OPTICAL REMOTE SENSING OF ABOVEGROUND FOREST BIOMASS AND CARBON STOCKS IN RESOURCE-CONSTRAINED AFRICAN ENVIRONMENTS

Dube Timothy

A thesis submitted to the College of Agriculture, Engineering and Science, at the University of KwaZulu-Natal, in fulfilment of the academic requirements for the degree of Doctor of Philosophy in Environmental Science
(Specialization: Remote Sensing)

Pietermaritzburg
South Africa

June 2015

ABSTRACT

Accurate and reliable information on forest stand volume, aboveground biomass (AGB) and carbon stocks is valuable in understanding and monitoring ecosystem response and its contribution to the global carbon cycle. This knowledge is also critical in climate change modelling and for appropriate planning and monitoring conservation efforts. So far, research has established that the current carbon balance estimates in African ecosystems are characterised by great uncertainties and a very unstable source, with a carbon sink of about 0.3 Pg Cyr^{-1}. This is mainly due to the lack of accurate and robust methods, together with the reliable data sources necessary for quantifying and monitoring forest AGB and carbon stocks at regional scales. Based on this premise, there is a need to swiftly identify affordable, timely and readily-available data sets, together with robust data processing techniques, for the accurate retrieval of forest stand volume, AGB and carbon stocks in the African context. The advancement in remote sensing technologies provides one of the promising primary-data sources or methods urgently required for improving stand volume, AGB and carbon stock estimation accuracies at regional scales. Specifically, the advent of new generation sensors provides new outstanding prospects for improving AGB and carbon stock estimation accuracies in resource-constrained regions (*i.e.* sub-Saharan Africa), where the availability of airborne satellite data (*i.e.* hyperspectral, lidar and radar) remains a challenging task. The main purpose of this study was therefore to examine the utility and strength of cheap, and sometimes freely-available, new generation multispectral sensors. These included the Landsat-8 Operational Land Imager (OLI), SPOT, RapidEye, Sentinel Multispectral Imager (MSI) and WordView-2 for quantifying and mapping AGB and carbon stocks of three plantation forest species (*i.e. Eucalyptus dunii, Eucalyptus grandis* and *Pinus taeda*), a previously challenging task with traditional broadband satellite sensors. To achieve this objective, an overview on AGB and carbon stock studies in African ecosystems was first provided. Secondly, various tree-structural attributes (stand volume, diameter at breast height, and tree height), using medium resolution SPOT-5 image in conjunction with ancillary data sets, were estimated. Finally, AGB estimates were derived, based on the RapidEye Spaceborne sensor and the newly-launched Landsat-8 OLI remote sensing data sets.

The results demonstrate that remote sensing AGB in sub-Saharan Africa accurately and reliably remains a challenge, due to the unavailability of high resolution data sets, as well as high saturation problems associated with the readily-available data sets. In that premise, our results underscored the immediate need to develop robust methods that are capable of enhancing AGB and carbon stock estimates, using cheap and freely-available new generation multispectral sensors. The results obtained using image spectral information and spectral vegetation indices derived from SPOT-5, combined with ancillary data, namely, rainfall metrics and stand age, showed improved predictive accuracies of tree structural attributes, such as tree diameter at breast height (DBH), height and stand volume. For example, SPOT-5 remotely sensed data, combined with ancillary data had better accuracies of mean diameter at breast height (DBH), mean height and stand volume for *Eucalyptu*s trees in plantation forests, compared to the use of remote sensing data as an independent model input data set.

Using vegetation indices and raw spectral bands derived from a fine spatial resolution RapidEye Spaceborne sensor, with strategically-positioned red-edge band, improved both inter- and intra-plantation forest species AGB estimation accuracies. Furthermore, the use of random forest (RF) and stochastic gradient boosting (SGB) regression ensembles significantly enhanced the performance of the new generation multispectral sensors in estimating and mapping AGB. The utility of these regression ensembles underscore the relevance of stochastic models in predicting AGB drawn from different species and genera, using the new generation multispectral sensors, with strategically-positioned bands, as well as fine spatial resolutions.

The results of the study further showed that the newly-launched Landsat-8 OLI multispectral push-broom scanner with a large swath width (185-km) and a 16-day temporal resolution, as well as an improved radiometric resolution (8 to 12 bits), and the signal-to-noise ratio, has the invaluable potential and strength of estimating AGB and carbon stocks, especially in resource scarce regions, where the availability of high spatial resolution data remains a challenge. The strength of Landsat-8 OLI derived texture parameters was further tested and results demonstrated that texture metrics can enhance AGB estimation accuracies, when compared to simple spectral reflectances, simple band ratios and the most popular spectral vegetation indices. The results of this study also demonstrated

that plantation forest aboveground carbon stocks vary significantly ($\alpha = 0.05$) across different tree species, structural attributes (*i.e.* stems, barks, branches and leaves) and across different age groups. These results also underscored the importance of remote sensing technology as an important tool that supports and enhances the quantification of forest carbon stock variations across different tree species, structural components and age groups.

Overall, the findings of this study have shown the unique potential of new generation remote sensing sensors in quantifying AGB and carbon stocks in resource-constrained regions, such as sub-Saharan Africa. In addition, the results provide the necessary insight and motivation to the remote sensing community, particularly in resource-constrained regions, to shift towards embracing the readily-available and cheap new generation multispectral sensors for regional application.

PREFACE

The research work described in this thesis was carried out in the School of Environmental Sciences (SES), University of KwaZulu-Natal, Pietermaritzburg, from January 2013 to June 2015, under the supervision of Prof. Onisimo Mutanga (School of Environmental Sciences, University of KwaZulu-Natal, South Africa).

I would like to declare that the research work reported in this thesis has never been submitted in any form to any other university. It therefore represents my original work, except where due acknowledgments are made.

Timothy Dube Signed: _____ Date: _____

As the candidate's supervisor, I certify the above statement and have approved this thesis for submission.

Prof. Onisimo Mutanga Signed: _____ Date: _____

DECLARATION 1 - PLAGIARISM

I, **Timothy Dube**, declare that:

1. the research reported in this thesis, except where otherwise indicated, is my original research,

2. this thesis has not been submitted for any degree or examination at any other university,

3. this thesis does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons,

4. this thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

    a. their words have been re-written, but the general information attributed to them has been referenced;
    b. where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced;

5. this thesis does not contain text, graphics, or tables copied and pasted from the Internet, unless specifically acknowledged and the source being detailed in the thesis and in the References section.

Signed:_____

DECLARATION 2 - PUBLICATIONS AND MANUSCRIPTS

1. **Dube, T,** Mutanga, O, Ismail, R. In press. Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs, *Journal of Tropical Ecology*. **58**, Issue 2.
2. **Dube, T,** Mutanga, O, Abdel-Rahman, E, Ismail, R, Slotow, R. 2015. Predicting *Eucalyptus spp.* stand volume in Zululand, South Africa: An analysis using stochastic gradient boosting regression ensemble with multisource datasets. *International Journal of Remote Sensing*. **33:** 4502-4526.
3. **Dube, T,** Mutanga, O., Ismail, R. 2014. Intra- and inter-species biomass prediction in a plantation forest: testing the utility of high spatial resolution spaceborne multispectral RapidEye sensor and advanced machine learning algorithms. *Sensors*. **14**: 15348-15370.
4. **Dube, T,** Mutanga, O. 2015. Evaluating the utility of the medium-spatial resolution Landsat-8 multispectral sensor in quantifying aboveground biomass in uMngeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing.* **101:** 36‑46.
5. **Dube, T,** Mutanga, O. 2015. Investigating the robustness of the newly launched Landsat-8 push-broom sensor derived texture indices in estimating medium-density plantation forest species aboveground biomass in KwaZulu-Natal South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing,* **108**: 12–32.
6. **Dube, T,** Mutanga, O. 2015. Quantifying the variability and allocation patterns of aboveground carbon stocks across plantation forest types, structural attributes and age in sub-tropical coastal region of KwaZulu-Natal, South Africa using remote sensing. *Journal of Applied Geography,* **64:** 55–65**.**

Signed: _____

DEDICATION

This thesis is dedicated to Phetho, Prudence, Lisbeth and my late brother Jacob Dube. You passed on early before you could witness this endeavour. Your presence will always be greatly missed….!

To my parents, (Mr K and Mrs M Dube), as well as the rest of the family.

ACKNOWLEDGEMENTS

"*Can God furnish a table in the wilderness? Yes he can….!*" Having grown up from the dusty and remote rural village of Gwanda, Zimbabwe, studying for a PhD was not in the list of my dreams. I would like to therefore take this opportunity and thank the almighty God for giving me this lifetime opportunity.

"If you want to Walk Fast, Walk Alone, but if you want to Walk Far, Walk together…!" Based on this African proverb, I would like to express my appreciation for the support I received from various organisations and individuals during the entire period of my studies. Firstly, I would like to thank the School of Agricultural, Earth and Environmental Sciences through the University of KwaZulu-Natal for affording me the opportunity to pursue my studies and, more importantly, ACCESS-SA, the Applied Center for Climate and Earth Systems Science in South Africa, for funding this research under the theme "Land Use and Land Cover Change". Secondly, a word of thanks is also extended to NASA and the DLR-Germany Aerospace Centre for the provision of Landsat images and high resolution RapidEye images of the area under study. I would also like to express my gratitude to the Sappi Forest Company in South Africa for providing up-to-date forest mensuration data.

Above all, my dream to study for a PhD could not have been realized without my first promoter and supervisor, Prof. Onisimo Mutanga. Soon after completing my MSc with the University of Twente (Netherlands), he began suggesting potential PhD topics, which led to a sound scientific proposal that later led to this PhD thesis. ***"It was indeed a great honour and privilege to work with you Prof. Onnie."*** You taught me not only how to simplify and logically solve scientific problems, but also ***how to confront real life challenges***. Besides, you taught me the scientific writing style and this is confirmed by the number of manuscripts accepted in internationally-recognised ISI journals, with only minor corrections. Truly, I do not know how to thank you Prof, but I certainly know and trust that the good LORD will reward you immensely.

I would also like to extend my words of thanks to Dr Riyad Ismail (Sappi Forest Company, South Africa), for the scientific and technical expertise and guidance you always shared with me throughout the entire period of study. You always made sure that you drove from Sappi to the Department, to attend to our occasionally-scheduled discussions. It was not easy, I have to confess. Your scientific criticism at first watered down my confidence, but later I realised it was worth enduring. ***Thank you, Dr Ismail, for the knowledge you imparted to me.***

In addition, it would be ungrateful for me not to thank Dr John Odindi, Dr Mark Dent, Prof Heinz Beckedahl and Dr Sumaiya Desai (Geography Department UKZN), Dr Elfatih M. Abdel-Rahman (ICIPE, Kenya) and Dr Adam Elhadi (Department of Geography, Wits University), and Dr Samuel Adelabu "Ogar" (University of the Free State), who were always available to share their expertise, which ***provided the most needed stimulation to keep me going***. I would also like to thank my MSc lecturer and thesis supervisor, Prof. Suhyb Mhd. Salama, who consistently encouraged me to further study for a PhD. To my colleagues dearest Dr Khoboso E. Seutloali, Victor Bangamwabo and Mbulisi Sibanda (Geography Department UKZN), together we undoubtedly formed a good and formidable team for any challenge, whether scientific or social. ***"That's what true friends are there for. Thank you guys…"***

I would also like to thank my former undergraduate mentors Mrs Idah Mbengo, Dr Lazarus Zanamwe, Prof. TPZ Mpofu and, above all, Prof. Amon Murwira, who groomed me to be who I am from my early days as an undergraduate Geography (GIS and Remote Sensing) student. You always inspired us with your GIS and remote sensing antics from the first moment you taught us. May the good LORD protect you and give you more strength to even inspire many. A word of thanks also goes to the rest of my former undergraduate mentors from the University of Zimbabwe and my MSc lecturers from the ITC Faculty, at the University of Twente. To my dearest friends, Tawanda Winmore, Gara (currently a PhD candidate: ITC-UT), Emmerson Chivhenge and Charity Nyelele (currently a PhD candidate-SUNY-ESF Environmental Science), I still remember the times we used to share our ambitions when we were still doing our undergraduate and MPhil studies at the University of Zimbabwe. Although life was not that rosy, we never lost focus and now here we are.

To my colleagues, at the Department of Geography, UKZN, Fadzai Pwiti, Lucky Nkomo, Dr Kabir Peerbhay, Romano Lottering, Dr Mercy Ojoyi, Gelal Omer, Charles Otunga, Kuhle Ndyamboti (CSIR), as well as Samuel Kusangaya (Water Resource Commission, UKZN), thank you for your moral and brotherly support all the time. I would also take this opportunity to thank my friends and colleagues, Webster Gumindoga my mentor ("Prof"), Mary Chawira, Henry Ndaimani, Mark Zvidzai, Godfrey Pachavo, Terence Mushore, Tsitsi Bangira (Tete) and Abel Chemura.

My since gratitude also goes to all members of staff at the Department of Geography, UKZN and Higher Degrees Office (particularly Sibongile), for their overwhelming support and friendship. My special thanks go to Mrs. Shanita Ramroop for her support during the entire period of study, *"**that was marvellous Shani"***. Your office was always wide open for me. May God reward you abundantly. I am also thankful to Mr Donavan de Vos and Brice Gijsbertsen for being there for me and also making sure that all logistics were available. Above all, thanks to the Cleaning Department, who tirelessly kept the PhD research office clean.

Similarly, I appreciate my Pastor (T Siziba -*"The unconquerable")* and the entire AFM-Blessed Family Church, in Pietermaritzburg, South Africa, for their stead-fast prayers during the entire period of study. ***May the good LORD protect and bless you.*** I am also greatly indebted to my parents and the entire family for their moral support, prayers and encouragement. Most importantly, special thanks also go to the one who stole my heart, Cletah Shoko, for her moral support and prayers. I had little time to always be with you my dear; however, you kept encouraging and praying for me to work hard and to never lose hope. ***Now I can safely say it is over….!***

Timothy Dube

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTERS ONE AND TWO

GENERAL OVERVIEW AND CONTEXTUALISATION

1. GENERAL INTRODUCTION

Figure 1.1 Plantation forest ecosystems in the sub-tropical coastal region of KwaZulu-Natal, South Africa

1.1 African Forest Ecosystems and the Global Carbon Cycle

Forest ecosystems play an important role in the global carbon cycle and are known to be the prime drivers of both regional and local climate systems through biosphere-atmospheric interactions (Baccini *et al.,* 2008; Field *et al.,* 1998; Martinic and Sporcic, 2005). Recent global carbon assessments, based on atmospheric carbon dioxide (CO_2) observations, inverse modelling techniques and land observations by the Intergovernmental Panel on Climate Change (IPCC), show that terrestrial ecosystems sequester approximately 2.0-3.4 Pg C $year^{-1}$ (Laurin *et al.,* 2014; Wolf *et al.,* 2011). The increased global recognition and importance of forest ecosystems in the global carbon cycle recently triggered international negotiations and agreements. It was concluded that an accurate and reliable understanding of the current and possible future role of forests as carbon sinks, in both managed and unmanaged ecosystems, was fundamental for worldwide climate modelling, due to their significance in the net carbon sink and emission calculations (IPCC, 2003b; Patenaude *et al.,* 2004; UNFCCC, 2001; 2011). For instance, international agreements and regional markets have since started accrediting an economic value to forest stand volume, aboveground biomass (AGB) and carbon stocks, encouraging signatory countries to intensify and maintain forests for climate mitigation purposes (Patenaude *et al.,* 2004; UNFCCC, 1998b). Moreover, countries that ratified the Kyoto Protocol to the United Nations Framework Convention on Climate Change (UNFCCC) are required to provide annual reports on their carbon sequestrations from Land Use, Land Use Change and Forestry (LULUCF) activities (Patenaude *et al.,* 2004). Information on the spatial distribution of forests, forest stand volume and biomass storage, as well as carbon storage dynamics, is therefore crucial for the continuous and accurate quantification of CO_2 greenhouse gas fluxes, policy development and implementation. In addition, knowledge on forest stand volume, AGB and carbon storage is essential in evaluating forest productivity, ensuring well-informed sustainable conservation practices, deriving future atmospheric carbon projections, as well as understanding the general functioning of the planet and the environment.

While there is an increasing understanding of the role of forest ecosystems in global carbon cycle, research on forest stand volume, AGB and carbon stocks quantification and monitoring, particularly in managed plantation forests, across individual African countries and at continental scale, still remains scarce (Dube *et al.,* 2015; Laurin *et al.,* 2014). The reason for the scarcity of up-to-date and accurate forest carbon information in Africa is partly due to: (i) limited resources for national and/or continent level forest carbon monitoring, (ii) a lack of political will and prioritization, (iii) a

lack of continental cooperation, and (iv) to some extent, the lack of technical and scientific expertise (Boahene, 1998; Amuzu-Sefordzi *et al.,* 2015). Therefore, to meet the demands of the Kyoto Protocol, as well as to monitor and understand the contribution of both managed and unmanaged forest ecosystems across the African continent, there is need to develop robust techniques and identify suitable and optimal spatial data sets for the quantification and monitor of forest carbon stocks at a regional scale. This will also enhance a critical understanding of forest ecosystem responses to the global carbon cycle and climate change in forest ecosystems in the African context.

The recent recognition of forests as a potential sink of atmospheric carbon has resulted in numerous studies being conducted in estimating AGB or carbon stocks across varying scales, with large efforts noted in the developed world (*i.e.* in Europe, North America and other parts of South America and Asia) (Baccini *et al.,* 2008; Henry *et al.,* 2011; Lu, 2006). However, other developing parts of the world, particularly Africa, are still lagging behind, despite the immediate need for knowledge on forest stand volume and the amount of AGB or carbon stored in their ecosystems. For example, the study by Woollen et al. (2012b) shows that the current knowledge of Africa's carbon pools is limited, despite its importance in the global carbon budget. This knowledge gap is further highlighted by other studies that indicate relatively high forest carbon storage uncertainties across African ecosystems (Laurin *et al.,* 2014; Wolf *et al.,* 2011). For example, a study by Wolf et al. (2011) on the latest net long-term forest carbon balance showed a great uncertainty and an unstable carbon source. Moreover, most of these estimates are mainly based on forest stand volume, AGB and carbon predictions derived from indigenous forests only. It is therefore against this premise that many questions currently remain unanswered regarding the actual contribution of African ecosystems (both managed and unmanaged forest ecosystems) to the global carbon cycle (*i.e.* whether the continent's forest ecosystems are net atmospheric carbon sources or sinks). In this regard, it is necessary to identify quick, efficient, robust techniques and timely data sets with a spatial coverage that would permit an accurate assessment of the African forest ecosystems' contribution to the global and regional carbon cycle and its fluxes, a previously challenging task. So far, two broad approaches, namely, field-based traditional methods, which involve field measurements or harvesting and recent remote sensing techniques, have been developed to quantify forest stand volume (i.e. a proxy for aboveground forest biomass) and AGB. This facilitates the derivation of forest carbon stocks by converting the attained biomass by a factor of 0.5 (Dixon *et al.,* 1994; Houghton, 2005).

1.2 Quantification of Aboveground Forest Biomass and Carbon Stocks

As mentioned earlier, forest carbon is quantified and monitored, based on either traditional field-based surveys or remote sensing (Gibbs *et al.,* 2007; Henry *et al.,* 2011; Lu, 2006). Field-based methods involve tree-harvesting and weighing or the use of available generalized or species-specific allometric equations, as well as the use of biomass expansion factors and stand volume (Henry *et al.,* 2011). Although the results from these methods are regarded as highly accurate for AGB quantification, their use remains restricted, as they are not practical over large areas. Besides, field-based AGB methods are exceptionally labour-intensive, time-consuming, and difficult to implement across regional and remote areas. Lu (2006) describes field-based methods as practically and spatially restricted to a small tree sample size and they require a sufficient number of samples, which in most cases is not feasible. Recently, scientists, ecologists, environmentalists and the remote sensing community have begun to discredit the use of field-based methods for AGB quantification, due to a number of associated inaccuracies since its inception in favour of the new and cutting-edge remotely sensed technologies (Koch, 2010; Lu, 2006).

In contrast, a significant body of literature illustrates that remote sensing technologies are a reliable source of data (Dube *et al.,* in press; Gonzalez *et al.,* 2010; Koch, 2010; Zhang *et al.,* 2014). Remote sensing provides valuable and relatively cheap primary data necessary for the timely and accurate quantification of forest structural attributes (*i.e.* stand volume, AGB and carbon stocks), when compared to the use of convectional field surveys (Anaya *et al.,* 2009; Koch, 2010; Sivanpillai *et al.,* 2006). Furthermore, remote sensing technologies allow robust landscape scale stand volume, AGB and carbon stocks retrieval, as well as minimizing costs associated with field data collection (Adam *et al.,* 2014; Dube *et al.,* 2014a; Güneralp *et al.,* 2014). Consequently, the application of remote sensing technologies in quantifying stand volume, AGB and carbon stocks as the prime data-source, has gained increasing attention, especially for natural forest AGB across the African continent and the world over (Dube *et al.,* in press; Gonzalez *et al.,* 2010; Koch, 2010; Zhang *et al.,* 2014).

1.3 Remote Sensing of Aboveground Forest Biomass and Carbon Stocks

The recent need for understanding the role of terrestrial ecosystems on the global carbon cycle has prompted the need for the accurate, timely and reliable mapping and monitoring of stand volume, AGB (proxies for forest carbon) and carbon stocks, using remote sensing data sets. The data sets

range from coarse-to-medium spatial resolution, known as broadband sensors (Baccini *et al.,* 2004), high spatial resolution hyperspectral (Clark *et al.,* 2011; Swatantran *et al.,* 2011), radar and airborne light detection and ranging (lidar) instruments (Popescu *et al.,* 2011; Sun *et al.,* 2011). The use of these data sets, particularly airborne sensors (*i.e.* hyperspectral, lidar and radar), has proven useful in vegetation mapping and monitoring. Airborne sensors provide the primary spatial data necessary for the accurate estimation of important tree structural attributes, such as wood volume and AGB or carbon stocks, when compared to field-based methods (Koch, 2010).

Literature, however, demonstrates that the use of broadband optical sensors, such as the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Moderate Resolution Imaging Spectroradiometer (MODIS), with moderately coarse spatial resolution, results in poor stand volume, AGB and carbon stock estimates due to the occurrence of mixed pixel information (Basuki *et al.,* 2011). Basuki *et al.* (2011) observed that sensor signal within a single pixel is actually a spectral aggregation of radiance or reflectance of all components *e.g.* mixture of forest canopy cover and bare soil, *etc*. This aggregation hinders the accurate quantification of important vegetation structural attributes *e.g.* biomass, especially when using spectral vegetation indices (Basuki *et al.,* 2011). It has been established that optical sensing techniques have a track record of having saturation problems when used in biomass-related studies (le Maire *et al.,* 2011; Mutanga and Skidmore, 2004b) and also suffer greatly from high signal-to-noise ratio. Adding to that, these sensors are generally characterized by a decreasing sensitivity whenever tree canopy structural heterogeneity and age increases (Lima *et al.,* 2003; Patenaude *et al.,* 2004). When applied to 100% vegetation canopy cover conditions, or when the leaf area index (LAI) is around 3–4 m^2m^{-2}, AGB estimates from multispectral data saturates, although under normal circumstances, forest LAI can be above $4m^2m^{-2}$ (le Maire *et al.,* 2011; Wang *et al.,* 2005).

In view of the above, it is clear that the applicability of broadband optical remote sensing alone on forest AGB estimation remains restricted to simpler forests with limited canopy closure, where LAI is less than $4m^2m^{-2}$, such as coniferous forests (le Maire *et al.,* 2011), hence the need to be integrated with high resolution images or ancillary data sets. On the contrary, findings from other remote sensing AGB studies demonstrate high accurate AGB and carbon estimates obtained, using high spatial resolution hyperspectral, radar and airborne lidar sensors (Montesano *et al.,* 2014; Zhang *et al.,* 2014). Using the Geoscience Laser Altimeter System (GLAS) sensor, Zhang *et al.* (2014) estimated AGB at a national scale with high accuracy, producing an R^2 of 0.97 and the root

mean square error (RMSE) of 9.11 t ha^{-1}. Although useful in providing the primary spatial data sets necessary for the accurate estimation of important tree structural attributes, the use of hyperspectral, radar and airborne lidar sensors in vegetation-related studies comes with its own challenges, particularly in data-scarce areas.

1.4 Implications of Remote Sensing of Aboveground Forest Biomass and Carbon Stocks in Resource-Scarce Areas

The challenges currently faced by the remote sensing community with regard to the application of hyperspectral, radar and airborne lidar sensors in vegetation-related studies (*e.g.* stand volume, AGB and carbon stocks) include limited spatial coverage, high image costs, availability, huge data-volumes, the need for excessive number of field samples, as well as high data pre-processing costs, particularly in resource-constrained areas (Gara *et al.,* 2014b; Mutanga *et al.,* 2015; 2012). Despite the detailed spectral information, hyperspectral data processing is complex, with geometrical and statistical properties associated with high data dimensionality. The use of such data sets, particularly lidar for large-scale forest stand volume, or AGB across Africa, is still greatly under-developed. It is against this background that the abovementioned remote sensing technologies are considered impractical for ‘‘wall-to-wall’’ AGB and carbon stocks quantification in resource-scarce areas, such as the sub-Saharan Africa (Carreiras *et al.,* 2012; Colgan *et al.,* 2012; Dube *et al.,* 2014a; Wulder *et al.,* 2008b). Recent studies have demonstrated that the use of airborne sensors (hyperspectral, lidar) for field campaigns is constrained by operational restrictions on data-acquisition flight campaigns, the variability of data accessibility on a country basis and multicollinearity problems (Mathieu *et al.* 2013). Consequently, these data sets are more appropriate for small-scale, or ‘‘project-based applications’’, or for use in sampling strategies, especially in resource-constrained areas (Carreiras *et al.,* 2012). For instance, hyperspectral, radar and airborne lidar sensors are critical for the mapping and monitoring of plot-level stand volume, AGB and carbon stocks over small forest units. These limitations therefore prompt the need to shift towards the use of new generation multispectral sensors (*i.e.* Landsat-8 OLI and RapidEye), with more appealing and robust properties necessary for large-scale AGB estimation.

New generation multispectral sensors are cheap and readily-available, with a large swath-width and are capable of providing large-scale stand volume and AGB mapping and monitoring (Gara *et al.,* 2014b; Mutanga *et al.,* 2015; 2012). This advancement in imaging technologies in recent years

provides cost-effective, commendable and reliable opportunities for monitoring, inventorying and understanding changes in forest attributes *i.e.* AGB and carbon from plot to regional scales. The presence of improved radiometric, spatial and spectral properties associated with the new generation multispectral sensors are therefore hypothesised to be capable of permitting accurate and reliable regional scale or "wall-to-wall" AGB and carbon accounting (Gibbs *et al.,* 2007; Hall *et al.,* 2011; Houghton *et al.,* 1996; Laurin *et al.,* 2014), as well as sustaining forest resources inventorying and conservation (Næsset, 2007). These new generation multispectral sensors have a significant number of lucrative improvements in terms of their radiometric, spatial and spectral resolutions that are deemed imperative for forest-related studies (Gara *et al.,* 2014b; Kross *et al.,* 2015; Machwitz *et al.,* 2014; Mutanga *et al.,* 2015; Rana *et al.,* 2013; Zandler *et al.,* 2015). In addition, some of these sensors have a wide swath-width and a high revisit time (*i.e.* Landsat-8 OLI, with 185 km swath width and a 16-day temporal resolution), which is of paramount importance for continuous and large-scale AGB and carbon stock estimates. New generation multispectral sensors also contain refined spectral bands with unique and strategically-positioned band-settings at high spatial resolution, such as the red-edge band of the RapidEye. This band is known to contribute noticeably to AGB estimation at high density canopies (Gara *et al.,* 2014b; Mutanga *et al.,* 2015; 2012).

Similar to airborne hyperspectral sensors, new generation multispectral sensors (*e.g.* RapidEye) are associated with high spatial resolutions of 4-m, surpassing those of the old broadband multispectral sensors (Gara *et al.,* 2014b; Mutanga *et al.,* 2015; 2012). Studies using hyperspectral spectral bands and spectral vegetation indices computed from these bands (red-edge) have demonstrated their sensitivity to variations in plant biophysical properties (Gara *et al.,* 2014b; Mutanga *et al.,* 2015; *et al.,* 2012). It is therefore hypothesized that comparable forest stand volume, AGB and carbon stock estimates can be similarly achieved using the new generation multispectral sensors, which also have improved radiometric, spatial and spectral properties (Gara *et al.,* 2014b; Mutanga *et al.,* 2015; 2012). The presence of a limited key number of spectral bands is critical in reducing superfluous redundancy and minimising noise associated with airborne hyperspectral data sets.

To address the general limitations facing Africa and to ensure accurate and reliable forest stand volume, AGB and carbon stocks in these areas, various techniques and new generation multispectral sensors have to be tested. This research endeavour will help to provide precise recommendations on cheap, robust and remote sensing sensors or optimal spectral information

suitable for forest: i) stand volume, and ii) aboveground biomass (important proxies of forest carbon stocks) mapping and monitoring in data-scarce areas. The emphasis of this thesis was therefore restricted to the integration of multispectral SPOT-5 and ancillary data, as well as the use of new generation multispectral sensors, such as the Landsat-8 OLI and RapidEye Spaceborne sensors, in mapping and monitoring important forest structural attributes in resource-constrained areas across different scales. The task was simply to identify, develop and test robust methods and techniques for effective forest stand volume, AGB and carbon stocks mapping and monitoring in resource-scarce areas at local, regional and global scales.

1.3 Aim

Informed by the aforementioned observations, the main aim of the study was to assess the performance and strength of using multisource data and the new generation medium-to-high spatial resolution multispectral sensors in mapping and monitoring plantation forest species stand volume, AGB and carbon stocks in the uMgeni Catchment, KwaZulu-Natal, South Africa.

1.4 Objectives of the thesis

The main objectives of the study were:

1. to provide a detailed overview on the trade-offs between satellite data availability, cost and predictive accuracy in quantifying aboveground biomass in African environments,.
2. to predict *Eucalyptus spp.* stand volume, using stochastic gradient boosting regression ensemble with multisource data sets,
3. to test the utility of high spatial resolution spaceborne multispectral RapidEye sensor and advanced machine learning algorithms in estimating intra- and inter-species biomass prediction in plantation forests,
4. to evaluate the utility of the medium-spatial resolution Landsat-8 OLI multispectral sensor in quantifying aboveground biomass,
5. to investigate the robustness of the newly-launched Landsat-8 OLI push-broom sensor derived texture indices in estimating medium-density plantation forest species aboveground biomass, and
6. to quantify the variability and allocation patterns of aboveground carbon stocks across plantation forest types and structural attributes in sub-tropical coastal region of KwaZulu-Natal, South Africa, using remote sensing.

1.5. Scope of the study

This thesis was aimed at examining the capability of new generation sensors in estimating the structural attributes (*i.e.* stand volume, AGB and carbon stocks) of important plantation forest species in the uMgeni catchment, KwaZulu-Natal, South Africa. The study mainly highlights the potential and strength of integrating ancillary data with medium spatial resolution SPOT-5 data in predicting stand volume. The study further demonstrates the utility of new generation multispectral sensors with unique band-settings and design (*i.e.* RapidEye and Landsat-8 OLI sensors) in accurately estimating inter- and intra-species AGB. The strength and performance of new generation multispectral sensors in estimating and mapping forest stand volume and AGB is examined, using two robust and effective non-parametric machine learning algorithms, such as the random forest (RF) and stochastic gradient boosting (SGB). These algorithms (RF and SGB) have the potential to integrate the strength of regression trees and to boost and involve a probabilistic component, which decreases the final model variance and improves the predictive accuracy (Breiman, 2001; De'ath, 2007; Elith *et al.,* 2008).

1.6. Description of Study Area

The study was conducted on eucalyptus and pine plantation forests currently managed for pulpwood production by the Sappi Pulp and Paper Company within the uMgeni Catchment, South Africa (Figure 1.1). The most predominant species found within the catchment include the fast-growing stocks of *Eucalyptus dunii* (*E.dunii*), *Eucalyptus grandis* (*E. grandis*) and *Pinus taeda (P. taeda)*. These plantations are evenly scattered along the coast, stretching from the north of Pietermaritzburg (Latitude 29° 24'S, Longitude 30° 18'E) to the south around Richards Bay (Latitude 28˚ 48'S, Longitude 32˚ 02'E) and to the north, including the town of Mtubatuba (Latitude 28° 25'S, Longitude 32° 10'E) (Figure 1.1). The uMgeni Catchment is characterised by moderately steep and undulating topography, ranging from ~50 to 1266 m above sea-level. The climate of the area is predominantly subtropical, with high humid temperatures (*i.e.* mean annual temperature is approximately 21.7ºC) and high summer rainfall during the summer months (*i.e.* October - February), varying between 730 to 1 500 mm. It also consists of deep-fertile lithic soils, such as the leptosols, cambisols, acrisols and lixisols, favourable for the rapid-growth of *Eucalyptus* and *Pine* stands are the most predominant (Sappi, 1993; Scott and Lesch, 1997).

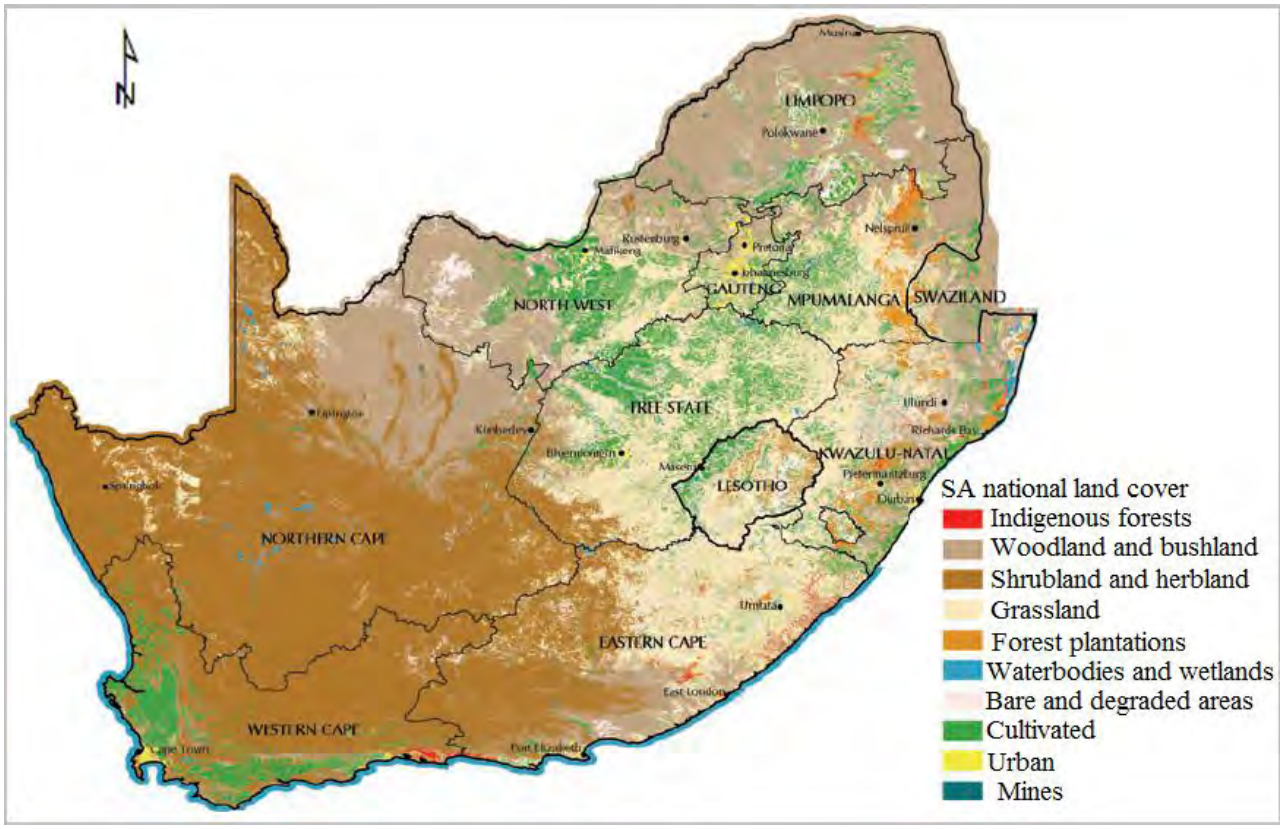


Figure 1.2 The spatial distribution of plantation forests in the south-eastern regions of South Africa (CSIR, 2000)

1.7 Thesis Outline

This thesis is comprised of six stand-alone articles, which have been submitted to internationally-recognised peer-reviewed journals. Five of the manuscripts have already been published online, one is in press. In this thesis, each article has been presented as a stand-alone chapter that can be read and considered autonomously, from the entire dissertation, but it contributes to the overall introduction (Chapter One) and synthesis (Chapter Eight). It is also critical to note that the content of most of the manuscripts submitted to peer-reviewed journals has been retained. This means that each of the stand-alone chapters consist of its abstract and conclusion, which relate it to the subsequent chapter, hence the presence of duplications and overlaps, particularly in the 'introduction' and 'methods' sections, of the various thesis chapters. This duplication is assumed to be of little consequence when considering that these are peer-reviewed scientific articles, which are stand-alone chapters that can be read separately, without losing the overall context. The entire thesis is made up of eight chapters. These chapters can be split into four major sections: (i) General overview and contextualisation, (ii) Local scale mapping, (iii) Regional scale mapping, and (iv) Forestry application and synthesis.

1.7.1 General Overview and Contextualisation

1.7.1.1 Chapter One

This chapter serves as an introduction and a contextualization of the study. It highlights the importance and environmental relevance of the mapping and monitoring of key forest structural attributes, such as tree volume, AGB and carbon stocks, which serve as a proxy for assessing forest contribution to the global carbon cycle. The chapter further elaborates on key methods and data sets, as well as the challenges faced by researchers in undertaking this type of study in areas with limited resources. In addititon, the detailed research problem, aim and objectives are provided.

1.7.1.2 Chapter Two

This chapter provides a detailed overview of remote sensing applications for AGB and carbon stocks mapping and monitoring in sub-Saharan Africa. It further presents the research strides and challenges (*e.g.* particularly the trade-offs between availability, cost and predictive accuracy) on AGB mapping and monitoring in Africa environments. The chapter also highlights the gap and need for identifying and developing cheap methods for accurate AGB mapping and monitoring in resource limited areas.

1.7.2 Local Scale Mapping

1.7.2.1 Chapter Three

The chapter focuses on predicting stand volume and other related tree structural attributes in managed plantation forest, using remote sensing and ancillary data sets. Information on forest stand volume is fundamentally important in understanding the role of forests as carbon sinks and their contribution to the global carbon cycle. Besides, this information is a pre-requisite for a detailed evaluation of commercial forest resources and their sustainable management. In this chapter, the application of multisource data (SPOT-5, rainfall data and stand age) in predicting plantation forest stand volume is examined, using a stochastic gradient boosting (SGB) algorithm.

