

Optical Remote Sensing for Biomass Estimation in the Tropics: the case study of Uganda

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt dem Rat der Chemisch-Geowissenschaftlichen Fakultät
der Friedrich-Schiller-Universität Jena

von Dipl. Valerio Avitabile
geboren am 31 January 1975 in Roma (Italy)

Gutachter:

1. Prof. Dr. Christiane Schmullius, Friedrich-Schiller-Universität Jena
2. Prof. Dr. Martin Herold, Wageningen University

Datum der Verteidigung: 18 Juli 2012

“For every complex problem,
there is an answer
that is clear, simple, and wrong”

H.L. Mencken

Abstract

The assessment of tropical forest biomass is gaining increasing interest mainly for national resources planning, global carbon modeling, climate change research and mitigation activities, such as reducing emissions from deforestation and forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries (REDD+). It is widely recognized that remote sensing provides the key source of data for updated, consistent and spatially explicit assessment of biomass and its dynamics, especially in large countries with limited accessibility. However, there is no agreement in the international community on what method should be used to reliably determine this parameter.

This study investigates the capabilities and limitations of freely available optical satellite data at medium resolution to estimate aboveground biomass density of vegetation at national scales in the tropics, and compares this approach with existing methodologies to understand and quantify the sources of variability in the estimations. Uganda was chosen as a case-study because it presents a reliable national biomass reference dataset.

As a result of this thesis, aboveground woody biomass for circa-2000 was mapped at national scale in Uganda at 30-m spatial resolution on the basis of Landsat ETM+ images, a national land cover dataset and field data using an object-oriented approach. A regression tree-based model (Random Forest) produced good results (cross-validated R^2 0.81, RMSE 13 Mg/ha) when trained with a sufficient number of field plots representative of the vegetation variability at national scale. The Random Forest model effectively captured non-linear relationships between satellite data and biomass density, and was able to use categorical data (land cover) in the regression to improve the results.

This study demonstrated that in certain contexts Landsat data provide the capability to effectively spatialize field biomass measurements and produce accurate and detailed estimates of biomass distribution for all vegetation types at national scale. This approach tended to provide conservative biomass estimates and its limitations were mainly related to the saturation of the optical signal at high biomass density and to the cloud cover, which hindered the compilation of radiometrically consistent multi-temporal satellite datasets. For this reason, the Landsat mosaic created for Uganda with images acquired in the dry season during 1999 – 2003 did not contain phenological information useful for discriminating some vegetation types, such as deciduous formations. Fusion of satellite and land cover data improved model results because it combined the high spatial and thematic resolution of the Landsat data with the phenological information provided by the land cover data. However, to avoid error propagation, accurate, detailed and up-to-date land cover or other ancillary data are necessary.

When compared with the reference biomass map of the Forest Department of Uganda, the map produced in this study presented higher agreement than other five regional/global biomass maps. Moreover, the comparative analysis showed strong disagreement between the products, with estimates of total aboveground biomass of Uganda ranging from 343 to 2201 Tg and different spatial distribution patterns. Compared to the reference map based on country-specific field data and a national land cover dataset (estimating 468 Tg), maps based on biome-average biomass values, such as the Intergovernmental Panel on Climate Change default values, and global land cover datasets strongly overestimated biomass availability of Uganda (ranging from 578 to 2201 Tg), while maps based on satellite data and regression models provided conservative estimates (ranging from 343 to 443 Tg). The

comparison of the maps predictions with field data, upscaled to map resolution using land cover data, was in accordance with the above findings.

This study also demonstrated that the biomass estimates were primarily driven by the biomass reference data while the type of spatial maps used for their stratification had a smaller, but not negligible, impact. The differences in format, resolution and biomass definition used by the maps, as well as the fact that some datasets were not independent from the reference data to which they were compared, were carefully considered in the interpretation of the results.

The strong disagreement between existing products and the large impact of biomass reference data on the estimates indicated that the first, critical step to improve the accuracy of the biomass maps consists of the collection of accurate biomass field data for all relevant vegetation types. However, detailed and accurate spatial datasets are crucial to properly spatialize the field information and to obtain accurate estimates at specific locations.

Zusammenfassung

Für die nationale Ressourcenplanung, globale Kohlenstoffmodellierung, Klimawandelforschung oder auch im Rahmen internationaler Aktivitäten zur Abschwächung der Klimawandels, wie z.B. REDD+ (*Reducing Emissions from Deforestation and Forest Degradation and the Role of Conservation, Sustainable Management of Forests and Enhancement of Forest Carbon Stocks in Developing Countries*), ist die Bewertung und Quantifizierung tropischer Waldbiomasse von zunehmender Relevanz. In diesem Zusammenhang gelten Fernerkundungsdaten als geeignete Grundlage, um eine aktuelle, konsistente und räumlich explizite Abschätzung von Biomasse und deren Dynamik zu erhalten. Allerdings bestehen derzeit noch keine internationalen Vereinbarungen und Abkommen, welche die Auswahl und Verwendung zuverlässiger Methoden zur Ableitung von Biomasse definieren.

Die vorliegende Arbeit zeigt die Möglichkeiten und Limitationen frei verfügbarer, optisch moderat aufgelöster Satellitendaten zur Abschätzung oberirdischer Biomassendichte in den Tropen auf nationaler Ebene. Dabei ist insbesondere der Vergleich der entwickelten Methodik mit existierenden Ansätzen von Bedeutung, um die Variabilität bezüglich der Abschätzung von Biomasse quantifizieren zu können. Als Fallbeispiel für diese Arbeit wurde Uganda aufgrund seiner zuverlässigen und national vorliegenden Biomassereferenzdaten gewählt.

Innerhalb dieser Arbeit wurde mittels eines objektorientierten Ansatzes für das Jahr 2000 auf Grundlage von Landsat ETM+ Daten, einem nationalen Landbedeckungsdatensatz und Felddaten die oberirdische Holzbiomasse für Uganda mit einer räumlichen Auflösung von 30 x 30 m kartiert. Unter Verwendung eines regressions-basierten Entscheidungsmodelles (*Random Forest*) und eines, die Vegetationsvariabilität ausreichend repräsentierenden, Trainingsdatensatzes, wurden gute Ergebnisse erreicht (kreuz-validiertes R^2 0.81, RMSE 13 Mg/ha). Das *Random Forest* Modell ermöglicht hierbei die Beschreibung des nicht-linearen Zusammenhanges zwischen Satellitendaten und Biomassendichte und erlaubt darüber hinaus zur einer weiteren Verbesserung der Ergebnisse die Integration kategorischer Daten (z.B. Landbedeckung) in die Regression.

Die vorliegende Studie demonstriert, dass im betrachteten Kontext Landsat-Daten für eine räumlich explizite Bestimmung der Biomassenverteilung für alle Vegetationstypen sowie deren Variabilität auf nationaler Ebene geeignet sind. Der verwendete Ansatz liefert eine konservative Abschätzung der Biomasse, wobei Limitationen hauptsächlich durch die Sättigung des optischen Signals in hohen Biomassedichten und der Wolkenbedeckung gegeben sind. So verhinderte insbesondere die Wolkenbedeckung die Erstellung von radiometrisch konsistenten, multi-temporalen Satellitendatensätzen. Aus diesem Grund war die Datengrundlage zur Erstellung des Landsat-Mosaikes für Uganda auf Szenen beschränkt, welche in den Trockenperioden zwischen 1999 – 2003 akquiriert wurden. Dementsprechend beinhaltete das Landsat-Mosaik keine phänologischen Informationen zur Unterscheidung der Vegetationstypen, wie z.B. sommergrüne Formationen. Durch die Fusionierung von Satelliten- und Landbedeckungsdaten konnten die räumlich und thematisch hochaufgelösten Landsat-Daten mit den phänologischen Informationen des nationalen Landbedeckungsdatensatzes für eine signifikante Verbesserung der Modellergebnisse kombiniert werden. Um die Fehlerfortpflanzung innerhalb des Modells zu minimieren, besteht daher die Notwendigkeit der Verwendung akkurater, detaillierter und aktueller Landbedeckungsinformationen sowie zusätzlicher Referenzdaten.

Die vergleichende Analyse zeigt, dass die in dieser Arbeit erstellte Biomassenkarte im Vergleich zu fünf existierenden regionalen und globalen Biomasseprodukten eine höhere Übereinstimmung mit der Referenzbiomassenkarte des Waldministeriums (*Forest Department*) von Uganda aufweist. Weiterhin konnten deutliche Unstimmigkeiten zwischen den existierenden Produkten hinsichtlich der absoluten oberirdischen Biomasse (343 - 2201 Tg) sowie deren räumlichen Verteilungsmuster aufgezeigt werden. Vergleichend zu der auf Felddaten und dem nationalen Landbedeckungsdatensatz basierenden Referenzkarte, erfolgte eine deutliche Überschätzung der Biomasse (578 – 2201 Tg) durch die Produkte, welche auf Durchschnittswerten von Biomen basieren (z.B. Standardwerte des *Intergovernmental Panel on Climate Change*; globale Landbedeckungsdatensätze). Produkte basierend auf Satellitendaten und Regressionsmodellen hingegen liefern eher konservative Schätzungen (343 – 443 Tg). Bestätigt werden diese Ergebnisse durch den Vergleich der verschiedenen Produkten mit Felddaten die mittels eines Landbedeckungsdatensatzes auf die entsprechende Kartenauflösung hochskaliert wurden.

Die Studie zeigte weiterhin den grundlegenden Einfluss der Biomassereferenzdaten (Feldmessungen) auf die Biomasseschätzungen, während die Art der räumlichen Karte zur Stratifizierung (z.B. Landbedeckung, Biome) eine geringere, wenn auch nicht vernachlässigbare Relevanz hat. Die Unterschiede im Format, der Auflösung sowie der jeweiligen Biomassedefinitionen der verschiedenen Produkte wurden hierbei im gleichen Maße berücksichtigt wie auch die teilweise Abhängigkeit der Produkte zu den verwendeten Referenzdaten, mit denen die vergleichende Analyse durchgeführt wurde.

Die deutliche Unstimmigkeiten zwischen existierenden Biomassenprodukten und der große Einfluss der Biomassereferenzwerte auf die Schätzung zeigen, dass die Erhebung von akkuraten Biomassereferenzen im Gelände für alle relevanten Vegetationstypen der zunächst kritischste Schritt zur Verbesserung der Genauigkeiten von Biomassekarten darstellt. Allerdings sind hierfür detaillierte und räumlich akkurate Datensätze notwendig, die sowohl die spezifische und effektive Lokalisierung der Feldmessungen als auch deren Regionalisierung ermöglichen.

Table of Contents

ABSTRACT	IV
ZUSAMMENFASSUNG.....	VI
ACKNOWLEDGMENTS.....	X
1. INTRODUCTION	1
1.1. Relevance of tropical forest biomass	1
1.2. The role of sub-Saharan Africa	2
1.3. Uganda: a case study	2
2. SCIENTIFIC BACKGROUND	4
2.1. Biomass and climate change	4
2.2. The role of forest biomass in the UNFCCC and REDD+	6
2.2.1. The UNFCCC and the Kyoto Protocol	6
2.2.2. REDD+	7
2.3. Approaches for biomass retrieval	8
2.4. Biomass field measurements	9
2.5. Biomass retrieval from Remote Sensing	10
2.5.1. Sensors: optical, radar or lidar?	12
2.5.2. Optical Remote Sensing for biomass retrieval	14
2.5.2.1. Physical basis: the spectral properties of vegetation	14
2.5.2.2. Image processing and variable selection	15
2.5.2.3. High and Low resolution sensors	17
2.6. Modeling approaches	18
2.6.1. Parametric models	18
2.6.2. Non-parametric models	18
2.7. Comparison and validation of biomass datasets	19
3. RESEARCH OBJECTIVES.....	21
3.1. Scientific needs	21
3.2. Research questions and objectives	22
3.3. Methodology	23
3.4. Outline of the thesis.....	24
4. RETRIEVING A NEW BIOMASS MAP OF UGANDA	26
4.1. Study area.....	26
4.1.1. Uganda and the National Biomass Study (NBS)	26
4.1.2. Budongo Forest Reserve.....	27
4.2. Data and methods	28
4.2.1. Landsat data.....	28
4.2.1.1. Image pre-processing	30
4.2.1.1.1. Image registration	30
4.2.1.1.2. Radiometric Correction.....	30
4.2.1.1.3. Image segmentation	32
4.2.1.1.4. Water and clouds mask	33
4.2.2. Land Cover data.....	34
4.2.3. Field data	34
4.2.3.1. The NBS field plots.....	34
4.2.3.1.1. Selection of the NBS field plots.....	36
4.2.3.2. Field campaign in the Budongo Forest Reserve.....	38
4.2.3.2.1. Selection of the allometric equation.....	39
4.2.3.2.2. Biomass field dataset in the Budongo Forest Reserve	43

4.2.3.3.	The biomass field reference dataset for Uganda	44
4.2.4.	Modeling approach	45
4.3.	Results	47
4.3.1.	Models results.....	47
4.3.2.	Model comparison	51
4.3.3.	A new biomass map for Uganda.....	52
4.4.	Discussion	54
5.	COMPARISON OF EXISTING BIOMASS MAPS OF UGANDA	60
5.1.	Biomass datasets	60
5.1.1.	Maps description.....	60
5.1.2.	Mapping approaches.....	61
5.2.	Methods.....	63
5.2.1.	Analysis of biomass maps	63
5.2.1.1.	Map pre-processing.....	63
5.2.1.2.	Map comparison statistics.....	64
5.2.2.	Integrating biomass maps with field data	65
5.2.3.	Analysis of biomass reference values and spatial data	66
5.3.	Results	67
5.3.1.	Comparison of biomass maps	67
5.3.1.1.	Total aboveground biomass of Uganda.....	67
5.3.1.2.	Other estimates of Uganda’s biomass	69
5.3.1.3.	Spatial distribution of biomass.....	69
5.3.1.4.	Comparison of biomass stocks in forest and non-forest areas.....	74
5.3.1.5.	Comparison of biomass density values in forest areas.....	75
5.3.2.	Comparison of maps with field data.....	76
5.3.3.	Comparison of biomass reference values and spatial data.....	77
5.4.	Discussion	79
6.	CONCLUSIONS.....	82
6.1.	Optical Remote Sensing for biomass retrieval	82
6.2.	Comparison of biomass maps of Uganda	83
6.3.	Future research	84
	REFERENCES	86
	LIST OF FIGURES.....	100
	LIST OF TABLES.....	102
	LIST OF ACRONYMS	103
	APPENDIX A.....	105

Acknowledgments

This thesis is the result of a long journey, during which many people walked with me. Without doubts, I would not have reached this destination without their company, support, or just presence. The following acknowledgments do not follow any specific order because I cannot rank the invaluable support of these persons.

This journey started thank to Christiane Schmullius, who believed in me since our first meeting and heartily welcomed me in her group. Nonetheless her uncountable duties and travels, she always found the time to advice and support me, and gave me the freedom and space to wander and benefit from many opportunities.

One of these opportunities was the time spent at Boston University. This was such an intense period where I received an excellent training and learned to think scientifically. This “impossible” stay became possible thanks to Curtis Woodcock, who sponsored my stay for so long. Curtis, Mark Friedl and Alan Strahler welcomed me very nicely in their classes and with great patience supported my slow progresses. Without doubts, their teachings form the foundation that guided and will guide my research work. In Boston, I also benefitted from the teachings of Suchi Gopal and I truly enjoyed the company of Xiaoman, Manish, Nyanga, Megan, Sangram, Mitch, Alex, Moses, Craig and all the LCA group.

Even before Boston, and after it, Alessandro Baccini was present. During such long time, Ale constantly inspired me, provided crucial contribution to my research and supported me, especially during difficult times. I feel really thankful for his help and friendship.

Martin Herold was another great source of inspiration: his enthusiasm, energy, positive thinking and smart suggestions helped me since the beginning of my research. Martin’s support grew constantly during this journey and more and more I am learning from him and enjoying his guidance.

The EO Jena group welcomed me very nicely in their department and, nonetheless my non-existent German-speaking skills, I never felt alone thank to the company of Nico, Ralf, Roman, Reik, Michi, Claudia, Jacqueline, Johannes, Robert, Tanvir and many others. I also benefitted from the smart conversation and sharp suggestions of Sören Hese and Markus Reichstein. Apart the university gang, I shared lots of good time and pasta with Basti, Sophie and Mirko, and enjoyed the presence of Christian. Then, Anna appeared, and Jena took a new light.

Back in Italy, I would like to heartily thank Paolo Sarfatti and Luca Ongaro, who made this journey possible and supported me in many ways for a long time, along with Antonio Ciuchini, who always somehow managed to fix various administrative issues. I would like to thank a lot Riccardo Valentini for giving me this great chance, and Antonio Bombelli for his friendly help and prompt assistance, especially regarding the non-linear European reporting bureaucracy. During various project meetings, I also had the luck to share time and interesting conversations with Matieu Henry and Danae Maniatis.

The great people that I met in Woods Hole provided me inspiration and confidence to start the PhD, and during my stay I shared and received a lot especially from Jared, Nadine, Emily, Wayne, Josef, Scott and Dan.

I would like to gratefully thank the Uganda field team, who worked so hard in the beautiful forests and savannas of Budongo to acquire the data necessary for my research.

I deeply thank the person from CGIAR that shared his dataset with me, nonetheless he did not know me. This dataset represents the backbone of my research and no meaningful results would have been achieved without it.

Throughout this long time, Sara was always there. There are just no words to express my gratitude and many other beautiful feelings for her constant support, love and smile. Grazie di cuore.

And, my family, who always supported me. Wherever part of the world I was living, they were always close to me, no matter what; they constitute my foundation.

1. Introduction

1.1. Relevance of tropical forest biomass

Vegetation biomass is a proxy for several ecosystems services, affects the local and regional climate and provides a variety of products such as timber, fuelwood and biofuel. Biomass is also relevant information for carbon cycle budgeting because it consists of approximately 50% carbon (Malhi et al., 2004).

Large amount of biomass are stored in the tropics (Houghton, 2005). Within the climate change debate tropical forests are attracting increasing attention because of their role as carbon sink (Stephens et al., 2007; Malhi, 2010) and the large Greenhouse Gases (GHG) emissions associated with their disappearance (Houghton, 2007; Le Quéré et al., 2009). Therefore, assessment of the amounts and dynamics of biomass in the tropics is crucial for reducing uncertainties in global carbon modeling (Le Quéré et al., 2009; Houghton, 2005) as well for national planning in many tropical countries, where vegetation is a primary source of products and energy for local communities (Masera et al., 2006).

Accurate biomass estimates are also required for the implementation of a reliable mechanism under the United Nations Framework Convention on Climate Change (UNFCCC) to reduce emissions from tropical deforestation and forest degradation below an historical reference level and for the conservation, sustainable management and enhancement of forest carbon stocks in developing countries (REDD+) through appropriate financial incentives or carbon markets (Gullison et al., 2007, UNFCCC, 2010). The implementation of such mechanism, currently under discussion by the UNFCCC, poses several challenges (e.g. political, economic, social) (Stickler et al., 2009) but first requires the accurate monitoring of the forest carbon stocks and their dynamics at national level (UNFCCC, 2009). While there is high interest in seeing such initiatives take form, monitoring forest biomass stocks and stock changes is identified as a key challenge for developing countries wishing to take part in the expected REDD+ mechanism (Herold & Skutsch, 2011).

Currently, there are still large uncertainties associated with the amount and distribution of carbon stocks in the tropics (Waggoner, 2009; Houghton and Hackler, 2006). Monitoring and assessment of biomass resources is an expensive and time consuming task, complicated by the large extents and difficult accessibility of tropical forests. As a result, few developing countries have put in place an efficient monitoring system, and current biomass estimates for these countries are highly variable (Houghton et al., 2001; Houghton, 2005; Gibbs et al., 2007).

Given the relevance of tropical forest biomass for several applications and the high uncertainty of its distribution, this topic is currently a priority research area.

It is widely recognized by the scientific community that remote sensing provides the key source of data for updated, consistent and spatially explicit assessment of biomass and its dynamics, especially in large countries with limited accessibility (Penman et al., 2003; Herold and Johns, 2007; De Fries et al., 2007; UNFCCC, 2008; GOFCC GOLD, 2010) but there is no agreement on what method should be used to reliably determine this parameter (Goetz et al., 2009, Gibbs et al., 2007).

1.2. The role of sub-Saharan Africa

Most research on tropical forest biomass carried out during the last years has focused on the Brazilian Amazon (Houghton et al., 2001; Lu, 2005; Saatchi et al., 2007; Stickler et al., 2009; Li, 2010; Asner, 2010) and Southeast Asia (Foody et al., 2003; Field et al., 2008; Inoue et al., 2010; Kohler et al., 2010; Descloux et al., 2011) because in these regions large-scale deforestation activities caused major GHGs emissions. Instead, despite the African continent was identified as the most vulnerable to climate change and a priority region by the UNFCCC (2006), sub-Saharan Africa has been largely ignored in large-scale biomass assessment because the contribution of this region to global GHG emissions is relatively small (Malhi, 2010). In addition, assessing biomass in Africa is severely constrained by difficult access to forest areas due to scarce infrastructure and political instability. As a consequence, Sub-Saharan Africa is currently one of the regions with lowest availability of in-situ biomass data (Bombelli et al., 2009, Ciais et al., 2009) and with the highest uncertainty regarding the amount and distribution of carbon stocks (Houghton and Hackler, 2006).

However, having the fastest population growth in the world and a rising economic growth rate, Africa's share of global GHG emissions is expected to increase in the coming decades (Canadell et al., 2009). Since deforestation and forest degradation are already responsible for a substantial part of Africa's greenhouse gas emissions (Houghton and Hackler, 2006; Bombelli et al., 2009), in the near future African forests will have a growing impact on climate change, or a large potential for its mitigation.

In order to quantify such impact/potential, uncertainties on amount and distribution of carbon stocks in African tropical forests need to be significantly reduced. Such need has stimulated recent studies to focus on Sub-Saharan Africa and updated biomass estimates at regional scale have been recently produced (Henry, 2010; Baccini et al. 2008; Gibbs & Brown, 2007; Drigo, 2006). However, while these products can support global and regional carbon modeling, their resolution and accuracy are generally not sufficient for the implementation of activities at national level, such as REDD+ or monitoring bio-energy resources. For this reason, studies at country-scale are highly required.

1.3. Uganda: a case study

Within sub-Saharan Africa, Uganda is a country highly dynamic in terms of biomass stock and represents a case study of primary interest. The high economic and population growth experienced in the last 25 years has had a dramatic impact on the forestry resources, which have been reduced by agricultural expansion and growing demand for charcoal, fuelwood and timber (Drichi, 2003; FAO, 2003). Only between 1990 and 2005, the country's total biomass stock decreased by 26% (FAO, 2006).

Most importantly, Uganda is expected to experience continued rapid population and economic growth in the coming years, with associated smallholder agricultural expansion into forested areas and increasing demand for forest products that will likely accelerate deforestation (Kanabahita, 2001; FAO, 2003). With 66% of the total Aboveground Biomass (AGB) located outside of protected areas (Drichi, 2003) and insufficient incentives to pursue sustainable forest management (Kanabahita, 2001), Uganda is therefore highly relevant to REDD policies and biomass monitoring activities.

In addition, Uganda presents a long history of biomass assessment, supported by extensive biomass and land cover field measurements (Drichi, 2003). Since the availability of a reliable national biomass reference dataset is a key factor to successfully develop and test new methods for biomass retrieval, Uganda represented an optimal case study.

2. Scientific background

2.1. Biomass and climate change

Climate change is (almost) unanimously considered the most important environmental crisis of our time. The relevance of climate change is due to its major ecological impacts on natural environments and the tremendous social and economic consequences on human systems. Climate change is a consequence of the increase of the atmospheric concentration of GHGs compared to pre-industrial levels due to human activities, mainly burning of fossil fuels and land cover changes. Due to their radiative forcing, GHGs alter the energy balance of the climate and tend to warm the atmosphere. Warming of the climate is unequivocal, and there is very high confidence that the net effect of human activities is that of warming (IPCC, 2007).

Carbon dioxide (CO₂) is the most important anthropogenic GHG, accounting for almost 77% of global anthropogenic GHG emissions (IPCC, 2007). Atmospheric CO₂ concentration increased of 36% (from 280 to 379 ppmv) between 1750 and 2005 and it continues to increase exponentially. Only between 1970 and 2004, CO₂ emissions increased of 80%. Fossil fuel combustion and land use changes are the main sources of CO₂, contributing respectively to 65% and 35% of global historical emissions since 1750 (IPCC, 2007).

Within the land use sector, the vast majority of emissions are due to the alteration of global vegetation through deforestation and forest degradation.

Plants absorb CO₂ through photosynthesis and release it through respiration, decomposition and combustion. By storing almost as much C as the atmosphere and being the driver for the C accumulation in soils, terrestrial vegetation plays a critical role in the global C cycle and its changes are highly relevant for climate change (Figure 1). In addition, being the most dynamic component of the terrestrial C cycle, forest disturbances as logging or fires have a rapid impact on atmospheric CO₂ concentration.

Globally, vegetation is both a strong C sink and a relevant C source by sequestering about 2.6 GtC and releasing about 1.6 GtC yearly during the 1990s (IPCC, 2007). The main source of CO₂ emissions from the land use sector is tropical deforestation (Le Quéré et al., 2009), contributing to 20% of global emissions during the 1990s (IPCC, 2007). Considering the rise of fossil fuel emissions of the last decade, the share of emissions from tropical forests has decreased but still remained substantial, accounting for 15% of global emissions during the period 2000 – 2005 (Malhi, 2010) and 12% in 2008 (van der Werf et al., 2009) without considering emissions from tropical peatlands. Tropical forests are also a strong C sink and almost counterbalance emissions from tropical deforestation, absorbing about 12% of global emissions during the period 2000 – 2005 (Malhi, 2010). It is important to notice that the net carbon balance of tropical forests varies considerably among the continents: tropical Asia appears to be a large C source, South America a moderate C sink and tropical Africa a large C sink (Malhi, 2010).

Nonetheless the relevance of the land use sector on the global C budget, the land use C source has the largest uncertainties, with estimates of emissions ranging from 0.5 to 2.7 GtC year⁻¹ during the 1990s (IPCC, 2007). Most of the uncertainty is due to tropical deforestation (Le Quéré et al., 2009), which contribution to global emissions in 2008 ranged between 6% and 17% (van der Werf et al., 2009). Similarly, changes in the C sink

are highly uncertain and reducing their uncertainties is crucial considering the significant impact of this sector on future atmospheric CO₂ levels (Le Quéré et al., 2009).

Since biomass of vegetation consists of approximately 50% carbon (Malhi et al., 2004), the amount of C stored in vegetation is directly related to their biomass density.

Aboveground biomass is defined as the total amount of aboveground living organic matter in trees expressed in units of oven-dry weight (Brown, 1997). Biomass density represents the mass per unit area and is usually expressed as tons (Mg) per hectare, while total biomass of a certain region is obtained by multiplying its mean biomass density with the corresponding area.

Carbon emissions from tropical forests depend on the area of forest change and the associated biomass loss. Consensus regarding the area affected by deforestation and degradation has yet to be reached (van der Werf et al., 2009), but the accuracy of the estimates increased during the last decades thanks to the use of satellite data, and operational forest monitoring systems are now a feasible goal for most developing countries (DeFries et al., 2007).

Instead, the amount of biomass stored in tropical forests and its distribution remains highly uncertain at regional and national level (DeFries et al., 2007; Houghton, 2005). Since assessing biomass stocks is as important as quantifying the rate of deforestation to reduce the error of estimated net fluxes of carbon (Houghton, 2005), improving the knowledge of tropical biomass distribution has become a crucial component to understand the role of tropical forests on the C cycle and their potential for mitigation and adaptation to climate change.

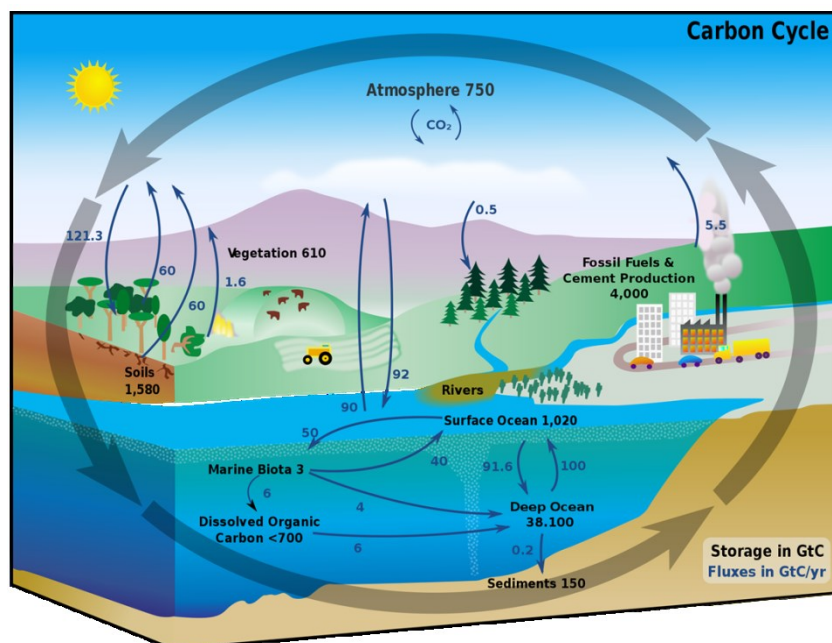


Figure 1: The Carbon cycle (Source: NASA, Earth observatory, http://earthobservatory.nasa.gov/Features/CarbonCycle/carbon_cycle4.php)

2.2. The role of forest biomass in the UNFCCC and REDD+

2.2.1. The UNFCCC and the Kyoto Protocol

The United Nations Framework Convention on Climate Change (UNFCCC) is an international environmental treaty produced during the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992, also known as the "Earth Summit". The UNFCCC was conceived along with the Convention on Biological Diversity (CBD) and the United Nations Convention to Combat Desertification (UNCCD) but it quickly gained a role of primary importance regarding environmental issues and, compared to the other conventions, obtained stronger recognition and support from the international community.

The UNFCCC is aimed at stabilizing GHG concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system. All signatory countries are required to periodically report to the Convention an inventory of all anthropogenic GHG emissions from sources and removals from sinks, including information on deforestation and forest degradation at national scale. Reporting guidelines for national inventories of CO₂ emissions and removals by land use sector were provided by the Intergovernmental Panel on Climate Change (IPCC) (IPCC 1996, 2006).

While the UNFCCC is legally non-binding because it does not set mandatory limits on GHG emissions nor contain enforcement mechanisms, such aspects were introduced in 1997 with the Kyoto Protocol. On the basis of the principle of "common but differentiated responsibilities", the Kyoto Protocol requires 37 developed countries (industrialized countries and economies in transition, also called the "Annex I countries") to reduce their collective GHG emissions by 5.2% with respect to the 1990 levels during the period 2008 - 2012. Flexible mechanisms, as the Clean Development Mechanism (CDM), were included to allow Annex I countries to acquire emission credits in order to fulfill their reduction commitments. The Kyoto Protocol recognizes the relevance of forests in reducing GHG emissions (Article 3.3 and 3.4) but the accounting rules were highly controversial and ultimately tropical forests were excluded from the Protocol, apart for the afforestation and reforestation activities under the CDM.

In the almost 20 years since the Earth Summit, the conditions of the natural resources have constantly deteriorated and, nonetheless the efforts led by the United Nations, the implementation of the conventions and their impact on the environment was limited by several challenges (Millennium Ecosystem Assessment, 2005). Moreover, while the rise of 2°C in global temperatures above pre-industrial levels is widely considered as a limit to avoid serious effects of climate change, basing on the pledges for voluntary emissions reduction of the Copenhagen Accord and related projected emissions, the United Nations Environment Programme (UNEP) estimated changes in global mean temperature between 2.5 to 5 °C relative to pre-industrial temperature levels until the end of the 21st century (UNEP, 2010). Therefore, more efforts need to be done in the coming years to stabilize the GHG concentrations within levels that allow natural adaptation of the ecosystems, sustainable economic development and secure food production.

2.2.2. REDD+

During the last years, the recognition of the significant contribution of tropical deforestation to global GHG emissions has brought this topic at the center of the UNFCCC negotiations. The political and scientific communities recognized the importance of forests for both mitigating and adapting to climate change (IPCC, 2007). However, while an effective strategy against climate change must include tropical forests, according to the principle of “common but differentiated responsibilities” reducing deforestation in developing countries should not limit their development.

A proposal to include financial incentives to reduce emissions from deforestation in developing countries (REDD) was presented during the 11th Conference of the Parties (COP-11) of Montreal in 2005 and then included in the negotiation of the post-Kyoto agreement at the 2007 COP-13 in Bali (UNFCCC, 2008). During the following negotiations the issue of reducing emissions from deforestation and forest degradation in developing countries gained increasing consensus and was expanded to include also the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries (REDD+). The discussions focused on social, economic, technical and methodological issues, as well as policy approaches and modes of compensation. During the 2009 COP-15 in Copenhagen there was an agreement at the methodological level on REDD+ (UNFCCC, 2009), followed in 2010 by commitments from several developed countries during the Oslo Climate and Forest Conference to provide consistent funds for supporting REDD+ policies and measures. Lastly, the policy and mechanisms for implementing REDD+ were specified and agreed during the 2010 COP-16 in Cancun (UNFCCC 2010), assuring that the mechanism will be included in the post-Kyoto agreements.

The methodological framework defined in the COP decisions indicates that developing countries wishing to take part in the REDD+ mechanism will have to monitor their forests and associated carbon stocks and stock changes. Specifically, the UNFCCC requires the establishment of national Monitoring, Reporting and Verification (MRV) systems and the use of the most recent IPCC Guidance and Guidelines to assess anthropogenic GHG emissions by sources and removals by sink in forest areas, forest carbon stocks and forest area changes. Using a combination of remote sensing and ground observations, MRV systems will need to report to the UNFCCC transparent, consistent and accurate estimates of forest carbon stocks and forest area change, and ultimately gain the confidence of international donors and carbon markets.

Currently, while there is high interest in seeing the successful implementation of national-level REDD+ programs, more work remains to be done to solve policy and scientific issues, such as addressing permanence, additionality, leakage or the development of reference emission levels. While the design of MRV systems is still under discussion (Herold and Skutsch, 2011), monitoring forest biomass at national scale remains one of the key steps for developing REDD+ activities, and a priority research topic for the scientific community.

2.3. Approaches for biomass retrieval

The large disagreement among biomass estimates for tropical countries provided by different studies have stimulated a debate regarding what method should be used to reliably determine this parameter (Goetz et al., 2009, Gibbs et al., 2007).

Traditionally, biomass stock of a certain area is estimated by multiplying the mean biomass density of each forest type with its extent, and by summing the corresponding products. Field data, usually acquired from the national forest inventory, are used to estimate the average biomass density of the forest types while their extent is derived from the cadastral system, based on field survey of land uses. By using summary statistics of land uses, statistically robust estimates of biomass stock can be obtained without the need to describe their spatial distribution. This “non-spatial” approach is widely applied in developed countries where cadastral systems are well-established and frequently updated but it is often not applicable in developing countries where land tenure and cadastral infrastructure are poorly developed and not regularly maintained. In tropical regions and particularly in sub-Saharan Africa most of the countries do not have the technical and financial capacities to assess land use area through field measurement, and national forest inventories are often rare and outdated. In such countries land use information are mainly obtained from indirect measurements using remotely sensed data, such as satellite images or aerial photos. Considering the costs and technical difficulties associated with the acquisition of aerial photos, satellite images are often the only viable solutions for large developing countries.

While the use of remotely sensed spatial datasets requires sophisticated computer analysis and may introduce artifacts, this approach has the advantage to provide temporally consistent and spatially explicit estimates (maps) with complete coverage of the study area. Remote sensing data can be used to map land cover/use or to directly estimate biomass density (Goetz et al., 2009). In both cases, maps describing the spatial distribution of biomass density are produced.

With regard to national greenhouse gases inventories under the UNFCCC, the Good Practice Guidance (GPG) of the Intergovernmental Panel on climate Change (IPCC) indicates that both strategies (spatial and non-spatial estimates of biomass density) fulfill the requirements for reporting emissions in the Land Use, Land Use Change and Forestry (LULUCF) sector. With regard to REDD+ there is a growing interest on spatially explicit biomass estimates because biomass density at specific locations, such as areas affected by deforestation, can be significantly different from average values (Houghton, 2005; GOFCC GOLD, 2010). Therefore, high resolution biomass maps can better quantify emissions from forests with more (or less) than average biomass density where assuming average biomass values would underestimate (or overestimate) the corresponding emissions (Houghton, 2005; Houghton and Hackler, 2006).

The use of spatial datasets to complement and spatialize existing biomass ground measurements has become more frequent during the last years, when satellite data have been especially used for biomass assessment of large countries with high biomass variability and limited field data. For instance, MODIS data were used to estimate forest biomass throughout the Russian Federation (Houghton et al., 2007) and woody biomass across tropical Africa (Baccini et al., 2008). Remotely sensed land cover datasets as GLC2000 (Gibbs et al., 2007) and GlobCover (Henry, 2010) were the main predictors to map biomass distribution in sub-Saharan Africa. Similarly, the Africover land cover database was used by Drigo (2006) to assess wood resources for Eastern Africa. A

combination of MODIS, land cover and ancillary information was used to map forest biomass for the conterminous United States (Blackard et al., 2008) while a combination of metrics derived from optical (MODIS) and radar (JERS-1, QSCAT and SRTM) data were employed to estimated AGB distribution of the Amazon basin (Saatchi et al., 2007).

2.4. Biomass field measurements

There is unanimous consensus that field measurements are essential to estimate biomass density at specific locations and for calibrating and validating remote sensing models for large-area biomass mapping. Since the accuracy of remotely sensed biomass estimates ultimately depends on the accuracy of the reference field data, acquiring reliable ground measurements is crucial for large area biomass assessment.

Aboveground biomass, as considered in this study, is defined as the total amount of aboveground living organic matter in trees expressed in units of oven-dry weight (Brown, 1997). Biomass density represents the mass per unit area and is expressed as tons (Mg) per hectare, while total biomass of a certain region (Mg) is obtained by multiplying its mean biomass density with the corresponding area.

Tree biomass can be accurately quantified on the ground by logging, oven-drying and weighing all tree components but, since this procedure is destructive and extremely time consuming, tree biomass is usually derived from plant parameters as diameter at breast height, tree height or wood density using allometric equations. Still, measuring plant parameters is a time consuming and expensive task that can be carried out only on a limited sample area. In addition, while biomass values derived from field measurements are often considered as “ground-truth”, these values are estimates, not without errors, of actual vegetation biomass. In fact, when biomass over a certain area is estimated from sample plots, there are several sources of errors, mainly incorrect tree measurements, inappropriate allometric model, insufficient plot size and scarce representativeness of the plots in a heterogeneous landscape (Chave et al., 2004). Uncertainties regarding wood density (or wood specific gravity), a species-specific coefficient representing the ratio of dry weight to green volume, are also relevant for biomass estimation (Chave et al., 2006).

Among these factors, the most important source of error is related to the allometric model (Chave et al., 2004). Thus, the selection of an appropriate equation is crucial to reduce the uncertainty of the biomass estimates. While species and site-specific allometric relationships can provide accurate estimates of forest biomass, such equations are usually not available for tropical forests, where ecological variability and species richness is very high. In such contexts, reliable biomass estimates can be obtained using generalized allometric equations (i.e. equations applicable to all species) stratified by forest type or ecological region, which are based on large datasets and can provide robust estimates over large areas (Brown, 1997, Chave et al., 2005). Similarly, while the amount of plant parameters that are input to the allometric model is proportional to the accuracy of the biomass estimate, measuring the height or identifying the species (to derive wood density) of every tree is very time consuming and often not possible in dense tropical forests due to the scarce visibility of the canopy or the large variability of tree species. However, since tree Diameter at Breast Height (DBH) is the most important predictor of tree biomass (Chave et al., 2005) and can explain alone a large part of its variation (up to 95% in

specific conditions) (Brown, 2002), reliable biomass estimates can also be obtained on the basis of the DBH only, which is an easy and time-effective measurement.

Generalized allometric equation for the tropics available in the literature differ significantly regarding their representativeness and can provide very different biomass estimates for given plant parameters. For example, the comparison of 16 allometric models estimating AGB as function of DBH for different environmental conditions showed that the biomass estimates at tree level vary by several orders of magnitude (Figure 2) (Baccini et al., 2009). Therefore, the applicability of generic allometric equations to specific conditions must be critically evaluated, especially when the equation is applied in contexts not included in the sample plots used to develop or validate the allometric model (Brown, 2002). This is especially relevant when generalized equations are applied to African forests because most of the published allometric models are based on field data collected in South America and Southeast Asia (Chave et al., 2005).

