applicable, and flexible, Monte Carlo approach for fieldwork sites. The resulting error maps visually illustrate areas where data needs to be improved and/or more fieldwork should be undertaken.

Regarding areas for further work:

- While a neural network system was trained to emulate human decision making, providing as an optimal trade-off between cost and accuracy, it might take a long time before enough studies are published to get sufficient training data for accurate predictions. In addition, the practical scenarios showed that the total error and the error components varied based on the system components. Therefore, predicting the final accuracy remains complicated.

- More tree-height based allometric equations need to be provided in order to make use of the higher relative accuracy of TLS based tree height compared to the relative accuracy of TLS based tree diameter.

- The relationship between the branch and total AGB needs to be explored further. When small-area field samples were used, the branch maps showed a higher correlation to the very high spatial resolution remote sensing data. However, the conversion from the branch to the total AGB maps remains a significant source of uncertainty.

- It is recommended that future AGB assessment studies/projects report the input details and the accuracy of their outputs, which can then be included within future versions of the accuracy prediction tool.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J. and Devin, M. (2016) 'Tensorflow: Large-scale machine learning on heterogeneous distributed systems', *arXiv preprint arXiv:1603.04467*.

Achard, F., Stibig, H.-J., Eva, H. D., Lindquist, E. J., Bouvet, A., Arino, O. and Mayaux, P. (2010) 'Estimating tropical deforestation from Earth observation data', *Carbon Management,* 1(2), pp. 271-287.

Agency, E. (2016) *Environment Agency LIDAR data. Technical Note*: Environment Agency.

Ahlstrom, A., Schurgers, G., Arneth, A. and Smith, B. (2012) 'Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections', *Environmental Research Letters,* 7(4), pp. 044008.

Ahmed, R. (2012) *Accuracy of biomass and structure estimates from radar and lidar.* PhD., University of Massachusetts.

Ahmed, R., Siqueira, P., Hensley, S. and Bergen, K. (2013) 'Uncertainty of Forest Biomass Estimates in North Temperate Forests Due to Allometry: Implications for Remote Sensing', *Remote Sensing,* 5(6), pp. 3007-3036.

Alberti, G., Candido, P., Peressotti, A., Turco, S., Piussi, P. and Zerbi, G. (2005) 'Aboveground biomass relationships for mixed ash (Fraxinus excelsior L. and Ulmus glabra Hudson) stands in Eastern Prealps of Friuli Venezia Giulia (Italy)', *Annals of Forest Science,* 62(8), pp. 831-836.

Andrae, R. (2010) 'Error estimation in astronomy: A guide', *arXiv preprint arXiv:1009.2755*.

Andrae, R., Schulze-Hartung, T. and Melchior, P. (2010) 'Dos and don'ts of reduced chi-squared', *arXiv preprint arXiv:1012.3754*.

Argamosa, R. J. L., Blanco, A. C., Baloloy, A. B., Candido, C. G., Dumalag, J. B. L. C., Dimapilis, L. L. C., & Paringit, E. C. (2018). 'Modelling Above-ground Biomass of Mangrove Forest Using Sentinel-1 Imagery'. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences,* 4(3).

Askne, J. and Santoro, M. (2012) *Experiences in Boreal Forest Stem Volume Estimation from Multitemporal C-Band InSAR.* INTECH Open Access Publisher.

Askne, J. I. H., Fransson, J. E. S., Santoro, M., Soja, M. J. and Ulander, L. M. H. (2013) 'Model-based biomass estimation of a hemi-boreal forest from multitemporal TanDEM-X acquisitions', *Remote Sensing,* 5(11), pp. 5574-5597.

Atkinson, P., Sargent, I., Foody, G. and Williams, J. (2007) 'Exploring the geostatistical method for estimating the signal-to-noise ratio of images', *Photogrammetric Engineering & Remote Sensing,* 73(7), pp. 841-850.

Attarchi, S. and Gloaguen, R. (2014) 'Improving the Estimation of Above Ground Biomass Using Dual Polarimetric PALSAR and ETM+ Data in the Hyrcanian Mountain Forest (Iran)', *Remote Sensing,* 6(5), pp. 3693-3715.

Baghdadi, N., Le Maire, G., Fayad, I., Bailly, J. S., Nouvellon, Y., Lemos, C. and Hakamada, R. (2014) 'Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in Eucalyptus plantations in Brazil', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 7(1), pp. 290-299.

Bailey, S. (2012) 'Principal component analysis with noisy and/or missing data', *Publications of the Astronomical Society of the Pacific,* 124(919), pp. 1015.

Bao, Y., Gao, W. and Gao, Z. (2009) 'Estimation of winter wheat biomass based on remote sensing data at various spatial and spectral resolutions', *Frontiers of Earth Science in China,* 3(1), pp. 118-128.

Becek, K. (2010) *Biomass Representation in Synthetic Aperture Radar Interferometry Data Sets.* postdoctoral, The University of Brunei Darussalam, Darussalam

Berger, A., Gschwantner, T., McRoberts, R. E. and Schadauer, K. (2014) 'Effects of measurement errors on individual tree stem volume estimates for the Austrian National Forest Inventory', *Forest Science,* 60(1), pp. 14-24.

Berra, E. F., Fontana, D. C. and Pereira, R. S. (2014) 'Accuracy of forest stem volume estimation by TM/Landsat Imagery with different geometric and atmospheric correction methods', *International Journal of Applied,* 4(3).

Bodansky, D. (2016) 'The Paris climate change agreement: a new hope?', *American Journal of International Law,* 110(2), pp. 288-319.

Bortolot, Z. J. and Wynne, R. H. (2005) 'Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data', *ISPRS Journal of Photogrammetry and Remote Sensing,* 59(6), pp. 342-360.

Bollandsas, O. M., Ene, L. T., Gobakken, T., & Næsset, E. (2018). Estimation of biomass change in montane forests in Norway along a 1200 km latitudinal gradient using airborne laser scanning: a comparison of direct and indirect prediction of change under a model-based inferential approach. *Scandinavian Journal of Forest Research, 33(2),* 155-165.

Bourbigot, M., Johnson, H. and Piantanida, R. (2016) 'Sentinel-1 Product Definition', *MDA Document Number: SEN-RS-52-7440.*

Breidenbach, J., Antón-Fernández, C., Petersson, H., McRoberts, R. E. and Astrup, R. (2014) 'Quantifying the model-related variability of biomass stock and change estimates in the Norwegian National Forest Inventory', *Forest Science,* 60(1), pp. 25-33.

Breiman, L. and Cutler, A. (2003) 'Random forests manual v4', *Technical report*: UC Berkel.

Bresnahan, P., Powers, R. and HenryVazquez, L. 'WorldView-3 Geolocation Accuracy and Band Co-Registration Analysis', *Joint Agency Commercial Imagery Evaluation (JACIE) Workshop*: USGS.

Brolly, G. and Kiraly, G. (2009) 'Algorithms for stem mapping by means of terrestrial laser scanning', *Acta Silvatica et Lignaria Hungarica,* 5, pp. 119-130.

Caflisch, R. E. (1998) 'Monte carlo and quasi-monte carlo methods', *Acta numerica,* 7, pp. 1-49.

Cai, S., Kang, X. and Zhang, L. (2013) 'Allometric models for aboveground biomass of ten tree species in northeast China', *Annals of Forest Research,* 56(1), pp. 105-122.

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., Culvenor, D., Avitabile, V., Disney, M. and Armston, J. (2015) 'Nondestructive estimates of above- ground biomass using terrestrial laser scanning', *Methods in Ecology and Evolution,* 6(2), pp. 198-208.

Castel, T., Guerra, F., Caraglio, Y. and Houllier, F. (2002) 'Retrieval biomass of a large Venezuelan pine plantation using JERS-1 SAR data. Analysis of forest structure impact on radar signature', *Remote Sensing of Environment,* 79(1), pp. 30-41.

Chakraborty, T., Saha, S. and Reif, A. (2016) 'Biomass equations for European beech growing on dry sites', *iForest-Biogeosciences and Forestry,* 9(5), pp. 751.

Chaudhary, B. (2015) *Tkinter GUI Application Development Blueprints.* Packt Publishing Ltd.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S. and Perez, R. (2004) 'Error propagation and scaling for tropical forest biomass estimates', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* 359(1443), pp. 409-420.

Chen, C. W. (2001) *Statistical-cost network-flow approaches to two-dimensional phase unwrapping for radar interferometry.* Stanford University.

Chen, G. and Hay, G. J. (2011) 'A support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and QuickBird data', *Photogrammetric Engineering \& Remote Sensing,* 77(7), pp. 733-741.

Chen, Q. (2013) 'LiDAR Remote Sensing of Vegetation Biomass', *Remote Sensing of Natural Resources, CRC Press: Taylor & Francis, Group*, pp. 399-420.

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

Chen, Q. and McRoberts, R. 'Statewide mapping and estimation of vegetation aboveground biomass using airborne Lidar'. *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*: IEEE, 4442-4444.

Chen, W., Zorn, P., Chen, Z., Latifovic, R., Zhang, Y., Li, J., Quirouette, J., Olthof, I., Fraser, R. and Mclennan, D. (2013) 'Propagation of errors associated with scaling foliage biomass from field measurements to remote sensing data over a northern Canadian national park', *Remote Sensing of Environment,* 130, pp. 205-218.

Cocard, M., Geiger, A., Meier, E. and Nuesch, D. 'X-SAR geometric error budget analysis'. *Geoscience and Remote Sensing Symposium, 1991. IGARSS'91. Remote Sensing: Global Monitoring for Earth Management., International*: IEEE, 247-251.

Colgan, M. S., Asner, G. P. and Swemmer, T. (2013) 'Harvesting tree biomass at the stand level to assess the accuracy of field and airborne biomass estimation in savannas', *Ecological Applications,* 23(5), pp. 1170-1184.

Congalton, R. G. (1991) 'Remote Sensing and Geographic Information System Data Integration: Error Sources and', *Photogrammetric Engineering & Remote Sensing,* 57(6), pp. 677-687.

Congedo, L. 2013. Semi-automatic classification plugin for QGIS. Tech. rep. Rome: Sapienza.

Copernicus *Open Access Hub.*: Scihub.copernicus.eu. . Available at: https://scihub.copernicus.eu/ (Accessed: 1 Jan. 2016 2016).

Côté, J.-F., Fournier, R. A. and Egli, R. (2011) 'An architectural model of trees to estimate forest structural attributes using terrestrial Lidar', *Environmental Modelling & Software,* 26(6), pp. 761-777.

Coulibaly, L., Migolet, P., Adegbidi, H. G., Fournier, R. and Hervet, E. 'Mapping aboveground forest biomass from IKONOS satellite image and multi-source geospatial data using neural networks and a kriging interpolation'. *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*: IEEE, III-298-III-301.

Covello, F., Battazza, F., Coletta, A., Lopinto, E., Fiorentino, C., Pietranera, L., Valentini, G. and Zoffoli, S. (2010) 'COSMO-SkyMed an existing opportunity for observing the Earth', *Journal of Geodynamics,* 49(3), pp. 171-180.

Culjak, I., Abram, D., Pribanic, T., Dzapo, H. and Cifrek, M. 'A brief introduction to OpenCV'. *MIPRO, 2012 proceedings of the 35th international convention*: IEEE, 1725-1730.

Curran, P. J. and Dungan, J. L. (1989) 'Estimation of signal-to-noise: a new procedure applied to AVIRIS data', *IEEE Transactions on Geoscience and Remote sensing,* 27(5), pp. 620-628.

Czapla-Myers, J., McCorkel, J., Anderson, N., Thome, K., Biggar, S., Helder, D., Aaron, D., Leigh, L. and Mishra, N. (2015) 'The ground-based absolute radiometric calibration of Landsat 8 OLI', *Remote Sensing,* 7(1), pp. 600-626.

Daliakopoulos, I. N., Grillakis, E. G., Koutroulis, A. G. and Tsanis, I. K. (2009) 'Tree crown detection on multispectral VHR satellite imagery', *Photogrammetric Engineering & Remote Sensing,* 75(10), pp. 1201-1211.

Dalponte, M., Bruzzone, L. and Gianelle, D. (2011) 'A system for the estimation of single-tree stem diameter and volume using multireturn Lidar data', *Geoscience and Remote Sensing, IEEE Transactions on,* 49(7), pp. 2479-2490.

Dalponte, M., Coops, N. C., Bruzzone, L. and Gianelle, D. (2009) 'Analysis on the use of multiple returns Lidar data for the estimation of tree stems volume', *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* 2(4), pp. 310-318.

Dassot, M., Barbacci, A., Colin, A., Fournier, M. E. R. I. E. M. and Constant, T. H. I. E. R. Y. (2010) 'Tree architecture and biomass assessment from terrestrial Lidar measurements: a case study for some Beech treesFagus Sylvatica', *Silvilaser,Freiburg*, pp. 206-215.

Del Frate, F. and Solimini, D. (2004) 'On neural network algorithms for retrieving forest biomass from SAR data', *Geoscience and Remote Sensing, IEEE Transactions on,* 42(1), pp. 24-34.

Deutscher, J., Perko, R., Gutjahr, K., Hirschmugl, M. and Schardt, M. (2013) 'Mapping tropical rainforest canopy disturbances in 3D by COSMO-skymed spotlight InSAR-stereo data to detect areas of forest degradation', *Remote Sensing,* 5(2), pp. 648-663.

Dong, J., Kaufmann, R. K., Myneni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., Buermann, W., Alexeyev, V. and Hughes, M. K. (2003) 'Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks', *Remote Sensing of Environment,* 84(3), pp. 393-410.

Dube, T. and Mutanga, O. (2015) 'Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa', *ISPRS Journal of Photogrammetry and Remote Sensing,* 101, pp. 36-46.