1.7.2.2 Chapter Four

While the integration of ancillary data with SPOT-5 spectral information produces high quality stand volume estimates, there is need to identify remote sensing data that can be independently applied in areas where ancillary data sets are unavailable, to derive accurate estimates at a local or regional scale. This chapter therefore focusses on inter-and-intra-species AGB estimation at a local scale, using the high spatial resolution RapidEye Spaceborne sensor. The potential of the new high

resolution RapidEye image with unique band-settings is evaluated in estimating intra-and-inter-species AGB, using two advanced machine learning algorithms (SGB and RF).

1.7.3 Regional Scale Mapping

1.7.3.1 Chapter Five

Although the use of the high resolution RapidEye sensor yields better AGB estimates, this data set has a narrow swath-width and is spatially restricted to a localized scale application. In order to ensure a regional scale mapping and monitoring of forest biomass, satellite data with a wide swath-width is needed. This chapter focuses on identifying cheap, readily-available and appropriate remote sensing data sets for mapping and monitoring AGB over regional scales in resource-constrained areas. The chapter examines the utility of the newly-launched 30-m Landsat-8 OLI sensor with improved signal-to-noise ratios, a wide swath width (185 km) and a 16-day revisit time, for timely and regional scale AGB estimation for three plantation forest species. The study further establishes if the Landsat-8 OLI's "push-broom" design can enhance the detection of the most important forest structural properties, when compared to the whiskbroom Landsat 7 ETM+.

1.7.3.2 Chapter Six

As observed in Chapter Five, the use of Landsat-8 OLI-derived spectral bands and vegetation indices in AGB estimation results in slightly weak AGB estimates, due to saturation challenges at high density canopies. It is important to devise methods that can help to improve the estimates based on this sensor, particularly in resource-constrained areas, given the cost and the sensor's spatial fidelity. This chapter therefore concentrates on improving plantation forest AGB estimation accuracy, using Landsat-8 OLI-derived texture metrics. The accurate quantification of forest AGB, using cheap and readily-available data sets, is fundamental for carbon accounting and ensuring sustainable forest management. The models in this chapter are informed by the findings obtained in Chapter Five.

1.7.4 Forestry Application and Synthesis

1.7.4.1 Chapter Seven

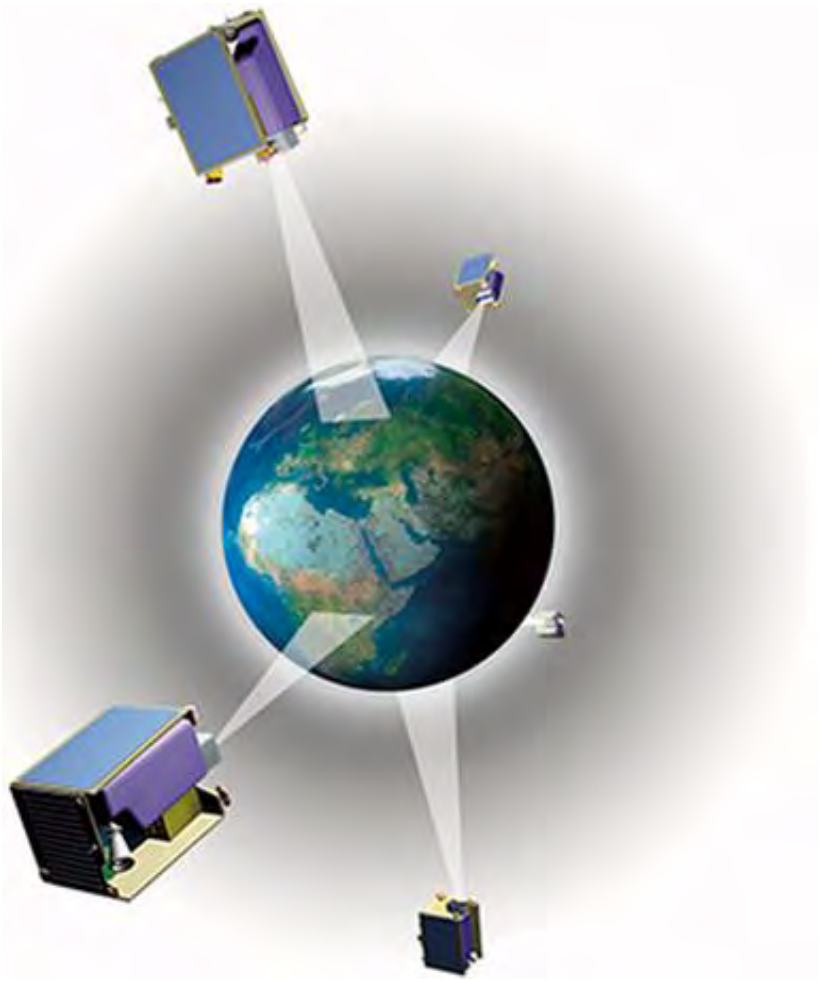
Having successfully identified and developed robust methods for stand volume and AGB computations, particularly for applications in resource-scarce areas, it would be important to use these methods to further quantify forest aboveground carbon stocks in resource-scarce areas. This information is critical, because it indicates the clear and direct role of forest ecosystems in the

global carbon cycle. The objective of this chapter is therefore to quantify the variability of plantation forest aboveground carbon stocks across different tree species, structural attributes (*i.e.* stems, barks, branches and leaves), as well as across different age groups, using remotely sensed data. The chapter focuses on estimating the AGB of stems, barks, branches and leaves, derived from models that have been developed, based on remotely sensed data. Statistical analysis is applied in determining the variability across different structural attributes.

1.7.4.2 Chapter Eight

This chapter provides a synthesis of the findings and conclusions drawn, based on the preceding chapters. It makes further recommendations for future research, based on the highlighted limitations of this study. Finally, a single reference list is provided at the end of the thesis.

2. QUANTIFYING ABOVEGROUND BIOMASS IN AFRICAN ENVIRONMENTS: A REVIEW OF THE TRADE-OFFS BETWEEN SATELLITE DATA AVAILABILITY, COST AND PREDICTIVE ACCURACY



This chapter is based on:

Dube. T, Mutanga, O. Ismail, R. In press. Quantifying aboveground biomass in African environments: A review of the trade-offs between satellite data availability, cost and predictive accuracy, *Journal of Tropical Ecology,* 58, Issue 2.

Abstract

Increased global recognition of the role of forests in regulating the biosphere-atmospheric carbon cycle through carbon sequestration, has resulted in a wide range of scientific studies on the estimation, mapping, monitoring and prediction of AGB at various scales in sub-Saharan Africa. In many parts of the developing world, specifically in sub-Saharan Africa, the accurate quantification of AGB, although still a challenge, is important for national carbon accounting, REDD+ project payments, sustainable forest management and strategic policy-making. In this review, an overview of remote sensing applications for AGB estimation in sub-Saharan Africa, including research challenges and basic information related to the trade-offs between satellite data availability, cost and predictive accuracy is provided. It is assumed that this review is timely, due to a relative increase in the number of remotely sensed forest carbon studies in the recent years (specifically the period between 1998 and 2013). Remotely sensed data is particularly appealing, due to its robustness, instantaneity and repeated spatio-temporal coverage and hence, its ability to successfully estimate and map AGB. However, estimation accuracy and image acquisition cost vary with sensor resolution and type and this has largely restricted AGB estimation in Africa. It is assumed that this study will provide guidance for future national carbon accounting studies, which is one of the main objectives of the Kyoto Protocol and the REDD+ (Reducing Emissions from Deforestation and forest Degradation) project, housed under the United Nations Framework Convention on Climate Change (UNFCCC), particularly for the developing world.

Keywords: accessibility, biomass, carbon stocks, forest plantation, predictive error, remote sensing, trade-offs

2.1 Introduction

Forest aboveground biomass is a measure of tree or shrub cumulative Net Primary Productivity, (NPP). The mapping of forest AGB, in terms of their status and development, is becoming increasingly important, due to the growing need to improve the conservation of forests in the face of a changing climate. Forest ecosystems have been recognized by the Inter-Governmental Panel on Climate Change through the Kyoto Protocol, as capable of mitigating climate change problems. Therefore, the measurement and monitoring of AGB has become an important topic in international climate change negotiations. For instance, the 11th Conference of Parties of the UNFCCC under the Kyoto Protocol initiated the REDD+ project in developing countries. The main aim of the REDD+ project was to highlight the need for possible climate change mitigation measures through sound forest conservation actions in developing countries (UNFCCC, 1998a; 2001; 2011). In sub-Saharan Africa thus far, few countries, have fully embraced the aims and objectives highlighted by the Kyoto Protocol, which advocates sustainable forest (*i.e.* both indigenous and emerging plantation forests, namely, the *Pinus, Eucalyptus* and *Acacia spp.*), management and assessing their contributions to biosphere-atmospheric carbon cycles, as potential carbon sinks. Studies conducted elsewhere in the world (*e.g.* South America) demonstrate that forest plantations occupy a significant spatial extent and are capable of storing meaningful amounts of atmospheric carbon content (le Maire *et al.,* 2011).

The recognition of forests as a potential sink of atmospheric carbon content has resulted in many AGB quantification methods (*i.e.* direct and indirect methods) being developed (Chinembiri *et al.,* 2013b; Henry *et al.,* 2011). Direct AGB estimation methods are broadly classified into (i) Tier-1: basic methods, based on generalized equations, (ii) Tier-2: intermediate approaches, based on volume equations and wood gravity, and (iii) Tier-3: complexity methods, based on biomass equations (Henry *et al.,* 2011). However, the lack of suitable parameters is one of the major issues and challenges associated with the direct methods of estimating and mapping AGB or carbon in places, such as sub-Saharan Africa and south-east Asia. In sub-Saharan Africa, numerous studies have utilised traditional methods (*i.e.* direct methods) to estimate AGB (Dovey, 2009; Henry *et al.,* 2011; Schönau and Boden, 1982a). Most parameters for estimating AGB (*i.e.* allometric equations, wood density values, yield tables and biomass expansion factors) have, however, been derived from studies performed outside Africa, in countries, such as Costa Rica, Brazil and Mexico (Henry *et al.,* 2011). Since most allometric equations were developed outside sub-Saharan Africa, the major challenge is finding ways to implement these parameters in Africa, with limited uncertainities.

Consequently, very few published sources exist for forest types in sub-Saharan Africa (Henry *et al.,* 2011). In addition, besides the scarcity of suitable and key parameters, traditional methods are environmentally destructive and impractical for large-scale implementation. Moreover, these methods require intensive field work and large volumes of ancillary data for analysis, which are labor-intensive, costly and time-consuming (Gara *et al.,* 2014b; Gonzalez *et al.,* 2010; Henry *et al.,* 2011). When using traditional methods, site access in protected areas is also poor, due to complex terrain and organizational restrictions. By using remotely sensed data, these limitations may be addressed in a range of scales and remote sensing technology offers suitable means for the independent verification of the forest carbon pool estimates (Muukkonen and Heiskanen, 2005).

Remote sensing, unlike traditional approaches, provides spatial and temporal data that are useful in mapping AGB at different spatial scales in a more robust, quick and efficient manner (Boyd *et al.,* 1999; Carreiras *et al.,* 2012a). It allows for repeated image acquisitions over the same locations, which are necessary for the detection of temporal changes in carbon stocks. In addition, remotely sensed data are stored in digital format, so that they can be easily integrated with ancillary data in a Geographic Information System (GIS) for further analysis. In the light of these advantages, researchers have used optical sensors (Boyd *et al.,* 1998; 1999; Foody and Boyd, 2002) and active sensors (Carreiras *et al.,* 2012a; 2013; Colgan *et al.,* 2013; Mitchard *et al.,* 2012; 2011; 2009) to estimate AGB in sub-Saharan Africa, with varying degrees of accuracy. Therefore, the utility of remote sensing in estimating AGB necessitates a review of the extent to which the technology has been utilized within the African context. This information is important for sustainable forest management and the identification of readily-available data sets for the accurate estimation of AGB on a regional scale. The current prevailing economic situation in most countries in sub-Saharan Africa requires cost-effective and accurate methods for quick, accurate and efficient AGB estimation, particularly on a national or regional scale (Gara *et al.,* 2014b). This article therefore seeks to: (i) provide a critical evaluation of the literature on AGB estimation, using different remote sensing platforms, and (ii) review the trade-offs between sensor estimation accuracy and costs.

In order to achieve the above objectives, a variety of keywords were used to gather relevant literature related to AGB from selected peer-reviewed journals. All possible keywords or word combinations, were used and these were performed in the "abstract, title, keywords and topic" to enhance the chances of identifying all AGB studies that have been conducted in Africa. To the best of our knowledge, no study on AGB estimation, using remotely sensed data, was recorded prior to

1998, hence only the literature obtained between the period from 1998-2013 was considered for this study. The keywords used included: “aboveground biomass”, “aboveground carbon stocks”, “forest carbon”, “remote sensing biomass”, “remote sensing forest carbon stocks”, “lidar and biomass”, “tree structural attributes”, “SPOT and biomass”, “stand level biomass or carbon mapping” and “pixel-based carbon estimation”, “aboveground biomass in Africa”, “aboveground biomass sub-Saharan Africa”, amongst others. The remote sensing journals searched included: Remote Sensing of Environment, ISPRS Journal of Photogrammetry and Remote Sensing, IEEE Transactions on Geoscience and Remote Sensing, Applied Earth Observation and Geoinformation, IEEE Geoscience and Remote Sensing Letters, Photogrammetric Engineering and Remote Sensing, IEEE Applied Earth Observations and Remote Sensing, International Journal of Remote Sensing, The Photogrammetric Record, Canadian Journal of Remote Sensing, GIScience and Remote Sensing, Remote Sensing Letters, Journal of Applied Remote Sensing, Sensors and Remote Sensing. Other selected journals searched included: International Journal of Geographical Information Science, Transactions in GIS, Computers and Geosciences, Journal of Spatial Science, International Journal of Digital Earth and Geocarto International, and Ecological Applications. Figure 2.1 shows a visual textual summary of the frequently appearing single terms in this document, with higher frequency results in a larger font size, and vice versa.

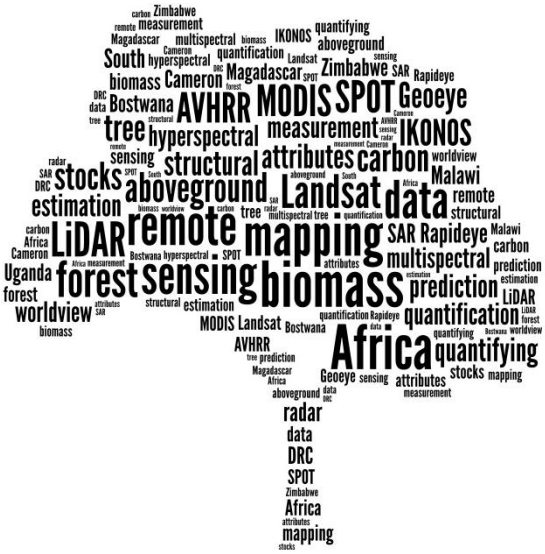


Figure 2.1 A textual summary of this review

2.2 Overview of Aboveground Biomass Studies in Sub-Saharan Africa

To the best of our knowledge, no study has reviewed the literature on the estimation of different forest AGBs in an African environment, using remote sensing. For instance, literature shows that only two studies reviewed forest AGB, based on traditional methods (Gibbs *et al.,* 2007; Henry *et al.,* 2011). However, literature gathered from selected major peer-reviewed remote sensing journals demonstrates a relative increase in AGB studies across the continent (Figure 2.2). For this review, forty-nine (49) publications, with two-thirds of them published in the past decade (2003-2013), were gathered (Figure 2.2). Overall, it was observed that more AGB studies, using remote sensing data sets, were published between 2004 and 2005 and between 2011 and 2013, when compared to the other years. This increasing trend of publications is expected to grow in the near future, making this a critical time to provide a detailed overview of the already available literature on AGB.

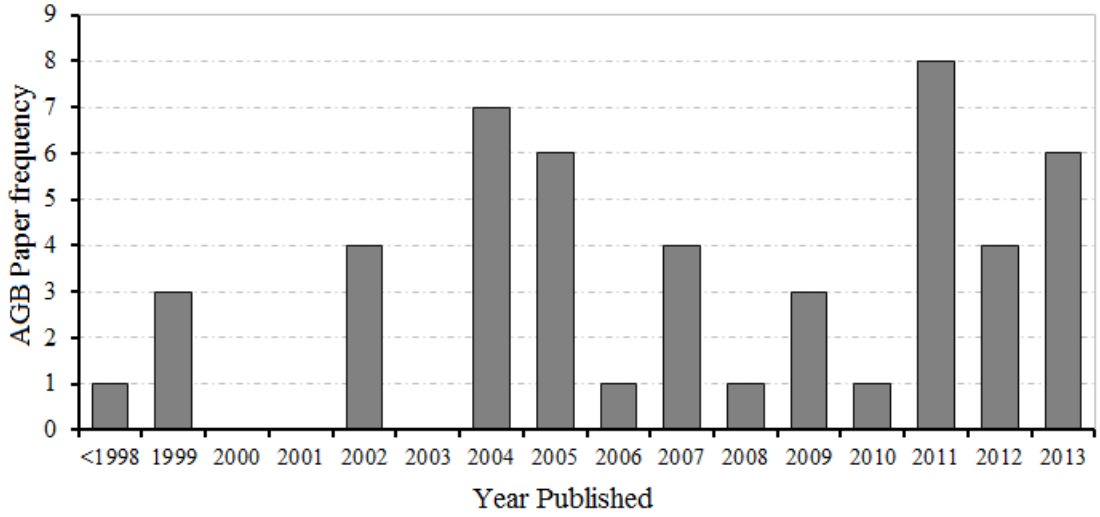


Figure 2.2 Growth of remote sensing popularity in AGB mapping in sub-Saharan Africa between 1998 and 2013

Based on the reviewed literature (Table 2.1), most AGB studies utilised coarse multispectral (31.25%) and radar (22.92%) data sets. Only one study by Thenkabail et al. (2004a) utilised hyperspectral data sets. Table 2.1 shows that 70% of the total AGB studies were conducted in West Africa, particularly in Cameroon, Benin and Ghana. Two-thirds of these studies were conducted in Cameroon, due to the presence of expansive, continuous and highly heterogenous plantations and moist, semi-evergreen regenerating and mature tropical forests (above 50 000 km^2), as well as the availability of direct foreign funding from the Terrestrial Iniative in Global Environmental Research (TIGER) programme of the Natural Environmental Research Council in the United Kingdom (UK) for mapping forest carbon within the country (Boyd *et al.,* 1998, 1999).

The remaining 30% were conducted in the central and southern African countries, namely; Gabon, the Democratic Republic of Congo, Uganda, Mozambique, Madagascar, Namibia, South Africa and Zimbabwe (Carreiras *et al.,* 2012a; 2013; Colgan *et al.,* 2013; Cryus and Tanja, 2004; Eckert, 2012b; Mitchard *et al.,* 2012). Lidar was used to estimate AGB in countries like South Africa and Gabon (Colgan *et al.,* 2013; Mitchard *et al.,* 2011) and radar in Cameroon, Guinea-Bissau, Mozambique and Uganda (Carreiras *et al.,* 2012a; *et al.,* 2013; Mitchard *et al.,* 2012). Mulitispectral data sets have been applied in countries, such as Zimbabwe (Cryus and Tanja, 2004) and Cameroon (Boyd *et al.,* 1998, 1999). All these studies focussed on estimating AGB from indigenous forests. None of them attempted to estimate AGB in commercial forest plantations, despite the vast tracts of land occupied by commercial forests in most of the sub-Saharan countries. It can also be noted that countries like Malawi, Angola, Kenya, Swaziland, Tanzania, Zambia, as well as those in the East, still lag behind in terms of forest AGB estimation. Having provided an overview of AGB studies in Africa, Figure 2.3 provides a detailed spatial representation of AGB studies across sub-Saharan Africa. In general, it can be observed that these studies have been conducted in areas with high vegetation cover, as indicated by the high normalized difference vegetation index (NDVI) values, rather than in areas with low NDVI. However, there is need for future studies to also quantify AGB and carbon stocks in areas charecterised with low vegetation cover, using high spatial resolution sensors that are capable of detecting individual tree species.

Table 2.1 A summary of remotely sensed aboveground biomass average accuracies derived from studies done across sub-Saharan Africa

Sensor type	No. AGB studies	%age studies	Avg. R^2	Avg. predictive error
Active:				
lidar (0.5cm-5m)	8	15.67	0.89	14 %
radar (1cm-10m)	11	22.92	0.74	25 %
Hyperspectral	1	2.08	0.83	-
Multispectral:				
FSR (< 5m)	5	10.42	0.75	27 %
MSR (10m-60m)	9	17.67	0.68	32 %
CSR (>250-1000m)	15	31.25	0.58	42 %
Total	49	100	-	-

**FSR = Fine spatial resolution; MSR = Medium spatial resolution and CSR = Coarse spatial resolution

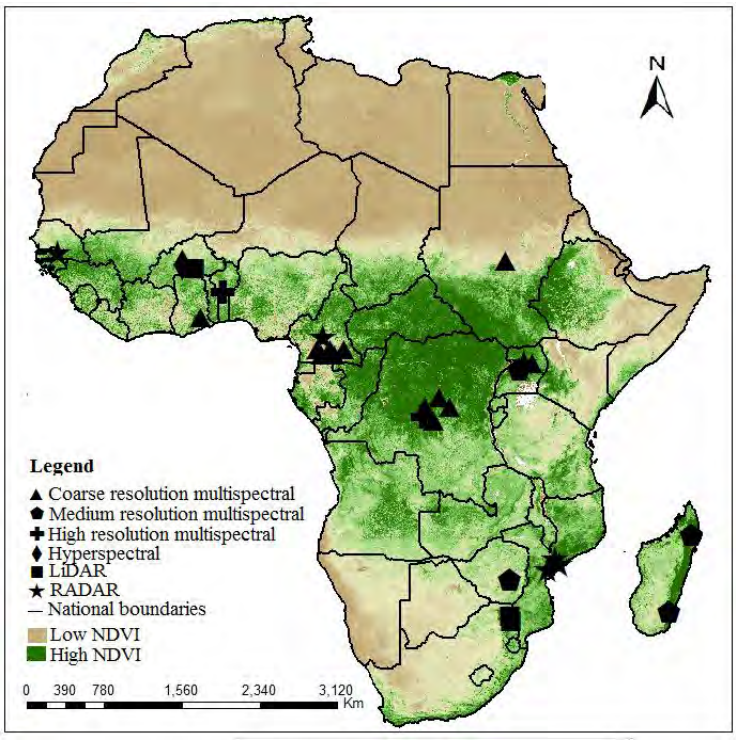


Figure 2.3 Spatial distribution of AGB studies based on remotely sensed data in sub-Saharan Africa derived from the information gathered from selected peer-reviewed journals

2.3 Active Sensors

2.3.1 Aboveground Biomass Estimation Using Lidar and Radar Remote Sensing Sensors

Lidar is an active remote sensing technology that utilizes a laser, which transmits a light pulse towards a target and a receiver, to measure the backscattered or reflected light. This remote sensing data set is regarded as one of the most robust and appropriate means of data collection and has the ability to characterize vertically-distributed forest attributes (Gumindoga *et al.,* 2015). Literature shows that the most important tree-structural attributes necessary for AGB estimation (*i.e.* tree height or forest structural information) can be easily extracted from lidar with great accuracy (Dube and Mutanga, 2014). For instance, work conducted in the developed world demostrates successful AGB estimation from lidar for a single tree, producing high r-square and extremely low root mean square errors (Koch *et al.,* 2010; Patenaude *et al.,* 2005b).

The unique ability provided by the lidar data set has seen few studies testing its strength in estimating AGB in sub-Saharan Africa (Mitchard *et al.,* 2012; 2011). Work by Mitchard *et al.* (2011) is one of the pioneering studies in sub-Saharan Africa. In this study, authors mapped AGB in Gabon with a 25% uncertainty error, using spaceborne lidar (ICESat GLAS footprints) and ground data. The use of generic tropical allometric equations and GLAS data to estimate tree height, however, increased the error term.

In addition, using spaceborne lidar, Colgan *et al.* (2013) successfully (R^2 = 0.91) investigated the influence of hill-slope topography and soil properties in estimating AGB in the Kruger National Park, South Africa. In a different study, Colgan *et al.* (2013) further assessed the accuracy of airborne lidar and field data in estimating AGB at a tree level in the savannas adjacent to the Kruger National Park. The results of their study showed that the use of the object-based model to identify and classify individual tree crowns, as well as to estimate AGB, significantly improved the estimation accuracy, compared to the use of existing airborne methods. Overall, lidar-based AGB studies demonstrate an enhanced capability of quantifying AGB at plot level (in sub-Saharan Africa), especially when compared to estimates derived using passive sensors, such as Landsat products.

Similar to lidar, radar systems are active remote sensors (*i.e.* they independently generate their own source of energy during image acquisition process). These sensors are operational between approximately one cm and 10 m for VHF5 in the microwave portion of the electromagnetic

spectrum (Patenaude *et al.,* 2005b). This unique characteristic permits radar systems to capture images without any obstruction, due to moderate precipitation to cloudy conditions (Koch, 2010; Patenaude *et al.,* 2005b). In Africa, for instance, radar data is increasingly becoming popular in mapping AGB, specifically in tropical environments, with extreme weather conditions (Carreiras *et al.,* 2012a; 2013). Thus far, a review of the literature shows that a limited number of AGB research studies have been conducted across sub-Saharan Africa using radar data sets (Carreiras *et al.,* 2012a; 2013; Mitchard *et al.,* 2012). For example, based on Advanced Land Observing Satellite (ALOS), Phased Array L-band Synthetic Aperture Radar (PALSAR) and two regression algorithms (*i.e.* bagging stochastic gradient boosting (BagSGB) and semi-empirical models), Carreiras et al. (2012a) successfully estimated AGB in Guinea-Bissau. This study yielded a very high correlation coefficient (R^2) value of 0.95 and a low root mean square error (RMSE) value of 26.62 Mg ha^{-1}. In addition, Dube (2013), using ALOS PALSAR, detected AGB losses and gains in the Miombo woodland ecosystems in Mozambique. In a different study, Mitchard *et al.* (2009) examined the relationship between field-measured AGB and cross-polarized radar backscatter values, with high prediction accuracies. However, higher saturation levels from ALOS PALSAR were noted and this was attributed to the structural features of African savannas (Mitchard *et al.,* 2009). Similar observations were noted by Mitchard et al. (2011), who investigated the relationship between the radar backscatter (*i.e.* ALOS PALSAR HV) and AGB plots, with a high R^2-value of 0.86 and a RMSE varying between 25 - 40%.

Although the above studies have successfully attempted to estimate AGB in different parts of sub-Saharan Africa, using lidar and radar data sets, their main limitation is that all of them were implemented at a local or small scale. AGB estimates are currently required at regional or global scale, because "wall-to-wall" estimates are more effective in providing a comprehensive understanding of the global carbon pool than local-scale. Unlike other parts of the world (*i.e.* the developed world), the application of these data sets on a regional or global scale remains one of the major challenges in sub-Saharan Africa. The main reasons for this limitation is the cost, the scarcity of data for operational applications and limited image pre-processing technical expertise, amongst others. Although AGB estimates from lidar, hyperspectral and radar data sets are more accurate and reliable, when compared to those derived using traditional multispectral data sets, such data sets may be restricted spatially to smaller areas. For this reason, there is a need for future studies in sub-Saharan Africa, to develop and identify appropriate sampling techniques that can help enhance the AGB estimation accuracy at a minimal cost within the regions.

2.3.2 Aboveground Biomass Estimation Using Hyperspectral Remote Sensing Data

Recent advances in sensor technology (*i.e.* the presence of unique and strategically-positioned bands) have enhanced AGB monitoring in regions with high forest canopy closure or high biomass. This enables the sensor to accurately characterize different forest-structural properties at optimal temporal and spatial scales. This has consequently resulted in the great appreciation and use of the data sets for AGB estimates, particularly outside sub-Saharan Africa, given that traditional multispecral sensors suffer from saturation problems.

Nonetheless, the utility of hyperspectral remote sensing platforms for estimating AGB in sub-Saharan Africa is still developing and its usage has gained limited acceptance for operational use, due to costs and restricted accessibility (Dube *et al.,* forthcoming; Gara *et al.,* 2014b; Thenkabail *et al.,* 2004a). Thus far, only one AGB study has been conducted, using the hyperspectral remote sensing sensor (Thenkabail *et al.,* 2004a). The study found that Hyperion narrowband data yielded good models, which explained between 36–83% more of the variability in rainforest biomass, when compared to Landsat products and Ikonos, which explained between 13–60%. The study further concluded that the use of narrowband Hyperion data has a greater advantage and strength over broad-band Ikonos and ETM+ data sets in estimating AGB in the rainforest vegetation.

2.4 Passive Sensors

2.4.1 Aboveground Biomass Estimation Using Old Generational Multispectral Sensors

Coarse, medium and fine spatial resolution multispectral images have since become more popular and attractive for estimating AGB in resource-constrained sub-Saharan Africa (Boyd *et al.,* 1998, 1999; Eckert, 2012a; Foody and Boyd, 2002; Wu *et al.,* 2013) and in other parts of the world. Coarse-to-medium multispectral data (*i.e.* MODIS, Landsat, ASTER and SPOT) are currently provided freely or are available at a low cost (Table 2.1 and 2.2). Some of these sensors (*i.e.* Landsat products) provide images with a wide swath-width (above 185 km) and have a repeated global coverage, which is necessary for continuous AGB assessment. These unique characteristics have resulted in the sensors gaining more popularity for regional or global AGB mapping (Boyd *et al.,* 1998; 1999; Foody and Boyd, 2002). Currently, there are huge volumes of archived data sets (*i.e.* NOAA AVHRR, Landsat products and MODIS) with large spatial coverage and, dating back to 1972, when the first Earth Resources Technology Satellite (ERTS-1) Landsat 1 was launched by the National Aeronautics and Space Administration.

In sub-Saharan Africa thus far, numerous AGB studies have been conducted, using the afore-mentioned sensors, with fairly reasonable accuracy (Boyd *et al.,* 1998; 1999; Foody and Boyd, 2002). Using the middle infrared (MIR) wavelengths (3.0 - 5.0μm) of the National Oceanic and Atmospheric Administration-Advanced Very High Resolution Radiometer (NOAA-AVHRR), Boyd et al. (1998) successfully estimated tropical forests AGB in Cameroon. This study demonstrated the potential of exploiting under-utilised electromagnetic portions between near infrared (NIR) and thermal infrared (TIR) wavelengths, when estimating AGB. The findings from the above study are also confirmed by Boyd *et al.* (1999), whose work demonstrated a strong relationship between total AGB, MIR reflectance and vegetation indices in Cameroon. The two studies concluded that MIR spectral wavelengths are more sensitive to changes in forest properties than the visible and NIR reflectances. Carreiras *et al.* (2012), using canopy cover derived from the MODIS NDVI data (MOD13Q1) time-series, estimated AGB with a high coefficient of determination (R^2-value = 0.91). AGB estimation accuracies obtained by Carreiras *et al.* (2012) are comparable to those of Carreiras *et al.* (2012a) and Mitchard *et al.* (2009), which were derived using lidar and radar data sets. Mitchard *et al.* (2009) concluded that the proposed approach was relevant for operational use in the tropical savannah woodlands and that it contributes to the low-cost and large-scale assessment and monitoring of forest carbon fluxes in sub-Saharan Africa.

Although the above AGB studies, utilising multispectral data, demonstrated reasonable estimation accuracies, their application at plot level or in densely-forested areas, imposes serious challenges. For instance, the application of medium-to-coarse spatial resolution multispectral sensors for AGB estimation in the tropics is neither feasible nor accurate, due to spectral saturation challenges associated with dense canopy closure (Thenkabail *et al.,* 2004a). The reflectance signal in the visible and NIR is mostly correlated with the green leaf area index (LAI) and canopy cover of the vegetation. Furthermore, AGB generally becomes decoupled from LAI after a given stand age, as AGB continues to increase after canopy closure. The decoupling from LAI after a given stand age hinders accurate AGB estimation, when using multispectral sensors. The utility of all Landsat satellite products and MODIS for estimating AGB has thus, to a large extent, resulted in inaccurate results in the tropics, due to the presence of mixed pixels and the mismatch between pixel-size and field-plot area (Gibbs *et al.,* 2007; Thenkabail *et al.,* 2004a). Given the poor financial situations in sub-Saharan Africa and the immediate need for AGB estimates on a regional scale, the use of cheap or readily-available and suitable sensors with appropriate optimal spectral characteristics is necessary, if this objective is to be achieved.

For continuous and accurate large-scale AGB estimation, there is a need for the remote sensing research community and environmentalists to shift towards embracing the newly-launched, readily-available and cheap multispectral sensors (*i.e.* 30-m Landsat-8 OLI sensor, 5-m RapidEye, Sentinel-2, SPOT-6 and 2-m Worldview-2). This emerging group of new generation sensors, with refined radiometric, spatial, spectral and temporal properties, presents new opportunities for the detection and segmentation of the structural attributes of forest. Literature shows that these sensors provide more granular measurements of plant productivity, which is ideal for accurately estimating AGB in resource-limited environments (Eckert, 2012a; Gara *et al.,* 2014b; Mutanga *et al.,* 2012).

Currently, the utility of these sensors has not been fully explored. For instance, only a single study estimated forest AGB, using Worldview-2 image in sub-Saharan Africa (Eckert, 2012a), as shown in Table 2.1. The results of the study indicated a higher correlation between image texture parameters and AGB, when compared to spectral information. Mutanga et al. (2012), demonstrated the utility of three normalized difference vegetation indices (NDVIs) computed from the WorldView-2 red edge and NIR bands in the estimation and mapping of wetland vegetation AGB in South Africa. An average biomass of 3.44 kg m^{-2} and RMSE of prediction of 0.441 kg m^{-2} was obtained, which was much higher than the standard AGB saturation level reported in literature. This task was previously challenging, when using broad-band satellite imagery. The authors finally concluded that the indices calculated from the additional spectral bands of Worldview-2 performed better than the traditional multispectral indices, due to their enhanced sensitivity to vegetation properties, namely, canopy AGB and chlorophyll content. This observation is confirmed by Kokaly and Clark (1999), who showed that vegetation indices, calculated from the hyperspectral red-edge and NIR spectral bands, minimize atmospheric and water absorption influence, as well as soil background, thereby enhancing estimation accuracy, a characteristic of some new generation sensors. Despite the promising performance from the new generation sensors, thus far little is known on AGB estimation, using new generation sensors, such as RapidEye, Landsat-8 OLI, SPOT-6 and Sentinel-2. It is thus critical to explore the possibility of utilizing these sensors for AGB estimation, as well as investigating the potential of them being up-scaled to landscape level, to meet the objectives of the Kyoto Protocol and REDD+. Studies conducted outside sub-Saharan Africa have already shown improved forest AGB estimates from Worldview-2 imagery (Ozdemir and Karnieli, 2011; Shamsoddini *et al.,* 2013). Ozdemir and Karnieli (2011), using WorldView-2 texture metrics to derive different forest structural parameters in Israel, concluded that this data set offers new opportunities for AGB estimation. Shamsoddini *et al.* (2013) mapped pine tree structural

parameters from WorldView-2 data in New South Wales, with an estimation error of 30%. The results demonstrated variability in the sensitivity of the two separate NIR bands over numerous tree structural parameters and this shows the utility of new bands and provides the opportunity to develop new indices for the better estimation of AGB.

Although the new generation sensors indicate a very promising opportunity for regional scale AGB estimation in resource-limited environments, they do have some limitations. The major limitations of the new generation sensors include: (i) their inability to penetrate forest canopies in order to provide information on critical reflected solar energy, and (ii) the ineffectiveness to map AGB in tropical regions, where cloud cover is often a challenge. Like their predecessors, new generation multispectral sensors provide two-dimensional views of forest canopy surfaces. This ability only allows spectral properties from these sensors to relate more appropriately to the percentage canopy cover and LAI, rather than directly to AGB. The lack of the most important shortwave infrared (SWIR) band (except for Landsat-8 OLI and Sentinel-2 sensors), which has proved important for AGB estimation, especially in boreal forests (Lu, 2006), also remains one of the major challenges. Many studies have demonstrated the relevance of SWIR in volume and AGB mapping. For example, work by Hyyppä *et al.* (2000) has demonstrated that the SWIR spectral range has the highest correlations of all the examined forest attributes. Muukkonen and Heiskanen (2005) also found the strongest correlation between forest AGB and the ASTER SWIR spectral band. It would therefore be important to ensure that the development of future generation sensors includes this band, since it is more critical in AGB estimation.

2.5 Trade-offs Between Satellite Data Availability, Costs and the Predictive Accuracy

Research shows that active sensors, such as lidar and radar, provide high accuracy for estimating AGB, with an average R^2-value of ±0.74 and ±0.89, when compared to multispectral data sets. It can also be noted in Table 2.1 that only one study tested the utility of hyperspectral data in estimating AGB in sub-Saharan Africa. The limited number of studies using this sensor can be largely attributed to the cost, availability, as well as the Hughes phenomenon, or "the curse of dimensionality", associated with this data set. Although, lidar and radar data provide accurate and reliable AGB estimates, due to their cost and limited availability (Table 2.1 and 2.2), the two data types have since gained more popularity for local scale-based applications (*e.g.* small-scale: <10 000 ha), rather than "wall-to-wall" mapping (± 56 000 km^2) in sub-Saharan Africa (Carreiras *et al.,* 2012; Colgan *et al.,* 2013). The existing lidar and radar-based AGB studies in sub-Saharan

Africa targeted mainly small-scale areas, despite the REDD+ project requiring regional or global scale estimates. The full application of these expensive and complex sensors in sub-Saharan still remains a challenge.

It is therefore important to establish sensors and methods that are cheap, fast and robust for application in developing countries, in order to be able to meet the requirements of the REDD+ project. However, not all African countries have the capacity to access funding and hence cannot use active sensors for estimating AGB. It is therefore essential that the systems for measuring and estimating AGB for REDD+ projects in sub-Saharan Africa are available at a low cost, with acceptable and reliable accuracy. As a result, the cheap, free and readily-available multispectral data sets remain crucial for AGB estimation in sub-Saharan Africa. The moderately high average R^2- value (±0.68) and slightly lower average predictive errors (±32%) in AGB results, as shown in Table 2.1, indicate the potential strength and possible avenues for improving regional estimates, using medium resolution sensors. Currently, most of these multispectral sensors (*i.e.* MODIS and Landsat products) are readily accessed from the NASA and USGS Earth Resources Observation and Science (EROS) Center archive (http://earthexplorer.usgs.gov/), dating back to 1972. In the future, the European Space Agency (ESA) will also make their images freely-available, covering wide expanses of land area, which is important for regional or global AGB mapping. Further improvement is shown by fine spatial resolution multispectral sensors with higher predictive averages, almost comparable to those obtained using radar and lidar data sets (Table 2.1). The fine spatial resolution multispectral sensors are, for instance, available at a fairly reasonable cost, especially when compared to lidar, hyperspecral and radar data sets (Table 2.2).