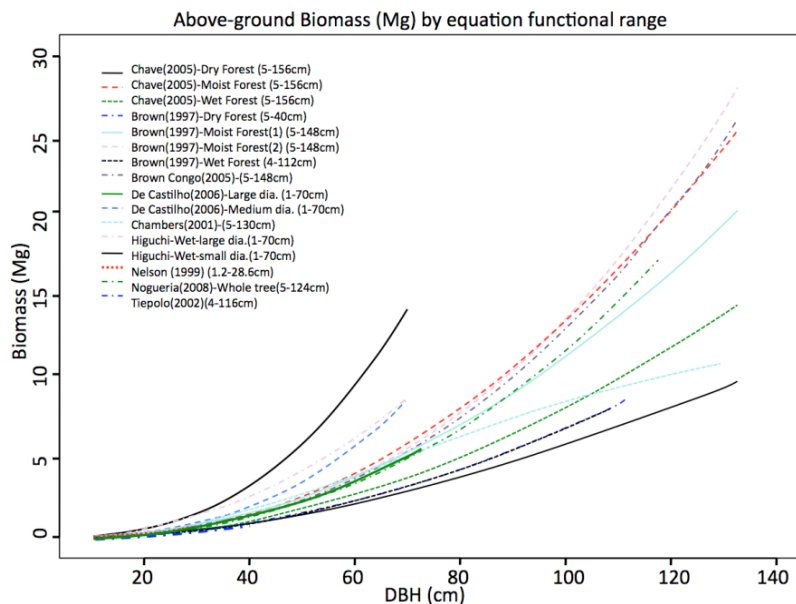


Figure 2: Tree biomass for Diameter at Breast Height (DBH) according to 16 published allometric equations. The author, the ecozone and applicable diameter range for each equation are reported in the figure (from Baccini et al., 2009)

2.5. Biomass retrieval from Remote Sensing

Remote Sensing consists on obtaining information of an object, area or phenomenon using a device that is not in contact with the studied object, area or phenomenon (Lillesand et al., 2008). Applied to Earth Observation, remote sensing consists on measuring properties of objects on the earth's surface using data acquired from sensors located onboard of aircraft or satellites (Schowengerdt, 2007). Specifically, earth observation satellites acquire repeated and consistent spatial information on land surface characteristics and such information can be related to biophysical parameters, as aboveground biomass (AGB) density of vegetation.

Application of remote sensing to tropical forests is particularly challenging because of complex and variable forest structure (Nelson et al., 2000; Steininger, 2000; Lu, 2005) and difficulty obtaining high quality remote sensing data and corresponding ground data sets (Foody et al., 2003). Nonetheless, considering that extensive biomass field campaigns are costly and time consuming, remote sensing is the only source of data for biomass monitoring in large countries with scarce land use information and poor infrastructure (Herold and Johns, 2007; De Fries et al., 2007; GOFCC GOLD, 2010).

To this end, a number of methods have been proposed to map biomass distribution from satellite observations. These methods, reviewed by Boyd and Danson (2005), Lu (2006) and Goetz et al. (2009), are based on a direct or indirect approach depending if they relate the remotely sensed data to biomass density in a direct (mapping biomass stock) or indirect way (mapping forest cover and forest parameters). The direct approach derives AGB directly from the satellite data using classification techniques (as neural network or regression trees) that empirically relate the surface signal to AGB density. Linear regression or model inversion techniques are also employed, especially in combination with active remote sensing sensors (Saatchi et al., 2007; Mitchard et al., 2009). Differently, the indirect approach uses satellite observations to map vegetation attributes as tree density or forest type. In many cases satellite data are also combined with other spatial datasets, such as population density, climatic parameters or surface topography, to derive a thematic map with finer-grained strata used to stratify the field data. The biomass stock is then calculated by combining the area assessment of the different vegetation classes with their biomass stocks, which is obtained by averaging the corresponding field measurements (Goetz et al., 2009).

In both approaches, remote sensing data are ultimately a mean to spatialize existing ground observations. In the direct approach ground data are needed to parameterize the model that relates the satellite signal to the biomass density while in the indirect approach the field observations are used to quantify the average biomass density of the vegetation strata.

Currently, there is no agreement in the scientific community whether AGB can be effectively monitored by remote sensing with the direct approach. As a matter of fact, existing satellites cannot measure biomass directly but instead they detect vegetation parameters that are related to AGB as tree density, tree height or vegetation structure. As a consequence, the relation between the satellite signal and AGB is empirical and site-specific, and the transferability over time and space of a model calibrated on a specific training dataset is limited (Foody et al., 2003).

Nonetheless these limitations, there is growing interest on the direct approach because it can provide continuous biomass estimates describing the full range of biomass variability at a spatial resolution usually higher than the indirect approach. Conversely, the indirect approach can identify only a limited number of strata (i.e. biomass classes) and cannot explain the intra-class variability, which can be large when the strata do not reflect accurately the biomass distribution, resulting in large errors for areas that diverge from average conditions.

For these reasons, space agencies are currently evaluating new satellite missions dedicated to direct biomass mapping, as BIOMASS from ESA (Le Toan et al., 2010). However, such advanced satellites are currently in the design stage and will not be operative in the near future, and alternative solutions must be found to quantify amount and dynamics of AGB in the tropical forests

2.5.1. Sensors: optical, radar or lidar?

Remote sensing sensors measure the energy contained in the electromagnetic radiation (i.e. radiance, measured as watts per unit source area (m^2), per unit solid angle (steradians, sr) and per unit wavelength (μm)) reflected or emitted by the Earth. Passive sensors (i.e. optical) measure the energy that is naturally reflected in the Visible, Near Infrared (NIR) and ShortWave Infrared (SWIR) spectral regions (from 0.4 μm to about 3 μm) or emitted in the Thermal Infrared (TIR) region ($> 5 \mu m$) while active sensors (i.e. lidar and radar) use an artificial source of radiation and measure the signal scattered back to the sensor in a wide range of wavelengths, from 0.7 μm to 1 m (Figure 3). At short wavelengths, active and passive sensors measure also the energy reflected or emitted from the atmosphere and clouds (Schowengerdt, 2007).

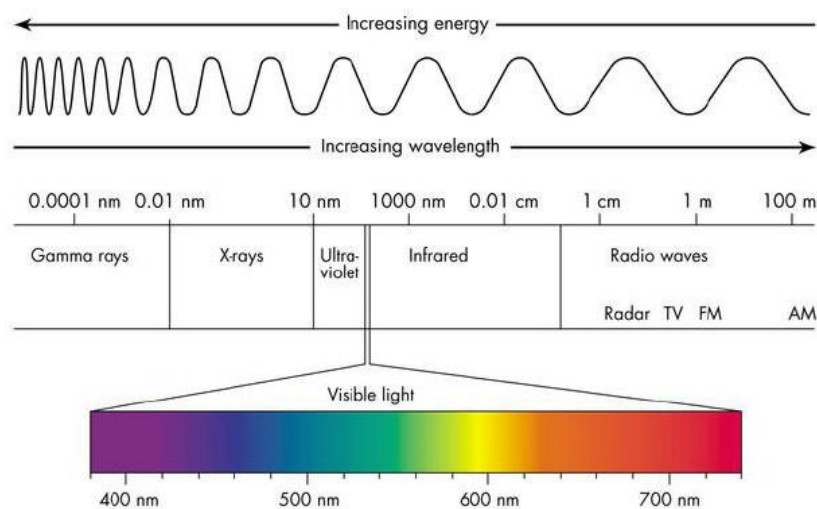


Figure 3: Overview of the electromagnetic spectrum

Scientific research carried out during the last decades has shown that existing satellite sensors, namely optical, radar and lidar, have potential to estimate biomass from space and that each sensor presents its specific capabilities and limitations.

Optical data have been widely used for vegetation monitoring since the '70s and several studies have demonstrated the sensitivity of visible and shortwave infrared wavelengths to vegetation density and structure, which in turn are related to AGB (Gemmell, 1995; Steininger, 2000; Foody et al., 2003; Lu et al., 2004; Baccini et al., 2004; Lu et al., 2005). The limitations of optical systems are mainly related to the fact that the signal is affected by atmospheric conditions and that is sensitive to the leaf characteristics more than to the woody components of vegetation. As a result, optical data are more correlated to vegetation density than to its vertical structure, which causes their saturation in closed canopy forest (Gemmell, 1995).

In comparison with optical sensors, radar presents the advantage of being an active sensor with a signal transparent to atmospheric moisture, thus radar acquisitions are not affected by cloud coverage or day/night cycles. Synthetic Aperture Radar (SAR) systems operating at large wavelengths (i.e. L-band, P-band) are suitable for biomass mapping because the microwaves penetrate into the forest canopy and interact directly with the woody elements

(stem, large branches), providing information on forest structural attributes including forest height and aboveground biomass (Kasischke et al., 1997; Rosenqvist et al., 2003). However, as optical sensors, radar data also shows an asymptotic relationship with biomass due to the saturation of the signal at high biomass values. In addition, the radar signal is affected by topography and is highly sensitive to the moisture content of the target.

LiDAR (Light Detection And Ranging) is an active sensor that uses laser pulses to directly detect vegetation height and structure. Because waveform LiDAR metrics can accurately characterize vertical forest structure and the interconnection of the latter with biomass (Drake et al. 2002; Lefsky et al. 2002), several studies have found a strong linear correlation between LiDAR data and AGB (Drake et al. 2002; Drake et al. 2003; Lefsky et al. 1999; Nelson et al. 1988). Moreover, LiDAR has shown no saturation effects at high biomass values (Drake et al., 2003; Sun et al., 2008) but, by operating at infrared wavelengths, the LiDAR signal is sensitive to cloud coverage.

Recently, the use of LiDAR remote sensing in biomass estimation has increased (Koch 2010). So far, most of the large-scale investigations have focused on sub-boreal forests (Boudreau et al. 2008; Nelson et al. 2009; Ranson et al. 2007) and few studies have applied airborne (Asner 2009; Drake et al. 2003) or spaceborne (Baccini et al. 2008; Lefsky et al. 2005) LiDAR data in tropical areas. Airborne sensors provide highly accurate biomass estimates (Ni-Meister et al. 2010), but the associated large data volume and high costs usually limit their application to local scales. As an alternative, the Geoscience Laser Altimeter System (GLAS) sensor on board the Ice, Cloud and Elevation satellite (ICESAT) satellite has proven to be valuable for biomass and canopy height estimation over large areas (Lefsky 2010; Lefsky et al. 2005). However, GLAS acquires data with a sampling strategy, which not allows spatially continuous representation of biomass or canopy variability, and the large ground footprint is not optimal for vegetation analysis. In addition, due to the failure of the last of its three lasers, GLAS data acquisition ended in October 2009 and ICESat ended its science mission in February 2010 (<http://icesat.gsfc.nasa.gov/icesat/index.php>).

By combining their different capabilities, the synergistic use of different sensors is expected to overcome most of their limitations. Ad example, GLAS data can be integrated with imaging optical systems to overcome the sampling coverage of the LiDAR system (Baccini et al. 2008; Goetz et al. 2010; Nelson et al. 2009). However, at present time the current acquisition strategies of radar and LiDAR satellites limit the compilation of multi-sensor datasets over large areas. While spaceborne LiDAR data (i.e. GLAS) are acquired with a sampling strategy, radar satellites (e.g. ALOS) provide continuous coverage but multi-temporal global consistent datasets suitable for biomass mapping are still not available.

The availability of radar and LiDAR data is expected to increase largely in the future thanks to the ongoing acquisition of the existing satellites (e.g. ALOS) and the launch of new satellites (e.g. ICESat II, BIOMASS, LIST, Tandem-X). Nonetheless, at present time optical satellites provide the largest archive of data in terms of spatial and temporal coverage (Cohen and Goward, 2004), with globally consistent datasets at resolution equal or lower than 30 m available free of charge (e.g. Landsat, MODIS) (Woodcock et al., 2008). For this reason, optical data still provide the preferable data source for large scale vegetation analysis.

2.5.2. Optical Remote Sensing for biomass retrieval

2.5.2.1. *Physical basis: the spectral properties of vegetation*

Optical remote sensing uses the energy reflected by the earth's surface in the visible, NIR and SWIR wavelengths to identify surface materials, as vegetation, soil or water. Specifically, each material is identified on the basis of its spectral signature, which describes its reflectance as a function of wavelength. While spectral signatures of most materials are available in spectral libraries, in practice there is not a perfect correspondence between a certain object type and a pre-defined spectral signature mainly because of natural variability of each material, atmospheric effects and coarse spectral quantization of the remote sensing system (Schowengerdt, 2007).

Spectral properties of plants have been thoroughly studied since the 1960s. Using laboratory and field analysis, pioneer studies investigated the eco-physiological mechanisms related to leaf absorption, transmittance and reflectance of light at different wavelengths (Gates et al., 1965), the seasonal variation of leaf reflectance (Olson et al., 1962; Gausman, 1974) and its correlation with leaf water content (Tucker, 1980). Spectral analysis showed that vegetation can be easily distinguished from other materials because plants have a specific spectral signature, characterized by low reflectance in the green and red wavelengths due to chlorophyll absorption, a sharp increase near 710 μm (also called red edge), high reflectance in the NIR due to chlorophyll reflectance and two dips in the SWIR due to the leaf water absorption.

The spectral signature of vegetation is not constant among different plant communities but shows high variability according to leaf type, seasonality, growth stage, plant health and moisture content. For this reason, spectral analysis not only identifies presence of vegetation but can also provide additional information on species composition, stage, health or water content. However, when vegetation reflectance is measured from aircraft or satellites, more variability is due to the fact that reflectance at canopy level is significantly different from leaf reflectance. Moreover, most spaceborne sensors acquire data at a spatial resolution where the different spectra of several species and other materials present within the pixel, as soil and shadow, are mixed (Li and Strahler, 1992). As a consequence, the discrimination of physiognomy and species with optical data is problematic (Wulder and Franklin, 2003).

Instead, optical sensors are strongly correlated with vegetation structural parameters. Specifically, optical spectral bands are highly sensitive to vegetation canopy cover because increasing vegetation density corresponds to higher foliage absorption, which causes a decrease in reflectance in the visible wavelengths, and especially in red channel (Wulder and Franklin, 2003). Retrieval of vegetation canopy height or biomass density is more complicated because these parameters cannot be directly measured with optical data but are indirectly correlated with canopy structural parameters that can be detected with optical sensors, as tree density, crown size, leaf area index, texture, canopy water content and shadow fractions (Gemmell, 1995; Phua and Saito, 2003; Wulder and Franklin, 2003; Zhang and Kodragunta, 2006).

Being more related to the horizontal canopy characteristics than to its vertical structure, the optical signal presents a complex relation with vegetation biomass or vertical structural parameters, and tends to saturate in closed canopy forest. However, due to the variability

of canopy structural characteristics, optical data have shown sensitivity to AGB also in closed forests. Moving from monoplane to multi-layered forest, canopy reflectance properties change and so does biomass density (Lu et al., 2005). Spectral bands in the SWIR region have been found to be mostly correlated (with an inverse relationship) to volume and AGB because of their higher sensitivity and larger dynamic spectral range than other bands to canopy water content and amount of shadows in the canopy, which are directly correlated with AGB (Gemmel, 1995, Steininger, 2000, Baccini et al., 2004). Similarly, texture information becomes a useful biomass predictor in dense forests with articulated structure of the upper canopy (Lu, 2005). The correlation between detectable forest structural complexity and biomass saturates in mature forests after a certain threshold, which depends on forest characteristics but usually is around 150 – 200 Mg ha⁻¹ for tropical forests (Steininger, 2000; Lu, 2005).

2.5.2.2. Image processing and variable selection

Satellite measurements are affected by atmospheric, topographic and calibration influences. In order to extract reliable information on land surface characteristics from satellite data, it is first necessary to correct the spatial and spectral distortions present in the images. Distortions are eliminated (or, in most of the cases, reduced) through image processing procedures as image registration, topographic correction, radiometric normalization and atmospheric correction.

Remote sensing imagery presents spatial distortions due to scanner characteristics, satellite orbital geometry, flight disturbances and topographic effects. Such geometric errors can be systematic or random, and can be corrected modeling the sources of distortion (for platform-induced distortion) or applying a polynomial correction based on Ground Control Points (GCP) with known location. In order to correct for topographic effects, the use of a Digital Elevation Model (DEM) is required (Schowengerdt, 2007). Geometric distortions are especially relevant when satellite data are related to other spatial data because large errors are introduced by misregistration of the datasets. In particular, high geolocation accuracy is required when satellite data are compared with other images (e.g. for change detection analysis) and/or when are related to ground observations (e.g. for biomass analysis).

The spectral information contained in satellite images consists of Digital Numbers (DNs), which are values without physical meaning. Images acquired from different sensors or at different times cannot be directly compared because the DNs are affected by sensor and atmospheric characteristics.

While earth remote sensing avoid spectral regions (e.g. 2.5 – 3 μm or 5 – 8 μm) where radiation is absorbed by atmospheric constituents as water vapor or carbon dioxide, radiation in the visible, NIR and SWIR spectral bands is scattered and absorbed by air molecules, aerosol and particulate during their path from the sun to the earth and again from the earth to the sensor (Schowengerdt, 2007). For this reason, the radiance at sensor does not correspond to the radiance at surface, and radiometric calibration procedures are required when using several images acquired under different atmospheric conditions.

Specifically, inter-sensor data comparisons require only the sensor calibration, which consists on the conversion of image values (DNs) to at-sensor radiance using sensor calibration information. Instead, the comparison of satellite data acquired at different times

as well as the comparison of satellite data with laboratory or field reflectance data requires both the atmospheric correction, i.e. the conversion of at-sensor radiance to at-surface radiance using information on the view-path atmospheric conditions, and the solar and topographic corrections, i.e. the conversion of at-surface radiance to surface reflectance on the basis of solar irradiance, solar path transmittance and topographic data.

The complete radiometric calibration procedure is complex and requires data on atmospheric conditions at the time of satellite overpass that are usually not available and that are often substituted by default parameters or are extracted from the image itself, introducing uncertainty and errors in the calibration procedure. To avoid error accumulation, simple absolute or relative radiometric correction methods (or their integration) are recommended over more sophisticated approaches based on atmospheric modeling when independent atmospheric data at the time and location of the image acquisition are not available (Song et al., 2001; Schroeder et al., 2006; Schowengerdt, 2007). In addition, since several applications do not require the precise quantification of surface reflectance while it is sufficient that the images to be compared are on the same relative scale, relative radiometric normalization procedures are preferred. Relative correction consists on the normalization (i.e. conversion to the same relative scale) of the DN's of a set of images to a reference image using different approaches, as histogram normalization or selection of pseudo-invariant features (Schott et al., 1988, Hall et al., 1991).

After calibration, the spectral information provided by the satellite data can be used in their original spectral space (i.e. the original bands) or can be redistributed into a new feature space. The new spectral spaces are created using linear and non-linear transformation of the original bands and do not add new information to the original data but may present some advantages for specific applications (Schowengerdt, 2007).

Several spectral indices, such as the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974) or the Tasseled Cap components (Kauth and Thomas, 1976), were developed to effectively derive vegetation biophysical parameters (e.g. canopy cover) from continuous spectral values. Vegetation indices have the capability to reduce the impact of variable environmental conditions as sun view angles, shadows, soil background or atmospheric conditions on the vegetation reflectance, and therefore can improve the correlation of the satellite signal with studied parameter (Lu, 2006).

Vegetation indices have been widely used for biomass estimation. While some indices provided a significant contribution to biomass estimation in sites with complex vegetation structure, in most of the cases they were not better related to stand parameters than the original spectral bands (Foody et al., 2003; Lu et al., 2004). Similarly, image texture bands were found to provide additional information to the spectral bands only in some specific forest types (e.g. multi-layered forests) but did not provide relevant contribution in other conditions, especially in forests with simple structures. In addition, identifying texture metrics that are significantly correlated to AGB is difficult and varies greatly according to the site and image characteristics (Lu, 2005). Instead, some studies found that the inclusion of biophysical variables derived from the raw spectral bands substantially improved the accuracy of the biomass predictions. For example, Hall et al. (2006) included stand height and crown closure to increase the model results compared to those based only on the original bands in a case study in Canada, while Zheng et al. (2004) achieved higher accuracies using stand age as a predictor variable in a case study in Wisconsin (USA). However, most of these approaches suffer from the one-time one-space syndrome (Woodcock and Ozdogan, 2004), i.e. cannot be easily generalized to different conditions.

2.5.2.3. *High and Low resolution sensors*

Space-borne optical data for land application are mainly acquired at high resolution (≤ 30 m) and at moderate resolution (≥ 300 m). High resolution data have the advantage to provide spatial detail compatible with the size of vegetation units and biomass field observations but compiling a temporally and radiometrically consistent cloud-free dataset over large areas is not always possible. This issue is addressed by the large swath and frequent repeat cycle of moderate resolution sensors but retrieving AGB in mixed pixels is problematic, and the limited spatial detail misses the small-scale biomass variability (Lu, 2006). In addition, it is often difficult to relate field data with satellite observations because of mismatch in measurement scale or resolution (Baccini et al., 2007).

Very high resolution (VHR) data (≤ 5 m) provide detailed information on tree canopy and may be able to detect single tree crowns but at such resolution canopy and topography creates high spectral variability and shadows that cannot easily be related to AGB. In addition, SWIR bands, which are most important for biomass estimation, are not available at VHR, and creating consistent datasets over large area is not possible and/or very expensive. For this reasons, VHR data are mainly used as reference data to validate biomass predictions obtained with coarser resolution data (Lu, 2006).

For these reasons, high resolution data such as Landsat, ASTER and SPOT data are usually employed for biomass analysis at local scales (Sader et al., 1989; Roy and Ravan, 1996; Fazakas et al., 1999; Foody et al., 2001; Phua and Saito, 2003; Zheng et al., 2004; Lu, 2005; Muukkonen and Heiskanen, 2005; Labrecque et al., 2006; Hall et al., 2006; Zheng et al., 2007; Powell et al., 2010) while moderate to coarse resolution data (e.g. MODIS) are applied at regional scale (Dong et al., 2003, Baccini et al., 2004; Zhang and Kondragunta, 2006; Saatchi et al., 2007; Blackard et al., 2008; Baccini et al., 2008). Some studies have integrated both high and moderate resolution data for sub-national analysis (Tomppo et al., 2002; Muukkonen and Heiskanen, 2007). Instead, the capability of optical data to directly estimate AGB at high spatial resolution and at national scale in the tropics has not yet been satisfactorily investigated, and represents a primary research area.

Among the available high resolution satellites, Landsat is the only instrument providing global coverage and a nearly 40 years of data record (Williams et al., 2006; Goward et al., 2006). The revolutionary opening and free distribution of its archive (Woodcock et al., 2008) greatly facilitates the use of Landsat data for large-area analysis of vegetation and its dynamics (Powell et al., 2010, Cohen et al., 2010). Most importantly, data acquired by this satellite are sensitive to vegetation structure (Gemmell, 1995) and their spatial resolution (30 m) is compatible with the size of vegetation stands and biomass field plots. In addition, considering the difficulties to acquire and process consistent datasets over large areas, Landsat resolution represents an optimal compromise between spatial detail and large area monitoring. For these reasons, Landsat data represent a preferable data source for high-resolution and low-cost AGB analysis at national level for developing countries.

2.6. Modeling approaches

2.6.1. Parametric models

Parametric models assume a specific form of the probability distribution of the input data, usually the normal distribution, and a specific type of relation between the predictors and the response variable, usually the linear relation (Gotelli and Ellison, 2004). The most common parametric model for classification purposes (the maximum-likelihood) uses a probability model (normal distribution) to define the decision boundaries (Schowengerdt, 2007). The most common parametric models for regression purposes assume a linear relationship between one and another variable (or set of variables). In some cases, logarithmic and exponential relations can be transformed to linear relationships using mathematical functions. Linear models include univariate regression, analysis of variance, generalized linear models, multivariate regression and principal component analysis.

Linear models have been widely used to relate remote sensing measurements to biophysical parameters. A variation of linear regression called Reduced Major Axis regression (Larsson, 1993) was recently applied to predict forest parameters as percent cover (Schroeder et al, 2007) and basal area (Healey et al., 2006). With regard to biomass estimation, a review of the literature indicated that linear regression models have been used in several different contexts with variable results, but usually not achieving high model performance statistics. Multiple regression analysis is arguably the most common approach, especially for studies based on remote sensing data with medium spatial resolution, because of its capability to use simultaneously the information contained in several input bands (Lu, 2005). Multiple linear regression were used to estimate biomass from Landsat data in the tropics, such as in study areas in Brazil (Steininger, 2000; Lu, 2005), in Malaysia and Thailand (Foody et al., 2003), in India (Roy et al., 1996), as well as in temperate and boreal regions, such as in the USA (Zheng et al., 2004), Canada (Labreque et al., 2006) and Finland (Hame et al., 1997).

Parametric models have also been commonly employed to estimate tree volume and biomass from active remote sensing sensors. For example, linear regression models were employed with radar data in the Amazon (Saatchi et al., 2007), in Africa (Mitchard et al., 2009), in Australia (Austin et al., 2003), in the USA (Harrell et al., 1997), and in Sweden (Fransson et al., 2001) while linear regressions were employed with LiDAR data in Costa Rica (Drake et al., 2002; Nelson et al., 1997).

2.6.2. Non-parametric models

Several studies carried out during the last decade have shown that the assumptions of parametric models (e.g. linear relations, uncorrelated inputs) are usually not satisfied for the case of estimating biomass with remote sensing data, and instead demonstrated that the relation between the variables is not linear (Foody et al., 2001; Foody et al., 2003; Baccini et al., 2004; Muukkonen and Heiskanen, 2005; Li et al., 2010). For this reason, empirical non-parametric models, such as tree-based models, neural networks, K-nearest neighbors or support vector machines, have been increasingly used by the remote sensing community during the last years to relate biophysical measurements with remote sensing data (Fazakas

et al., 1999; Foody, 2003; Makela and Pekkarinen, 2004; Labrecque et al., 2006; Walker et al., 2007; Tomppo et al., 2009). Their strength (or weakness, from another point of view) compared to parametric models (e.g. linear regression) is that they do not assume any a-priori statistical distribution of the input data nor any specific form in the relation (e.g. linear) between the predictors and the response variable.

Tree-based models (Breiman, 1984) are simple yet robust and mature tools that have previously been successfully applied for biomass estimation using remote sensing data in different contexts (Baccini et al., 2004; Saatchi et al., 2007; Blackard et al., 2008; Baccini et al., 2008; Powell et al., 2010). These models can handle both continuous and categorical variables to predict continuous (regression trees) or categorical (classification trees) response variables.

Tree-based models use binary splits to recursively partition the dataset in subsets (“nodes”) with progressively higher homogeneity with regard to the response variable, and then assign the mean value (regression trees) or the most frequent class (classification trees) of the response variable as the node prediction. In order to reduce model over-fitting, the algorithms usually employ cross-validation to identify the optimal tree size that perform best on independent data (Breiman, 1984; De’ath and Fabricius, 2000).

Tree-based model performance has improved significantly with the application of ensemble methods. These methods generate several models using procedures such as bagging (Breiman, 1996) and boosting (Freund and Shapire, 1996) to resample with replacement from the training dataset and then aggregate the model results in predictions that have lower variance than those obtained from a single model.

Random Forest (Breiman, 2001) is an extension of tree-based models that extends bagging (i.e. constructing each tree with a different bootstrap sample of the data) by splitting the nodes using the best among a random subset of predictors chosen at each node. This somehow counterintuitive random choice of predictors ultimately produces a model more robust against overfitting and improves prediction accuracy (Breiman, 2001; Liaw and Wiener, 2002). The algorithm, as implemented in the open source software R (R Development Core Team, 2009), produces statistics on model performance and predictor importance.

Model performance is computed on the basis of a cross-validation approach. For each tree the algorithm makes predictions on the data not included in the bootstrap sample, called out-of-bag (OOB) data, and aggregates these predictions to compute the overall error rate that, in the case of regression trees, is expressed as the percent of variance explained (R^2) and Root Mean Squared Error (RMSE). The OOB error rate is considered a reliable estimate of the actual model performance on independent data if the training data are not autocorrelated (De’ath and Fabricius, 2000; Liaw and Wiener, 2002).

2.7. Comparison and validation of biomass datasets

It is widely recognized that accuracy assessment is an essential component of any mapping activity (Card, 1982; Cihlar, 2000). Validation of remote sensing maps quantifies their quality, identifies their limitations and ultimately builds confidence in the user community on the capabilities of remote sensing data for specific applications. With regards to biomass estimates, information on uncertainty is also critical in the REDD+ context, and the UNFCCC requires the estimates of carbon emissions to be transparent, consistent and

as accurate as possible in order to reduce uncertainties, as far as national capacities and capabilities allow (UNFCCC, 2009).

A proper validation of remote sensing estimates requires their comparison with accurate, independent and comparable ground observations. However, in the case of large-area biomass maps of tropical forests, accuracy assessment becomes a challenging task because of the lack of field validation datasets with comparable coverage and resolution. The limited validation of the biomass products is a critical issue that is somehow hindering their operational use for national assessment of biomass and C stocks. Since maps based on global or regional datasets may not be tailored to country specific circumstances, their applicability at national scales needs to be better understood with appropriate case-studies.

Biomass density of tropical forests has a high level of spatial variability and assessing the uncertainty of its estimates requires a large number of sample points (GOFD GOLD, 2010). However, measuring biomass in the field is an expensive and time consuming task, and ground observations are usually scarce in comparison with the variability of this parameter over the study area. For this reason, in most cases the available field data are used to produce the biomass estimates and the validation phase is performed using a small independent subset of the reference data or applying statistical procedures, as cross-validation or jack-knifing of the training data. However, such validation datasets are usually not sufficient to adequately represent the variability of the estimated parameter or are highly correlated with the training dataset. In addition, field data are usually affected by errors, and their quality has a large impact on the validation results: it has been demonstrated that even small errors in the reference data can introduce large errors in the accuracy assessment of the remote sensing products (Foody, 2010). Therefore, the quality of the accuracy assessment is influenced by the non-optimal sampling scheme, by errors in the ground data and by presence of spatial autocorrelation (Congalton, 1991; Friedl, et al., 2000; Foody, 2002), while the differences in spatial resolution between the ground observations (field plots) and the map units (pixels or polygons) require up-scaling procedures that introduce assumptions and approximations (Baccini et al., 2007).

All these issues affect the quality of the validation results and the understanding of the reliability of the biomass products, which in turn reduce the confidence of the international community in the biomass datasets and ultimately hinder their adoption for operational national assessment of biomass/carbon stocks or for legally binding agreements, as the UNFCCC post-Kyoto agreement.

Spatial comparison of biomass products is an alternative approach to evaluate their reliability (Houghton et al., 2001). Clearly, comparing remote sensing products does not assess their accuracy but high level of agreement among several products likely reflects reliable predictions while low level of agreement identifies problematic areas (Herold et al., 2008).

Equally important, a spatially-explicit comparative analysis of maps based on different approaches reveals common trends and dissimilarities in the datasets, and these trends can be related to the input data and methodologies used by the maps. Additional analysis can isolate and quantify the effect of specific input layers on the biomass predictions. The relevance of this process relies on the fact that it deepens the understanding of the strengths and weaknesses of the compared approaches, quantifies the impact of the input data and methods on the maps estimates, and provides results that can ultimately be translated in recommendations for developing an operational biomass mapping methodology.

3. Research objectives

3.1. Scientific needs

Accurate, detailed and spatially consistent monitoring of biomass stock and its dynamics at national scale in tropical countries is necessary for natural resources management, carbon cycle budgeting and implementation of a REDD+ mechanism. Specifically, REDD-related climate change mitigation policies require reliable estimates of forest biomass in order to increase the confidence in the emission reduction estimates, and related carbon credits, due to activities reducing forest loss or forest degradation.

The official figures on forest biomass stocks of developing countries provided by the national authorities and reported in the United Nations Food and Agriculture Organization (FAO) Forest Resource Assessment (FRA) reports are usually derived from forest inventory data that are designed for purposes other than biomass assessment and that are often incomplete or outdated. In addition, the biomass data are usually provided in an aggregated format (i.e. as national average values) that is not adequate for the above mentioned objectives, since forest biomass density presents a high spatial variability related to climate, soil and topographical variations as well as natural and human disturbance events, which usually occur at fine scales (Brown, 1997; Gaston et al., 1998). Moreover, due to the scarce transparency and poor validation procedures, the reliability of these figures has been questioned (Houghton, 2005; Gibbs et al., 2007).

For these reasons, in the last years a number of spatially explicit biomass and carbon stock datasets with moderate resolution were produced using satellite and ancillary data for tropical regions (Drigo, 2006; Saatchi et al., 2007; Gibbs and Brown, 2007; Baccini et al., 2008; Henry, 2010, Saatchi et al., 2011) or with global coverage (Gibbs, 2006; Reusch and Gibbs, 2008; Kindermann et al., 2008). However, the estimates provided by different datasets differ significantly for several tropical countries (Gibbs et al., 2007) and the resolution and accuracy provided by these products are not compatible with country-scale applications.

Therefore, there is an urgent need to develop methods capable to estimate reliably biomass distribution. Besides accuracy, biomass estimates need to be provided in a coherent, consistent and transparent way. Moreover, regarding scale factors biomass estimates need to be provided at:

- Spatial scales that are compatible with the areas affected by forest disturbance processes, i.e. at high spatial resolution
- Temporal scales that are compatible with the timing of the forest disturbance events, i.e. at frequent and repeated temporal intervals
- National scales, in order to be compatible with national management policies and REDD+ related carbon accounting procedures

3.2. Research questions and objectives

Within this framework and in order to contribute to the need of better spatially-explicit biomass estimates, **the overall goal of the present research was to explore, understand and quantify the capabilities and limitations of optical satellite data and ancillary information for producing detailed, accurate and spatially explicit AGB estimates at national scale in the tropics**, to be used in support of national planning and REDD-related activities.

Two main research objectives were identified:

- a) develop a methodology to retrieve aboveground biomass at country scale with medium spatial resolution (30 m)
- b) compare the novel methodology with alternative approaches and understand the main sources of variability in the estimates

According to the rationale presented above, the research was carried out with Landsat data for the case study of Uganda and was aimed at producing a new biomass map for Uganda. Uganda was selected as a case-study because of the availability of an extensive biomass field dataset collected by the National Biomass Study program (Drichi, 2003), necessary for model training and validation. The comparison of different approaches was performed by comparing a set of available biomass/carbon stock maps for the area of Uganda. This analysis allowed to better understand the reliability of existing biomass datasets and the advantages and disadvantages of the novel method with respect to existing approaches. The availability of a high-quality reference dataset produced by the National Authority of Uganda allowed drawing reliable conclusions from the comparison analysis.

The research questions underlying the first objective were the following:

- Are freely available Landsat data sufficient to compile a consistent national dataset needed to derive biomass density?
- Are, and to what extent, Landsat data sensitive to variation in biomass density of tropical vegetation?
- What is the most appropriate modeling approach to retrieve biomass from satellite data?
- Do, and to what extent, land cover data provide additional information useful for biomass estimation?
- What are the capabilities and limitations of this approach?

The research questions underlying the second objective were the following:

- What is the accuracy and spatial similarity of the existing biomass datasets?
- What are the causes of disagreement among the datasets?
- What are the effects of different input data on the biomass predictions?

3.3. Methodology

In accordance with the research objectives, the methodology employed in the present study consisted of two main components:

- a) Development of a novel methodology and its application to map aboveground biomass density of Uganda
- b) Comparison of existing biomass maps with a reference dataset

The operational steps employed for each of the two methodological components are reported below and are graphically summarized in Figure 4.

- a) Development of a novel methodology and its application to map aboveground biomass density of Uganda
 - Review of existing literature on remote sensing applications for biomass retrieval
 - Collection, pre-processing and analysis of field, remote sensing and ancillary data for Uganda. A field campaign was carried out to integrate the existing field dataset, and different allometric models to derive biomass from field parameters were compared
 - Comparison of different modeling approaches and input variables, and selection of the most appropriate model and predictor variables to map biomass density of Uganda. Model statistics were also used to assess the accuracy of the biomass predictions
 - Critical evaluation of the capabilities and limitations of the novel approach
- b) Comparison of existing biomass maps with a reference dataset
 - Selection, pre-processing and harmonization of a set of existing biomass/carbon maps for Uganda
 - Comparison of the biomass maps with a reference map and with selected field data to quantify the accuracy and spatial similarity of the biomass estimates.
 - Quantitative evaluation of the relative impact of the biomass reference data and spatial datasets on the biomass estimates
 - Critical evaluation of the sources of disagreement between the biomass datasets (i.e. input data, mapping approach)

It is important to notice that, due to the availability of field data for a wide range of vegetation types (e.g. woodland, shrubland, savanna), the present study was not limited to forest biomass but included all woody formations. This is especially important in the dry and semi-dry tropics where non-forest vegetation types store substantial amounts of biomass because their low biomass density is counterbalanced by coverage over large areas. Throughout the study, the terms biomass and AGB refer to live woody aboveground biomass of vegetation.

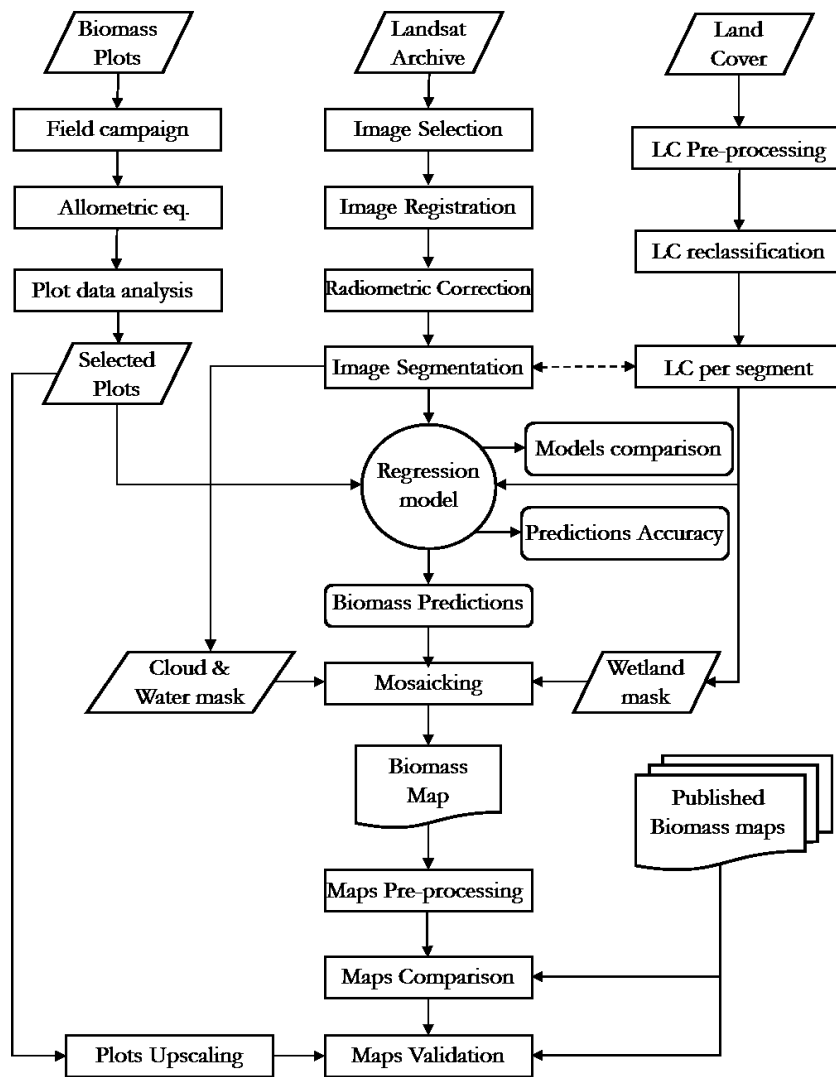


Figure 4: Flowchart representing methodological and processing steps

3.4. Outline of the thesis

The thesis is organized in four main sections and consists of six chapters.

The first section includes the introduction, the scientific background and the research objectives of this study. The introduction (Chapter 1) provides an overview of the topic, delineating the relevance of tropical forest biomass, the key role of sub-Saharan Africa in the global carbon cycle and the need of monitoring biomass in a highly dynamic country as Uganda. The scientific background (Chapter 2) summarizes the role of tropical biomass in the climate change context and reviews the relevant literature to provide an overview of existing approaches for biomass estimation, with focus on biomass retrieval through optical remote sensing and accuracy assessment of the biomass predictions. On the basis of this review, Chapter 3 (research objectives) describes the scientific needs and identifies the research gaps, which are used to delineate the research questions and objectives. The chapter provides also an overview of the methodology.

The second section (Chapter 4) addresses the first objective of this thesis, which is the retrieval of aboveground biomass of Uganda with Landsat and land cover data. This chapter describes the data and methods applied, the results obtained (e.g. the biomass map of Uganda), and provides a detailed discussion on the potential and limitations of the novel approach.

The third section (Chapter 5) addresses the second objective of this thesis, that is the comparison of existing biomass maps (including the map produced in the previous section) using a number of spatial statistic metrics to quantify the maps agreement with a national reference dataset. Additional analyses are then used to understand the sources of disagreement among the maps and to identify the optimal input data for biomass estimation from field and remote sensing data.

The fourth section (Chapter 6) reports the conclusions of this thesis, responds to the research questions raised in Chapter 3 and synthesizes the main findings of the two research topics, highlighting the innovative aspects of this study. This section also provides the recommendations for future research regarding biomass retrieval in the tropics at national scales using a combination of field and remote sensing data.

4. Retrieving a new biomass map of Uganda

4.1. Study area

4.1.1. Uganda and the National Biomass Study (NBS)

Uganda, located along the equator in East Africa at the eastern border of the Congo basin, is a country very diverse in terms of vegetation types, ranging from the dense humid forests in the south-west to the dry savanna and shrublands in the north-east.

Uganda lies mainly between 900 and 1500 m above sea level and presents a tropical climate with two rainy seasons in the South and one in the North. The total country area is equal to 241,551 Km², of which subsistence cropland is the most widespread land cover type (35%), followed by grassland (21%) and woodland (16%). Water bodies cover 15% of total area (Drichi, 2003). Vegetation is mainly represented by shrubland and grassland in the north-east (yearly precipitation 900 mm), woodlands in the north and west, and forest in the south and west (yearly precipitation 1500 mm) (Figure 5).