Dubois-Fernandez, P., Champion, I., Guyon, D., Cantalloube, H., Garestier, F., Dupuis, X. and Bonin, G. 'Forest biomass estimation from P-band high incidence angle data'. *ESA Special Publication*, 8.

Ebrahimi, H., Aslebagh, S. and Jones, L. 'Use of Monte Carlo simulation in remote sensing data analysis'. *Southeastcon, 2013 Proceedings of IEEE*, 1-4.

Eckert, S. (2012) 'Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data', *Remote Sensing,* 4(4), pp. 810-829.

Eckert, S., Ratsimba, H. R., Rakotondrasoa, L. O., Rajoelison, L. G. and Ehrensperger, A. (2011) 'Deforestation and forest degradation monitoring and assessment of biomass and carbon stock of lowland rainforest in the Analanjirofo region, Madagascar', *Forest Ecology and Management,* 262(11), pp. 1996-2007.

Edson, C. and Wing, M. G. (2011) 'Airborne Light Detection and Ranging (Lidar) for individual tree stem location, height, and biomass measurements', *Remote Sensing,* 3(11), pp. 2494-2528.

Efron, B. and Tibshirani, R. J. (1994) *An introduction to the bootstrap.* CRC press.

Englhart, S., Keuck, V. and Siegert, F. (2012) 'Modeling aboveground biomass in tropical forests using multi-frequency SAR data—A comparison of methods', *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* 5(1), pp. 298-306.

EPSG (2010) *OSGB 1936 to WGS 84 Petroleum*: EPSG Geodetic Parameter Registry. Available at: http://www.epsg-registry.org/report.htm?type=selection&entity=urn:ogc:def:coordinateOperation:EPSG::1314&reportDetail=long&style=urn:uuid:report-style:default-with-code&style_name=OGP%20Default%20With%20Code&title=Petroleum.

ESA, 2016, ‘Sentinel-2: Resolution and Swath’. Available at: https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath (Accessed: 5 Jan. 2018).

Evans, J. P., Scheffers, B. R., & Hess, M. (2014) 'Effect of laurel wilt invasion on redbay populations in a maritime forest community'. Biological invasions, 16(8), 1581-1588.

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

Feng, Y., Lu, D., Chen, Q., Keller, M., Moran, E., Nara dos-Santos, M., Luis Bolfe, E. and Batistella, M. (2017) 'Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon', *International Journal of Digital Earth*, pp. 1-21.

Ferraz, A., Saatchi, S., Mallet, C. e., ment and Meyer, V. (2016) 'Lidar detection of individual tree size in tropical forests', *Remote Sensing of Environment,* 183, pp. 318-333.

Forestry_Commission (2001) *National inventory of woodland and trees: England.* Forestry Commission.

Forestry_Commission (2002a) *National Inventory of Woodland and Trees- Cornwall*, United Kingdom: Forestry_Commission. Available at: https://www.forestry.gov.uk/pdf/cornwall.pdf/$FILE/cornwall.pdf.

Forestry_Commission (2002b) *National Inventory of Woodland and Trees- Devon*, United Kingdom: Forestry_Commission. Available at: https://www.forestry.gov.uk/pdf/devon.pdf/$FILE/devon.pdf.

Forrester, D. I., Tachauer, I., Annighoefer, P., Barbeito, I., Pretzsch, H., Ruiz-Peinado, R., Stark, H., Vacchiano, G., Zlatanov, T. and Chakraborty, T. (2017) 'Generalized biomass and leaf area allometric equations for European tree species incorporating stand structure, tree age and climate', *Forest Ecology and Management,* 396, pp. 160-175.

Foster, J. R., Kingdon, C. C. and Townsend, P. A. 'Predicting tropical forest carbon from EO-1 hyperspectral imagery in Noel Kempff Mercado National Park,Bolivia'. *Geoscience and Remote Sensing Symposium,2002. IGARSS'02. 2002 IEEE International*: IEEE, 3108-3110.

Fraser, R. and Li, Z. (2002) 'Estimating fire-related parameters in boreal forest using SPOT VEGETATION', *Remote Sensing of Environment,* 82(1), pp. 95-110.

Fu, Y., Lei, Y., Zeng, W., Hao, R., Zhang, G., Zhong, Q. and Xu, M. (2017) 'Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error', *Canadian Journal of Forest Research,* 47(8), pp. 1095-1103.

Gale, J., Bradshaw, J., Chen, Z., Garg, A., Gomez, D., Rogner, H., Simbeck, D. and Williams, R. (2005) 'Sources of CO_2', *Carbon dioxide capture and storage, Cambridge University Press, UK*, pp. 94.

Ge, S., Xu, M., Anderson, G. L. and Carruthers, R. I. (2007) 'Estimating Yellow Starthistle (Centaurea solstitialis) leaf area index and aboveground biomass with the use of hyperspectral data'.

Ghasemi, N., Sahebi, M. R. and Mohammadzadeh, A. (2013) 'Biomass Estimation of a Temperate Deciduous Forest Using Wavelet Analysis', *Geoscience and Remote Sensing, IEEE Transactions on,* 51(2), pp. 765-776.

Ghasemi, N., Sahebi, M. R., Mohammadzadeh, A. and Bayani, R. (2011) 'Mapping biomass of a temperate deciduous forest using SAR data', *Geomatics 90.*

Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S. and Marinaro, M. (2010) 'Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs', *Geoscience and Remote Sensing, IEEE Transactions on,* 48(10), pp. 3769-3779.

Gjertsen, A. K. (2007) 'Accuracy of forest mapping based on Landsat TM data and a k-NN-based method', *Remote Sensing of Environment,* 110(4), pp. 420-430.

Goering, D. J., Chen, H., Hinzman, L. D. and Kane, D. L. (1995) 'Removal of terrain effects from SAR satellite imagery of Arctic tundra', *Geoscience and Remote Sensing, IEEE Transactions on,* 33(1), pp. 185-194.

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R. A. and Sun, M. (2009) 'Mapping and monitoring carbon stocks with satellite observations: a comparison of methods', *Carbon balance and management,* 4(1), pp. 2.

Gonçalves, F., Treuhaft, R., Law, B., Almeida, A., Walker, W., Baccini, A., dos Santos, J. R. and Graça, P. (2017) 'Estimating Aboveground Biomass in Tropical Forests: Field Methods and Error Analysis for the Calibration of Remote Sensing Observations', *Remote Sensing,* 9(1), pp. 47.

Goulden, T., Hopkinson, C., Jamieson, R. and Sterling, S. (2016) 'Sensitivity of DEM, slope, aspect and watershed attributes to Lidar measurement uncertainty', *Remote Sensing of Environment,* 179, pp. 23 - 35.

Gu, Y., Wylie, B. K., Boyte, S. P., Picotte, J., Howard, D. M., Smith, K. and Nelson, K. J. (2016) 'An optimal sample data usage strategy to minimize overfitting and underfitting effects in regression tree models based on remotely-sensed data', *Remote Sensing,* 8(11), pp. 943.

Hackenberg, J., Morhart, C., Sheppard, J., Spiecker, H. and Disney, M. (2014) 'Highly accurate tree models derived from terrestrial laser scan data: A method description', *Forests,* 5(5), pp. 1069-1105.

Hall, F. G., Bergen, K., Blair, J. B., Dubayah, R., Houghton, R., Hurtt, G., Kellndorfer, J., Lefsky, M., Ranson, J. and Saatchi, S. (2011) 'Characterizing 3D vegetation structure from space: Mission requirements', *Remote Sensing of Environment,* 115(11), pp. 2753-2775.

Hall, R. J., Skakun, R. S., Arsenault, E. J., & Case, B. S. (2006). 'Modelling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand' volume. *Forest ecology and management*, 225(1-3), 378-390.

Han, Z., Tian, X., Jiang, H., Ling, F. and Yan, M. 'Uncertainty analysis of forest above-ground biomass increments in southern Qilian Mountains'. *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, 4430-4433.

Handcock, R., Gillespie, A., Cherkauer, K., Kay, J., Burges, S. and Kampf, S. (2006) 'Accuracy and uncertainty of thermal-infrared remote sensing of stream temperatures at multiple spatial scales', *Remote Sensing of Environment,* 100(4), pp. 427-440.

Hansen, E. H., Gobakken, T., Bollandsås, O. M., Zahabu, E. and Næsset, E. (2015) 'Modeling aboveground biomass in dense tropical submontane rainforest using airborne laser scanner data', *Remote Sensing,* 7(1), pp. 788-807.

He, Q., Chen, E., An, R. and Li, Y. (2013) 'Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest', *Forests,* 4(4), pp. 984-1002.

Hein, S. and Spiecker, H. (2008) 'Crown and tree allometry of open-grown ash (Fraxinus excelsior L.) and sycamore (Acer pseudoplatanus L.)', *Agroforestry systems,* 73(3), pp. 205-218.

Heiskanen, J. (2006) 'Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data', *International Journal of Remote Sensing,* 27(6), pp. 1135-1158.

Hensley, S., Oveisgharan, S., Saatchi, S., Simard, M., Ahmed, R. and Haddad, Z. (2014) 'An error model for biomass estimates derived from polarimetric radar backscatter', *IEEE Transactions on Geoscience and Remote Sensing,* 52(7), pp. 4065-4082.

Heuvelink, G. B. M. (1998) *Error propagation in environmental modelling with GIS.* CRC Press.

Heuvelink, G. B. M. (1999) 'Propagation of error in spatial modelling with GIS', *Geographical information systems,* 1, pp. 207-217.

Hill, T. C., Williams, M., Bloom, A. A., Mitchard, E. T. and Ryan, C. M. (2013) 'Are inventory based and remotely sensed above-ground biomass estimates consistent?', *PloS one,* 8(9), pp. e74170.

Hirata, Y., Tabuchi, R., Patanaponpaiboon, P., Poungparn, S., Yoneda, R. and Fujioka, Y. (2014) 'Estimation of aboveground biomass in mangrove forests using high-resolution satellite data', *Journal of forest research,* 19(1), pp. 34-41.

Hodgson, M. E. and Bresnahan, P. (2004) 'Accuracy of Airborne Lidar-Derived Elevation', *Photogrammetric Engineering & Remote Sensing,* 70(3), pp. 331-339.

Holopainen, M., Haapanen, R., Karjalainen, M., Vastaranta, M., Hyyppä, J., Yu, X., Tuominen, S. and Hyyppä, H. (2010) 'Comparing accuracy of airborne laser scanning and TerraSAR-X radar images in the estimation of plot-level forest variables', *Remote Sensing,* 2(2), pp. 432-445.

Houghton, R. (2005) 'Aboveground forest biomass and the global carbon balance', *Global Change Biology,* 11(6), pp. 945-958.

Hunt Jr, E. R., Fahnestock, J., Kelly, R. D., Welker, J. M., Reiners, W. A. and Smith, W. K. (2002) 'Carbon sequestration from remotely-sensed NDVI and net ecosystem exchange', *From laboratory spectroscopy to remotely sensed spectra of terrestrial ecosystems*: Springer, pp. 161-174.

Imhoff, M. L., Johnson, P., Holford, W., Hyer, J., May, L., Lawrence, W. and Harcombe, P. (2000) 'BioSAR< TM>: An inexpensive airborne VHF multiband SAR system for vegetation biomass measurement', *IEEE transactions on geoscience and remote sensing,* 38(3), pp. 1458-1462.

Israelsson, H., Ulander, L. M., Askne, J. I., Fransson, J. E., Frolind, P.-O., Gustavsson, A. and Hellsten, H. (1997) 'Retrieval of forest stem volume using VHF SAR', *Geoscience and Remote Sensing, IEEE Transactions on,* 35(1), pp. 36-40.

Jacobsen, K. 'Characteristics of very high resolution optical satellites for topographic mapping'. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences:[ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging For Geospatial Information] 38-4 (2011), Nr. W19*: Göttingen: Copernicus GmbH, 137-142.

Jager, H. I. and King, A. W. (2004) 'Spatial uncertainty and ecological models', *Ecosystems,* 7(8), pp. 841-847.

Jandl, R., Lindner, M., Vesterdal, L., Bauwens, B., Baritz, R., Hagedorn, F., Johnson, D. W., Minkkinen, K. and Byrne, K. A. (2007) 'How strongly can forest management influence soil carbon sequestration?', *Geoderma,* 137(3), pp. 253-268.

Jung, H.-S., Lu, Z. and Zhang, L. (2013) 'Feasibility of along-track displacement measurement from Sentinel-1 interferometric wide-swath mode', *IEEE Transactions on Geoscience and Remote Sensing,* 51(1), pp. 573-578.

Kankare, V., Holopainen, M., Vastaranta, M., Liang, X., Yu, X., Kaartinen, H., Kukko, A. and Hyyppä, J. (2017) 'Outlook for the Single-Tree-Level Forest Inventory in Nordic Countries', *The Rise of Big Spatial Data*: Springer, pp. 183-195.

Karsoliya, S. (2012) 'Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture', *International Journal of Engineering Trends and Technology,* 3(6), pp. 713-717.

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

Kellndorfer, J., Walker, W., Pierce, L., Dobson, C., Fites, J. A., Hunsaker, C., Vona, J. and Clutter, M. (2004) 'Vegetation height estimation from shuttle radar topography mission and national elevation datasets', *Remote sensing of Environment,* 93(3), pp. 339-358.

Ketterings, Q. M., Coe, R., van Noordwijk, M., Palm, C. A. and others (2001) 'Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests', *Forest Ecology and management,* 146(1), pp. 199-209.

Khati, U., Singh, G. and Mohanty, S. 'Forest biophysical parameter estimation using space-borne bistatic PolInSAR measurements'. *SPIE Asia-Pacific Remote Sensing*, 98771B-98771B.

Koch, B. (2010) 'Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment', *ISPRS Journal of Photogrammetry and Remote Sensing,* 65(6), pp. 581-590.