Table 2.2 Overview on trade-offs between aboveground biomass sensor predictive accuracy and costs

Sensor resolution	Available satellite data	Utility for mapping	Acquisition Cost
Very fine (between 0.5cm-10cm)	Lidar, radar Hyperspectral	Pixel scale	Highly expensive
Fine (<5m)	Ikonos, Quickbird, RapidEye, Worldview-2	Validation at a localised scale	High
Medium (10-60 m)	SPOT Landsat	Small and large scale mapping of aboveground biomass	Low or free
Course (250-1000m)	AVHRR MODIS	Large scale estimation and mapping of aboveground biomass	Free

*Multispectral optical sensors despite accuracy problems have been considerably utilised operationally in estimating and mapping aboveground boimass in Africa. Active sensors, such as lidar and radar (*i.e*. ALOS PALSAR L-band cross-polarised (HV) radar data ALSOR PALSAR), are not yet used operationally for aboveground biomass estimation in an African environment, except for small-scale use only.

Previous studies conducted outside sub-Saharan Africa demonstrate a greater potential of employing multi-source (*e.g.* multispectral data with ancillary data sets) or multi-date data sets on multispectral platforms, when estimating AGB (Baccini *et al.,* 2004; Dube *et al.,* in press). Multi-date multispectral data perform well in discriminating AGB classes and allow the examination of the spatial and temporal contribution of forests to carbon sequestration (le Maire *et al.,* 2011). In addition, the radiometric and temporal variations can contain new information about forest characteristics important for AGB estimation (le Maire *et al.,* 2011; Powell *et al.,* 2010). To the best of our knowledge, few investigations have been carried out using multi-date techniques for AGB estimation and most of these have been applied outside Africa. For instance, a recent study by le Maire et al. (2011) in Brazil showed that it is beneficial to use multi-temporal data sets in estimating AGB (R^2-value of 0.82) at global or regional scales. The study accurately demonstrated the use of MODIS products (Vegetation Indices 16-Day L3 Global 250 m) and ancillary data in monitoring AGB over 15 000 ha of *Eucalyptus* plantations, using stepwise linear or nonlinear (Random Forest) regression models. The integration of ancillary data, such as tree age, with multispectral data, enhanced the AGB estimation and prediction accuracy. However, in Africa, multi-source data sets have not been fully embraced in AGB studies.

2.6 Future Investigation

Although considerable progress has been made in sensor development and the application of remote sensing technology for AGB estimation, it still remains a challenge. Firstly, to the best of our knowledge, no work has been conducted, based on remote sensing data sets, to estimate and map forest plantations AGB in sub-Saharan Africa. Plantation forests, such as the *Pinus*, *Eucalyptus* and *Acacia* spp, occupy quite a significant portion of the land area and house a significant amount of unknown carbon, which plays a vital role on biosphere-atmospheric carbon fluxes. However, for accurate, unbiased and reliable national carbon accounting to meet the objectives of the Kyoto Protocol and the REDD+ project, the focus of future research must be directed towards estimating and mapping AGB across all terrestrial ecosystems in sub-Saharan Africa.

Meanwhile, to achieve the above objective, the challenge still exists to develop or identify effective, cheap, accurate and operational techniques to estimate and map AGB on a large-scale, capitalizing on the improved sensor characteristics and processing techniques. A new crop of studies, focusing on the use of remote sensing in forest ecology or vegetation mapping, demonstrates the adoption of new tree-based statistical ensemble methods (*i.e.* machine learning algorithms), namely, stochastic gradient boosting and random forests, as suitable for successful and improved estimation accuracies, when using the readily-accessible multispectral sensors (Carreiras *et al.,* 2012; Mutanga *et al.,* 2012) and when combined with ancillary data (Baccini *et al.,* 2004; Dube *et al.,* in press; le Maire *et al.,* 2011). Although active sensors provide optimal AGB estimates, possible variations and changes over time and space cannot be satisfied in the sub-Saharan Africa, due to available economic constraints and other related challenges, mentioned earlier. Challenges posed by active sensors for successful application in sub-Saharan Africa are a clear indication that optical sensors remain the main possible solution for AGB quantification across the continent (Dube *et al.,* 2015; Powell *et al.,* 2010). The major challenge that researchers have to resolve when dealing with multispectral sensors is improving their direct relationship with AGB, which is difficult to establish in areas characterized with wood volumes exceeding 100 m^3 ha^{-1}. There is, therefore, a need for future research, to move towards the integration of multispectral remotely sensed with ancillary data sets.

Another promising future research challenge would be to test the utility of new generation moderately fine spatial-resolution multispectral sensors (*i.e.* Landsat-8 OLI, RapidEye, Sentinel-2

and Worldview-2), which are moderately cheap and readily available. The WorldView-2 optical satellite sensor, launched in October 2009, provides panchromatic data at a geometric resolution of 0.5 m and multispectral data, divided into eight spectral bands, at a geometric resolution of 2 m. Furthermore, future research in sub-Saharan Africa should focus on testing the potential of new generation sensors as a substitute for active sensors, because of their enhanced spatial resolution and the increased number of bands. Moreover, some of these sensors are comprised of an increased number of spectral bands, together with key and strategically-positioned vegetation wavelengths (*i.e.* the red edge bands: 690–730 nm). These strategically-positioned bands are not currently available on multispectral satellite sensors, except for hyperspectral sensors that are expensive to acquire and also require complex pre-processing methods. In addition, more research is needed to find the best variables (ancillary data) and predictive models that can be integrated with cheap, and sometimes free, multispectral data sets. Overall, the AGB estimation, using coarse spatial-resolution data, is still very limited because of the common occurrence of mixed pixels and the huge difference between the size of field measurements and image pixel size. This results in difficulties in the integration of ground-based sample data and remote sensing-derived variables. A synthetic analysis of multi-source data, with a combination of different modelling approaches, may be needed for accurate AGB estimations in a large area.

2.7 Conclusion

Literature demonstrates that there is a decline in the number of studies using conventional methods to estimate AGB, compared to remote sensing methods. Conventional methods, although accurate, are time-consuming, too costly and practically impossible to apply on a broader scale. Although active sensors, such as lidar and radar, provide higher and more reliable AGB estimates than coarse multispectral data, they are still not operational in the African environments, due to the cost of their acquisition. Given the poor economic situation of most sub-Saharan African countries, multispectral data still remain relevant for AGB quantification, regardless of saturation problems in densely-closed canopies, the occurrence of mixed pixels and a huge mismatch between the size of field measurements and the image pixel size. Previous work outside sub-Saharan Africa shows that AGB estimates can be greatly improved by the use of multi-date multispectral data sets, the integration of remotely sensed data with ancillary data and spectral decomposition. Therefore, there is a need for further investigations into the applicability of the above approaches in quantifying stand volume and AGB in sub-Saharan Africa, using new generation sensors, particularly in

managed plantation forests. In addition, there is a need to identify efficient and robust predictive models that can help improve stand volume, AGB and carbon stock estimates from these data sets.

CHAPTERS THREE AND FOUR

LOCAL SCALE MAPPING

3. *EUCALYPTUS* STANDS VOLUME PREDICTION USING STATISTICAL METHODS AND WITH MULTISOURCE DATA

This chapter is based on:

Dube, T, Mutanga, O, Abdel-Rahman, E, Ismail, R, Slotow, R. 2015. Predicting *Eucalyptus spp.* stand volume in Zululand, South Africa: An analysis using stochastic gradient boosting regression ensemble with multisource data sets. *International Journal of Remote Sensing.* **33,** 4502-4526**.**

Presented at **African Association of Remote Sensing of the Environment** 31-10-2014, University of Johannesburg, South Africa.

Abstract

Accurate, reliable and up-to-date forest stand volume information is a pre-requisite for a detailed evaluation of commercial forest resources and their sustainable management. Commercial forest responses to global climate change is still uncertain, hence mapping stand volume as carbon sinks is fundamentally important in order to understand the role of these forests in stabilizing the effects of climate change. The aim of this study was to examine the utility of stochastic gradient boosting (SGB) and multisource data to predict the stand volume of a *Eucalyptus* plantation in South Africa. The SGB ensemble, random forest (RF) and stepwise multiple-linear regression (SMLR) were used to predict *Eucalyptus* stand volume and other related tree-structural attributes, such as mean tree height and mean diameter at breast height (DBH). Multisource data consisted of SPOT-5 raw spectral features (4 bands), 14 spectral vegetation indices, rainfall data and stand age. When all variables were used, the SGB algorithm showed that stand volume can be accurately estimated (R^2 = 0.78 and RMSE = 33.16 m^3ha^{-1} (23.01% of the mean)). The competing RF ensemble produced an R^2 value of 0.76 and a RMSE value of 37.28 m^3ha^{-1} (38.28% of the mean). SMLR, on the other hand, produced an R^2 value of 0.65 and an RMSE value of 42.50 m^3ha^{-1} (42.50% of the mean). Our study further showed that *Eucalyptus* mean tree height (R^2 = 0.83 and RMSE = 1.63 m (9.08% of the mean)) and mean diameter at breast height (R^2 = 0.74 and RMSE = 1.06 (7.89% of the mean)) can also be reasonably predicted, using SGB and multisource data. Furthermore, when the most important SGB model selected variables were used for prediction, the predictive accuracies improved significantly for mean DBH (R^2 = 0.81 and RMSE = 1.21 cm (6.12% of the mean)), mean tree height (R^2 = 0.86 and RMSE = 1.39 m (7.02% of the mean)) and stand volume (R^2 = 0.83 and RMSE = 29.58 m^3ha^{-1} (17.63% of the mean)). These results underscore the importance of integrating multisource data with remotely sensed data for predicting *Eucalyptus* stand volume and related tree-structural attributes.

Keywords: Eucalyptus plantation, multiple-linear regression, stand age, random forest ensemble, SPOT-5, stand volume, tree-structural attributes

3.1 Introduction

Plantation forests (1.5 million ha) are more prevalent than indigenous forests (530 000 ha) in South Africa (Sagi, 2012). These intensively-managed forests consist mainly of *Eucalyptus* species (700,000 ha), contribute approximately 7.3% (ZAR 20.4 billion) to the gross domestic product (GDP) and employ around 437 400 people (Godsmark, 2010). Besides commercial utilization, plantation forests are environmentally important in regulating bio-geochemical (le Maire et al. 2011) and bio-geophysical processes (Dube and Mutanga, 2014). Through biochemical processes and photosynthesis, forests also extract and house additional atmospheric carbon (Canadell et al. 2007; Henry et al. 2011). This natural phenomenon is viewed as a clean mechanism for mitigating climate change (Canadell *et al.,* 2007). To fully understand forest contributions in the carbon cycle, accurate, reliable and current information is essential (Geldenhuys, 2000). Information more specifically related to mean diameter at breast height, mean tree height and stand volume is central for deriving stand level, or total forest carbon stocks and associated fluxes (Santoro *et al.,* 2006; Somogyi et al.,*et al.,* 2008).

Direct stand volume estimation methods are based on species-specific, or general allometric, equations derived from terrestrial enumerations of related tree structural variables, such as basal area (BA), height and diameter at breast height (DBH) (Chinembiri *et al.,* 2013; Gara *et al.,* 2014). Direct methods, however, are prone to error propagation (Gebreslasie *et al.,* 2010) and are time-consuming, tedious, limited in extent and expensive to implement (Tesfamichael *et al.,* 2010). Remote sensing techniques have been well-documented as a promising alternative to field-based methods for stand or plot level tree volume estimation (Gebreslasie *et al.,* 2010; Tesfamichael *et al.,* 2010).

A number of studies have thus far explored the utility of multispectral (Gebreslasie *et al.,* 2010), hyperspectral (Canavesi *et al.,* 2010) and active imaging sensors (Fransson *et al.,* 2000; Tesfamichael *et al.,* 2010) for determining stand volume or aboveground forest carbon stocks (Henry *et al.,* 2011; le Maire *et al.,* 2011). Using lidar data and the local maxima (LM) filtering approach, Tesfamichael *et al.* (2010) estimated *Eucalyptus* volume with R^2 values, ranging from 0.82 to 0.94, and with root mean square error (RMSE) values, ranging from 20% to 43%. Canavesi et al. (2010) used eight Hyperion EO-1 (total bands = 220) derived vegetation indices to estimate *Eucalyptus* stand volume, based on multiple-linear regression models. Using multiple-linear predictive models, the authors found that 71% of the variation in stand volume can be explained

with a RMSE value of 43.73 m^3 ha^{-1}. Fransson et al. (2000) estimated tree stem volume from airborne CARABAS-II VHF (20–90MHz) synthetic aperture radar (SAR) data, using linear predictive models, and obtained RMSE values between 4% and 22%. Although hyperspectral and lidar data provide accurate and reliable forest stand volume estimates, due to high acquisition costs and limited availability, these data types have gained more popularity for project-based applications (limited extent), rather than for "wall to wall" applications in sub-Saharan Africa and other developing countries (Carreiras *et al.,* 2012). Medium-to-coarse resolution multispectral imagery thus remains crucial for estimating stand volume and related tree-structural attributes, such as mean tree height and mean diameter at breast height (DBH), measured at 1.3 m above the base of the tree, primarily because of data availability at limited or no cost. In additionally, these data types do not necessarily require complex pre-processing and analysis, when compared to hyperspectral and lidar data sets. Given such advantages related to the use of multispectral sensors, there is a need for a swift prediction of stand level tree-structural attributes *i.e.* mean tree height, mean DBH, stand volume and biomass, to facilitate a more accurate and precise assessment of sub-Saharan Africa's carbon cycle (Laurin *et al.,* 2014).

It would be beneficial if the mapping of stand level tree-structural attributes, especially stand volume, were carried out for large areas, in order to understand the role of these forests in stabilizing climate change effects. Medium spatial resolution multispectral sensors thus represent a possible alternative, due to their large coverage (*i.e.* swath width), despite having limitations, such as the presence of mixed pixel information (Sarker *et al.,* 2011). Researchers now advocate incorporating multispectral data with ancillary stand data sets, to obtain more accurate estimates (Baccini *et al.,* 2004; le Maire *et al.,* 2011; Main-Knorn *et al.,* 2011). Moreover, machine-learning techniques, such as SGB, have also proven to be useful in estimating structural parameters, such as volume at plot level, biomass and carbon stocks (Carreiras *et al.,* 2012a; 2013; Pierce *et al.,* 2012; Shataee *et al.,* 2011). For example, Carreiras et al. (2013) used a modified bagging SGB (BagSGB) approach in Mozambique and estimated aboveground forest carbon stocks with an R^2 value of 0.95 and a RMSE value of 5.03 $Mg \cdot ha^{-1}$ from Advanced Land Observing Satellite (ALOS) Phased Array L-band Synthetic Aperture Radar (PALSAR) data. Similarly, Carreiras *et al.* (2012) used a BagSGB approach to derive biomass with an R^2 value of 0.95 and an RMSE value of 26.62 Mg ha^{-1} from ALOS PALSAR data in Guinea-Bissau. Shataee *et al.* (2011) compared random forest (RF) and boosting regression trees (BRT) for plot-level forest volume estimation, using height and intensity metrics data generated from aerial laser scanner (ALS) and Landsat TM 5. The results of

their study demonstrated that BRT is the most effective model, producing plausible volume estimates (RMSE = 40.56%), when compared to RF (RMSE = 42.93%).

However, although tree-based ensemble methods have found popularity in studies that use radar and high spatial resolution multispectral sensors (such as Quickbird-2 and Ikonos-2), their utility with medium spatial resolution multispectral data sets is rather limited. The study by Pierce *et al.* (2012) is an exception. Using RF regression with R^2 values ranging from 0.55 to 0.68, Pierce *et al.* (2012) quantified spatial patterns of surface and canopy fuel loads across a large heterogeneous landscape in California, by integrating plot level data, topographic characteristics and Landsat 5 TM data. Therefore, it would be interesting to test the utility of tree-based ensembles in predicting different tree-structural attributes *i.e. Eucalyptus* stand volume, mean tree height and mean diameter at breast height from medium spatial resolution multispectral data sets and for selecting the optimal set of variables that best explain variations in stand level volume, amongst others. If successful, the accurate prediction of mean DBH, mean tree height and stand volume, based on integrating medium spatial resolution multispectral sensors, such as SPOT-5 (10m pixel resolution), together with ancillary data sets, can thus provide further insight into, and an understanding of African ecosystem carbon balances and regional climatic systems.

Therefore, this study aims to provide an effective and easily-applicable solution for the prediction of *Eucalyptus* volume and related tree-structural attributes (*i.e.* mean DBH and mean tree height) at a stand level. To achieve this goal, the strength of the SGB algorithm, which is a hybrid between boosting (Schapire, 2003) and bagging procedures (Friedman, 2002), was tested in predicting volume and related tree-structural attributes from multisource data (SPOT-5 raw spectral information, vegetation indices and ancillary variables derived from rainfall and stand age). Our study also compared the utility of the SGB ensemble with the RF ensemble and a classical parametric method *i.e.* the stepwise multiple-linear regression model.

3.2 Materials and Methods

3.2.1 Study Area

The *Eucalyptus* plantations currently managed for pulpwood production by the Sappi Pulp and Paper Company in the Zululand region of South Africa (Figure 3.1) were investigated in this study. These *Eucalyptus* plantations occupy approximately 25 840 ha of the area. The growing stock consists predominantly of four- to eleven- year old *Eucalyptus grandis* (*E. grandis*), *Eucalyptus*

grandis × *Eucalyptus camaldulensis* (*E.gxc*) and *Eucalyptus grandis* × *Eucalyptus urophylla* (*E.gxu*) hybrid clones. These plantations are evenly scattered along the coast, stretching from the south around Richards Bay (Latitude 28˚ 48'S, Longitude 32˚ 02'E) to the north, including the town of Mtubatuba (Latitude 28° 25'S, Longitude 32° 10'E), with an altitude averaging between 50 and 100 m above sea level. The climate is subtropical, with high humidity and temperatures (the mean annual temperature is approximately 21.7ºC) and high summer rainfall, reaching ±1200 mm, favourably important for fast growing *Eucalyptus* stands (Scott and Lesch, 1997). Soils are mainly of Aeolian origin (*i.e.* grey and yellowish quaternary sands) and contain little organic carbon; hence, there is a suitable deep rooting medium for the growing stock (Dube *et al.,* 2014b). These fast-growing *Eucalyptus* species are planted as clones or seedlings and harvested every six to seven years. Stands are managed on a pulpwood regime (*i.e.* established at 1667 trees ha^{-1}) and intensive soil preparation and weed control measures are practiced, until crown closure occurs between 1 to 1.5 years. Due to the species composition, intensive silviculture, flat terrain and relatively even soil composition, stands are fairly uniform with regard to crown cover and tree density (Little and Toit, 2003; Viero, 2002).

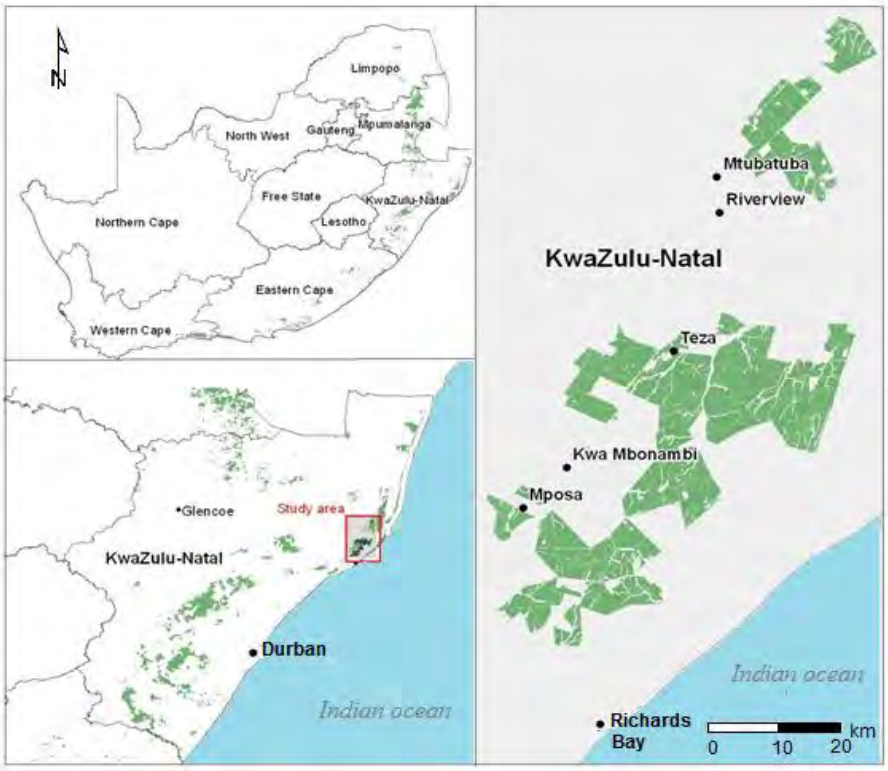


Figure 3.1 The spatial distribution of Eucalyptus plantation forest in the Zululand region of KwaZulu-Natal, South Africa

3.3.2 Image Acquisition

A SPOT-5 image (Scene ID 5 142-407 12/05/14 07:50:26 1 J) covering the study area was acquired on 14 May 2012. SPOT-5 consists of green (500-590 nm), red (610-680 nm) and near-infrared (NIR: 780-890 nm) bands, with a pixel resolution of 10 m, whereas the shortwave infrared (SWIR: 1580-1750 nm) band is captured at a pixel resolution of 20 m, but is resampled to 10 m, using the nearest neighbour method. The image was geo-referenced, using 18 evenly distributed ground control points (GCPs) and a 10 m digital elevation model (DEM). The geographical error in terms of RMSE was less than one pixel. In addition the SPOT-5 image was atmospherically corrected and converted to reflectance values, using the ATCOR 3 module in ERDAS Imagine 2011, in order to reduce the effect of haze and other atmospheric influences.

3.3.3 Multispectral SPOT-5 Spectral Bands and Derived Vegetation Indices

Initially, the SPOT-5 spectral bands and selected spectral vegetation indices (VIs), as shown in Table 3.1, were used to predict *Eucalyptus* volume and related tree-structural attributes (*i.e.* mean diameter at breast and mean tree height) at a stand level. VIs were selected, based on previous studies that predicted tree volume and structural attributes, using remotely sensed data (Lee *et al.,* 2004; Swatantran *et al.,* 2011; Zheng *et al.,* 2004).

Table 3.1 SPOT-5 spectral parameters and vegetation indices used for this study

Parameters	Formula	References
Spectral band reflectance		
Green, Red, NIR and SWIR	-	-
Vegetation Indices		
1. Difference Vegetation Index (DVI)	NIR – Red	Tucker (1979)
2. Global Environmental Monitoring Index (GEMI)	$\eta(1-0.25\eta)\frac{\mathrm{Re}d-0.125}{1-\mathrm{Re}d}$; $\eta=\frac{2\left(NIR^2-\mathrm{Re}d^2\right)+1.5*NIR+0.5*\mathrm{Re}d}{NIR+\mathrm{Re}d+0.5}$	Pinty and Verstraete (1992)
3. Green normalized difference vegetation index (GNDVI)	$\frac{NIR-Green}{NIR+Green}$	Gitelson et al. (1996)
4. Moisture Stress Index (MSI)	SWIR/NIR	Rock et al. (1986)
5. Normalized Difference Vegetation Index (NDVI)	(NIR – Red)/(NIR + Red)	Jordan (1969)
6. Corrected NDVI ($NDVI_c$) $SWIR_{min}$ and $SWIR_{max}$ are the minimum and maximum reflectances observed in the field plots	$\frac{NIR-\mathrm{Re}d}{NIR+\mathrm{Re}d}\left(1-\frac{SWIR-SWIR_{\min}}{SWIR_{\max}-SWIR_{\min}}\right)$,	Nemani et al. (1993)
7. Normalized difference infrared index (NDII)	$\frac{NIR-SWIR}{NIR+SWIR}$	Kimes et al. (1981)
8. Optimized Soil-Adjusted Vegetation Index (OSAVI)	$(1+L)\frac{NIR+\mathrm{Re}d}{NIR+\mathrm{Re}d+L}$; L = 0.16	Rondeaux et al. (1996)
9. Renormalized Difference Vegetation Index (RDVI)	$\sqrt{NDVI*DVI}$	Roujean and Breon (1995)
10. Reduced Simple Ratio (RSR)	$\frac{NIR}{\mathrm{Re}d}\left(1-\frac{SWIR-SWIR_{\min}}{SWIR_{\max}-SWIR_{\min}}\right)$; $SWIR_{min}$ and $SWIR_{max}$ as defined in $NDVI_c$	Brown et al. (2000)
11. Simple Ratio (SR)	NIR/Red	Jordan (1969)
12. Soil adjusted vegetation index (SAVI)	$\frac{(1+L)(NIR+\mathrm{Re}d)}{NIR+\mathrm{Re}d+L}$; L = 0.5	Huete (1988a)
13. Soil adjusted vegetation index (SAVI2)	$\frac{NIR}{\mathrm{Re}d+b/a}$; b was set to 0.025 and a to 1.25	Major et al. (1990)
14. Modified Soil Adjusted Vegetation Index (MSAVI2)	$NIR+0.5-\sqrt{(NIR+0.5)^2-2(NIR-\mathrm{Re}d)}$	Qi et al. (1994)

3.3.4 Mean Diameter at Breast Height, Mean Tree Height and Stand Volume

Variables extracted from the Sappi forestry database consisted of species composition, age, mean stem density per hectare, mean DBH and mean tree height for 150 individual stands (*i.e.* the area of each stand is about 20 ha), representing approximately 12% of the total planted area (1 292 stands). These variables (*i.e.* tree height in meters, DBH in centimetres) were collected during the months of May and June 2012. Tree height and diameter at breast height (DBH) were measured, using the Haglof Digitech calliper and the Vertex IV laser instrument, respectively. Using a sampling method adopted by most South African forestry companies, these tree attributes were recorded, utilizing circular sample plots of about 400 m^2 in size that were systematically distributed (usually every 100 m) within the stand. Sample intensities may vary between 2% and 10%, depending on the species composition, stand size or local forest conditions (Wessels *et al.,* 1985).

Subsequently, stand level volume estimates were calculated based on the averaged sample plot measurements, following a generalised allometric model for all *Eucalyptus* species (*i.e. E. grandis* and the hybrids *E.gxc* and *E.gxu* etc.) grown in this area (Bredenkamp, 2000; Kassier, 2012). To quantify stand volume (m^3ha^{-1}), mean stand height (Ht), computed basal area and species form factor were used as follows:

$$Vol = BA \times Ht \times f \qquad \text{Equation 3. 1}$$

Where Vol = stand volume (m^3ha^{-1}); BA =Basal Area (m^2ha^{-1}); Ht = mean stand height (m); and f = species form factor, which is the fraction of tree diameter at the point of measurement and the DBH (Bredenkamp, 2000; Kassier, 2012).

Descriptive statistics for the three tree-structural attributes *i.e.* stand volume, mean tree height and mean DBH is shown in Table 3.2. The average stand volume in Zululand plantation forest was 164 $m^3\ ha^{-1}$, whereas the average DBH was 13.45 cm. The average mean tree height for the entire study area was 19.75 m.

Table 3.2 Descriptive statistics for the three measured tree-structural attributes (DBH, tree height and stand volume)

Species properties	Min.	Mean	Max.	Stdev
Diameter at breast height (cm)	8.34	13.43	17.55	2.06
Tree height (m)	10.35	19.75	24.46	3.37
Eucalyptus stand volume (m/ha^3)	23.43	163.60	326.40	72.63

3.3.5 Ancillary Stand Data

Monthly rainfall data (in millimetres) corresponding to the stand age was gathered from the South African Weather Service and Institute for Soil, Climate and Water weather stations, via the Agro-Met data information system located at the Agricultural Research Council (http://www.arc.agric.za). The Agricultural Research Council computes the rainfall data set, by assigning a rainfall value to a specific location based on spatial interpolation. The method considers the measured rainfall values present at the five rainfall stations closest to a location and the satellite rainfall estimate at that specific location. The measured rainfall data is obtained from 550 automatic near-real time rainfall stations that are evenly distributed across the country, with a spatial resolution of 1 km x 1 km (de Coning, 2013; de Coning and Poolman, 2010). A map illustrating the rainfall distribution patterns within the study area is shown in Figure 3.2. As shown in the Figure, there are significant variations in the amount of rainfall received amongst the stands located in the south (*i.e.* wetter) and stands located in the north (*i.e.* drier).

Based on the period of rainfall data corresponding to the stand age, the mean rainfall (meanR), maximum rainfall (maxR), minimum rainfall (minR) and the coefficient of variation of rainfall (covR) were then calculated. The derived rainfall metrics, together with the tree age (EnumAge), were then used as ancillary variables to predict *Eucalyptus* stand volume and the related tree-structural attributes (*i.e.* mean DBH and mean tree height).

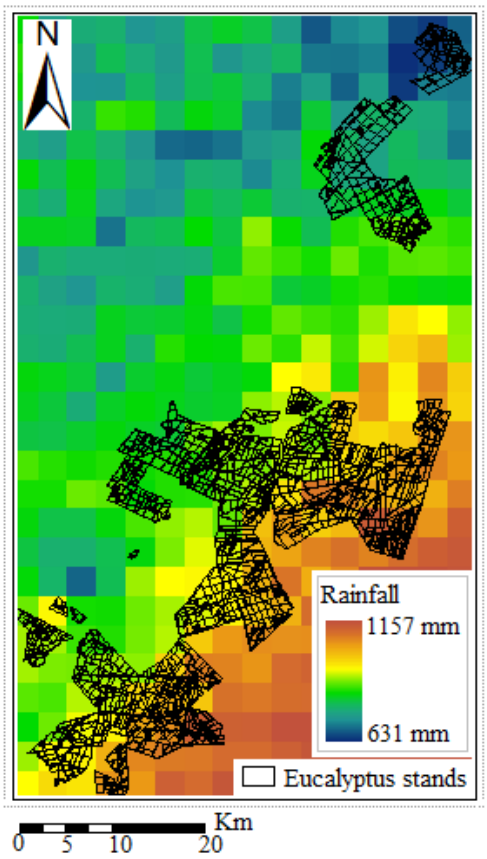


Figure 3.2 Rainfall distribution map of the study area

3.4 Modelling Techniques

3.4.1 Stochastic Gradient Boosting

Stochastic gradient boosting (Friedman, 2001; 2002; Friedman and Meulman, 2003) is an advanced machine-learning algorithm related to bagging and boosting techniques. The SGB algorithm builds numerous regression trees by sequentially modelling "pseudo" residuals, using a stepwise model fitting approach (recommended for reducing bias) and model averaging, which decreases model variance (Elith *et al.,* 2008; Freund and Schapire, 1997). At each iteration, a regression tree is grown from a random sub-sample of the data set selected, without replacement (Elith *et al.,* 2008; Freund and Schapire, 1997). A residual deviance is then calculated on data not used in the model fitting process. Trees are added, until the total residual deviance calculated from the withheld data ceases to decrease. The number of trees giving the lowest total residual deviance represents the most appropriate model for predicting the response variable (De'ath, 2007; Leathwick *et al.,* 2006). The main strengths of the SGB algorithm include: (i) the resistances to outliers, (ii) the use of predictor variables without transformation, (iii) the capability to fit complex nonlinear relationships, and (iv) the automatic handling of interaction effects among predictors. Finally, the

algorithm introduces an element of stochasticity, which improves the model accuracy and reduces over-fitting (Leathwick *et al.,* 2006; De'ath, 2007; Elith *et al.,* 2008).

The SGB algorithm is normally governed by three important user-defined parameters which encompass (i) the learning rate (*lr*), (ii) tree complexity (*tc*), and (iii) the number of regression trees (*ntree),* which depends on *lr* and *tc* for an optimal prediction (Hastie *et al.,* 2001; Leathwick *et al.,* 2006; Elith *et al.,* 2008). The *lr* parameter determines the contribution of each tree to the growing model, whereas *tc* is the number of the samples which are used in each final node (De'ath, 2007; Elith *et al.,* 2008). The SGB algorithm was implemented, using the gbm package within the R statistical package (R Development Core Team, 2008). A comprehensive guide on appropriate model settings is summarized in the literature (Friedman 2001; 2002; De'ath, 2007; Elith *et al.,* 2008). The guidelines presented in these studies resulted in the subsequent parameter settings. In the present study, *lr* was set between 0.0001 and 0.1, whereas *tc* was set between 1 and 5. In addition, the bagging fraction, which determines a fraction of the training data selected randomly for computing each tree, was set at 0.3 during the analysis process.

3.4.2 Variable Importance and Selection

The SGB algorithm estimates the relative influence of each variable in predicting a response, based on the formula developed by Friedman (2001). For regression, the algorithm determines the variable importance through quantifying the sum of squared improvements at all splits, determined by the predictor, and then averages the relative influence of each predictor variable over the collection of trees (Friedman and Meulman, 2003; Elith *et al.,* 2008). The relative contribution of the individual variable is then scaled (*i.e.* so that the sum adds to 100), with higher numbers indicating a stronger influence on the response variable and *vice versa*.

The SGB algorithm implements variable selection by successively removing predictor variables that do not improve model accuracies, as determined by the residual deviance. In the present study, variable selection was performed, using a backward feature elimination (BFE) approach. The BFE approach develops a model by first utilizing all the predictor variables (n = 23 in this study) and then progressively eliminating variables, based on their relative influence. This process creates multiple models and the model with the optimal subset of variables is then selected, based on the lowest residual deviance obtained (Elith *et al.,* 2008).

3.4.3 Random Forest

The results of the SGB ensemble were then compared with the random forest ensemble, a machine-learning technique developed by Breiman (2001). RF improves the classification and regression trees (CART) method, by combining a large set of regression trees. The RF ensemble, which is also a modified version of adaboosting (bootstrap aggregation), has an additional modification of selecting only a random subset of candidate features (*mtry*), to determine the split at each node of a tree. This ensemble method uses recursive partitioning, to create multiple regression trees (*ntree*), and then averages the results of all trees (Breiman, 2001). Each tree is independently grown to maximum size without any pruning, based on a bootstrap sample from the training data set (approximately 2/3 of the original data). RF optimization was performed, based on two parameters, namely, *ntree*, the number of regression trees grown, based on a bootstrap sample of the observations and *mtry*, the number of different predictors tested at each node. To determine the optimal *ntree* and *mtry* values that can best predict stand volume, mean DBH and mean tree height, the two parameters (*mtry* and *ntree*) were optimized, based on the lowest ten-fold cross-validation prediction error. The *ntree* values were tested from 500 to 2500, with a 500 interval, whereas *mtry* was tested from 1 to 23, based on a single interval. For a detailed description of RF, see Breiman (2001). The RF ensemble was implemented, using the "randomForest" library (Liaw and Wiener, 2002) within the R statistical package (R Development Core Team, 2008)

3.4.4 Stepwise Multiple-linear Regression

The results obtained from the above-mentioned algorithms were also compared with those derived using stepwise multiple-linear regression. Previous studies indicate that the SMLR model has been extensively used in modelling the relationship between remotely sensed data and different tree-structural attributes, with reasonable accuracy rates (Jain *et al.,* 2007; Jones and Ricciardi, 2005).

3.4.5 Model Accuracy Assessment

The coefficient of determination (R^2) and the root mean square error (RMSE) were used in the study to assess the predictive accuracy of all the three predictive methods (*i.e.* SGB, RF and SMLR models).

3.5 Results

3.5.1 Predicting Eucalyptus Stand Volume

Table 3.3 demonstrates the prediction results, based on the R^2 and RMSE values obtained from SGB, RF and SMLR models, respectively. The required parameters for RF and SGB, derived from optimization of the models, are shown in Table 3.3. Stand volume was estimated using: (i) SPOT 5 bands, (ii) VIs, (ii) ancillary stand data (*i.e.* rainfall metrics, and stand age), and (iv) multisource data (*i.e.* all data sets combined). For all of the above modelling stages, the SGB algorithm performed better than the RF and SMLR models. Based on only the SPOT-5 bands, the SGB model produced an R^2 value of 0.41 and an RMSE value of 54.40 m^3ha^{-1}, the competing RF model produced an R^2 value of 0.39 and an RMSE value of 55.06 m^3ha^{-1}, while the SMLR model produced an R^2 value of 0.24 and an RMSE value of 63.70 m^3ha^{-1}. It can also be noted that, for the three modelling stages (i, ii and iii), SPOT-5 produced less accurate stand volume (m^3ha^{-1}) prediction results (Table 3.3). Overall, the best results were obtained using the multisource data sets for all three modelling techniques considered in this study. The SGB model yielded an R^2 value of 0.78 and an RMSE value of 33.16 m^3ha^{-1}, the RF model produced an R^2 value of 0.76 and an RMSE value of 37.26 m^3ha^{-1} and the SMLR model produced an R^2 value of 0.65 and an RMSE value of 42.50 m^3ha^{-1}. The SMLR model produced the least accurate results for all the modelling stages.

Table 3.3 Comparisons of *Eucalyptus* stand volume prediction results, using three predictive models (stochastic gradient boosting, random forest and stepwise multiple-linear regression)

Input data set	Statistical methods	*tc*	*lr*	*mtry*	*ntree*	R^2	RMSE (m^3 ha^{-1})
SPOT-5 bands	SGB	3	0.005	-	1300	0.41	54.40 (37.75%)
(n = 4)	RF	-	-	3	500	0.39	55.06 (38.21%)
	SMLR	-	-	-	-	0.24	63.70 (44.21%)
Vegetation indices	SGB	5	0.003	-	1000	0.44	52.74 (36.60%)
(n=14)	RF	-	-	12	500	0.41	54.62 (37.90%)
	SMLR	-				0.42	52.97 (36.76%)
Ancillary stand data	SGB	5	0.005	-	800	0.50	52.30 (36.29%)
(n = 5)	RF	-	-	3	1000	0.46	52.30 (36.29%)
	SMLR	-	-	-	-	0.42	55.37 (38.42%)
Multisource data	SGB	5	0.005	-	1350	0.78	33.16 (23.01%)
(n = 23)	RF	-	-	6	500	0.76	37.28 (25.87%)
	SMLR	-	-	-	-	0.65	42.50 (29.49%)

*For stochastic gradient boosting (SGB), random forest (RF) and stepwise multiple regression (SMLR), the R^2 and RMSE values were generated using the 30% holdout test data set (n = 45) and the optimization results are based on a 10-fold cross validation method using the training data set (n = 105). RMSE values were also provided as a percentage (%) of the mean. The SLMR method selected nine variables in predicting stand volume.