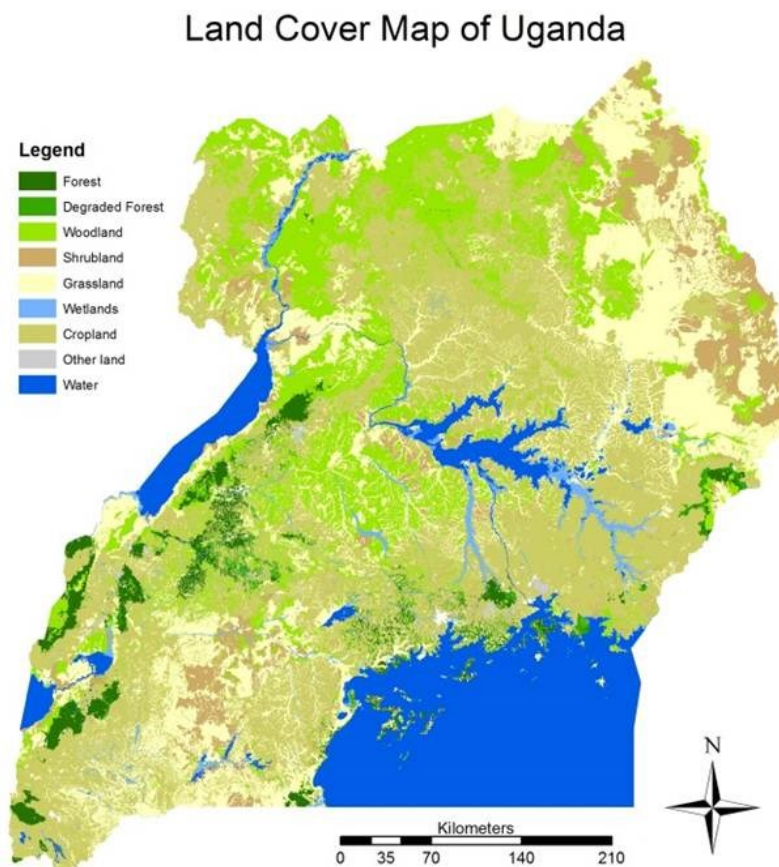


Figure 5: The NBS Land Cover map of Uganda. Only the main land cover classes are displayed

Uganda is one of the few African nations with a long term program aimed at the accurate assessment of biomass resources. In 1989, the government of Uganda established the National Biomass Study (NBS), a program aimed at the assessment of biomass resources

and their dynamics at national level using country-specific data and methodology. Using a nation-specific data and methodology, the NBS can be considered compliant with Tier 3 of the IPCC GPG Guidelines for estimating carbon stock changes on forest lands (IPCC, 2006).

The NBS of Uganda produced in 2003 a national biomass map based on the combination of an extensive biomass field campaign (over 5,000 plots), country-specific allometric equations and an ad-hoc land cover/use map (Drichi, 2003). The land cover map identified and stratified the 13 main land cover types by agro-ecological regions and biomass stock density classes, assessed through visual analysis of satellite images and extensive field validation. The biomass map was eventually produced by applying to each land cover stratum the average biomass density of the field plots located within the unit (Drichi, 2003). Given its high quality, the NBS represents an optimal reference dataset to better understand the capabilities of existing biomass or C maps and related methodologies for national level applications.

The country biomass stock is estimated at 468 Tg of air-dry AGB, mainly located in tropical forests (29%) and woodlands (27%) (Drichi, 2003). Areas classified as cropland and grassland also store relevant amount of biomass (24% and 10%, respectively) because of the abundance of scattered trees (Drichi, 2003; FAO, 2003).

Monitoring activities from national and international organizations reported that the forest area of Uganda and related biomass stocks have been severely reduced during the last three decades as a consequence of the large population and economic growth, which has fuelled the request for new agricultural areas and forest products, such as charcoal, timber and fuelwood (Drichi, 2003; FAO, 2006). With the country's growth being expected to continue in the coming years, the remaining forest resources of Uganda are in danger of further encroachment and degradation if proper forest practices and protection activities are not put in place in a timely manner (FAO, 2003).

4.1.2. Budongo Forest Reserve

Budongo Forest Reserve is located in the north-western part of Uganda, on the western edge of the Albertine rift valley, in the districts of Masindi and Hoima (geographic coordinates between 1° 35'-1° 55' N and 31° 18' – 31° 42' E) (Figure 6). Budongo Forest Reserve was gazetted in 1932 and is the largest tropical forest reserve of Uganda, covering an area of about 82,500 ha. Budongo is considered a reserve of high importance for global biodiversity, ranking third in overall importance of Ugandan forests (Howard et al., 1997), and contains the largest and most valuable timber forests of Uganda (Howard, 1991).

The climate is characterized by high temperatures, ranging from 23 to 32 degrees, and high rainfall, ranging from 1,400 to 1,500 mm annually and mainly concentrated in two rainy seasons occurring in April – May and October – November (Forest Department Uganda, 1997).

Located at medium altitude (1100 m mean elevation) in the transitional area between the rain forests of the Congo basin and the dry savannas of Eastern Africa, the vegetation of Budongo includes moist semi-deciduous high forest, woodlands and savanna grasslands. 53% of the Reserve is covered by continuous tropical forest (including early successional forests, mixed forest, *Cynometra* forest and swamp forest) while the remaining part is

dominated by grassland and woodlands communities (Hamilton, 1984). The tropical high forest, dominated by mahogany species, presents lower species diversity compared to the Congo basin forest and its structure has been altered by mechanical logging, pit-sawing and selected poisoning of trees with arboricides (Plumptre, 1996). Instead, the woodlands are heavily affected by fire, originated from natural (lightening) and human (land clearing, hunting) causes.

Due to past resettlements of people coming from other parts of the country and from Sudan and Congo, the local population consists on several agro-pastoral ethnic groups with heterogeneous culture, language and nationality, which main economic activity consists on agriculture (Langoya et al., 1998). The pressure of local communities on the forest (including hunting) has increased during the last years as a consequence of the population growth.

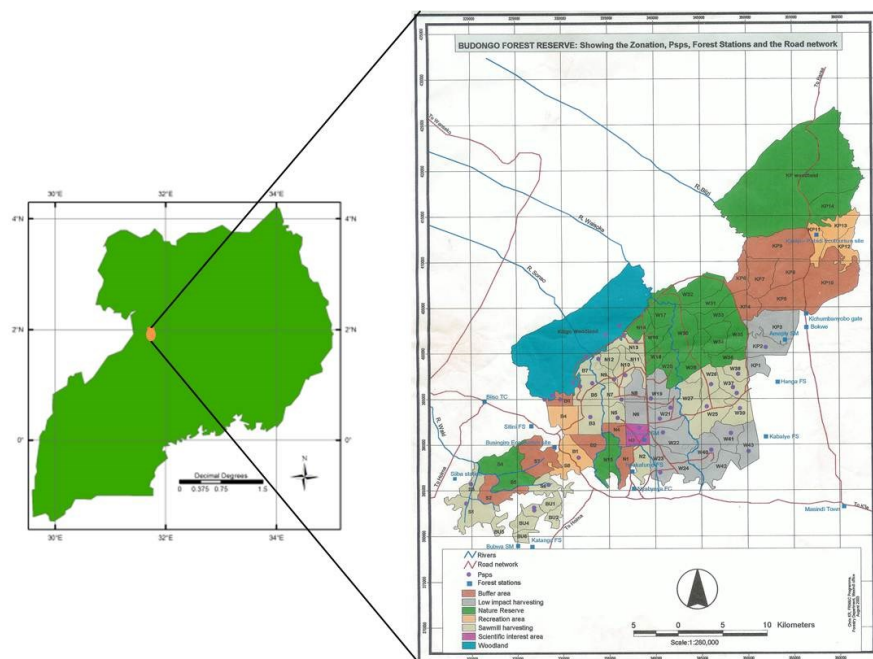


Figure 6: Location (left) and Zonation map (right) of Budongo Forest Reserve, Uganda (source: Forestry Department of Uganda)

4.2. Data and methods

4.2.1. Landsat data

The remote sensing dataset consisted a mosaic of 17 Landsat ETM+ images, format L1T, acquired in the period 1999 – 2003 during the dry season (December to March) when cloud coverage is at a minimum and spectral separability between trees and grass/shrub is at a maximum (Table 1, Figure 7, Figure 8). Landsat TM data were not employed because they were not available for the period of interest due to missing receiving station for central Africa (Goward et al., 2006).

The frequent cloud cover did not allow compiling a single phenologically consistent dataset for the whole country. Instead, the Landsat mosaic included 2 phenologically consistent sets of images, hereafter referred to as image “blocks”, which were used to develop two separate biomass models. The four images located in the South-Western part of the country form “Block 2” while all the other images form “Block 1” (Table 1).

Table 1: Landsat acquisition date and image block number

Path	Row	Date	Image Block
170	58	10-Jan-03	1
170	59	10-Jan-03	1
170	60	10-Jan-03	1
171	57	17-Jan-03	1
171	58	17-Jan-03	1
171	59	17-Jan-03	1
171	60	18-Feb-03	1
172	57	6-Feb-02	1
172	58	6-Feb-02	1
172	59	6-Feb-02	1
172	60	31-Dec-99	2
172	61	31-Dec-99	2
173	57	15-Jan-03	1
173	58	15-Jan-03	1
173	59	27-Mar-00	1
173	60	31-Mar-03	2
173	61	31-Mar-03	2

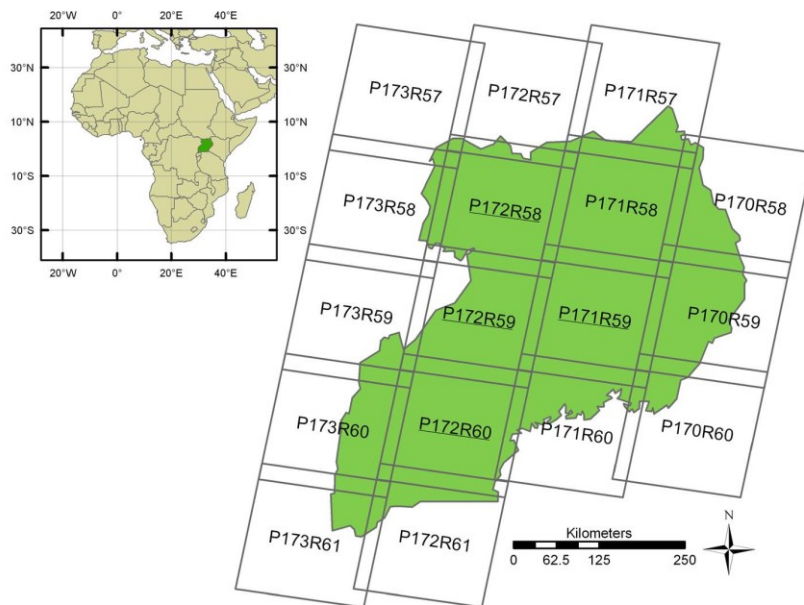


Figure 7: Landsat Path and Row for Uganda. The underlined scenes were used as reference for the radiometric normalization of the neighboring scenes

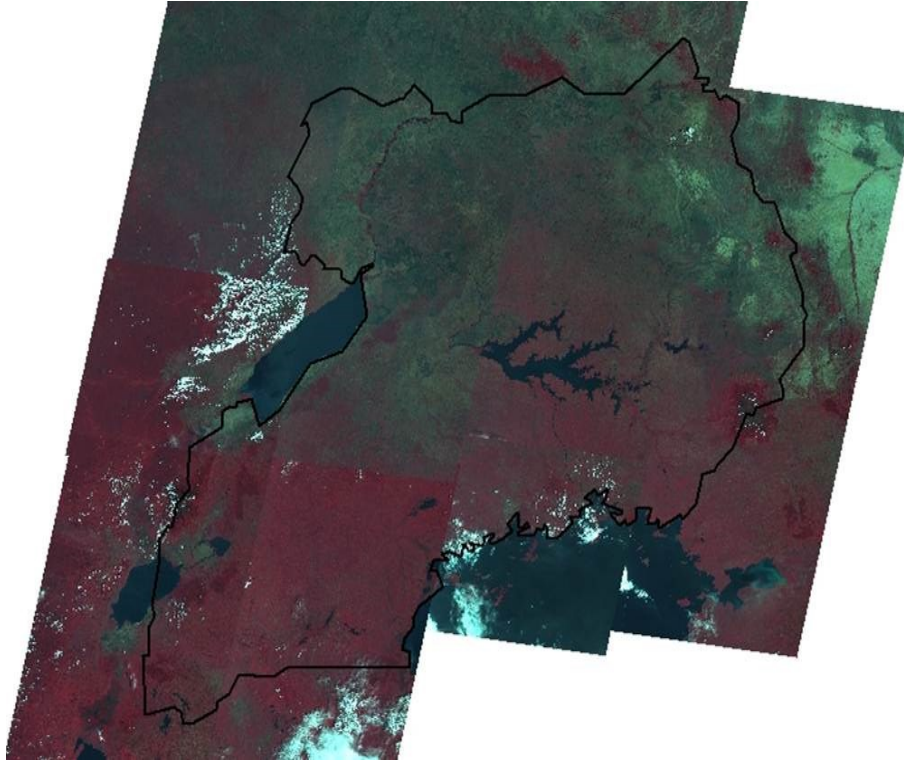


Figure 8: Landsat ETM+ mosaic for Uganda for the year circa-2000

4.2.1.1. Image pre-processing

4.2.1.1.1. Image registration

Landsat images were re-projected to a common reference system (UTM, WGS84, zone 36N) and then co-registered. The Landsat L1T correction process utilizes both ground control points (GCP) and digital elevation models (DEM) to attain improved geodetic accuracy and a geometrically rectified product free from distortions (NASA, 2009). Therefore, image registration was a minor task, necessary only to correct for systematic shifts occurring in a few cases between neighboring images.

4.2.1.1.2. Radiometric Correction

For several remote sensing applications the accurate estimation of surface reflectance is not necessary while the radiometric consistency among different images is crucial to obtain consistent results when the same regression model is applied to the images. In the absence of independent atmospheric data at the time and location of the image acquisition, simple absolute or relative radiometric correction methods (or their integration) are recommended over more sophisticated approaches based on atmospheric modeling (Song et al., 2001; Schroeder et al., 2006; Schowengerdt, 2007). In order to identify the most appropriate approach to calibrate the Landsat images of Uganda, the performance of an absolute and a relative correction method was first evaluated over a representative pair of neighboring Landsat images, and the best correction method was applied to the complete dataset.

The absolute radiometric correction was performed using an empirical, image-based Dark Object Subtraction (DOS) approach, where the digital numbers (DNs) are converted to surface reflectance on the basis of atmospheric parameters derived from the image itself (Chavez, 1996; Song et al., 2001). The DOS method quantifies the upwelling path radiance on the basis of dark objects (e.g. deep clear water bodies, shadow of low reflectance objects) present in the image, and uses constant atmospheric transmittance values. In particular, the lowest DN in each spectral band with at least one thousand pixels is assumed to have 1% reflectance.

The quality of the DOS output depends on the presence of appropriate dark objects, which may not always be present in each scene. For this reason, when more than one image in the Landsat mosaic was acquired on the same date under identical atmospheric conditions, the darkest objects present in any of these images were used.

The relative radiometric calibration was performed among images located in the same image block. This procedure, also known as radiometric normalization, assumes a band-specific linear relationship between the DN of the same objects across time (i.e. different images). The linear equations, which are computed on the basis of radiometrically stable objects (Pseudo-Invariant Features, PIFs), transform the images to be calibrated as if they were acquired under the same atmospheric conditions and sensor characteristics of the reference image (Hall et al., 1991; Schott et al., 1988).

The correction equations were computed by selecting stable PIFs for each spectral band located on the overlapping area of each pair of neighboring Landsat images. Since the equation coefficients may be affected by the nature of dark and bright objects available within the images, to obtain stable equations several PIFs covering the full range of brightness intensity (bright, medium and low) were first selected and then the ordinary least-square (OLS) method was used to identify a linear equation fitting the input data. The reference images were chosen on the basis of minimal path radiance (i.e. lower values for dark objects) and centralized position with respect to other scenes in order to minimize second-order calibration, which provides calibration of one image to a corrected image rather than to the reference image.

The performances of the two correction approaches (absolute and relative) were evaluated on the basis of changes of the Jeffrey-Matusita (JM) index (Richards, 1993) for some testing PIFs on a pair of Landsat images before and after the correction.

The JM index ranges from 0 to 2 and is inversely related to the multi-dimensional spectral similarity of the objects. In order to minimize changes due to seasonal (phenology) or permanent (land cover change) effects, two images acquired during the dry season, only 7 days apart, were selected (Path 170 Row 58 and Path 171 Row 58). 53 polygons were identified on the overlapping area of the two images and the JM index was computed for each pair of polygons before and after the radiometric correction. The results showed that the average value of the JM index for all polygons was 1.70 before the correction. This parameter increased to 1.99 after the DOS correction while it decreased to 0.94 after the relative calibration (DN matching).

The results suggested that the DOS reduced, instead of increasing, the spectral similarity between the two images, while the relative calibration successfully matched the images. Similar results have been observed in other studies (Schroder et al., 2006; Song et al., 2001). The poorer performance of the DOS method may be due to the fact that the original images were already spectrally similar and the errors introduced by the DOS assumptions were larger than the atmospheric effects. As mentioned above, the DOS method is

dependent on the type of dark objects present in the images and does not correct for the view path transmittance, which may be relevant for ETM bands 4, 5 and 7 (Schowengerdt, 2007). It was noted that, nonetheless the two images were acquired only 7 days apart, some testing polygons were already partially affected by phenological changes (e.g. start of greening), causing an increase of the standard deviation of the JM index after the relative calibration (from 0.27 to 0.47).

Based on these results, a relative normalization of the Landsat images was performed within each image block and the calibrated images were used as input to the regression models. Since the images located on the same Path were often acquired on the same date (see Table 1), the normalization was mainly performed between different Paths. The images located on Path 172 were used as reference and the images on Path 171 and 173 were normalized to the neighboring images on Path 172 while the images located on Path 170 were normalized to the neighboring images on Path 171 (Figure 7).

4.2.1.1.3. Image segmentation

This analysis employed an H-resolution approach (or object-oriented approach) because the study elements in the ground scene (vegetation units uniform for biomass density class) were usually larger than the image resolution cell (30 m for Landsat) (Strahler et al., 1986). The segmentation algorithm employed for this study is a multiple-pass region-growing method based on Euclidean distance in n-dimensional space and a minimum region size (Woodcock and Harward, 1992). Since the segmentation algorithm was mainly sensitive to input bands and object size parameters, different inputs were tested to identify the combination maximizing the segments homogeneity: Landsat bands (band 3, 4, 5 and 7), Tasseled Cap components (Kauth and Thomas, 1976; Crist and Cicone, 1984) and Texture channels. Texture, computed as the local minimum variance using an adaptive-window approach (Woodcock and Ryherd, 1989), describes the variability of the pixels within a certain window. It has shown capability to improve segmentation performance if the objects present textural differences (Ryherd and Woodcock, 1996). Texture was computed for Landsat bands 3, 4 and 5 with a window size ranging from 3-by-3 to 9-by-9 pixels.

The performance of the segmentation algorithm was evaluated for a test image (Path 172 Row 59) using the segment variance (σ^2_{segm}) as an indicator of the segment spectral homogeneity.

Areas affected by cloud and smoke were first masked because their extreme spectral values could significantly affect the computation of the mean segment variance. Second, the average variance of all m segments was computed for each Landsat band. Then, this value was standardized (i.e. divided) by the band variance (σ^2_{band}) to compensate for different DN dynamic ranges of each band. Finally, the standardized variances were averaged through all n bands to compute the mean standardized segment variance (MSSV).

$$MSSV = \sum_{i=1}^n \left(\frac{\sum_{i=1}^m \sigma^2_{segm}}{m_{segm}} \times \frac{1}{\sigma^2_{band}} \right) \times \frac{1}{n_{band}} \quad (4.1)$$

Several different combinations of segmentation inputs were tested on the Landsat image Path 172 Row 59 to identify the combination producing the most spectrally homogeneous segments, according to the MSSV index.

The results (Table 2) showed that the optimal combination of inputs was given by Landsat bands 3, 4, 5 and 7, while the Tasseled Cap components Brightness, Greeness, Wetness (BGW) minimized the standard deviation of the segment variance. In both cases, the addition of a texture band did not reduce the MSSV. Among possible texture bands tested on a representative subset of the image Path 172 Row 59, those computed from band 3 and using a larger window size (9-by-9) performed better than texture channels derived from bands 4 and 5 and smaller window sizes (Table 3).

On the basis of these results, the 17 Landsat images were segmented using bands 3, 4, 5 and 7, a maximum multi-dimensional spectral distance of 40 units, a maximum segment size for merging two adjacent regions of 500 pixels, and a minimum segment size of 20 pixels (i.e. 1.8 ha). The minimum segment size determined the Minimum Mapping Unit (MMU) of the biomass map.

The mean Landsat value for each band and the predominant NBS Land Cover class (according to a plurality rule) were computed for each segment and used as inputs to the biomass models.

Table 2: MSSV and its Standard Deviation (SD) for selected segmentation inputs on image p172r59. Inputs are coded as follow: “b” indicates Landsat bands, “Tex” indicates the Landsat band input for texture, “w” is the texture window size

INPUTS	MSSV	SD of MSSV
b3457	0.1133	0.1316
b3457Tex3w9	0.1142	0.1360
b347	0.1157	0.1306
BGW	0.1158	0.1279
b345	0.1163	0.1346
b347Tex3	0.1283	0.1484

Table 3: MSSV for different window sizes and texture bands, using input Landsat bands 3, 4, 5 and 7 on a representative subset of the image p172r59

Window size	MSSV	Texture band	MSSV
Tex3w9	0.2110	Tex3w9	0.2110
Tex3w7	0.2123	Tex4w9	0.2197
Tex3w5	0.2138	Tex5w9	0.2266
Tex3w3	0.2180	Tex7w9	0.2256

4.2.1.1.4. Water and clouds mask

Water bodies, clouds and their shadows were identified and masked before the biomass analysis. For this purpose, the images were first classified at pixel level using a Maximum Likelihood classifier trained on the basis of visually-identified training areas representative of each spectral class present in the Landsat mosaic. The classified images were then thematically aggregated into 4 classes (water, clouds, shadows and everything else) and spatially aggregated at segment level using a plurality rule (i.e. selecting the most frequent class within the segment).

The outcome was first manually edited to correct for minor classification errors and then used to mask water, clouds and shadows in the Landsat images. In the final biomass map, the biomass density of water was set to 0 Mg ha⁻¹ while clouds and shadows were set to “No Data”.

4.2.2. Land Cover data

Two land cover datasets were used for the present analysis: the Africover Multipurpose Land Cover database for Uganda (www.africover.org) and the NBS Land Cover map (Figure 5) (Drichi, 2003). Both datasets provide a detailed land cover/use classification of Uganda for the year circa-2000, with the Africover map having higher thematic resolution (67 classes) and the NBS map having higher spatial resolution (Minimum Mapping Unit ranging from 1 to 50 ha, according to the class).

The two datasets were pre-processed, which consisted in their re-projection to a common reference system (UTM, WGS84, Zone 36N), vector to raster conversion and legend reclassification.

The NBS Land Cover map, given its higher spatial resolution, was used to identify the predominant land cover class of each Landsat segment, and this information was then used as a predictor in the biomass models.

The Africover dataset, given its higher thematic detail, was used to identify the herbaceous wetlands, which were masked in the biomass models. Masking of herbaceous wetlands was necessary because preliminary analysis showed that this vegetation type received erroneous biomass estimates by the regression models as a consequence of its spectral similarity with tropical humid forests and the absence of specific training data. The wetlands were first identified by aggregating the Africover classes with dominance of herbaceous vegetation on permanently flooded land and then a constant biomass value, instead of the model predictions, was assigned to these areas. The woody biomass density of wetlands, mainly due to the presence of sparse trees, was computed as the average of the NBS biomass map within these areas and was set equal to 4 Mg ha⁻¹.

4.2.3. Field data

4.2.3.1. The NBS field plots

The NBS programme conducted an extensive field campaign aimed at measuring biomass density on a large number of samples representative of the biomass variability at national scale. The study employed a systematic sampling approach and identified about 5,000 plots throughout the country located on a systematic grid. In each plot, land cover characteristics and a number of plant parameters (e.g. species, diameter at breast height (DBH), height and crown width) were recorded for all trees with DBH \geq 3 cm. The sampling intensity was not uniform throughout the country but it was weighted by population density, with higher intensity in more densely populated areas. Plot size was 50 x 50 m for all land cover types except plantations, where it was 20 x 20 m. The plots were systematically located

throughout the country at 5 x 10 Km distance and, at each grid intersection, clusters of 1 to 5 plots were located at 300 m distance. The Northern part of the country was under-sampled because of political instability and ongoing conflicts, while Forest Reserves were often not sampled because of difficult accessibility.

Tree biomass was then derived from the plant parameters using country-specific allometric equations computed on the basis of destructive sampling of 3,477 trees (Drichi, 2003).

A subset of the complete field dataset was made available for this study. The subset comprised 3510 biomass plots measured between 1995 and 2005 well distributed throughout the country, with the exception of the northern region (Figure 9). The analysis of the frequency distribution of the plot values per biomass class revealed that the field biomass dataset was severely skewed towards low values, with 50% of the field plots having biomass density lower than 8 Mg ha⁻¹ and 90% of the plots having biomass density lower than 40 Mg ha⁻¹ (Figure 10). While dense high tropical forest of Uganda can reach biomass density over 300 Mg ha⁻¹ (Drichi, 2003), high-biomass forests were clearly under-represented in the NBS field dataset because mainly located within Forest Reserves, which were under-sampled during the NBS sampling scheme because of their difficult accessibility

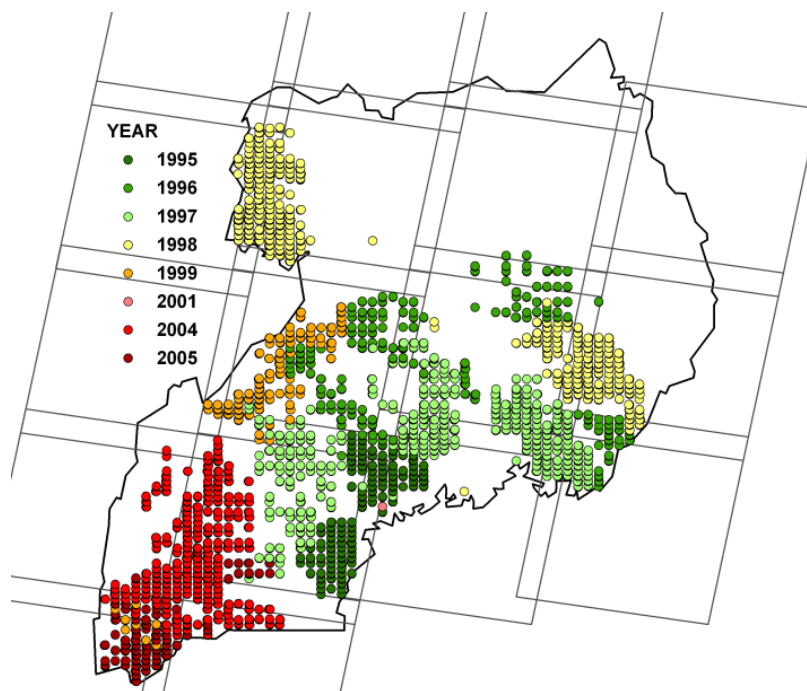


Figure 9: Location and year of acquisition of the NBS field plots, overimposed the Landsat scenes

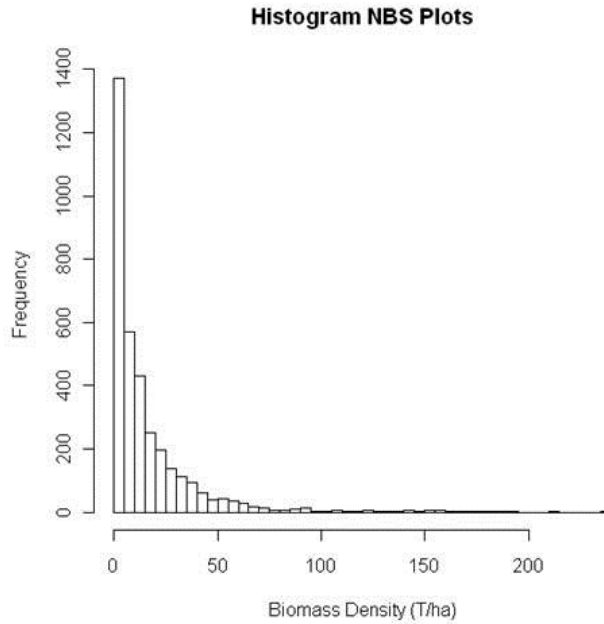


Figure 10: Histogram of the available subset of NBS field plots (N= 3510)

4.2.3.1.1. Selection of the NBS field plots

4.2.3.1.1.1. Selection procedure

The field data were provided in raw format, i.e. in form of plant parameters and related biomass for each tree, and plot coordinates were reported in different reference systems. Therefore, the plot data needed to be converted to plot air-dry biomass density and the plot coordinates were re-projected to a common reference system (UTM, WGS84, Zone 36N).

Considering the temporal difference between the plot measurements and the image acquisition dates (up to 8 years) and possible image or plot geolocation errors, it was necessary to screen the field data to select only the plots representative of the satellite data. The temporal mismatch was heterogeneous: while in image Block 1 the field plots were measured before the Landsat acquisition dates, in image Block 2 some plots were measured before and other plots were measured after the satellite acquisition dates. Theoretically, the field biomass values could be updated to the image acquisition dates using vegetation growth models. However, due to the lack of information necessary to parameterize such models with sufficient accuracy to dry tropical vegetation types and to the specific local conditions, it was considered that the use of growth models would introduce large uncertainty, and therefore they were not employed.

Most importantly, there were several other sources of temporal mismatch apart from the natural growth of vegetation, as changes due to fires or logging, as well as spatial mismatch due to geolocation errors. Such temporal and spatial mismatches required an accurate selection of the plot to be used in the biomass models.

The plot selection was performed through visual analysis because, due to the many sources of mismatch between plots and images to be taken into account, several criteria needed to be applied, and an automatic system could not perform such complex operation in an efficient way. For example, the plots not completely located within one segment but

overlapping even marginally two or more segments should (theoretically) be discarded. However, applying such criterion using an automatic system would reduce dramatically the number of available plots, since as many as 64% of the selected plots would be discarded, and would not consider possible geolocation errors, eventually discarding several plots that could be used while retaining erroneous plots. Instead, the visual comparison of the plot biomass density with the satellite data allowed to better determine the plots to be discarded. In other words, nonetheless the visual selection approach introduces a level of subjectivity, which is unavoidable when using “expert knowledge”, it allowed to deal in the best possible way with the several sources of mismatch and to minimize the number of discarded plots.

Specifically, after visual comparison with the segmented Landsat images, the plots associated with clouds or shadows, fire scars or heterogeneous image segments were removed.

Heterogeneous segments were created due to the minimum segment size (20 pixels) (see section 4.2.1.1.3). There were 2 types of heterogeneous segments: (A) segments including a vegetation type that is naturally heterogeneous, as savanna woodland or scattered agriculture; (B) segments including two ecologically distinct vegetation types, as a small patch of forest (smaller than 20 pixels) surrounded by agricultural areas. Only the plots located within type-B heterogeneous segments were discarded because in such cases the plots may not be representative of the overall segments and their biomass density may not be related to the average segment spectral values. Instead, the plots located in type-A segments were retained because, due to the large area of the biomass plots (50 x 50 m), the plot biomass values were expected to adequately represent the biomass density of the heterogeneous segments.

Plots not corresponding to the images because of co-registration mismatch (e.g. low-biomass plot located inside a forest but near its edge) or land cover change such as deforestation (e.g. high-biomass plot located outside a forest but near its edge) were also removed. However, if the mismatch between the plot and the image was due to seasonal variability of the vegetation (e.g. deciduous vegetation), the plot was retained.

It should be noted that the screening procedure could not remove errors related to field measurements and to use of allometric equations, which may be relevant and propagate through the analysis (Chave et al., 2004).

4.2.3.1.1.2. Selection results

After screening the 3,510 NBS field plots with the Landsat data according to the procedure presented above, 976 (27.8%) of the plots were discarded because associated with clouds or shadows (63 plots, 1.8%), fire scars (46 plots, 1.3%), heterogeneous image segments (211 plots, 6%) or because of co-registration mismatch and land cover change (656 plots, 18.7%).

The discarded plots were uniformly distributed throughout the country and among the Land Cover classes, with the number of discarded plots being between about 20% and 30% for most land cover classes (apart few exceptions due to the small number of plots in some land cover classes) (Table 4) and uniformly distributed compared to the original distribution of the plots, as it can be seen by comparing Figure 11 (Location of discarded field plots) with Figure 9 (Location of all available field plots). Cloud coverage in the Landsat mosaic was very small (<1% area) and was a minor cause for discarding plots

while most of the plots were discarded because of co-registration mismatch or land cover change occurred between the plot measurement and the image acquisition dates.

Table 4: Number of available, selected and discarded NBS plots for each NBS land cover class

LC class	All plots	Selected plots	Discarded plots	
	N	N	N	%
Plantation	1	0	1	100%
Forest closed	46	22	24	52%
Forest degraded	43	30	13	30%
Woodland	348	251	97	28%
Shrubland	138	109	29	21%
Grassland	613	474	139	23%
Wetland	5	2	3	60%
Subsistence Agriculture	2303	1638	665	29%
Industrial Agriculture	6	4	2	33%
Urban	7	4	3	43%
Total	3510	2534	976	28%

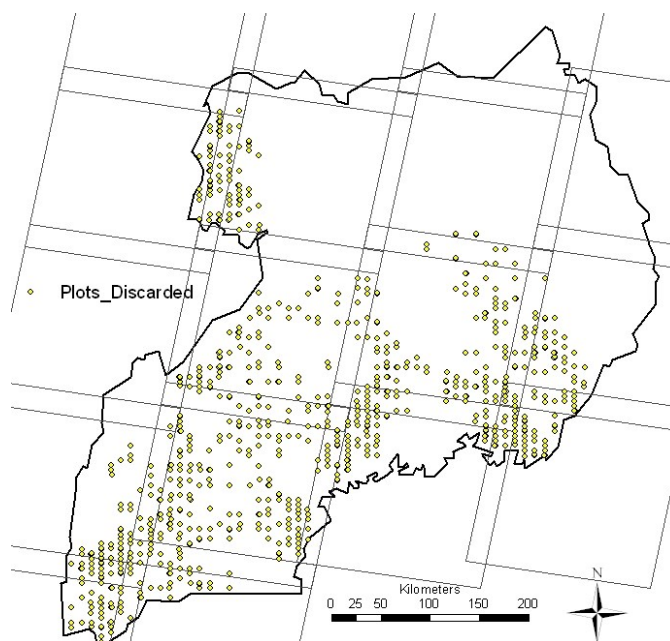


Figure 11: Location of discarded NBS field plots

4.2.3.2. Field campaign in the Budongo Forest Reserve

A field campaign aimed at measuring biomass density of vegetation was carried out in the Budongo Forest Reserve during November – December 2008 in collaboration with the Woods Hole Research Center (WHRC) (MA, USA). The Budongo dataset was then merged with the NBS dataset to obtain a field biomass reference database representative of the complete variability of biomass in Uganda.

The Budongo field campaign acquired data for 112 field plots located along a north-south transect crossing the entire Forest Reserve (Figure 12). The plots were located in pre-

defined locations and centered in correspondence of the center of the footprint of Lidar measurements acquired by the GLAS instrument on board the ICESAT satellite. The sampling scheme, designed by the WHRC according to their study on the relation between biomass density and Lidar metrics, allowed collecting data over a large vegetation gradient spanning from dense primary tropical forest to woodland and savannah ecosystems. In particular, the Budongo field campaign allowed the acquisition of a consistent number of biomass reference data in dense tropical forests, which were very poorly represented in the NBS dataset.

The field plots were squared with an extent of 40 x 40 m and area of 0.16 ha. Considering the environmental conditions of the Budongo Forest Reserve, and specifically the large number of species and the scarce visibility of forest canopy due to the multi-layered forest structure and dense understorey, the species and the height of the trees located within the plots were not recorded and only the DBH of all living trees with diameter above 10 cm was measured. Information on plot geolocation, land cover, land use, disturbance history, topography, soil and water regime were also acquired. In addition, the heights of the three tallest trees within the plot were recorded for the lidar analysis.

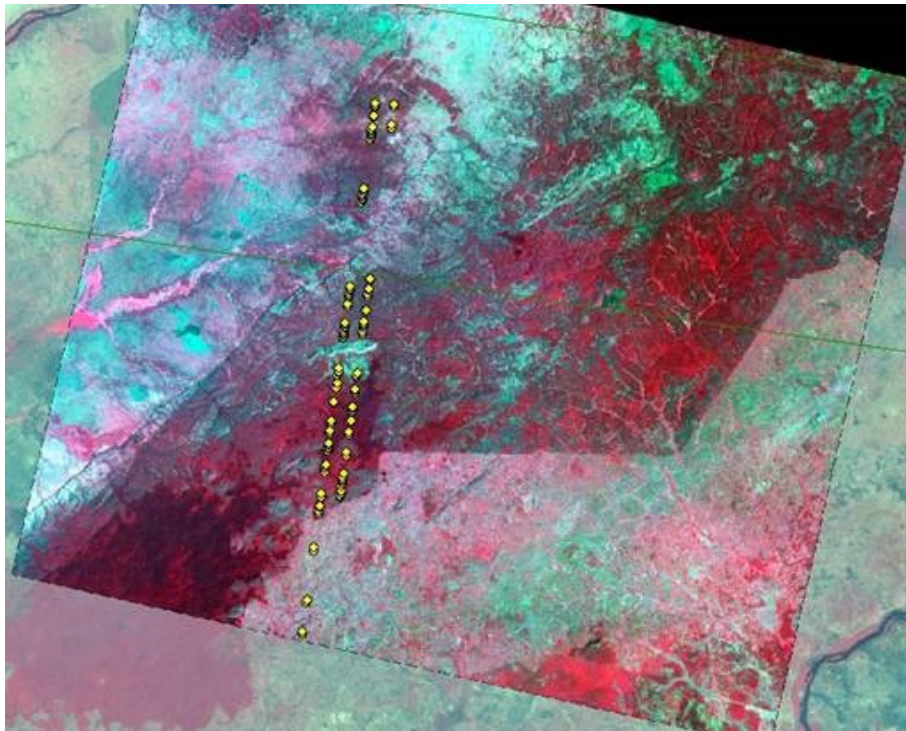


Figure 12: location of the field plots (yellow dots) measured in the Budongo Forest Reserve, overlaid on a SPOT5 image and a Landsat ETM+ image in the background

4.2.3.2.1. Selection of the allometric equation

The most important source of error in the estimates of tropical forest biomass is related to the choice of the allometric model (Chave et al., 2004). In order to select the most appropriate model for the Budongo dataset, four generalized allometric regressions for tropical forests were tested using the NBS dataset. Specifically, the allometric equations presented by Chave et al. (2001) for rainforest (4.2), by Chambers et al. (2001) for moist

forest (4.3) and by Chave et al. (2005) for dry forest (4.4) and for moist forest (4.5) based on DBH (d) and wood density (ρ) were applied to the NBS field dataset.

$$AGB_{\text{Chave2001}} = \exp(-2+2.42 \cdot \ln(d)) \quad (4.2)$$

$$AGB_{\text{Chambers}} = \exp(-0.37+0.333 \cdot \ln(d)+0.933 \cdot \ln(d)^2-0.122 \cdot \ln(d)^3) \quad (4.3)$$

$$AGB_{\text{Chave2005(dry)}} = \rho \times \exp(-0.667+1.784 \cdot \ln(d)+0.207 \cdot \ln(d)^2-0.0281 \cdot \ln(d)^3) \quad (4.4)$$

$$AGB_{\text{Chave2005(moist)}} = \rho \times \exp(-1.499+2.148 \cdot \ln(d)+0.207 \cdot \ln(d)^2-0.0281 \cdot \ln(d)^3) \quad (4.5)$$

The performance of the allometric models was assessed on the basis of the correlation of the biomass estimates with the NBS values. The NBS biomass values were considered as reference (i.e. most accurate estimates of the “real” tree biomass) because these values were computed on the basis of diameter at breast height (DBH), height (H), crown width (CR) using three size-dependent allometric equations specifically developed for Uganda, which accurately represent the allometric relations of the region.

The NBS is a very large dataset composed of 259,943 trees with a wide range of tree sizes and with tree biomass ranging from 0.1 to about 50 Mg. However, since small trees were much more frequent than medium and large trees (99.9% of the trees were below 10 Mg) and since the results were heavily affected by rare trees with very high biomass (i.e. outliers), the trees with biomass higher than 10 Mg were excluded from the analysis. In addition, it was noted that most of Budongo plots were located in forest or woodland while most of the NBS plots were located outside forest (i.e. cropland, grassland or shrubland), where the allometric relationships are different from those of forest species. For this reason, the analysis was also performed selecting only for the NBS plots located in forest and woodland, separately. To do this, the NBS plots located in areas classified as forest or woodland according to the NBS Land Cover map were selected. To take into account the minimum mapping unit of the NBS map, only the plots with biomass density higher than 65 and 13 Mg ha⁻¹ were further selected for the forest and woodland datasets, respectively. These biomass thresholds were identified on the basis of the consideration that the Budongo field plots located within forests and woodlands presented a minimum biomass density of 65 and 13 Mg ha⁻¹, respectively. This selection procedure identified 18,803 trees for the forest dataset and 51,889 trees for the woodland dataset.