Krause, K. S. (2008) 'WorldView-1 pre and post-launch radiometric calibration and early on-orbit characterization', *Optical Engineering+ Applications*, pp. 708116-708116.

Kyriou, A., and Nikolakopoulos, K. (2018). 'Assessing the suitability of Sentinel-1 data for landslide mapping'*, European Journal of Remote Sensing, 51(1)*, 402-411.

Lau, S. W. (2011) 'Comparison and Fusion of space borne L-, C-and X-Band SAR Images for Damage Identification in the 2008 Sichuan Earthquake'.

Laurin, G. V., Puletti, N., Hawthorne, W., Liesenberg, V., Corona, P., Papale, D., Chen, Q. and Valentini, R. (2016) 'Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data', *Remote Sensing of Environment,* 176, pp. 163-176.

Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., Manning, A. C., Boden, T. A., Tans, P. P., Houghton, R. A., Keeling, R. F., Alin, S., Andrews, O. D., Anthoni, P., Barbero, L., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Currie, K., Delire, C., Doney, S. C., Friedlingstein, P., Gkritzalis, T., Harris, I., Hauck, J., Haverd, V., Hoppema, M., Klein Goldewijk, K., Jain, A. K., Kato, E., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Melton, J. R., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M.

S., Nakaoka, S. I., O'Brien, K., Olsen, A., Omar, A. M., Ono, T., Pierrot, D., Poulter, B., Rödenbeck, C., Salisbury, J., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Sutton, A. J., Takahashi, T., Tian, H., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J. and Zaehle, S. (2016) 'Global Carbon Budget 2016', *Earth Syst. Sci. Data,* 8(2), pp. 605-649.

Le Quéré, C., Moriarty, R., Andrew, R., Peters, G., Ciais, P., Friedlingstein, P., Jones, S., Sitch, S., Tans, P. and Arneth, A. (2014) 'Global carbon budget 2014', *Earth System Science Data Discussions,* 7(2), pp. 521-610.

Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J. and Boden, T. (2015) 'Global carbon budget 2015', *Earth System Science Data,* 7(2), pp. 349-396.

Le Quéré, C., Peters, G., Andres, R., Andrew, R., Boden, T., Ciais, P., Friedlingstein, P., Houghton, R., Marland, G. and Moriarty, R. (2013) 'Global carbon budget 2013', *Earth System Science Data Discussions,* 6(2).

Le Toan, T., Quegan, S., Davidson, M., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S. and Shugart, H. (2011) 'The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle', *Remote sensing of environment,* 115(11), pp. 2850-2860.

Leboeuf, A., Beaudoin, A., Fournier, R., Guindon, L., Luther, J. and Lambert, M.-C. (2007) 'A shadow fraction method for mapping biomass of northern boreal black spruce forests using QuickBird imagery', *Remote Sensing of Environment,* 110(4), pp. 488-500.

Lee, J. (1993) 'New Monte Carlo algorithm: Entropic sampling', *Phys. Rev. Lett.,* 71, pp. 211-214.

Lever, J., Krzywinski, M. and Altman, N. (2016) 'Points of significance: Model selection and overfitting', *Nature Methods,* 13(9), pp. 703-704.

Li, M., Tan, Y., Pan, J. and Peng, S. (2008) 'Modeling forest aboveground biomass by combining spectrum, textures and topographic features', *Frontiers of Forestry in China,* 3(1), pp. 10-15.

Li, W., Chen, E., Li, Z., Zhang, W. and Jiang, C. (2016) 'Assessing Performance of Tomo-SAR and Backscattering Coefficient for Hemi-Boreal Forest Aboveground Biomass Estimation', *Journal of the Indian Society of Remote Sensing,* 44(1), pp. 41-48.

Lim, K. S. and Treitz, P. 'Estimation of aboveground forest biomass using airborne scanning discrete return Lidar in Douglas-fir'. *Proc. of the ISPRS working group VII/2 "Laser-Scanners for Forest and Landscape, Assessment*, 149-152.

Lim, K. S. and Treitz, P. M. (2004b) '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), pp. 558-570.

Liu, J., Liang, X., Hyyppä, J., Yu, X., Lehtomäki, M., Pyörälä, J., Zhu, L., Wang, Y. and Chen, R. (2017) 'Automated matching of multiple terrestrial laser scans for stem mapping without the use of artificial references', *International Journal of Applied Earth Observation and Geoinformation,* 56, pp. 13-23.

Liu, Q., Chen, C., Zhang, Y. and Hu, Z. (2011) 'Feature selection for support vector machines with RBF kernel', *Artificial Intelligence Review,* 36(2), pp. 99-115.

Lo, E. (2005) 'Gaussian error propagation applied to ecological data: post-ice-storm-downed woody biomass', *Ecological monographs,* 75(4), pp. 451-466.

Lokupitiya, E., Lefsky, M. and Paustian, K. (2010) 'Use of AVHRR NDVI time series and ground-based surveys for estimating county-level crop biomass', *International Journal of Remote Sensing,* 31(1), pp. 141-158.

Lophaven, S. N., Nielsen, H. B. and Søndergaard, J. (2002) *DACE-A Matlab Kriging toolbox, version 2.0.*

Lu, D. (2006) 'The potential and challenge of remote sensing- based biomass estimation', *International journal of remote sensing,* 27(7), pp. 1297-1328.

Lu, D. and Batistella, M. (2005) 'Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon', *Acta Amazonica,* 35(2), pp. 249-257.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G. and Saah, D. (2012) 'Aboveground forest biomass estimation with Landsat and lidar data and uncertainty analysis of the estimates', *International Journal of Forestry Research,* 2012.

Lu, D., Mausel, P., Brondizio, E. and Moran, E. (2002) 'Above-ground biomass estimation of successional and mature forests using TM images in the Amazon Basin', *Advances in Spatial Data Handling*: Springer, pp. 183-196.

Lucas, R., Armston, J., Fairfax, R., Fensham, R., Accad, A., Carreiras, J., Kelley, J., Bunting, P., Clewley, D. and Bray, S. (2010) 'An evaluation of the ALOS PALSAR L-band backscatter—Above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure', *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* 3(4), pp. 576-593.

Lucas, R. M., Lee, A. and Bunting, P. J. (2008) 'Retrieving forest biomass through integration of CASI and LiDAR data', *International Journal of Remote Sensing,* 29(5), pp. 1553-1577.

Luther, J., Fournier, R., Piercey, D., Guindon, L. and Hall, R. (2006) 'Biomass mapping using forest type and structure derived from Landsat TM imagery', *International Journal of Applied Earth Observation and Geoinformation,* 8(3), pp. 173-187.

Marinelli, M., Corner, R. and Wright, G. 'A comparison of error propagation analysis techniques applied to agricultural models'. *Accuracy.*

Mauya, E. W., Hansen, E., Gobakken, T., Bollandsas, O., Malimbwi, R. and Næsset, E. (2015) 'Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania', *Carbon balance and management,* 10(1), pp. 10.

McGlinchy, J., van Aardt, J. A. N., Erasmus, B., Asner, G. P., Mathieu, R., Wessels, K., Knapp, D., Kennedy-Bowdoin, T., Rhody, H., Kerekes, J. P. and others (2014) 'Extracting structural vegetation components from small-footprint waveform lidar for biomass estimation in savanna ecosystems', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 7(2), pp. 480-490.

McKinney, W. (2011) 'pandas: a foundational Python library for data analysis and statistics', *Python for High Performance and Scientific Computing*, pp. 1-9.

McRoberts, R. E., Magnussen, S., Tomppo, E. O. and Chirici, G. (2011) 'Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data', *Remote Sensing of Environment,* 115(12), pp. 3165-3174.

McRoberts, R. E., N\ae, s., Erik and Gobakken, T. (2015) 'Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data', *Remote Sensing of Environment,* 163, pp. 13-22.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., Harmon, M. E. and McKee, W. A. (1999) 'Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon', *Remote Sensing of Environment,* 67(3), pp. 298-308.

Melaas, E. K., Friedl, M. A. and Zhu, Z. (2013) 'Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data', *Remote Sensing of Environment,* 132, pp. 176-185.

Mercer, B., Zhang, Q., Schwaebisch, M. and Denbina, M. 'Estimation of Forest Biomass from an Airborne Single-pass L-band PolInSAR System'. *ISPRS J. Photogramm. In Proc. of the PolInSAR Workshop*, 24-28.

Mette, T., Papathanassiou, K. and Hajnsek, I. 'Biomass estimation from polarimetric SAR interferometry over heterogeneous forest terrain'. *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International*: IEEE, 511-514.

Migolet, P., Coulibaly, L., Adegbidi, H. G. and Hervet, E. 'Utilization of neural networks for the estimation of aboveground forest biomass from Ikonos satellite image and multi-source geo-scientific data'. *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*: IEEE, 4339-4342.

Milano, F. 'A python-based software tool for power system analysis'. *Power and Energy Society General Meeting (PES), 2013 IEEE*: IEEE, 1-5.

Minh, D. H. T., Le Toan, T., Rocca, F., Tebaldini, S., Villard, L., R\'e, j.-M. e., chain, Maxime, Phillips, O. L., Feldpausch, T. R., Dubois-Fernandez, P., Scipal, K. and others (2016) 'SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana', *Remote Sensing of Environment,* 175, pp. 138-147.

Minh, D. H. T., Tebaldini, S., Rocca, F., Le Toan, T., Villard, L. and Dubois-Fernandez, P. C. (2015) 'Capabilities of BIOMASS tomography for investigating tropical forests', *IEEE Transactions on Geoscience and Remote Sensing,* 53(2), pp. 965-975.

Mitchard, E. T., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. and Brown, S. (2013) 'Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps', *Carbon balance and management, 8(1),* 10.

Mohan, S., Das, A., Haldar, D. and Maity, S. 'Monitoring and retrieval of vegetation parameter using multi-frequency polarimetric SAR data'. *Synthetic Aperture Radar (APSAR), 2011 3rd International Asia-Pacific Conference on*: IEEE, 1-4.

Molto, Q., Rossi, V. and Blanc, L. (2013) 'Error propagation in biomass estimation in tropical forests', *Methods in Ecology and Evolution,* 4(2), pp. 175-183.

Montesano, P. M., Nelson, R. F., Dubayah, R. O., Sun, G., Cook, B. D., Ranson, K. J. R., N\ae, s., E. and Kharuk, V. (2014) 'The uncertainty of biomass estimates from LiDAR and SAR across a boreal forest structure gradient', *Remote Sensing of Environment,* 154, pp. 398-407.

Morfitt, R., Barsi, J., Levy, R., Markham, B., Micijevic, E., Ong, L., Scaramuzza, P. and Vanderwerff, K. (2015) 'Landsat-8 Operational Land Imager (OLI) radiometric performance on-orbit', *Remote Sensing,* 7(2), pp. 2208-2237.

Moskal, L. M. and Zheng, G. (2011) 'Retrieving forest inventory variables with terrestrial laser scanning (TLS) in urban heterogeneous forest', *Remote Sensing,* 4(1), pp. 1-20.

Müller, H., & Quenzel, H. (1985) 'Information content of multispectral lidar measurements with respect to the aerosol size distribution' *Applied optics,* 24(5), 648-654.

Muukkonen, P. (2007) 'Generalized allometric volume and biomass equations for some tree species in Europe', *European Journal of Forest Research,* 126(2), pp. 157-166.

Næsset, E., Bollandsås, O. M., Gobakken, T., Solberg, S. and McRoberts, R. E. (2015) 'The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data', *Remote Sensing of Environment,* 168, pp. 252-264.

Namayanga, L. N. (2002) 'Estimating terrestrial carbon sequestered in aboveground woody biomass from remotely sensed data', *International Institute for Geo-Informatıon Science and Earth Observation Enschede, The Nederlands*.

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996) *Applied linear statistical models.* Irwin Chicago.

Nichol, J. E. and Sarker, M. R. (2011) 'Improved biomass estimation using the texture parameters of two high-resolution optical sensors', *Geoscience and Remote Sensing, IEEE Transactions on,* 49(3), pp. 930-948.

Njana, M. A. (2017) 'Indirect methods of tree biomass estimation and their uncertainties', *Southern Forests: a Journal of Forest Science,* 79(1), pp. 41-49.

Nolè, G., Murgante, B., Calamita, G., Lanorte, A., & Lasaponara, R. (2015). 'Evaluation of Urban Sprawl from space using open source technologies'. *Ecological informatics*, 26, 151-161.

OSGeo_Project *GDAL home page*. Available at: http://www.gdal.org/ (Accessed: 17/05/2017.

Pacala, S. W., Canham, C. D., Saponara, J., Silander Jr, J. A., Kobe, R. K. and Ribbens, E. (1996) 'Forest models defined by field measurements: estimation, error analysis and dynamics', *Ecological monographs,* 66(1), pp. 1-43.

Pairman, D. and McNeill, S. 'Efficient calculation in the map domain of SAR layover and shadow masks'. *Geoscience and Remote Sensing, 1997. IGARSS'97. Remote Sensing-A Scientific Vision for Sustainable Development., 1997 IEEE International*: IEEE, 2057-2059.

Pandas, (2017) Available at: http://pandas.pydata.org/pandas-docs/version/0.21/pandas.pdf (Accessed: 17/05/2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., (2011). 'Scikit-learn: Machine learning in Python'. *Journal of machine learning research*, pp.2825-2830.

Peng, X., Tan, W., Wang, Y., Hong, W. and Wu, Y. (2012) 'Convolution back-projection imaging algorithm for downward-looking sparse linear array three dimensional synthetic aperture radar', *Progress In Electromagnetics Research,* 130.