3.5.2 Predicting Mean Diameter at Breast Height and Mean Tree Height

Mean DBH and mean tree height prediction results, obtained from the SGB, RF and SMLR models, are presented in Table 3.4. Overall, the SGB model performed better than the other models in terms of the RMSE of prediction.

Table 3.4 Comparisons of *Eucalyptus* mean DBH and mean tree height prediction results using three predictive models (stochastic gradient boosting, random forest and stepwise multiple-linear regression)

		mean DBH		mean tree height	
Input data set	Statistical methods	R^2	RMSE (cm)	R^2	RMSE (m)
SPOT-5 bands	SGB	0.46	1.38 (10.28%)	0.49	2.53 (14.09%)
(n = 4)	RF	0.40	1.47 (10.95%)	0.49	2.68 (14.92%)
	MLR	0.32	1.49 (11.09%)	0.43	2.73 (15.20%)
Vegetation indices	SGB	0.45	1.57 (11.69%)	0.42	2.69 (14.98%)
(n=14)	RF	0.43	1.62 (12.06%)	0.46	2.84 (15.81%)
	MLR	0.31	1.73 (12.88%)	0.40	3.05 (16.98%)
Ancillary stand data	SGB	0.54	1.23 (9.16%)	0.56	2.56 (14.25%)
(n = 5)	RF	0.49	1.38 (10.28%)	0.48	2.69 (14.98%)
	MLR	0.47	2.63 (19.58%)	0.42	2.73 (15.20%)
Multisource data	SGB	0.74	1.06 (7.89%)	0.83	1.63 (09.08%)
(n = 23)	RF	0.67	1.19 (8.86%)	0.78	1.67 (09.30%)
	MLR	0.68	1.32 (9.83%)	0.72	1.83 (10.19%)

*For stochastic gradient boosting (SGB), random forest (RF) and stepwise multiple regression (SMLR), the R^2 and RMSE values were generated using the 30% holdout test data set (n = 45) and the optimization results are based on a 10-fold cross validation method using the training data set (n = 105). RMSE values were also provided as a percentage (%) of the mean. The SLMR method selected 10 variables in predicting mean DBH, and mean tree height.

Predicting mean DBH and mean tree height, using SPOT-5 reflectance bands, resulted in low accuracies, with SGB producing the best results (mean DBH: R^2 = 0.46 and RMSE = 1.38 cm; mean tree height: R^2 = 0.49 and RMSE = 2.53 m). Mean DBH and mean tree height predictions, using the VIs and ancillary stand data sets, show a similar trend, that is, the SGB model produces the best accuracies, followed by RF and SMLR, although the RF model produced better accuracies (R^2 = 0.46 and RMSE = 2.84 m) when predicting mean tree height, using the VIs data set. The multisource data set produced the best predictions of mean DBH and mean tree height for all models considered in this study. However, the best models for mean tree height (R^2 = 0.83 and RMSE = 1.63 m) and mean DBH (R^2 = 0.74 and RMSE = 1.06 cm) were obtained, using the SGB model.

3.5.3 Variable Selection and Model Improvement

Since the SGB models, using the multisource data sets, produced the best accuracies for predicting mean tree height, mean DBH and stand volume, this section only focuses on improving the SGB

model accuracies by implementing a backward variable selection process. In assessing the relative importance of variables, as determined by the SGB algorithm, Figure 3.3 shows that the SWIR, minR, EnumAge and covR variables are ranked highly, when predicting mean DBH (Figure 3.3a), mean tree height (Figure 3.3b) and stand volume (Figure 3.3c). In addition the relative importance of the SWIR variable is greater than 20%, when predicting mean DBH, mean tree height and stand volume. Following these variable importance rankings, the next challenge was to select the optimal number of variables: (i) that could produce the best accuracies, when predicting mean DBH, mean tree height and stand volume, and (ii) that would be helpful in the interpretation of the final models.

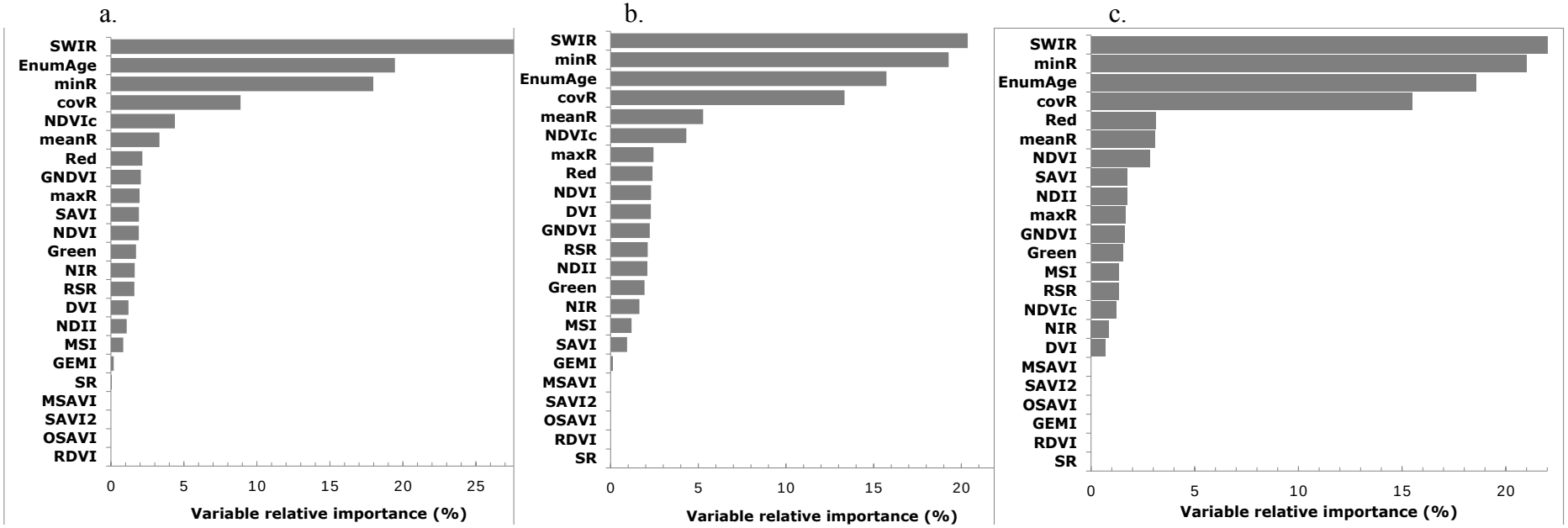


Figure 3.3 The relative importance of variables in the multisource data set (n = 23). The variables are ranked based on their contribution to the stochastic gradient boosting model. a, b and c represent the relative importance of the variables, when predicting mean DBH, mean tree height and stand volume

Table 3.5 Comparisons of *Eucalyptus* mean DBH, mean tree height and stand volume prediction results, using stochastic gradient boosting and backward feature elimination

Attribute	No. of variables	R^2	RMSE
mean DBH	10	0.81	1.21 cm (6.12%)
mean tree height	10	0.86	1.39 m (7.02%)
Stand volume	9	0.83	29.58 m^3ha^{-1} (17.63%)

*For stochastic gradient boosting, the R^2 and RMSE values were generated using the 30% holdout test data set (n = 45) and the optimization results are based on a 10-fold cross validation method, using the training data set (n = 105). RMSE values were also provided as a percentage (%) of the mean.

Table 3.5 presents the predictions of mean DBH, mean tree height and stand volume and the model parameters for the SGB model and variable selection, using the backward feature elimination (BFE) method. In this regard, the BFE method selected ten predictor variables (*i.e.* SWIR, EnumAge, minR, covR, NDVIc, meanR, red, GNDVI, maxR and SAVI) from all the data sets (Figure 3.3a) that were optimal for predicting mean DBH. The selected variables produced an R^2 of 0.81 and an RMSE of 1.21 cm, which are better than those achieved using all variables (Figure 3.4a and Table 3.4). In comparison, the SGB model using all the variables produced an R^2 value of 0.74 and RMSE value of 1.06 cm (Table 3.4). The BFE method selected SWIR, minR, EnumAge, CovR, meanR, NDVIc, maxR, Red, NDVI and DVI for predicting mean tree height (Figure 3.3b). Based on these variables, mean tree height prediction was improved (R^2 = 0.86 and a RMSE = 1.39 m), compared to SGB using all the variables (see Table 3.4). In predicting stand volume, the BFE method selected SWIR, minR, EnumAge, covR, Red, meanR, NDVI, SAVI and NDII. Moreover, the selection of SWIR, minR, EnumAge, covR, Red, meanR, NDVI, SAVI and NDII resulted in an R^2 of 0.83 and an RMSE of 29.58 m^3 ha^{-1}, which are better than those achieved from the use of all variables (see Table 3.3). Overall, SGB models that used variables selected by the BFE method showed improved prediction results for mean DBH, mean tree height and stand volume.

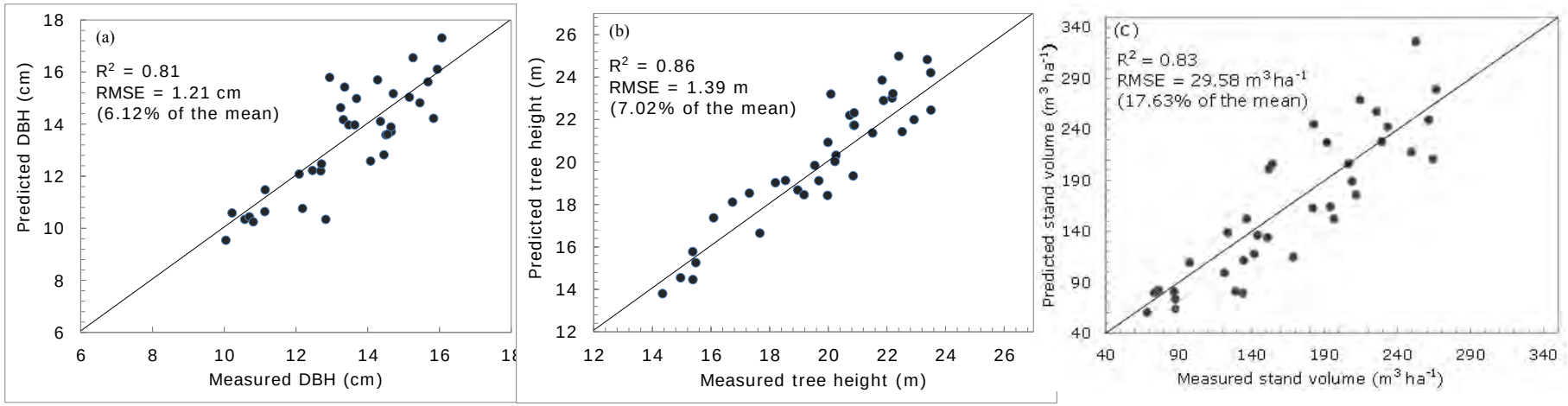


Figure 3.4 Prediction results for (a) mean DBH, (b) mean tree height and (c) stand volume, based on variables selected from the multisource data set using the backward feature elimination approach and stochastic gradient boosting

3.6 Discussion

One of the major challenges in predicting tree-structural attributes (*i.e.* stand volume, mean DBH and mean tree height), using medium spatial resolution multispectral data sets, is the inability to overcome the problem of saturation in areas with high canopy cover. Studies have explored the utility of high spectral and spatial resolution sensors with strategically-located bands to overcome saturation problems, as well as to enhance the estimation accuracies (Gara *et al.,* 2014b). Nevertheless, the cost, volume and availability of high spatial and spectral resolution sensors, such as hyperspectral, Woldview-2, RapidEye, lidar and radar data sets, remain one of the major setbacks in resource-limited environments. The solution therefore lies in identifying possible ways of enhancing the prediction accuracy of the most important tree-structural attributes, based on the cost and availability of multispectral sensors. This research therefore provides important evidence on the potential of predicting of *Eucalyptus* stand volume and related tree-structural attributes (*i.e.* mean DBH and mean tree height) from multisource data sets, based on advanced predictive models (*i.e.* SMLR, RF and SGB) in the Zululand region of South Africa. The need for developing a cheap and quick operational approach is necessary for improving our understanding of the contribution of forests towards the regional carbon pool.

3.6.1 Combining Remotely Sensed Data with Ancillary Stand Data

The results of this study have indicated the potential and effectiveness of predicting mean DBH, mean tree height and stand volume with higher accuracy, using relatively cheap and readily-available medium-spatial resolution SPOT-5 derived data (*i.e.* spectral bands and vegetation indices), combined with ancillary stand data sets (*i.e.* rainfall metrics and stand age). The combination of these data sets (*i.e.* multisource data set) yielded better results, when compared to using the individual data sets on their own. For instance, when all predicted variables (*i.e.* medium spatial resolution SPOT-5 spectral information, vegetation indices, rainfall metrics and stand age) were used, the overall accuracies increased significantly for all the predictive models (*i.e.* SMLR, RF and SGB), as opposed to when the variables were used separately. Furthermore, the results obtained after implementing variable selection further improved the final model prediction accuracy for all tree-structural attributes considered in this study, compared to those derived from all variables.

These results are consistent with the results obtained in previous work that integrated remote sensing data with ancillary data sets, to estimate various tree-structural attributes (Baccini *et al.,*

2004; Zheng *et al.,* 2007; Gebreslasie *et al.,* 2010; le Maire *et al.,* 2011). For example, Zheng *et al.* (2007) estimated forest AGB in China, using field forest inventory data, leaf area index and vegetation indices derived from Landsat enhanced thematic mapper images. Results showed that the SMLR model, using LAI and simple ratio (SR) index, as predictors, explained 93% of the AGB variance for mixed forests, whereas the model based on LAI and stand age captured only 79% of AGB for broad-leaved forests. Recently, Zhang *et al.* (2014) estimated AGB in Canada, using a spatially explicit data set derived from a combination of forest inventory data from 1 968 plots and space borne lidar canopy height data, 10 climatic variables and elevation. Using four regression models, namely, spatial interpolation, non-spatial and spatial regression models, as well as the RF algorithm, their results showed that the integration of field-based inventory data, space borne lidar data, land cover classification, as well as climatic and environmental variables enhanced the quantification of forest AGB across large regions, explaining 62% of the variation in biomass. The work conducted by Baccini *et al*. (2004) observed that the use of coarse spatial resolution multispectral data, in combination with appropriate climatic and topographic factors, can effectively be used to map aboveground forest biomass for National Forest lands in California. Furthermore, le Maire *et al.* (2011) estimated different tree-structural attributes at stand scale with high accuracy, using MODIS NDVI time series and bioclimatic data. Stand-age alone explained more than 82% of stand wood volume variability and 87% of stand dominant height variability. The results from this study further confirm the importance of combining multispectral data, together with ancillary variables, as an important alternative for enhancing the prediction accuracy of stand volume and tree-structural attributes, such as the mean tree height and mean DBH.

3.6.2 Comparison of Stochastic Gradient Boosting, Random Forest and Stepwise Multiple-linear Regression

It is important to also note that one of the main objectives in this study was to compare the strength of three predictive models (*i.e.* SLMR, RF and SGB). The SGB, RF ensembles and SMLR model were compared on their ability to accurately predict mean DBH, mean tree height and stand, using multisource data. Results show that the SGB ensemble performed better than the competing RF ensemble and SMLR model. This has been confirmed by earlier studies, which suggest that in applications, where the number of field observations is small, SGB tends to outperform the RF regression ensemble (Dube and Mutanga, 2014). Previous studies indicate that the RF algorithm requires a larger number of field observations to strengthen the randomization process, especially when the sample data set varies significantly (Gara *et al.,* 2014b; Özçift, 2011). In addition, a study

by Krahwinkler and Rossman (2011) has shown that the RF algorithm uses a combination of soft linear boundaries and that it may not function with few observations at the decision surface. Therefore, RF does not efficiently perform when the number of observations is small.

Researchers have further noted that the SGB and RF algorithms are better than conventional regression methods, such as multiple-linear regressions and stepwise regression methods, because they have elements of stochasticity, responsible for improving model accuracy and reducing model overfitting (Leathwick *et al.,* 2006; Elith *et al.,* 2008). SGB can also handle voluminous and highly multicollinear data sets (De'ath, 2007). In addition, when used in conjunction with the BFE method, the SGB ensemble provided valuable information, by ranking and selecting the most important predictor variables (Leathwick *et al.,* 2006; Elith *et al.,* 2008). Overall, SGB models that used variables selected by the BFE method showed improved prediction results for mean DBH, mean tree height and stand volume (Table 3.5). These results can be explained by the fact that the inclusion of less important variables in analysis can cause the convergence and instability of models, or influence the random errors, due to noise in information predictor variables that have no relation to the response variable (Mutanga et al. 2012).

A closer examination of the variables selected for predicting the mean DBH, mean tree height and stand volume (Figure 3.3 and Table 3.5) revealed that the SPOT-5 SWIR band is the most highly ranked variable for those predictions. The SWIR spectral band is reported to be more sensitive to vegetation properties, such as canopy biomass and water content, when compared to other electromagnetic spectrum regions (Dube, 2012; Dube *et al.,* 2012; Main-Knorn *et al.,* 2011; Main-Knorn *et al.,* 2013). Main-Knorn *et al*. (2013) demonstrated that the inclusion of the SWIR spectral band in modelling tree-structural attributes is critical in enhancing the predictive accuracies for coniferous forest species in the Western Carpathian Mountains. In addition, rainfall has been noted to have a significant influence on forests, as it accelerates structural growth, especially in water-limited environments (Gebreslasie *et al.,* 2010; Gracia *et al.,* 2010). Water supply does not only affect plantation forest yield, but also determines the spatial distribution of plantation species across the landscape. For instance, water scarcity proves to be a critical constraint to successful and health primary productivity (Gracia *et al.,* 2010). Literature shows that for effective carbon assimilation within the whole plant system, the presence of an optimal amount of water, to allow carbon diffusion, is a pre-requisite (Chaves *et al.,* 2009).

In addition to the above three most important selected variables (*i.e.* SPOT-5 SWIR band and minimum rainfall), our study indicated that stand-age and vegetation indices, such as the NDII, NDVIc, NDVI, DVI, GNDVI and SAVI, play a critical role in enhancing the prediction of mean DBH and mean tree height. These results are also in accordance with the general knowledge regarding the relationship of spectral vegetation indices with various tree-structural attributes. Literature shows that the remotely sensed spectral properties are very sensitive to tree structural attributes (Tucker 1979; Thenkabail *et al.,* 2004; Main-Knorn *et al.* 2013; Dube, and Mutanga, 2014). However, the weak performance of vegetation indices computed from SPOT-5 multispectral image can be associated with saturation problems, especially where there is high canopy cover (Mutanga and Skidmore, 2004a). On the other hand, the results of this study show that the mean DBH predictions are better than those of mean tree height, which is not always the case in most studies available from literature. The improved mean DBH prediction results obtained in this study, when compared to those of mean tree height, can be attributed to the incorporation of ancillary data sets. Thus far, none of the studies available in literature has combined rainfall, age with spectral information in predicting mean DBH and mean tree height. Stand age is important for predicting volume, as shown by the relative importance graph in Figure 3.3. However, in this study it may not be selected as the most important variable because of the lack of variability in monoculture stands, as opposed to natural forests.

Overall, the results of our study have shown that SGB is a useful approach in predicting mean DBH, mean tree height and stand volume, using SPOT-5 derived data, combined with multisource data. This is particularly helpful in the African context, where the availability of higher resolution imagery is rather limited. However, for the approach presented in this study to be operational, various available and newly-launched multispectral sensors, together with other environmental variables, such as topography and temperature, should be investigated and tested in heterogeneous forests. Although the results of this study seem to be promising, the terrain in Zululand is relatively flat and therefore this method should be tested in rugged terrain, where both the land cover and topography would affect the spectral values of remotely sensed imagery (Dorren *et al.,* 2003).

3.7 Conclusion

The results have shown that:

1. predictions of mean DBH, mean height and stand volume for *Eucalyptu*s trees are improved when SPOT-5 derived data (*i.e.* spectral bands and vegetation indices) are combined with ancillary stand data (*i.e.* rainfall metrics and stand age),
2. the SGB ensemble performed better than the competing RF ensemble and SMLR, when predicting mean DBH, mean height and stand volume, and
3. the best results for predicting mean DBH (R^2 = 0.81 and RMSE = 1.21 cm), mean height (R^2 = 0.86 and RMSE = 1.39 m) and stand volume (R^2 = 0.83 and RMSE = 29.58 m^3ha^{-1}) were obtained by the SGB model, based on variables selected by the BFE method.

While the integration of ancillary data with SPOT-5 spectral information produces high quality stand volume estimates, there is a need to identify remote sensing data that can be independently applied in areas where ancillary data sets are unavailable, in order to derive accurate estimates.

4. QUANTIFYING INTRA- AND INTER-SPECIES ABOVEGROUND BIOMASS

This chapter is based on:

Dube, T, Mutanga, O, Ismail, R. 2014. Intra-and-inter species biomass prediction in a plantation forest: testing the utility of high spatial resolution spaceborne multispectral RapidEye sensor and advanced machine learning algorithms. *Sensors,* 14, 15348-15370.

Abstract

The quantification of aboveground biomass, using remote sensing, is critical for better understanding the role of forests in carbon sequestration and for informed sustainable management. Although remote sensing techniques have been proven to be useful in assessing forest biomass in general, more is required to investigate their capabilities in predicting intra- and inter-species biomass, which are mainly characterised by non-linear relationships. In this study, two machine-learning algorithms, SGB and RF regression trees, were tested in predicting intra-and-inter-species biomass, using high resolution RapidEye reflectance bands, as well as the derived vegetation indices in a commercial plantation. The results showed that the SGB algorithm yielded the best performance for intra-and-inter-species biomass prediction, using all the predictor variables, and was based on the most important selected variables. For example, using the most important variables, the algorithm produced an R^2 of 0.80 and RMSE of 16.93 t ha^{-1} for *E.grandis,* an R^2 of 0.79 and RMSE of 17.27 t ha^{-1} for *P.taeda* and an R^2 of 0. 61 and RMSE of 43.39 t ha^{-1} for the combined species data sets. Comparatively, RF yielded plausible results only for *E.dunii* (R^2 of 0.79; RMSE of 7.18 t ha^{-1}). The study demonstrated that although the two statistical methods were able to predict biomass accurately, RF produced weaker results, compared to SGB, when applied to the combined species data sets. The result underscores the relevance of stochastic models in predicting biomass drawn from different species and genera, using the new generation high resolution RapidEye sensor with strategically-positioned bands.

Keywords: bag fraction; biosphere-atmospheric interactions; learning rate; high resolution RapidEye imagery; tree complexity; variable importance and variable selection

4.1 Introduction

Forests serve as an important key driver of regional and local climate systems through biosphere-atmospheric interactions (Muukkonen and Heiskanen, 2007; Poulain *et al.,* 2011; Yang *et al.,* 2011). Information on forest spatial distribution, biomass levels and dynamics is therefore required, for the accurate estimation of greenhouse gases flux, policy development and implementation (Gibbs *et al.,* 2007). In addition, knowledge on intra-and-inter commercial forest biomass is central in: (i) determining their productive capacity, (ii) ensuring informed sustainable management practices, and (iii) understanding the functioning of the planet, earth and the environment (de Jong *et al.,* 2003; Heiskanen, 2006). Therefore, the continuous estimation, mapping and monitoring of aboveground forest biomass (tonnes ha^{-1}), which is the amount of living plant matter (de Jong *et al.,* 2003), is central in climate modelling worldwide, due to its significance in net carbon emission computations (Chinembiri *et al.,* 2013b; Kurz and Apps, 2006; St - Onge *et al.,* 2008).

Currently, there are two approaches for forest biomass estimation, namely, field-based traditional methods (*i.e.* field measurements or harvesting) and remotely sensed methods (Lu, 2006). So far, traditional methods have been side-lined in favour of remotely sensed techniques, since its inception. Although regarded as highly accurate (Henry *et al.,* 2011; Lu, 2006), the traditional methods are exceedingly time-consuming, labour-intensive and difficult to implement, especially in remote areas, and they are practically and spatially limited to a small tree sample size and require a sufficient number of samples (Adam *et al.,* 2014; Lu, 2005b). Recent evidence suggests that remote sensing seems to be a valuable and low-cost tool for determining forest biophysical attributes, when compared to field surveys (Carreiras *et al.,* 2012b; Gebreslasie *et al.,* 2011; Mutanga *et al.,* 2012). Remotely sensed data permits robust biomass retrieval, which is critical for assessing the ecosystem yield and carbon accounting. As a result, biomass estimation, using remotely sensed data, as the primary source, has gained increasing interest in the past decades, especially for natural forests, at both local and regional scales (Lu, 2005b).

Although biomass cannot be directly quantified from space, satellite image reflectance permits the extraction of biomass estimates, especially when integrated with field-based measurements (Dong *et al.,* 2003). Consequently, various remotely sensed studies concerning forest biomass estimation have been applied at different scales. It has been discovered that coarse spatial resolution optical sensors are useful for biomass mapping at continental and global scale, rather than at local scale (van der Werf *et al.,* 2006b; Zhang and Kondragunta, 2006), due to the plausible trade-off between spatial resolution, image coverage and frequency in data acquisition (Baccini *et al.,* 2004;

Heiskanen, 2006; Hyyppa and Hyyppa, 2001; Lu, 2006; Muukkonen and Heiskanen, 2007). The main limitation with the broadband multispectral sensors for biomass estimation is the fact that they are characterised by mixed pixels, which occur as a result of large sensor footprint (Basuki *et al.,* 2011). In that regard, the huge difference between the satellite data pixel size and the ground reference data makes these sensors inapplicable for intra-and-inter-species biomass prediction in commercial forest plantations.

Recent efforts have been geared towards the use of high resolution sensors, such as narrow band hyperspectral, radar and lidar data, for estimating aboveground forest biomass (AGB) to reduce the limitations associated with the broadband multispectral data sets (Carreiras *et al.,* 2012a; 2013; Colgan *et al.,* 2013; Mitchard *et al.,* 2012; 2011; 2009; Saatchi *et al.,* 2011). Results have shown that hyperspectral, radar and lidar sensors have robust means of data collection and the subsequent characterization of the vertically-distributed forest attributes, hence they can be regarded as an appropriate primary data source for forest inventorying. The use of these data sets nonetheless comes with its own limitations in terms of cost, availability, spectral contiguity, processing and analysis complexity, especially in the African context, given its economic situation and lack of the required technical expertise. For example, processing hyperspectral data for vegetation applications is a major challenge, due to the Hughes phenomenon or "the curse of dimensionality". This problem often introduces a high degree of multicollinearity as a result of the similarities in the biophysical spectral reflectance properties (Adjorlolo, 2013; Clevers *et al.,* 2007; Ferwerda *et al.,* 2005; Knox *et al.,* 2011).

The utility of new generation sensors, such as the RapidEye, with strategic bands, is therefore seen as a trade-off between the advantages of coarse multispectral data, hyperspectral, lidar and radar data in predicting intra-and-inter-species AGB (Eckert, 2012b; Ozdemir and Karnieli, 2011; Thenkabail *et al.,* 2004b). Currently, RapidEye, together with WorldView-2 sensors, are the only commercial multispectral satellite sensors which provide a reasonable number of spectral bands that are configured in unique portions of the electromagnetic spectrum and provide global, high-resolution access to the red- edge spectral band (Cheng and Chaapel, 2008). In remote sensing, the "red-edge" is the region of abrupt change in the leaf reflectance between 680 and 780 nm, due to the combined effects of strong chlorophyll absorption in red wavelengths and a high reflectance in the NIR wavelengths, due to leaf internal scattering (Horler *et al.,* 1983). The new generation RapidEye image containing strategically-positioned bands, with a fine spatial resolution of 5 m, is

hypothesized to be critical for vegetation mapping, when compared to the traditional broadband satellite images, such as ASTER, SPOT and Landsat Thematic Mapper. Above all, the RapidEye reduces unnecessary redundancy, a problem associated with hyperspectral data (Mutanga *et al.,* 2012). Recently, the strategically-positioned bands of the RapidEye imagery have successfully been applied extensively in detecting different levels of insect defoliation in Mopane woodlands (Adelabu *et al.,* 2013a; Adelabu *et al.,* 2013b), whereas other studies have demonstrated that the strategically-positioned RapidEye bands allow for enhanced vegetation mapping (Schuster *et al.,* 2012; Tigges *et al.,* 2013).

However, the rich spectral information contained in this data set has not been exploited for estimating intra-and-inter-species biomass in managed commercial plantations. For instance, commercial forests with mixed species (interspecies) are characterised by significant biomass variations, making it difficult for national carbon accounting. Taxonomical and structural differences are a major problem for intra-species biomass estimation (Carreiras *et al.,* 2012b). More importantly, different species and genera result in high biomass variations that are associated with non-linear relationships, making algorithm applications a significant challenge in estimating AGB in such environments. Due to the intra-and-inter-species variability, there is a high probability of outliers and unbalanced data sets in the collected training data. It is therefore critical for biomass studies to identify robust models that could overcome the failure to estimate biomass in forests characterised by intra-and-inter-species (Atta-Boateng and William, 1998; Carreiras *et al.,* 2012b; Dube *et al.,* 2015; Lu, 2005a).

In this study, the potential of two machine-learning algorithms, SGB and RF, in predicting intra-and-inter-species biomass in a commercial plantation forest in the midlands region of KwaZulu-Natal, South Africa, using the strategically-positioned spectral information derived from 5m RapidEye imagery, is therefore assessed. Previous studies have shown that non-parametric statistical techniques, such as the SGB and RF, simplify the biomass estimation process, when compared to other statistical regression methods (Adam *et al.,* 2014; Carreiras *et al.,* 2012b; Mutanga *et al.,* 2012). Both regression ensembles have received considerable attention, due to a number of statistical modeling properties. For instance, the SGB method produces results with plausible and highly robust estimates in regression studies, due to its ability to handle outliers, inaccurate training data, missing and unbalanced data sets (De'ath, 2007; Lawrence *et al.,* 2004; Leathwick *et al.,* 2006). Moreover, the model's stochastic characteristic in modelling non-linear

relationships and the inherent ability to handle, identify and select critical variables from large amounts of data, is expected to provide the best model accuracies (Carreiras *et al.,* 2012b; De'ath, 2007; Elith *et al.,* 2008; Lawrence *et al.,* 2004). Most importantly, SGB uses a stage-wise additive model fitting procedure that enhances the predictive performance of weak learning algorithms.

On the other hand, RF provides other appealing statistical properties, such as the useful internal estimates of error, strength, correlation and variable importance (Prasad *et al.,* 2006; Strobl *et al.,* 2008). In addition, Strobl *et al.* (2008) describe random forest algorithm as an effective tool, which performs simple and complex regressions with the modest fine-tuning of parameters, resulting in accurate predictions. The highlighted characteristics of SGB and RF, as well as the probability of intra-and-inter-species biomass variability, have therefore prompted an investigation into their capabilities (SGB and RF) in predicting AGB from a commercial forest in the midlands of KwaZulu-Natal, South Africa. Although both machine-learning techniques have been found to be robust under certain conditions, in this mixed species environment of KwaZulu-Natal, it is expected that SGB would perform better, due to its capabilities in modelling possible outliers and unbalanced data sets, as well as non-linear relationships. To the best of our knowledge, so far no study has assessed the SGB and RF for intra-and-inter-species biomass prediction in a commercial forest and, in particular, using the strategically-positioned bands of the new generation sensors, such as RapidEye. Therefore, our main objective was to investigate the robustness of the two machine-learning algorithms in predicting intra-and-inter-species biomass from plantation forests, using the recent high spatial resolution spaceborne RapidEye multispectral imagery. A secondary objective was to evaluate the relative importance of the high resolution RapidEye reflectance bands, as well as the derived vegetation indices, in the prediction of intra-and-inter species biomass.

4.2 Materials and Methods

4.2.1 Study Area

The study was conducted at the Sappi Clan forest, located approximately 27 km away from Pietermaritzburg city, the provincial capital of KwaZulu-Natal Province, South Africa (Figure 4.1). The plantation is located between Latitude 29°24'46.74"S and 29°17'45.94"S and Longitude 30°18'32.89"E and 30°28'28.21"E. South Africa is home to vast tracks of commercial plantation forests, both hardwood and softwood, covering approximately one percent of the total land area (DAFF, 2008). Specifically, the Clan forest used in this study covers about 6700 ha. The forest is characterized by extensive commercial forestry dominated by *Pinus* (*P. taeda*) and *Eucalyptus spp.*

(*e.g. E. grandis and E. dunii*). The climate in the study area is sub-tropical, with the mean annual rainfall varying between 700 mm and 1500 mm (Scott and Lesch, 1997). These fast-growing *Eucalyptus* species are planted with clones or seedlings and harvested every six to seven years. Stands are managed on a pulpwood regime (*i.e.* established at 1 667 trees/ha); and intensive soil preparation and weed control measures are practiced, until crown closure occurs between 1 to 1.5 years.

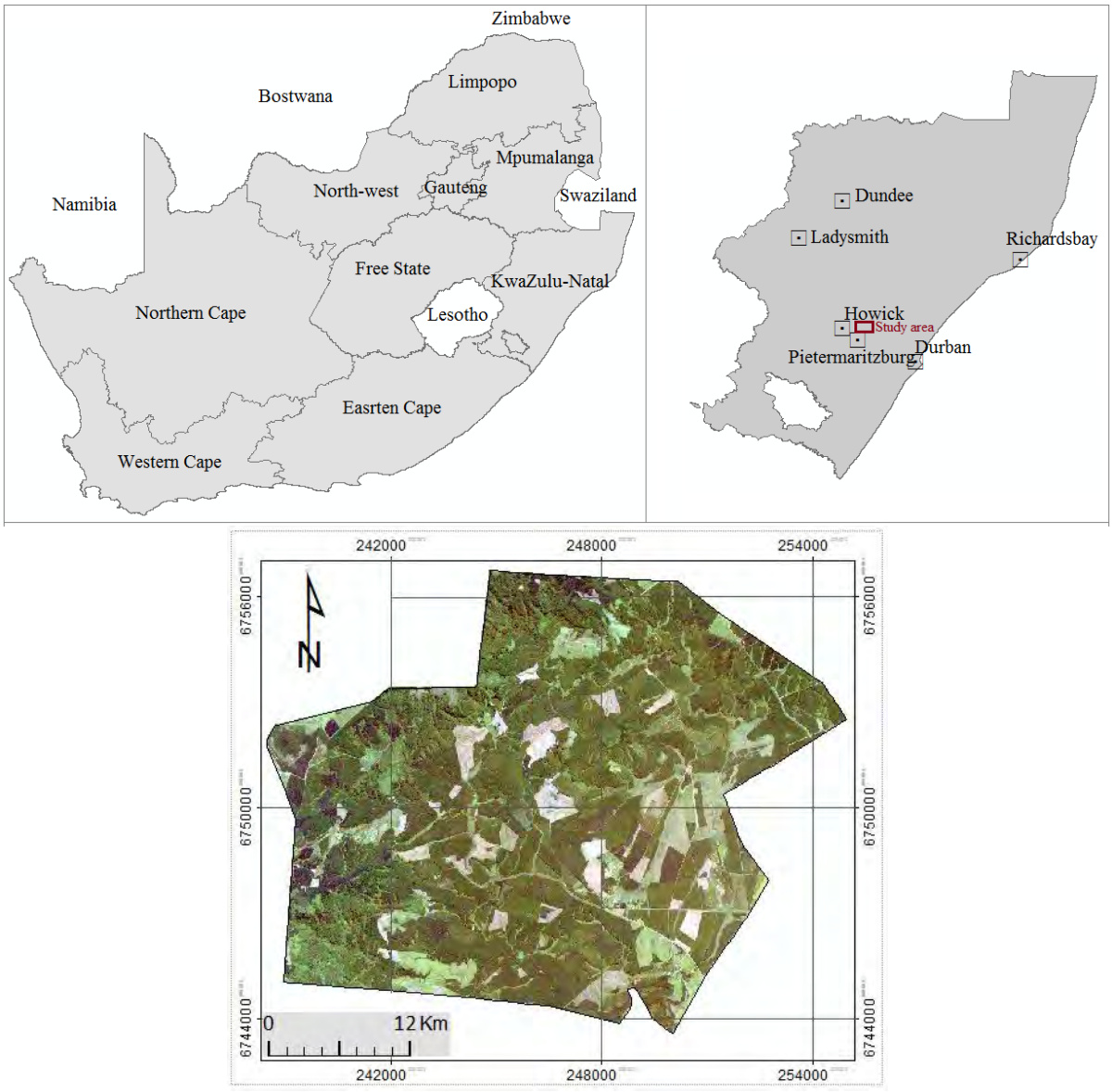


Figure 4.1 Pseudo colour composite map of the study area derived using Landsat-8 OLI bands 5, 6 and 1

4.2.2 Field Data Collection and Sampling Design

The field campaign was carried out between the 30th of July and the 22nd of August 2013, in conjunction with the Sappi annual routine field surveys. Sampling was conducted on *Eucalyptus grandis, Eucalyptus dunii* and *Pinus taeda* forests aged between 8 and 20 years (Figure 4.2). Selected tree structural variables, namely, tree diameter-at-breast height (DBH) and tree height (H) were measured for each plot (181), using the Haglof Digitech Calliper and Vertex IV laser instrument, respectively. A total of 181 plots were selected for field surveys, using vector maps, by courtesy of Sappi. The selection criteria were based on species type, age and the spatial location of compartments. The measurements were collected using a grid-based systematic sampling technique, utilizing circular sample plots, approximately 400 m^2 in size. These plots were systematically distributed (usually every 100 m) within the stand. Sample intensities varied between 2 and 10%, depending on the species composition, stand size or local forest conditions (Wessels and Kassier, 1985).