The Chave et al. (2005) equations require wood density as input parameter, but this information was not available for the Budongo dataset. While in the tropics wood density varies considerably at species level (from 0.25 to 1.1), its mean value over several species, different vegetation types and geographical areas has lower variability. It has been showed that when information on wood density at species level is not available, reliable biomass estimates can be obtained using wood density values at genus-level (Chave et al., 2004) and, in their absence, constant average values can also be used (Brown, 1997). In the case of Budongo, a constant wood density value was applied.

In order to identify the most suitable wood density value for the Budongo dataset, the Chave et al. (2005) equation for dry forest was applied to the NBS dataset using different wood density values, and the value providing estimates more similar to the NBS was chosen. Values between 0.5 and 0.6 produced biomass estimates very similar to the NBS, with the optimal wood density being equal to 0.65 for forest areas, 0.54 for woodlands and 0.54 for all plots ($R^2 = 0.899$). The appropriateness of these values was confirmed by the fact that they were very similar to the mean wood density value for all trees in the NBS dataset (0.57) and to the mean wood density for commons species in tropical Africa (0.58)

reported by Brown et al. (1997). In addition, 11 tree species that were occasionally recorded in Budongo presented a mean wood density equal to 0.62. On the basis of this analysis, the value of 0.54, which provided the highest correlation with the NBS dataset, was used as input parameter for the allometric equations.

The four allometric equations presented above were applied to the NBS dataset and the biomass estimates were compared with the NBS values. The equation providing biomass estimates more similar to the NBS values was identified on the basis of the squared correlation coefficient (R^2) and the Root Mean Square Error (RMSE). The results (Table 5, Figure 13) showed that the pan-tropical equation from Chave et al. (2005) for dry forest provided biomass estimates more correlated to the NBS values than any other equation for the three cases, i.e. considering all plots, only the forest plots and only the woodland plots. In addition, the distribution of the residuals (considered here as the difference between the estimates of one generic equation and the reference NBS values) for the Chave (2005) equation for dry forest (Figure 14) showed that the predictions for large trees had lower accuracy and tended to slightly underestimate biomass estimates, but were not heavily biased because they were almost equally distributed between positive and negative values. The results (Table 5) also showed that the other generic equations performed better when selecting only plots in forest areas than with the complete dataset, confirming the different allometry for trees outside forest.

Table 5: results of the comparisons of the biomass estimates obtained with the NBS allometric equations and with four generic equations for all vegetation types (All plots), forest areas only (Forest plots) and woodlands only (Woodland plots)

Equation	All Plots		Forest Plots		Woodland Plots	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Chave (2001)	0.438	184	0.796	210	0.650	95
Chambers	0.240	214	0.681	262	0.150	147
Chave (2005) Dry	0.899	78	0.920	131	0.906	49
Chave (2005) Moist	-0.140	262	0.534	317	0.232	140

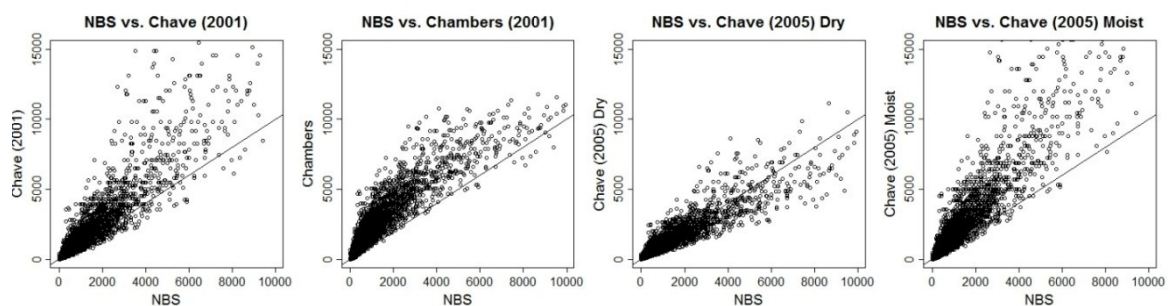


Figure 13: comparison of the biomass estimates obtained with the NBS allometric equations (x axis) and with four generic allometric equations (y axis), considering all plots. The biomass is computed for each tree and is reported in Kg. The black line is 1:1

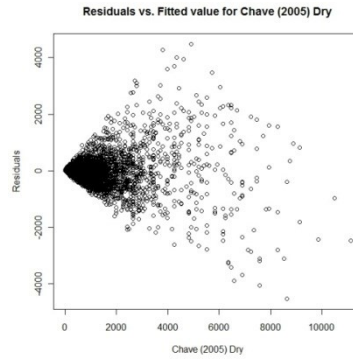


Figure 14: distribution of the residuals obtained comparing the NBS values with the Chave (2005) dry values, considering all plots. The tree biomass is reported in Kg.

The NBS dataset was also used to quantify the impact of the additional predictors. Specifically, the performances of the Chave et al. (2005) equations based on DBH, tree height (H) and wood density (ρ) for dry forest (4.6) and moist forest (4.7) were tested using the NBS dataset. Since the NBS provides the air-dry wood density while the above equations require the oven-dry wood density, the NBS air-dry coefficients were converted to oven-dry by multiplying them with the correction coefficient 0.872 (Chave et al., 2006).

$$AGB_{Chave2005(dry)} = \exp(-2.187 + 0.916 \cdot \ln(\rho \cdot DBH^2 \cdot H)) \quad (4.6)$$

$$AGB_{Chave2005(moist)} = \exp(-2.977 + \ln(\rho \cdot DBH^2 \cdot H)) \quad (4.7)$$

The results (Table 6, Figure 15) show that using three predictors (DBH, H and ρ) the two equations perform similarly and provide biomass estimates very similar (but systematically smaller) to the NBS values. Similar results were also achieved using two generic equations for tropical forests based on DBH, H and ρ (Brown et al., 1989). If tree height is not included as predictor, the Chave et al. (2005) equation for moist forest strongly overestimates tree biomass while the equation for dry forest provides more stable results. In addition, the equation for dry forest shows that decreasing the number of predictors increases the “spread” of the biomass estimates but reduces the systematic underestimation (bias) of biomass for large trees. Therefore, while tree height is a critical predictor for the Chave (2005) Moist equation and it is necessary to reliably apply this equation in Uganda, the Chave (2005) Dry equation provides reliable estimates also without this parameter.

Table 6: results of the comparison of the biomass estimates obtained with the NBS allometric equations and two generic equations using different predictor variables: Diameter at Breast Height (DBH), tree height (H) and wood density (ρ) (left column); DBH and ρ only (center column); DBH only (right column)

Equation	DBH, H, ρ		DBH, ρ		DBH	
	R ²	RMSE	R ²	RMSE	R ²	RMSE
Chave (2005) Dry	0.873	87	0.932	64	0.899	78
Chave (2005) Moist	0.905	75	0.001	245	-0.14	262

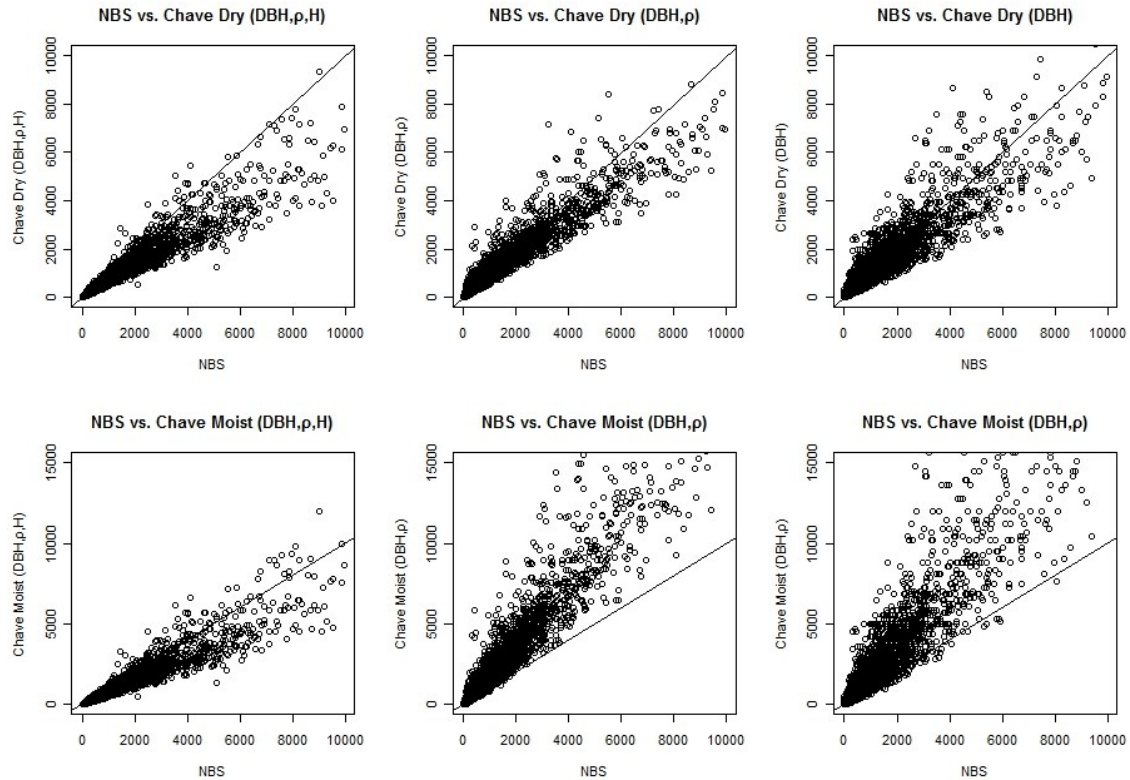


Figure 15: comparison of the biomass estimates obtained with the NBS allometric equations (x axis) and with two generic allometric equations (y axis), using different predictor variables: Diameter at Breast Height (DBH), tree height (H) and wood density (ρ) (left graphs); DBH and ρ only (center graphs); DBH only (right graphs). The biomass is computed for each tree and is reported in Kg. The black line is 1:1

4.2.3.2.2. Biomass field dataset in the Budongo Forest Reserve

The analysis of the allometric equations showed that the equation published by Chave et al. (2005) for dry tropical forest based on DBH and with a constant wood density of 0.54 provides biomass estimates comparable with those obtained from more sophisticated country-specific equations, as the NBS equations. Therefore, this allometric equation was used to convert the field measurements collected in Budongo to tree biomass. The plot biomass density was then computed dividing the sum of the biomass of each tree located within the same plot with the plot area.

Woody aboveground biomass density at plot level in the Budongo Forest Reserve ranged between 0 and 489 Mg ha⁻¹. The frequency distribution of the plot values per biomass class was bi-modal, with the two peaks corresponding to the biomass density of woodlands (< 140 Mg ha⁻¹) and of dense forest (> 220 Mg ha⁻¹) (Figure 16).

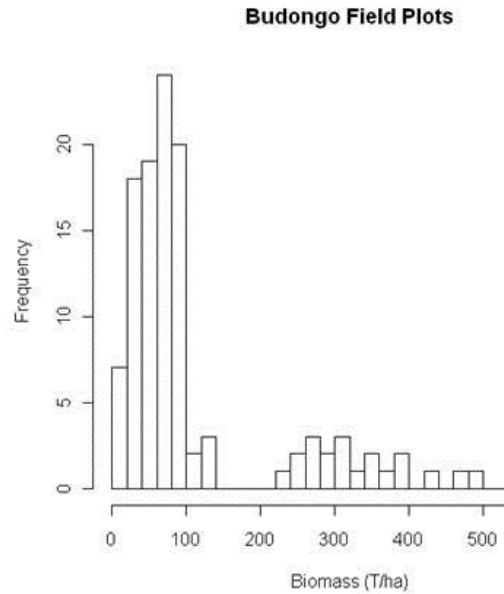


Figure 16: frequency distribution of the biomass density values of the Budongo field plots (N = 112)

4.2.3.3. The biomass field reference dataset for Uganda

The field plots selected by the NBS dataset (section 4.2.3.1) were combined with the plots collected in the Budongo Forest Reserve (section 4.2.3.2) to form the biomass reference dataset of this study.

Since the predictor variables (satellite and land cover data) were computed at the level of Landsat segments, when more than one field plot was located within the same Landsat segment their values were averaged and the mean biomass values per segment, instead of the original plot values, were input to the biomass models. On the basis of this procedure, the biomass reference dataset consisted of 2527 segments that were used to predict biomass density on the basis of satellite and ancillary data.

The analysis of the training data (Figure 17) revealed that their distribution was severely skewed toward smaller biomass values. The high frequency of low-biomass field plots was a direct consequence of using systematic sampling in a country dominated by croplands and grasslands. Since the distribution of the training data somehow reflected the spectral variability of different biomass classes (i.e. in Uganda low-biomass areas are much more spectrally diverse than medium and high-biomass areas), the skewed distribution allowed for better estimates at low values.

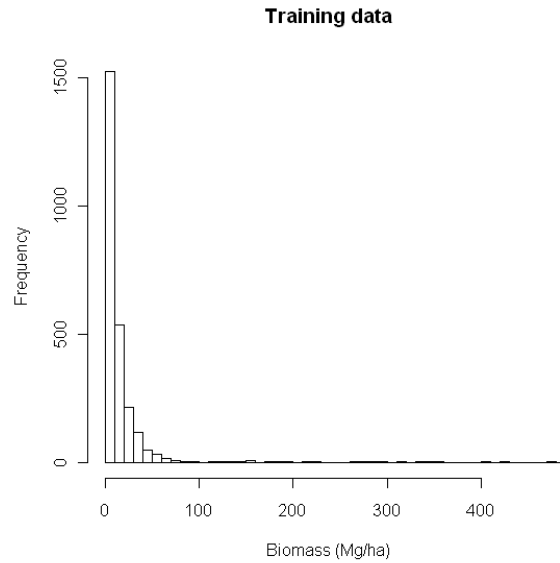


Figure 17: Frequency distribution of selected training data (N = 2527)

The biomass reference values were used as training data for two biomass models according to their location: the plots located within the image Block 1 were used to develop Model 1 and the plots located within image Block 2 were used to develop Model 2.

Given the scarcity of field plots in dense forests, all training data with biomass higher than 100 Mg ha^{-1} were used in both models. This was necessary to provide each model with enough training data to adequately represent high-biomass classes. The use of training data derived from images that were not radiometrically calibrated was possible because of the distinct spectral signature of dense forests in comparison with that of lower biomass classes: dense forests presented much lower spectral values in the Short Wave Infrared (SWIR) bands than other vegetation types, and small variations in their values among uncalibrated images did not affect the capability of the models to effectively separate them from other vegetation types.

In conclusion, the number of training data used as input to the two models was larger than the number of selected field plots for the following reasons: a) when more than one plot was located within the same Landsat segment, their average biomass density was used as an input value; b) the plots with biomass density higher than 100 Mg ha^{-1} were used by both models; c) in some cases the plots located on the overlapping area of two Landsat images acquired on different dates were used as training data for both images.

As a result of this procedure, 1627 training data were used to develop Model 1 and 1140 training data for Model 2. Considering the training data without duplicates, a total number of 2734 reference values were used to develop the two biomass models.

4.2.4. Modeling approach

On the basis of the literature review, a non-parametric modeling approach was considered the most appropriate method to estimate biomass density from remote sensing data. Specifically, this study employed Random Forest (Breiman, 2001), an extension of tree-

based models that uses the ensemble method to estimates large number of models using different bootstrap samples of the data (see section 2.6.2)

Random Forest applies a cross-validation approach to estimate model performance, using the predictions on the data not included in the bootstrap sample, called out-of-bag (OOB) data, to compute the model error rate. The OOB error rate is considered a reliable estimate of the model performance on independent data. To test this assumption, the prediction accuracy was assessed on randomly selected subsets (30%, 40% and 50%) of the original dataset held out before the model development. In addition, the effect of the skewed distribution of the input data (see Section 4.2.3.3) on the model RMSE was investigated by randomly selecting a uniformly distributed dataset. To account for some variability in the results obtained from a random selection of original data, model performance was calculated as the average of 50 runs on different random subsets.

Random Forest can be fine-tuned by setting the model parameters, such as the total number of trees to grow (*ntrees*), the number of variables randomly sampled at each split (*mtry*) and the minimum size of the terminal nodes (*nodesize*) (Liaw, 2009). The model sensitivity to different values of *ntrees* (from 100 to 2000), *mtry* (from 1 to 5) and *nodesize* (from 1 to 20) was evaluated by changing one parameter at each run. The analysis confirmed that the model performance was not very sensitive to the setting of these parameters (Liaw and Wiener, 2002) and that the optimal values for our training dataset (*ntrees* = 1000, *mtry* = 2, *nodesize* = 10) produced results very similar to those based on the default parameters computed by the algorithm (*ntrees* = 500, *mtry* = 2, *nodesize* = 5).

To evaluate the contribution of the predictor land cover to biomass estimation, the performances of two models for each “block” of the Landsat mosaic were computed and compared: the first model was based only on Landsat data (the Landsat spectral bands excluding the panchromatic and the thermal channels) and the second model was based on Landsat and land cover data.

Similarly, to evaluate the advantage of using an advanced regression tree model versus more traditional approaches, the performance of Random Forest (library *randomForest* in R) was computed and compared with that of a conventional tree-based model (library *rpart* in R) and a stepwise multivariate linear regression model (library *stats* in R).

Finally, the image Path 172 Row 59 was used to test if additional predictors derived from Landsat (texture bands) or topographic data (altitude, slope, aspect) improved the model performance.

The analyses were always performed at segment level by computing the average value of each variable within the Landsat segments and by relating this value to the average biomass density of the plots located within the training segments.

The percent variance explained (R^2) and the RMSE were used to evaluate the model performance on the complete datasets. In addition, the relative RMSE (RMSEr), the Bias and the Coefficient of Variation (CV) were also computed to assess and compare the model performance for different biomass classes. These statistics were calculated as follow, where \hat{y}_i was the predicted biomass on the i th plot, y_i the observed biomass on the i th plot, \bar{y} the mean biomass of n plots and $\hat{\bar{y}}$ the mean predicted biomass of n plots:

$$R2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (4.9)$$

$$RMSEr = RMSE/\bar{y}_i \quad (4.10)$$

$$Bias = \hat{\bar{y}} - \bar{y} \quad (4.11)$$

$$CV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \hat{\bar{y}})^2}{n}} \times \frac{1}{\hat{\bar{y}}} \quad (4.12)$$

4.3. Results

4.3.1. Models results

A biomass model was developed for each block of the Landsat mosaic (Model 1 and Model 2) and their performance was tested with and without land cover as a predictor.

The results (Table 7) indicated that the four models achieved very similar results, explaining between 83% and 86% of the variance of the OOB data and presenting a RMSE between 13.9 and 16.9 Mg ha⁻¹, depending on the model and predictors. Model 2, relative to the more humid part of the country, presented slightly larger errors at low biomass values because, due to higher rainfall, spectral reflectance of low-biomass vegetation types was more variable and hence more difficult to predict. On the other hand, model predictions at high values (≥ 100 Mg ha⁻¹) were almost identical since the models shared the same input data.

Statistics for the complete dataset (Model 1 + 2) were computed by merging predictions from the two models without replicates (i.e. considering input data ≥ 100 Mg ha⁻¹ only once) in one dataset (Table 7, Figure 18). The performance for the complete dataset was slightly lower than that of individual models, with R² between 78% and 81% and RMSE between 12.6 and 13.5 Mg ha⁻¹, depending on the predictors.

In all cases, the inclusion of land cover only slightly increased the overall model performance but improved biomass predictions, especially in deciduous formations where the model based only on Landsat data tended to underestimate biomass density (see Figure 26). For this reason, land cover was included as a predictor variable in the biomass models.

Regarding the variable importance, which was computed as the average increase of node purity (i.e. residual sum of squares) on the OOB data that results from including each

variable (Liaw, 2009), SWIR spectral bands and land cover provided the maximum contribution to the model predictions (Figure 19) because they allowed effective separation between high and low biomass data. The important role of SWIR wavelengths in biomass prediction was consistent with previous studies (Steininger, 2000; Lu et al., 2004; Baccini et al., 2004; Baccini et al., 2008).

Table 7: Performance results on the OOB data of the Random Forest models with and without the predictor land cover (LC). Model 1 + 2 includes predictions from the two models without replicates. N is the number of input data

	N	Landsat + LC		Landsat	
		R ²	RMSE	R ²	RMSE
Model 1	1627	0.847	13.9	0.826	14.8
Model 2	1140	0.864	15.2	0.834	16.9
Model 1 + 2	2734	0.811	12.6	0.781	13.5

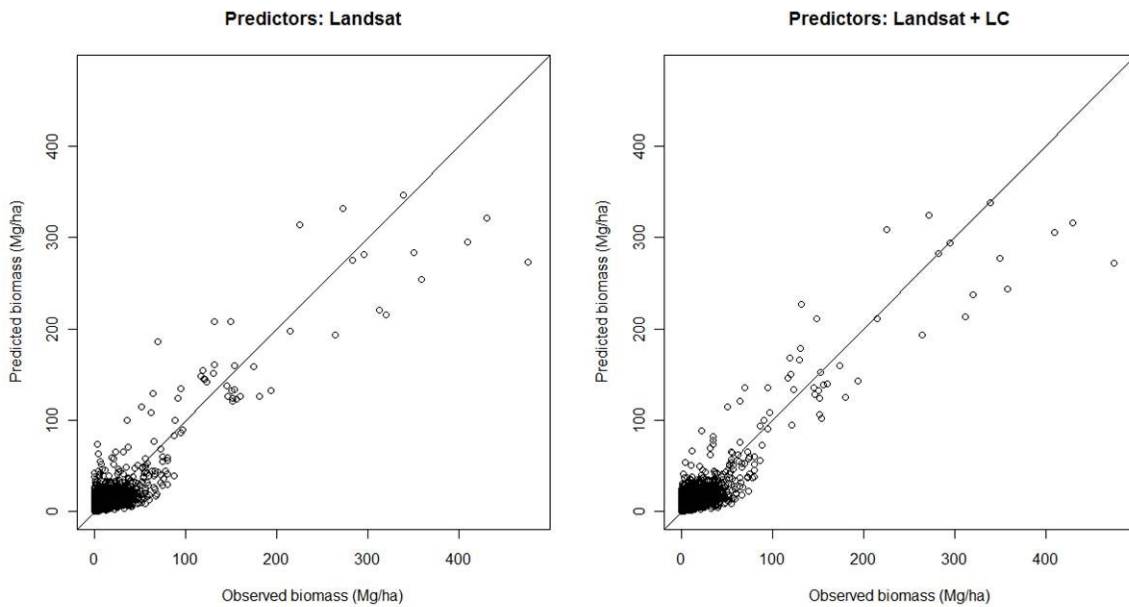


Figure 18: Predictions of all input data (Model 1 + 2) on OOB data without (left) and with (right) the predictor land cover (LC)

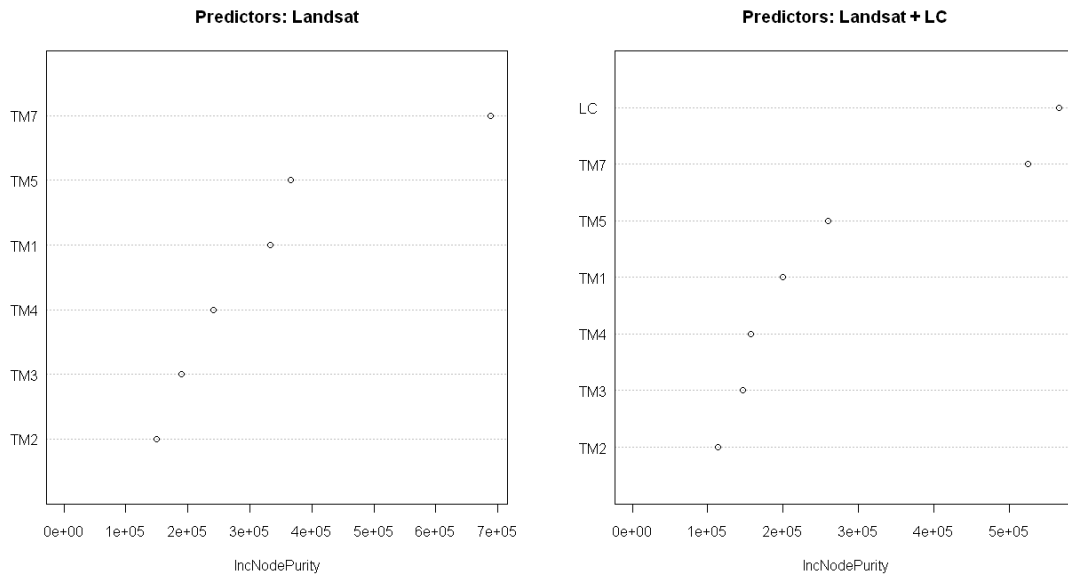


Figure 19: Variable Importance plots for Model 1 without (left) and with (right) the predictor land cover (LC). IncNodePurity represents the variable importance. The Variable Importance plot for Model 2 (not reported) showed the same ranking

The percentage variance explained by the models (R^2) was influenced by the range of the response variable, and the values reported above were positively affected by the large range of biomass values present in the training data.

Considering the result from both models (Model 1 + 2), the R^2 was 0.42 for biomass values ranging from 0 to 100 Mg ha⁻¹ and 0.59 for values higher than 100 Mg ha⁻¹. Model performance dropped severely for biomass values ≥ 175 Mg ha⁻¹. This result likely reflects the small sample size ($N = 16$) but it may also be caused by the unequal distribution of the training data, which affected the construction of the tree-based model creating a larger number of nodes for low biomass values compared to those for medium and high values.

Additional model statistics computed for each biomass class (Table 8, Table 9) showed that the absolute errors (RMSE) increased for higher biomass classes, but the relative errors (RMSEr) and the normalized dispersion of the predictions (CV) decreased. While the overall bias was very small, the model tended to underestimate biomass at high values.

Considering that model residuals in the dataset tended to increase at higher biomass values, the predominance of low-biomass training data biased the estimate of the model RMSE, which was clearly over-optimistic for medium and high biomass predictions. By selecting a subset of training data with uniform frequency for each biomass class ($N = 44$ plots), the model RMSE increased to 63.7 Mg ha⁻¹ (relative RMSE 36%), which is about an average value of this parameter for the full range of values.

Table 8: Performance statistics for all data (Model 1 + 2) for biomass classes. Due to scarcity of high biomass values, all data > 200 Mg ha⁻¹ were grouped in one class

	Biomass range (Mg ha ⁻¹)			
	0 - 475	0 - 100	100 - 200	200 - 475
N	2734	2699	21	14
RMSE	12.6	10.2	41.5	88.8
RMSEr	91%	92%	29%	27%
Bias	0.4	0.6	1.6	-47.3
CV	176%	91%	23%	16%

Table 9: Performance statistics for all data (Model 1 + 2) for biomass classes in the range 0 – 100 Mg ha-1

	Biomass range (Mg ha ⁻¹)			
	0 - 25	25 - 50	50 - 75	75 - 100
N	2388	249	50	12
RMSE	7.6	18.8	30.9	27.2
RMSEr	104%	55%	51%	32%
Bias	2.6	-14.0	-17.6	-11.3
CV	67%	57%	61%	42%

In order to assess the reliability of the OOB error rate estimate, the performance of Model 1 was also tested against independent datasets excluded from model development by randomly selecting 30% to 50% of original data. The results (Table 10) showed that the explained variance remained above 80% and was only slightly smaller to that of the OOB data.

Finally, the capabilities of additional predictors derived from Landsat or topographic data to improve model performance were tested on the test-image Path 172 Row 59. Specifically, the mean values for each Landsat segment were computed for the metrics Altitude, Slope, Aspect and Texture. Texture was derived from Landsat bands 3, 4, 5, 7 and computed as standard deviation of pixel values in a 3x3 adaptive window. The standard deviation of texture values within each segment as well as spectral values from band 6 (thermal band) were included as explanatory variables.

The performance of Random Forest, computed as an average of 50 model runs, remained stable ($R^2 = 0.80$, $RMSE = 15 \text{ Mg ha}^{-1}$) after the inclusion of the additional predictors. This result suggested that even though some of these newly added variables were modestly important (Figure 20), they did not contain any additional information for biomass estimation compared to Landsat and land cover, and therefore were not employed further.

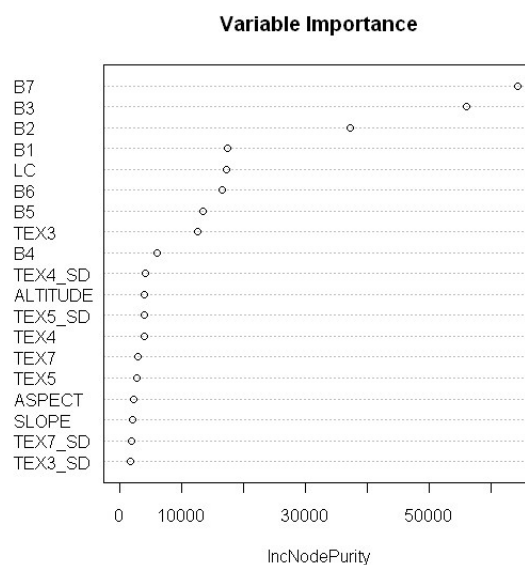


Figure 20: Variable Importance Plot for a model including the additional predictors for the image p172r59. The number after “TEX” indicates the Landsat band used to compute the Texture, the suffix “_SD” indicates the standard deviation of the texture band

4.3.2. Model comparison

The comparison of different modeling approaches showed that Random Forest performed better than a single regression tree: the variance explained by a single tree model on the independent dataset was between 14% and 17% lower than that of Random Forest (Table 10). Another important difference between the two models was that the biomass predictions of a single regression tree were discrete, similar to those of a classification algorithm, while the predictions from Random Forest were continuous (see Figure 21).

It is important to notice that, since the single regression trees are pruned to avoid overfitting, the number of final nodes was considerably smaller than that of Random Forest: for image Block 1 the single regression tree identified only nine biomass prediction values, whereas Random Forest provided almost continuous estimates. However, due to the predominance of low biomass data in our dataset (85% of the data were below 25 Mg ha⁻¹), Random Forest did not achieve a lower RMSE than a single tree model (Table 10), showing that for very small biomass ranges the accuracy obtained with an appropriate number of discrete predictions (RPART identified 5 biomass classes smaller than 25 Mg ha⁻¹) is comparable with that of continuous estimates.

Table 10: Performance of Random Forest (RF) and a single Regression Tree (RPART) for different training and testing datasets from image Block 1. Predictors: Landsat bands (without LC)

Model Dataset	Training			Testing				
	N	RF R ²	RF RMSE	N	RF R ²	RF RMSE	RPART R ²	RPART RMSE
30% Test	1139	0.824	14.6	488	0.826	14.7	0.709	10.0
40% Test	976	0.825	15.1	651	0.812	14.5	0.683	12.4
50% Test	814	0.809	15.5	813	0.813	15.1	0.672	14.6

Linear models did not achieve satisfactory results when applied to this dataset. For example, a stepwise multivariate linear regression using the Landsat bands as predictors explained only 23% of the variance of the data in image Block 1 (RMSE 31 Mg ha⁻¹, p value < 0.0001). However, such poor performance was biased downward by the skewed distribution of the training data: the variance explained by the linear model increased to 65% (RMSE 71.8 Mg ha⁻¹, p value < 0.0001) using training data with equal frequency for each biomass class (N = 44 plots) (Table 11), and was 40% when the model was applied to 20% of independent data randomly sampled from the uniform training dataset.

The dataset with uniform distribution for biomass class was used to compare the prediction capabilities of the three model types (Random Forest, single regression tree, stepwise multiple linear regression) using the training data, instead than on the OOB data (Table 11, Figure 21).

Since regression trees do not perform as well as linear models in predicting strongly linear relationships (De'ath and Fabricius, 2000), the lower performance of linear models supported the assumption of a complex, non-linear relationship between biomass and remotely sensed data (Foody et al., 2003; Baccini et al., 2004; Muukkonen and Heiskanen, 2005).

Table 11: Model performance on the training data (not on the OOB data) of the uniformly distributed dataset for Random Forest (RF), single Regression Tree (RPART), stepwise multiple linear regression (LM). Predictors: Landsat bands (without LC)

Dataset	Model	N	R ²	RMSE
Uniform distribution	RF	44	0.943	29.1
	RPART	44	0.803	53.8
	LM	44	0.651	71.8

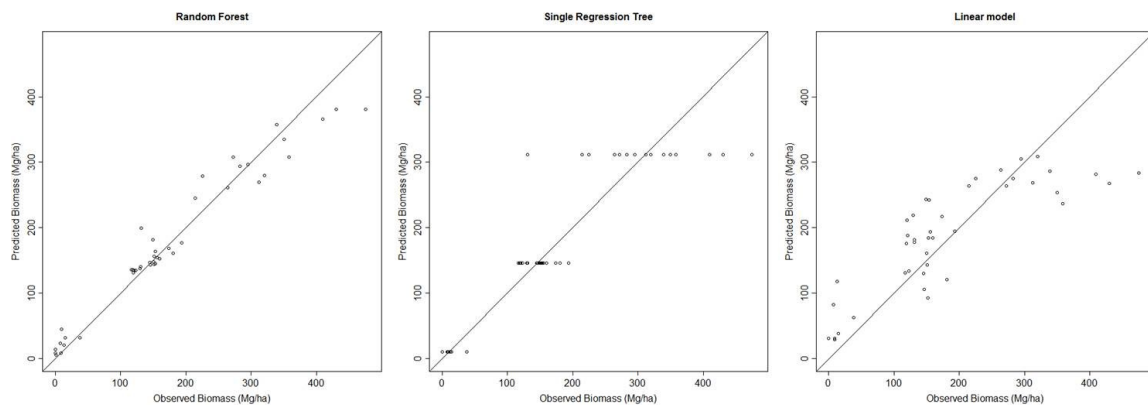


Figure 21: Model predictions for training data with uniform distribution per biomass class of (left to right): Random Forest, single regression tree, stepwise multiple linear regression. The black line is 1:1

4.3.3. A new biomass map for Uganda

Landsat and land cover data were used to predict AGB for each image segment using Random Forest. The biomass predictions were then mosaicked to create a map that represents the spatial distribution of live aboveground woody biomass density of Uganda for the year circa-2000 at Landsat resolution (Figure 23).

According to this analysis, the aboveground biomass stock of Uganda was equal to 343 Tg. The biomass density ranged from 0 to 383 Mg ha⁻¹ at the resolution of the Landsat segments; 1% of the total area (246,000 ha) was covered by clouds and shadows, and classified as “No Data”.

The frequency distribution of predicted biomass was similar to that of the training data (Figure 22). The map indicated that most of the country presents low biomass density, but the contribution of these areas to the total AGB stock was important. Areas with biomass density lower than 25 Mg ha⁻¹ cover about 87% of land area and store about half (53%) of total AGB stock. The remaining AGB (47%) is stored in medium and high biomass density areas covering less than 13% of land surface. Areas with medium biomass density (25 – 100 Mg ha⁻¹) cover about 10% of land area and store 25% of total stock, while areas with high biomass density (> 100 Mg ha⁻¹) cover less than 3% of land area and store 22% of total stock. Low and medium biomass areas correspond to cropland, grassland, shrubland and woodland, while high biomass areas correspond to dense forests, mainly located within national parks and protected areas along the western border of the country.

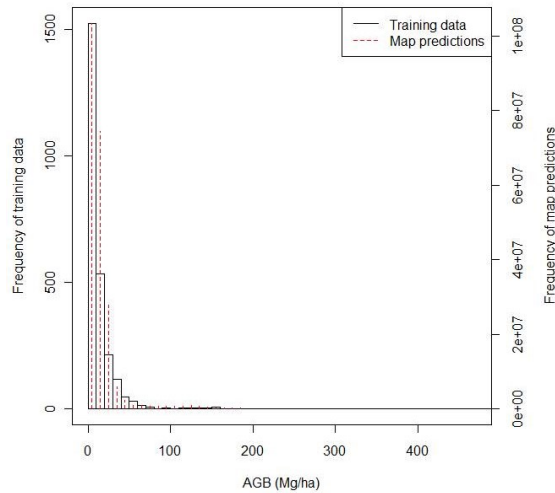


Figure 22: Frequency distribution of the selected training data (N = 2527) and of the map predictions

This estimate of the total stock of Uganda was 27% lower than the NBS value (468 Tg), suggesting that the presented approach provided conservative biomass predictions (see Section 4.4). The FAO (2006) indicated that the NBS figures are relative to 1992 and reported a decrease of AGB in forests of 50 Tg (from 294 to 244 Tg) between 1992 and 2000. The NBS Report (Drichi, 2003) also indicated a decrease in biomass stock in areas outside forest during the second half of 1990's, but did not provide updated figures for the year 2000. Under the conservative assumption that biomass outside forest remained constant, the reference value of total AGB stock of Uganda for the year 2000 would be 418 Tg, reducing the difference with the estimate of this study to 18%.

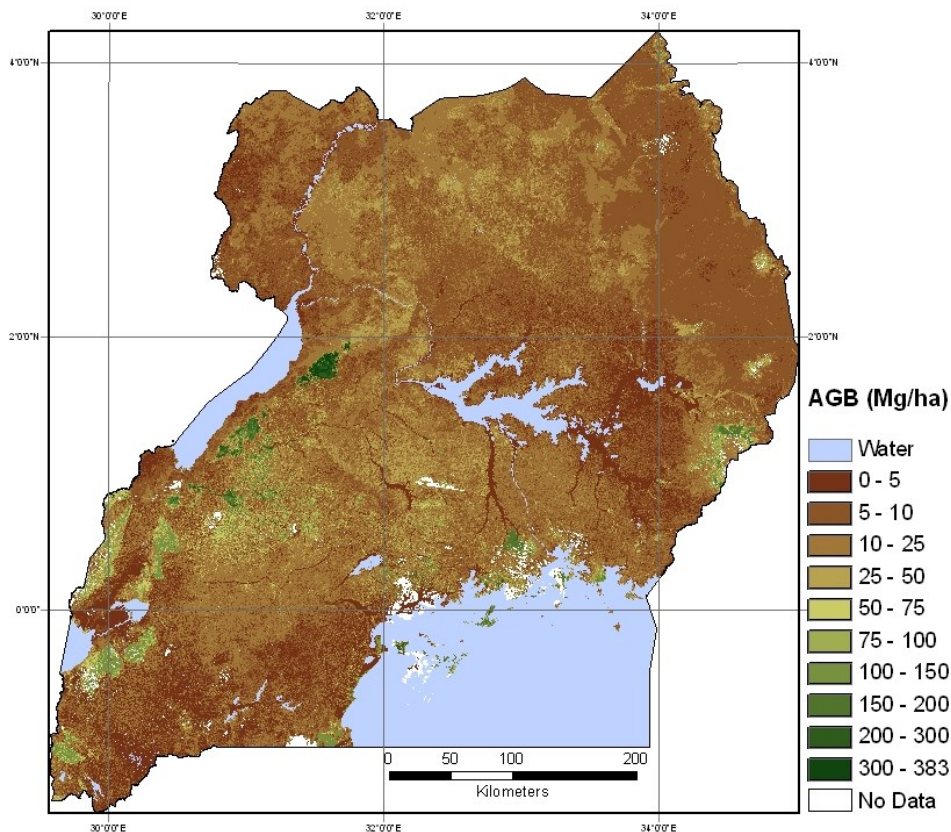


Figure 23: AGB density map of Uganda for the year circa-2000

4.4. Discussion

The results of this study suggest that it is possible to produce spatially explicit biomass estimates at high resolution over large areas if adequate field data, satellite images and ancillary data are available. In Uganda, Landsat and land cover data were used to spatialize field biomass measurements and produce accurate (R^2 0.81) and detailed (MMU 1.8 ha) estimates at national scale that may be used for national planning or REDD-related monitoring of forest biomass. Figure 25 shows the unprecedented spatial resolution attained with this approach in comparison with other published biomass maps.

Using a non-parametric regression model was crucial to obtain these results. Random Forest was able to capture the complex, non-linear relationship between spectral data and AGB, but required training data representative of the spectral variability within each main biomass class.

Training data must be screened because the model predictions, computed as mean values within each terminal node, are sensitive to outliers.

Accurate selection of the satellite data and image pre-processing were equally important to remove noise from the regression. For example, selection of images acquired during the dry season helped to avoid confusion between evergreen woody and deciduous non-woody vegetation.

Thanks to the large number of low-biomass field plots, it was possible to observe that the satellite signal varies largely in low-biomass vegetation types as savannah, subsistence agriculture and agroforestry. Therefore, it is crucial to collect training data for each spectral class to avoid biomass overestimation in these areas, as it was the case for wetlands where the high intensity of greenness confounded the model.

Given the empirical nature of the model, it was not necessary to atmospherically correct and convert the remote sensing data to surface reflectance values, but it was necessary to radiometrically calibrate the satellite images to a common scale. While transferability over time and space of a model calibrated on a specific training dataset is limited (Foody et al., 2003), the approach, rather than the algorithm, can be easily replicated when sufficient data are available.

Landsat observations provided an effective way to spatialize field biomass measurements because the data acquired by this satellite are sensitive to vegetation structure and their spatial resolution is compatible with stand and plot size.