Pérez-Salicrup, D. R., & Barker, M. G. (2000) 'Effect of liana cutting on water potential and growth of adult Senna multijuga (Caesalpinioideae) trees in a Bolivian tropical forest'. *Oecologia, 124(4),* pp. 469-475.

Persson, H. J. (2016) 'Estimation of Boreal Forest Attributes from Very High Resolution Pléiades Data', *Remote Sensing,* 8(9), pp. 736.

Persson, H. J. and Fransson, J. E. S. (2016) 'Comparison between TanDEM-X and ALS based estimation of above ground biomass and tree height in boreal forests', *Scandinavian Journal of Forest Research,*, pp. 1-37.

Phuong, T. M., Lin, Z. and Altman, R. B. 'Choosing SNPs using feature selection'. *Computational Systems Bioinformatics Conference, 2005. Proceedings. 2005 IEEE*: IEEE, 301-309.

Poli, D., Remondino, F., Angiuli, E. and Agugiaro, G. (2015) 'Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction', *ISPRS Journal of Photogrammetry and Remote Sensing,* 100, pp. 35-47.

Popescu, S. C. and Hauglin, M. (2014) 'Estimation of biomass components by airborne laser scanning', *Forestry Applications of Airborne Laser Scanning*: Springer, pp. 157-175.

Pouliot, D., King, D. and Pitt, D. (2005) 'Development and evaluation of an automated tree detection delineation algorithm for monitoring regenerating coniferous forests', *Canadian Journal of Forest Research,* 35(10), pp. 2332-2345.

Pratihast, A. K. (2010) '3D tree modelling using mobile laser scanning data', *University of Twente, Faculty of Geo-Information and Earth Observation ITC, Enschede*.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M. and Lewis, P. (2013) 'Fast automatic precision tree models from terrestrial laser scanner data', *Remote Sensing,* 5(2), pp. 491-520.

Réjou- Méchain, M., Tanguy, A., Piponiot, C., Chave, J. and Hérault, B. (2017) 'BIOMASS: An R Package for estimating aboveground biomass and its uncertainty in tropical forests', *Methods in Ecology and Evolution*.

Reubens, B., Poesen, J., Danjon, F., Geudens, G., & Muys, B. (2007) 'The role of fine and coarse roots in shallow slope stability and soil erosion control with a focus on root system architecture: a review'. *Trees,* 21(4), pp. 385-402.

Richards, M. A. (2007) 'A beginner's guide to interferometric SAR concepts and signal processing [AESS Tutorial IV]', *IEEE Aerospace and Electronic Systems Magazine,* 22(9), pp. 5-29.

Risvik, H. (2008) '"PCA module for Python', *Internet: http://folk. uio. no/henninri/pca_module*.

Robinson, C., Saatchi, S., Neumann, M. and Gillespie, T. (2013) 'Impacts of spatial variability on aboveground biomass estimation from L-band radar in a temperate forest', *Remote Sensing,* 5(3), pp. 1001-1023.

Rosette, J., Cook, B., Suárez, J., North, P., Nelson, R. and Los, S. (2012) *Lidar remote sensing for biomass assessment.* INTECH Open Access Publisher.

Saatchi, S., Marlier, M., Chazdon, R. L., Clark, D. B. and Russell, A. E. (2011) 'Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass', *Remote Sensing of Environment,* 115(11), pp. 2836-2849.

Sabia, R., Camps, A., Vall-Llossera, M. and Talone, M. 'Extended ocean salinity error budget analysis within the SMOS mission'. *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*: IEEE, IV-534-IV-537.

Santi, E., Tarantino, C., Amici, V., Bacaro, G., Blonda, P., Borselli, L., Rossi, M., Tozzi, S. and Torri, D. (2014) 'Fine-scale spatial distribution of biomass using satellite images', *Journal of Ecology and The Natural Environment,* 6(2), pp. 75-86.

Santos, J. R., Freitas, C. C., Araujo, L. S., Dutra, L. V., Mura, J. C., Gama, F. F., Soler, L. S. and Sant'Anna, S. J. (2003) 'Airborne P-band SAR applied to the aboveground biomass studies in the Brazilian tropical rainforest', *Remote Sensing of Environment,* 87(4), pp. 482-493.

Sarker, L. R. and Nichol, J. E. (2011) 'Improved forest biomass estimates using ALOS AVNIR-2 texture indices', *Remote Sensing of Environment,* 115(4), pp. 968-977.

Scaramuzza, P. L., Markham, B. L., Barsi, J. A. and Kaita, E. (2004) 'Landsat-7 ETM+ on-orbit reflective-band radiometric characterization', *IEEE transactions on geoscience and remote sensing,* 42(12), pp. 2796-2809.

Schilling, A., Schmidt, A. and Maas, H.-G. 'Automatic tree detection and diameter estimation in terrestrial laser scanner point clouds'. *Proceedings of the 16th Computer Vision Winter Workshop*: Citeseer, 75-83.

Schlund, M., von Poncet, F., Kuntz, S., Schmullius, C. and Hoekman, D. H. (2015) 'TanDEM-X data for aboveground biomass retrieval in a tropical peat swamp forest', *Remote Sensing of Environment,* 158, pp. 255-266.

Schubert, A., Miranda, N., Geudtner, D. and Small, D. (2017) 'Sentinel-1A/B Combined Product Geolocation Accuracy', *Remote Sensing,* 9(6), pp. 607.

Schubert, A., Small, D., Meier, E., Miranda, N. and Geudtner, D. 'Spaceborne SAR product geolocation accuracy: A Sentinel-1 update'. *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*: IEEE, 2675-2678.

Schubert, A., Small, D., Miranda, N., Geudtner, D. and Meier, E. (2015) 'Sentinel-1a product geolocation accuracy: Commissioning phase results', *Remote Sensing,* 7(7), pp. 9431-9449.

Scipy, (2006) 'NumPy'. Available at: http://www.numpy.org/ (Accessed: 17/05/2017.

Seabold, S. and Perktold, J. (2010) 'Statsmodels: Econometric and statistical modeling with python'. *Proceedings of the 9th Python in Science Conference*, 61.

Shao, Z. and Zhang, L. (2016) 'Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China', *Sensors,* 16(6), pp. 834.

Shataee, S., Weinaker, H. and Babanejad, M. (2011) 'Plot-level Forest Volume Estimation Using Airborne Laser Scanner and TM Data, Comparison of Boosting and Random Forest Tree Regression Algorithms', *Procedia Environmental Sciences,* 7, pp. 68-73.

Sherrill, K., Lefsky, M., Battles, J., Waring, K. and Gonzalez, P. 'An Error Propagation Analysis of Estimates of Aboveground Biomass Estimates From Lidar Remote Sensing'. *AGU Fall Meeting Abstracts*, 07.

Smith, S. and Gilbert, J. (2003) 'National inventory of woodland and trees: Great Britain', *National inventory of woodland and trees: Great Britain*.

Smola, A. J. and Schölkopf, B. (2004) 'A tutorial on support vector regression', *Statistics and computing,* 14(3), pp. 199-222.

Soja, M. J., Sandberg, G. and Ulander, L. M. (2013) 'Regression-based retrieval of boreal forest biomass in sloping terrain using P-band SAR backscatter intensity data', *Geoscience and Remote Sensing, IEEE Transactions on,* 51(5), pp. 2646-2665.

Stelmaszczuk, G., Martyna, A., Rodriguez-Veiga, P., Ackermann, N., Thiel, C., Balzter, H. and Schmullius, C. (2015) 'Non-Parametric Retrieval of Aboveground Biomass in Siberian Boreal Forests with ALOS PALSAR Interferometric Coherence and Backscatter Intensity', *Journal of Imaging,* 2(1), pp. 1.

Storey, J., Choate, M. and Dekota, S. 'Geometric and spatial performance of Landsat 8'. *ASPRS 2014 Annual Conference & co-located JACIE Workshop.*

Storey, J., Choate, M. and Lee, K. (2014b) 'Landsat 8 Operational Land Imager on-orbit geometric calibration and performance', *Remote Sensing,* 6(11), pp. 11127-11152.

Sun, G. and Ranson, K. J. (2001) 'Terrain Effects on Forest RADAR Backscatter: Modeling and Correlation', *CEOS SAR Workshop.*

Sun, G., Ranson, K. J. and Kharuk, V. I. (2002) 'Radiometric slope correction for forest biomass estimation from SAR data in the Western Sayani Mountains,Siberia', *Remote Sensing of Environment,* 79(2), pp. 279-287.

Sun, G., Ranson, K. J., Masek, J., Fu, A. and Wang, D. 'Predicting tree height and biomass from GLAS data'. *10th International Symposium on Physical Measurements and Signatures in Remote Sensing.*

Suresh, M., Kiran Chand, T., Fararoda, R., Jha, C. and Dadhwal, V. (2014) 'Forest above ground biomass estimation and forest/non-forest classification for Odisha, India, using L-band Synthetic Aperture Radar (SAR) data', *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 1, pp. 651-658.

Svob, S., Arroyo-Mora, J. P., & Kalacska, M. (2014). 'A wood density and aboveground biomass variability assessment using pre-felling inventory data in Costa Rica'. *Carbon Balance and Management, 9,* 9. URL: http://doi.org/10.1186/s13021-014-0009-y

Tanase, M. A., Panciera, R., Lowell, K., Tian, S., Hacker, J. M. and Walker, J. P. (2014) 'Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests', *Remote Sensing of Environment,* 145, pp. 93-104.

Tanase, M. A., Santoro, M., de La Riva, J., Perez-Cabello, F. and Le Toan, T. (2010) 'Sensitivity of X-, C-, and L-band SAR backscatter to burn severity in Mediterranean pine forests', *Geoscience and Remote Sensing, IEEE Transactions on,* 48(10), pp. 3663-3675.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Legg, C. and De Dieu, M. J. (2004) 'Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests', *Remote Sensing of Environment,* 90(1), pp. 23-43.

Tibshirani, R. (1996) 'A comparison of some error estimates for neural network models', *Neural Computation,* 8(1), pp. 152-163.

Tomppo, E., Nilsson, M., Rosengren, M., Aalto, P. and Kennedy, P. (2002) 'Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass', *Remote Sensing of Environment,* 82(1), pp. 156-171.

Tuominen, S. and Haapanen, R. (2013) 'Estimation of forest biomass by means of genetic algorithm-based optimization of airborne laser scanning and digital aerial photograph features', *Silva Fennica,* 47(1).

USGS (2017) *Landsat Data Access | Landsat Missions*: Landsat.usgs.gov. Available at: http://landsat.usgs.gov/landsat-data-access (Accessed: 1 Jan. 2016 2016).

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

Van Aardt, J. A. N. (2004) *An object-oriented approach to forest volume and aboveground biomass modeling using small-footprint lidar data for segmentation,estimation,and classification.* Faculty of the Virginia Polytechnic Institute and State University [Online] Available at: https://vtechworks.lib.vt.edu/bitstream/handle/10919/28789/vanAardt_Dissertation.pdf?sequence=1&isAllowed=y (Accessed: 1/5/2014).

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E. and Yu, T. (2014) 'scikit-image: image processing in Python', *PeerJ,* 2, pp. e453.

Vierling, L. A., Vierling, K. T., Adam, P. and Hudak, A. T. (2013) 'Using Satellite and Airborne LiDAR to Model Woodpecker Habitat Occupancy at the Landscape Scale', *PloS one,* 8(12), pp. e80988.

Walker, W. S., Stickler, C. M., Kellndorfer, J. M., Kirsch, K. M. and Nepstad, D. C. (2010) 'Large-area classification and mapping of forest and land cover in the Brazilian Amazon: a comparative analysis of ALOS/PALSAR and Landsat data sources', *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* 3(4), pp. 594-604.

Wei, S. and Wang, Y. 'A Robust Cylinder Fitting Method with Parameterized Distance Function Based on Gauss Image'. *Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009. SNPD'09. 10th ACIS International Conference on*: IEEE, 83-86.

Weisbin, C. R., Lincoln, W. and Saatchi, S. (2014) 'A Systems Engineering Approach to Estimating Uncertainty in Above-Ground BiomassAGB,Derived from Remote-Sensing Data', *Systems Engineering,* 17(3), pp. 361-373.

Wencheng, X., Xiaoling, Z. and Jun, S. 'MIMO antenna array design for airborne down-looking 3D imaging SAR'. *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*: IEEE, V2-452-V2-456.

Westra, E. (2016) *Python geospatial development.* Packt Publishing Ltd.

Wu, C., Shen, H., Wang, K., Shen, A., Deng, J. and Gan, M. (2016) 'Landsat Imagery-Based Above Ground Biomass Estimation and Change Investigation Related to Human Activities', *Sustainability,* 8(2), pp. 159.

Xie, Y., Sha, Z., Yu, M., Bai, Y. and Zhang, L. (2009) 'A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China', *Ecological Modelling,* 220(15), pp. 1810-1818.

Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G. and Ni-Meister, W. (2011) 'Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar', *Remote sensing of Environment,* 115(11), pp. 2965-2974.

Yuan, X., Li, L., Tian, X., Luo, G. and Chen, X. (2016) 'Estimation of above-ground biomass using MODIS satellite imagery of multiple land-cover types in China', *Remote Sensing Letters,* 7(12), pp. 1141-1149.

Zaki, N. A. M., Latif, Z. A., Suratman, M. N. and Zainal, M. Z. 'Aboveground biomass and carbon stocks modelling using non-linear regression model'. *IOP Conference Series: Earth and Environmental Science*: IOP Publishing, 012030.

Zhang, G., Ganguly, S., Nemani, R. R., White, M. A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y. and Myneni, R. B. (2014) 'Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data', *Remote Sensing of Environment*.

Zhao, F., Guo, Q. and Kelly, M. (2012) 'Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA', *Agricultural and Forest Meteorology,* 165, pp. 64-72.

Zhao, X. (2012) *Random sets to model uncertainty in remotely sensed objects.* PhD., University of Twente, Enschede.