4.2.3 Field Aboveground Forest Biomass Computation

For individual species biomass (t ha^{-1}) calculation, two approaches were applied for the three selected species (*E. grandis, E. dunii* and *P. taeda*). The first method was only used for *Eucalyptus spp.* and it involved the use of volume and biomass expansion factors found in literature, specifically for South African species (Dovey, 2009). Volume was derived and reported at stand level, following the allometric method explained by Bredenkamp (2000). For *P. taeda*, a general allometric equation was used for biomass computation, as proposed by the Intergovernmental Panel on Climate Change (IPCC), (IPCC, 2007). The basis for the application of this allometric equation for this species *(P. taeda)* in particular, is the fact that the rainfall (800 - 1500 mm) and temperature range (21-34^oC) are similar to the climatic conditions prevailing in the study area. The equation used for the species was also formulated using a diameter-at-breast height (DBH), ranging from 0.6 cm-56 cm at a rainfall of 800 to 1500 mm and temperatures were similar to the study area. Species difference prompted the use of different approaches for computing biomass, because of the existing different structural and taxonomical characteristics (Atta-Boateng and William, 1998; Henry *et al.,* 2011; Lu, 2005a). Moreover, literature shows that different allometric equations exist for the biomass computation of the selected species (Henry *et al.,* 2011; Schönau and Boden, 1982a). The biomass results from the two approaches were finally standardised to the same unit of measurement, which is tonnes per hectare (t ha^{-1}).

a  b 

Figure 4.2 Typical field site showing (a) *Eucalyptus spp.* and (b) *P. taeda* in early August 2013

4.3 Image Acquisition and Data Pre-processing

A recent high spatial resolution spaceborne multispectral sensor (*i.e.* RapidEye imagery), with zero percent cloud cover, covering the study area, was obtained on the 25th of August 2013 from Germany. The RapidEye image comprised of five multispectral bands with a spatial resolution of 5 m. The spectral ranges of the five bands are 440 – 510 nm (B1-blue), 520 – 590 nm (B2- green), 630 – 685 nm (B3-red), 690 – 730 nm (B4-red-edge) and 760 – 850 nm (B5-near infrared). All the RapidEye products are collected by a 12-bit imager. Radiometric corrections were applied to the RapidEye image, subsequently converting the image digital numbers (DN) into values directly related to absolute radiances, using a constant factor (originally determined during launch) (RapidEye, 2011). Earlier experimentation by Naughton *et al.* (2011) demonstrated that the image registration error was within a single pixel, hence further geometric processing was not implemented. Radiance image was atmospherically corrected and transformed to canopy reflectance, using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm built in ENVI 4.7 software.

4.3.1 Spectral information and vegetation indices derived from strategically-positioned multispectral Spaceborne RapidEye image bands

A point map of the biomass plots was developed, using the field data and GPS recordings. This map was then overlaid on the RapidEye images to generate a region-of-interest (ROI) map, using the central GPS point for each plot (n = 181). A 3×3 pixels window (*i.e.* 15 m × 15 m) was used to collect vegetation image spectra from each band (n = 5), using ArcGIS 10 software. The 3×3 pixels window size was used, in order to avoid the inclusion of pixels located outside the plot (Cho *et al.,* 2007; Mutanga *et al.,* 2012). Hence, only pixels that fall entirely within the ROIs were included in

the spectral data set, while the pixels that partially fall inside the ROIs were discarded (Cho *et al.,* 2007; Mutanga *et al.,* 2012; Wang *et al.,* 2007). The spectra were collected and averaged for each plot. All derived parameters that were related to the field plot data are listed in Table 4.1. The indices were chosen, based on previous research dealing with forest biomass estimation from remote sensing data.

Table 4.1 Selected strategically-positioned RapidEye spectral parameters and vegetation indices used for this study

Parameters	Formula	References
Single band reflectance		
Blue, green, red, NIR and Red-edge	-	
Vegetation Indices		
Simple Ratio	NIR/Red	Jordan (1969a)
RVI.RE (Ratio Vegetation Index)	Red-edge/ NIR	de Sousa et al. (2012)
NDVI (Normalized Difference Vegetation Index)	(NIR – Red)/(NIR + Red)	Rouse et al. (1974); Jordan (1969)
NDVI.RE	(NIR – Red-edge)/(NIR + Red-edge)	Mutanga et al. (2012)
DVI (Difference Vegetation Index)	NIR – Red	Tucker (1979)
MSR (Modified Simple Ratio)	(NIR/Red)-1/ (NIR/Red)^$^{0.5}$+1	Qi et al. (1994)
MSR.RE	(NIR /Red-edge)-1/ (NIR /Red-edge)^$^{0.5}$+1	
TVI (Triangular Vegetation Index)	0.5*[120*(NIR−Green)−200*(Red-Green)]	Broge and Leblanc (2001)
TVI.RE	0.5*[120*(NIR−Green)−200*(Red-edge-Green)]	
IPVI (Perpendicular Vegetation Index)	NIR/(NIR + Red)	Crippen (1990)
IPVI.RE	NIR /(NIR+ Red-edge)	
GI (Greenness Index)	Green/Red	Zarco-Tejada et al. (2005)
GI.RE	Green/Red-edge	
PSSR (Pigment specific simple ratio)	NIR/Red-edge	Blackburn (1998)

4.4 Intra- and Inter-species Biomass Training and Test Data sets

To validate the performance of the SGB and RF algorithms (Adam *et al.,* 2014; Lawrence *et al.,* 2006), the data sets (*E.dunii*: n = 63, *E.grandis*: n = 65, *P.taeda*: n = 53 and all species: n = 181) were randomly split into 70%, for a training data set and 30% for a test (independent) data set. Moreover, the training data sets were used in optimizing both regression algorithms (SGB and RF) and to train the prediction models, whereas the test data set was used to examine the performance and reliability of the prediction model.

4.5 Statistical Analysis

Two main data analysis techniques were implemented and these included stochastic gradient boosting (SGB) and random forest (RF) regression algorithms. The two algorithms are discussed in detail below.

4.5.1 Stochastic Gradient Boosting Regression Model

Stochastic gradient boosting is a powerful machine learning technique, producing competitive, highly robust and interpretable procedures for both regression and classification applications (Friedman, 2001). The tree ensemble has the ability to accommodate different types of explanatory variables and data with missing values (De'ath, 2007). The ensemble is immune to outlier effects; it can fit complex nonlinear relationships and automatically handles interaction effects among predictors. The algorithm also introduces an element of stochasticity, thus improving model accuracy and reducing over-fitting (Elith *et al.,* 2008; Leathwick *et al.,* 2006).

SGB predicts the response variables by combining regression tree and boosting algorithms (De'ath, 2007; Elith *et al.,* 2008; Leathwick *et al.,* 2006; Pinkerton *et al.,* 2010; Ridgeway, 2006). The ensemble uses a backward stage-wise approach, by fitting regression tree models iteratively to a subset of the training data (50%) that is randomly selected, without replacement. A residual deviance is then calculated on data not used in the model fitting process. Trees are added until the total residual deviance calculated from the withheld data ceases to decrease. The number of trees giving the lowest total residual deviance represents the most appropriate model for prediction.

During model fitting, SGB is governed by three important user-defined parameters (Elith *et al.,* 2008; Hastie *et al.,* 2001; Leathwick *et al.,* 2006), that is: (i) the learning rate (*lr*), which determines the contribution of each tree to the final model, (ii) the tree complexity (*tc*), which is the number of

independent variables interacting to determine each split, and (iii) the number of regression trees (*nt)* in the ensemble. The learning rate controls the increase in model complexity, with smaller values resulting in fitting a larger number of trees (De'ath, 2007). For each combination of *nt*, *tc* and *lr,* the combination producing the lowest cross-validated deviance is then identified, using the training data set. For this study, SGB models were fitted with varying values for *nt* (100–10 000), *lr* (0·1–0·0001), *tc* of 1 and 5, as well as a bag fraction of 0.2-0.75, and the results were evaluated across all categories of species biomass. The gbm library (Ridgeway, 2006) for the R statistical package for statistical analysis (R Development Core Team, 2008) was utilized to implement SGB.

4.5.2 Stochastic Gradient Boosting and Relative Variable Importance

For the accurate and simple prediction of inter- and intra-species biomass, the relative individual variable influence was determined, to identify the smallest number of input variables (p =19) that yielded the best predictive performance. This information is important, because not all model input variables are equally relevant in the modelling process. In this regard, it is often suitable to learn or determine the relative influence of each input variable in predicting inter- and intra-species biomass. Based on SGB, the relative influence of model terms was calculated by the contribution of each variable in reducing overall model deviance (De'ath, 2007; Elith *et al.,* 2008). Subsequently, variable selection was achieved by implementing a backward feature elimination approach, to determine the most important spectral bands and vegetation indices required for accurate biomass prediction. More precisely, the approach develops a model which utilizes all the input predictor variables and then progressively eliminates input predictor variables with the least relative influence. In addition all SGB models are optimized in terms of their *lr*, *tc* and *nt* hyper-parameters. The SGB model for predicting inter-and-intra-species biomass was initially run, using nineteen variables.

4.5.3 Random Forest Regression Algorithms

Random forest (RF) is a machine-learning technique developed by Breiman (2001) that employs bootstrap aggregation, where a number of trees (*ntree*) are constructed, based on a random subset of samples derived from the training data. The RF regression algorithm utilizes bootstrap samples from the training data, without pruning, to grow a large number of decision trees (Dye *et al.,* 2011; Ismail and Mutanga, 2010; Prasad *et al.,* 2006). These trees assign each variable (RapidEye band reflectance or vegetation index) to a response value (biomass), using the averaged estimates that the value receives from the collection of all trees (Prasad *et al.,* 2006). The algorithm has an additional

modification of selecting only a random subset of candidate features (*mtry*) to determine the split at each node of a tree. This ensemble method uses recursive partitioning, to create multiple regression trees (*ntree*), and then averages the results of all trees (Breiman, 2001). RF algorithm is easy to implement as only two parameters (*ntree* and *mtry*) need to be optimized, based on the lowest root mean square error (RMSE) of prediction (Breiman, 2001; Coulston *et al.,* 2012; Özçift, 2011). The *ntree* parameter, the number of regression trees grown, based on a bootstrap sample of the observations (the default value is 500 trees), and *mtry* is the number of different predictors tested at each node (the default value is 1/3 of the total number of the variables). Thus, in this study the *ntree* parameter values were tested in increments of 500 to 2500, with a 500 interval, whereas the *mtry* was tested in increments of 1 to 19.

Approximately one-third of the data, that is not included in the bootstrapped training sample, called the out-of-bag (OOB) samples, is then used to evaluate the RF model. A number of researchers have shown that the OOB samples offer unbiased estimates of the training error (Adam *et al.,* 2014; Adelabu *et al.,* 2013a; Breiman, 2001; Palmer *et al.,* 2007; Powell *et al.,* 2010; Prasad *et al.,* 2006). The permutation based variable importance follows the rationale that the random permutation of a predictor variable represents the absence of the variable from the model. Hence, the difference in prediction accuracy, prior to and after permuting a variable, is used as a measure of importance. The number of observations predicted correctly, decreases substantially if the permuted variable is strongly associated with the response values. Grömping (2009) provides a more detailed account of the random forest's variable importance measures, both from the theoretical understanding and from the perspective of computational advantages. The ensemble was implemented, using the random Forest package (Liaw and Wiener, 2002) within the R statistical package version R-3.0.2 (R Development Core Team, 2008).

4.5.4 Variables Selection using Random Forest

Random forest measures the importance of each predictive variable, using the mean decrease in accuracy that is calculated from the OOB sample data. However, the challenge was to select the fewest number of predictors that offer the best predictive power and that help in the interpretation of the final model. In this regard, a backward feature elimination method (BFE), integrated with random forest regression as part of the evaluation process, was implemented (RF), based on 1000 model runs. The BFE uses the ranking to identify the sequence in which to discard the least important predictors from the input data sets. The method starts with the entire variables (p = 19)

and then progressively eliminates the least promising variable from the list. For each iteration, the model is optimized by selecting the best *mtry* and *ntree*. The least promising variable is then eliminated and the root mean square error is calculated. The smallest subset of variables, with the lowest RMSE, is then selected to predict inter-species biomass. A comprehensive analysis of the predictive performance of the different subsets of extracted RapidEye reflectance and vegetation indices was implemented, to explore the role of the new generation sensor in predicting interspecies biomass, as well as to test if the variables selection method implemented in this study can enhance the predictive performance of random forest regression model.

4.6 Effectiveness of SGB and RF in Predicting Intra- and inter-species Biomass

To assess the effectiveness of SGB and RF algorithms in predicting either intra- or inter-species biomass in a commercial forest environment, the r-square (R^2) and the root mean square error (RSME) were computed (Eq. 4.1). A one-to-one relationship between measured and predicted AGB values was fitted with coefficients of determination (R^2) and root mean square error (RMSE) values were reported.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{measured} - X_{predicted})^2}{n}} \qquad \text{Equation. 4.1}$$

Where $X_{measured}$ is measured biomass values, $X_{predicted}$ is predicted biomass values and i represent each of the predictor variables included in the summation process (p = 19).

4.7 Results

4.7.1 Intra- and Inter-species Aboveground Biomass (t ha^{-1})

Table 4.2 shows descriptive statistics for each category of the target species (*e.g. E. dunii* (n = 63), *E. grandis* (n = 65), *P. taeda* (n = 53) and for the all species-data sets (n = 181)). High biomass was observed for *P. taeda*, followed by *E. grandis,* and *E. dunii* having the least biomass.

Table 4.2 Descriptive statistics of the measured above ground biomass (t ha^{-1})

Species Type	Total	Min.	Max.	Mean	Std dev.
E. dunii	63	33.24	96.49	52.86	16.39
E. grandis	65	106.03	225.07	170.30	29.94
P. taeda	53	137.11	298.04	206.07	42.83
All species	181	33.24	298.04	139.89	72.22

4.7.2 Intra-species AGB: SGB and RF Regression Predictive Performance based on all Variables

One-to-one relationship between measured and predicted intra-species biomass, using the SGB and RF regression models, are shown in Figure 4.3. For each model, the R^2 and RMSE were reported. A comparative analysis of the predictive performance of the two models shows that the SGB model yielded better predictions for the intra-species data set, producing R^2 of 0.75 and RMSE of 18.40 t ha^{-1} (10.80%) for *E. grandis*; R^2 of 0.77 and RMSE of 19.43 t ha^{-1} (19.18%) for *P. taeda.* Comparatively, the RF produced better results for *E. dunii* (R^2 of 0.74 and RMSE of 8.14 t ha^{-1}).

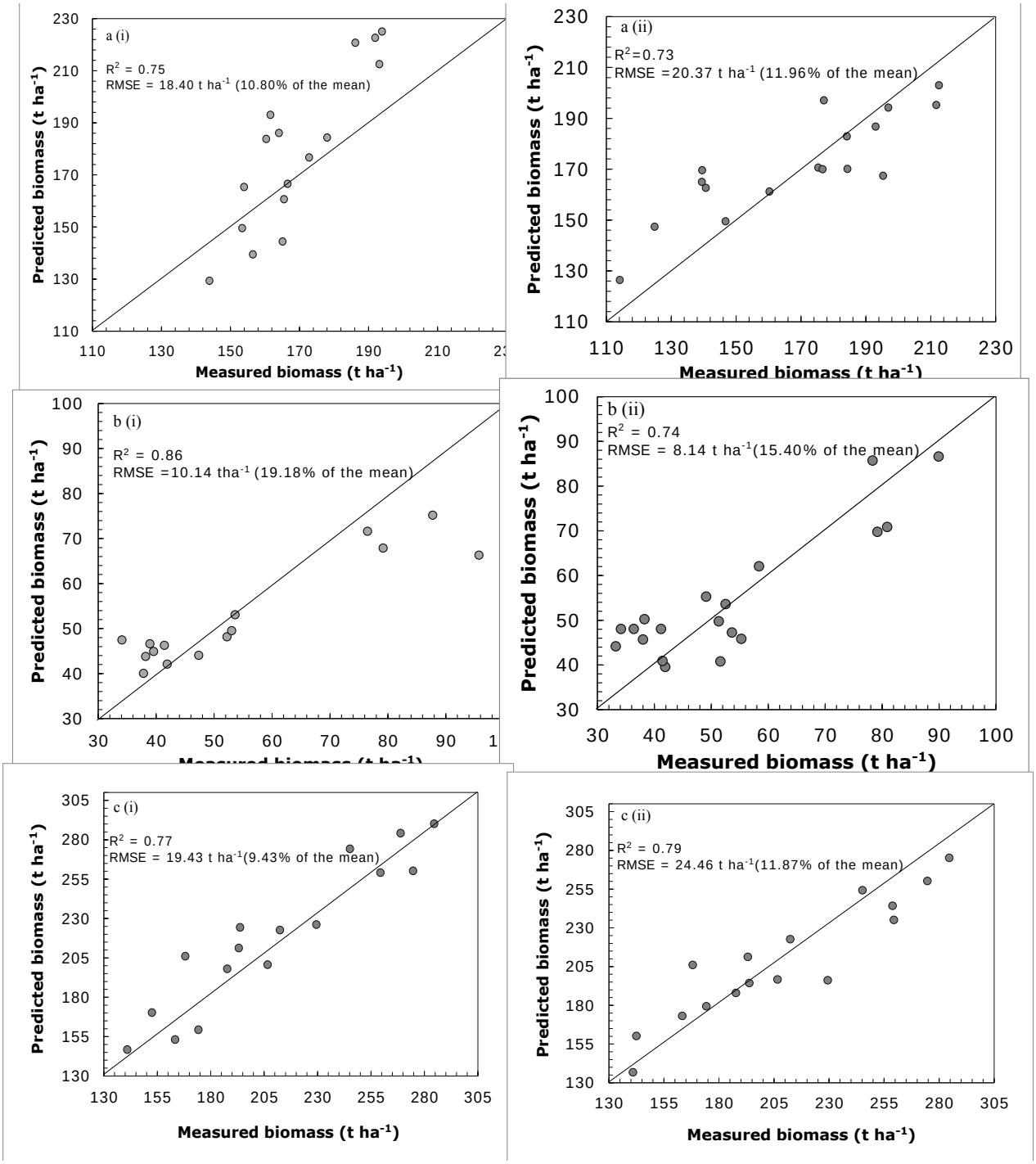


Figure 4.3 One-to-one relationship between measured and predicted intra-species biomass based on: (i) SGB, and (ii) RF algorithms. a, b, and c represent *E. grandis*, *E. dunii,* and *P. taeda* based on all the predictor variables (p = 19)

4.7.3 Inter-species AGB: SGB and RF Regression Predictive Performance based all on Variables

In testing the potential of SGB and RF in predicting inter-species biomass it can be observed that SGB produced plausible results based on the R^2 of 0.58 and RMSE of 46.51 t ha^{-1}; 33.25% of the mean, compared to RF, which had an R^2 of 0.33 and RMSE of 64.27 t ha^{-1}; 45.94% of the mean (Figure 4.4).

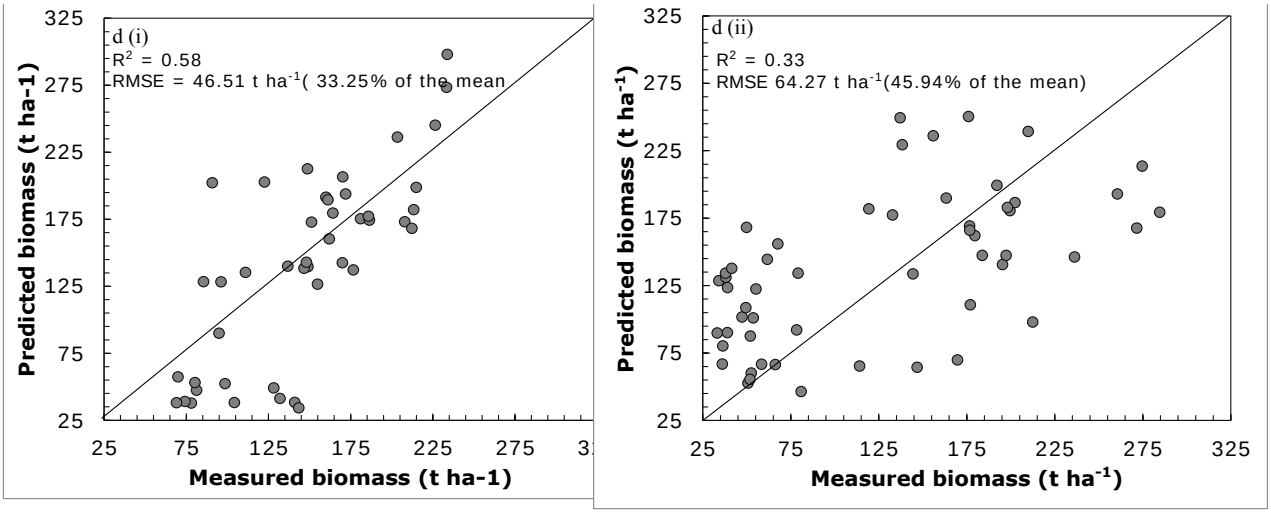


Figure 4.4 The one-to-one relationship between measured and predicted inter-species biomass for all species data combined, based on: (i) SGB, and (ii) RF algorithms without variable selection

4.7.4 Variable Selection using SGB and RF Models

The SGB and RF variable importance measures were used to explore the relevance of model input variables (strategically-positioned RapidEye spectral bands and derived vegetation indices). The backward variable selection provided by the two algorithms (SGB and RF) have successfully explored and defined the relative importance of the individual input variables (predictors). In addition the methods further managed to select the optimal number of the input variables for predicting intra-and-inter-species AGB. For SGB, better results were achieved after variable selection was implemented, see Table 4.4. The SGB backward variable selection method selected a few optimal numbers of important variables for: (a) *E. grandis* (p = 4), (b) *E. dunii* (p = 7), (c) *P. taeda* (p = 6), and (d) all species data combined (p = 19), using the optimal *nt* and *lr,* which resulted in deviance reduction (Table 4.3). For instance, *E. grandis* achieved the lowest predictive deviance (0.27), based on *nt* = 2350, *lr* = 0.001 and *tc* = 3. *E. dunii,* on the other hand, yielded better results (lowest deviance value), based on a value of *lr* = 0·001*, nt* = 3750 and *tc*= 3. Similarly, for *P. taeda*

and all species combined, a value of *nt* = 2850, *lr* = 0.001 and *tc* = 3 produced the best results, with the lowest deviance.

For the RF ensemble, the optimal number of variables was determined, based on the lowest averaged RMSE obtained after running the backward feature elimination process 1000 times. The process selected four variables for predicting (a) *E. grandis,* based on averaged RMSE of 26.10 t ha^{-1}*;* seven predictor variables for (b) *E. dunii,* based on an averaged RMSE value of 10.87 t ha^{-1}*;* six variables for (c) *P. taeda,* with an averaged RMSE value of 31.65 t ha^{-1}, and lastly, nineteen variables for (d) all species data sets, based on a RMSE value of 50.76 t ha^{-1} (Figure 4.5). The findings in Figure 4.5 further demonstrate that the RMSE error generally decreased as the least important variables were progressively removed from the model. The use of the most important RF selected variables produced the lowest RMSE across all species categories. To conclude, important variables selected by SGB and RF (Table 4.3) were used in the final model for predicting biomass across all species categories, using the test data set (Table 4.4).

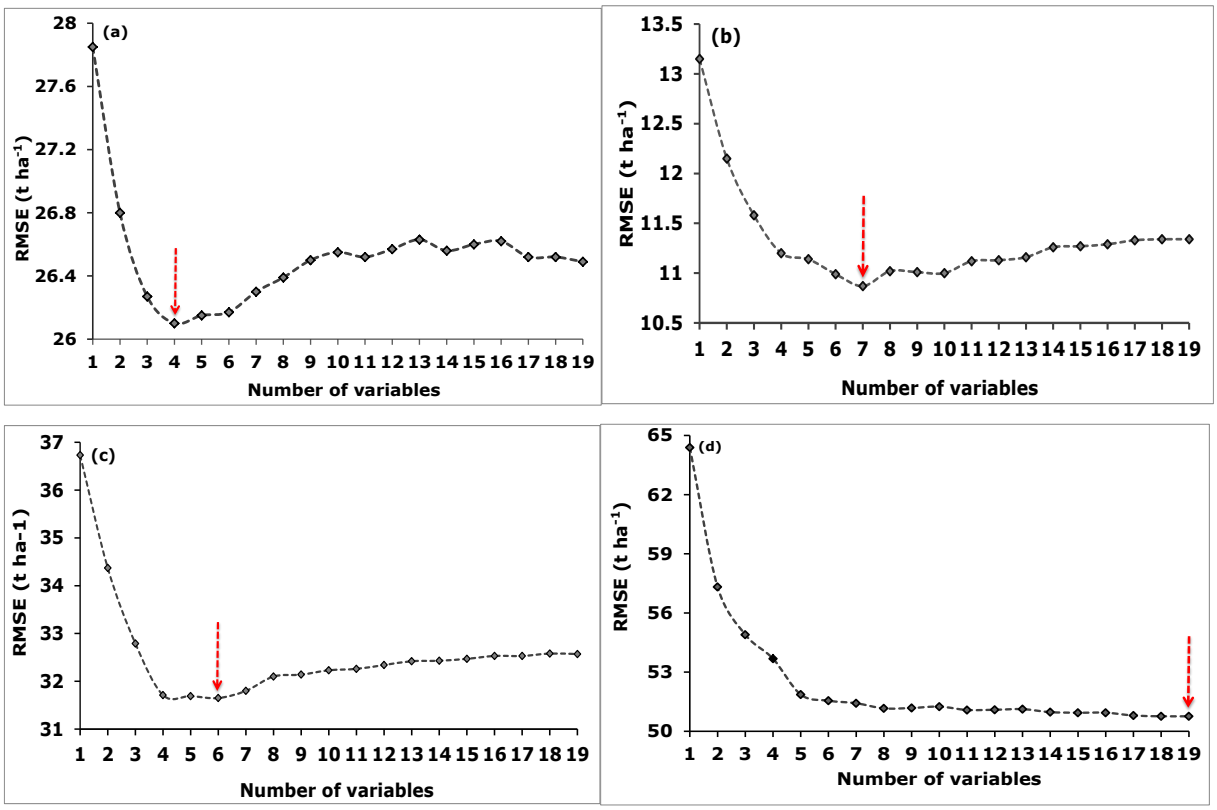


Figure 4.5 The optimal number of variables (spectral bands and VIs), based on the backward feature elimination search function for estimating intra-and-inter species using Random Forest (based on 1000 repetitions). (a), (b), (c) and (d) represent *E. grandis, E. dunii, P. taeda* and inter-species data set. *The best number of variables with the

lowest error rate is shown by the arrows and the RMSE is calculated from the training data set.

The results in Table 4.3 show the most important predictor variables that were selected for estimating intra- and inter-species biomass. Most interestingly, the results from both models show that a limited and similar number of input variables contribute to intra-and-inter-species biomass prediction. It can be observed that in predicting intra- and inter-species biomass, the NIR, red-edge, and Red bands are selected across all categories by both algorithms (Table 4. 3).

Table 4.3 Illustrates the most important variables retained by SGB and RF after implementing variable selection.

	E. grandis		*E. dunii*		*P. taeda*		*All species*	
Variables selected	SGB	RF	SGB	RF	SGB	RF	SGB	RF
1	RE	NIR	NIR	RE	NIR	NIR		
2	PSSR	RE	RE	PSSR	Green	Green		
3	GI.RE	PSSR	Red	GI.RE	RE	RE	All variables	
4	NIR	DVI	GI.RE	NIR	Red	Red	selected	
5	Green	-	Green	Green	DVI	DVI		
6	-	-	Blue	DVI		Blue		
7	-	-	-	Blue	-	-		

4.7.5 Intra-species AGB: SGB and RF regression predictive performance, using selected variables

Table 4.4 demonstrates the aboveground biomass prediction results obtained using the SGB and RF algorithms and the most important variables are shown in Table 4.3. It can be observed that inter-and-intra species biomass predictions, based on the most important variables, provide better predictive accuracies, when compared to the SGB and RF models that use all the predictor variables (Figure 4.3 and 4.4) The SGB model produced good accuracies in predicting *E. grandis* (R^2 = 0.80, RMSE = 16.93), *P. taeda* (R^2 = 0.79, RMSE = 17.27 t ha^{-1}) and the all species data (R^2 = 0.61, RMSE = 43.39 t ha^{-1}). The RF ensemble, however, demonstrated better results (R^2 = 0.79; RMSE 7.18 t ha^{-1}) in predicting *the biomass E. dunii* (Table 4.4).

Table 4.4 Inter-and-intra species biomass prediction results using the most important variables selected by the two regression models SGB and RF

Species	Statistical methods	tc	lr	mtry	nt/ntree	R^2	RMSE (t ha^{-1})
E. grandis	SGB (n = 5)	3	0.001	-	3750	0.80	16.93
	RF (n = 4)	-	-	4	500	0.76	18.61
E.dunii	SGB (n = 6)	5	0.001	-	2350	0.88	09.23
	RF (n = 7)	-	-	7	500	0.79	07.18
P.taeda	SGB (n = 5)	5	0.001	-	800	0.79	17.27
	RF (n = 6)	-	-	6	2800	0.80	22.43
All species	SGB (n = 19)	5	0.001	-	2800	0.61	43.39
	RF (n = 19)	-	-	19	750	0.37	59.27

4.8 Discussion

The accurate, reliable and timely quantification of intra-and-inter-species AGB, using remote sensing technologies, is critical for better understanding the role of forests in local climate systems through biosphere-atmospheric interactions, for a detailed evaluation of commercial forest resources, as well as for informed sustainable management. In this study, two machine learning regression algorithms, namely, SGB and RF based on 1000 model runs, were assessed in predicting intra-and-inter-species biomass in a commercial plantation forest located in the midlands region of KwaZulu-Natal, South Africa, using the RapidEye Spaceborne sensor.

4.8.1 RapidEye image potential in predicting intra- and inter-species aboveground biomass

One of the most critical challenges in predicting biomass in plantation forests, using remote sensing, is the complexity of species structural and taxonomic composition, as well as the presence of dense vegetation canopies, resulting in significant inter-species biomass variations. It is therefore critical to identify remote sensing data sets with critical spectral information that can overcome the saturation problems and produce better intra-and-inter-species biomass prediction accuracies. In this study, it was shown that high spatial resolution RapidEye image data, with strategically-positioned bands, can accurately predict intra-and-inter-species biomass in commercial forests, when compared to the existing broadband multispectral data, which have high spectral variation and saturation problems at high density biomass. Furthermore, this study demonstrated new generation multispectral sensors as having the capability to provide a better and cost-effective alternative for predicting inter-species biomass, when compared to existing broadband multispectral images (Adam *et al.,* 2014; Eckert, 2012b; Mutanga *et al.,* 2012). Most importantly, the presence of the red-edge band, which has been unavailable in existing multispectral sensors, provides very critical and sensitive measurements of vegetation properties, such as chlorophyll content, necessary

for predicting forest metrics, such as biomass etc. (Mutanga *et al.,* 2012; Ozdemir and Karnieli, 2011). The findings from this study therefore largely support the claim that strategically-positioned bands (*e.g.* red-edge) found in new generation RapidEye multispectral imagery, contains more spectral information critical for vegetation mapping, when compared to other broadband multispectral sensors.

4.8.2 SGB and RF prediction performance using different RapidEye spectral parameters

Stochastic gradient boosting has increasingly been used in ecological modelling, with limited application in remote sensing studies (*e.g.* Buston and Elith, 2011; De'ath, 2007; Elith *et al.,* 2008; Froeschke and Froeschke, 2011; He *et al.,* 2009; Kint *et al.,* 2012; Leathwick *et al.,* 2006; Lewin *et al.,* 2014; Pinkerton *et al.,* 2010; Soykan *et al.,* 2014). On the other hand, random forest has been applied mainly in classification (*e.g.* Adam *et al.,* 2011; Adelabu *et al.,* 2013a; 2013b; Chan and Paelinckx, 2008; Gislason *et al.,* 2006; Guo *et al.,* 2011; Lawrence *et al.,* 2006; Pal, 2003; Stumpf and Kerle, 2011), hence there are limited remote sensing studies that utilize SGB and RF for regression analysis (*e.g.* Baccini *et al.,* 2004; Carreiras *et al.,* 2012a; 2012b; Mutanga *et al.,* 2012). The results of the present work have demonstrated the applicability and strength of the two algorithms (SGB and RF) for variable selection and intra-and-inter-species biomass prediction, using the spaceborne RapidEye imagery.

Moreover, for the two different algorithms applied, the better results, based on the R^2 and RMSE, were obtained from the SGB model across all species categories, except for the *E. dunii* data set. The results of the present study further demonstrated that SGB and RF models are useful and robust for intra-species biomass prediction, using remotely sensed data. For the prediction of all inter-species biomass (species data combined), the RF model performed poorly, when using all the variables. This poor performance of RF is attributed to the high variability in biomass, as a result of the existing differences amongst the tree species considered in this study. The results of this study have shown that RF is less robust in an environment with mixed species, when compared to the SGB algorithm. Furthermore, literature shows that the RF regression algorithm results in AGB underestimates, when the data set is large and variable, as well as overestimates, when the data is small with less variability (Horning, 2010; Mutanga *et al.,* 2012).

For the SGB model algorithm, plausible inter-species biomass prediction results were observed, indicating the model's robustness in handling non-linear inter-species biomass relationships. The

good performance of the SGB regression algorithm can be associated with the model's internal regularization process and the element of stochasticity, which are well-known for enhancing its predictive performance (Elith *et al.,* 2008; Ganjisaffar *et al.,* 2011; Hastie *et al.,* 2001; Suen *et al.,* 2005). These results are further supported by Carreiras *et al*. (2012a) whose work demonstrated that the simple base learner, in our case, decision trees, built by running the SGB model using a random sub-sample of the training data without replacement, substantially improved the prediction accuracy. However, the effectiveness and robustness of the SGB algorithm in variable selection, based on remotely sensed data sets, still needs to be tested in the mapping and understanding of other vegetation metrics, such as aboveground carbon stocks, in non-homogenous forested areas. This information would aid in assessing the contribution of forests to carbon sequestration, as well as for a comprehensive evaluation of commercial forest resources, which is a pre-requisite for informed sustainable management.

4.9 Conclusion

This paper investigated: (i) the robustness of two machine-learning algorithms, Stochastic Gradient Boosting and Random Forest regression trees, to predict intra- and-inter-species biomass in plantation forests, using RapidEye multispectral imagery in KwaZulu-Natal, South Africa, and (ii) the performance and the strength of the SGB and RF regression algorithms as variable selection and prediction methods.

The results have demonstrated that:

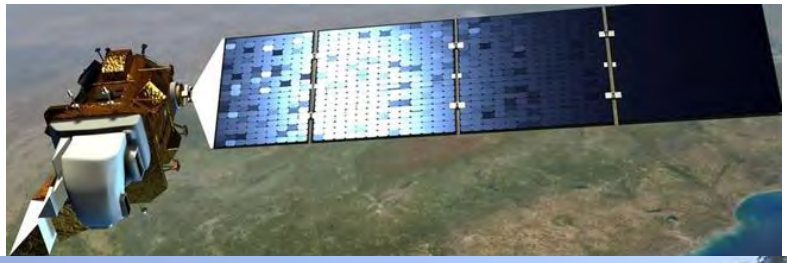
1. stochastic Gradient Boosting regression tree is more robust in predicting both intra-and-inter-species biomass in plantation forests, when integrated with the strategically-positioned bands of the multispectral spaceborne RapidEye imagery, compared to the Random Forest ensemble,
2. the new generation Spaceborne multispectral sensors (*e.g.* RapidEye) with a high spatial resolution have the potential of predicting intra-and-inter-species biomass in areas of closed and dense vegetation,
3. both machine-learning algorithms (SGB and RF regression trees) were able to provide a valuable screening tool for the identification of the most important spectral bands and derived vegetation indices required, to accurately predict inter-and-intra-species biomass.

Overall, results of the present study demonstrate the utility, great potential and robustness of the high spatial resolution RapidEye satellite data, with strategically-positioned spectral information and SGB regression algorithm, in modelling non-linear biomass relationships for mixed forests, which was a previously challenging task with broadband satellite sensors. Considering the relatively high inter- and intra-species AGB accuracies, derived from high cost and spatially restricted RapidEye satellite data, there is, however, the need to identify data sets that are relatively cheap and that provide a regional spatial coverage required for AGB estimation in resource-scarce areas. Such information is necessary, if regional understanding of the contribution of forest to the global carbon cycle is to be achieved.

CHAPTERS FIVE AND SIX

REGIONAL SCALE MAPPING

5. THE UTILITY OF THE MEDIUM-SPATIAL RESOLUTION LANDSAT-8 OLI MULTISPECTRAL SENSOR IN QUANTIFYING ABOVEGROUND BIOMASS



This chapter is based on:

Dube, T, Mutanga, O. 2015. Evaluating the utility of the medium-spatial resolution Landsat-8 OLI multispectral sensor in quantifying aboveground biomass in the uMngeni Catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing***. 101:** 36–46.

Presented at the **2nd National Conference on Global Change,** Nelson Mandela Metropolitan University, Port Elizabeth, South Africa.

Abstract

Aboveground biomass estimation is critical in understanding forest contribution to regional carbon cycles. Despite the successful application of high spatial and spectral resolution sensors in aboveground biomass estimation, there are challenges related to high acquisition costs, small area coverage, multi-collinearity and limited availability. These challenges hamper the successful regional scale aboveground biomass quantification. The aim of this study was to assess the utility of the newly-launched medium-resolution multispectral Landsat-8 OLI data set, with a large swath width, in quantifying aboveground biomass in a forest plantation. Different sets of spectral analysis (Test I: spectral bands; Test II: spectral vegetation indices and Test III: spectral bands + spectral vegetation indices) were applied in testing the utility of Landsat-8 OLI, using two non-parametric algorithms, stochastic gradient boosting and the random forest ensembles. The results of the study show that, when compared to Landsat ETM+, the medium-resolution multispectral Landsat 8-OLI data set provides better aboveground biomass estimates for *Eucalyptus dunii*, *Eucalyptus grandis* and *Pinus taeda,* especially when using the extracted spectral information together with the derived spectral vegetation indices. It was also noted that incorporating the optimal subset of the most important selected medium-resolution multispectral Landsat-8 OLI bands, improved aboveground biomass accuracies. Medium-resolution multispectral Landsat-8 OLI aboveground biomass estimates were compared with Landsat 7 ETM+ estimates and the latter yielded lower estimation accuracies. Overall, this study demonstrates the invaluable potential and strength of applying the relatively-affordable and readily-available newly-launched medium-resolution, Landsat-8 OLI data set, with a large swath width (185 km) in precisely estimating aboveground biomass. This strength of the Landsat-8 OLI data set is crucial, especially in sub-Saharan Africa, where high-resolution remote sensing data availability remains a challenge.

Keywords: biomass estimation; Landsat data continuity mission; Landsat 7 ETM+, medium spatial resolution; Operational land imager

5.1 Introduction

The great improvement in understanding the nexus between AGB and climate change in recent years has prompted the need for timely and reliable AGB mapping, using numerous remote sensing techniques (Gara *et al.,* 2014b; Gibbs *et al.,* 2007; Houghton *et al.,* 1996; IPCC, 2003a). Airborne hyperspectral instruments, new generation multispectral sensors and airborne light detection and ranging (lidar) instruments have proven useful in providing fine spatial and spectral resolution data sets that are primarily required to accurately estimate structural attributes, such as wood volume and AGB or carbon content of different tree species (Nelson *et al.,* 2003; Patenaude *et al.,* 2004; Zhao *et al.,* 2009). However, they are associated with limited spatial coverage, due to high costs, availability, huge data-volumes, as well as high data pre-processing costs. In this regard, these remote sensing technologies are not practical for "wall-to-wall" AGB or carbon applications (Carreiras *et al.,* 2012b; Colgan *et al.,* 2012; Dube *et al.,* 2014a; Wulder *et al.,* 2008b).