It is well known that optical sensors are more sensitive to leaf characteristics than to the woody component of vegetation. As a result, optical data are more correlated to vegetation density than to its vertical structure, which leads to their saturation above certain biomass values. However, SWIR bands have sensitivity to canopy water content and shadow fraction, which tend to increase moving from open and monolayer to closed multi-layered forest canopies (Gemmell, 1995; Steininger, 2000). If, as was the case in our study area, these canopy parameters (water content and shadow fraction) at the resolution of image segments are associated with biomass density, a non-parametric model that exploits spectral variations of SWIR bands may also have satisfactory prediction capabilities in closed canopy forests, up to a certain biomass value. In addition, image segmentation had a positive effect on the model accuracy at extreme (very high and very low) values because reducing spatial resolution also reduced the variability of the parameter to estimate.

In the present study, the models tended to saturate around 150 – 200 Mg ha⁻¹. However, the number of high-biomass training data was too small to clearly identify the saturation threshold.

Cloud cover severely limits the amount of satellite data available in the tropics. Cloud cover in the Landsat mosaic was less than 1% of land area, but cloud contamination can be expected to be a bigger issue for larger tropical regions located in areas with persistent cloud cover (e.g. Congo basin, South East Asia). Thanks to the opening of the Landsat archive and improved image processing techniques, cloud free and radiometrically consistent Landsat mosaic have been produced for large humid tropical areas, as in the Congo basin (Hansen et al., 2008), but the creation of consistent multi-temporal datasets remains challenging.

For the case of Uganda, the Landsat mosaic did not contain multi-temporal information because cloud coverage hindered the compilation of a consistent multi-temporal dataset, which may be important for discriminating some vegetation types (e.g. seasonal or deciduous formations).

It is likely that the use of land cover as a predictor increased the model performance because it provided information related to vegetation phenology missing in the Landsat mosaic. For example, the biomass map derived only from Landsat data (acquired in the dry season) tended to underestimate the biomass content of dry deciduous shrubland and woodland compared to the NBS values, while the inclusion of land cover provided more accurate estimates for these formations (Figure 26).

On the other hand, errors and limitations of land cover data were propagated in the biomass map. For example, since in Uganda deforestation is mainly driven by small-scale local activities, it affects areas that are often smaller than the minimum mapping unit of the land cover map. As a result, the inclusion of land cover in the model caused biomass overestimation for small or partially deforested areas (Figure 27).

The availability of land cover maps suitable for biomass mapping varies from country to country and may be a limitation in the application of this approach to other countries. However, besides the fact that national high resolution datasets are now being produced more frequently in tropical countries also thanks to coordinated international efforts, such as the FAO Global Land Cover Network that directly sustains a number of national and regional mapping activities, for the case of Uganda the model results obtained without land cover data were overall statistically similar to those including land cover data (Table 7). Hence, while land cover data improve model predictions (Figure 26), the presented approach is not strictly dependent on their availability.

It is important to stress that Landsat data present higher spatial and thematic resolution compared to land cover data (Figure 24). The use of Landsat data allowed the computation of spatially detailed and continuous biomass estimates. Conversely, a biomass map based only on land cover data would assign the same average value to all spatial units belonging to the same class, reducing the map resolution and missing substantial spatial variability within the land cover polygons.

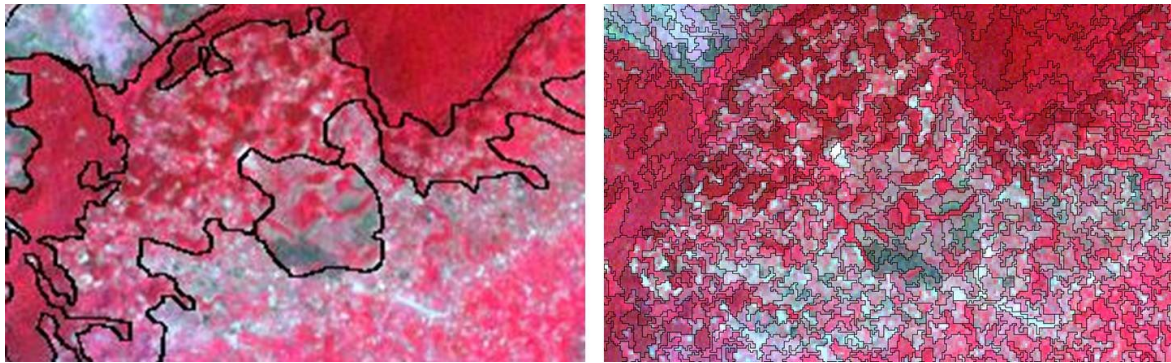


Figure 24: Comparison between land cover polygons (left) and image segments (right) for an heterogeneous area in Western Uganda. The image segments are able to identify vegetation units more homogeneous and with higher spatial resolution compared to land cover units

In addition, using average biomass values for broad land cover classes can cause serious underestimation of emissions from deforestation when it occurs in forests with higher than average biomass density (Houghton, 2005; Houghton and Hackler, 2006). On the contrary, using average biomass values derived from traditional forest inventories, which are usually aimed at assessing the merchantable timber volume, may overestimate the total country stock because the field data are mainly collected in the most productive forests, while disturbed low-biomass forests are under-sampled as a consequence of their low economic value and the difficult accessibility due to the regrowth of dense understory developing after the disturbance. In such cases, the average biomass values derived from the inventories do not represent the overall class variability but are biased towards the high-biomass forests.

Considering the fact that model performance decreased when applied to a smaller biomass range (Table 8) and that the predictions from tree-based models tended to underestimate biomass at high values and overestimate at low values, the accuracy of biomass estimates at local scales may be lower than the overall model statistics suggest.

In particular, saturation of the optical signal in closed canopy forests, model under-prediction of high biomass values, acquisition of images during the dry season and missing predictions for areas affected by cloud coverage were the main factors that tended to reduce the accuracy of the predictions and were responsible for the lower estimates of total biomass stock compared to the NBS values. Even if in some cases estimates based only on land cover data were overly optimistic for the reasons mentioned above and the true values cannot be known, it is likely that the estimates produced in this study were somehow conservative because the factors that may bias the predictions downward were predominant compared to those that may bias the predictions upward.

In conclusion, the significance of the presented approach is related to the fact that it can be applied to vegetation located in tropical and subtropical dry and semi-dry regions, which store relevant amounts of carbon. For example, the IPCC (2000) reports that carbon stored in vegetation of tropical savannah is comparable to the amounts stored in temperate forests, because the low biomass density of savannah ecosystems is counterbalanced by their large extents. Therefore, better monitoring methods can provide a significant contribution for a better understanding and for reducing the uncertainties of the global carbon cycle.

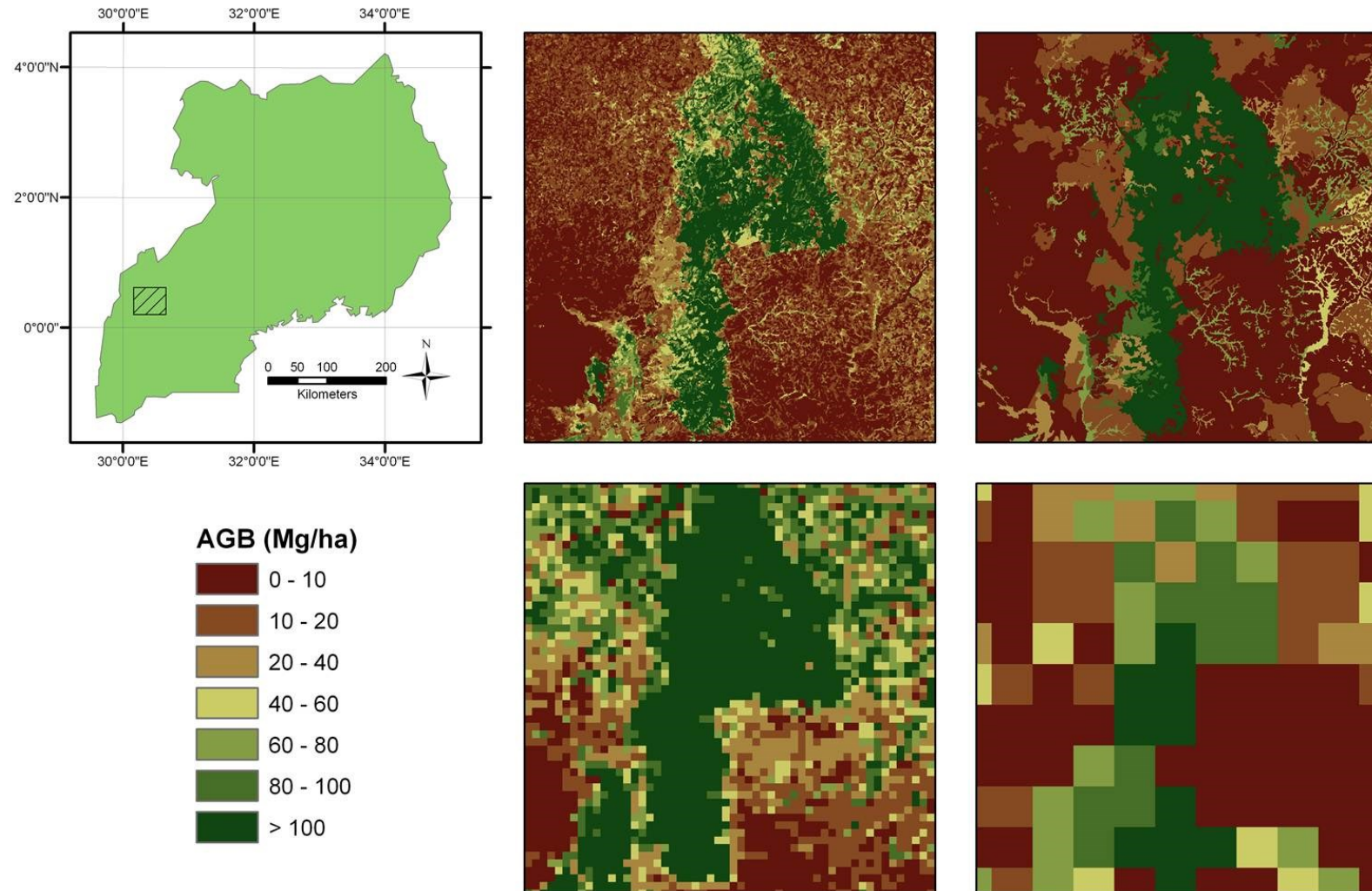


Figure 25: Comparison of biomass maps at different resolution for the test area indicated by the black frame in the map of Uganda. The biomass maps, from up-left clockwise, are derived from Landsat and LC data at 30 m pixel resolution (this study), from national LC data (adapted from Drichi, 2003), from LC data at 5 km resolution (Gibbs and Brown, 2007) and from MODIS data at 1 km resolution (Baccini et al., 2008)

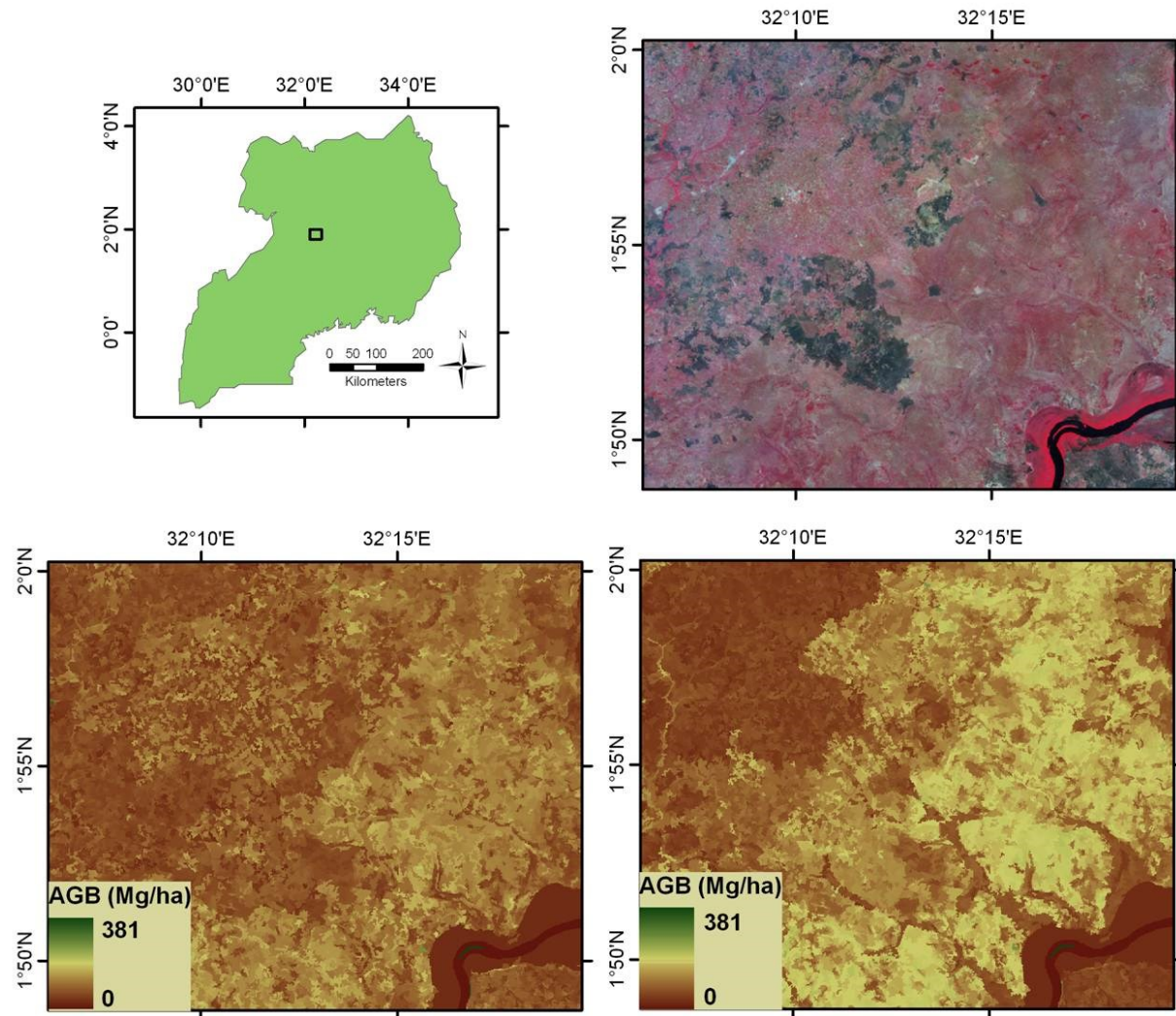


Figure 26: example of biomass underestimation using only Landsat data as predictor for the test area indicated by the black frame in the map of Uganda. Biomass density of the woodlands ($10 - 20 \text{ Mg ha}^{-1}$) located in the east and south part of the Landsat image (upper-right) was underestimated by the model not including LC (lower-left), while the higher estimates ($30 - 40 \text{ Mg ha}^{-1}$) provided by the model including LC (lower-right) were in agreement with the field data

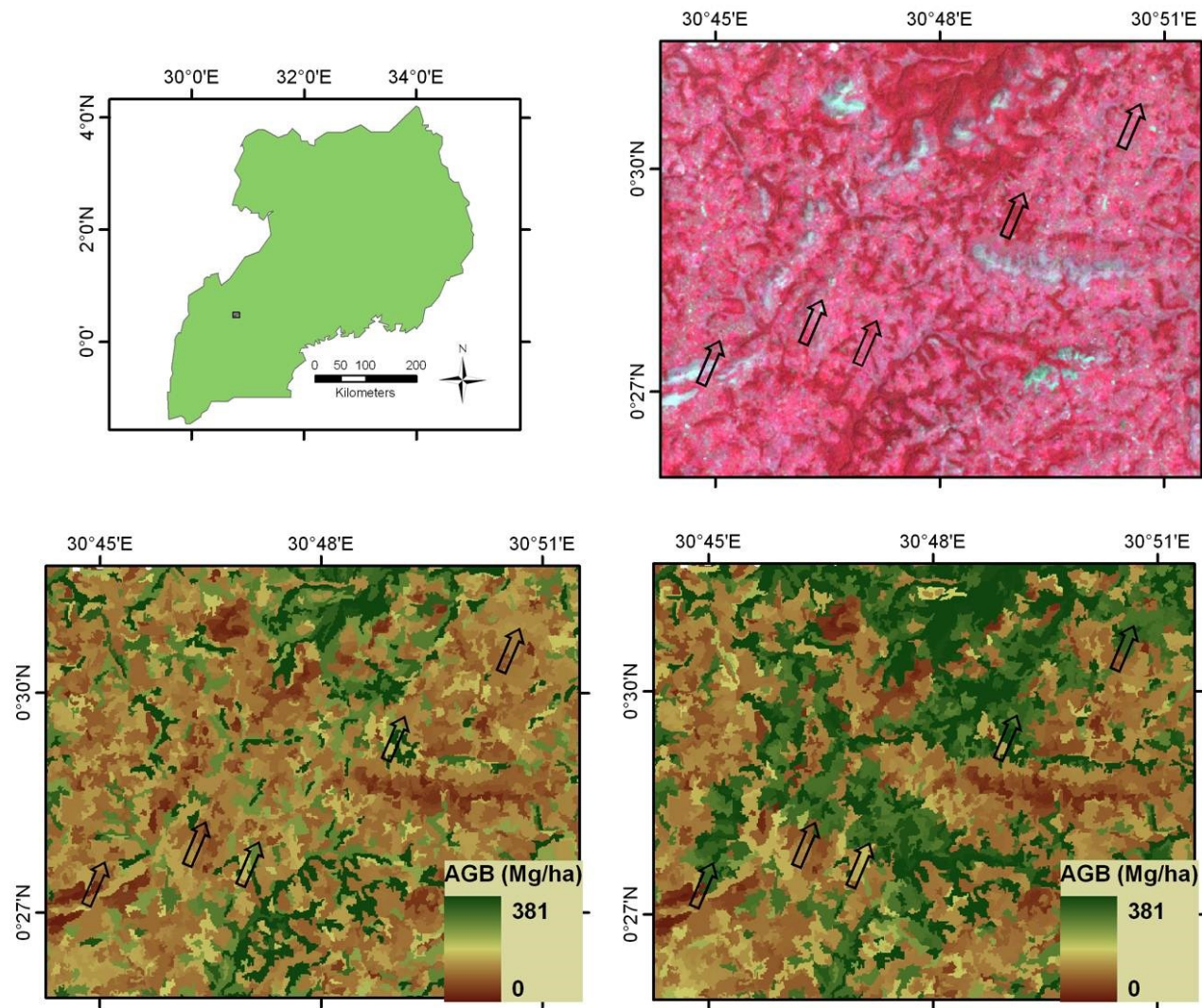


Figure 27: example of biomass overestimation using LC as predictor for the test area indicated by the black frame in the map of Uganda. Some deforested areas indicated by arrows in the Landsat image (upper-right) are classified as forest in the LC map, increasing the biomass estimates in the model including LC (lower-right) compared to the estimates based only on Landsat data (lower-left)

5. Comparison of existing biomass maps of Uganda

5.1. Biomass datasets

In order to better understand the reliability of existing biomass products and to compare the biomass map produced for Uganda (section 4.3.3) with published biomass datasets, the second part of this thesis consisted on the investigation of the similarity and accuracy of a set of biomass maps for the case study of Uganda. The analysis was aimed at understanding the causes of agreement and disagreement among the datasets and the effect of different input data and methodologies on the biomass predictions. In this chapter, the map produced in this thesis is referred to as the Avitabile map.

Six biomass or C stock maps, namely the Avitabile, Baccini et al. (2008), Drigo (2006), Gibbs & Brown (2007), Henry (2010) and Reusch & Gibbs (2008) maps, were compared to the NBS biomass map (Drichi, 2003) for the area of Uganda (Table 12). While some maps were based on global and regional datasets, others were based on country-specific data. The datasets with resolution lower than 5 Km (i.e. Kindermann et al., 2008; Olson et al., 1985) were not included in the present comparison because were considered too coarse in comparison with the limited extent of the country.

While it was not possible to define with certainty the most accurate map because the true biomass values cannot be known, the NBS biomass map was considered as reference because based on the largest number of nation-specific field data and a widely accepted methodology. Therefore, the similarity and accuracy of the maps were evaluated by comparing the biomass products to the NBS map and to a set of field data collected by the NBS. Specifically, the objectives of this comparative analysis were to use the NBS dataset to: (1) compare existing biomass datasets; (2) quantify their accuracy; (3) assess the effect of different input data and methodologies on the biomass estimates.

5.1.1. Maps description

The NBS map was not available for this study and was derived from the NBS Land Cover map (see section 4.2.2) by applying to each land cover stratum the average biomass density indicated in the NBS Report (Table 5.1 and 5.2 in Drichi, 2003). The land cover strata are defined on the basis of the land cover type, the ecological region and the biomass stock density class. Since the NBS Report did not provide a value for each class present in the map but reports the standard deviation per class, the biomass density of the missing strata was obtained from the map unit belonging to the same land cover type and ecozone but higher (lower) biomass density class by subtracting (summing) one standard deviation from (to) its value. Therefore, in this study the NBS biomass map does not refer to the map produced by the Uganda National Authority but to the slightly modified version obtained with the above procedure.

The NBS biomass map was compared to five biomass/carbon datasets. A brief description of the input data and methodology used by each dataset is reported below and in Table 12.

- 1) Avitabile (this thesis) derived AGB of Uganda from Landsat data acquired in the period 1999 - 2003, the NBS national land cover map and a subset of the NBS field plots using an object-oriented approach and a regression tree based model.

- 2) Baccini et al. (2008) derived AGB of tropical Africa from a pixel-based best-quality mosaic of MODIS data acquired in the period 2000 - 2003 and forest inventories data using a regression tree model. Training data were also derived from the NBS biomass map of Uganda.
- 3) Drigo (2006) derived AGB of Eastern Africa from the Africover national land cover datasets (www.africover.org), the FAO Global Ecological Zone (GEZ) map (FAO, 2001a) and biome average values derived from forest inventories.
- 4) Gibbs & Brown (2007) derived AGB of tropical Africa using the Global Land Cover map for the year 2000 (GLC2000) (Mayaux et al., 2004) and the CIESIN (2005) population density dataset to update the Brown & Gaston (1996) dataset, which in turn estimated AGB for the year 1980 on the basis of vegetation, population, soil, climate and elevation datasets in combination with biome average values and forest inventories from Tropical Africa and South East Asia.
- 5) Henry (2010) derived Aboveground Carbon (AGC) stock of sub-Saharan Africa from the Globcover 2005 land cover map (Arino et al., 2008), MODIS Vegetation Continuous Field (VCF) tree cover values (Hansen et al., 2003), climatic parameters, ecoregion maps and forest inventory data, using a Monte Carlo procedure to compute the mean AGC of each stratum from the field data.
- 6) Reusch & Gibbs (2008) derived above- and below-ground carbon density at global scale using the GLC2000 map, the FAO GEZ map, a frontier forest map (Bryant et al., 1997) and the IPCC Tier 1 default carbon stock density values (IPCC, 2006).

5.1.2. Mapping approaches

The biomass datasets can be classified according to the approach used to relate satellite or other spatial data to ground observations. Goetz et al. (2009) identified three main approaches:

- a) the Stratify & Multiply (SM) approach, where satellite data are used to derive a thematic map (e.g. vegetation map, land cover map) and the field data located within each strata are averaged;
- b) the Combine & Assign (CA) approach, where satellite and other spatial datasets are integrated to derive a thematic map with finer-grained strata;
- c) the Direct Remote Sensing (DR) approach, where satellite data are directly converted to AGB using classification techniques (e.g. neural network, regression trees).

Another possible approach, not considered in Goetz et al. (2009), is the Model approach, where satellite data are input to ecosystem models that predict a range of ecosystem parameters, including the standing biomass. According to this classification, the Baccini map is based on the DR approach whereas the Avitabile map is an extension of the DR approach (DR+) because it integrates satellite with LC data using a statistical model. Similarly, the Henry map is based on an extended CA approach (CA+) using a Monte Carlo procedure and satellite inputs in combination with categorical data (i.e. LC map) to obtain almost continuous biomass estimates. All other maps are mainly based on LC data but also employ additional datasets to identify strata with higher detail than the original LC map, and therefore follow the CA approach.

Table 12: Main characteristics of the biomass and C maps used for the comparative analysis. The spatial resolution of vector-based maps is given by a Variable Minimum Mapping Unit, which is class-specific. The spatial datasets used by the biomass/C maps included the Africover LC map, the Global Land Cover map for the year 2000 (GLC2000), the Globcover 2005 map, the FAO Global Ecological Zone (GEZ) map, the MODIS Vegetation Continuous Field (VCF) tree cover products, the CIESIN's Gridded Population of the World dataset and the Frontier Forest map. The number of biomass classes of the maps with continuous values was calculated after their conversion to integer values (i.e. classes of 1 Mg ha⁻¹ interval)

Map	Coverage	Format	Spatial resolution	Number of classes	Pool	Variable	Biomass data	Spatial data	Ancillary data	Period	Approach
NBS	Uganda	Vector	4-50 ha	48	Aboveground	Biomass	NBS Field data	NBS LC map	NBS Ecozone map	circa-2000	CA
Avitabile	Uganda	Raster	30 x 30 m	376	Aboveground	Biomass	subset of NBS Field data	Landsat	None	1999-2003	DR+
Baccini	Tropical Africa	Raster	1 x 1 Km	343	Aboveground	Biomass	Field data	MODIS	None	2000-2003	DR
Drigo	Eastern Africa	Vector	40-120 ha	143	Aboveground	Biomass	Biome average	Africover	FAO GEZ	circa-2000	CA
Gibbs & Brown	Tropical Africa	Raster	5 x 5 Km	176	Above and belowground	Biomass	Biome average	GLC2000	CIESIN Population	2000	CA
Henry	sub-Saharan Africa	Raster	300 x 300 m	346	Aboveground	Carbon	Field data	Globcover	FAO GEZ, MODIS VCF	2005	CA+
Reusch & Gibbs	Global	Raster	1 x 1 Km	14	Above and belowground	Carbon	IPCC Biome average	GLC2000	FAO GEZ, Frontier forest map	2000	CA

5.2. Methods

5.2.1. Analysis of biomass maps

5.2.1.1. *Map pre-processing*

Before the pair-wise comparison of each map with the NBS map, the datasets were standardized with regard to measurement variable, reference system, spatial coverage and spatial resolution. In order to investigate the effect of spatial resolution on the maps similarity, the datasets were also aggregated and compared at different resolutions (from 1 to 50 Km). Aggregation was performed through spatial averaging and resampling using the nearest neighbor method.

AGB was used as common measurement unit. The Gibbs & Brown, Henry and Reusch & Gibbs maps presented different units and were converted to AGB using the conversion factors originally applied in the map: the maps reporting carbon densities were converted to biomass densities using a carbon ratio of 0.47 for the Reusch & Gibbs map and 0.50 in the Gibbs & Brown and Henry maps; the maps reporting total (above- and below-ground) biomass were converted to AGB using the root-to-shoot ratios for each ecological zone and life form provided by IPCC (2006) for the Reusch & Gibbs map and the ratios for each vegetation type reported by Gaston et al. (1998) for the Gibbs & Brown map.

The maps, provided in different geographic systems, were projected to a common reference system before their comparison. Re-projection of a raster dataset inevitably causes deformation of its grid, which is corrected with the successive resampling that reduces the map geolocation accuracy. In order to minimize geolocation errors, which introduce artifacts when comparing maps, higher resolution datasets were re-projected to the reference system of the dataset at lower resolution. In addition, the maps were co-registered when their visual comparison indicated the presence of shifts between corresponding map features (e.g. water bodies, forest edges).

The maps were then aggregated to a common resolution on the basis of the following procedures, depending on the nature (vector, raster) and native resolution of the datasets.

When the maps to be compared were both raster-based, the dataset at higher resolution was aggregated and resampled to the grid and pixel size of the dataset at lower resolution.

When the maps to be compared maps were both vector-based, the map with smaller Minimum Mapping Unit (MMU) (i.e. higher spatial detail) was first converted to high-resolution (30 m) raster and then aggregated within the polygons of the map with larger MMU.

When the comparison was done between a vector-based and a raster-based map, a variable strategy was employed, depending on their relative resolution and based on the assumption that a vector polygon is mainly homogeneous with regard to the measurement variable (e.g. AGB) while a raster cell may correspond to a heterogeneous area. Hence, when a raster cell overlaps two or more polygons, its value should be compared to the area-weighted average of the polygon values. On these premises, when the raster-based map had a lower resolution than the vector-based map, the latter was converted to high-resolution raster and then aggregated to the cell size and grid of the low-resolution raster. Instead, when the raster-based map had a higher resolution than the vector-based map, the raster was aggregated within the vector polygons and the comparison was performed at polygon level.

The reference NBS map, which was vector-based with a variable MMU ranging from 0.04 Km² for forests to 0.5 Km² for grasslands, was considered as having higher resolution than the raster maps with cell size ≥ 1 Km but lower resolution than the raster maps with cell size ≤ 300 m.

5.2.1.2. *Map comparison statistics*

The level of agreement of the biomass maps with the reference NBS map was assessed by comparing total and spatial distribution of the biomass estimates. Specifically, the comparison was performed on the basis of the following statistics:

- Total biomass
- Bias
- Map histogram
- Scatterplot
- Fuzzy Numerical index
- Fuzzy Numerical maps
- Difference maps
- Variograms

The first three parameters are not spatially explicit while the other statistics measure the similarity of the biomass spatial distribution. Each statistic was computed at the map native resolution and, after aggregation through averaging, also at the following resolutions: 1 Km, 5 Km, 10 Km, 20 Km, and 50 Km.

The first statistic, total biomass, is the total AGB (Tg) within the country, obtained by summing the product of the biomass density of each map unit (pixel, polygon) with its area. The bias (Mg ha⁻¹) is the difference between the mean biomass densities of two maps, where the mean densities were obtained by dividing the total biomass by the national land area. The map histograms represent the frequency distribution of the map values in biomass classes while the scatterplots visually represent the difference between corresponding map cells. The difference maps, obtained by subtracting corresponding map cells, provide the spatial distribution of the difference of the map estimates at pixel level.

The Fuzzy Numerical (*FN*) index (van Vliet et al., 2009; Hagen-Zanker, 2006), equivalent to the Fuzzy Kappa statistic (Hagen-Zanker et al., 2005) for continuous data, measures the similarity of spatial patterns between two numerical raster maps, ranging between 0 (fully distinct) and 1 (fully identical). The index is computed as the average of the numerical similarity s between each pair of corresponding cells (a and b) in the two maps, which in turn is computed cell-by-cell as follow:

$$s(a, b) = 1 - \frac{|a - b|}{\max(|a|, |b|)} \quad (5.1)$$

where the cell values (a and b) are re-computed considering the neighboring cells within a specified window. Specifically, the cell values are re-computed as weighted average of a 4-by-4 window, using an exponential decay function for the weights of the neighboring pixels. Sensitivity analysis showed that the absolute value of the index is sensitive to the algorithm

parameters (i.e. the window size and weighting function) but the relative agreement among the maps did not change using different algorithm values.

By considering similarity between areas larger than single pixels, the *FN* index reduces the effect of co-registration errors between maps and focuses on the similarity between spatial patterns more than corresponding cells. The *FN* index provides a single measure of the maps' overall agreement while the *FN* map represents the numerical similarity at pixel-level. These statistics were obtained using the open source software Map Comparison Kit v. 3.2 (www.riks.nl/mck). The current version of the *FN* index does not account for autocorrelation, causing lower agreement for maps with strong autocorrelation (Hagen-Zanker, 2009).

Image variograms provide information on the spatial variation of the measurement variable in the image, with their behavior being related to the density, the size of the objects and the variance present in the scene (Woodcock et al., 1988a, 1988b). Variograms can be derived from data points (punctual variogram) or from images (regularized variogram).

In this study, the variograms of the biomass maps were computed to represent and compare the spatial variation of areas with homogeneous biomass density (the "objects") in the datasets. Since the variograms are sensitive to the spatial resolution of the dataset (Woodcock et al., 1988a), the maps were converted to raster format at 1 Km resolution before the analysis and the Gibbs & Brown map was excluded because of its lower native resolution (5 Km).

The map variograms were also compared with the punctual variogram derived from the NBS field plots to identify the map most able to maintain the spatial variation represented by the field data.

The NBS plots were systematically located at 5 x 10 Km distance and, at each grid intersections, clusters of 1 to 5 plots were located at 300 m distance. Since the variogram analysis depends on the spatial distribution of the input data and the input data were irregularly distributed due to the cluster sampling strategy, the plot distribution was regularized by computing the average biomass density within each cluster and referring this value to the coordinates of the cluster center (or grid intersection). On the basis of this procedure, 1184 data points were originated from the original 3510 plots, and these data points at a constant 5 x 10 km distance were used to compute the NBS plot variogram.

The plot distribution was not homogeneous throughout the country because the Northern part of Uganda and the Forest Reserves were under-sampled (see section 4.2.3.1). In addition, the variogram derived from the NBS plots presents a different spatial scale and coverage than the 1 km biomass maps but, due to the systematic sampling strategy and the large number of data points, it provides the best available approximation of the biomass variability on the ground.

The variograms were computed using the package *gstat* (Pebesma, 2004) of the open source software R (R Development Core Team, 2009). Water bodies were masked before the analysis.

5.2.2. Integrating biomass maps with field data

Comparing remote sensing products, as biomass maps, may reveal common trends and dissimilarities in the datasets but does not assess their accuracy. A proper validation requires the comparison of the remote sensing estimates with accurate, independent and comparable ground observations. However, suitable reference data comparable with medium-coarse resolution maps are usually not available and must be upscaled to match the spatial resolution of the remotely sensed data.

In this study, the biomass maps were compared with the NBS field data. Before the comparison with the biomass maps, the field plots were up-scaled to the resolution of the remote sensing datasets using the following procedure.

For vector-based maps, the biomass density values of the plots located within each polygon were averaged.

For raster-based maps, the field plots were upscaled to raster resolution using the NBS national land cover map. The upscaling procedure was based on the assumption that the plots were representative of the biomass density of the land cover polygon in which they were located. The procedure consisted on selecting the pixels where the field plots represented, through land cover polygons, at least 90% of the pixel area. Then, the plots located within each selected pixel were area-weighted averaged on the basis of the NBS land cover map (i.e. weighting the plots according to the fraction of the corresponding LC polygon area within each pixel). If, within a pixel, there was more than one plot inside the same polygon, the average plot biomass was used. The Avitabile and Henry maps were aggregated to 1 Km resolution before the analysis.

The upscaled plot values were then compared with the estimates of the corresponding unit (pixel or polygon) of the biomass maps. Since categorical maps (i.e. maps based on the SM or CA approach) attribute an average value to the map units belonging to the same class, the averages of all upscaled field plots located within the same biomass class were also compared with the class value itself.

Lastly, in order to test the assumption that 1 plot (area of 0.0025 Km²) may not be comparable with the larger map units (area of 1 Km² for most datasets), the comparison statistics were re-computed selecting only the map units with at least 2 field plots.

The reliability of this validation procedure depends on the accuracy of the reference dataset (Congalton, 1991). In order to use field data to validate large-scale biomass maps, biomass reference values at map resolution are estimated from the plant parameters measured at plot level. During this process, errors are originated from several sources, in particular from the use of allometric equations (Chave et al., 2004) and from the upscaling procedure (Baccini et al., 2007). In the present analysis, these errors were reduced using the NBS country-specific allometric equations (Drichi, 2003) and a large field dataset (3510 plots), upscaled on the basis of a high-resolution map (the NBS land cover map).

5.2.3. Analysis of biomass reference values and spatial data

The biomass maps considered in this study were obtained by combining biomass reference values with spatial datasets.

In order to separate and quantify the impact of these two components on the map estimates, the same set of biomass reference values was applied to different spatial maps, and vice-versa. Specifically, the most detailed biomass reference values (the NBS plots) and then the most general values (the IPCC Tier 1 values) were applied to the NBS LC map, to the GLC2000 map stratified by the FAO GEZ map (as in the Reusch & Gibbs map) and to the Globcover map stratified by the FAO GEZ map (as in the Henry map). The NBS plots were also applied to Landsat data in the Avitabile map.

In order to apply the NBS reference data to other spatial datasets, the NBS field plots were averaged within each class of the spatial map, and then the average values were multiplied with the area extent of the corresponding class. Instead, in order to apply the IPCC Tier 1 values to the NBS land cover map, the NBS LC map needed to be harmonized with the IPCC Tier 1

classification scheme. To do this, the NBS LC map was first aggregated to the IPCC land cover classes (according to the reclassification scheme reported in Table 13) and then stratified by the GEZ ecozones. Eventually, the IPCC Tier 1 values were multiplied with the area extents of the corresponding classes.

The biomass estimates of each map were also compared for forest and non-forest areas separately. In addition, the NBS biomass reference values for forest areas only were detailed compared with the reference values used by the other maps in order to identify the causes of disagreement.

Table 13: Reclassification scheme used to convert the NBS land cover (LC) classes to the IPCC classes.

NBS LC classes	IPCC Classes
1 - Plantation	Forest
2 - Plantation	Forest
3 - Closed forest	Forest
4 - Open forest	Forest
5 - Woodland	Forest
6 - Shrubland	Shrub
7 - Grassland	Grass
8 - Wetland	Grass
9 - Small cropland	Crop
10 - Large cropland	Crop
>10 - No Data	No Data

5.3. Results

The comparison results are to be interpreted keeping in consideration that the Avitabile and Baccini maps were not independent from the reference data. Specifically, the Avitabile map used the NBS field plots and the NBS LC map to estimate biomass density while the Baccini map used the NBS biomass map to derive training data.

On the other hand, the Drigo, Gibbs & Brown, Henry and Reusch & Gibbs maps were independent from the reference data and the comparison results represent an independent validation of their performance for the area of Uganda.

However, the results were partially affected by differences in map formats and resolutions. For example, coarser maps, (e.g. Gibbs & Brown, Drigo) were favored in the comparison because the similarity between the datasets tended to increase at lower resolution (see Figure 31).

5.3.1. Comparison of biomass maps

5.3.1.1. *Total aboveground biomass of Uganda*

The comparison of biomass stock of Uganda revealed strong disagreement between the remote sensing products, with estimates of total AGB ranging from 343 to 2201 Tg (Table 14, Figure 28). The Baccini map provided the closest estimate (443 Tg) to the NBS reference value (468 Tg) while the Reusch & Gibbs map provided the most different value (2201 Tg).

Estimates from maps based on the DR approach were conservative (i.e. negative bias) while maps based on the CA approach provided higher values than the NBS (i.e. positive bias).

The map histograms (Figure 29) showed large differences among the distribution of the biomass estimates, with the Avitabile, Baccini and NBS maps as well as the NBS field plots being concentrated at low values ($< 100 \text{ Mg ha}^{-1}$) and the Drigo, Gibbs & Brown, Henry and Reusch & Gibbs maps presenting higher frequencies at higher values.

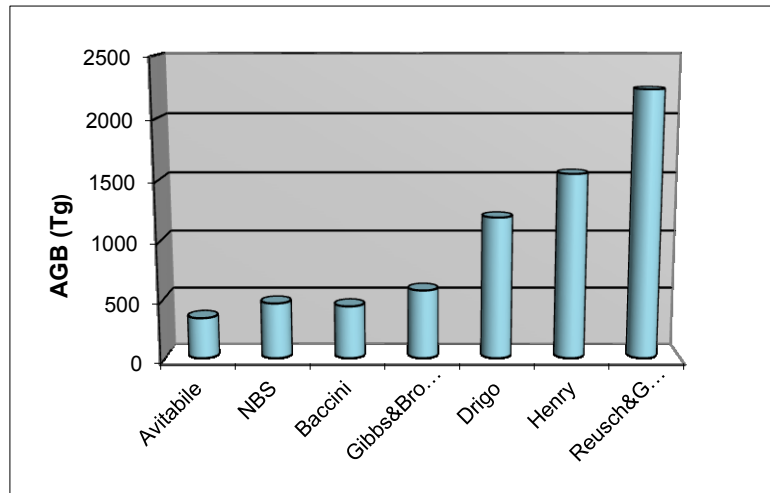


Figure 28: Comparison of the estimates of total AGB stock of Uganda

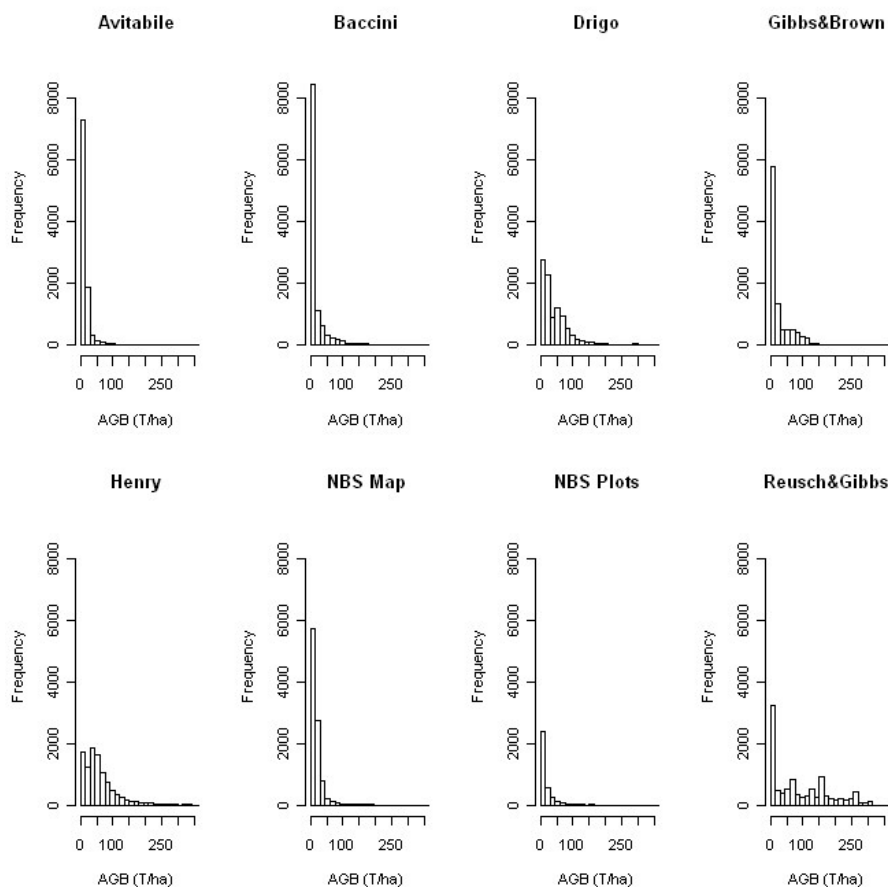


Figure 29: Histograms of the biomass maps and NBS field plots. The histograms represents the frequency (i.e. number of occurrence) (y axis) for each AGB class (x axis). AGB is reported in Mg ha^{-1} . The histograms are derived from maps aggregated at the resolution of the coarser map ($5 \times 5 \text{ Km}$) for consistent representation of frequency values

5.3.1.2. Other estimates of Uganda's biomass

Similar level of disagreement among AGB estimates of Uganda, as the tendency of maps based on biome average values and coarse spatial datasets to overestimate this parameter, was observed in Gibbs et al. (2007). This study estimated total above- and belowground C stock for several tropical countries combining four sets of biomass reference values with the GLC2000 map stratified by the FAO GEZ map. After converting the Gibbs et al. (2007) values for Uganda to AGB by applying a biomass conversion factor of 0.47 and the root-to-shoot ratio for rainforest (the most common ecozone of Uganda) of 0.37 (IPCC, 2006), the estimates were between 674 and 1921 Tg, within the range found in the present comparison but consistently above the NBS reference value. Higher AGB values were obtained using root-to-shoot ratios for the other ecozones present in Uganda.