Zhao, X., Stein, A., Chen, X. and Zhang, X. (2011) 'Quantification of extensional uncertainty of segmented image objects by random sets', *Geoscience and Remote Sensing, IEEE Transactions on,* 49(7), pp. 2548-2557.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J. and Ryu, S.-R. (2004) 'Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA', *Remote Sensing of Environment,* 93(3), pp. 402-411.

Zheng, S., Cao, C., Dang, Y., Xiang, H., Zhao, J., Zhang, Y., Wang, X. and Guo, H. (2014) 'Retrieval of forest growing stock volume by two different methods using Landsat TM images', *International Journal of Remote Sensing,* 35(1), pp. 29-43.

Zianis, D., Muukkonen, P., Mäkipää, R. and Mencuccini, M. (2005) *Biomass and stem volume equations for tree species in Europe.* FI.

Appendix-A: Studies used as inputs to the accuracy prediction tool.

1. Radar studies.

Study	*Bands*	*Polarisations*	*Resolution*	*Incidence angle (degrees)*	*Study area scale*	*Land cover heterogeneity (number of species)*	*Field allometric equation*	*Model*	*Average biomass*	*Best RMSE/average AGB*
(Ahmed, 2012)	L,C	HH,HV,VH,VV	1.6	25-65	forest	6	species	Parametric	200	0.17
(Askne et al., 2003)	C	VV	30	23	forest	2	species	Parametric	140	0.07
(Askne et al., 2003)	L	HH	100	23	forest	2	species	Parametric	140	0.25
(Minh et al., 2014)	P	HH,HV,VH,VV	1	24-62	forest	4	general	Parametric	350	0.09
(Soja et al., 2013)	P	HH,VV	50	26-35	forest	3	species	Parametric	182	0.22
(Rauste et al., 2009)	L	HH,VV	100	40	county	3	general	Parametric	66	0.53
(Askne et al., 2013)	X	VV	11	41	forest	2	general	Parametric	105	0.18
(Askne et al., 2013)	X	VV	12	34	forest	3	general	Parametric	105	0.21
(Rauste et al., 2009)	X	HH,HV,VH,VV	16	40	county	3	general	Parametric	66	0.70
(Sandberg et al., 2011)	L,P	HH,HV	4	28–50	country-state	3	species	Parametric	111	0.31
(Sandberg et al., 2011)	L	HH,HV	4	28–50	country-state	3	species	Parametric	111	0.18
(Sandberg et al., 2011)	P	HH,HV	4	28–50	country-state	3	species	Parametric	111	0.27
(Tian et al., 2013)	L	HV	0.25	18-43	forest	2	species	physical model	90	0.26
(Tian et al., 2013)	L	HH	0.25	18-43	forest	2	species	physical model	90	0.31
(Hyde et al., 2007)	X, P	HH,VV	5	20-60	forest	3	species	Parametric	125	0.42
(Hyde et al., 2007)	UHF	HH,HV,VV	0.75	20-60	forest	3	species	Parametric	125	1.00
(Robinson et al., 2013)	L	HH,HV,VV	5	20-50	forest	6	species	Parametric	158	0.30
(Robinson et al., 2013)	L	HH,HV,VV	5	20-50	forest	6	species	Parametric	75	0.31
(Tsui, 2013)	C	HH,VV	50	39.2	forest	3	general	Parametric	223	0.54
(Tsui, 2013)	L	HH,HV	19	34.3	forest	2	general	Parametric	224	0.44
(Tsui, 2013)	L,C	VH	20	34.3-39.2	forest	3	general	Parametric	224	0.36

(Tsui, 2013)	L,C	HH,HV,VV	20	34.3-39.2	forest	3	general	Parametric	224	0.40
(Tan, 2012)	X	HH,HV,VH,VV	7	24	county	15	species	Parametric	75	0.38
(Tan, 2012)	L	HH,HV,VH,VV	10	24	county	15	species	Parametric	75	0.31
(Tanase et al., 2014b)	L	HH	10	8 - 30.0	forest	4	species	Parametric	50	0.22
(Tanase et al., 2014b)	L	HV	10	8 - 30.0	forest	4	species	Parametric	50	0.23
(Tanase et al., 2014b)	L	HH,HV,VV	10	8 - 30.0	forest	4	species	Parametric	50	0.21
(Tanase et al., 2014b)	L	HH	10	8 - 30.0	forest	4	species	Non-parametric	50	0.26
(Tanase et al., 2014b)	L	HV	10	8 - 30.0	forest	4	species	Non-parametric	50	0.26
(Tanase et al., 2014b)	L	HH,HV,VV	10	8 - 30.0	forest	4	species	Non-parametric	50	0.26
(Tanase et al., 2014b)	L	HH	10	8 - 30.0	forest	4	species	Non-parametric	50	0.24
(Tanase et al., 2014b)	L	HV	10	8 - 30.0	forest	4	species	Non-parametric	50	0.24
(Tanase et al., 2014b)	L	HH,HV,VV	10	8 - 30.0	forest	4	species	Non-parametric	50	0.24
(Tanase et al., 2014b)	L	HH	10	8 - 30.0	forest	4	species	Non-parametric	50	0.23
(Tanase et al., 2014b)	L	HV	10	8 - 30.0	forest	4	species	Non-parametric	50	0.24
(Tanase et al., 2014b)	L	HH,HV,VV	10	8 - 30.0	forest	4	species	Non-parametric	50	0.24
(Banskota et al., 2011)	VHF	VV	30	5–45	county	8	general	Parametric	140	0.22
(Mougin et al., 1999)	P	HV	1.5	35	forest	4	species	Parametric	200	0.37
(Mougin et al., 1999)	P,L,C	HH,HV,VV	1.5	35	forest	4	species	Parametric	175	0.23
(Jing et al., 2013)	C	HH	3	10.0-60.0	forest	1	species	Non-parametric	19	0.11
(Jing et al., 2013)	C	HV	3	10.0-60.0	forest	1	species	Non-parametric	19	0.06
(Jing et al., 2013)	C	HH,HV	3	10.0-60.0	forest	1	species	Non-parametric	19	0.08
(Sarker et al., 2012)	L	HH,HV	10	8.0-60.0	country-state	10	general	Parametric	175	0.24
(Sarker et al., 2012)	L	HH	10	8.0-60.0	country-state	10	general	Parametric	175	0.43
(Sarker et al., 2012)	L	HV	10	8.0-60.0	country-state	10	general	Parametric	175	0.39
(Ghasemi et al.)	L	HH,HV	12.5	8.0-60.0	forest	6	general	Parametric	200	0.13
(Sun et al., 2011)	L	HH,HV,VH,VV	15	21.5-34.3	forest	3	general	Parametric	150	0.19
(Tanase et al., 2014a)	L	HH	6	15-45	forest	2	species	Parametric	45	0.68
(Tanase et al., 2014a)	L	HV	6	15-45	forest	2	species	Parametric	43	0.60
(Tanase et al., 2014a)	L	VV	6	15-45	forest	2	species	Parametric	48	0.75
(Tanase et al., 2014a)	L	HH,HV	6	15-45	forest	2	species	Parametric	45	0.63
(Tanase et al., 2014a)	L	HH,HV,VH,VV	6	15-45	forest	2	species	Parametric	45	0.68
(Dobson et al., 1995)	L,C,X	HH,HV,VH,VV	7.4	29-34	forest	5	species	Parametric	125	0.11

(Avtar et al., 2013)	L	HH,HV	40	7.9-60.0	country-state	20	general	Parametric	140	0.45
(Avtar et al., 2013)	L	HH,HV	40	7.9-60.0	country-state	20	general	Parametric	90	0.21
(Avtar et al., 2013)	L	HH,HV	40	7.9-60.0	country-state	20	general	Parametric	110	0.19
(Hamdan et al., 2014)	L	HV	25	7.9-60.0	country-state	2	general	Parametric	227	0.19
(Hamdan et al., 2014)	L	HH	25	7.9-60.0	country-state	2	general	Parametric	227	0.22
(Hamdan et al., 2014)	L	HH,HV	25	7.9-60.0	country-state	2	general	Parametric	227	0.24
(Avtar et al., 2014)	L	HH,HV	12.5	35	country-state	5	general	Parametric	90	0.26
(Englhart et al., 2011)	X	VV	8.25	29.9-36.2	county	5	species	Parametric	180	0.65
(Englhart et al., 2011)	L	HH,HV	12.5	38.8	county	5	species	Parametric	180	0.57
(Englhart et al., 2011)	L,X	HH,HV,VV	8.25	29.9-36.2	county	5	species	Parametric	180	0.44
(Englhart et al., 2011)	L,X	HH,HV,VV	8.25	29.9-36.2	county	5	species	Parametric	200	0.30
(Schlund et al., 2015)	X	HH	3.3	46.8–48.5	forest	1	species	Parametric	314	0.17
(Treuhaft et al., 2015)	X	HH	2.5	40	forest	4	species	physical model	179	0.29
(Solberg et al., 2013)	X	HH	3	36-42	forest	3	species	physical model	338	0.20
(Neumann et al., 2012)	L	HV	50	25-55	country-state	3	species	Parametric	94	0.38
(Neumann et al., 2012)	P	HV	50	25-55	country-state	3	species	Parametric	94	0.43
(Neumann et al., 2012)	L	HV	50	25-55	country-state	3	species	Parametric	94	0.23
(Neumann et al., 2012)	P	HV	50	25-55	country-state	3	species	Parametric	94	0.30
(Neumann et al., 2012)	L	HV,VV,HH	50	25-55	country-state	3	species	Parametric	94	0.22
(Neumann et al., 2012)	P	HV,VV,HH	50	25-55	country-state	3	species	Parametric	94	0.28
(Neumann et al., 2012)	L	HV,VV,HH	50	25-55	country-state	3	species	Non-parametric	94	0.21
(Neumann et al., 2012)	P	HV,VV,HH	50	25-55	country-state	3	species	Non-parametric	94	0.26
(Neumann et al., 2012)	L	HV,VV,HH	50	25-55	country-state	3	species	Non-parametric	94	0.26
(Neumann et al., 2012)	P	HV,VV,HH	50	25-55	country-state	3	species	Non-parametric	94	0.24
(Tsui et al., 2012)	C	HH,HV,VH,VV	8	39.2	forest	2	species	Parametric	223	0.54
(Tsui et al., 2012)	L	HH,HV	20	34.3	forest	2	species	Parametric	223	0.44
(Tsui et al., 2012)	L,C	HH,HV,VH,VV	8	34-39	forest	2	species	Parametric	223	0.45
(Suresh et al., 2014)	L	HV	50	36.6-40.9	county	5	species	Parametric	100	0.47
(Montesano et al., 2014)	L	HH, HV, VV	5	NA	county	20	general	Parametric	100	0.45
(Montesano et al., 2014)	L	HH, HV, VV	5	NA	county	20	general	Parametric	20	0.60
(Minh et al., 2015)	P	HV	1.245	25	forest	20	general	Parametric	350	0.10
(Tian et al., 2012)	L	HV	10	8 - 30.0	forest	1	species	Parametric	96	0.42

(Tian et al., 2012)	L	HV	10	8 - 30.0	forest	1	species	Non-parametric	96	0.26
(Næsset et al., 2011)	X	HH	15	57	forest	3	species	Parametric	110	0.29
(Vastaranta et al., 2014)	X	HH,VV	2	35.8-52	forest	3	species	Non-parametric	150	0.28
(Saatchi et al., 2011)	L	HH	5	20-60	forest	11	species	Parametric	139	0.35
(Saatchi et al., 2011)	L	HV	5	20-60	forest	11	species	Parametric	139	0.32
(Saatchi et al., 2011)	L	VV	5	20-60	forest	11	species	Parametric	139	0.33
(Saatchi et al., 2011)	P	HH	10	20-60	forest	11	species	Parametric	139	0.31
(Saatchi et al., 2011)	P	HV	10	20-60	forest	11	species	Parametric	139	0.27
(Saatchi et al., 2011)	P	VV	10	20-60	forest	11	species	Parametric	139	0.28
(Saatchi et al., 2011)	L	HH,HV	5	20-60	forest	11	species	Parametric	139	0.31
(Saatchi et al., 2011)	L	HH,HV,VV	5	20-60	forest	11	species	Parametric	139	0.28
(Saatchi et al., 2011)	P	HH,HV	10	20-60	forest	11	species	Parametric	139	0.21
(Saatchi et al., 2011)	P	HH,HV,VV	10	20-60	forest	11	species	Parametric	139	0.19

2. Lidar studies.

Study	*Records*	*Footprint size m*	*Study area size*	*Land cover heterogeneity (No. Of species)*	*Allometric equation*	*Model*	*Average biomass*	*Best RMSE/average AGB*
(Ahmed, 2012)	Full Waveform	20	forest	6	species	Parametric	200	0.17
(Baghdadi et al., 2013)	Full Waveform	70	forest	1	species	Parametric	64	0.25
(Bortolot and Wynne, 2005)	Multi Returns	0.5	forest	4	species	Parametric	70	0.16
(Drake et al., 2002a)	Full Waveform	25	forest	3	age	Physical	175	0.36
(Drake et al., 2002a)	Full Waveform	25	forest	4	age	Parametric	175	0.11
(He et al., 2013)	Multi Returns	0.4	forest	4	species	Parametric	104	0.14
(Lim and Treitz, 2004a)	Multi Returns	0.3	forest	1	general	Parametric	84	0.35
(Lim and Treitz, 2004b)	Multi Returns	0.2	forest	4	site-specific	Parametric	120	0.40
(Lu et al., 2012)	Multi Returns	0.25	forest	7	species	Parametric	239	0.34