The limitations of using hyperspectral remote sensing technologies have also been cited by Mathieu *et al.* (2013), who described airborne-sensor data collection as constrained by operational restrictions on data-acquisition flight campaigns and the variability of data accessibility on a country basis. These data sets are therefore more suitable for "project-based applications", or for use in sampling strategies (Carreiras *et al.,* 2012b). The abovementioned limitations have therefore seen a shift towards the use of free and readily-available broadband multispectral sensors, with a large swath width, such as Landsat data sets. These data sets allow timely and regional scale AGB or carbon accounting (Gibbs *et al.,* 2007; Hall *et al.,* 2011; Houghton *et al.,* 1996; Vaglio Laurin *et al.,* 2014) and sustainable forest resources management and inventorying (Næsset, 2007).

The newly-launched Landsat-8 OLI multispectral sensor is assumed to present a number of invaluable opportunities for understanding the regional contribution of forest ecosystems to the carbon cycle. Moreover, the opening and free distribution of its archive digital data set, with a wider swath width of 185 km and a 16-day temporal resolution, since its launch on the 11th of February 2013, makes the Landsat-8 OLI sensor one of the key primary data sources, highly suitable and practical for regional AGB analysis, especially in resource-limited areas. Furthermore, the newly-launched medium-resolution multispectral Landsat-8 OLI sensor provides: (i) a refined spectral range for certain bands, that is critical for improving the vegetation spectral responses across the near-infrared (NIR) and panchromatic bands, (ii) improved radiometric resolution from 8 bits to 12 bits, which is critical in enabling the characterization of different forest conditions (El-Askary *et al.,*

2014; Pahlevan and Schott, 2013), (iii) the changes in sensor design that have also resulted in substantial improvements in signal to noise ratios (SNR), almost twice as good as Landsat 7 Enhanced Thematic Mapper plus: ETM+ (Irons *et al.,* 2012), and (iv) the narrow NIR band, which prevents water vapour absorption effect at 0.825 μm, critical in allowing accurate surface reflectance acquisition. In addition increased Landsat-8 OLI sensor dwell-time in sampling radiation at each ground pixel is perceived to improve radiometric resolution, which, in turn, increases the spectral record precision and eliminates sensor spectral saturation problems, a common phenomenon to prior Landsat products (Irons *et al.,* 2012; Jia *et al.,* 2014; Verbyla, 1995). Furthermore, the freely-available medium-resolution multispectral Landsat-8 OLI sensor's large swath width of 185-km and frequent repeat cycle provide critical spatial information that is compatible with the size of vegetation units and AGB field samples (Hall *et al.,* 2006).

In addition, the Landsat-8 OLI employs long detector arrays, with more than 7 000 detectors per spectral band, aligned across its focal plane to view across the swath width, unlike the earlier Landsat satellite instruments. For instance, the earlier Landsat satellite instrument utilized oscillating whiskbroom mirrors that scanned the field of view across-track and then transmitted light to a few detectors. This "push-broom" (along track) design is believed to enhance the sensitivity of the instrument, enabling improved detection of the most important forest structural properties. Based on the newly-launched Landsat-8 OLI sensor's technical advancements, this study is therefore of the view that the sensor provides the primary data source required for spatially explicit and accurate assessment of AGB and its dynamics, mainly at regional scales. To the best of our knowledge, the newly-launched medium-resolution multispectral Landsat-8 OLI has not been tested for the remote estimation of AGB, despite its documented abilities.

In this study, the utility of the medium-resolution multispectral Landsat-8 OLI was assessed in estimating plantation forests AGB, using different error measures, namely, root mean square error, percentage bias and the mean absolute error. Specifically, our first objective was to estimate AGB for *Eucalyptus spp.* and *Pinus spp.* in the Mgeni Catchment, KwaZulu-Natal Province, South Africa, using the medium-resolution multispectral Landsat-8 OLI sensor. Secondly, the effectiveness of Landsat-8 OLI for AGB estimates was assessed by comparing it with Landsat 7 ETM+, to confirm if the improved changes have significantly improved the Landsat series data.

5.2 Materials and Methods

5.2.1 Study Area

The study was conducted at the Clan Sappi Forests (a paper and pulp company) area (between Latitudes 29°24'46.74"S, 29°17'45.94"S and Longitude 30°18'32.89"E, 30°28'28.21"E), part of the uMgeni Catchment, situated between Greytown and Pietermaritzburg in the province of KwaZulu-Natal, South Africa (Figure 5.1). The Clan area is characterised by gentle undulating terrain, with apedal and plinthic soil classes of the ecca group (Sappi, 1993). Sub-tropical climatic conditions prevail, with the rainy season occurring from October to February and the mean annual rainfall variability ranging from 700 mm to 1500 mm. The mean annual temperature is approximately 21.7ºC, which, together with high summer rainfall, provides favourable conditions for the production of various commercial forests (Scott and Lesch, 1997). For example, the area is currently home to various *Eucalyptus* and *Pinus* trees, occupying approximately 6 700 ha, that are mainly grown for pulpwood.

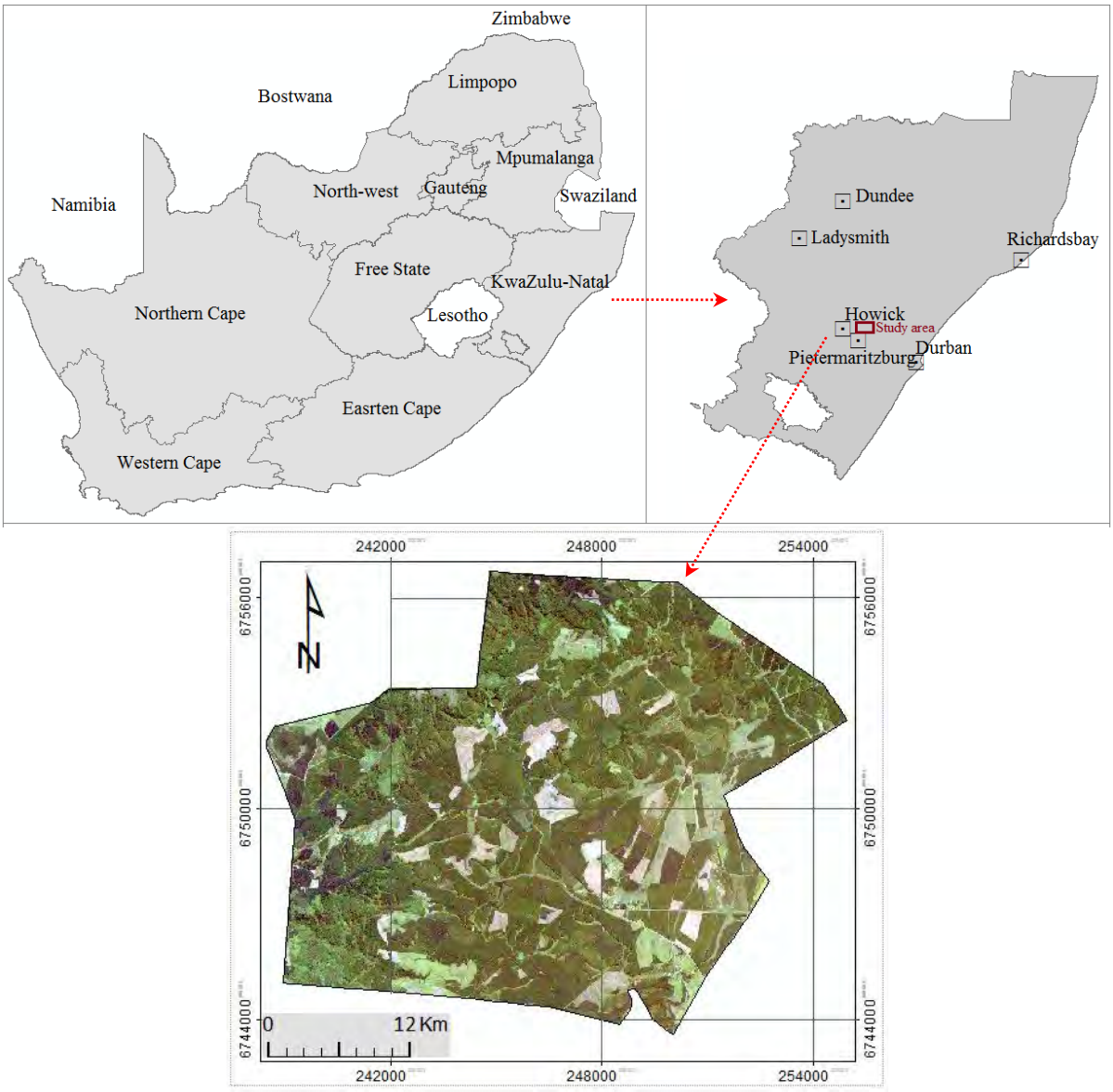


Figure 5.1 Location of the study area in uMgeni Catchment, South Africa.

5.2.2 Field Measurements

Tree diameter-at-breast height (DBH) and tree height (Ht) were measured on 181 plots. DBH and Ht were used to generate *in-situ* AGB estimates to train the remote AGB-retrieval models and assess estimation accuracy. The field data collection was primarily conducted from July 30th to August 22nd 2013, in conjunction with the Sappi annual routine field surveys. The measurements were mainly done on three commercial forest species found in the Clan forest area: (a) *Eucalyptus grandis* (EG), (b) *Eucalyptus dunii* (ED), and (c) *Pinus taeda* (PT), aged between 8 and 20 years. DBH and Ht were measured, using the Haglof Digitech Calliper and Vertex IV laser instrument, respectively. The measurements were taken, using a grid-based systematic sampling technique, with

approximately 400 m^2 circular sample plots, distributed systematically (~100 m) within the stand (Wessels and Kassier, 1985). The 181 plots were selected based on vector maps, by courtesy of Sappi forest. The selection criterion was a function of species type, age and spatial location of the compartments.

5.2.3 Aboveground Forest Biomass Computation

The Intergovernmental Panel on Climate Change (IPCC) recommends the use of standard techniques for the field measurements of different vegetation structural attributes. For this study, three different tree species (*i.e.* EG, ED and PT), with varying structural and taxonomical characteristics, were assessed for aboveground forest biomass, based on two different aboveground biomass (t ha^{-1}) computational techniques; (i) volume (m^3ha^{-1}) and Biomass Expansion Factors (BEFs), and (ii) general allometric equations (Atta-Boateng and William, 1998; Henry *et al.,* 2011; Lu, 2005a; Schönau and Boden, 1982). The volume (m^3ha^{-1}) and BEFs approach was only applied for the *Eucalyptus species* and it involved the use of the volume derived and reported at stand level, following the allometric method explained by Bredenkamp (2000), and BEFs available in literature, specifically for South African species (Dovey, 2009). For *P. taeda*, a general allometric equation was used for biomass computation, as proposed by the IPCC (2007). The basis for the application of this allometric equation, for this species *(P. taeda)* in particular, is the fact that the rainfall (800 - 1500 mm) and temperature range (21^o-34^oC) are similar to the climatic conditions prevailing in the study area. Similarly, the equation used for the species was also formulated, using diameter-at-breast height (DBH), ranging from 0.6 cm-56 cm, at rainfall and temperatures similar to the study area. Species differences prompted the use of different approaches for computing biomass because of the existing differences in species structural and taxonomical characteristics (Atta-Boateng and William, 1998; Henry *et al.,* 2011; Lu, 2005a). Moreover, literature shows that different allometric equations exist for the biomass computation of the selected species (Chinembiri *et al.,* 2013a; Gara *et al.,* 2014b; Henry *et al.,* 2011; Schönau and Boden, 1982a). The biomass results from the two approaches were finally standardised to the same unit of measurement, which are tonnes per hectare (t ha^{-1}).

5.2.4 Remote Sensing Data Acquisition and Pre-processing

The study region is covered by one Landsat tile (path/row: 168/80). Cloud-free Landsat-7 ETM+ and Landsat Data Continuity Mission (LDCM)'s Landsat-8 OLI satellite images covering the region of interest (Clan Sappi plantation area) were acquired during the time that coincided with the field

measurement dates, that is the 3rd and 11th of August 2013. The Landsat images were acquired during sunny and clear sky day conditions, with cloud cover less than 10%, a sun azimuth angle of 39.78 and a sun elevation angle of 33.14. The satellite images were obtained from the USGS Earth Resources Observation and Science (EROS) Center archive (http://earthexplorer.usgs.gov/). The Landsat-8 OLI sensor launched on the 11th of February 2013 has a 16-day temporal resolution. On board the Landsat-8 OLI sensor are two push-broom instruments: (i) the Operational Land Imager (OLI), consisting of nine spectral bands, (Table 5.1), and (ii) the Thermal Infrared Sensor (TIRS), which encompasses thermal bands 10 and 11 at a 100 m spatial resolution. On the other hand, ETM+ on Landsat 7 satellite has spectral bands similar to those of the prior TM 5, except for the inclusion of the high resolution (15m) panchromatic band eight.

Table 5.1 OLI and ETM+ spectral and spatial characteristics

OLI spectral bands			ETM+ Spectral bands		
Band #	Bandwidth (λ ; µm)	GSD (m)	Band #	Bandwidth (λ ; µm)	GSD (m)
1	0.433-0.453	30			
2	0.450-0.515	30	1	0.450 - 0.515	30
3	0.525-0.600	30	2	0.525 - 0.605	30
4	0.630-0.680	30	3	0.630 - 0.690	30
5	0.845-0.885	30	4	0.750 - 0.900	30
6	1.560-1.660	30	5	1.550 - 1.750	30
7	2.100-2.300	30	7	2.090 - 2.350	30
8	0.500-0.680	15	8	0.520 - 0.900	30
9	1.360-1.390	30			

Source: http://landsat.usgs.gov

The two Landsat scenes were obtained in digital number (DN), hence the need to be converted to reflectance values. Firstly, Landsat ETM+ was converted to Top-Of-Atmosphere (TOA) spectral radiances and then to at-sensor reflectances, following procedures provided by Chander *et al.* (2009) using the reflectance rescaling coefficients provided in the image's metadata files. Since all Landsat TM+ images, acquired from 2003 to the present, have scan-line errors (stripes), the image bands were corrected for errors, using the Landsat toolbox in ArcGIS 10.2 (Law and Collins, 2013), before being converted to reflectance (Walawender *et al.,* 2012; Wulder *et al.,* 2008a). For Landsat-8 OLI, conversion from DN to reflectance was implemented in ENVI environment, following the approach described on the USGS website (http://landsat.usgs.gov). Atmospheric correction of Landsat ETM+ and Landsat-8 OLI images to surface reflectance was performed, using the Fast Line-of-sight

Atmospheric Analysis of Spectral Hypercube (FLAASH) radiative transfer model (Perkins *et al.,* 2005). The two images were then geometrically corrected, using ground control points (GCPs).

5.3 Regression Algorithms

Commercial forest aboveground biomass estimation was determined using two powerful regression algorithms, stochastic gradient boosting (Elith *et al.,* 2008) and random forest (Breiman, 2001). Stochastic gradient boosting algorithm integrates the strengths of regression trees and boosting, which constructs and integrates a group of different constructed models and involves a probabilistic component to decrease final model variance, at the same time improving, on the predictive accuracy (De'ath, 2007; Elith *et al.,* 2008; Gara *et al.,* 2014b). On the other hand, RF is an ensemble learning algorithm after Breiman (2001), designed to increase the classification and regression trees (CART) method by integrating a large set of decision trees based on a deterministic technique, by selecting a random set of variables and a random sample from the training data set. SGB is currently gaining popularity in remote sensing applications (Carreiras *et al.,* 2012b; Güneralp *et al.,* 2014; Moisen *et al.,* 2006) and comprises of numerous small regression trees sequentially constructed from "pseudo" residuals (*i.e.* gradient of the loss function of the previous tree). The main advantage of the SGB algorithm is that it does not require the transformation of predictor variables, besides being resistant to outliers, owing to the use of the steepest-gradient algorithm. Moreover, the sequential model-fitting builds on trees fitted previously and increasingly focuses on the most difficult prediction cases, simultaneously providing a stability advantage over regression trees (De'ath, 2007; Elith *et al.,* 2008; Friedman, 2002; Leathwick *et al.,* 2006; Moisen *et al.,* 2006). The RF regression algorithm utilizes bootstrap samples from the training data without pruning, to grow a large number of decision trees (Dye *et al.,* 2011; Ismail and Mutanga, 2010; Prasad *et al.,* 2006). The algorithm has an additional modification of selecting only a random subset of candidate features (*mtry*), to determine the split at each node of a tree. In addition this method employs recursive partitioning, to create multiple regression trees (*ntree*), and then averages the results of all trees (Breiman, 2001). The SGB and RF algorithms were implemented, using the R statistical package (R Development Core Team, 2008).

5.3.1 Experiments

The strength of Landsat-8 OLI and 7 ETM+ data sets in quantifying AGB was examined, based on four tests (Tests I–IV) implemented, using the SGB and RF regression algorithms (Table 5.2). For each test conducted, the set of predictor variables varied, with Test I involving only the extracted

image spectral information, Test II: involving image derived vegetation indices, Test III: involving extracted image spectral information and image derived vegetation indices, and Test IV: involving the most important variables chosen, using the variable selection method. The field measured biomass data set was split into training (2/3) and validation (1/3) for the two algorithms (Riggins *et al.,* 2009). For the SGB algorithm, values for the three important user-defined parameters, that is: (i) the learning rate (*lr*), which determines the contribution of each tree to the final model, (ii) the tree complexity (*tc*), which is the number of independent variables interacting to determine each split, and (iii) the number of regression trees (*nt)* in the ensemble, were set to accommodate the relatively small data set (Elith *et al.* 2008; Hastie *et al.,* 2001; Leathwick *et al.,* 2006). In the present study, the shrinkage rate was set between 0.0001 and 0.1, whereas the tree complexity was set between 1 and 5. Furthermore, the bagging fraction, which determines a fraction of the training data selected randomly for computing each tree, was set at 0.3 and 0.5 during the data analysis. For the RF algorithm, the optimal *ntree* and *mtry* values that could best predict AGB were determined, based on the two parameters (viz. *mtry* and *ntree*). The *mtry* and *ntree* parameters were optimized, using the lowest ten-fold cross-validation prediction error. In this study, the *ntree* values were tested from 500 to 2 500, with a 500 interval, whereas *mtry* was tested from 1 to 14, based on a single interval.

Table 5.2 OLI and ETM+ predictor variables used in forest plantation AGB estimation

Data type	Data source	Details	Experimentation
Image spectral information (ISI)	OLI ETM+	1-6 (blue, green, red, NIR, SWIR I & II) 1-6 (blue, green, red, NIR, SWIR I & II)	I
Vegetation indices (VIs)	OLI ETM+	DVI, GEMI, GNDVI, MSAVI, MSI, NDII, NDVI, NDVIc, OSAVI, RDVI, RSR, SAVI, SAVI2, SR	II
ISI + VIs	OLI ETM+	(1-6) + (DVI, GEMI, GNDVI, MSAVI, MSI, NDII, NDVI, NDVIc, OSAVI, RDVI, RSR, SAVI, SAVI2, SR)	III
More OLI and ETM using the most important selected variables			IV

[a]DVI: Difference Vegetation Index; GEMI: Renormalized Difference Vegetation Index; GNDVI: Green Normalized Difference Vegetation Index; MSAVI2: Modified Soil Adjusted Vegetation Index; MSI: Moisture Stress Index; NDII: Normalized Difference Infrared Index; NDVI: Normalized Difference Vegetation Index; NDVIc: Corrected NDVI; OSAVI: Optimized Soil-Adjusted Vegetation Index; RDVI: Renormalized Difference Vegetation Index; RSR: Reduced Simple Ratio; SR: Simple Ratio; SAVI: Soil Adjusted Vegetation Index; SAVI2: Soil Adjusted Vegetation Index.

5.4 AGB Prediction Error Reporting

For both the SGB and RF algorithms, the explained variance (R^2), root mean square error (RMSE) and RMSE% (measured vs. predicted AGB measurements), for the test data set (see Eq.1), the mean of the absolute error (MAE), and percentage bias (BIAS%) were reported. The BIAS% measures the average tendency of the model simulated biomass estimates to be larger or smaller than the measured biomass values (Yapo *et al.,* 1996). The optimal value of BIAS% is 0.0, with positive values (+ve) illustrating over-estimation and negative values (-ve) demonstrating model under-estimation. Overall, the low values indicate accurate model simulation, and vice-versa. On the other hand, the MAE measures the absolute magnitude of the error and the closer to zero the MAE, the more accurate the simulation (Heiskanen, 2006).

5.5 Results

5.5.1 Measured aboveground biomass descriptive statistics (t ha^{-1})

Table 5.3 shows descriptive statistics of the individual species aboveground biomass measured in the Clan Forests. High aboveground biomass was observed for *Pinus taeda* (298.04 t ha^{-1}) and the least for *Eucalyptus dunii* (96.49 t ha^{-1}). From the statistic summary reported in Table 5.3, it can also be observed that the measured aboveground biomass varies greatly amongst the three species under study (*i.e. Eucalyptus dunii, Eucalyptus grandis* and *Pinus taeda*).

Table 5.3 Measured aboveground biomass descriptive statistics (t ha^{-1})

	Commercial species of monitored		
	Eucalyptus dunii	*Eucalyptus grandis*	*Pinus taeda*
No. of samples	63.00	65.00	53.00
Min.	33.24	106.03	137.11
Max.	96.49	225.07	298.04
Avg.	52.86	170.30	206.07
Std dev.	16.39	29.94	42.83

5.5.1 Comparison of AGB estimates from Landsat-8 OLI and ETM+ extracted Spectral Information

AGB estimation results for ED, EG and PT, in terms of the coefficient of determination (R^2), root mean square error (RMSE), RMSE%, percent bias and mean of the absolute forecast error (MAE), are shown in Tables 5.4, 5.5 and 5.6. The results indicate that the Landsat-8 OLI spectral data set provides better biomass estimates, when compared to the Landsat ETM+ spectral data sets. For example, the results indicate that Landsat-8 OLI spectral information, utilizing the SGB algorithm,

produced an R^2 value of 0.47, the lowest RMSE (16.35 t ha^{-1}; 30.93% of the mean), the lowest percentage bias of -1.40 and the lowest MAE (14.23) for *E. dunii* species. For *E. grandis*, the extracted OLI spectral information again, based on the SGB, yielded the highest R^2 value of 0.37, the lowest RMSE (28.94 t ha^{-1}; 17.02% of the mean), the lowest percentage bias of -0.40 and the lowest MAE (23.03). Similar prediction performance, using the extracted OLI spectral information, was also observed for *P.taeda,* with an R^2 value of 0.44, an RMSE of 28.93 t ha^{-1} (14.04% of the mean), the lowest percentage bias value of -0.40 and the least MAE value of 23.03. On the other hand, extracted Landsat ETM+ spectral information produced slightly weaker estimates in terms of RMSE, Bias% and MAE for all species (see Table 5.4).

Table 5.4 Commercial forest species AGB estimation errors and biases obtained, using Landsat-8 OLI and ETM+ extracted spectral information

			R^2	RMSE t ha^{-1}	Bias%	MAE
E.dunii	L8	SGB	0.47	16.35 (30.93%)	-1.40	14.23
		RF	0.40	18.13 (34.30%)	-4.00	15.62
	L7	SGB	0.43	18.13 (34.30%)	-4.00	15.62
		RF	0.21	23.15 (43.79%)	-5.10	18.62
E.grandis	L8	SGB	0.37	28.94 (17.02%)	-0.40	23.03
		RF	0.32	29.13 (17.13%)	-3.70	23.17
	L7	SGB	0.36	28.62 (16.83%)	-3.70	25.08
		RF	0.29	30.15 (17.73%)	-4.50	29.33
P.taeda	L8	SGB	0.37	28.93 (14.04%)	-0.40	23.03
		RF	0.30	32.83 (15.93%)	1.70	27.70
	L7	SGB	0.34	37.06 (17.98%)	-4.20	28.72
		RF	0.29	35.25 (17.11%)	0.50	30.04

*L8 and L7 denotes spectral information extracted, using Landsat-8 OLI and ETM+ data sets, respectively

5.5.2 A Comparison of AGB Estimates from Landsat-8 OLI and ETM+ derived VIs

The best AGB estimates for Test II were obtained for the three species (ED, EG and PT), using the newly-launched Landsat-8 OLI derived vegetation indices (see Table 5.5). The findings in Table 5.5 show a good performance of the Landsat-8 OLI derived vegetation indices in estimating AGB for: (a) *E.dunii* (R^2 = 0.58; RMSE = 14.03 t ha^{-1} (26.54% of the mean), Bias% = -0.20 and MAE = 11.00), (b) *E. grandis* (R^2 = 0.48; RMSE = 25.23 t ha^{-1} (14.83% of the mean), Bias% = -1.10 and MAE = 19.42), and (c) *P. taeda* (R^2 = 0.43; RMSE = 26.54 t ha^{-1} (12.88% of the mean); Bias% = -1.10 and MAE = 22.92) using the SGB algorithm. Similarly, the Landsat-8 OLI derived vegetation indices produced the best performance, using the RF algorithm, although the estimates were lower than those obtained using the SGB algorithm. Comparatively, results from Landsat ETM+ were

weaker for all species considered, when compared to estimates derived using the Landsat-8 OLI data sets. For instance, Landsat ETM+ derived vegetation indices applying the SGB algorithm produced the RMSE of 18.66 t ha^{-1} (35.30 % of the mean), a bias% of -4.30 and the MAE value of 14.65 for *E. dunii* species, whereas for *E. grandis*, an RMSE of 26.54 t ha^{-1}; 15.60 % of the mean), a percentage bias of -2.26 and the MAE of 22.92 were obtained. A similar trend was observed for *P. taeda,* where the Landsat ETM+ derived estimates were slightly lower than those obtained using Landsat-8 OLI derived vegetation indices.

Table 5.5 Commercial forest species AGB estimation errors and biases generated, using Landsat-8 OLI and ETM+ derived vegetation indices

			R^2	RMSE t ha^{-1}	Bias%	MAE
E.dunii	L8	SGB	0.58	14.03 (26.54%)	-0.20	11.00
		RF	0.53	18.66 (35.30%)	-4.30	14.65
	L7	SGB	0.47	15.56 (29.44%)	-7.10	12.52
		RF	0.33	19.12 (36.17%)	-5.30	14.91
E.grandis	L8	SGB	0.48	25.23 (14.83%)	-1.10	19.42
		RF	0.36	26.54 (15.60%)	-2.26	22.92
	L7	SGB	0.44	36.23 (21.30%)	-2.70	22.82
		RF	0.38	27.67 (16.27%)	-3.10	27.14
P.taeda	L8	SGB	0.43	26.54 (12.88%)	-1.10	22.92
		RF	0.37	29.48 (14.31%)	-0.30	23.43
	L7	SGB	0.37	36.23 (17.58%)	-4.20	28.27
		RF	0.41	31.92 (15.49%)	0.80	26.98

*L8 and L7 denotes vegetation indices computed using Landsat-8 OLI and ETM+ data, respectively

5.5.3 A Comparison of AGB Estimates from Landsat-8 OLI and ETM+ extracted Spectral bands and derived VIs

The results in Table 5.6 show AGB estimates for ED, EG and PT derived from Landsat-8 OLI and ETM+ extracted spectral information and vegetation indices. Firstly, it can be noted that high AGB estimates were obtained from the use of the combined extracted spectral information and vegetation indices, when compared to the use of these variables separately, as shown in Tables 5.4 and 5.5. It can be observed that Landsat-8 OLI also produced high AGB estimates, compared to those obtained using Landsat ETM+. For instance, high R^2, RMSE, Bias% and MAE results can be noted for *E. dunii*, *E. grandis* and *P. taeda* AGB estimation, based on Landsat-8 OLI, than when Landsat ETM+ was used.

Table 5.6 Commercial forest species AGB estimation errors and biases derived, using combined Landsat-8 OLI and 7 extracted spectral information and vegetation indices

			R^2	RMSE t ha^{-1}	Bias%	MAE
E.dunii	L8	SGB	0.66	12.63 (23.89%)	-1.80	9.56
		RF	0.65	15.35 (29.04%)	-7.40	11.42
	L7	SGB	0.56	13.18 (24.93%)	-2.19	9.83
		RF	0.51	18.76 (35.49%)	-4.00	13.63
E.grandis	L8	SGB	0.49	22.13 (12.99%)	-2.40	15.17
		RF	0.52	23.29 (13.68%)	-2.70	18.52
	L7	SGB	0.48	24.38 (14.32%)	-2.50	19.49
		RF	0.43	25.36 (14.89%)	-3.20	19.23
P.taeda	L8	SGB	0.72	25.64 (12.44%)	-1.00	21.77
		RF	0.55	24.51 (11.89%)	-2.30	20.09
	L7	SGB	0.52	28.67 (13.91%)	-2.10	24.33
		RF	0.46	28.41 (13.79%)	-3.65	23.79

*L8 and L7 denotes spectral information and vegetation indices extracted, using Landsat-8 OLI and ETM+ images, respectively

5.5.4 Combined Plantation Forest AGB Estimates from Landsat-8 OLI and ETM+ extracted Spectral Bands and derived VIs

The results in Table 5.7 indicate combined species AGB estimates obtained from Landsat-8 OLI and ETM+ extracted spectral information and vegetation indices. The results show that both Landsat-8 OLI and ETM+ sensors estimated combined plantation forest AGB with weaker accuracy. It can also be observed that although Landsat-8 OLI data set yielded a better performance in estimating forest plantation AGB, when compared to Landsat ETM+, some of its estimates were generally low. Overall, the combined Landsat-8 OLI sensor spectral information and vegetation indices produced better results in terms of R^2, RMSE, Bias% and MAE, when compared to the Landsat ETM+ sensor.

Table 5.7 Combined plantation forest species AGB estimation errors and biases derived, using Landsat-8 OLI and ETM+ extracted spectral information and vegetation indices

			R^2	RMSE t ha^{-1}	Bias%	MAE
Spectral bands	L8	SGB	0.31	66.41 (47.68%)	-6.70	59.27
		RF	0.27	67.15 (48.21%)	-7.10	58.65
	L7	SGB	0.29	66.85 (47.99%)	-6.80	57.38
		RF	0.22	70.81 (50.83%)	-4.40	61.44
VIs	L8	SGB	0.30	63.42(45.53%)	-5.10	50.89
		RF	0.23	68.28 (49.02%)	-6.00	57.48
	L7	SGB	0.21	67.52 (48.47%)	-5.20	57.89
		RF	0.23	69.30 (49.75%)	-7.50	60.21
Spectral bands + VIs	L8	SGB	0.35	62.17 (44.63%)	-4.90	50.11
		RF	0.33	63.61 (45.67%)	-5.40	56.42
	L7	SGB	0.26	65.50 (47.02%)	-5.70	55.76
		RF	0.25	66.14 (47.48%)	-5.80	56.92

*L8 and L7 denotes spectral information and vegetation indices extracted, using Landsat-8 OLI and ETM+ images, respectively

5.5.5 AGB Estimation using Important Landsat-8 OLI and ETM+ Selected Variables

Variable selection was mainly implemented, using the stochastic gradient boosting method, as it produced better overall results, when compared to the random forest algorithm. The results in Figures 5.2a, 5.2b, 5.2c and 5.3a show AGB estimates for *E. dunii*, *E grandis, P. taeda* and for the combined species data set, using the most important Landsat-8 OLI and ETM+ variables, selected using the stochastic gradient boosting method. The most important selected variables included SWIR bands, NDVIc, NDVI, NDII, NIR, SAVI, red band, MSI and green band. The error of estimation reduced significantly for both remote sensing data sets, with Landsat-8 OLI sensor producing the least errors for all three species. For instance, for *P. taeda*, an R^2 of 0.69, a RMSE of 21.65 t ha^{-1} (10.50% of the mean), using Landsat-8 OLI and an R^2 of 0.65, a RMSE of 22.33 t ha^{-1}(10.83% of the mean), using Landsat ETM+ selected variables, were observed. Similarly, for *E. dunii,* an R^2 of 0.71, a RMSE of 10.66 t ha^{-1}; (6.26% of the mean), based on Landsat-8 OLI, and R^2 of 0.68, RMSE of 11.81 t ha^{-1}, (6.93% of the mean), based on Landsat ETM+, were obtained. High accuracies were also observed for *E. grandis* (R^2 = 0.67; RMSE = 17.35 t ha^{-1}; (10.19% of the mean)) using the most important selected variables of Landsat-8 OLI and Landsat ETM+ (R^2 = 0.69; RMSE = 19.83 t ha^{-1}; (11.82% of the mean)).

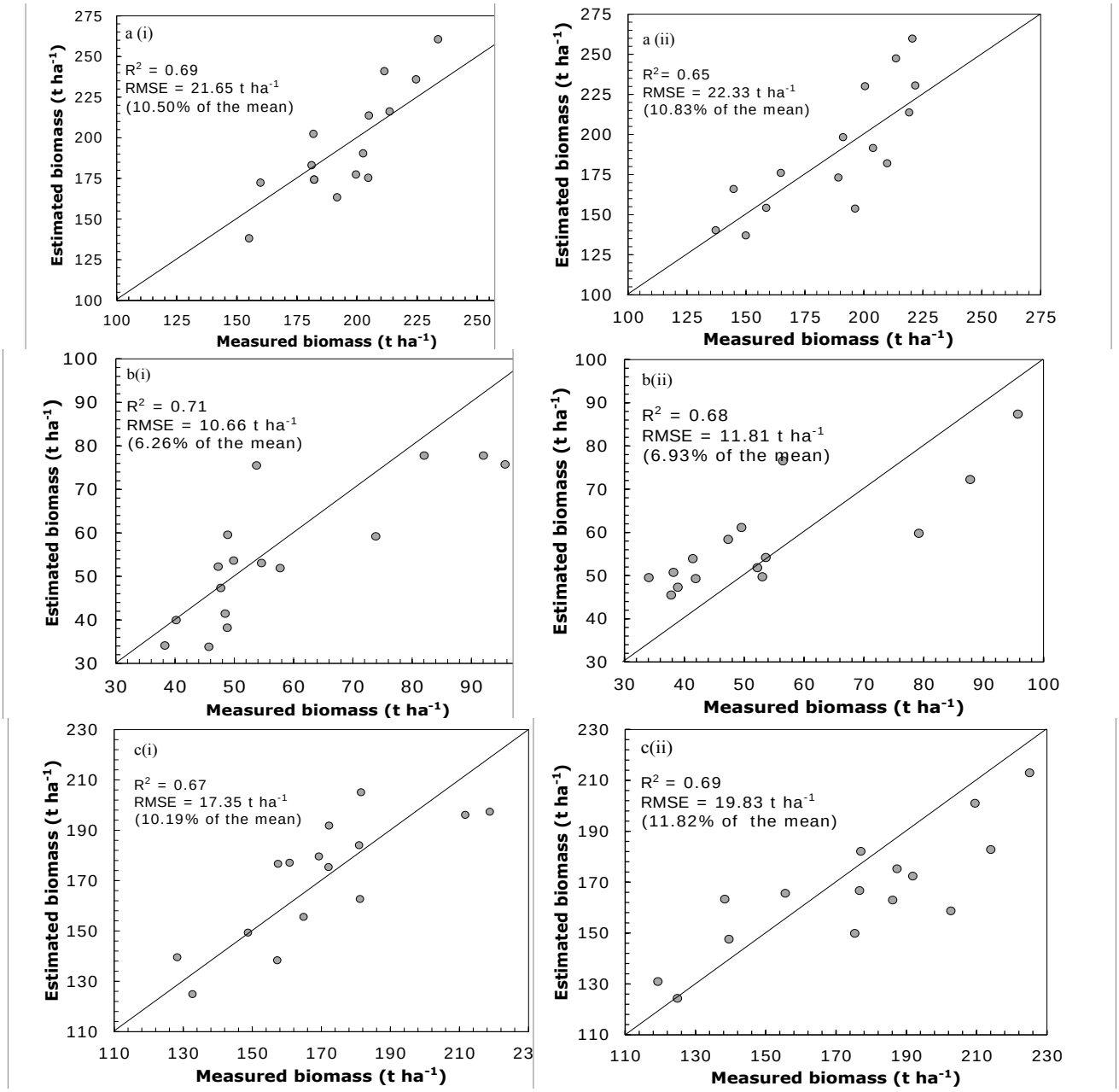


Figure 5.2 One-to-one relationship between measured and estimated species AGB based on Landsat-8 OLI and ETM+ important variables. (a), (b), and (c) represent *P. taeda, E. dunii* and *E. grandis,* based on the most important (i) OLI and (ii) ETM+ predictor variables.

The results in Figure 5.3 show combined species AGB estimates based on the Landsat-8 OLI and Landsat ETM+ important selected variables. A comparative analysis of the AGB estimates for combined species from both sensors shows that the Landsat-8 OLI sensor yielded better, but weaker estimates, producing R^2 of 0.42 and RMSE of 55.32 t ha^{-1} (41.13%), when compared to the Landsat ETM+ sensor, which yielded an R^2 of 0.32 and a RMSE of 64.26 t ha^{-1} (43.86%). Overall,

the findings from the four experiments confirmed the strength of the newly-launched Landsat-8 OLI multispectral sensor in estimating AGB, compared to Landsat ETM+ sensor. Although it was not the main aim of this study to compare the performance of the two regression algorithms (RF and SGB), overall it can be observed that the SGB algorithm outperformed the random forest ensemble across all four modelling stages.