High variability in the estimates of forest biomass was also noticed comparing the FAO Forest Resource Assessment (FRA) statistics for Uganda for the year 2000, which indicate 681 Tg in the FRA 2000 Report (FAO, 2001b) and only 244 Tg in the FRA 2005 Report (FAO, 2006). The difference, due to the adoption of the NBS data in the FRA 2005 Report, indicates the sensitivity of this parameter to the input data and methodology used to estimate it. It should be noted that the value reported in the FRA 2005 refers to forest AGB and is about half of the NBS value because in Uganda large amount of biomass are stored in areas outside forest land.

Table 14: Total and mean AGB of Uganda for different biomass maps. The map bias is relative to the NBS value. The values for the maps indicated with the symbol (a) are derived from Gibbs et al. (2007)

Map	Total AGB (Tg)	Mean AGB (Mg/ha)	Bias (Mg/ha)
Avitabile	343	14.2	-5.2
Baccini	443	18.4	-1.0
NBS	468	19.4	-
Gibbs & Brown	579	24.0	4.6
Drigo	1191	49.3	29.9
Henry	1550	64.2	44.8
Reusch & Gibbs	2201	91.1	71.7
Houghton(1999)/DeFries et al. (2002) ^a	674	27.9	8.5
Brown(1997)/Achard et al. (2002, 2004) ^a	744	30.8	11.4
Olson et al.(1983)/Gibbs(2006) ^a	832	34.5	15.1
IPCC(2006) ^a	1921	79.5	60.2

5.3.1.3. Spatial distribution of biomass

Differently from summary statistics (e.g. total biomass) where positive and negative differences may compensate each other, spatial comparison of the maps reveals their similarity at local level. Different but related information regarding the agreement of the biomass maps with the NBS reference map were provided by the scatterplots (Figure 30), the difference maps (Figure 32), the Fuzzy Numerical (*FN*) index (Figure 31), the Fuzzy Numerical maps (Figure 33) and the variograms (Figure 34). The results of the maps spatial comparisons (i.e. scatterplots, difference maps, Fuzzy Numerical maps) computed at variable resolution (1, 5, 10, 20 and 50 Km) are reported in Appendix A.

The scatterplots, difference maps and *FN* maps showed that the Reusch & Gibbs map strongly overestimated AGB distribution over most of the country and especially at low biomass values. The Gibbs & Brown map overestimated AGB in the northern part of Uganda and underestimated it in the south-eastern region. Conversely, both the Drigo and Henry maps overestimated AGB in the southern region and at high biomass values, and presented estimates similar to the NBS in the northern region. The spatial analysis of the Baccini map, which total AGB was very similar to the NBS value, revealed the presence of local differences, with overestimation in the western region counterbalanced by underestimation in the center-eastern region. The Avitabile map presented high spatial agreement with the NBS map and provided lower biomass densities over most of the country, with higher differences in the forest areas. The *FN* index confirmed these findings and showed that the spatial agreement between the maps usually increased at lower spatial resolution.

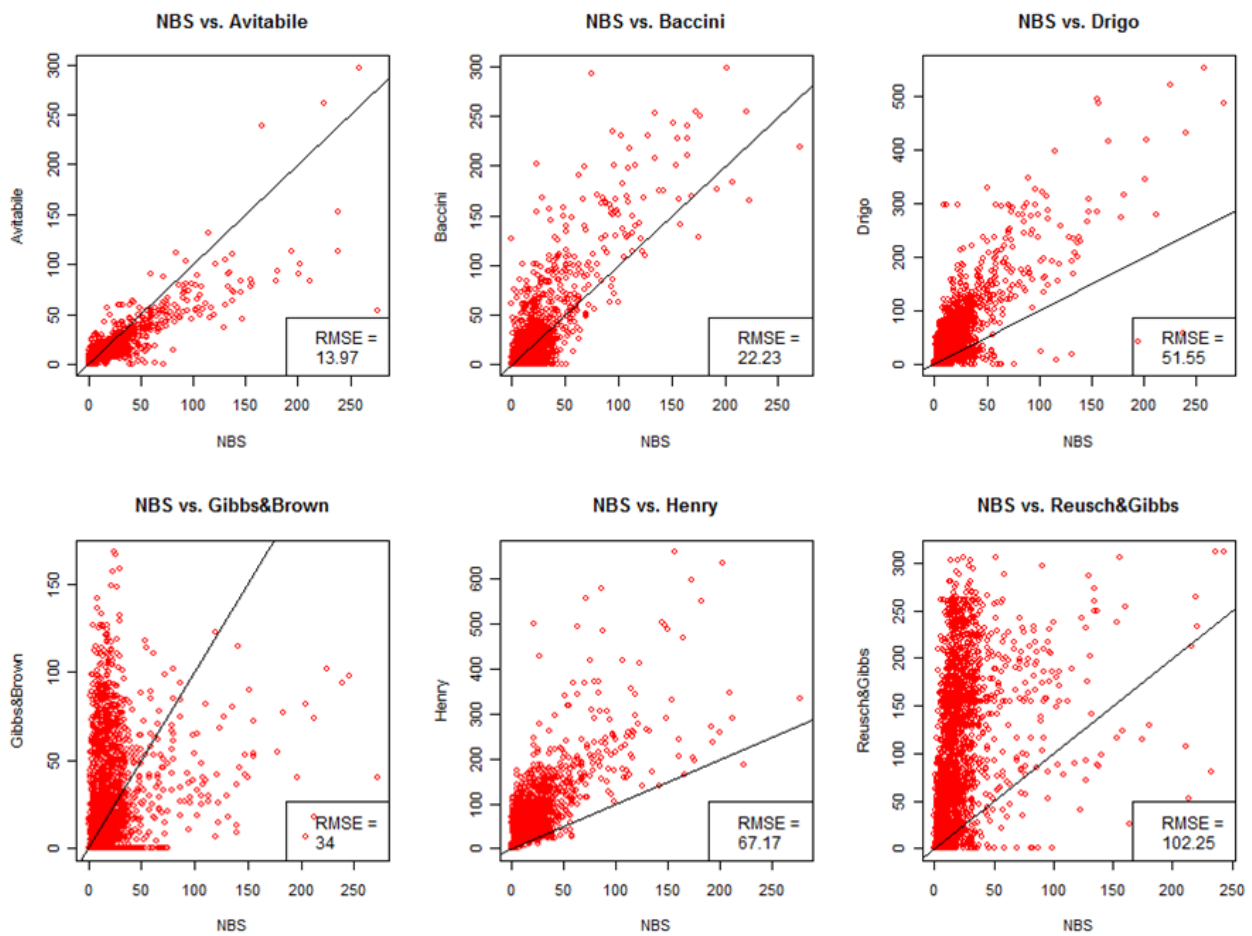


Figure 30: Comparison of AGB values between the NBS and the other biomass maps. The comparison is performed at pixel level and the results are reported for the maps aggregated at 10 Km resolution for graphical reasons. The distance along the axis of the points to the 1:1 line represents the difference between the AGB estimates, in Mg ha^{-1}

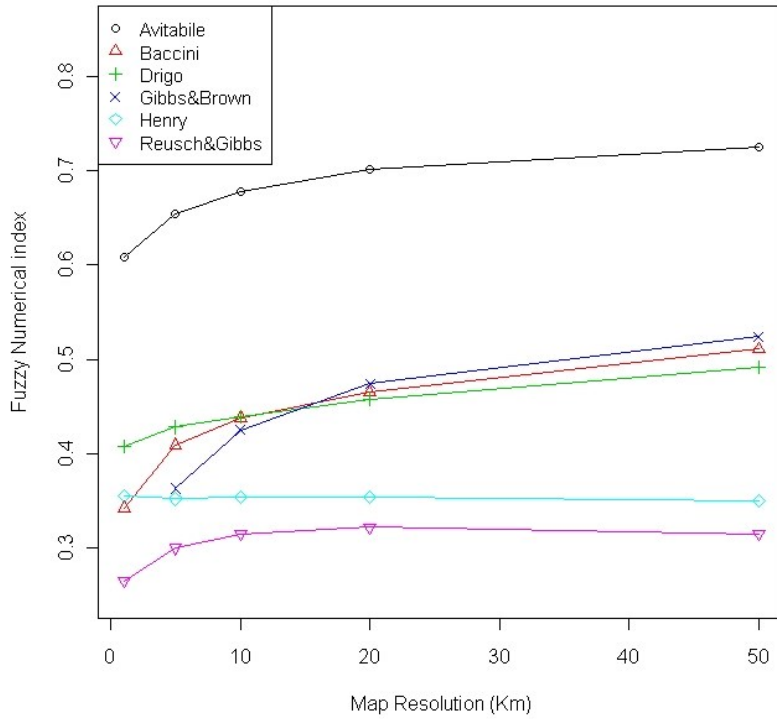


Figure 31: Fuzzy Numerical index, representing the mean similarity between the biomass maps and the NBS map computed at different spatial resolutions (from 1 to 50 Km)

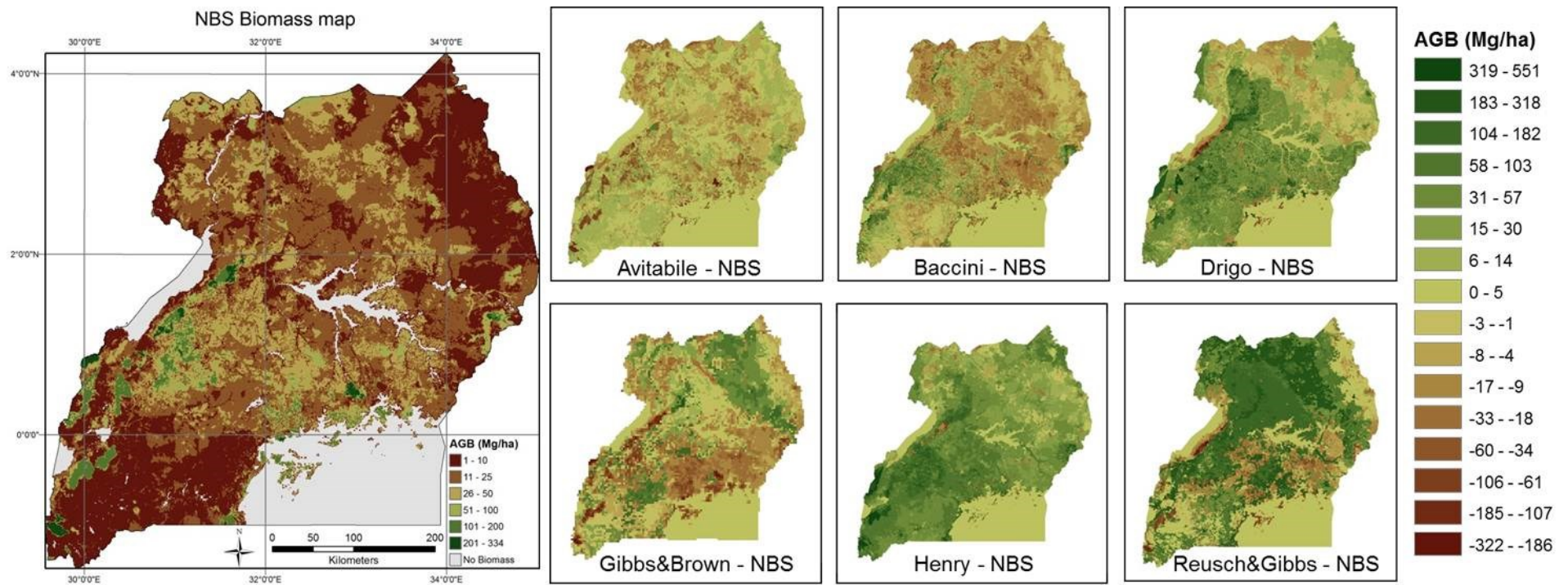


Figure 32: NBS reference biomass map (left) and difference of AGB values between the biomass maps and the NBS map (right). The difference maps, obtained by subtracting the corresponding map pixels, indicate overestimation with positive values (in green) and underestimation with negative values (in brown) in comparison to the NBS map

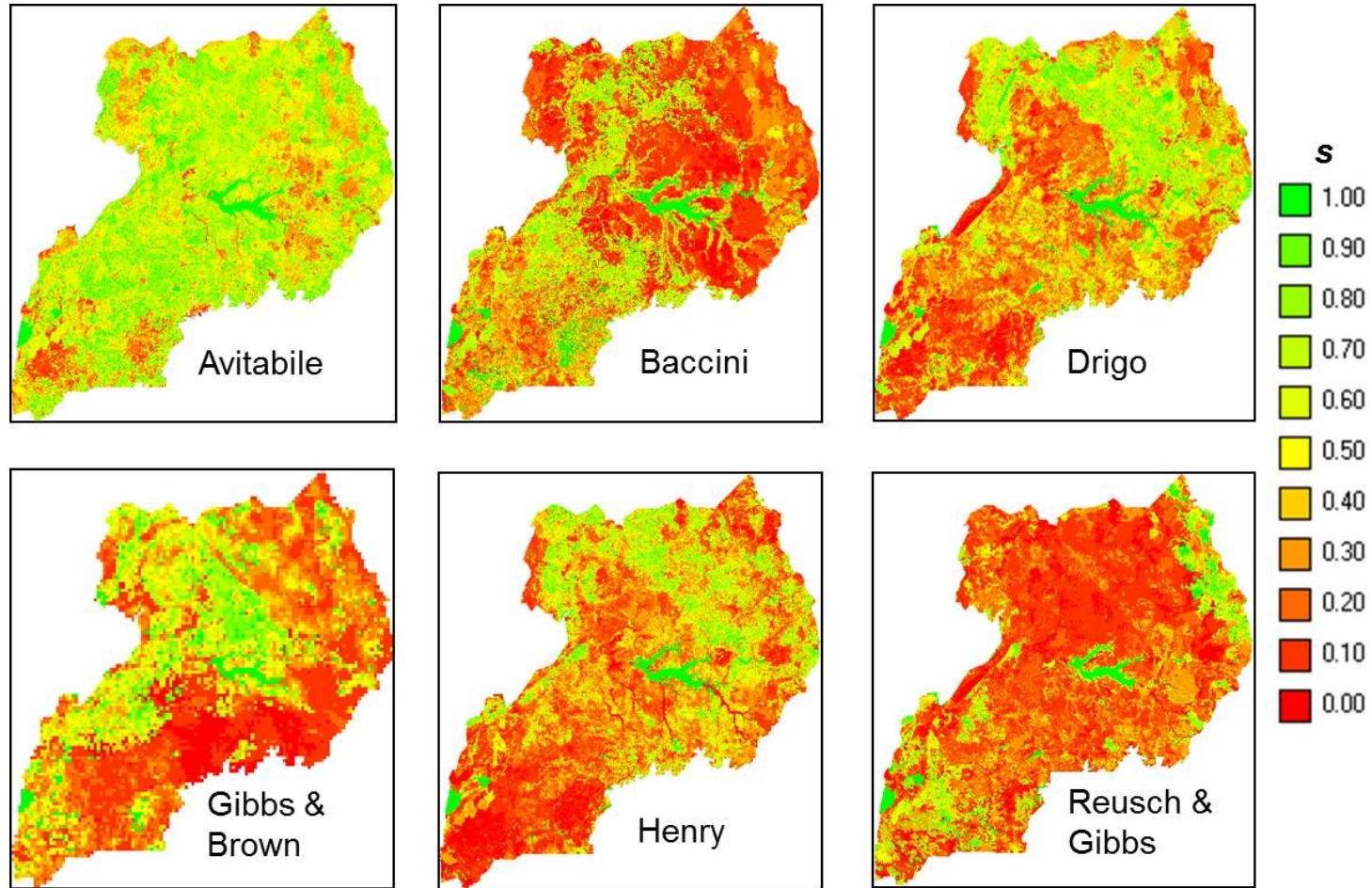


Figure 33: Fuzzy Numerical maps, representing the spatial distribution of the numerical similarity (s) between the biomass maps and the NBS map, ranging from 0 (fully distinct) to 1 (fully identical)

The comparison of variograms (Figure 34) indicated a large variability in the variance of the datasets. The field plot variogram was best approximated by the Avitabile and NBS variograms, suggesting that these two maps best maintained the biomass spatial variation represented by the field data. Instead, the other maps presented higher semivariance, indicating larger variations in biomass predictions compared to those observed from the field data at corresponding distances.

Variogram behavior was related to the stratification approach employed by the maps and showed that maps based on coarse stratification layers and biome average values (i.e. Reusch & Gibbs, Drigo) could not capture the high biomass spatial variability represented by the field plots but mapped large homogeneous areas or few, distinct biomass classes. On the contrary, maps based on a regression approach or several detailed strata (i.e. Avitabile, NBS) identified small biomass areas with a continuum of values, representing more closely the spatial characteristics of the field data. The variogram of the Baccini map, based on a regression approach but with coarse resolution, showed an intermediate behavior while the Henry map, despite its higher resolution and large number of biomass classes, represented a biomass spatial variation not matching with that of the field plots.

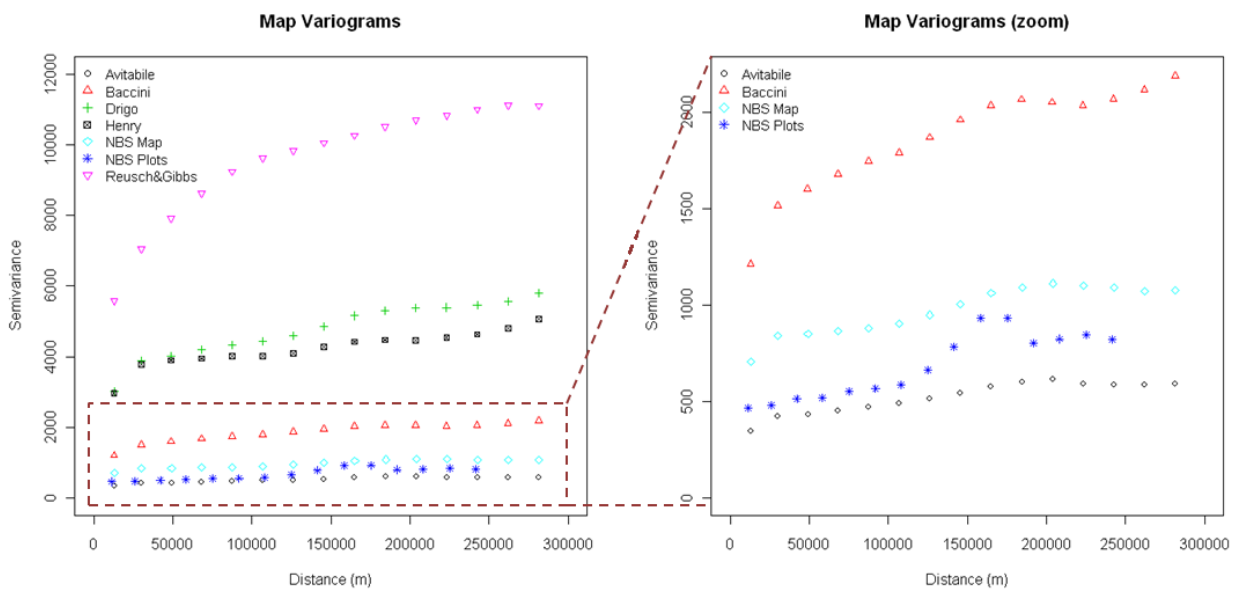


Figure 34: Variograms of the biomass maps and NBS field plots (left), with a zoom of the variograms for semivariance values between 0 and 2500 (right)

5.3.1.4. Comparison of biomass stocks in forest and non-forest areas

The AGB estimates of the biomass maps were also compared separately for forest and non-forest areas. Since the maps presented different mapping approaches and forest definitions (resulting in different distributions of forest areas), the NBS map was used to identify forest areas in all maps in order to obtain comparable results. Therefore, the forest and woodland areas in the NBS Land Cover map were aggregated to derive a forest mask that was applied to the other maps, and the biomass estimates for the areas classified as forest and as non-forest were compared separately.

The results indicated that most of the difference between the biomass maps and the NBS values was located in areas classified as non-forest by the NBS (Table 15, Figure 35). With regards to the maps that overestimated total AGB of Uganda, the overestimation was mostly concentrated in areas outside forests (as classified by the NBS) and it was due to the combination of higher

biomass reference values for non-forest vegetation types with a larger mapping of forest areas. The larger extent of forest areas was due to different forest definitions and mapping methodologies (e.g. data sources, minimum mapping unit). The relative contribution of the two factors (different biomass reference values and forest mapping) to the overall difference could not be separated in this analysis.

Table 15: Comparison of biomass estimates with the NBS values for all areas (left columns), forest areas only (center columns) and non-forest areas only (right columns). Forest areas were defined according to the NBS land cover map

Map	Total AGB		Forest AGB			Non – Forest AGB		
	Tg	Difference to NBS	Tg	% of Total AGB	Difference to NBS	Tg	% of Total AGB	Difference to NBS
Avitabile	343	-27%	186	54%	-29%	157	46%	-25%
NBS	468	-	260	56%	-	208	44%	-
Baccini	443	-5%	242	55%	-7%	201	45%	-3%
Gibbs&Brown	579	24%	216	37%	-17%	363	63%	74%
Drigo	1191	154%	548	46%	111%	643	54%	209%
Henry	1550	231%	636	41%	144%	914	59%	339%
Reusch&Gibbs	2234	377%	793	35%	205%	1441	65%	592%

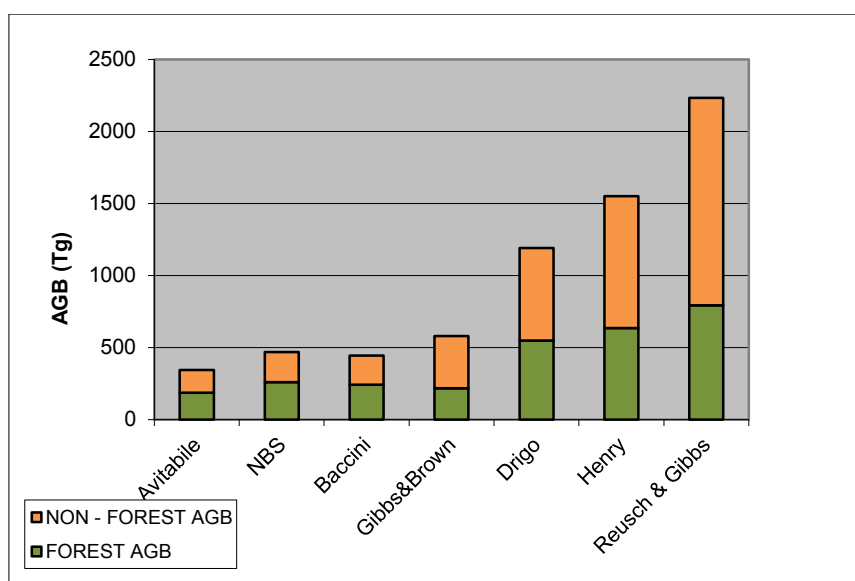


Figure 35: Biomass stock (Tg) in forest and non-forest areas for the biomass maps. Forest areas were defined according to the NBS land cover map

5.3.1.5. Comparison of biomass density values in forest areas

The biomass density values reported by the NBS in forest areas were usually lower than the corresponding values reported by most of the other maps. The average biomass density reported by the NBS for dense forests was equal to 223 Mg ha⁻¹ while the corresponding value was equal to 310 Mg ha⁻¹ for the Reusch & Gibbs map and 562 Mg ha⁻¹ for the Drigo map. Similarly, the maximum NBS value for highly stocked dense forests was equal to 334 Mg ha⁻¹ with the

maximum value at plot level equal to 523 Mg ha⁻¹, while other studies (e.g. Mitchard et al., 2009) found higher values at forest and plot levels.

The lower NBS values may be due to the fact that the objective of the NBS was to estimate biomass mainly for energy purposes (bioenergy) and therefore the study focused on the main accessible sources of biomass, under-sampling forest areas of difficult accessibility. Since forests located in accessible areas are likely to have lower biomass than remote forests, the NBS values for forests may be biased towards low values. On the other hand, the biomass density values of the NBS map are related to the map spatial resolution, which presented a minimum mapping unit of 4 ha for forests and 10 ha for woodlands. While biomass density in forests at plot resolution can reach very high values, this parameter usually decreases rapidly when computed on larger areas. Most importantly, the NBS map reports biomass density per land cover class. Therefore the value of 223 Mg ha⁻¹ represents the average biomass density of the class “forest”, and higher-than-average map units (mature, rich forests) should be compensated by lower-than-average units (degraded forests). Instead, the higher biomass values reported by other studies are likely to represent specific sites and may not represent the average forest conditions of Uganda.

5.3.2. Comparison of maps with field data

The comparison of the biomass maps with the field plots (Table 16) confirmed the large differences among the datasets.

The error estimate was higher for the Reusch & Gibbs, Henry, Drigo and Gibbs & Brown maps, which presented a RMSE equal to 66.7, 62.2, 57.2 and 43.1 Mg ha⁻¹ respectively. Since the reference field data were independent from these maps, the error estimates quantify the map accuracies for the area of Uganda. However, the RMSEs were reduced by the skewed distribution of the field data (Figure 17), which focused the comparison on low biomass areas where prediction errors tended to be smaller than those in high biomass areas.

The other three maps, directly or indirectly related to the field data, presented lower error estimates.

The RMSE of the Avitabile map aggregated at 1 Km resolution (17.7 Mg ha⁻¹) was comparable with the value computed at the map's native resolution using independent NBS field plots (15 Mg ha⁻¹) (see section 4.3.1). Similarly, the RMSE of the NBS map (24.1 Mg ha⁻¹) was comparable with the mean standard deviation of the biomass strata (16.9 Mg ha⁻¹) reported by Drichi (2003). Instead, the RMSE of the Baccini map for Uganda (24.7 Mg ha⁻¹) was lower than the value reported by Baccini et al. (2008) for tropical Africa computed using independent reference data (50.5 Mg ha⁻¹), possibly because in the present analysis the map was indirectly related to the reference data.

When the comparison was performed selecting only the map units (pixels or polygons) with 2 or more field plots, there was only a small increase in the map accuracies, while the number of units available for the comparison decreased considerably (Table 16). When the comparison was performed for biomass classes (as defined in section 5.2.2) instead of map units, the RMSE was higher (Table 17), possibly because the aggregation in classes reduced the effect of the skewed distribution of the field data.

Table 16: Comparison of the biomass maps with the NBS field data. The comparison is performed using all map units with at least 1 field plot (“All plots”, left columns) and selecting only the map units with 2 or more field plots (“Plots ≥ 2 ”, right columns). N is the number of map units used for the comparison

Map	All plots			Plots ≥ 2		
	N	Bias (Mg/ha)	RMSE (Mg/ha)	N	Bias (Mg/ha)	RMSE (Mg/ha)
Avitabile	850	-2.9	17.7	173	-1.7	13.6
Baccini	888	-4.8	24.7	184	-7.5	23.5
Drigo	985	39.3	57.2	664	38.8	54.0
Gibbs&Brown	92	9.2	30.4	31	6.1	28.0
Henry	878	51.1	62.2	183	57.1	69.9
NBS	1129	1.1	24.1	693	0.9	19.5
Reusch&Gibbs	778	35.5	66.7	160	31.5	65.1

Table 17: Comparison of the biomass maps with the NBS field data by biomass classes. The comparison is performed using all map units with at least 1 field plot (“All plots”, left columns) and selecting only the map units with 2 or more field plots (“Plots ≥ 2 ”, right columns). N is the number of biomass classes with field plots for each map

Map	All plots			Plots ≥ 2		
	N	Bias (Mg/ha)	RMSE (Mg/ha)	N	Bias (Mg/ha)	RMSE (Mg/ha)
Avitabile	52	-0.9	24.7	45	-0.9	20.0
Baccini	75	26.7	45.0	56	14.5	31.5
Drigo	94	47.7	64.5	90	45.5	61.5
Gibbs&Brown	36	36.2	47.9	14	9.3	34.0
Henry	163	75.4	88.5	130	67.4	79.6
NBS	34	8.6	28.7	33	7.8	28.5
Reusch&Gibbs	11	70.7	104.3	11	70.7	104.3

5.3.3. Comparison of biomass reference values and spatial data

The comparison of AGB estimates obtained using different combinations of the input data showed that applying different biomass reference values to the same spatial map caused very large variations in the biomass estimates (219 - 504%) (Table 18). On the contrary, using different spatial maps to stratify a set of biomass reference value caused much smaller variations (20 - 33%) (Table 19). However, the spatial datasets influenced the distribution of the estimates and, according to the *FN* index, the overall spatial agreement of the biomass maps with the NBS reference map decreased using global LC datasets instead of Landsat data or the NBS LC map (Table 19).

Table 18: Total AGB using different combinations of spatial and biomass data. The table also reports the percentage difference of total AGB using different biomass reference data for each spatial dataset

Spatial data	Biomass data	Total AGB (Tg)	Difference (%)
NBS LC	NBS plots	468	-
	IPCC Tier1	1,492	219%
GLC2000 + FAO GEZ	NBS plots	377	-
	IPCC Tier1	2,234	493%
GlobCover + FAO GEZ	NBS plots	363	-
	IPCC Tier1	2,191	504%

Table 19: Total AGB using different combinations of biomass and spatial data. The values indicated with the symbol ^(a) and ^(b) are derived from Driichi (2003) and Avitabile (this thesis), respectively. The table also reports the percentage difference of total AGB using different spatial dataset for each biomass reference data and the Fuzzy Numerical index, which is computed by comparing each map with the NBS reference map (NBS plots + NBS LC) at 1 Km resolution

Biomass data	Spatial data	Total AGB (Tg)	Difference (%)	Fuzzy Numerical
NBS plots	NBS LC	468 ^a	-	-
	GLC2000+FAO GEZ	377	-19.5%	0.568
	GlobCover+FAO GEZ	363	-22.5%	0.565
	Landsat	343 ^b	-26.7%	0.608
IPCC Tier1	GLC2000+FAO GEZ	2,234	-	0.265
	GlobCover+FAO GEZ	2,191	-1.9%	0.292
	NBS LC	1,492	-33.2%	0.410

The results also indicated that the use of the IPCC Tier 1 values in combination with different spatial maps always over-estimated AGB of Uganda (Table 18). The comparison of the biomass estimates separately per vegetation type revealed that most of the overestimation was due to the IPCC reference values for forest (ranging 115 – 310 Mg ha⁻¹), which were significantly higher than the corresponding NBS values (ranging 35 – 223 Mg ha⁻¹) and, when applied to the NBS LC map, estimated forest biomass to 1279 Tg while this was only 294 Tg according to the NBS (Table 20).

In addition, the comparison of the LC maps revealed large differences among the datasets. Specifically, the GLC2000 mapped larger areas of forest and shrubland and smaller extents of grassland and agriculture in comparison to the NBS (Table 20). Therefore, by using the IPCC Tier 1 values in combination with the GLC2000 map (as in the Reusch & Gibbs map), the higher biomass reference values for forest and shrubland were applied to larger area coverage of these two classes, causing the strong overestimation of AGB stock found in the Reusch & Gibbs map.

Table 20: Comparison of biomass and area statistics for the main LC classes of Uganda. For each LC class, the comparison is performed between the NBS and the IPCC Tier 1 biomass reference values (left column), between the total AGB estimates of Uganda obtained by applying the NBS and IPCC reference values to the NBS LC map (central column), and between the area mapped by the NBS LC and the GLC2000 maps (right column)

LC class	AGB Ref. values (Mg/ha)		Total AGB (Tg)		Area (Km ² x 1000)	
	NBS Plots	IPCC Tier1	NBS Plots + NBS LC	IPCC Tier1 + NBS LC	NBS LC	GLC2000
Forest	35 - 223	115 - 310	294	1279	49.3	67.3
Shrubland	14	70	14	98	14.2	49.2
Grassland	8	2 - 6	47	31	51.2	15.8
Agriculture	11	10	112	84	84.7	70.7

5.4. Discussion

The comparison of six biomass and C stock maps with the NBS reference dataset revealed large differences regarding estimates of total AGB of Uganda and its spatial distribution. These differences were related to the biomass reference data (biome average values versus field data), the spatial datasets (global versus national maps or satellite data) and the statistical approach (averaging versus regression models) employed by the maps.

The biomass reference data were responsible for most of the variability in the AGB estimates (Table 18). Maps based on biome average biomass values (i.e. Reusch & Gibbs) and averages from forest inventories acquired in several countries (i.e. Drigo, Gibbs & Brown) agreed less with the reference dataset than maps based on national forest inventory data, strongly overestimating the country's AGB stock. This was due to the fact that biome average values, representative of large areas and broad vegetation types, were applied to areas of Uganda that diverge from the average biome conditions. For example, the IPCC Tier 1 value of 260 Mg ha⁻¹ for AGB in forests in the tropical moist deciduous ecozone was not appropriate to the woodlands in northern Uganda, located in an area much drier and with lower tree density and dimension than in the average ecozone condition, where the NBS reference value was 35 Mg ha⁻¹.

In addition, while biomass density in mature forest and at specific locations can reach higher values than the average NBS values (e.g. Mitchard et al., 2009), applying default AGB values to 1 Km resolution maps is not appropriate in a country like Uganda, which is characterized by a highly fragmented landscape.

The applicability of the biomass reference values also depends on the definition of forest. In the case of the IPCC Tier 1 values, the biomass reference values for forests, ranging from 115 to 310 Mg ha⁻¹, were not always matching the definition of forest, which includes any area with more than 10% tree canopy cover and with tree height exceeding 5 m at maturity. Moreover, while the NBS identifies 3 main forest types (further stratified by biomass stocking class), the IPCC identifies only 1 forest class, which includes the NBS classes of closed forest, open forest and woodland (see Table 13). Therefore, forest areas at the lower end of the canopy cover range (e.g. 10% – 40% cover) are likely to have lower biomass density values than the minimum Tier 1 value (115 Mg ha⁻¹).

The use of biome average values tended to overestimate the total AGB of Uganda but in some cases their application to local level underestimated AGB density (Figure 32). Therefore, in terms of C assessment the use of average values would cause underestimation of emissions from deforestation occurring in forests with more than average biomass density (Houghton, 2005; Houghton and Hackler, 2006).

Regarding the spatial datasets, maps based on global LC datasets agreed less with the reference data than maps based on national LC or satellite data. Global datasets represent the distribution of the main LC types at regional level using a limited number of classes and, as a consequence of the small thematic detail, the variability of a biophysical parameter as biomass within the LC classes is usually large. For example, in the GLC2000 dataset all areas dominated by trees (>40% cover) are classified as forest, but it can be expected that even after their stratification by ecological regions the biomass density within a forest type varies considerably according to tree size, tree density and specific conditions related to local climate and land use history. Instead, national LC maps identified different forest types that allowed reducing the variability in the AGB strata. Satellite data were also able to identify strata with low AGB variability and, in comparison with LC maps, presented the advantage of higher spatial resolution, higher thematic content (i.e. continuous values) and were not affected by ambiguities in class definition or propagation of LC errors in the biomass estimates.

Regarding the statistical approach, biomass estimates derived from averaging methods (CA approach) agreed less with the reference data than predictions based on regression models (DR approach). While this result was clearly affected by the fact that the Avitabile and the Baccini maps (DR approach) were not independent from the reference datasets, similar conclusions were reached by Goetz et al. (2009) on the basis of independent comparison data. Using LiDAR metrics closely related to AGB density, Goetz et al. showed that in central Africa the Baccini map (DR approach) had a narrower range of variability of LiDAR values within each AGB class and hence a smaller uncertainty for any AGB estimate than the Gibbs & Brown map (CA approach).

It is important to note that these results are due to the input data employed by the CA maps (i.e. biome average values and coarse spatial maps) more than to the approach itself. Ultimately, the quality of a biomass map depends on the input data used to generate it. However, the factors reducing the map accuracy were strictly associated with the approach: for instance, the DR approach could hardly be applied using few average biomass values, and it employs satellite data that are certain to have higher spatial and thematic resolutions than any derived map. In addition, since all the map units belonging to the same class receive the same biomass value, the averaging approach cannot explain the intra-class variability, which can be large when the strata do not accurately reflect the biomass distribution. Instead, maps based on regression models provide continuous estimates that can describe the full range of biomass variability. Moreover, while the categorical data (such as LC) that are used to identify the strata can only represent the dominant class within a certain unit (unless there is no dominant class and the unit is defined as a “mosaic”), the continuous nature of remotely sensed surface reflectance accounts also for the minor components, but this capability is limited by the complexity of retrieving biomass from a mixed signal.

Similarly, the capability of providing continuous estimates is limited when the magnitude of the error for continuous values is such that AGB estimates must be aggregated in classes to achieve satisfactorily accuracies.

The comparative analysis also indicated that, while maps based on the CA approach tended to overestimate biomass density, maps based on the DR approach showed the tendency to underestimate biomass density. This effect was mainly a consequence of three factors related to the use of optical data and decision tree models in the DR approach (see section 4.4).

First, the optical signal tends to saturate in closed canopy forests resulting in the tendency to underestimate areas with high biomass densities (i.e. above 150 - 200 Mg ha⁻¹). Second, in order to minimize atmospheric effects (i.e. cloud coverage, haze) satellite images are often acquired during the dry season when deciduous vegetation is usually without leaves and may be confounded with low-vegetation areas. Third, by averaging the data within each terminal node, decision tree models intrinsically underestimate at high values (and overestimate at low values).

As mentioned above, the differences between the approaches tends to reduce when the maps based on the CA approach identify several strata with high spatial and thematic resolution, and when the continuous AGB estimates provided by the DR approach are aggregated in classes to achieve satisfactorily accuracies.

The differences among the maps were also affected by the lack of a common definition of AGB. Some maps (NBS and Avitabile) refer to air-dry biomass while others refer to oven-dry biomass (Henry, Reusch & Gibbs) or do not provide this information (Baccini, Drigo, Gibbs & Brown). Similarly, there are differences regarding the minimum diameter and the inclusion or exclusion of dead trees, non-woody plants (herbaceous plants, lianas) or non-woody components (foliage, seeds). Specifically, the choice of the minimum diameter may have a larger impact in low-

biomass vegetation types, where small trees provide a higher contribution (in relative terms) to the total biomass compared to high-biomass forests. However, it is not obvious that the smaller classes of diameter always represent a larger proportion of the total biomass: in woodlands, for example, grazing and farming activities (if present) may significantly reduce the understorey vegetation and limit the contribution of smaller diameter classes to total biomass. Most importantly, considering that oven-dry biomass is equivalent to about 80% of air-dry biomass (Brown & Gaston, 1997) and that in some ecosystems tree biomass can represent more than 95% of the total AGB (Henry, 2010), the contribution of different AGB definition can only account for a limited portion of the large differences present among the maps.

The comparison of the biomass maps with ground reference data confirmed the above findings but also demonstrated the difficulty to perform a consistent validation of remote sensing products. Compatible maps as well as an accurate, representative and comparable field dataset are required to obtain comparable and reliable estimates of map accuracy, but such datasets are usually not available. In the present analysis the validation results were affected by differences in the format (vector, raster), spatial resolution and biomass definition used by the maps, by the skewed distribution of the field data and, most importantly, by the fact that some datasets were not independent from the testing data.

Lastly, this study demonstrates that the AGB estimates were primarily driven by the biomass reference values while the type of spatial datasets used for their stratification had a smaller, but not negligible, impact. The results highlight the importance of the applicability of biomass reference values to the study area but also indicate that the resolution and accuracy of the spatial data are still critical to obtain reliable AGB estimates at local level, which are necessary for land management or estimating emissions from deforestation at specific locations.