(Mitchard et al., 2013)	Full Waveform	70	global	20	general	Parametric	100	0.23
(Vaglio et al., 2014)	Multi Returns	0.1	forest	4	general	Parametric	172	0.39
(Popescu et al., 2011)	Full Waveform	70	forest	2	general	Parametric	150	0.25
(Hyde et al., 2007)	Multi Returns	2	forest	3	species	Parametric	125	0.21
(Tsui, 2013)	Multi Returns	0.5	forest	3	general	Parametric	315	0.18
(Tan, 2012)	Multi Returns	0.25	forest	15	species	Physical	75	0.20
(Tanase et al., 2014b)	Multi Returns	0.1	forest	4	species	Non-parametric	61	0.28
(Banskota et al., 2011)	Multi Returns	1	county	8	general	Parametric	140	0.20
(Banskota et al., 2011)	profiling	0.6	county	8	general	Parametric	140	0.23
(Riegel, 2012)	Multi Returns	0.16	forest	3	general	Parametric	1	0.13
(Sun et al., 2011)	Full Waveform	20	forest	3	general	Parametric	150	0.21
(Clark et al., 2011)	Multi Returns	0.3	forest	3	general	Parametric	121	0.30
(Clark et al., 2011)	Multi Returns	0.3	forest	3	general	Parametric	121	0.28
(Clark et al., 2011)	Multi Returns	0.3	forest	3	general	Parametric	30	0.36
(Clark et al., 2011)	Multi Returns	0.3	forest	3	general	Parametric	30	0.27
(Persson et al., 2012)	Multi Returns	0.5	forest	5	general	Parametric	111	0.40
(Johnson et al., 2014)	Multi Returns	2	county	20	general	Parametric	208	0.17
(Johnson et al., 2014)	Multi Returns	2	county	20	general	Non-parametric	208	0.35
(Popescu et al., 2004)	Multi Returns	0.65	forest	1	species	Parametric	132	0.33
(Tanase et al., 2014a)	Multi Returns	0.15	forest	2	species	Parametric	57	0.30
(Drake et al., 2002b)	Full Waveform	25	forest	2	general	Parametric	161	0.14
(Drake et al., 2002b)	Full Waveform	25	forest	2	general	Parametric	161	0.10
(Drake et al., 2003)	Full Waveform	25	forest	1	general	Parametric	224	0.14
(Drake et al., 2003)	Full Waveform	25	forest	2	general	Parametric	161	0.14
(Means et al., 1999)	Full Waveform	10	forest	2	species	Parametric	500	0.26
(Means et al., 1999)	Full Waveform	10	forest	2	species	Parametric	500	0.18
(Swatantran et al., 2011)	Full Waveform	12.5	forest	5	species	Parametric	200	0.35
(Kronseder et al., 2012)	Full Waveform	0.25	county	2	general	Parametric	228	0.21
(Schlund et al., 2015)	Multi Returns	0.3	forest	1	species	Parametric	315	0.11
(Skowronski et al., 2014)	Multi Returns	1	forest	5	species	Parametric	88	0.26
(Tsui et al., 2012)	Multi Returns	0.52	forest	2	species	Parametric	315	0.18
(Hyde et al., 2007)	Multi Returns	5	forest	3	species	Parametric	125	0.21

(Hyde et al., 2006)	Full Waveform	12.5	forest	5	species	Parametric	500.00	0.15
(Sheridan et al., 2015)	Multi Returns	0.25	forest	10	species	Parametric	114	0.31
(Hansen et al., 2015)	Multi Returns	0.22	forest	20	general	Parametric	462	0.32
(Skowronski et al., 2014)	Multi Returns	1	forest	5	species	Parametric	93	0.26
(Neigh et al., 2013a)	Full Waveform	50	global	20	general	Physical	75	0.78
(Montesano et al., 2014)	Full Waveform	50	contenant	20	general	Parametric	100	0.50
(Montesano et al., 2014)	Full Waveform	50	contenant	20	general	Parametric	20	1.00
(Montesano et al., 2014)	Full Waveform	20	country	20	general	Parametric	100	0.50
(Montesano et al., 2014)	Full Waveform	20	country	20	general	Parametric	20	0.50
(Montesano et al., 2014)	Multi Returns	1	country	20	general	Parametric	100	0.45
(Montesano et al., 2014)	Multi Returns	1	country	20	general	Parametric	20	0.60
(St-Onge et al., 2008)	Multi Returns	1	forest	6	species	Parametric	175	0.23
(Anderson et al., 2008)	Full Waveform	10	forest	8	species	Parametric	243	0.23
(Tian et al., 2012)	Multi Returns	0.5	forest	1	species	Parametric	97	0.13
(Nelson et al., 2007)	Multi Returns	1	forest	1	species	Parametric	200	0.17
(He et al., 2012)	Multi Returns	1	forest	3	species	Parametric	100	0.16
(Næsset et al., 2011)	Multi Returns	1.4286	forest	3	species	Parametric	110	0.14
(Vastaranta et al., 2014)	Multi Returns	2	forest	3	species	Non-parametric	150	0.22

3. Optical studies.

Study	*Resolution*	*Spectral resolution*	*Study area size*	*Land cover heterogeneity (No. of species)*	*Allometric equation*	*Model Type*	*Average biomass*	*Best RMSE/average AGB*
(Dong et al., 2003)	8000	Multi-spectral	country-state	8	general	Physical	61	0.5
(Eckert, 2012)	2	Multi-spectral	county	4	general	Parametric	50	0.08
(Heiskanen, 2006)	15	Multi-spectral	forest	3	general	Parametric	8	0.41

(Leboeuf et al., 2007)	0.65	Panchromatic	forest	12	general	Physical	50	0.22
(Migolet et al., 2007)	4	Multi-spectral	forest	5	general	Non- parametric	9	0.01
(Migolet et al., 2007)	4	Multi-spectral	forest	5	general	Parametric	8	0.18
(Nichol and Sarker, 2011)	10	Multi-spectral	forest	3	species	Parametric	175	0.17
(Sarker and Nichol, 2011)	10	Multi-spectral	forest	3	general	Parametric	175	0.26
(Tomppo et al., 2002)	30	Multi-spectral	forest	3	general	Non- parametric	71	0.53
(Xie et al., 2009)	30	Multi-spectral	forest	7	general	Non- parametric	121	0.42
(Xie et al., 2009)	30	Multi-spectral	forest	7	general	Parametric	147	0.53
(Zhang and Kondragunta, 2006)	1000	Multi-spectral	Continent	20	general	Parametric	125	0.32
(Nichol and Sarker, 2010)	10	Multi-spectral	county	3	general	Parametric	175	0.14
(Labrecque et al., 2006)	30	Multi-spectral	county	3	general	Non- parametric	116	0.51
(Labrecque et al., 2006)	30	Multi-spectral	county	3	general	Parametric	107	0.53
(Labrecque et al., 2006)	30	Multi-spectral	county	3	general	Physical	153	0.52
(Chung et al., 2009)	30	Multi-spectral	forest	2	general	Non- parametric	131	0.50
(Chung et al., 2009)	30	Multi-spectral	forest	3	general	Parametric	136	0.41
(Muukkonen and Heiskanen, 2005)	15	Multi-spectral	country-state	6	species	Non- parametric	104	0.41
(Muukkonen and Heiskanen, 2005)	15	Multi-spectral	country-state	6	species	Parametric	104	0.45
(Reese et al., 2002)	30	Multi-spectral	county	4	species	Non- parametric	66	0.53
(Reese et al., 2002)	30	Multi-spectral	county	4	species	Non- parametric	84	0.69
(Reese et al., 2002)	30	Multi-spectral	county	4	species	Non- parametric	95	0.69
(Reese et al., 2002)	30	Multi-spectral	county	4	species	Non- parametric	82	0.79
(Soenen et al., 2010)	10	Multi-spectral	forest	2	species	Parametric	170	0.41
(Soenen et al., 2010)	10	Multi-spectral	forest	2	species	Parametric	80	0.41
(Clark et al., 2011)	1.6	Hyper-spectral	forest	3	general	Parametric	121	0.58
(Clark et al., 2011)	1.6	Hyper-spectral	forest	3	general	Parametric	121	0.53
(Clark et al., 2011)	1.6	Hyper-spectral	forest	1	general	Parametric	30	0.75
(Clark et al., 2011)	1.6	Hyper-spectral	forest	1	general	Parametric	30	0.67
(Persson et al., 2013)	2.5	Panchromatic	forest	5	general	Parametric	110	0.33
(Persson et al., 2013)	10	Multi-spectral	forest	5	general	Parametric	110	0.31
(Byrd et al., 2014)	30	Multi-spectral	forest	2	species	Parametric	15	0.21
(Byrd et al., 2014)	1	Multi-spectral	forest	2	species	Parametric	15	0.22
(Byrd et al., 2014)	30	Hyper-spectral	forest	2	species	Parametric	15	0.34

(Cortés et al., 2014)	30	Multi-spectral	forest	3	species	Parametric	86	0.52
(Hamdan et al., 2014)	5	Multi-spectral	county	2	general	Parametric	227	0.19
(Swatantran et al., 2011)	3.3	Multi-spectral	forest	5	species	Parametric	200	0.46
(Hyde et al., 2006)	2	Multi-spectral	forest	5	species	Parametric	500	0.21
(Hyde et al., 2006)	30	Multi-spectral	forest	5	species	Parametric	500	0.222
(Neigh et al., 2013b)	30	Multi-spectral	forest	8	species	Parametric	150	0.31
(Mutanga et al., 2012)	1	Multi-spectral	forest	4	general	Non- parametric	34	0.13
(Mutanga et al., 2012)	1	Multi-spectral	forest	4	general	Parametric	34	0.16
(Günlü et al., 2014)	30	Multi-spectral	forest	1	general	Parametric	83	0.10
(Ramoelo et al., 2014)	1	Multi-spectral	forest	1	species	Parametric	4	0.17
(Ramoelo et al., 2014)	1	Multi-spectral	forest	1	species	Parametric	2	0.19
(Ramoelo et al., 2014)	1	Multi-spectral	forest	1	species	Parametric	4	0.3
(Korom et al., 2014)	4	Multi-spectral	county	1	general	Parametric	73	0.10
(Korom et al., 2014)	4	Multi-spectral	county	1	general	Parametric	140	0.04
(Korom et al., 2014)	4	Multi-spectral	county	3	general	Parametric	135	0.10
(Wallin et al., 2008)	30	Multi-spectral	country-state	5	species	Parametric	448	0.52
(Güneralp et al., 2014)	30	Multi-spectral	forest	4	species	Parametric	125	0.23
(Güneralp et al., 2014)	10	Multi-spectral	forest	4	species	Parametric	125	0.18
(Dube and Mutanga, 2015)	30	Multi-spectral	forest	3	species	Parametric	139	0.45
(Anderson et al., 2008)	3.3	Multi-spectral	forest	8	species	Parametric	243	0.23
(Tian et al., 2012)	10	Multi-spectral	forest	1	species	Non- parametric	97	0.24
(Tian et al., 2012)	10	Multi-spectral	forest	1	species	Parametric	97	0.40

4. Combined studies.

Study	*Data source 1*	*Resolution of data source 1*	*Data source 2*	*Resolution of data source 2*	*Study area size*	*Land cover heterogeneity (No. Of species)*	*Allometric equation*	*Model*	*Average biomass*	*Best RMSE/average AGB*
(Attarchi and Gloaguen, 2014)	Radar	10	optical	30	forest	5	species	Parametric	196	0.16
(Ghasemi et al., 2013)	Radar	12.5	optical	10	forest	5	species	Non-parametric	250	0.11
(Lucas et al., 2008)	Lidar	0.75	optical	1	forest	4	species	Parametric	85	0.14
(Lucas et al., 2008)	Lidar	0.75	optical	1	forest	5	species	Physical	85	0.30
(Tanase et al., 2014c)	Lidar	0.15	Radar	6	forest	1	general	Parametric	48	0.48
(Tuominen and Haapanen, 2013)	Lidar	20	optical	0.5	county	8	species	Non-parametric	96	0.23
(Vaglio et al., 2014)	Lidar	1	optical	1	forest	4	general	Parametric	172	0.36
(Zhang et al., 2014)	Lidar	70	optical	30	state	8	general	Parametric	300	0.25
(Amini and Tetuko Sri Sumantyo, 2009)	Radar	50	optical	10	forest	7	general	ANN	92	0.20
(Popescu et al., 2011)	Lidar	70	optical	2.4	forest	5	general	Parametric	150	0.25
(Tian et al., 2013)	Lidar	0.15	Radar	0.25	forest	2	species	Physical	90	0.26
(Hyde et al., 2007)	Lidar	2	Radar	5	forest	3	species	Parametric	125	0.19
(Goh et al., 2014)	Radar	100	optical	10	forest	1	general	Parametric	420	0.36
(Tsui, 2013)	Lidar	0.5	Radar	20	forest	3	general	Parametric	315	0.07
(Treuhaft et al., 2003)	Radar	20	optical	20	forest	2	general	Parametric	161	0.16
(Banskota et al., 2011)	Lidar	0.6	Radar	30	county	8	general	Parametric	140	0.15