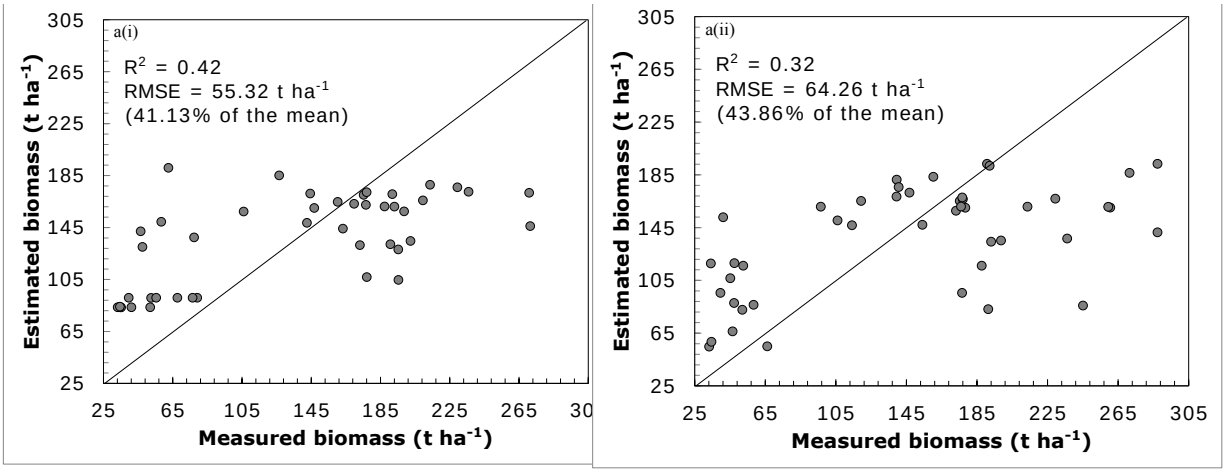


Figure 5.3 One-to-one relationship between measured and estimated combined species AGB based on: (i). Landsat-8 OLI and (ii) Landsat ETM+ selected important variables

5.5.6 Aboveground Biomass Mapping Landsat-8 OLI data

The newly-launched Landsat-8 OLI data set was selected for mapping aboveground biomass within the Clan Sappi Forest, using the SGB algorithm (Figure 5.4). The choice of the Landsat-8 OLI sensor for aboveground biomass was based on the fact that it produced overall plausible aboveground biomass estimates, when compared to the long serving Landsat ETM+ 7, yielding, for instance, high R^2 values, low root mean square errors and relative root mean square errors. The results in Figure 5.4 also show distinct aboveground biomass spatial distribution patterns available within the area under study.

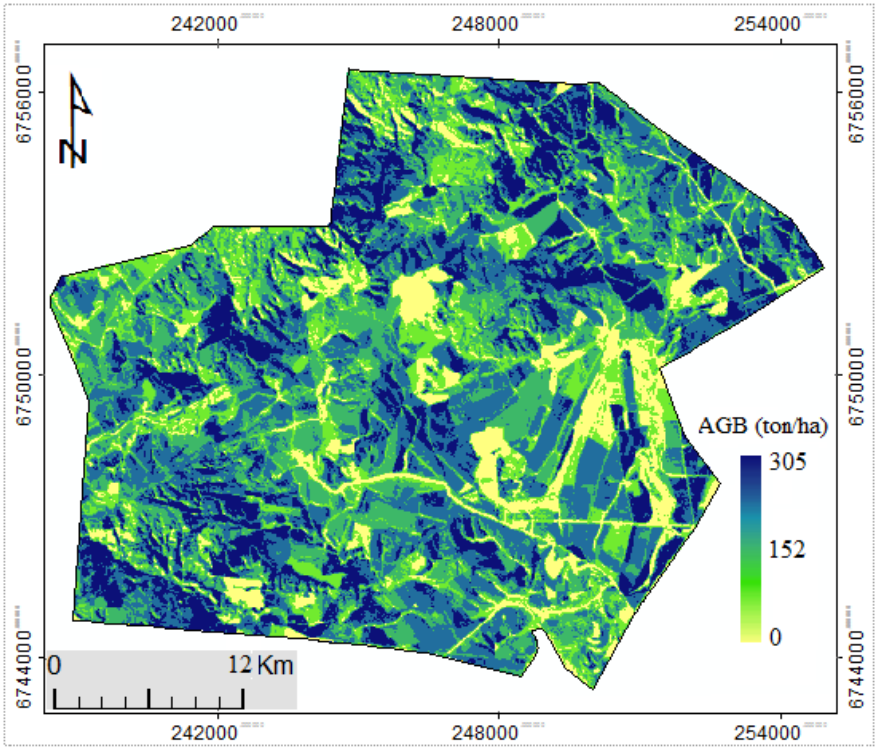


Figure 5.4 Aboveground biomass map of the Clan Sappi Forest in uMgeni Catchment, South Africa, produced using the newly-launched Landsat-8 OLI data set and stochastic gradient boosting algorithm

5.6 Discussion

The main essence of the present study was to examine the potential of the newly-launched Landsat-8 OLI sensor in estimating forest aboveground biomass in the uMgeni Catchment, South Africa. The accurate estimation of aboveground biomass provides a critical input data set required for ecological modelling and accurate carbon quantification. This study therefore aims to investigate whether the unique Landsat-8 OLI sensor design, or data acquisition approach as a push-broom scanner, has the capability of improving the quantification of aboveground biomass over that of the Landsat ETM+ sensor.

Although multispectral sensors provide an attractive alternative for estimating aboveground biomass at regional scale, especially in environments with limited access to high resolution data and the necessary technical expertise, one of their primary challenges is the inability to reduce error of estimation. The present study has demonstrated that the newly–launched Landsat-8 OLI multispectral sensor has the potential to enhance aboveground biomass estimation accuracy. For instance, when the extracted Landsat-8 OLI spectral information was used to estimate aboveground

biomass for the three different plantation forest species (ED, EG and PT), high accuracies in terms of the root mean square error, (RMSE), percentage bias (Bias%) and mean absolute error (MAE) were observed, compared to when the Landsat ETM+ sensor spectral information was used. Similar results were also observed, when a combination of spectral vegetation indices and extracted spectral information were used. The Landsat-8 OLI sensor outperformed the Landsat ETM+ sensor in deriving aboveground biomass, producing a comparatively high R^2, a low RMSE and the relative root mean square error of the mean. Moreover, it was observed that Landsat 8 outperformed Landsat 7 ETM+, when its performance was tested in estimating aboveground biomass for combined species data sets. However, when compared to individual species AGB estimates, both the Landsat-8 OLI and ETM+ sensors produced weaker results for combined species data sets.

The higher accuracy of Landsat-8 in OLI estimating AGB can be attributed to the unique sensor design. For instance, the Landsat-8 OLI makes use of numerous elongated sets of detectors for each waveband that are capable of a detailed scan of the surface along track. The along track design augments the sensitivity of the sensor to most critical vegetation biophysical metrics. The presence of a refined spectral range for particular bands, such as the near-infrared, also plays a fundamental function in boosting forest spectral responses, whereas the enhanced image radiometric resolution (12 bits) permits a precise detection of various vegetation conditions (El-Askary *et al.,* 2014; Pahlevan and Schott, 2013). Above all, the prolonged sensor radiation sampling residence-period for each field-of-view enhances precision during spectral detection and this subsequently minimizes saturation challenges (Irons *et al.,* 2012; Jia *et al.,* 2014; Verbyla, 1995). The observed strength of the newly-launched Landsat-8 OLI multispectral sensor therefore makes it a better alternative for biomass application than the previous Landsat ETM+ products.

Previous studies have examined and documented the utility of the 30-m spatial resolution Landsat ETM+ in estimating aboveground biomass across various scales, with reasonable success (Cohen *et al.,* 2003; Foody *et al.,* 2003a; Hall *et al.,* 2006; Hudak *et al.,* 2002; Kraus and Samimi, 2002; Ren-dong and Ji-yuan, 2002; Zheng *et al.,* 2004). However, all of the Landsat ETM+ acquisitions after May 31, 2003, have since developed an anomaly caused by the failure of the Scan Line Corrector (SLC), which compensated for the forward motion of the spacecraft, so that all the scans were aligned parallel with each other. The malfunction of the SLC mirror assembly has since resulted in approximately 22% data loss of the normal scene area (Chander *et al.,* 2009; Storey *et al.,* 2005). Therefore, the use of SLC-off Landsat ETM+ images have since become a challenge, especially for

large-scale remote sensing applications, due to the cost associated with scan line errors. However, due to non-availability of a sensor with relatively similar characteristics, researchers and practitioners continued to use the data obtained from Landsat ETM+, irrespective of its defects. In addition, the whisky-broom design of Landsat ETM+ has its own challenges in terms of the weak signal-to-noise ratios and low radiometric resolution of 8-bits, making the sensor less sensitive to critical vegetation properties (El-Askary *et al.,* 2014; Pahlevan and Schott, 2013), and the broad near-infrared spectral band permits the water vapour absorption effect (*e.g.* at 0.825 μm), inhibiting the accurate acquisition of surface reflectance and short-sensor dwelling-time in target radiation sampling at each ground-pixel, which is hypothesized to reduce the radiometric resolution of the sensor, thus reducing the spectral record precision and introducing sensor spectral saturation challenges (Irons *et al.,* 2012; Jia *et al.,* 2014; Verbyla, 1995). The newly-launched Landsat-8 OLI multispectral sensor was therefore developed as an alternative, to improve on terrestrial matter characterisation.

Furthermore, it was observed that the Landsat-8 OLI derived aboveground biomass estimation accuracy was higher than Landsat ETM+ estimates before and after the failure of the Scan Line Corrector on the 13th of May 2003, as recorded in the literature (Fazakas *et al.,* 1999; Hall *et al.,* 2006; Kajisa *et al.,* 2009; Luther *et al.,* 2006; Zheng *et al.,* 2004). For example, Kajisa *et al.* (2009) estimated forest biomass from Landsat ETM+ with an R^2 value of 0.67 and an Akaike's information criterion (AIC) value of 484.94, using the pixel-based approach in Kampong Thom Province, Cambodia. On the other hand, Hall *et al.* (2006), using Landsat ETM+ sensor, estimated aboveground biomass with a RMSE of 33.7 t ha^{-1} in west-central Alberta, whereas Luther *et al.* (2006) produced an RMSE value of 63.6 t ha^{-1} in the western Newfoundland ecoregion. Cortés *et al.* (2014) using Landsat ETM+ estimated *Eucalyptus* species aboveground biomass with a root mean square error of 25.81 t.ha^{-1}. The estimation accuracy from the above results was much lower, when compared with the results of the present study, using Landsat-8 OLI sensor. This can be attributed to the push-broom design of Landsat-8 OLI sensor, which is characterized by improved signal-to-noise ratios and high radiometric resolution of 12 bits, which makes the sensor more sensitive to critical vegetation properties (El-Askary *et al.,* 2014; Pahlevan and Schott, 2013). Moreover, the narrow or refined near-infrared spectral band inhibits water vapour absorption effect at 0.825 μm, as well as other sensor related properties, therefore allowing accurate surface spectral acquisition, which is critical in reducing sensor spectral saturation challenges.

Although Landsat ETM+ results of this study are weaker, when compared to those obtained using Landsat-8 OLI imagery, they are comparable to the results obtained in previous aboveground biomass estimation studies, using similar data sets (Cortés *et al.,* 2014). What is therefore unique in this study is that the push-broom multispectral Landsat-8 OLI imagery produced overall plausible aboveground estimates, compared to the long serving Landsat ETM+. These results are also in line with the assumptions that the push-broom Landsat-8 OLI sensor, with improved 12-bit quantization, is bound to have an impact on vegetation studies, when compared to the long-serving Landsat ETM+ (Güneralp *et al.,* 2014). In this study, based on four experimental levels, the push-broom multispectral Landsat-8 OLI sensor was established as the most attractive multispectral sensor. It can produce plausible aboveground estimates, compared to the long-serving and malfunctioning Landsat ETM+ sensor. However, despite Landsat-8 OLI performing well, the results in the study seem to be moderately low, when compared to hyperspectral- or lidar-derived biomass estimates.

For resource-scarce environments, Landsat-8 OLI seem to be posing a lucrative opportunity for regional scale AGB estimation, especially when compared to high resolution sensors, such as hyperspectral, lidar and sometimes radar sensors. This is despite these sensors having robust means of vegetation data collection and the subsequent characterization of the vertically distributed attributes. The application of the above-mentioned remote sensing sensors is associated with numerous limitations, such as availability, cost, contiguity, spectral processing and analysis complications, particularly in sub-Saharan Africa, given the financial challenges, among other factors (Gara *et al.,* 2014b; Mutanga *et al.,* 2012). For instance, the application of a hyperspectral data set for AGB estimation is a major challenge due to the huge data volume and the Hughes phenomenon. This challenge often introduces multicollinearity, as a result of similarities in the biophysical spectral reflectance properties. Similarly, the use of the radar data set in sub-Saharan Africa remains a challenge due to: (i) saturation problems, and (ii) the limited technical expertise required for image data handling and pre-processing, despite these images being provided at low or no cost.

Although the objective of this study was not to compare algorithms, the study has shown that stochastic gradient boosting outperformed the random forest algorithm in all aboveground biomass estimation stages, namely, (i) spectral bands, (ii) vegetation indices, (iii) combined data sets (*e.g.* spectral information and vegetation indices), and (iv) with most important selected variables. The

robustness of the two algorithms (*i.e.* RF and SGB) in estimating AGB is also confirmed by previous studies (Carreiras *et al.,* 2012b; Dube *et al.,* 2015; Güneralp *et al.,* 2014). For instance, the study by Güneralp *et al.* (2014) showed that the SGB significantly outperform other modelling techniques, such as Cubist, producing high R^2 and low RMSE values. The authors concluded that the algorithm enabled more accurate AGB estimation in floodplains characterized by uneven-aged forests. Moreover, a study by Dube *et al.* (2015) showed that both the RF and SGB regression models provide a potential avenue for improving operational AGB estimates across different scales. For example, the findings by Dube *et al.* (2015), using the 5-m spatial resolution RapidEye sensor, showed the SGB algorithm as having the potential to outperform the most popular RF ensemble, producing an R^2 value of 0.80 and a RMSE of 16.93 $t \cdot ha^{-1}$ for *Eucalyptus grandis*. The strength of stochastic gradient boosting over the random forest is explained by the model's ability to handle outliers, inaccurate training data, as well as missing and unbalanced data sets (De'ath, 2007; Lawrence *et al.,* 2004; Leathwick *et al.,* 2006). In addition, the stochastic gradient boosting regression tree is proven to have the capability to handle, identify and select critical variables from large amounts of data, making it more robust in improving model accuracies (Carreiras *et al.,* 2012b; De'ath, 2007; Elith *et al.,* 2008; Gara *et al.,* 2014b; Lawrence *et al.,* 2004). Moreover, the fact that the SGB algorithm uses a stage-wise additive model fitting procedure, enhances the predictive performance of weak learning algorithms. Literature shows that the random forest algorithm normally requires a larger number of observations, to improve on the randomization concept, especially when the sampled data set is varying (Adelabu *et al.,* 2014; Özçift, 2011). In addition, Krahwinkler and Rossman (2011) have stated that the random forest algorithm uses a combination of soft linear boundaries at the decision surface thus, it may not work with small sample sizes.

5.7 Conclusion

This study examined the utility of the newly-launched push-broom Landsat-8 OLI multispectral sensor for estimating the AGB of forests in uMgeni Catchment, KwaZulu-Natal, South Africa, as part of large-scale monitoring, to understand forest contribution to the regional carbon cycle.

The results have shown that:

1. the newly-launched Landsat-8 OLI multispectral sensor with a wide swath width of 185 km, coupled with improved signal-to-noise ratios, offers an invaluable primary data-source required for accurate aboveground biomass estimation, especially in data-scarce

environments, when compared to the malfunctioning whisky-broom Landsat ETM+ sensor, and

2. stochastic gradient boosting regression tree is a more powerful, robust and dependable technique that can estimate aboveground biomass, irrespective of the species, when compared to the competing random forest ensemble.

Overall, the results of this work provide the necessary insight and motivation to the remote sensing community, ecologists and environmentalists for shifting towards identifying the most suitable, cheap and readily-available remote sensing sensors necessary for reliable and accurate aboveground forest biomass monitoring, especially in data-scarce environments. However, given the relatively poor performance of the Landsat-8 OLI spectral and vegetation indices, it would be more worthwhile to devise further methods that can help to improve the estimates based on this sensor, principally in resource-constrained areas, given the cost and the sensor's spatial fidelity.

6. ENHANCING ABOVEGROUND BIOMASS ESTIMATION ACCURACY USING LANDSAT-8 OLI SENSOR TEXTURE METRICS

This Chapter is based on:

Dube, T, Mutanga, O. 2015. Investigating the robustness of the newly launched Landsat-8 pushbroom sensor derived texture indices in estimating medium-density plantation forest species aboveground biomass in KwaZulu-Natal South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing,* 108: 12–32.

Abstract

The successful launch of the 30-m Landsat-8 OLI push-broom sensor offers a new primary data source necessary for AGB estimation, especially in resource-limited environments. In this work, the strength and performance of Landsat-8 OLI image derived texture metrics (*i.e.* texture measures and texture ratios) in estimating plantation forest species AGB was investigated. It was hypothesized that the sensor's push-broom design, coupled with the presence of refined spectral properties, enhanced radiometric resolution (*i.e.* from 8 bits to 12 bits) and improved signal-to-noise ratio, have the potential to provide the detailed spectral information necessary for significantly strengthening AGB estimation in medium-density forest canopies. The relationship between image texture metrics and measurements of forest attributes can be used to help characterize complex forests and enhance fine vegetation biophysical properties, a difficult challenge when using spectral vegetation indices, especially in closed canopies. This study examines the prospects of using Landsat-8 OLI sensor-derived texture metrics for estimating AGB for three medium-density plantation forest species in KwaZulu-Natal, South Africa. In order to achieve this objective, three unique data pre-processing techniques were tested (Analysis I: Landsat-8 OLI raw spectral-bands vs. raw texture bands; Analysis II: Landsat-8 OLI raw spectral-band ratios vs. texture band ratios; and Analysis III: Landsat-8 OLI derived vegetation indices vs. texture band ratios). The Landsat-8 OLI derived texture parameters were examined for robustness in estimating AGB, using linear regression, stepwise-multiple linear regression and stochastic gradient boosting regression models. The results of this study demonstrated that all texture parameters, particularly band texture ratios calculated using a 3 x 3 window size, could enhance AGB estimation, when compared to simple spectral reflectance, simple band ratios and the most popular spectral vegetation indices. For instance, the use of combined texture ratios yielded the highest R^2 values of 0.76 (RMSE = 9.55 t ha^{-1} (18.07%) and CV-RMSE of 0.18); 0.74 (RMSE = 12.81 t ha^{-1} (17.72%) and CV-RMSE of 0.08); 0.74 (RMSE = 12.67 t ha^{-1} (06.15%) and CV-RMSE of 0.06) and 0.53 (RMSE = 20.15 t ha^{-1} (14.40%) and CV-RMSE of 0.15) overall for *Eucalyptus dunii*, *Eucalyptus grandis*, *Pinus taeda* individually, and all species, respectively. Overall, the findings of this study provide the necessary insight and motivation to the remote sensing community, particularly in resource-constrained regions, to shift towards embracing various texture metrics obtained from the readily-available and cheap multispectral Landsat-8 OLI sensor.

Keywords: estimation accuracy; Landsat-8 OLI texture metrics; push-broom sensor; signal-to-noise ratio; regression ensemble; swath-width; band texture ratios.

6.1 Introduction

Aboveground biomass is an important indicator of bio-physical processes related to forest dynamics. Information on AGB is valuable in understanding and monitoring ecosystem response and its contribution to the global carbon cycle and climate change for accurate greenhouse gas inventorying, as well as terrestrial carbon accounting (Chinembiri *et al.,* 2013b; Gara *et al.,* 2014a; Güneralp *et al.,* 2014; Lu, 2006). Accurate and regularly-repeated AGB estimation at regional or local scales is critical in reducing the uncertainty in estimating carbon sequestration and emissions (Güneralp *et al.,* 2014; Lu, 2006). Information on AGB is important in greenhouse gas lifecycle assessments and global climate change mitigation strategies. Moreover, the accurate and repeated monitoring of the status of forest ecosystems can also help in introducing appropriate planning and monitoring conservation efforts. Therefore, identifying affordable, timely and readily-available remote sensing data sets, together with robust image processing techniques, is essential to improve forest AGB estimates.

Although the application of remote sensing in forest AGB estimation remains one of the most favourable and invaluable potential approaches, the issue of image resolution (*i.e.* spectral and spatial properties) plays a major role in the accurate retrieval of AGB estimates. Spectral vegetation indices normally rely on the relationship between the red and near infra-red bands of the electromagnetic spectrum, to augment the spectral influence from the green vegetation, concurrently reducing contribution from the soil background, sun zenith angle, sensor viewing angle, senesced vegetation and the atmosphere (Bannari *et al.,* 1995; Carlson and Ripley, 1997; Choudhury, 1987; Elvidge and Chen, 1995; Foody *et al.,* 2003b; Huete, 1988b; Lu, 2006; Tarpley *et al.,* 1984; Teillet *et al.,* 1997; Tucker, 1979; Wiegand *et al.,* 1991). However, previous studies show that the use of spectral vegetation indices computed from medium-to-coarse spatial resolution multispectral sensors, in places with moderate or high canopy closure, produce poor results, primarily due to saturation challenges and the existing problem of multiple layering (Gara *et al.,* 2014b; Ingram *et al.,* 2005; Lu, 2006; Mutanga *et al.,* 2012; Mutanga and Skidmore, 2004b; Nichol and Sarker, 2011). Recent studies on AGB estimation advocate the use of texture parameters, instead of spectral vegetation indices. This is because texture parameters correlate much better with field data sets, since they allow for a finer distinction of vegetation structural details (Eckert, 2012a; Fuchs *et al.,* 2009; Nichol and Sarker, 2011; Sarker and Nichol, 2011).

The image-texture analysis technique can be applied in identifying spectrally unique objects, based on a function of local variance in the image, and it is also related to the spatial resolution and the size of the dominant scene objects (Haralick *et al.,* 1973). Image-texture is an important source of information, capable of identifying different aspects of forest stand structure, including density, age and leaf area index in medium-to-high spatial resolution images (Champion *et al.,* 2008). Some of the major strengths of applying image-texture measures include: (i) the ability to simplify and define complex forest canopy structures, such as multiple layering and even closed canopies; and (ii) enhancing the detection of fine vegetation biophysical properties, a complex challenge when using spectral vegetation indices, especially in closed canopies (Champion *et al.,* 2008). Moreover, earlier work (*i.e.* using high spatial resolution sensors) demonstrates that texture measures have the capability of enhancing vegetation discrimination (Dekker, 2003; Podest and Saatchi, 2002; Shimada *et al.,* 2014). However, the relationship between medium spatial resolution sensors and forest AGB has not yet been fully examined, especially when compared to the use of other image properties, such as raw spectral band information and vegetation indices. This is perhaps due to the difficulty of identifying and selecting proper texture parameters, together with the optimal window-size and offset.

Thus far, the available studies have mostly demonstrated the application of texture measures derived from high resolution sensors, such as synthetic aperture radar, Worldview-2, Advanced Visible and Near Infrared Radiometer type-2 (AVNIR-2) and SPOT-5, in estimating various tree-structural attributes, namely, stand age, leaf area index, stand density and AGB (Eckert, 2012a; Nichol and Sarker, 2011; Pandey *et al.,* 2010; Pinto *et al.,* 2012; Sarker and Nichol, 2011), and partly from medium resolution multispectral sensors, such as the Landsat products (Cutler *et al.,* 2012). The majority of these studies have shown that image-texture measures have the potential to accurately improve the detection of different forest stand characteristics. More importantly, image-texture measures have the capability of enhancing the discrimination of spatial information and simultaneously enhancing AGB detection levels, by increasing the saturation levels that could not be measured with spectral vegetation indices (Eckert, 2012; Kuplich *et al.,* 2005; Santos *et al.,* 2003; Sarker and Nichol, 2011; Vashum and Jayakumar, 2012; Xu *et al.,* 2011). For instance, Sarker and Nichol (2011) estimated AGB with a plausible adjusted r-square value of 0.88, using texture measures derived from the Advanced Land Observation Satellite, AVNIR-2. On the other hand, Eckert (2012) obtained high carbon estimates (R^2-value 0.84 and relative RMSE of 6.8%) for degraded forest, using WorldView-2 derived texture measures.

Thus, from the afore-mentioned studies, it is clear that image textural measures have the potential to provide an attractive opportunity for monitoring tree-structural attributes (*i.e.* AGB, leaf area index and chlorophyll content *etc*.) in areas with a moderate or high canopy closure. However, the problem with the use of high resolution images, in resource-constrained regions of sub-Saharan Africa, south-east Asia and South America, is the associated cost, limited availability and related technical challenges. Currently, the limited number of AGB studies in these regions, using high resolution images, for instance, is probably an indication of the limitations associated with these data sets (Dube *et al.,* 2014; Dube *et al.,* 2015; Gara *et al.,* 2014b; Koch, 2010). Nevertheless, considering the inevitability of forest AGB estimation at regional to global scales, the prospects of investigating the performance of image texture parameters, obtained from the multispectral sensors with medium-spatial resolution, a large swath-width and a repeated coverage, is necessary in resource-constrained regions.

The adoption of the free and readily-available remotely sensed data sets is critical in such resource-constrained regions. The newly-launched push-broom Landsat-8 OLI sensor, with 30-m spatial resolution, is one such data set. Thus far, the rich information contained in this sensor has not yet been fully exploited in order to understand the distribution of AGB. This is mainly due to the fact that this sensor has just recently been launched, possibly due to the purported challenges of saturation and the presence of a few spectral bands, which can be used to compute simple spectral vegetation indices. Among the different types of the readily-available multispectral remote sensing sensors, a derivation of texture measures from this remotely sensed data set (*i.e.* push-broom Landsat-8 OLI sensor) can improve the estimation of regional to local scale forest AGB. The newly-launched Landsat-8 OLI sensor is hypothesized to display great potential in estimating AGB across different scales, especially in data-scarce areas. This is due to the enhanced sensor's sensitivity to different vegetation properties (*e.g.* chlorophyll, leaf area index and AGB), the presence of a large swath-width (*i.e.* 185-km), coupled with improved signal-to-noise ratios. These sensor improvements enhance the radiometric sensitivity, which, in turn, improves spectral strength to detect the most important forest structural properties, thereby minimizing saturation problems. These saturation problems were normally common with the prior Landsat 7 ETM^{+} products. Although the new Landsat-8 OLI sensor presents a more attractive potential in vegetation studies than its counterparts (*i.e.* MODIS, earlier Landsat products etc.), previous work, using spectral vegetation indices computed from the Landsat-8 OLI, obtained relatively low AGB estimation accuracies (*i.e.* an R^2 of 0.69 between predicted and observed biomass and a moderately high

average RMSE of 14.91 t ha^{-1}). This shows that accurate forest AGB estimation is not only dependent on the Landsat-8 OLI data set, but also requires advanced and robust image processing techniques, such as texture metrics. Therefore, texture metrics are perceived to improve the estimation accuracy of AGB, when compared to the use of vegetation indices (Nichol and Sarker, 2011; Sarker and Nichol, 2011). The main aim of this paper was therefore to investigate the performance of texture metrics derived from the Landsat-8 OLI in estimating AGB in the Clan Sappi Forests in the KwaZulu-Natal Province, South Africa.

6.2 Materials and Methods

6.2.1 Study Area

The study was conducted at the Sappi Clan plantation forest (a paper and pulp company) in the midlands region of the KwaZulu-Natal Province, South Africa (Figure 6.1). The Clan Sappi Forests area is situated between Latitude 29°24'46.74"S, 29°17'45.94"S and Longitude 30°18'32.89"E, 30°28'28.21"E. The area is characterised by moderately steep and undulating topography, ranging from ~644 to 1266 m above sea-level, with predominantly deep-fertile lithic soils, such as the leptosols, cambisols, acrisols and lixisols (Sappi, 1993; Scott and Lesch, 1997). The area experiences sub-tropical climatic conditions, with the rainy season occurring during the summer months (*i.e.* October – February), with an annual rainfall ranging from 730 to 1500 mm and an average annual temperature of 21.7 C. Different species of eucalyptus and pine are present within the area and these trees are mainly grown for pulpwood production.

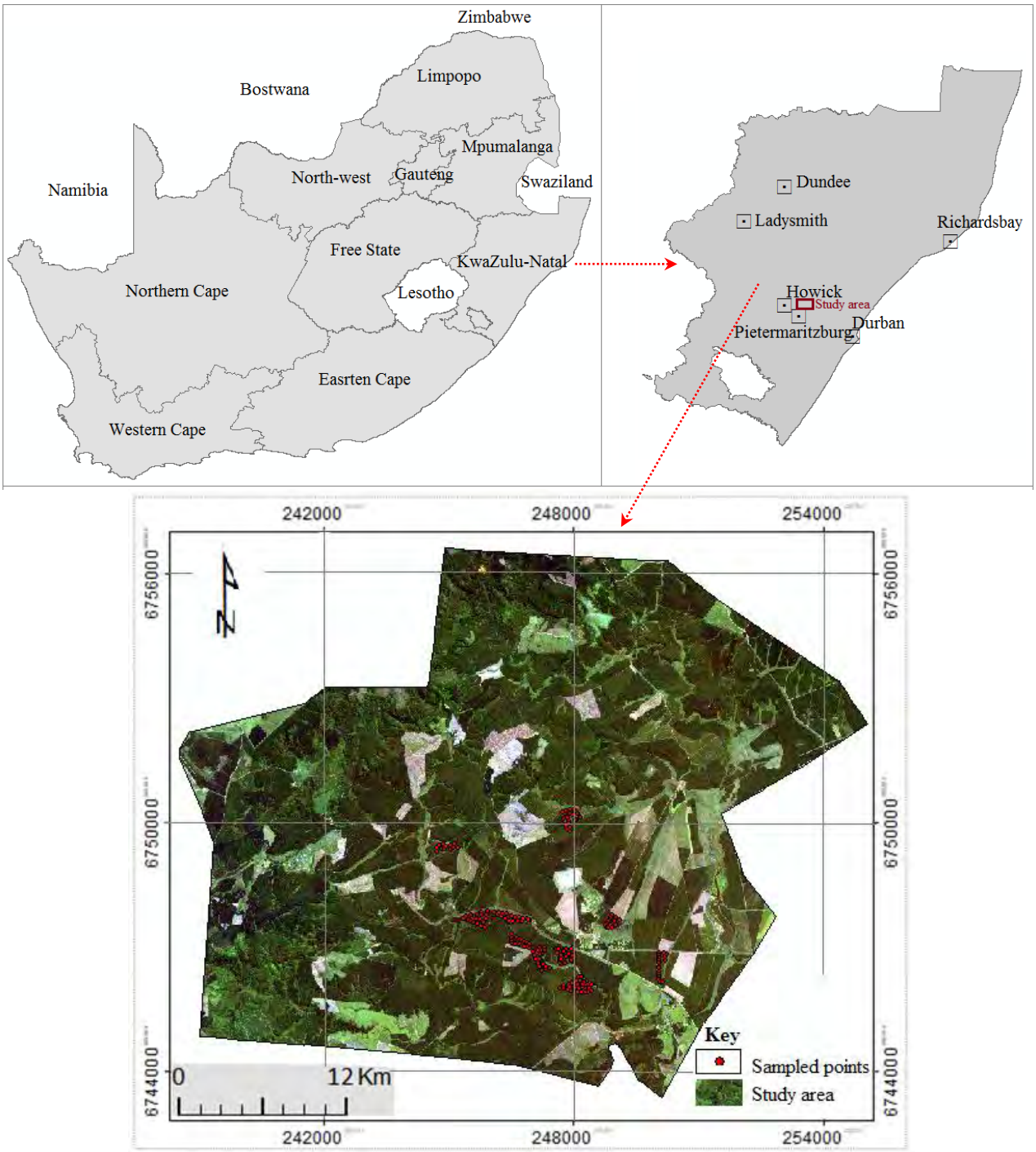


Figure 6.1 Study area location with the distribution and position of sampled plots

6.2.2 Field-data Inventorying

Field data collection was conducted from the 30th of July to 22nd August, 2013, coinciding with the Sappi annual routine-field surveys. A grid-based systematic sampling technique, with a 10 m-radius circular sample plots, distributed systematically (100 m), within the stand, was applied during the field data collection period (Wessels and Kassier, 1985). Diameter-at-breast height (DBH) and tree

height (Ht) were measured from the three dominant Clan Sappi Forests, namely, *Eucalyptus dunii* (ED), *Eucalyptus grandis* (EG) and *Pinus taeda* (PT). During data collection, trees with DBH ≥ 5 cm were considered. The vector maps of the study site, by courtesy of Sappi Forest, together with the aid of an expert forest specialist, were used to navigate and identify the selected plots of interest within the area. During the field campaign, plot centre GPS locations (n = 181 plots) were also recorded with sub-meter accuracy, using the Trimble GeoXH 6000 series handheld Global Position System (GPS). The measurement of DBH and Ht were taken, using the Haglof Digitech Calliper and Vertex IV laser instrument, respectively.

To derive *in situ* AGB for individual 181 sampled plots, two unique methods were chosen. AGB was calculated, using volume with species specific BEF, as well as the general functional group equations for the three plantation species. The use of different techniques in computing AGB in this study was driven by the prevailing unique structural and taxonomical properties of *Eucalyptus dunii, Eucalyptus grandis* and *Pinus taeda* (Atta-Boateng and William, 1998; Henry *et al.,* 2011; Lu, 2005a). Specifically, the general functional group equation was developed in an area with rainfall of approximately. 800 to 1500 mm and a temperature range between 21-34^{o}C, which is comparable to climate conditions of the study area. Furthermore, this equation was developed, based on a DBH range of 0.6 to 56 cm, a characteristic almost identical to the midlands region of KwaZulu-Natal, South Africa. Specifically, AGB for *Eucalyptus dunii* and *Eucalyptus grandis* species were derived using volume (m^3/ha), together with species-specific BEF available in literature (Dovey, 2009). Tree volume was computed based on mathematical function described in Bredenkamp (2000). For *Pinus taeda*, individual tree DBH measurements were converted to AGB via a general functional group equation and subsequently summed across the entire plot. Table 6.1 indicates descriptive statistics for individual species AGB, with *Pinus taeda* having high biomass (298.04 t ha^{-1}), when compared to the other species.

Table 6. 1 Measured aboveground biomass descriptive statistics (t ha^{-1})

Species type	# of samples	Min.	Max.	Avg.	Std dev.
E. dunii	63.00	33.24	96.49	52.86	16.39
E. grandis	65.00	106.03	225.07	170.30	29.94
P. taeda	53.00	137.11	298.04	206.07	42.83

6.2.3 Landsat-8 OLI Image Acquisition and Pre-processing

The Landsat-8 OLI sensor was launched on the 11th of February 2013 by the National Aeronautics and Space Administration and the United States Geological Survey (NASA-USGS). Officially, the Landsat-8 OLI sensor began normal operations on the 30th of May 2013 and presents a number of key improvements in design and spectral configuration. The novel and attractive spectral properties associated with the medium-resolution multispectral Landsat-8 OLI sensor include a refined spectral range for certain bands (*i.e.* critical for improving the vegetation spectral responses across the near-infrared) and enhanced radiometric resolution (*e.g.* from 8 bits to 12 bits, which significantly allows the characterization of different vegetation structural characteristics) (El-Askary *et al.,* 2014; Pahlevan and Schott, 2013). Moreover, the changes in the Landsat-8 OLI sensor design have also seen great advances in SNR, almost twice as good as Landsat 7 Enhanced Thematic Mapper plus: ETM+ (Irons *et al.,* 2012). It is because of these appealing unique sensor features that the utility of texture metrics, derived from the Landsat-8 OLI image, was tested in estimating plantation forest species AGB.

The Landsat-8 OLI image used in the study was acquired on the 3rd of August 2013, using the push-broom or "along track" Operational Land Imager sensor. The Clan Sappi Forests area is covered by one Landsat-8 OLI tile with path/row 168/80 and is obtained from the USGS Earth Resources Observation and Science (EROS) Centre archive (http://earthexplorer.usgs/). The image was acquired during a clear and sunny sky condition, with a sun azimuth angle of 39.78^{o} and a sun elevation angle of 33.14^{o}. The image spatial resolution was about 30-m with seven spectral bands ranging from 0.43 to 2.29 μm (VNIR-SWIR: Visible Near-Infrared and Short Wave Infrared) and bandwidths ranging between 0.3 to 18 μm. The seven image wavebands correspond to band one– coastal/aerosol (0.43 – 0.45 μm), band two–blue (0.45 – 0.51 μm), band three – green (0.53 – 0.59 μm), band four – red (0.64 – 0.67 μm), band five – near IR (0.85 – 0.88 μm), band six – SWIR (1.57 – 1.65 μm) and band seven – SWIR (2.11 – 2.29 μm). The Landsat-8 OLI sensor has a temporal resolution of 16 days.

The Landsat-8 OLI image bands were converted from digital number format (DN) to reflectance, using ENVI 5.1 software, following the methodology summarized on the USGS website (http://landsat.usgs.gov). Subsequently, the acquired Landsat-8 OLI image was atmospherically corrected, using the MODTRAN, based on the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH) radiative transfer algorithm (Matthew *et al.,* 2000; Perkins *et al.,* 2005)

which is embedded in the ENVI software. The Landsat-8 OLI image was then geometrically corrected (GCS Hartebeesthoek 1994), using 23 ground control points (GCPs) of major features (*e.g.* roads and rock outcrops). The first order polynomial function was used and a nearest-neighbour resampling protocol was applied. The total transformation root mean square error (RMSE) of less than a pixel was attained.

6.2.4 Landsat-8 OLI-derived Variables

Four simple spectral reflectances (*i.e.* blue, green, red and NIR), as well as six simple spectral band ratios (*i.e.* blue/green, blue/red, blue/NIR, green/red green/NIR and red/NIR), that were mostly applied in literature (Nichol and Sarker, 2011), were derived from the Landsat-8 OLI imagery. Five vegetation indices were also computed (Table 6.2), based on the previous findings, which highlighted their successful application as predictors for AGB estimation (Güneralp *et al.,* 2014). The texture parameters used here have been explored in Eckert (2012a) and Cutler *et al.* (2012). The spectral indices and texture information applied in this study were extracted at each location and the pixel containing a sample plot was identified from the Landsat-8 OLI imagery for analyses.

Table 6. 2 Selected spectral vegetation indices and image texture measures derived from Landsat-8 OLI applied in the estimation of aboveground biomass

Parameters	References or Equation
Single band reflectance	
band 2-blue, band 3-green, band 4-red, band 5-near-infrared	
Computed vegetation indices	
1. Moisture Stress Index (MSI) SWIR/NIR	Rock et al. (1986a)
2. Normalized Difference Vegetation Index (NDVI) (NIR − Red)/(NIR + Red)	Rouse et al. (1974); Jordan (1969b)
3. Corrected Normalized Difference Vegetation Index ($NDVI_c$) $\frac{NIR-\mathrm{Re}d}{NIR+\mathrm{Re}d}\left(1-\frac{SWIR-SWIR_{\min}}{SWIR_{\max}-SWIR_{\min}}\right)$	Nemani et al. (1993)
4. Normalized difference infrared index (NDII) (NIR – SWIR 1)/(NIR + SWIR 1)	(Hardisky et al. (1983); Kimes et al. (1981))
5. Soil adjusted vegetation index (SAVI) $\frac{(1+L)(NIR+\mathrm{Re}d)}{NIR+\mathrm{Re}d+L}$; $L=0.5$	Huete (1988b)
Grey level co-occurrence matrix texture measures	
1. Angular Second moment (ASM)	$\sum_{i,j=0}^{N-1} iP_{i,j}^2$

2. Contrast (CO)	$\sum_{i,j=0}^{N-1} iP_{i,j}(i-j)^2$
3. Correlation (CR)	$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(i-\mu_j)}{(\sigma_i^2)(\sigma_j^2)}\right]$
4. Data Range (DR)	$\max\{X\}-\min\{X\}$; $Where\ X=x_1,x_2,...,x_n$
5. Dissimilarity (DI)	$\sum_{i;j=0}^{N-1} iP_{i,j}\lvert i-j\rvert$
6. Entropy (EN)	$\sum_{i,j=0}^{N-1} iP_{i,j}(-\ln P_{i,j})$
7. Homogeneity (HO)	$\sum_{i,j=0}^{N-1} i\frac{P_{i,j}}{1+(i-j)^2}$
8. Mean (ME)	$\sum_{i,j=0}^{N-1} iP_{i;j}$
9. Standard deviation (Stdev)	$\sqrt{VA}$; $where\ VA=\sum_{i,j=0}^{N-1} iP_{ij}(i-ME)^2$
10 Variance (σ^2)	$\frac{\sum_{i,j}(x_{i,j}-\mu)^2}{n-1}$

*For equation 3 and 4 the SWIR represents the SWIR I band of Landsat 8 OLI.