6. Conclusions

6.1. Optical Remote Sensing for biomass retrieval

Accurate, updated and detailed maps of AGB distribution in tropical countries are urgently required for several applications, as REDD+ implementation activities, global C modeling and national resource planning. This study responds to the research questions raised on Chapter 4 and shows that remote sensing data can provide an effective way to spatialize field biomass measurements, and that optical data at 30-m resolution (Landsat) in association with a regression tree model can estimate AGB at national level with good accuracy and high spatial resolution.

The freely accessible Landsat archive allowed the compilation of a nationally consistent satellite dataset for the period circa-2000 but did not provide sufficient cloud-free images necessary for compiling consistent multi-date datasets (i.e. intra-annual datasets), necessary for a proper discrimination of seasonal or deciduous formations. Land cover data provided information related to vegetation phenology and their integration with satellite data moderately improved model results, but in certain cases affected negatively the predictions. For this reason, the use of ancillary information (as land cover) is recommended only when their accuracy, spatial detail and temporal resolution are comparable with those of remote sensing data.

Landsat data, and particularly SWIR bands, showed sensitivity to canopy water content and shadow fraction, which were associated with biomass density and allowed to identify a relation between spectral response and variation on biomass density, up to a saturation point identified around 150 – 200 Mg ha⁻¹. Since the relationship between spectral data and biomass was non-linear, non-parametric regression models represented the most appropriate modeling approaches, and satisfactory results were obtained employing the Random Forest model in combination with carefully selected training data.

In order to obtain reliable predictions, this approach required consistent satellite data, accurate field data for each biomass-related spectral class and appropriate ancillary information (e.g. land cover).

In the case of Uganda, the accuracy of the biomass predictions tended to be reduced by the saturation of the optical signal in closed canopy forests, model under-prediction at high biomass values and over-prediction at low values, availability of cloud-free images only during the dry season and missing predictions for areas affected by cloud cover.

For these reasons, the presented approach resulted to be particularly suited for countries with limited accessibility located in tropical and subtropical dry and semi-dry areas, where biomass density is usually below the saturation threshold of optical sensors and cloud cover is limited during the dry season. In such contexts, this methodology presents the capability to map spatially-detailed continuous biomass densities that, in comparison with conventional approaches mainly based on land cover classes, allow for more accurate estimates of carbon stocks and emissions from deforestation events, especially for changes occurring in forests with biomass densities substantially different from the average class values.

The innovative aspects of this study were mainly related to the following aspects:

- Relevance of the study area: while other studies have employed Landsat-type data for biomass mapping in various regions of the world, only few examples are available for

tropical African countries, nonetheless this region has the highest uncertainty regarding biomass stocks, and reducing such uncertainties is recognized as key for global carbon modeling (Houghton, 2005).

- The application at national level is critical since this is the scale required by IPCC for REDD+ monitoring, reporting and verification. Specifically, issues that are not present in smaller-scale studies can arise at national level because countries often include different ecological regions and very diverse vegetation types in their national territories. In this context, Uganda is a very interesting case-study because it includes several distinct vegetation types, from dry lowland savannah to dense humid mountain forests. Therefore, the relevance of the study also consists on investigating the capabilities and critically evaluating the limitations of a remote sensing dataset (Landsat) for mapping biomass consistently over different ecological vegetation types.
- The application to low-biomass areas and non-forest vegetation types is innovative, since most of existing studies are typically focused on forest areas. This is especially important in a country as Uganda, which is mainly covered by agriculture, savannah and woodland that store substantial amounts of biomass since their low biomass density is counterbalanced by coverage over large areas.
- The careful analyses and comparison of the performance of different models (Random Forest, single regression tree, linear regression) and input variables (land cover, texture, topographic metrics) strengthen the understanding of their capabilities and create consensus on the most appropriate modeling approach for biomass estimation using remote sensing data.

6.2. Comparison of biomass maps of Uganda

In order to better understand the reliability of existing biomass and C stock products, six maps (including the biomass map produced in this thesis) were compared with a reference dataset for the case study of Uganda.

The comparison revealed very large differences among the datasets and indicated that maps employing the CA approach in combination with biome average values (as the IPCC Tier 1 values) and global LC datasets strongly overestimated the biomass stock of Uganda. Instead, more reliable estimates were obtained using country-specific field data (i.e. forest inventories) in combination with satellite data and/or national LC and ancillary information.

Maps based on satellite data were able to provide continuous and spatially detailed biomass estimates. These maps tended towards conservative estimates mainly as a consequence of the processing techniques and the saturation of the satellite signal at high biomass values.

The comparison with ground reference data confirmed these findings but the validation results were not entirely comparable as they were affected by several factors.

The larger impact of biomass reference data than spatial maps on AGB estimates indicated that the first critical step to improve the accuracy of the biomass maps consists of the collection of accurate biomass field data for all relevant vegetation types. The collection of sufficient reference data is a critical issue that cannot be substituted with larger amount of remote sensing observations. However, detailed and accurate spatial datasets are crucial to obtain accurate estimates at specific locations and to correctly quantify emissions from deforestation, as required for REDD+ actions. The acquisition of field datasets comparable with the remote sensing

products as well as effective upscaling procedures are also necessary for the proper training and validation of satellite-based maps.

The innovative aspects of this study were mainly related to the following aspects:

- The spatially-explicit comparison of a number of biomass/carbon maps using a set of spatial statistics: several maps based on different data sources and approaches were recently published but, while previous studies noted the variability in the overall estimates at national levels (Gibbs et al., 2007), their spatial agreement was not evaluated before.
- The validation of a set of biomass maps at national scale using an unprecedented field dataset: the applicability of regional maps to national scales is uncertain because their proper validation is limited by the scarcity of field datasets with comparable coverage. The NBS field dataset allowed the consistent and extensive validation of the maps for the area of Uganda.
- The main sources of variability in the estimates were identified on the basis of specific simulations, which allowed quantifying separately the impact of biomass reference data and spatial datasets on the biomass estimates.

6.3. Future research

The need for better estimates of biomass stocks of tropical forests as well as the interest on remote sensing techniques capable to retrieve this parameter have increased tremendously during the last decade, stimulating a rapid development of this sector and the planning of dedicated satellite missions. With regard to the emerging topic of REDD+, the future research will have to respond to the challenges of designing national-scale systems capable to monitor, report and verify human-induced emissions in tropical forests in an accurate, consistent and transparent way.

While the present study has demonstrated the capability of optical satellite data for detailed, consistent and accurate estimation of biomass density at national level in the tropics, it has also highlighted the limitations of this approach, as the saturation of the optical signal at high biomass values, the impact of clouds and atmospheric contamination on data availability and data analysis, the effect of phenological changes on biomass estimations, and the limited transferability of the model to datasets belonging to different temporal or spatial domains. Such limitations indicate existing gaps that need to be addressed by future research in order to produce accurate biomass estimates over large areas at high spatial and temporal resolutions. Similarly, the comparison of different biomass/carbon maps and field reference datasets has stressed the large inconsistencies among existing products and indicated the need to identify the most appropriate methodology and input data according to the context and the purpose of the analysis. In addition, while most of the focus is currently on quantifying existing biomass stocks, new approaches will be required to better understand the biomass dynamics, and it can be expected that the quantification of the biomass changes and related carbon emissions and removals due to forest loss, forest degradation and forest regrowth will become a major research topic in the coming years.

By providing a key source of data, remote sensing will play a growing role for biomass assessment. The availability of remote sensing data is expected to increase considerably in the near future with a number of new satellites scheduled to be launched in the coming years. These satellites will carry onboard both passive (optical) and active (radar, lidar) sensors and will acquire data with unprecedented spatial and temporal resolutions. Such wealth of data, including polarimetric repeat pass radar data, hyperspectral optical data and fullwave laser data, will allow for a more detailed and accurate retrieval of biophysical parameters of vegetation as biomass density, but will require advanced data processing techniques and appropriate modeling approaches. Specifically, since no single spaceborne instrument can directly measure biomass, the fusion of different data types is considered the most promising approach because such integration may overcome the limitations of the individual datasets. However, effective integration models and technologies have not been sufficiently developed yet and will likely represent a key research area in the coming years. Moreover, it will be of primary importance to develop robust methods applicable in different ecological contexts in order to overcome the limitations of one-time one-place methodologies and provide estimates that are consistent through time and space.

Last (but not least), the acquisition of reliable reference datasets compatible with satellite-based products and their up-scaling to the resolution of the remote sensing datasets will be a key issue for the proper training and validation of spaceborne datasets. With new remote sensing based maps of forest parameters already produced over large areas, such as the global forest canopy height map (Lefsky, 2010) or the pan-tropical forest carbon stock map (Saatchi et al., 2011), methods capable to acquire reference datasets reliably and efficiently as well as appropriate up-scaling and comparison techniques will be crucial to better understand the capabilities and limitations of such datasets, to develop better remote sensing products and ultimately obtain robust biomass estimates.

References

- Achard, F., Eva, H.D., Stibig, H.J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J.P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*, 297, 999-1002
- Achard, F., Eva, H.D., Mayaux, P., Stibig, H.J., & Belward, A. (2004). Improved estimates of net carbon emissions from land cover change in the tropics for the 1990s. *Global Biogeochemical Cycles*, 18
- Arino, O., Bicheron, P., Achard, F., Latham, J., Witt, R., & Weber, J.L. (2008). GLOBCOVER The most detailed portrait of Earth. *Esa Bulletin-European Space Agency*, 24-31
- Asner, G.P., Keller, M., Pereira, R., & Zweede, J.C. (2002). Remote sensing of selective logging in Amazonia - Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. *Remote Sensing of Environment*, 80, 483-496
- Asner, G.P. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4, 034009
- Asner, G.P., Powell, G.V.N., Mascaro, J., Knapp, D.E., Clark, J.K., Jacobson, J., Kennedy-Bowdoin, T., Balaji, A., Paez-Acosta, G., Victoria, E., Secada, L., Valqui, M., & Hughes, R.F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 16738-16742
- Austin, J.M., Mackey, B.G., & Van Niel, K.P. (2003). Estimating forest biomass using satellite radar: an exploratory study in a temperate Australian Eucalyptus forest. *Forest Ecology and Management*, 176, 575-583
- Baccini, A., Friedl, M. A., Woodcock, C. E., & Warbington, R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, 31, 1-4.
- Baccini, A., Friedl, M. A., Woodcock, C. E., & Zhu, Z. (2007). Scaling field data to calibrate and validate moderate spatial resolution remote sensing models. *Photogrammetric Engineering and Remote Sensing*, 73(8), 945-954.
- Baccini, A., Laporte, N., Goetz, S.J., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, 3, 045011
- Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., et al. (2008). Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, 112, 1658-1677.
- Bombelli, A., Henry, M., Castaldi, S., Adu-Bredu, S., Arneeth, A., Grandcourt, A., et al. (2009). An outlook on the Sub-Saharan Africa carbon balance. *Biogeosciences*, 6, 2193-2205.

- Boudreau, J., Nelson, R., Margolis, H., Beaudoin, A., Guindon, L., & Kimes, D. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112, 3876-3890
- Boyd, D. S., & Danson, F. M. (2005). Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography*, 29(1), 1-26.
- Breiman, L., Friedman, J., Olshen, R. & Stone, J. (1984). *Classification and Regression Trees*. Wadsworth, California
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–23.
- Brown S. (1997). Estimating biomass and biomass change of tropical forests: a primer. In *FAO Forestry Paper*. Rome
- Brown, S., & Gaston, G. (1997). Tropical Africa: Land Use, Biomass, and Carbon Estimates for 1980. In Oak Ridge, Tennessee: Carbon Dioxide Information Center, Oak Ridge National Laboratory
- Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution*, 116, 363-372
- Bryant, D., Nielsen, D., & Tangle, L. (1997). The Last Frontier Forests: Ecosystems and Economies on the Edge. p. 42. Washington DC: World Resources Institute
- Canadell, J. G., Raupach, M. R., & Houghton, R. A. (2009). Anthropogenic CO₂ emissions in Africa. *Biogeosciences*, 6, 463-468.
- Card, D.H. (1982). Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy. *Photogrammetric Engineering and Remote Sensing*, 48, 431-439
- Center for International Earth Science Information Network (CIESIN), C.U., & CIAT, C.I.d.A.T. (2005). Gridded Population of the World Version 3 (GPWv3): Population Density Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University
- Chambers, J.Q., Santos, J.d., Ribeiro, R.J., & Higuchi, N. (2001). Tree damage, allometric relationships, and above-ground net primary production in central Amazon forest. *Forest Ecology and Management*, 152, 73-84
- Chave, J., Eacute, Ocirc, Me, Riéra, B., & Dubois, M.A. (2001). Estimation of biomass in a neotropical forest of French Guiana: spatial and temporal variability. *Journal of Tropical Ecology*, 17, 79-96
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., et al. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society*, 359(1443), 409-420.
- Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Folster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.P., Nelson, B.W., Ogawa, H., Puig, H., Riera, B., &

- Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, *145*, 87-99
- Chave, J., Muller-Landau, H.C., Baker, T.R., Easdale, T.A., Ter Steege, H., & Webb, C.O. (2006). Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecological Applications*, *16*, 2356-2367
- Chavez Jr., P. S. (1996). Image-Based Atmospheric Corrections - Revisited and Improved. *Photogrammetric Engineering & Remote Sensing*, *62*(9), 1025-1036.
- Ciais, P., Piao, S.L., Cadule, P., Friedlingstein, P., & Chedin, A. (2009). Variability and recent trends in the African terrestrial carbon balance. *Biogeosciences*, *6*, 1935-1948
- Cihlar, J. (2000). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing*, *21*, 1093-1114
- Cohen, W.B., & Goward, S.N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*, 535-545
- Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*, 35-46
- Crist E. P., & Cicone R. C. (1984). Application of the Tasseled Cap concept to simulated thematic mapper data. *Photogrammetric Engineering and Remote Sensing*, *50*(3), 343-352.
- De'ath, G., & Fabricius, K. E. (2000). Classification and Regression Trees: a Powerful Yet Simple Technique for Ecological Data Analysis. *Ecology*, *81*(11), 3178-3192.
- Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Niño, F., Schouten, L., et al. (2006). GLOBCOVER: a 300 m global land cover product for 2005 using Envisat MERIS time series. In, *Proceedings of the ISPRS Commission VII Symposium: Remote Sensing from Pixels to Processes*. Enschede, The Netherlands
- DeFries, R.S., Houghton, R.A., Hansen, M.C., Field, C.B., Skole, D., & Townshend, J. (2002). Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 14256-14261
- DeFries, R., Achard, F., Brown, S. L., Herold, M., Murdiyarso, D., Schlamadinger, B., et al. (2007). Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental Science & Policy*, *10*, 385-394.
- Descloux, S., Chanudet, V., Poilve, H., & Gregoire, A. (2011). Co-assessment of biomass and soil organic carbon stocks in a future reservoir area located in Southeast Asia. *Environmental Monitoring and Assessment*, *173*, 723-741
- Dong, J., Kaufmann, R.K., Myneni, R.B., Tucker, C.J., Kauppi, P.E., Liski, J., Buermann, W., Alexeyev, V., & Hughes, M.K. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sensing of Environment*, *84*, 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., & Prince, S.D. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, 79, 305-319
- Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing factors affecting the generality of relationships. *Global Ecology and Biogeography*, 12, 147-159
- Drichi, P. (2003). National Biomass Study, Technical Report. Forestry Department, Ministry of Water, Lands & Environment. PO Box 1613, Kampala, Uganda
- Drigo, R. (2006). WISDOM – East Africa. Woodfuel Integrated Supply/Demand Overview Mapping (WISDOM) Methodology. Spatial woodfuel production and consumption analysis of selected African countries. In, *Wood Energy Working Paper*: FAO Forestry Department
- FAO (2001a). Global ecological zoning for the global forest resources assessment 2000. In, *FAO FRA Working Paper*. Rome, Italy
- FAO (2001b). Global forest resources assessment 2000. In, *FAO Forestry Paper*. Rome, Italy
- FAO (2003). Forestry Sector Outlook Studies for Africa. Subregional Report East Africa. <ftp://ftp.fao.org/docrep/fao/005/Y8693E/Y8693E00.pdf>
- FAO (2005). Global Forest Resources Assessment 2005, Country Reports. Uganda. FRA/2005/223. <ftp://ftp.fao.org/docrep/fao/010/ai985E/ai985E00.pdf>
- FAO (2006). Global forest resources assessment 2005. *FAO Forestry Paper 147*. Rome, Italy.
- Fazakas, Z., Nilsson, M., & Olsson, H. (1999). Regional forest biomass and wood volume estimation using satellite data and ancillary data. *Agricultural and Forest Meteorology*, 98-99, 417–425.
- Field, R.D., & Shen, S.S.P. (2008). Predictability of carbon emissions from biomass burning in Indonesia from 1997 to 2006. *Journal of Geophysical Research-Biogeosciences*, 113
- Foody, G. M., Cutler, M. E., McMorrow, J., Pelz, D., Tangki, H., Boyd, D. S., et al. (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global Ecology and Biogeography*, 10, 379 - 387.
- Foody, G.M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185-201
- Foody, G. M., Boyd, D. S., & Cutler, M. E. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, 85, 463–474.
- Foody, G.M. (2010). Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment*, 114, 2271-2285

Forest Department Uganda (1997). Forest management plan for Budongo Forest Reserve for 1997 to 2007. 107 pp. Forest Department, Kampala, Uganda

Fransson, J.E.S., Smith, G., Askne, J., & Olsson, H. (2001). Stem volume estimation in boreal forests using ERS-1/2 coherence and SPOT XS optical data. *International Journal of Remote Sensing*, 22, 2777-2791

Freund, Y., & Shapire, R. (1996). Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148–156. San Francisco: IEEE.

Gaston, G., Brown, S., Lorenzini, M., & Singh, K.D. (1998). State and change in carbon pools in the forests of tropical Africa. *Global Change Biology*, 4, 97-114

Gates, D.M., Keegan, H.J., Schleter, J.C., & Weidner, V.R. (1965). Spectral Properties of Plants. *Appl. Opt.*, 4, 11-20

Gausman, H.W. (1974). Leaf Reflectance of near-Infrared. *Photogrammetric Engineering and Remote Sensing*, 40, 183-191

Gemmell, F. M. (1995). Effects of forest cover, terrain, and scale on timber volume estimation with Thematic Mapper data in a Rocky Mountain site. *Remote Sensing of Environment*, 51, 291–305

Gibbs, H.K. (2006). Olson's Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: An Updated Database Using the GLC2000 Land Cover Product. In. Oak Ridge, Tennessee: Carbon Dioxide Information Center, Oak Ridge National Laboratory

Gibbs, H.K., & Brown, S. (2007). Geographical Distribution of Woody Biomass Carbon in Tropical Africa: An Updated Database for 2000. In. Oak Ridge, Tennessee: Carbon Dioxide Information Center, Oak Ridge National Laboratory

Gibbs, H. K., Brown, S. L., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2, 1-13

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W. S., Kellndorfer, J. M., et al. (2009). Mapping and monitoring carbon stocks with satellite observations : a comparison of methods. *Carbon Balance and Management*, 4(2), 1-7

Goetz, S.J., Sun, M., Baccini, A., & Beck, P.S.A. (2010). Synergistic use of spaceborne lidar and optical imagery for assessing forest disturbance: An Alaska case study. *Journal of Geophysical Research*, 115

GOFC-GOLD (2010). A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals caused by deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. *GOFC-GOLD Report version COP16-1*, GOFC-GOLD Project Office, Natural Resources Canada, Alberta, Canada

Gotelli, N.J. & Ellison, A.M. (2004). A primer of ecological statistics. Sinauer. Sunderland, MA.

- Goward, S. N., Arvidson, T., Williams, D. L., Faundeen, J., Irons, J., Franks, S., et al. (2006). Historical Record of Landsat Global Coverage: Mission Operations, NSLRSDA, and International Cooperator Stations. *Photogrammetric Engineering & Remote Sensing*, 72(10), 1155-1169
- Gullison, R. E., Frumhoff, P. C., Canadell, J. G., Field, C. B., Nepstad, D. C., Hayhoe, K., et al. (2007). Tropical Forests and Climate Policy. *Science*, 316, 985-986
- Hagen-Zanker, A., Straatman, B., & Uljee, I. (2005). Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, 19(7), 769-785
- Hagen-Zanker, A. (2006). Comparing continuous valued raster data: A cross disciplinary literature scan. In: Maastricht: Research Institute for Knowledge Systems
- Hagen-Zanker, A. (2009). An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. *International Journal of Geographical Information Science*, 23(1), 61-73
- Hall, F. G., Strebel, D. E., Nickeson, J. E., & Goetz, S. J. (1991). Radiometric Rectification: Toward a Common Radiometric Response Among Multidate, Multisensor Images. *Remote Sensing of Environment*, 35, 11-27
- Hall, R. J., Skakun, R. S., Arsenault, E. J., & Case, B. S. (2006). Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, 225, 378-390
- Hame, T., Salli, A., Andersson, K. & Lohi, A. (1997). A new methodology for the estimation of biomass of conifer-dominated boreal forest using NOAA AVHRR data. *International Journal of Remote Sensing*, 18, 3211–3243
- Hamilton, A. (1984). Deforestation in Uganda. Oxford University Press, Nairobi
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C., & Sohlberg, R.A. (2003). Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. *Earth Interactions*, 7
- Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., Altstatt, A., et al. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*, 112, 2495–2513
- Harrell, P.A., Kasischke, E.S., Bourgeau-Chavez, L.L., Haney, E.M. & Christensen, N.L. Jr (1997). Evaluation of approaches to estimating aboveground biomass in southern pine forests using SIR-C data. *Remote Sensing of Environment*, 59, 223–233
- Healey, S. P., Yang, Z., Cohen, W. B., & Pierce, D. J. (2006). Application of two regression based methods to estimate the effects of partial harvest on forest structure using Landsat data. *Remote Sensing of Environment*, 101, 115–126
- Henry, M. (2010). Carbon stocks and dynamics in Sub Saharan Africa. In (p. 421): University of Tuscia, AgroParisTech/ENGREF

- Herold, M., & Johns, T. (2007). Linking requirements with capabilities for deforestation monitoring in the context of the UNFCCC-REDD process. *Environmental Research Letters*, 2, 1-7.
- Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112, 2538-2556
- Herold, M., & Skutsch, M. (2011). Monitoring, reporting and verification for national REDD+ programmes: two proposals. *Environmental Research Letters*, 6
- Houghton, R.A. (1999). The annual net flux of carbon to the atmosphere from changes in land use 1850-1990. *Tellus Series B-Chemical and Physical Meteorology*, 51, 298-313
- Houghton, R. A., Lawrence, K. T., Hackler, J. L., & Brown, S. (2001). The spatial distribution of forest biomass in the Brazilian Amazon: A comparison of estimates. *Global Change Biology*, 7, 731-746
- Houghton, R. A. (2005). Aboveground forest biomass and the global carbon balance. *Global Change Biology*, 11, 945-958
- Houghton, R. A., and Hackler J. L. (2006). Emissions of carbon from land use change in sub-Saharan Africa. *Journal of Geophysical Research*, 111, G02003
- Houghton, R. A. (2007). Balancing the Global Carbon Budget. *Annual Review of Earth and Planetary Sciences*, 35, 313-347
- Houghton, R.A., Butman, D., Bunn, A.G., Krankina, O.N., Schlesinger, P., & Stone, T.A. (2007). Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters*, 2
- Howard, P.C. (1991). Nature Conservation in Uganda's Tropical Forest Reserves. IUCN, Gland, Switzerland
- Howard, P.C., Davenport, T., Kigenyi, F. (1997). Planning conservation areas in Uganda's natural forests. *Oryx* 31, 253-264
- Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82, 397-416
- Kanabahita, C. (2001). Forestry Sector Outlook Studies. Country Report for Uganda. Kampala, Forestry Department, Ministry of Water, Lands & Environment. <http://www.fao.org/DOCREP/004/AC427E/AC427E00.htm>
- Kasischke, E.S., Melack, J.M., & Craig Dobson, M. (1997). The use of imaging radars for ecological applications--A review. *Remote Sensing of Environment*, 59, 141-156
- Kauth R. J., & Thomas G. S. (1976). The Tasselled Cap – A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat. *Symposium on Machine Processing of Remotely Sensed Data*, Purdue University, West Lafayette, IN, 4b41 - 4b51

- Kindermann, G.E., McAllum, I., Fritz, S., & Obersteiner, M. (2008). A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*, 42, 387-396
- 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, 581-590
- Kohler, P., & Huth, A. (2010). Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests. *Biogeosciences*, 7, 2531-2543
- Inoue, Y., Kiyono, Y., Asai, H., Ochiai, Y., Qi, J.G., Olioso, A., Shiraiwa, T., Horie, T., Saito, K., & Dounagsavanh, L. (2010). Assessing land-use and carbon stock in slash-and-burn ecosystems in tropical mountain of Laos based on time-series satellite images. *International Journal of Applied Earth Observation and Geoinformation*, 12, 287-297
- IPCC (1996). Climate change 1995: the science of climate change. Cambridge University Press, Cambridge, UK
- IPCC (2000). Land Use, Land Use Change and Forestry. Special Report, Inter-Governmental Panel on Climate Change. Cambridge University Press, Cambridge, UK
- IPCC (2006). IPCC Guidelines for National Greenhouse Gas Inventories. Ed. Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T. and K. Tanabe. Prepared by the National Greenhouse Gas Inventories Programme, IGES, Japan
- IPCC (2007). IPCC Fourth Assessment Report: Climate Change 2007 (AR4). Cambridge, United Kingdom and New York, NY, USA. Cambridge University Press
- Labrecque, S., Fournier, R. A., Luther, J. E., & 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, 129-144
- Langoya, C.D., Long, C., Grundy, I., Le Breton, G., Mejia, R., Benitez, R. (1998). Local communities and ecotourism development in Budongo Forest Reserve, Uganda. ODI, London
- Larsson, H. (1993). Linear regression for canopy cover estimation in Acacia woodlands using Landsat-TM, -MSS, and SPOT HRV XS data. *International Journal of Remote Sensing*, 14, 2129-2136
- Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, 2, 831-836
- Le Toan, T., Ulander, L., Dubois-Fernandez, P., Papathanassiou, K., Rocca, F., & Davidson, M. (2010). Retrieval of forest biomass from P-band SAR data: Prospects for the future BIOMASS mission. In, *In Proceedings ESA Living Planet Symposium 2010*. Bergen, Norway
- Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., & Shugart, H.H. (1999). Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. *Remote Sensing of Environment*, 67, 83 - 98

- Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., & Gower, S.T. (2002). Lidar remote sensing of aboveground biomass in three biomes. *Global Ecology and Biogeography*, *11*, 393-399
- Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., & de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, *32*
- Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo, F.D., Hunter, M.O., de Oliveira, R., & de Camargo, P.B. (2006). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, *33* L05501
- Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, *1*
- Lefsky, M.A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, *37*
- Li, X., & Strahler, A.H. (1992). Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *Geoscience and Remote Sensing, IEEE Transactions on*, *30*, 276-292
- Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18-22
- Liaw, A. (2009). Package “randomForest”, Breiman and Cutler’s random forests for classification and regression, v.4.5-30. <http://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Lillesand, T., Kiefer, R., & Chipman J. (2008). Remote Sensing and Image Interpretation, 6th edition. John Wiley & Sons, NY
- Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, *198*, 149-167
- Lu, D. (2005). Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *International Journal of Remote Sensing*, *26*(12), 2509–2525
- Lu, D., Batistella, M., Moran, E. (2005). Satellite Estimation of Aboveground Biomass and Impacts of Forest Stand Structure. *Photogrammetric Engineering & Remote Sensing*, *71*(8), 967-974
- Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, *27*(7), 1297-1328
- Malhi, Y., Baker, T.R., Phillips, O.L., Almeida, S., Alvarez, E., Arroyo, L., et al. (2004). The above-ground coarse wood productivity of 104 Neotropical forest plots. *Global Change Biology*, *10*, 563-591

- Malhi, Y. (2010). The carbon balance of tropical forest regions, 1990-2005. *Current Opinion in Environmental Sustainability*, 2, 237-244
- Masera, O., Ghilardi, A., Drigo, R., & Trossero, M.A. (2006). WISDOM: A GIS-based supply demand mapping tool for woodfuel management. *Biomass & Bioenergy*, 30, 618-637
- Mayaux, P., Bartholome, E., Fritz, S., & Belward, A. (2004). A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31, 861-877
- Mäkelä, H., and Pekkarinen, A. (2004). Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management*, 196, 245-255
- Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-Being: biodiversity Synthesis. World Resources Institute, Washington
- Mitchard, E.T.A., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N.S., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T.R., & Meir, P. (2009). Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, 36
- Muukkonen, P., & Heiskanen, J. (2005). Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment*, 99, 434 - 447
- Muukkonen, P., & Heiskanen, J. (2007). Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories. *Remote Sensing of Environment*, 107, 617-624
- Mäkelä, H., and Pekkarinen, A. (2004). Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management*, 196, 245-255
- NASA (2009). Landsat 7 Science Data Users Handbook. Last update August 4, 2009. http://landsathandbook.gsfc.nasa.gov/handbook/handbook_toc.html
- Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, 24, 247-267
- Nelson, R., Oderwald, R. & Gregoire, T.G. (1997). Separating the ground and airborne laser sampling phases to estimate tropical forest basal area, volume, and biomass. *Remote Sensing of Environment*, 60, 311–326
- Nelson, R., Kimes, D. S., Salas, W. A., & Routhier, M. (2000). Secondary Forest Age and Tropical Forest Biomass Estimation Using Thematic Mapper Imagery. *BioScience*, 50(5), 419-431
- Nelson, R., Ranson, K., Sun, G., Kimes, D., Kharuk, V., & Montesano, P. (2009). Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of Environment*, 113, 691-701
- Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C.E., Schaaf, C., Yao, T., Ranson, K.J., Sun, G., & Blair, J.B. (2010). Assessing general relationships between aboveground biomass and

vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research*, 115

Olson, C.E. Jr. & Good, R.W. (1962). Seasonal changes in light reflectance from forest vegetation, *Photogrammetric Engineering*, 28 (1), 107-114

Olson, J.S., Watts, J.A., & Allison, L.J. (1983). Carbon in live vegetation of major world ecosystems. In. Oak Ridge, Tennessee: Carbon Dioxide Information Center, Oak Ridge National Laboratory

Olson, J.S., Watts, J.A., & Allison, L.J. (1985). Major world ecosystem complexes ranked by carbon in live vegetation: A Database. In. Oak Ridge, Tennessee: Carbon Dioxide Information Center, Oak Ridge National Laboratory

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30: 683-691

Penman, J., et al. (2003). Good practice guidance for land use, land-use change and forestry. IPCC National Greenhouse Gas Inventories Programme and Institute for Global Environmental Strategies (Kanagawa, Japan). <http://www.ipcc-nggip.iges.or.jp/public/gpplulucf/gpplulucf.htm>

Phua, M. H., and Saito, H. (2003). Estimation of biomass of a mountainous tropical forest using Landsat TM data. *Canadian journal of remote sensing*, 29(4), 429-440

Plumptre, A.J. (1996). Changes following sixty years of selective timber harvesting in the Budongo Forest reserve, Uganda. *Forest Ecology and Management*, 89, 101-113

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., et al. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, 114, 1053-1068

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>

Ranson, K.J., Nelson, R., Kimes, D., Kharuk, V., Sun, G., & Montesano, P. (2007). Using MODIS and GLAS data to develop timber volume estimates in Central Siberia. *Igarss: 2007 IEEE International Geoscience and Remote Sensing Symposium, Vols 1-12*, 2306-2309

Ribeiro, N.S., Saatchi, S.S., Shugart, H.H., & Washington-Allen, R.A. (2008). Aboveground biomass and Leaf Area Index (LAI) mapping for Niassa Reserve, northern Mozambique. *Journal of Geophysical Research-Biogeosciences*, 113

Richards, J.A. (1993). *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, p. 240

Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., & Dobson, C. (2003). A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy*, 6, 441-455

- Rouse, J.W., Haas, R.H., Schell, J.A., Deering D.W. & Harlan J.C. (1974). Monitoring the vernal advancement of retrogradation of natural vegetation. *NASA/GSFC, Type III, Final Report (1974)*, 371 p. Greenbelt, MD
- Roy, P.S. & Ravan, S.A. (1996) Biomass estimation using satellite remote sensing data—an investigation on possible approaches for natural forest. *Journal of Bioscience*, 21, 535–561
- Ruesch, A., & Gibbs, H.K. (2008). New IPCC Tier-1 Global Biomass Carbon Map For the Year 2000. In. Oak Ridge, Tennessee: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory
- Ryherd, S., & Woodcock, C. E. (1996). Combining Spectral and Texture Data in the Segmentation of Remotely Sensed Images. *Photogrammetric Engineering & Remote Sensing*, 62(2), 181-194
- Saatchi, S. S., Houghton, R. A., Dos Santos Alvalá, R. C., Soares, J. V., & Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13, 816-837
- Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 9899-9904
- Sader, S.A., Waide, R.B., Lawrence, W.T., & Joyce, A.T. (1989). Tropical Forest Biomass and Successional Age Class Relationships to a Vegetation Index Derived from Landsat Tm Data. *Remote Sensing of Environment*, 28, 143
- Schott, J., Salvaggio, C., & Volchok, W. (1988). Radiometric scene normalization using pseudoinvariant features. *Remote Sensing of Environment*, 26, 1-16.
- Schowengerdt, R. A. (2007). Remote Sensing. Models and Methods for Image Preprocessing. Third edition. Academic Press, 560 p.
- Schroeder, T. A., Cohen, W. B., Song, C., Canty, M. J., & Yang, Z. (2006). Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. *Remote Sensing of Environment*, 103, 16 - 26
- Schroeder, T. A., Cohen, W. B., & Yang, Z. (2007). Patterns of forest regrowth following clearcutting in western Oregon as determined from a Landsat time-series. *Forest Ecology and Management*, 243, 259–273
- Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? *Remote Sensing of Environment*, 75, 230-244
- Steininger, M. K. (2000). Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International Journal of Remote Sensing*, 21(6&7), 1139-1157

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., et al. (2007). Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂. *Science*, 316(5832), 1732-1735

Stickler, C. M., Nepstad, D. C., Coe, M. T., McGrath, D. G., Rodrigues, H. O., Walker, W. S., et al. (2009). The potential ecological costs and cobenefits of REDD: a critical review and case study from the Amazon region. *Global Change Biology*, 15(12), 2803-2824

Strahler, A. H., Woodcock, C. E., & Smith, J. A. (1986). The nature of models in remote sensing. *Remote Sensing of Environment*, 20, 121-139

Sun, G., Ranson, K., Kimes, D., Blair, J., & Kovacs, K. (2008). Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of Environment*, 112(1), 107-117

Tomppo, E. O., Nilsson, M., Rosengren, M., Aalto, P., & 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, 156-171

Tomppo, E. O., Gagliano, C., Natale, F. D., Katila, M., & McRoberts, R. E. (2009). Remote Sensing of Environment Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery. *Remote Sensing of Environment*, 113, 500-517

Tucker, C.J. (1980). Remote sensing of leaf water content in the near infrared. *Remote Sensing of Environment*, 10, 23-32

United Nations Environment Programme (2010). The Emissions Gap Report: Are the Copenhagen Accord pledges sufficient to limit global warming to 2 °C or 1.5 °C? A preliminary assessment. <http://www.unep.org/publications/ebooks/emissionsgapreport/>

UNFCCC (2006). Impacts, vulnerability and adaptation to climate change in Africa. In: African Workshop on Adaptation Implementation of Decision 1/CP.10 of the UNFCCC Convention, Accra, Ghana, 21 - 23 September, 2006.