(Banskota et al., 2011)	Lidar	1	Radar	30	county	8	general	Parametric	140	0.17
(Gagliasso et al., 2014)	Lidar	0.25	optical	30	forest	5	species	Non-parametric	22	1.57
(Gagliasso et al., 2014)	Lidar	0.25	optical	30	forest	5	species	Non-parametric	22	1.26
(Gagliasso et al., 2014)	Lidar	0.25	optical	30	forest	5	species	Parametric	22	1.43
(Riegel, 2012)	Lidar	0.16	optical	1	forest	3	general	Parametric	3	0.05
(Sun et al., 2011)	Lidar	20	Radar	15	forest	3	general	Parametric	150	0.19
(Clark et al., 2011)	Lidar	0.3	optical	1.6	forest	3	general	Parametric	121	0.29
(Clark et al., 2011)	Lidar	0.3	optical	1.6	forest	3	general	Parametric	30	0.29
(Persson et al., 2013)	Lidar	0.5	optical	2.5	forest	5	general	Parametric	154	0.21
(Persson et al., 2012)	Lidar	0.5	optical	10	forest	5	general	Parametric	110	0.30
(Popescu et al., 2004)	Lidar	0.7	optical	4	forest	1	species	Parametric	80	0.36
(Cortés et al., 2014)	Lidar	0.2	optical	30	forest	3	species	Parametric	95	0.35
(Hamdan et al., 2014)	Radar	25	optical	5	county	2	general	Parametric	226	0.14
(Swatantran et al., 2011)	Lidar	12.5	optical	3.3	county	5	species	Parametric	200	0.32
(Kattenborn et al., 2015)	Radar	2.5	optical	0.5	forest	5	species	Non-parametric	196	0.15
(Tsui et al., 2012)	Lidar	0.52	Radar	8	forest	2	species	Parametric	299	0.08
(Hyde et al., 2007)	Lidar	5	Radar	5	forest	3	species	Parametric	125	0.19
(Hyde et al., 2006)	Lidar	5	optical	30	forest	5	species	Parametric	500	0.14
(Hyde et al., 2006)	Lidar	5	Radar	2.5	forest	5	species	Parametric	500	0.15
(Hyde et al., 2006)	Radar	2.5	optical	2	forest	5	species	Parametric	500	0.17
(Hyde et al., 2006)	Lidar	5	optical	2	forest	5	species	Parametric	500	0.15
(Neigh et al., 2013b)	Lidar	20	Radar	7.2	forest	8	species	Parametric	150	0.21

(St-Onge et al., 2008)	Lidar	1	optical	1	forest	6	species	Parametric	175	0.29
(Anderson et al., 2008)	Lidar	10	optical	3.3	forest	8	species	Parametric	243	0.21
(Tian et al., 2012)	Radar	10	optical	10	forest	1	species	Non-parametric	97	0.23
(Tian et al., 2012)	Radar	10	optical	10	forest	1	species	Parametric	97	0.29
(Amini and Tetuko Sri Sumantyo, 2009)	Radar	13	optical	2.5	forest	7	species	Non-parametric	51	0.11
(Nelson et al., 2007)	Lidar	1	Radar	30	forest	1	species	Parametric	200	0.16
(Tsui et al., 2013)	Lidar	1.4	Radar	8	forest	4	general	Non-parametric	450	0.45
(Næsset et al., 2011)	Lidar	1.4	Radar	15	forest	3	species	Parametric	110	0.39
(Simard et al., 2006)	Lidar	1.5	Radar	30	forest	1	species	Parametric	150	0.37
(Shendryk et al., 2014)	Lidar	0.5	optical	10	forest	4	species	Parametric	475	0.10
(Kandel et al., 2014)	Lidar	1	optical	5	county	3	species	Parametric	207	0.25

References for Appendix A

Ahmed, R. (2012) *Accuracy of biomass and structure estimates from radar and lidar.* PhD., University of Massachusetts.

Amini, J. and Tetuko Sri Sumantyo, J. (2009) 'Employing a method on SAR and optical images for forest biomass estimation', *Geoscience and Remote Sensing, IEEE Transactions on,* 47(12), pp. 4020-4026.

Anderson, J. E., Plourde, L. C., Martin, M. E., Braswell, B. H., Smith, M.-L., Dubayah, R. O., Hofton, M. A. and Blair, J. B. (2008) 'Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest', *Remote Sensing of Environment,* 112(4), pp. 1856-1870.

Askne, J., Santoro, M., Smith, G. and Fransson, J. E. (2003) 'Multitemporal repeat-pass SAR interferometry of boreal forests', *Geoscience and Remote Sensing, IEEE Transactions on,* 41(7), pp. 1540-1550.

Askne, J. I., Fransson, J. E., Santoro, M., Soja, M. J. and Ulander, L. M. (2013) 'Model-based biomass estimation of a hemi-boreal forest from multitemporal TanDEM-X acquisitions', *Remote Sensing,* 5(11), pp. 5574-5597.

Attarchi, S. and Gloaguen, R. (2014) 'Improving the Estimation of Above Ground Biomass Using Dual Polarimetric PALSAR and ETM+ Data in the Hyrcanian Mountain Forest (Iran)', *Remote Sensing,* 6(5), pp. 3693-3715.

Avtar, R., Suzuki, R. and Sawada, H. (2014) 'Natural Forest Biomass Estimation Based on Plantation Information Using PALSAR Data', *PloS one,* 9(1), pp. e86121.

Avtar, R., Suzuki, R., Takeuchi, W. and Sawada, H. (2013) 'PALSAR 50 m Mosaic Data Based National Level Biomass Estimation in Cambodia for Implementation of REDD+ Mechanism', *PloS one,* 8(10), pp. e74807.

Baghdadi, N., le Maire, G., Fayad, I., Bailly, J. S., Nouvellon, Y., Lemos, C. and Hakamada, R. (2013) 'Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in Eucalyptus plantations in Brazil'.

Banskota, A., Wynne, R. H., Johnson, P. and Emessiene, B. (2011) 'Synergistic use of very high-frequency radar and discrete-return lidar for estimating biomass in temperate hardwood and mixed forests', *Annals of Forest Science,* 68(2), pp. 347-356.

Bortolot, Z. J. and Wynne, R. H. (2005) 'Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data', *ISPRS Journal of Photogrammetry and Remote Sensing,* 59(6), pp. 342-360.

Byrd, K. B., O'Connell, J. L., Di Tommaso, S. and Kelly, M. (2014) 'Evaluation of sensor types and environmental controls on mapping biomass of coastal marsh emergent vegetation', *Remote Sensing of Environment,* 149, pp. 166-180.

Chung, S. Y., Yim, J. S., Cho, H. K. and Shin, M. Y. (2009) 'Comparison of Forest Biomass Estimation Methods by combining Satellite Data and Field Data', *Extending Forest Inventory and Monitoring*.

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

Cortés, L., Hernández, J., Valencia, D. and Corvalán, P. (2014) 'Estimation of Above-Ground Forest Biomass Using Landsat ETM+, Aster GDEM and Lidar', *Forest Res,* 3(117), pp. 2.

Dobson, M. C., Ulaby, F. T., Pierce, L. E., Sharik, T. L., Bergen, K. M., Kellndorfer, J., Kendra, J. R., Li, E., Lin, Y.-C. and Nashashibi, A. (1995) *Estimation of forest biophysical characteristics in northern Michigan with SIR-C/X-SAR*. Geoscience and Remote Sensing, IEEE Transactions on 877-895 (Accessed: 4 33).

Dong, J., Kaufmann, R. K., Myneni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., Buermann, W., Alexeyev, V. and Hughes, M. K. (2003) 'Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks', *Remote Sensing of Environment,* 84(3), pp. 393-410.

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., Chazdon, R. L., Weishampel, J. F. and Prince, S. (2002a) 'Estimation of tropical forest structural characteristics using large-footprint lidar', *Remote Sensing of Environment,* 79(2), pp. 305-319.

Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B. and Blair, J. B. (2002b) 'Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest', *Remote Sensing of Environment,* 81(2), pp. 378-392.

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B. and Hofton, M. (2003) 'Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships', *Global ecology and biogeography,* 12(2), pp. 147-159.

Dube, T. and Mutanga, O. (2015) 'Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa', *ISPRS Journal of Photogrammetry and Remote Sensing,* 101, pp. 36-46.

Eckert, S. (2012) 'Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data', *Remote Sensing,* 4(4), pp. 810-829.

Englhart, S., Keuck, V. and Siegert, F. (2011) 'Aboveground biomass retrieval in tropical forests—The potential of combined X-and L-band SAR data use', *Remote Sensing of Environment,* 115(5), pp. 1260-1271.

Gagliasso, D., Hummel, S. and Temesgen, H. (2014) 'A Comparison of Selected Parametric and Non-Parametric Imputation Methods for Estimating Forest Biomass and Basal Area', *Open Journal of Forestry,* 4, pp. 42.

Ghasemi, N., Sahebi, M., Mohammadzadeh, A. and Bayani, R. (2011) 'Mapping biomass of a temperate deciduous forest using SAR data', *Geomatics 90.*

Ghasemi, N., Sahebi, M. R. and Mohammadzadeh, A. (2013) 'Biomass Estimation of a Temperate Deciduous Forest Using Wavelet Analysis', *Geoscience and Remote Sensing, IEEE Transactions on,* 51(2), pp. 765-776.

Goh, J., Miettinen, J., Chia, A. S., Chew, P. T. and Liew, S. C. (2014) 'Biomass estimation in humid tropical forest using a combination of ALOS PALSAR and SPOT 5 satellite imagery', *Asian Journal of Geoinformatics,* 13(4).

Güneralp, I., Filippi, A. M. and Randall, J. (2014) 'Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling', *International Journal of Applied Earth Observation and Geoinformation,* 33, pp. 119-126.

Günlü, A., Ercanli, I., Başkent, E. and Çakır, G. (2014) 'Estimating aboveground biomass using Landsat TM imagery: A case study of Anatolian Crimean pine forests in Turkey', *Annals of Forest Research,* 57(2), pp. 289-298.

Hamdan, O., Hasmadi, I. M. and Aziz, H. K. 'Combination of SPOT-5 and ALOS PALSAR images in estimating aboveground biomass of lowland Dipterocarp forest'. *IOP Conference Series: Earth and Environmental Science*: IOP Publishing, 012016.

Hansen, E. H., Gobakken, T., Bollandsås, O. M., Zahabu, E. and Næsset, E. (2015) 'Modeling Aboveground Biomass in Dense Tropical Submontane Rainforest Using Airborne Laser Scanner Data', *Remote Sensing,* 7(1), pp. 788-807.

He, Q.-S., Cao, C.-X., Chen, E.-X., Sun, G.-Q., Ling, F.-L., Pang, Y., Zhang, H., Ni, W.-J., Xu, M. and Li, Z.-Y. (2012) 'Forest stand biomass estimation using ALOS PALSAR data based on LiDAR-derived prior knowledge in the Qilian Mountain, western China', *International journal of remote sensing,* 33(3), pp. 710-729.

He, Q., Chen, E., An, R. and Li, Y. (2013) 'Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest', *Forests,* 4(4), pp. 984-1002.

Heiskanen, J. (2006) 'Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data', *International Journal of Remote Sensing,* 27(6), pp. 1135-1158.

Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M. and Hunsaker, C. (2006) 'Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy', *Remote Sensing of Environment,* 102(1), pp. 63-73.

Hyde, P., Nelson, R., Kimes, D. and Levine, E. (2007) 'Exploring LiDAR–RaDAR synergy—predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR', *Remote Sensing of Environment,* 106(1), pp. 28-38.

Jing, Z.-x., Zhang, Y. and Wang, K.-j. 'Estimating Paddy Rice Biomass Using Radarsat-2 Data Based on Artificial Neural Network'. *2013 the International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE 2013)*: Atlantis Press.

Johnson, K. D., Birdsey, R., Finley, A. O., Swantaran, A., Dubayah, R., Wayson, C. and Riemann, R. (2014) 'Integrating forest inventory and analysis data into a LIDAR-based carbon monitoring system', *Carbon balance and management,* 9, pp. 3.

Kandel, P. N., Awasthi, K. D., Kauranne, T., Gautam, B., Gunia, K. and Baral, S. (2014) 'Estimation of above ground Forest biomass and Carbon stock by Integrating LiDAR, satellite image and field measurement in Nepal', *Research Journal of Agriculture and Forestry Sciences* 2(8), pp. 1-6.

Kattenborn, T., Maack, J., Faßnacht, F., Enßle, F., Ermert, J. and Koch, B. (2015) 'Mapping forest biomass from space–Fusion of hyperspectral EO1-hyperion data and Tandem-X and WorldView-2 canopy height models', *International Journal of Applied Earth Observation and Geoinformation,* 35, pp. 359-367.

Korom, A., Phua, M. H., Hirata, Y., Matsuura, T., Saito, H. and Wong, W. 'Aboveground biomass assessment of degraded rainforest using IKONOS-2: Specific forest cluster analysis'. *35th Asian Conference on Remote Sensing 2014: Sensing for Reintegration of Societies, ACRS 2014*: Asian Association on Remote Sensing.

Kronseder, K., Ballhorn, U., Böhm, V. and Siegert, F. (2012) 'Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data', *International Journal of Applied Earth Observation and Geoinformation,* 18, pp. 37-48.

Labrecque, S., Fournier, R., Luther, J. and Piercey, D. (2006) 'A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland', *Forest Ecology and Management,* 226(1), pp. 129-144.

Leboeuf, A., Beaudoin, A., Fournier, R., Guindon, L., Luther, J. and Lambert, M.-C. (2007) 'A shadow fraction method for mapping biomass of northern boreal black spruce forests using QuickBird imagery', *Remote Sensing of Environment,* 110(4), pp. 488-500.

Lim, K. S. and Treitz, P. 'Estimation of aboveground forest biomass using airborne scanning discrete return Lidar in Douglas-fir'. *Proc. of the ISPRS working group VII/2 "Laser-Scanners for Forest and Landscape, Assessment*, 149-152.

Lim, K. S. and Treitz, P. M. (2004b) '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), pp. 558-570.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G. and Saah, D. (2012) 'Aboveground forest biomass estimation with Landsat and lidar data and uncertainty analysis of the estimates', *International Journal of Forestry Research,* 2012.

Lucas, R. M., Lee, A. and Bunting, P. J. (2008) 'Retrieving forest biomass through integration of CASI and LiDAR data', *International Journal of Remote Sensing,* 29(5), pp. 1553-1577.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., Harmon, M. E. and McKee, W. A. (1999) 'Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon', *Remote Sensing of Environment,* 67(3), pp. 298-308.

Migolet, P., Coulibaly, L., Adegbidi, H. G. and Hervet, E. 'Utilization of neural networks for the estimation of aboveground forest biomass from Ikonos satellite image and multi-source geo-scientific data'. *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*: IEEE, 4339-4342.

Minh, D. H. T., Le Toan, T., Rocca, F., Tebaldini, S., d'Alessandro, M. M. and Villard, L. (2014) 'Relating P-band synthetic aperture radar tomography to tropical forest biomass', *IEEE Transactions on Geoscience and Remote Sensing,* 52(2), pp. 967-979.

Minh, D. H. T., Tebaldini, S., Rocca, F., Le Toan, T., Villard, L. and Dubois-Fernandez, P. C. (2015) 'Capabilities of BIOMASS tomography for investigating tropical forests', *IEEE Transactions on Geoscience and Remote Sensing,* 53(2), pp. 965-975.

Mitchard, E. T., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. and Brown, S. (2013) 'Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps', *Carb Bal Manage*.

Montesano, P. M., Nelson, R. F., Dubayah, R. O., Sun, G., Cook, B. D., Ranson, K. J. R., N\ae, s., E. and Kharuk, V. (2014) 'The uncertainty of biomass estimates from LiDAR and SAR across a boreal forest structure gradient', *Remote Sensing of Environment,* 154, pp. 398-407.

Mougin, E., Proisy, C., Marty, G., Fromard, F., Puig, H., Betoulle, J. and Rudant, J.-P. (1999) 'Multifrequency and multipolarization radar backscattering from mangrove forests', *Geoscience and Remote Sensing, IEEE Transactions on,* 37(1), pp. 94-102.

Mutanga, O., Adam, E. and Cho, M. A. (2012) 'High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm', *International Journal of Applied Earth Observation and Geoinformation,* 18, pp. 399-406.

Muukkonen, P. and Heiskanen, J. (2005) 'Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data', *Remote sensing of Environment,* 99(4), pp. 434-447.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Ståhl, G. and Weydahl, D. (2011) 'Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area', *Remote Sensing of Environment,* 115(12), pp. 3599-3614.

Neigh, C. S., Nelson, R. F., Ranson, K. J., Margolis, H. A., Montesano, P. M., Sun, G., Kharuk, V., Næsset, E., Wulder, M. A. and Andersen, H.-E. (2013a) 'Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR', *Remote Sensing of Environment,* 137, pp. 274-287.

Neigh, C. S. R., Nelson, R. F., Ranson, K. J., Margolis, H. A., Montesano, P. M., Sun, G., Kharuk, V., Næsset, E., Wulder, M. A. and Andersen, H.-E. (2013b) 'Taking stock of circumboreal forest carbon with ground measurements,airborne and spaceborne LiDAR', *Remote Sensing of Environment,* 137, pp. 274-287.

Nelson, R. F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M. L., Campbell, R. and Edwards, W. (2007) 'Investigating RaDAR–LiDAR synergy in a North Carolina pine forest', *Remote Sensing of Environment,* 110(1), pp. 98-108.

Neumann, M., Saatchi, S. S., Ulander, L. M. and Fransson, J. E. (2012) 'Assessing performance of L-and P-band polarimetric interferometric SAR data in estimating boreal forest above-ground biomass', *Geoscience and Remote Sensing, IEEE Transactions on,* 50(3), pp. 714-726.

Nichol, J. E. and Sarker, M. L. R. (2010) *Efficiency of Texture Measurement From Two Optical Sensors for Improved Biomass Estimation. ISPRS TC VII Symposium – 100 Years ISPRS*: na.

Nichol, J. E. and Sarker, M. R. (2011) 'Improved biomass estimation using the texture parameters of two high-resolution optical sensors', *Geoscience and Remote Sensing, IEEE Transactions on,* 49(3), pp. 930-948.

Persson, H., Wallerman, J., Olsson, H. and Fransson, J. E. 'Estimating biomass and height using DSM from satellite data and DEM from high-resolution laser scanning data'. *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*: IEEE, 1649-1651.

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

Popescu, S. C., Wynne, R. H. and Scrivani, J. A. (2004) 'Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA', *Forest Science,* 50(4), pp. 551-565.

Popescu, S. C., Zhao, K., Neuenschwander, A. and Lin, C. (2011) 'Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level', *Remote Sensing of Environment,* 115(11), pp. 2786-2797.

Ramoelo, A., Cho, M., Mathieu, R., Madonsela, S., van der Korchove, R., Kaszta, Z. and Wolff, E. (2014) 'Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data', *International Journal of Applied Earth Observation and Geoinformation*.

Rauste, Y., Lonnqvist, A. and Ahola, H. (2009) 'Mapping boreal forest biomass with imagery from polarimetric and semi-polarimetric SAR sensors', *Ambiencia,* 4(4), pp. 171-180.

Reese, H., Nilsson, M., Sandström, P. and Olsson, H. (2002) 'Applications using estimates of forest parameters derived from satellite and forest inventory data', *Computers and Electronics in Agriculture,* 37(1), pp. 37-55.

Riegel, J. B. (2012) *A comparison of Remote Sensing Methods for Estimating Above-Ground Carbon Biomass at a Wetland Restoration Area in the Southeastern Coastal Plain.* Thesis, Duke University.

Robinson, C., Saatchi, S., Neumann, M. and Gillespie, T. (2013) 'Impacts of spatial variability on aboveground biomass estimation from L-band radar in a temperate forest', *Remote Sensing,* 5(3), pp. 1001-1023.

Saatchi, S., Marlier, M., Chazdon, R. L., Clark, D. B. and Russell, A. E. (2011) 'Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass', *Remote Sensing of Environment,* 115(11), pp. 2836-2849.

Sandberg, G., Ulander, L. M., Fransson, J., Holmgren, J. and Le Toan, T. (2011) 'L-and P-band backscatter intensity for biomass retrieval in hemiboreal forest', *Remote Sensing of Environment,* 115(11), pp. 2874-2886.

Sarker, L. R. and Nichol, J. E. (2011) 'Improved forest biomass estimates using ALOS AVNIR-2 texture indices', *Remote Sensing of Environment,* 115(4), pp. 968-977.

Sarker, M. L. R., Nichol, J., Ahmad, B., Busu, I. and Rahman, A. A. (2012) 'Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation', *ISPRS Journal of Photogrammetry and Remote Sensing,* 69, pp. 146-166.

Schlund, M., von Poncet, F., Kuntz, S., Schmullius, C. and Hoekman, D. H. (2015) 'TanDEM-X data for aboveground biomass retrieval in a tropical peat swamp forest', *Remote Sensing of Environment,* 158, pp. 255-266.

Shendryk, I., Hellström, M., Klemedtsson, L. and Kljun, N. (2014) 'Low-Density LiDAR and Optical Imagery for Biomass Estimation over Boreal Forest in Sweden', *Forests,* 5(5), pp. 992-1010.

Sheridan, R. D., Popescu, S. C., Gatziolis, D., Morgan, C. L. and Ku, N.-W. (2015) 'Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest', *Remote Sensing,* 7(1), pp. 229-255.

Simard, M., Zhang, K., Rivera-Monroy, V. H., Ross, M. S., Ruiz, P. L., Castañeda-Moya, E., Twilley, R. R. and Rodriguez, E. (2006) 'Mapping height and biomass of mangrove forests in Everglades National Park with SRTM elevation data', *Photogrammetric Engineering & Remote Sensing,* 72(3), pp. 299-311.

Skowronski, N. S., Clark, K. L., Gallagher, M., Birdsey, R. A. and Hom, J. L. (2014) 'Airborne laser scanner-assisted estimation of aboveground biomass change in a temperate oak–pine forest', *Remote Sensing of Environment,* 151, pp. 166-174.

Soenen, S. A., Peddle, D. R., Hall, R. J., Coburn, C. A. and Hall, F. G. (2010) 'Estimating aboveground forest biomass from canopy reflectance model inversion in mountainous terrain', *Remote Sensing of Environment,* 114(7), pp. 1325-1337.

Soja, M. J., Sandberg, G. and Ulander, L. M. (2013) 'Regression-based retrieval of boreal forest biomass in sloping terrain using P-band SAR backscatter intensity data', *Geoscience and Remote Sensing, IEEE Transactions on,* 51(5), pp. 2646-2665.

Solberg, S., Astrup, R., Breidenbach, J., Nilsen, B. and Weydahl, D. (2013) 'Monitoring spruce volume and biomass with InSAR data from TanDEM-X', *Remote Sensing of Environment,* 139, pp. 60-67.

St-Onge, B., Hu, Y. and Vega, C. (2008) 'Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images', *International Journal of Remote Sensing,* 29(5), pp. 1277-1294.

Sun, G., Ranson, K. J., Guo, Z., Zhang, Z., Montesano, P. and Kimes, D. (2011) 'Forest biomass mapping from lidar and radar synergies', *Remote Sensing of Environment,* 115(11), pp. 2906-2916.

Suresh, M., Kiran Chand, T., Fararoda, R., Jha, C. and Dadhwal, V. (2014) 'Forest above ground biomass estimation and forest/non-forest classification for Odisha, India, using L-band Synthetic Aperture Radar (SAR) data', *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 1, pp. 651-658.

Swatantran, A., Dubayah, R., Roberts, D., Hofton, M. and Blair, J. B. (2011) 'Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion', *Remote Sensing of Environment,* 115(11), pp. 2917-2930.

Tan, C. P. (2012) *Comparing synthetic aperture radar and LiDAR for above-ground biomass estimation in Glen Affric, Scotland.* PhD., University of Edinburgh.

Tanase, M. A., Panciera, R., Lowell, K., Aponte, C., Hacker, J. M. and Walker, J. P. (2014a) 'Forest Biomass Estimation at High Spatial Resolution: Radar vs. Lidar sensors', *IEEE Transactions on Geoscience and Remote Sensing Letters*, pp. 1-5.

Tanase, M. A., Panciera, R., Lowell, K., Tian, S., Garcia-Martin, A. and Walker, J. P. (2014b) 'Sensitivity of L-band radar backscatter to forest biomass in semiarid environments: A comparative analysis of parametric and nonparametric models', *IEEE Transactions on Geoscience and Remote Sensing,* 52(8), pp. 4671-4685.

Tanase, M. A., Panciera, R., Lowell, K., Tian, S., Hacker, J. M. and Walker, J. P. (2014c) 'Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests', *Remote Sensing of Environment,* 145, pp. 93-104.

Tian, S., Tanase, M. A., Panciera, R., Hacker, J. and Lowell, K. 'Cooperative Research Centre for Spatial Information, The University of Melbourne, Australia'. *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*: IEEE, 2145-2148.

Tian, X., Su, Z., Chen, E., Li, Z., van der Tol, C., Guo, J. and He, Q. (2012) 'Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area', *International journal of applied earth observation and geoinformation,* 14(1), pp. 160-168.

Tomppo, E., Nilsson, M., Rosengren, M., Aalto, P. and Kennedy, P. (2002) 'Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass', *Remote Sensing of Environment,* 82(1), pp. 156-171.

Treuhaft, R., Gonçalves, F., dos Santos, J. R., Keller, M., Palace, M., Madsen, S. N., Sullivan, F. and Graça, P. M. (2015) 'Tropical-Forest Biomass Estimation at X-Band From the Spaceborne TanDEM-X Interferometer', *Geoscience and Remote Sensing Letters, IEEE,* 12(2), pp. 239-243.

Treuhaft, R. N., Asner, G. P. and Law, B. E. (2003) 'Structure-based forest biomass from fusion of radar and hyperspectral observations', *Geophysical Research Letters,* 30(9).

Tsui, O. W. (2013) *Integrating discrete-return scanning LiDAR and spaceborne RADAR to support aboveground biomass assessments.* The University of British Columbia, British Columbia, British Columbia.

Tsui, O. W., Coops, N. C., Wulder, M. A. and Marshall, P. L. (2013) 'Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass', *Remote Sensing of Environment,* 139, pp. 340-352.

Tsui, O. W., Coops, N. C., Wulder, M. A., Marshall, P. L. and McCardle, A. (2012) 'Using multi-frequency radar and discrete-return LiDAR measurements to estimate above-ground biomass and biomass components in a coastal temperate forest', *ISPRS Journal of Photogrammetry and Remote Sensing,* 69, pp. 121-133.

Tuominen, S. and Haapanen, R. (2013) 'Estimation of forest biomass by means of genetic algorithm-based optimization of airborne laser scanning and digital aerial photograph features', *Silva Fennica,* 47(1).

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

Vastaranta, M., Holopainen, M., Karjalainen, M., Kankare, V., Hyyppa, J. and Kaasalainen, S. (2014) 'TerraSAR-X stereo radargrammetry and airborne scanning LiDAR height metrics in imputation of forest aboveground biomass and stem volume', *Geoscience and Remote Sensing, IEEE Transactions on,* 52(2), pp. 1197-1204.

Wallin, D. O., Schroeder, T. A., Gray, A., Harmon, M. E. and Cohen, W. B. (2008) 'Estimating Live Forest Carbon Dynamics with a Landsat-based Curve-fitting Approach', *Journal of Applied Remote Sensing,* 2(023519 (9 May 2008)).

Xie, Y., Sha, Z., Yu, M., Bai, Y. and Zhang, L. (2009) 'A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China', *Ecological Modelling,* 220(15), pp. 1810-1818.

Zhang, G., Ganguly, S., Nemani, R. R., White, M. A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y. and Myneni, R. B. (2014) 'Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data', *Remote Sensing of Environment*.

Zhang, X. and Kondragunta, S. (2006) 'Estimating forest biomass in the USA using generalized allometric models and MODIS land products', *Geophysical Research Letters,* 33(9).