UNFCCC (2007). Report on the Second Workshop on Reducing Emissions from Deforestation in Developing Countries. <http://unfccc.int/resource/docs/2007/sbsta/eng/03.pdf>

UNFCCC (2008). Report on the workshop on methodological issues relating to reducing emissions from deforestation and forest degradation in developing countries. FCCC/SBSTA/2008/11

UNFCCC (2009). Reducing emissions from deforestation in developing countries: approaches to stimulate action. FCCC/SBSTA/2009/19/Add.1

UNFCCC (2010). Outcome of the work of the Ad Hoc Working Group on long-term Cooperative Action under the Convention—policy approaches and positive incentives on issues relating to reducing emissions from deforestation and forest degradation in developing countries; and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries

- van der Werf, G.R., Morton, D.C., DeFries, R.S., Olivier, J.G.J., Kasibhatla, P.S., Jackson, R.B., Collatz, G.J., & Randerson, J.T. (2009). CO2 emissions from forest loss, *2*, 737-738
- Van Vliet, J., Hagen-Zanker, A., Engelen, G., Hurkens, J., Vanhout, R., & Uljee, I. (2009). Map Comparison Kit 3: User Manual. In: Maastricht: Research Institute for Knowledge Systems
- Waggoner, P.E. (2009). Forest Inventories: Discrepancies and Uncertainties "The World's Forests: Design and Implementation of Effective Measurement and Monitoring" In. Washington, DC: Resources for the Future
- Walker, W. S., Kelldorfer, J. M., LaPoint, E., Hoppus, M. L., & Westfall, J. (2007). An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, *109*, 482 - 499
- Williams, D. L., Goward, S. N., & Arvidson, T. (2006). Landsat: Yesterday, Today, and Tomorrow. *Photogrammetric Engineering & Remote Sensing*, *72*(10), 1171-1178
- Woodcock, C.E., Strahler, A.H., & Jupp, D.L.B. (1988a). The use of variograms in remote sensing: I. scene models and simulated images. *Remote Sensing of Environment*, *25*, 323-348
- Woodcock, C.E., Strahler, A.H., & Jupp, D.L.B. (1988b). The use of variograms in remote sensing: II. real digital images. *Remote Sensing of Environment*, *25*, 349-379
- Woodcock, C. E., and Ryherd, S.L. (1989). Generation of texture images using adaptive windows. *Technical Papers, ASPRS/ACSM Annual Convention*, 2-7 April, Baltimore, Maryland, *2*, 11-22
- Woodcock, C.E., and V.J. Harward (1992). Nested-hierarchical scene models and image segmentation. *International Journal of Remote Sensing*, *13*(16), 3167-3187
- Woodcock, C.E., Ozdogan, M., Gutman, G., Janetos, A.C., Justice, C.O., Moran, E.F., Mustard, J.F., Rindfuss, R.R., Skole, D.L., Turner, B.L., & Cochrane, M.A. (2004). Trends in Land Cover Mapping and Monitoring Land Change Science. (pp. 367-377): Springer Netherlands
- Woodcock, C. A., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., et al. (2008). Free access to Landsat imagery. *Science*, *320*(5879), 1011
- Wulder, M. & Franklin, S. eds. (2003). Remote Sensing of Forest Environments: Concepts and Case Studies. Boston: Kluwer Academic
- Zhang, X., & Kondragunta, S. (2006). Estimating forest biomass in the USA using generalized allometric models and MODIS land products. *Geophysical Research Letters*, *33*(9), 1-5
- Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moin, J., et al. (2004). Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*, *93*, 402 - 411
- Zheng, G., Chen, J. M., Tian, Q. J., Ju, W. M., & Xia, X. Q. (2007). Combining remote sensing imagery and forest age inventory for biomass mapping. *Journal of Environmental Management*, *85*, 616-623

List of Figures

Figure 1: The Carbon cycle.....	5
Figure 2: Tree biomass for Diameter at Breast Height (DBH) according to 16 published allometric equations.....	10
Figure 3: Overview of the electromagnetic spectrum.....	12
Figure 4: Flowchart representing methodological and processing steps.....	24
Figure 5: The NBS Land Cover map of Uganda.....	26
Figure 6: Location (left) and Zonation map (right) of Budongo Forest Reserve, Uganda.....	28
Figure 7: Landsat Path and Row for Uganda.....	29
Figure 8: Landsat ETM+ mosaic for Uganda for the year circa-2000.....	30
Figure 9: Location and year of acquisition of the NBS field plots, over imposed the Landsat scene.....	35
Figure 10: Histogram of the available subset of NBS field plots (N= 3510).....	36
Figure 11: Location of discarded NBS field plots.....	38
Figure 12: Location of the field plots (yellow dots) measured in the Budongo Forest Reserve, overimposed a SPOT5 image and a Landsat ETM+ image in the background.....	39
Figure 13: Comparison of the biomass estimates obtained with the NBS allometric equations (x axis) and with four generic allometric equations (y axis), considering all plots.....	41
Figure 14: Distribution of the residuals obtained comparing the NBS values with the Chave (2005) dry values, considering all plots.....	42
Figure 15: Comparison of the biomass estimates obtained with the NBS allometric equations (x axis) and with two generic allometric equations (y axis), using different predictor variables.....	43
Figure 16: Frequency distribution of the biomass density values of the Budongo field plots.....	44
Figure 17: Frequency distribution of selected training data.....	45
Figure 18: Predictions of all input data (Model 1 + 2) on OOB data without (left) and with (right) the predictor land cover.....	48
Figure 19: Variable Importance plots for Model 1 without (left) and with (right) the predictor land cover.....	49
Figure 20: Variable Importance Plot for a model including the additional predictors for the image p172r59.....	50

Figure 21: Model predictions for the training data with uniform distribution per biomass class of (left to right): Random Forest, single regression tree and stepwise multiple linear regression	52
Figure 22: Frequency distribution of the selected training data (N = 2527) and of the map predictions	53
Figure 23: AGB density map of Uganda for the year circa-2000	53
Figure 24: Comparison between land cover polygons (left) and image segments (right) for an heterogeneous area in Western Uganda.	56
Figure 25: Comparison of biomass maps at different resolution for the test area indicated by the black frame in the map of Uganda.	57
Figure 26: Example of biomass underestimation using only Landsat data as predictor for the test area indicated by the black frame in the map of Uganda.	58
Figure 27: Example of biomass overestimation using LC as predictor for the test area indicated by the black frame in the map of Uganda.	59
Figure 28: Comparison of the estimates of total AGB stock of Uganda	68
Figure 29: Histograms of the biomass maps and NBS field plots.	68
Figure 30: Comparison of AGB values between the NBS and the other biomass maps.	70
Figure 31: Fuzzy Numerical index, representing the mean similarity between the biomass maps and the NBS map computed at different spatial resolutions (1 to 50 Km)	71
Figure 32: NBS reference biomass map (left) and difference of AGB values between the biomass maps and the NBS map (right).	72
Figure 33: Fuzzy Numerical maps, representing the spatial distribution of the numerical similarity (s) between the biomass maps and the NBS map	73
Figure 34: Variograms of the biomass maps and NBS field plots (left), with a zoom of the variograms for semivariance values between 0 and 2500 (right)	74
Figure 35: Biomass stock (Tg) in forest and non-forest areas for the biomass maps	75

List of Tables

Table 1: Landsat acquisition date and image block number	29
Table 2: MSSV and its Standard Deviation for selected segmentation inputs on image p172r59.	33
Table 3: MSSV for different window sizes and texture bands, using input Landsat bands 3, 4, 5 and 7 on a representative subset of the image p172r59	33
Table 4: Number of available, selected and discarded NBS plots for each NBS land cover class....	38
Table 5: Results of the comparisons of the biomass estimates obtained with the NBS allometric equations and with four generic equations for all vegetation types (All plots), forest areas only (Forest plots) and woodlands only (Woodland plots).....	41
Table 6: Results of the comparison of the biomass estimates obtained with the NBS allometric equations and two generic equations using different predictor variables	42
Table 7: Performance results on the OOB data of the Random Forest models with and without the predictor land cover.....	48
Table 8: Performance statistics for all data (Model 1 + 2) for biomass classes.....	49
Table 9: Performance statistics for all data for biomass classes in the range 0 – 100 Mg ha ⁻¹	50
Table 10: Performance of Random Forest (RF) and a single Regression Tree (RPART) for different training and testing datasets from image block 1.....	51
Table 11: Model performance on the training data (not on the OOB data) of the uniformly distributed dataset for Random Forest (RF), single Regression Tree (RPART), stepwise multiple linear regression (LM).	52
Table 12: Main characteristics of the biomass and C maps used for the comparative analysis.	62
Table 13: Reclassification scheme to convert the NBS land cover classes to the IPCC classes.	67
Table 14: Total and mean AGB of Uganda for different biomass maps.	69
Table 15: Comparison of biomass estimates with the NBS values for all areas (left columns), forest areas only (center columns) and non-forest areas only (right columns).....	75
Table 16: Comparison of the biomass maps with the NBS field data.	77
Table 17: Comparison of the biomass maps with the NBS field data by biomass classes.	77
Table 18: Total AGB using different combinations of spatial and biomass data.	77
Table 19: Total AGB using different combinations of biomass and spatial data.	78
Table 20: Comparison of biomass and area statistics for the main LC classes of Uganda.....	78

List of Acronyms

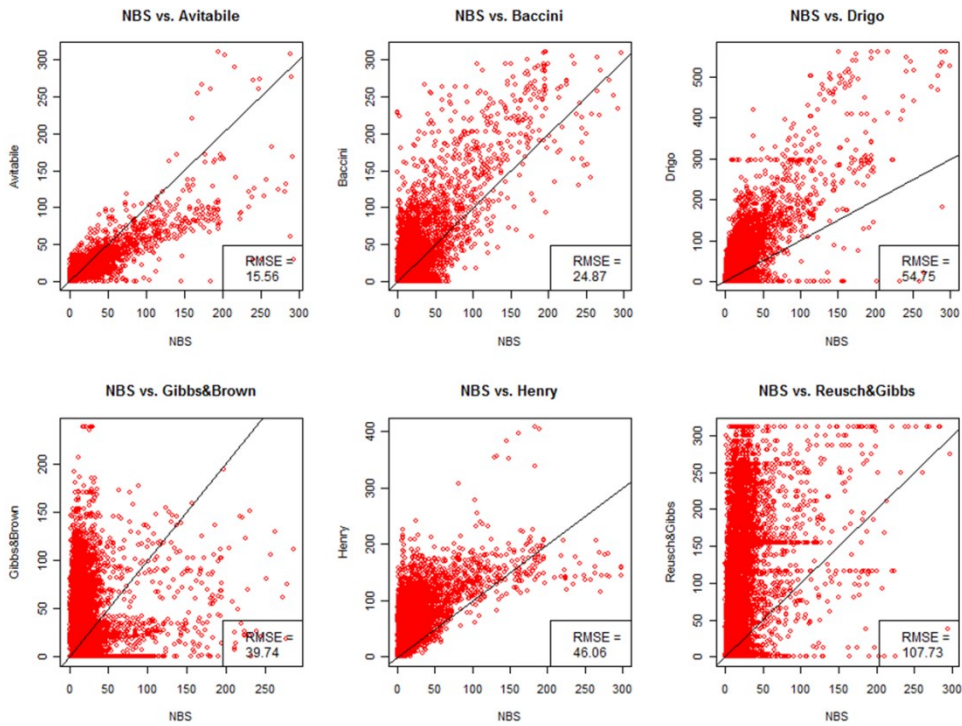
AGB	Aboveground Biomass
AGC	Aboveground Carbon Stock
ALOS	Advanced Land Observing Satellite
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BGW	Brightness, Greenness, Wetness
CDM	Clean Development Mechanism
CO ₂	Carbon dioxide
C	Carbon
CA	Combine & Assign
COP	Conference of the Parties
CR	Crown Width
CV	Coefficient of Variation
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DN	Digital Number
DOS	Dark Object Subtraction
DR	Direct Remote sensing
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus
FAO	The Food and Agriculture Organization of the United Nations
FN	Fuzzy Numerical
FRA	Forest Resource Assessment
GCP	Ground Control Points
GEZ	Global Ecological Zone
GHG	Greenhouse gases
GLAS	Geoscience Laser Altimeter System
GLC2000	Global Land Cover map for the year 2000
GPG	Good Practice Guidance
GtC	Giga tons Carbon
LC	Land Cover
LiDAR	Light Detection And Ranging
LULUCF	Land Use, Land-Use Change and Forestry
H	Height
Ha	Hectare
IPCC	Intergovernmental Panel on Climate Change
ICESAT	Ice, Cloud and Elevation satellite
JERS	Japanese Earth Resources Satellite
JM	Jeffrey-Matusita
Kg	Kilogram
Mg	Megagram
MMU	Minimum Mapping Unit
MODIS	Moderate Resolution Imaging Spectroradiometer
MRV	Monitoring, Reporting and Verification
MSSV	Mean Standardized Segment Variance
NBS	National Biomass Study
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
OLS	Ordinary Least Square

OOB	Out-Of-Bag
R^2	Squared correlation coefficient
REDD+	Reducing emissions from deforestation and forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SD	Standard deviation
SM	Stratify & Multiply
SPOT	Satellite Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SWIR	ShortWave Infrared
Tg	Teragram
TIR	Thermal Infrared
TM	Thematic Mapper
UNFCCC	United Nations Framework on Climate Change
UTM	Universal Transverse Mercator
VCF	Vegetation Continuous Field
WGS84	World Geodetic System 1984
VHR	Very High Resolution
WHRC	Woods Hole Research Center
ρ	Wood density

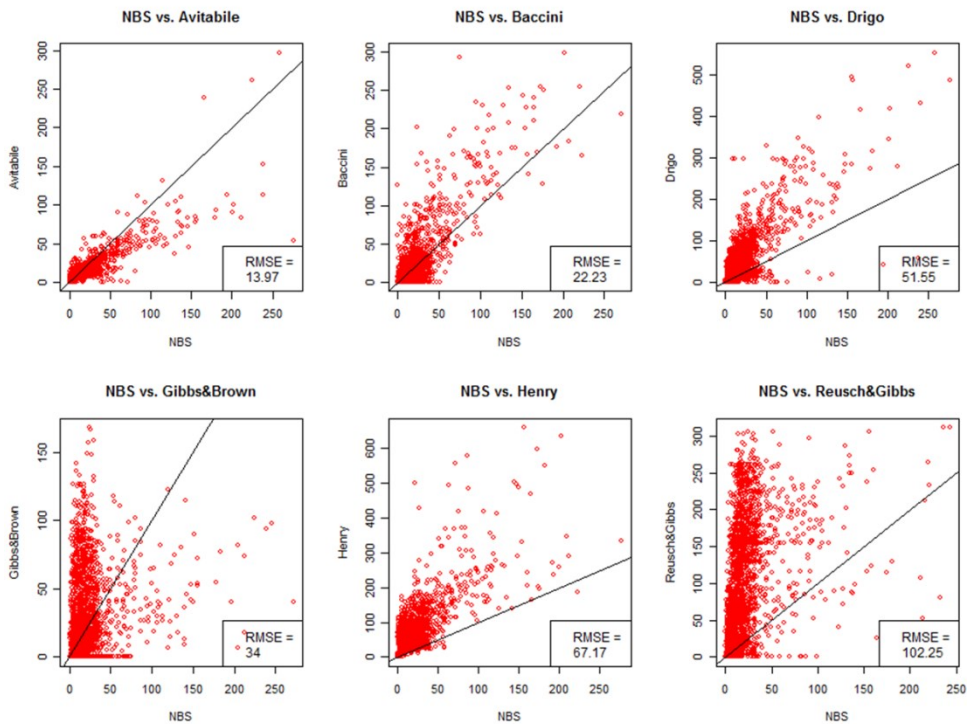
Appendix A

Spatial comparison of biomass maps at variable spatial resolutions (5 - 50 Km)

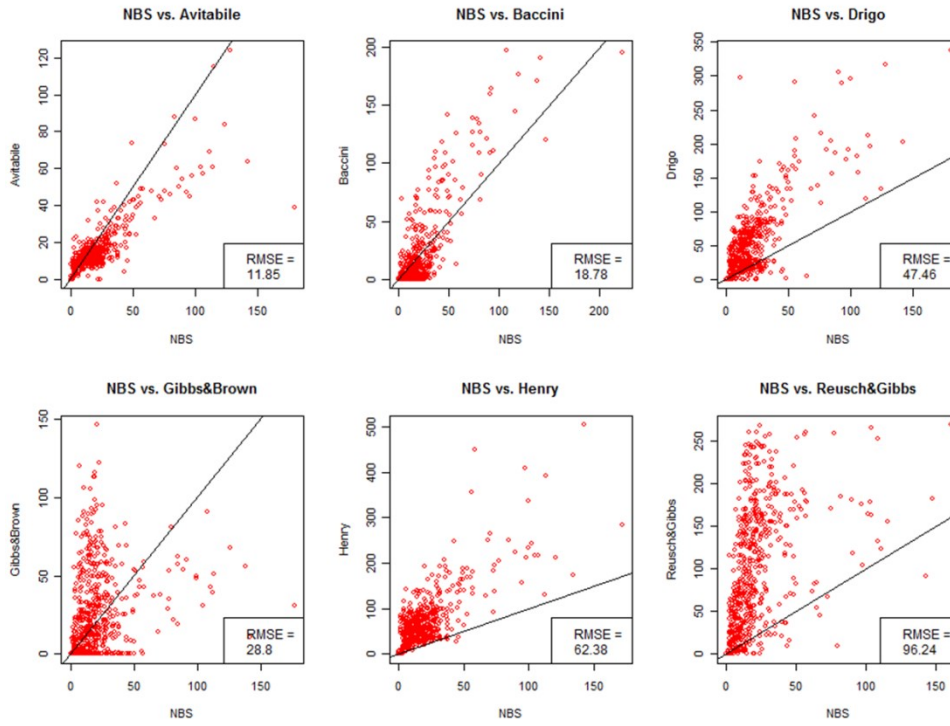
Maps Correlation at 5 Km resolution



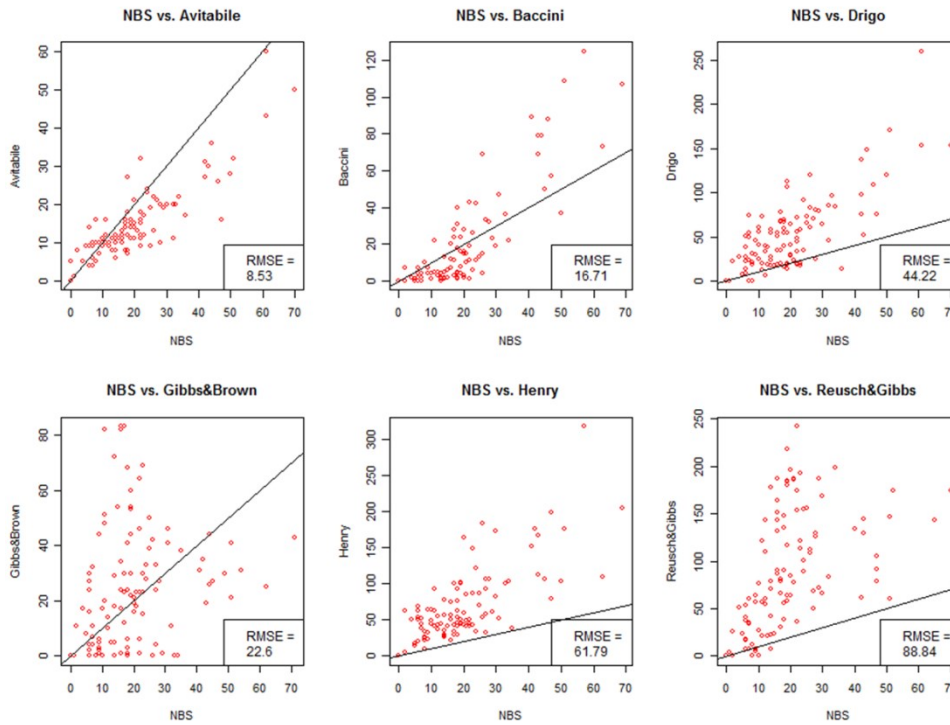
Maps Correlation at 10 Km resolution



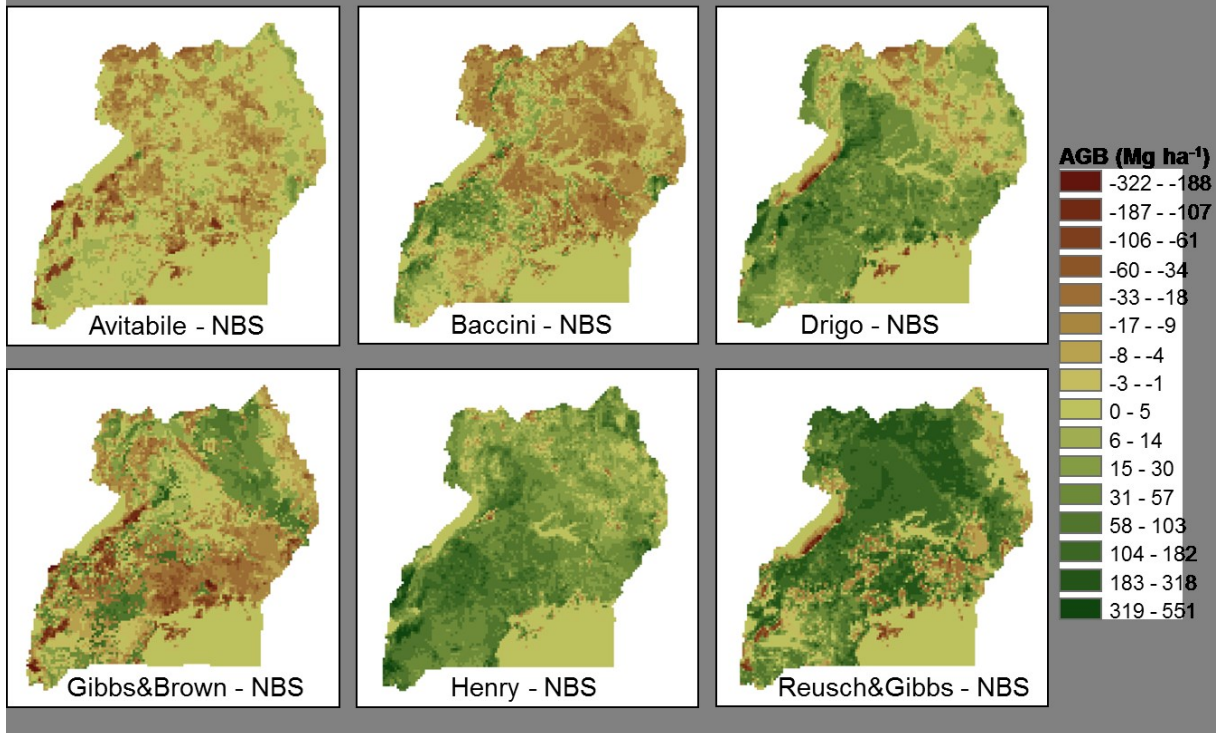
Maps Correlation at 20 Km resolution



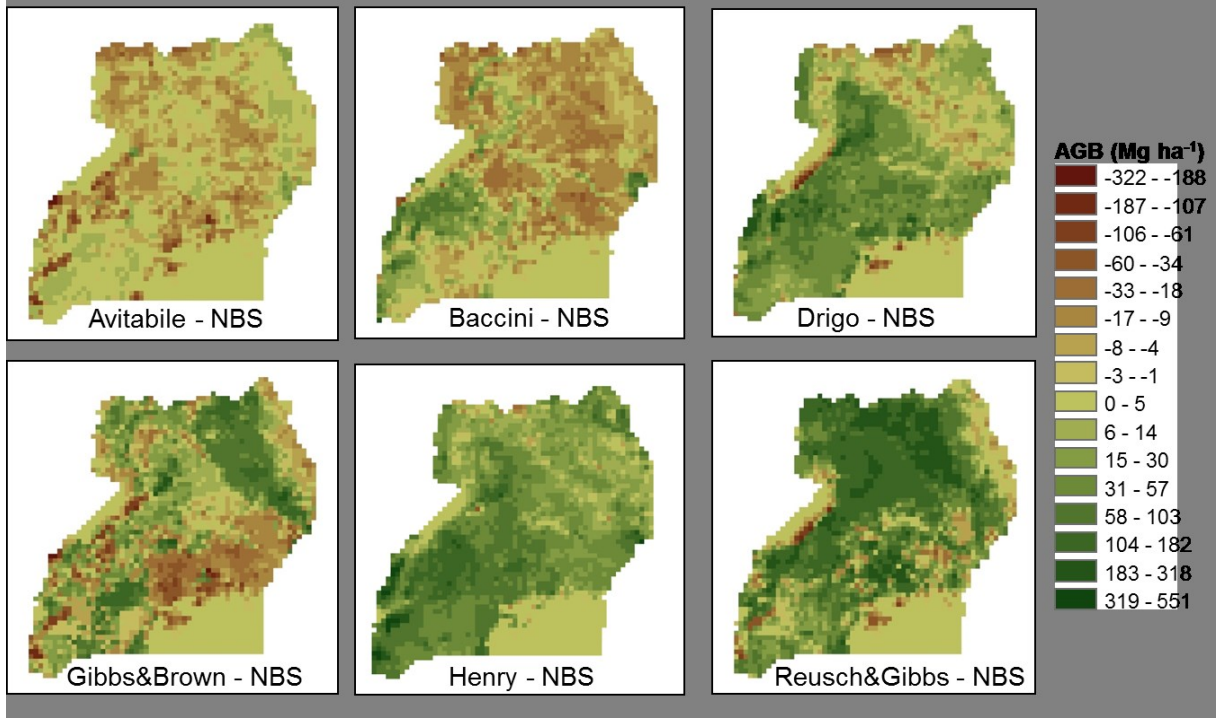
Maps Correlation at 50 Km resolution



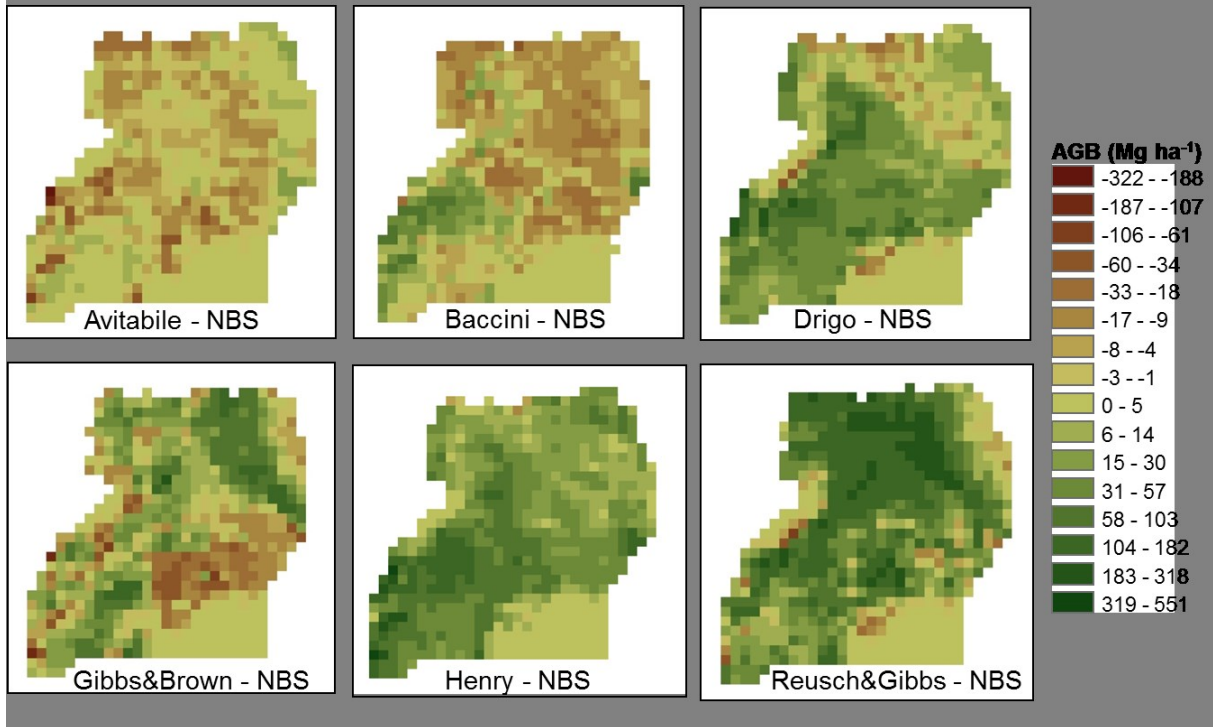
Difference Maps – 5 Km resolution



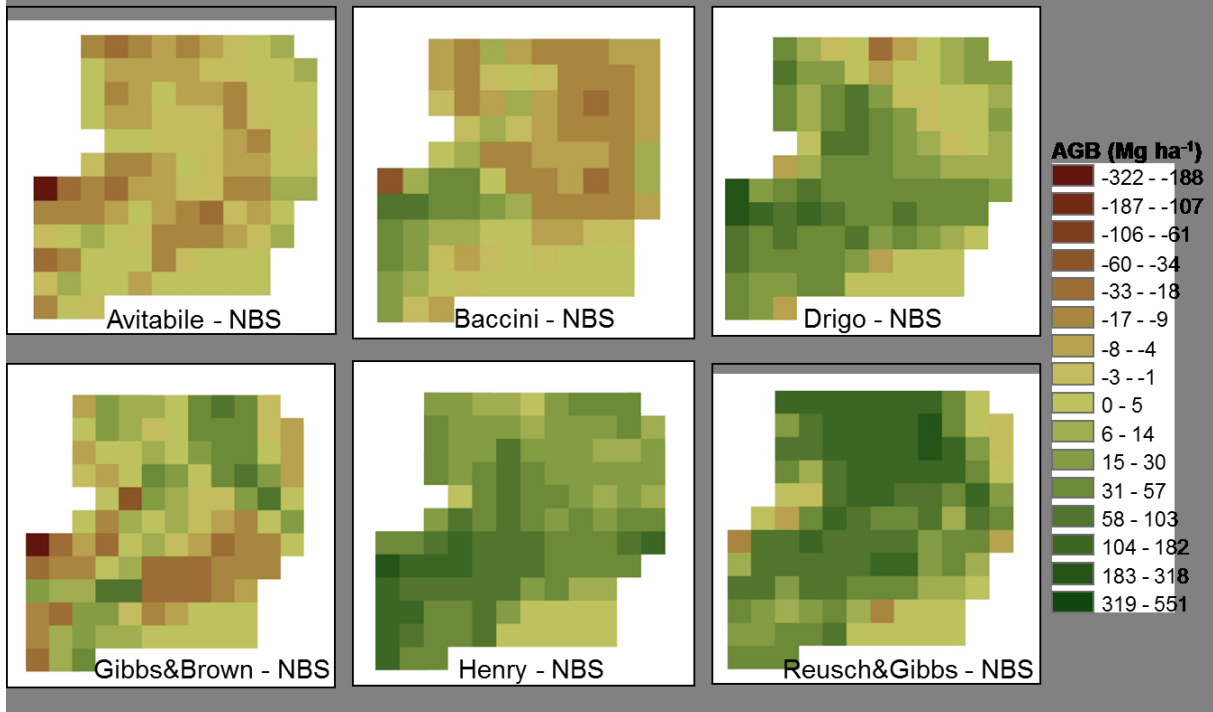
Difference Maps – 10 Km resolution



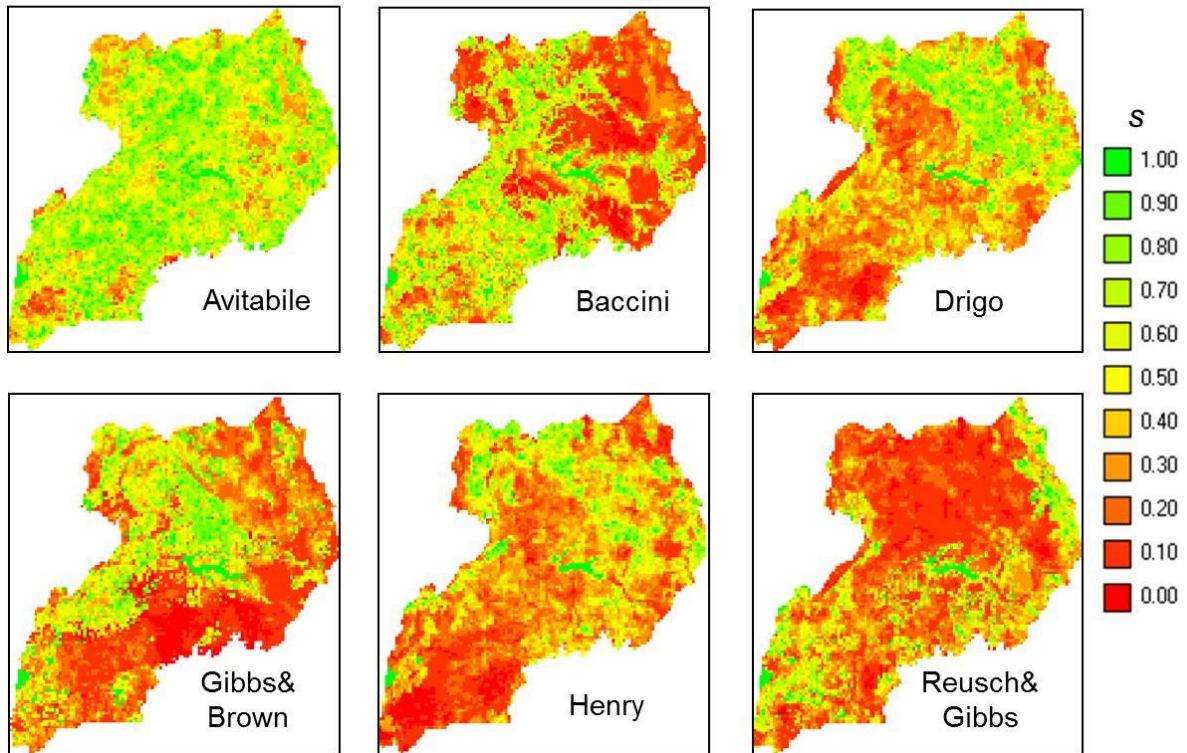
Difference Maps – 20 Km resolution



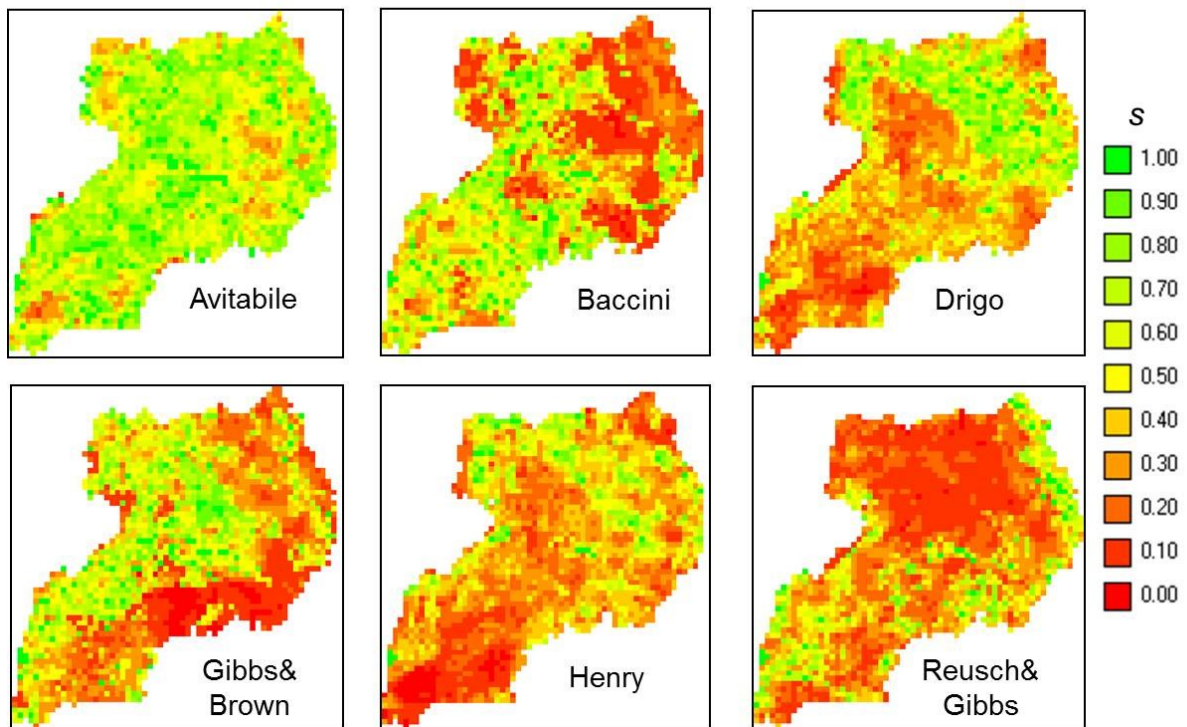
Difference Maps – 50 Km resolution



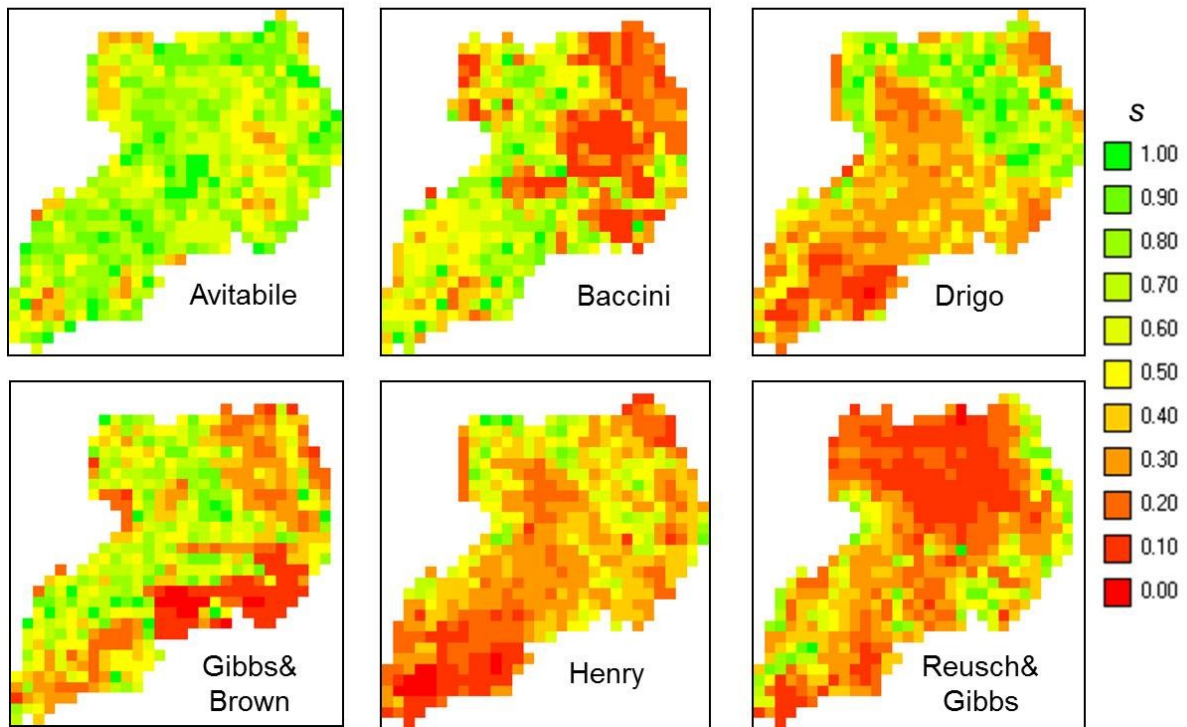
Fuzzy Numerical Maps – 5 Km resolution



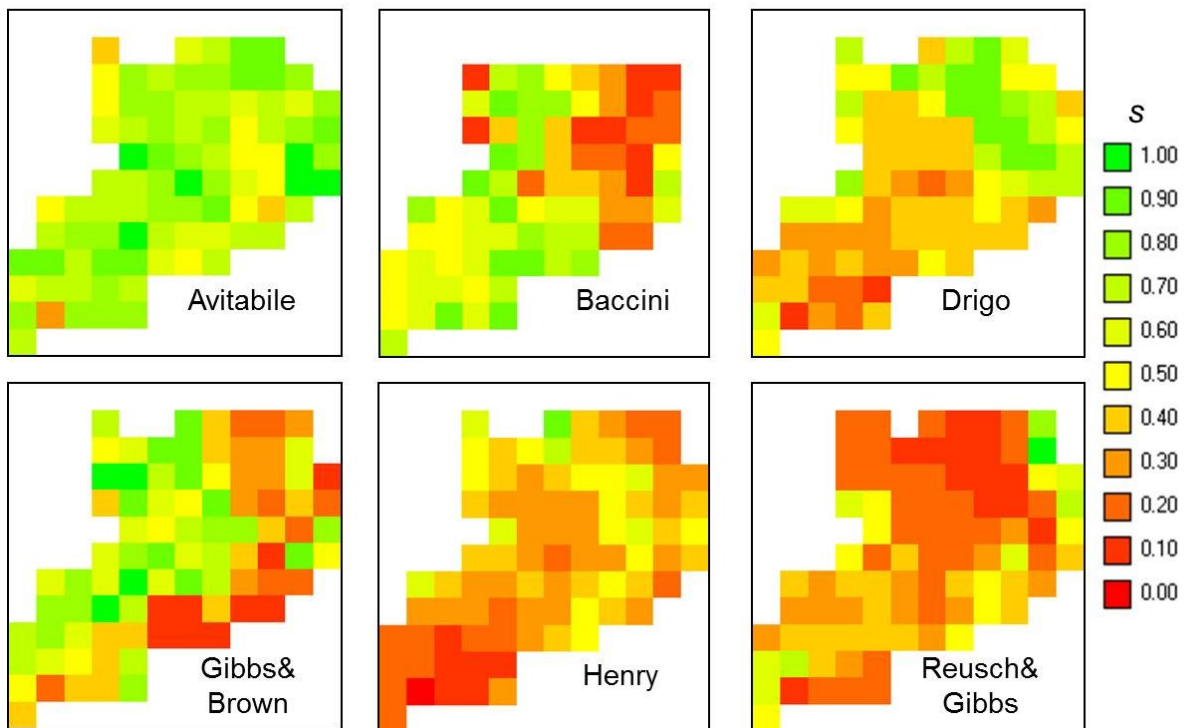
Fuzzy Numerical Maps – 10 Km resolution



Fuzzy Numerical Maps – 20 Km resolution



Fuzzy Numerical Maps – 50 Km resolution



Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorgelegte Arbeit selbständig und unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt habe.

Jena, den 19.12.2011

Valerio Avitabile

Curriculum Vitae

Personal data

Name: Valerio Avitabile
Nationality: Italian
Date of birth: 31 Jan 1975
Place of birth: Roma, Italy
Marital status: Single

Education and Qualifications

August 2010 - Currently

Researcher at the **Friedrich Schiller University** (Jena, Germany) for the project "Land Use and Climate Change Interactions in the Vu Gia Thu Bon River Basin, Central Vietnam" (LUCCI)

September 2007 – May 2008

Visiting Scholar at the Department of Geography of the **Boston University** (Boston, US) for training courses and research activities related to the PhD topic

September – December 2006

Visiting Scholar at the **Woods Hole Research Center** (Massachusetts, US) on the subject "Comparative analysis of approaches for biomass evaluation and mapping at a regional level", funded by the Organization for Economic Co-operation and Development (OECD) (Paris, France)

October 2003 – June 2004

Visiting Scholar at the Dep. of Environmental Resource Management, **University College Dublin** (Ireland) on the subject "Research on carbon stock evaluation in soil and wood products in Ireland", funded by the Italian Ministry of Foreign Affairs

September 2002

Member of the Italian Forest Professional Register

December 2001

Honours degree (5 years Diploma) in **Forest Science**, Tuscia University (Viterbo, Italy).
Specialisation: Environment management. Mark: 110/110 cum laude

July 1993

Scientific High School Diploma. Liceo Scientifico "Taletè" (Roma, Italy)

Language and Computer Skills

Languages

Italian: mother tongue
English: fluent (IELTS Certificate)

Computer Knowledge

ArcGIS, ArcINFO, Envi, R, Office, UNIX

Professional Experiences

January 2007 – March 2010

Research Fellow (Collaborator) at the **Istituto Agronomico per l'Oltremare** (Firenze, Italy) as Principal Investigator for the EU CarboAfrica project.

October 2004 – September 2006

Research Fellow at the **Istituto Agronomico per l'Oltremare** (Firenze, Italy) on the subject “Use of Land Cover data for the evaluation of carbon sequestration capability of terrestrial ecosystems” within the FAO/UNEP Global Land Cover Network initiative

August – October 2004

Technical consultant of the **NGO Movimondo** (Roma, Italy) for the “Mangochi Food Security and Land Care Project” in Malawi, Africa. Duties: Natural resource evaluation in the Namizimu Forest Reserve

March – June 2003

Collaborator for the professional foresters and agriculturists office **Agrifolia** (Roma, Italy)

Training Courses

March – April 2006

Advanced Course (36 hours) in **Carbon and Greenhouse Gases in Agricultural and Forest Ecosystems**, organized by the International Centre for Advanced Mediterranean Agronomic Studies, Spain

September 2005

Week course on **Application of GIS in meteorology and climatology**, organized by the Institute of Biometeorology, National Research Council, Firenze (Italy)

September - October 2003

Training course (60 hours) in **Remote sensing and GIS**, organized by AIT (Italian Remote Sensing Association) and IAO (Istituto Agronomico per l'Oltremare), Firenze (Italy)

July – August 2003

High training course (130 hours) in **Rural Development in Tropical Areas**, organized by Scuola Superiore Sant'Anna, Pisa (Italy)

March 2003

Week course in **Environmental and Forest Certification**, organized by Venetoagricoltura, Legnaro (PD – Italy)

Publications

Avitabile, V., Baccini, A., Friedl, M.A., & Schmillius, C., 2012. *Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda*. Remote Sensing of Environment, 117: 366–380

Avitabile, V., Herold, M., Henry, M., & Schmillius, C., 2011. *Mapping biomass with remote sensing: a comparison of methods for the case study of Uganda*. Carbon Balance and Management, 6, 7

Avitabile V., 2006. *The role of harvested wood products in the Kyoto Protocol*. Sherwood – Foreste ed Alberi Oggi, 121

Green C., Avitabile V., Byrne K., Farrell E.P., 2006. *Reporting carbon stocks in harvested wood products: implications for Ireland*. Biomass and Bioenergy, 30 (2006): 105-114

Saiz G., Green C., Butterbach-Bahlb K., Kiese R., Avitabile V., Farrell E.P., 2006 *Seasonal and spatial variability of soil respiration in a Sitka spruce chronosequence*. Plant and Soil, 287 (2006): 161-176

Avitabile V., Drigo R., Di Gregorio A., Sarfatti P., 2005. *Use of land cover data for woody biomass stock estimates in Eastern and Central Africa*. Proceedings of the conference AfricaGIS 2005, 31 October – 4 November, CSIR International Conference Centre, Pretoria, South Africa

Ciccarese L., Avitabile V., Brown S., Pettenella D., Schlamadinger B., 2003. *Could forest mitigate climate change?* Sherwood – Foreste ed Alberi Oggi, 94

Jena, den 18.07.2012

Valerio Avitabile