6.2.5 Landsat-8 OLI Sensor Texture Metrics derivation

For this particular study, the Landsat-8 OLI image texture metrics were statistically derived, using popular Gray Level Co-occurrence Matrix (GLCM) texture algorithms calculated from a relative displacement vector (*d, θ*), which explains the spatial distribution of the level pairs separated by *d* in direction *θ* (Haralick *et al.,* 1973). Although numerous image-texture metrics can be generated from the most popular Gray Level Co-occurrence Matrix, in this study, ten were selected including mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, correlation, standard deviation and data range (Baeza-Yates and Ribeiro-Neto, 1999; Blaschke *et al.,* 2014; Haralick, 1979; Haralick *et al.,* 1973; Sarker and Nichol, 2011) (Table 6.2). The choice of these texture metrics was prompted by their strength and successful application displayed in previous AGB estimation studies conducted in dense tropical forests, using high resolution image data sets (Bastin *et al.,* 2014; Cutler *et al.,* 2012; Eckert, 2012a; Sarker and Nichol, 2011; Sarker *et al.,* 2013).

Although the selected texture metrics have demonstrated successful application in most AGB studies, literature indicates that the identification and selection of the most appropriate texture

measures require an additional selection of the most suitable moving window sizes (Bastin *et al.,* 2014; Cutler *et al.,* 2012; Eckert, 2012a; Sarker and Nichol, 2011; Sarker *et al.,* 2013). Thus, in order to define the optimal window size required for accurate AGB estimation, the selected texture measures were computed from four Landsat-8 OLI spectral bands (*i.e.* band two–blue, band three – green, band four – red and band five – NIR), using four window sizes (*i.e.* 3 × 3, 5 × 5, 7 × 7 and 9 × 9) based on four offsets ([1,0], [1,1], [0,1] and [1,-1]). Literature indicates that texture measures derived from a small window size are sensitive to fine scale variations in pixel brightness, when compared to those derived using large window sizes. Thus, to determine a suitable window size for accurate AGB retrieval, a window with a strong relationship between estimated and measured AGB was considered, as indicated in literature. Furthermore, to evaluate the strength of image texture measures (*i.e.* best performing texture measures from the texture metrics on four window sizes on four bands with a combination of different offsets) and band texture ratios, the results obtained were compared with those derived, using selected spectral vegetation indices (Table 6.2).

6.2.6 Statistical Analysis

For this study, AGB was estimated using linear regression, stepwise multiple regression and SGB predictive methods. The choice and application of the SGB method in this study was based mainly on the algorithm's technical abilities over traditional parametric regression approaches, which include: (i) the non-parametric nature of the regressor, (ii) the potential to screen the most important predictors (*i.e.* capability to identify and select suitable predictor variables from the entire data set), (iii) the ability to handle non-linear and hierarchical relationships between predictor variables, (iv) acceptance of the missing values and the ability to improve final model accuracy through handling small perturbations in the training data set, and (v) the algorithm is resilient to overfitting (Carreiras *et al.,* 2012b; Gara *et al.,* 2014b; Güneralp *et al.,* 2014; Moisen *et al.,* 2006). In addition, the stochastic gradient boosting algorithm, unlike other approaches, utilises the combined strengths of bagging (a machine-learning algorithm, which has the capability to improve model stability and the final predictive accuracy) and boosting (an adaptive method for integrating multiple simple models, to provide improved predictive performance methods) (De'ath, 2007; Elith *et al.,* 2008; Friedman, 2002; Leathwick *et al.,* 2006; Moisen *et al.,* 2006). On the other hand, stepwise multiple linear regression was also applied in modelling AGB and the obtained results were compared to those obtained, using the SGB regression algorithm. In addition, statistical parameters, such as tolerance (Tol.) and variable inflator factor (VIF), were applied, to test for multicollinearity effects. To

demonstrate multicollinearity problems, a value less than 0.10 for tolerance (Belsley, 1991) and a value greater than 10 for VIF were considered as determinants (Sarker and Nichol, 2011).

6.2.7 Landsat-8 OLI data Preparation for AGB Estimation

Models of AGB were developed using the linear regression, stepwise multiple linear regression and stochastic gradient boosting regression models, with spectral/texture information as the predictor variables and AGB as the response variable. Specifically, the remotely sensed predictor variables were: (i) raw-spectral bands, (ii) raw-texture measures, (iii) spectral band ratios, (iv) vegetation indices, and (v) texture band ratios. Models of estimating aboveground biomass were conducted, following three analytic steps:

1. raw spectral-bands and spectral-band ratios derived from the Landsat-8 OLI sensor were used to model AGB individually in linear regression. The raw spectral-bands and spectral-band ratios were combined and collectively used in modelling AGB, using stepwise multiple linear regression, as well as using stochastic gradient boosting regression models. The strength of stepwise multiple linear regression and the stochastic gradient boosting algorithm (using R^2, RMSE, CV-RMSE and RMSE%) were then compared in estimating AGB and the best performing algorithm was then applied on texture metrics,
2. raw band texture measures (*i.e.* 10 texture metrics on four window sizes and four bands with some combination of offsets) from a single Landsat-8 OLI spectral band were individually tested in modelling AGB, using a SGB model. In addition, texture metrics were then combined and collectively used in modelling AGB, using stochastic gradient boosting regression model, and
3. vegetation indices and band-texture ratios (b2/b3, b2/b4, b2/b5, b3/b4, b3/b5 and b4/b5) generated from the Landsat-8 OLI sensor were tested individually in modelling AGB, using a SGB model. Similarly, texture metrics were then combined and collectively used in modelling AGB, using the SGB regression model.

In this study, field measured forest AGB values were used to train and validate the model. Thus, the field measured forest AGB data set was randomly split into 70% (two-thirds of field data) and 30% (one-third of field data) for model building and testing (Kohavi, 1995; Riggins *et al.,* 2009). The SGB model was parameterized, using the three most important user-defined hyper-parameters, namely: (i) the learning rate (*lr*), critical for determining the contribution of each tree to the final

model; (ii) the tree complexity (*tc*), which represents the number of independent variables interacting, to determine each split; and (iii) the number of regression trees (nt): (models that relate a response to their predictors by recursive binary splits) (Elith *et al.,* 2008; Hastie *et al.,* 2001; Leathwick *et al.,* 2006). For this work, the shrinkage rate (determines the contribution of each tree to the growing model) was set between 0.0001 and 0.1, whereas the tree complexity (which controls whether interactions are fitted) was set between 1 and 5. The bagging fraction, which determines a fraction of the training data set selected randomly for computing each tree, was set at 0.3 and 0.5 during data analysis. To assess the tested regression model accuracies, the coefficient of determination (R^2), root mean square error (RMSE), percentage root mean square error (RMSE%) and coefficient of variance of the root mean square error (RMSE-CV) were computed, using the test data set (Eq. 6.1-6.3).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{Equation 6.1}$$

Where; RMSE is the Root Mean Square Error; n is the number of observed values; y_i is the observed value, $\hat{y}_i$ is the estimated values, $\bar{y}$ mean of the observed values.

$$RMSE(\%) = \frac{\sqrt{\frac{1}{n}\sum_{1=n}^{n}\left(y_i - \hat{y}_i\right)^2}}{\bar{y}} \times 100 \qquad \text{Equation 6.2}$$

Where; RMSE% is the Percentage Root Mean Square Error; n is the number of observed values; y_i is the observed value, $\hat{y}_i$ is the estimated values, $\bar{y}$ mean of the observed values.

$$RMSE - CV = \left(\frac{RMSE}{\bar{y}}\right) \qquad \text{Equation 6.3}$$

Where; RMSE-CV is the Coefficient of Variance of the Root Mean Square Error; RMSE is the Root Mean Square Error; $\bar{y}$ mean of the observed values.

6.3 Results

6.3.1 Analysis I: Landsat-8 OLI raw spectral-bands vs. raw texture bands

A few texture parameters derived, using a 3 x 3 window size and an offset of [0, 1], were identified as the best variables for the accurate estimation of AGB (Table 6.3). These include mean, entropy, correlation, dissimilarity, homogeneity and variance. It was observed that raw Landsat-8 OLI bands yielded weaker AGB estimates across all tree species (*i.e.* ED, EG, PT and all species data (combined data)), when compared to estimates obtained using raw Landsat-8 texture measures (Table 6.3 and Figure 6.2). Despite poor performance from the raw Landsat-8 bands, the near-infrared and the green bands yielded slightly higher AGB estimates for ED, EG, PT and all species data combined. Using texture parameters derived from the NIR band, an R^2 value of 0.29, RMSE of 19.23 t ha^{-1} and CV-RMSE of 0.37 were obtained for ED. Similarly, using texture parameters derived from the green band, an R^2 value of 0.24, RMSE of 32.26 t ha^{-1} and CV-RMSE of 0.62 were obtained for ED species. A similar performance was observed for *Eucalyptus grandis, Pinus taeda* and all species. Comparatively, raw texture measures derived from Landsat-8 OLI yielded overall better AGB estimates than any individual spectral band. For instance, high average R^2 and RMSE and CV-RSME were attained from the use of Landsat-8 OLI raw texture measures derived from the NIR, green and blue bands, respectively. On the contrary, the use of the step-wise multiple linear regression, together with a stochastic gradient boosting algorithm, shows that the latter performed better, yielding the highest R^2 (0.51), the lowest RMSE (52.31 t ha^{-1}) and RMSE% (37.53%) for all species data sets combined (Table 6.3).

Moreover, it can be observed that, when all Landsat-8 OLI spectral bands (*i.e.* combined raw Landsat-8 OLI bands) were used, AGB estimates improved significantly, compared to the use of individual raw Landsat-8 OLI bands. Overall, the combined Landsat-8 OLI raw texture measures yielded plausible AGB estimates for ED, EG, PT and all species (Figure 6.2a – 6.2c and Table 6.3). Although, combined raw Landsat-8 OLI bands produced lower AGB estimates, their results were better, when compared with those obtained using individual bands (Figure 6.2d).

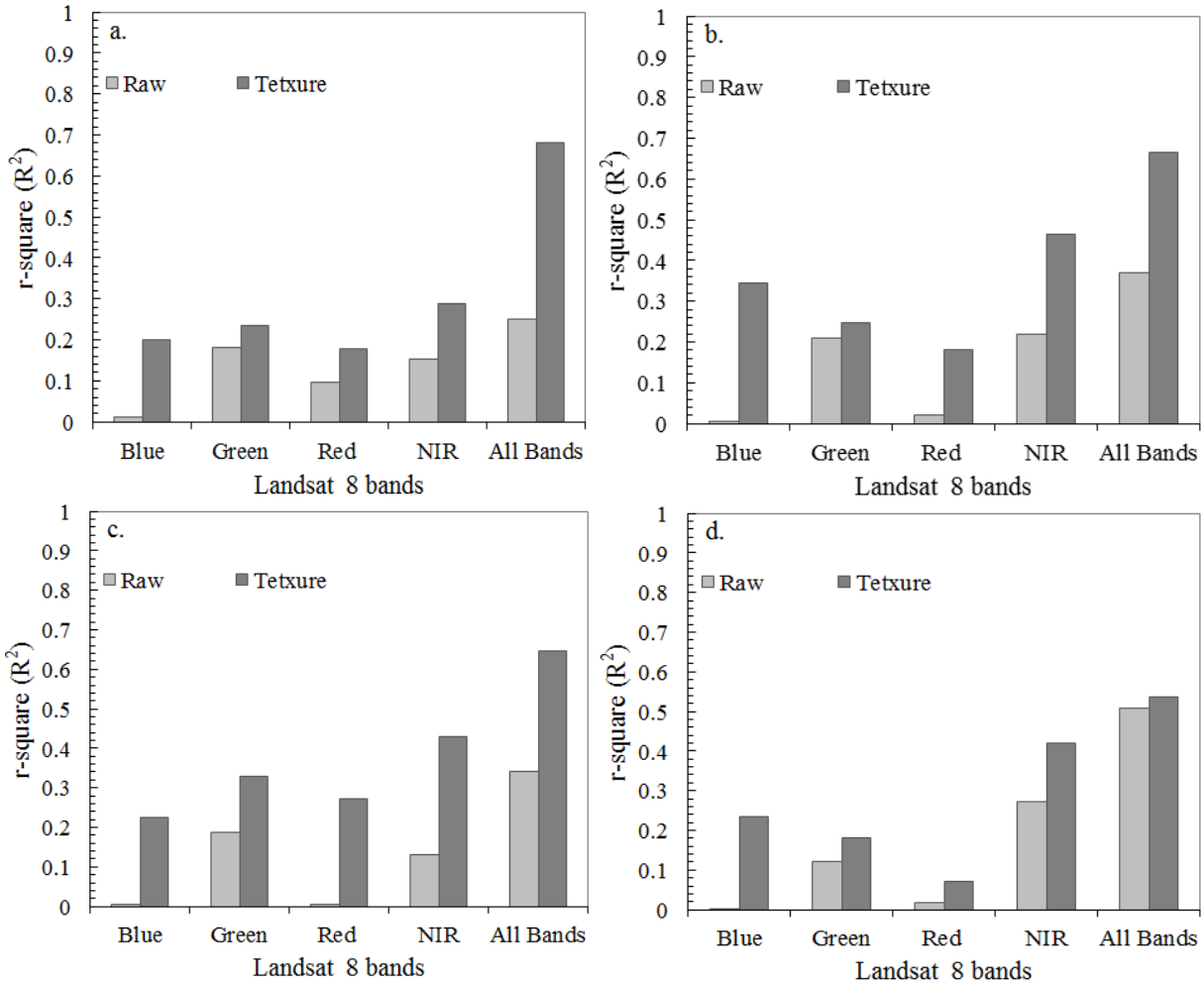


Figure 6.2 A comparison between the strength of Landsat-8 OLI sensor spectral bands and the best performing texture parameters derived using a 3 x 3 window size and an offset of [0, 1] in aboveground biomass estimation based on the three modelling techniques (*i.e.* linear regression, multiple linear regression, SGB algorithm). (a) represents *E dunii*,(b) stands for *E grandis*, (*c)* stands for *P taeda* and (*d)* represents all species data combined

Table 6. 3 Aboveground biomass estimates derived using Landsat-8 OLI spectral bands and the best performing texture parameters from a 3 x 3 window size and an offset [0, 1] using linear, stepwise-multiple linear and stochastic gradient boosting regression models

Model	Plant. species	Image data	Variable name and window size	R^2	RMSE t ha^{-1} (RMSE%)	CV-RMSE	Tol.	VIF
linear	E. dunii	Blue	—	0.01	39.64 (74.99%)	0.75	0.99	1.01
linear		Green	—	0.18	36.79 (69.60%)	0.70	0.82	1.22
linear		Red	—	0.09	40.31 (76.26%)	0.76	0.91	1.10
linear		NIR	—	0.15	38.97 (73.72%)	0.74	0.85	1.18
SMLR		All bands	—	0.19	35.12 (66.44%)	0.66	0.81	1.23
SGB		All bands	—	0.25	29.68 (56.15%)	0.56	0.75	1.33
linear	E. grandis	Blue	—	0.01	78.02 (45.81%)	0.46	1.00	1.00
linear		Green	—	0.21	61.93 (36.37%)	0.36	0.79	1.27
linear		Red	—	0.02	74.25 (43.60%)	0.44	0.98	1.02
linear		NIR	—	0.22	37.40 (21.96%)	0.22	0.78	1.28
SMLR		All bands	—	0.26	46.52 (27.32%)	0.27	0.74	1.35
SGB		All bands	—	0.33	55.46 (32.57%)	0.33	0.67	1.49
linear	P. taeda	Blue	—	0.01	78.53 (38.11%)	0.38	0.99	1.01
linear		Green	—	0.19	63.11 (30.63%)	0.31	0.81	1.23
linear		Red	—	0.01	74.89 (36.34%)	0.36	0.99	1.01
linear		NIR	—	0.13	59.47 (28.86%)	0.29	0.87	1.15
SMLR		All bands	—	0.20	46.52 (22.57%)	0.23	0.80	1.25
SGB		All bands	—	0.34	38.29 (27.37%)	0.27	0.66	1.52
linear	All species	Blue	—	0.01	80.73 (57.91%)	0.58	0.99	1.01
linear		Green	—	0.12	71.34 (51.18%)	0.51	0.88	1.14
linear		Red	—	0.02	84.13 (60.39%)	0.60	0.98	1.02
linear		NIR	—	0.27	69.66 (49.79%)	0.50	0.73	1.37
SMLR		All bands	—	0.24	61.73 (44.29%)	0.44	0.76	1.32
SGB		All bands	—	0.51	52.31 (37.53%)	0.37	0.49	2.04
SGB	E. dunii	b2 texture	1, 5, 6 (3 x 3)	0.20	36.34 (68.75%)	0.69	0.80	1.25
SGB		b3 texture	6, 1, 8, 3 (3 x 3)	0.24	32.26 (61.03%)	0.62	0.76	1.32
SGB		b4 texture	6, 1, 5 (3 x 3)	0.18	29.37 (55.56%)	0.56	0.82	1.22
SGB		b5 texture	1, 8, 6,2 (3 x 3)	0.29	19.23 (36.38%)	0.37	0.71	1.41
SGB		All data	1, 6, 8, 5, 3, 2 (3 x 3)	0.68	10.29 (19.47%)	0.20	0.32	3.13

SGB	E. grandis	b2 texture	1, 6, 5	(3 x 3)	0.35	37.35 (21.93%)	0.22	0.65	1.54
SGB		b3 texture	6, 1, 8, 3	(3 x 3)	0.25	33.42 (19.62%)	0.20	0.75	1.33
SGB		b4 texture	6, 1, 5	(3 x 3)	0.19	31.12 (18.27%)	0.18	0.81	1.23
SGB		b5 texture	8, 1, 2,6	(3 x 3)	0.46	29.30 (17.20%)	0.17	0.54	1.85
SGB		All data	2, 6, 8, 1, 3, 5	(3 x 3)	0.67	12.13 (07.12%)	0.07	0.33	3.03
SGB	P. taeda	b2 texture	5, 6, 1	(3 x 3)	0.23	41.55 (20.16%)	0.21	0.77	1.30
SGB		b3 texture	6, 1, 8, 3	(3 x 3)	0.33	35.23 (17.10%)	0.17	0.67	1.49
SGB		b4 texture	6, 5, 1	(3 x 3)	0.27	40.58 (19.69%)	0.20	0.73	1.37
SGB		b5 texture	8, 1, 6,2	(3 x 3)	0.43	21.19 (10.28%)	0.10	0.57	1.75
SGB		All data	1, 2, 3, 5, 8, 6	(3 x 3)	0.65	15.77 (07.65%)	0.08	0.35	2.86
SGB	All species	b2 texture	1, 6, 5	(3 x 3)	0.23	59.94 (43.00%)	0.43	0.77	1.30
SGB		b3 texture	6, 3, 8, 1	(3 x 3)	0.18	63.15 (45.30%)	0.46	0.82	1.22
SGB		b4 texture	5, 1, 6	(3 x 3)	0.07	71.03 (50.96%)	0.51	0.93	1.08
SGB		b5 texture	1, 2, 6, 8	(3 x 3)	0.42	29.03 (20.83%)	0.21	0.58	1.72
SGB		All data	1, 2, 8, 5, 3, 6	(3 x 3)	0.55	26.33 (18.89%)	0.19	0.45	2.22

**SMLR = stepwise multiple-linear regression; SGB = stochastic gradient boosting. 1-mean, 2-variance, 3-homogeneity, 4-contrast, 5-dissimilarity, 6-entropy, 7-second moment, 8-correlation, 9-standard deviation and 10-data range.

6.3.2 Analysis II: Landsat-8 OLI raw spectral – band ratios vs. texture band ratios

This study finds that texture-based models, derived using a stochastic gradient boosting algorithm, display enhanced accuracy in AGB estimation, relative to the use of spectral bands and spectral vegetation indices (Figures 6.3 a –d). Although the results derived, using both raw spectral band ratios and texture ratios, appear to be promising, raw spectral band ratios based on AGB estimates generally yielded low accuracy. For instance, weaker R^2 values were obtained from the use of raw spectral band ratios in estimating AGB for individual forest species and all species investigated. Figures 6.3a and 6.3b show that blue/ NIR, green/ NIR and red/ NIR raw spectral band ratios had slightly higher R^2 values, especially when compared to the use of single raw spectral bands. However, a comparison of AGB estimates between raw-spectral band ratios and raw-texture ratios indicate improved performance from the latter (Figure 6.3a and 6.3b). Overall, Landsat-8 OLI sensor texture-ratios derived from the following band combinations: band two/band three, band three/ band five and band four/ band five had a stronger correlation between predicted and observed biomass than other methods investigated.

The improved and higher R^2 values of 0.49 (RMSE = 20.45 t ha^{-1} and CV-RMSE of 0.39), 0.39 (RMSE = 35.60 t ha^{-1} and CV-RMSE of 0.21), 0.52 (RMSE = 21.85 t ha^{-1} and CV-RMSE of 0.11) were obtained for ED, EG and PT species ABG estimates, respectively, from red/NIR texture ratio. A similar performance from the red/NIR texture ratio was observed in estimating all species AGB, where an R^2 value 0.27 (RMSE = 35.50 t ha^{-1} and CV-RMSE of 0.26) was attained (Figures 6.3a – 6.3b and Table 6.4). Texture-based AGB estimates demonstrate a substantial enhancement, when compared with the best findings obtained, using Landsat-8 OLI raw spectral and band ratios. A further demonstration of good performance in AGB estimation was also observed from the results obtained, using the combined Landsat-8 OLI texture ratios (Figures 6.3a – 6.3b). The use of combined texture ratios yielded even higher R^2 values of 0.76 (RMSE = 9.55 t ha^{-1} and CV-RMSE of 0.18); 0.74 (RMSE = 12.81 t ha^{-1} and CV-RMSE of 0.08); 0.74 (RMSE = 12.67 t ha^{-1} and CV-RMSE of 0.06) and 0.53 (RMSE = 20.15 t ha^{-1} and CV-RMSE of 0.15) overall for ED, EG, PT and all species data sets, respectively. It is also important to note that the 3 x 3 window size and offset of [0,1], yielded the best overall results, when compared to the results obtained using other window sizes (*i.e.* 5 x 5, 7 x 7 and 9 x 9) and the other offsets, which include [1,1], [0,1] and [1,-1].

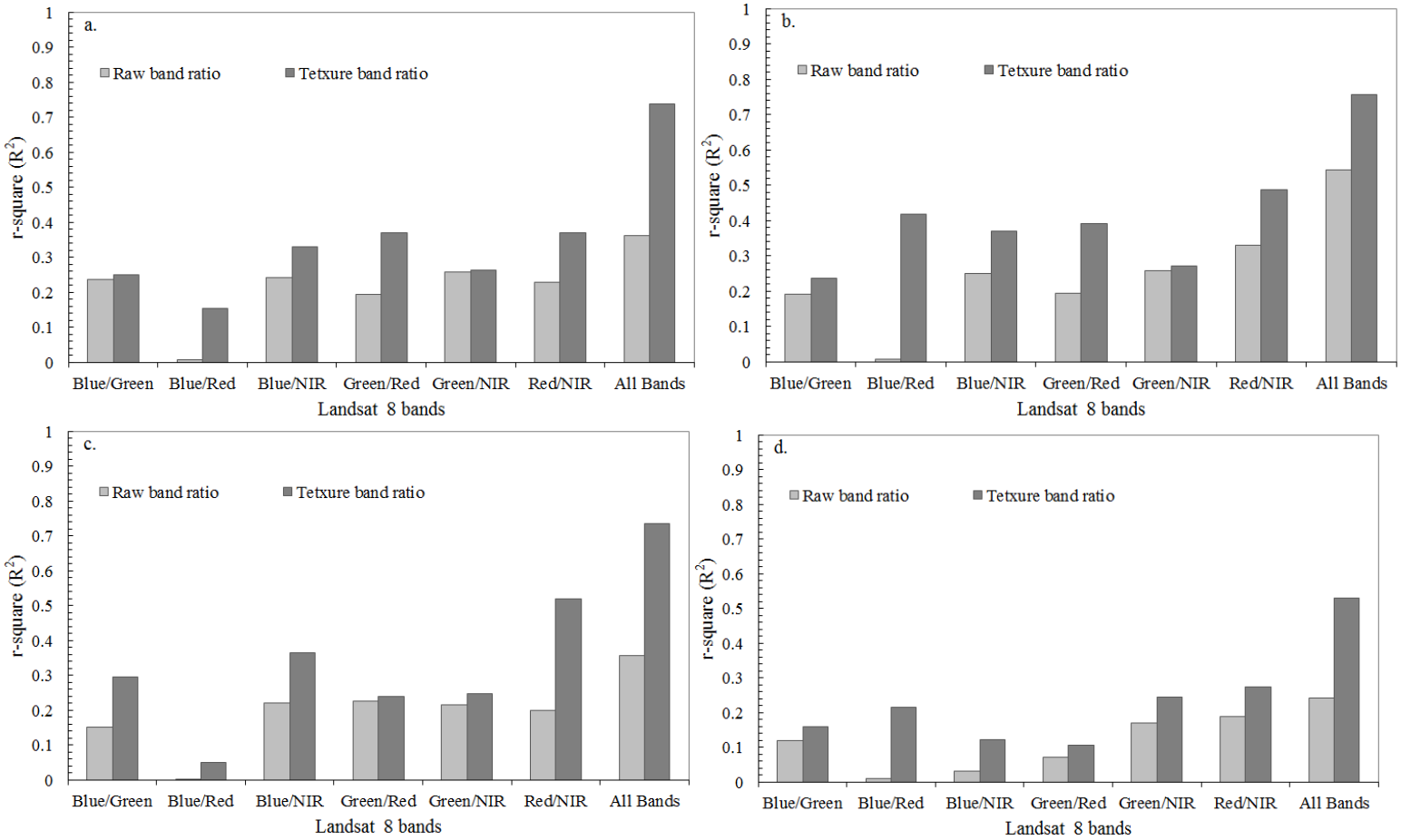


Figure 6.3 A comparison between the strength of Landsat-8 OLI sensor spectral band ratios and texture measure band ratios in aboveground biomass estimation using a stochastic gradient boosting algorithm. (a) represents *E dunii*, (b) stands for *E grandis*, (c) stands for *P taeda* and (d) represents all species data combined. The exact texture metrics used in each ratio calculation are presented in Table 6.3

Table 6. 4 Aboveground biomass estimates derived using Landsat-8 OLI band texture ratios from a 3 x 3 window size and an offset [0, 1], using stochastic gradient boosting regression models

Model	Plant. species	Image data	Selected texture ratios	R^2	RMSE t ha^{-1}(RMSE%)	CV-RMSE	Tol.	VIF
	E. dunii	b2/b3	1, 6, 2	0.24	32.28 (61.07%)	0.62	0.76	1.32
		b2/b4	6, 5, 8	0.42	30.11 (56.96%)	0.58	0.58	1.72
		b2/b5	6, 1	0.37	36.57 (69.18%)	0.70	0.63	1.59
SGB		b3/b4	8, 6, 2	0.39	35.63 (67.40%)	0.68	0.61	1.64
		b3/b5	1, 6, 8	0.27	27.92 (52.82%)	0.53	0.73	1.37
		b4/b5	6, 7, 5	0.49	20.45 (38.69%)	0.39	0.51	1.96
		All data	4, 1, 6, 5, 8, 7, 2	0.76	09.55 (18.07%)	0.18	0.24	4.17
	E. grandis	b2/b3	1, 2, 6	0.25	38.40 (24.27%)	0.23	0.75	1.33
		b2/b4	6, 8, 5	0.15	41.34 (19.55%)	0.24	0.85	1.18
		b2/b5	6, 1	0.33	33.29 (21.60%)	0.20	0.67	1.49
SGB		b3/b4	1, 8, 6	0.37	36.78 (22.45%)	0.22	0.63	1.59
		b3/b5	1, 6, 8	0.26	38.23 (20.90%)	0.22	0.74	1.35
		b4/b5	6, 5, 7	0.39	35.60 (07.52%)	0.21	0.61	1.64
		All data	6, 4, 1, 8, 5, 2, 7	0.74	12.81 (17.72%)	0.08	0.26	3.85
	P. taeda	b2/b3	1, 6, 2	0.30	36.52 (33.93%)	0.19	0.70	1.43
		b2/b4	6, 5, 8	0.05	69.92 (33.93%)	0.36	0.95	1.05
		b2/b5	6, 1	0.36	34.57 (16.78%)	0.18	0.64	1.56
SGB		b3/b4	8, 6, 2	0.24	37.69 (18.29%)	0.19	0.76	1.32
		b3/b5	1, 6, 8	0.25	39.74 (19.28%)	0.20	0.75	1.33
		b4/b5	6, 7, 5	0.52	21.85 (10.60%)	0.11	0.48	2.08
		All data	6, 4, 1, 8, 5, 2, 7	0.74	12.67 (06.15%)	0.06	0.26	3.85
	All species	b2/b3	1, 6, 2, 4	0.16	43.24 (30.91%)	0.31	0.84	1.19
		b2/b4	6, 5, 8	0.21	47.23 (33.76%)	0.34	0.79	1.27
		b2/b5	6, 1	0.12	67.96 (48.58%)	0.49	0.88	1.14
SGB		b3/b4	8, 6, 2	0.11	68.43 (48.92%)	0.50	0.89	1.12
		b3/b5	1, 6, 8	0.24	51.29 (36.66%)	0.37	0.76	1.32
		b4/b5	6, 7, 5	0.27	35.50 (25.38%)	0.26	0.73	1.37
		All data	1, 6, 8, 5, 2, 7, 4	0.53	20.15 (14.40%)	0.15	0.47	2.13

**SMLR = stepwise multiple-linear regression; SGB = stochastic gradient boosting. 1-mean, 2-variance, 3-homogeneity, 4-contrast, 5-dissimilarity, 6-entropy, 7-second moment, 8-correlation, 9-standard deviation and 10-data range.

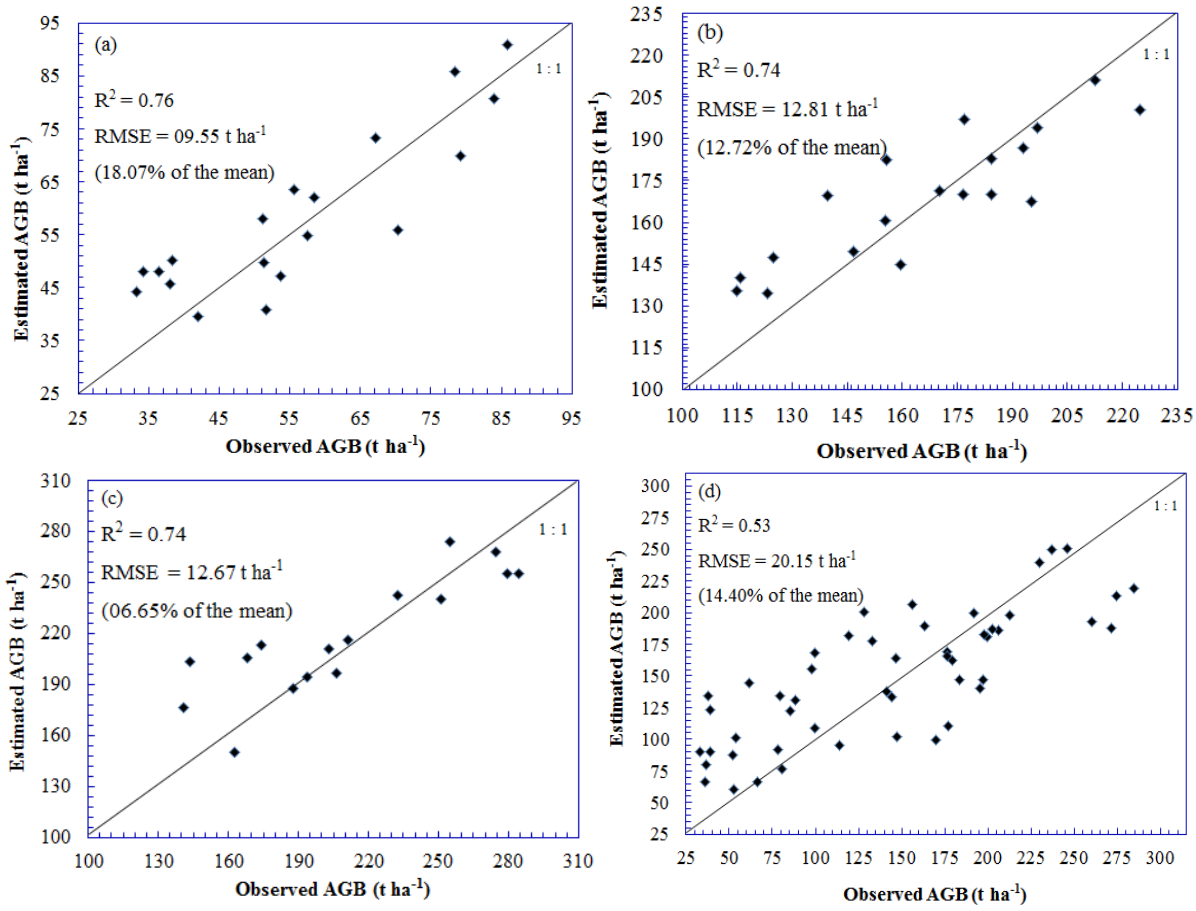


Figure 6.4 Scatterplots of the predicted vs. observed AGB of the best models. (a), (b), (c) and (d) represent *E. dunii, E. grandis, P. taeda* and all species data sets combined, respectively based Landsat-8 OLI texture ratios derived using a 3 x 3 window size, an offset of [0, 1] and stochastic gradient boosting algorithm

6.3.3 Analysis III: Landsat-8 OLI derived vegetation indices vs. texture band ratios

This work finds that Landsat-8 OLI derived texture ratios performed better than the selected vegetation indices, in terms of R^2; RMSE and CV-RMSE (Figure 6.5a). For instance, the use of combined texture ratios resulted in high R^2 values of 0.76 (RMSE = 9.55 t ha^{-1} and CV-RMSE of 0.18) for *eucalyptus dunii,* whereas, based on the five selected vegetation indices, weaker results were obtained, producing an R^2 of 0.71 and a RMSE of 10.66 t ha^{-1}. A similar performance was also observed for the other two species *i.e. Eucalyptus grandis* and *Pinus taeda,* as well as for all species data sets combined. Figure 6.5b shows the AGB map obtained using the best performing stochastic gradient boosting model, determined from Landsat-8 OLI texture band ratios (mean, entropy, dissimilarity, correlation, homogeneity, and variance), computed using the optimal window size of 3 x 3 and an offset of [0, 1].

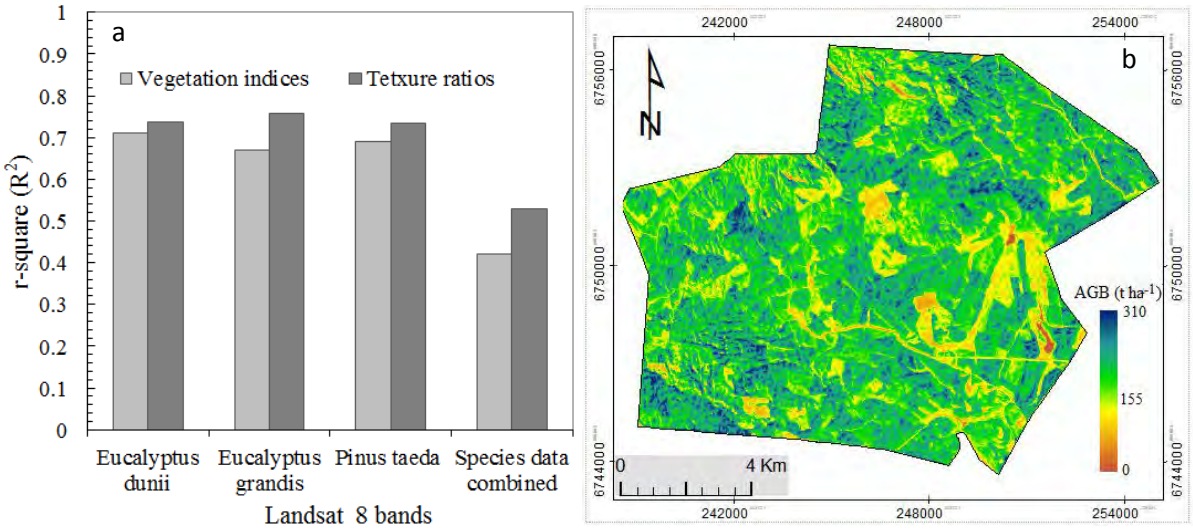


Figure 6.5 (a) A comparison between the strength of Landsat-8 OLI sensor spectral vegetation indices and band texture ratios in AGB estimation and (b) best performing Landsat-8 OLI image texture- derived aboveground biomass map

6.4 Discussion

The present study aimed to investigate the performance of texture metrics derived from the newly-launched Landsat-8 OLI sensor in estimating plantation forest species AGB in the midlands region of KwaZulu-Natal, South Africa. Specifically, this research investigated whether the various texture metrics obtained from the 30-m Landsat-8 OLI sensor with refined-spectral properties, as well as the improved signal-to-noise ratio, have the potential to enhance AGB estimation, when compared to other methods investigated *i.e.* spectral bands, bands ratios and vegetation indices.

This research shows that texture metrics (*i.e.* mean, entropy, dissimilarity, correlation, homogeneity and variance), derived from the 30-m Landsat-8 OLI multispectral sensor, using a 3 x 3 window size and an offset of [0, 1] overall, enhanced AGB estimation accuracy, when compared to the use of: (i) raw-spectral bands, (ii) raw-spectral band ratios, and (iii) spectral vegetation indices. The findings of this study are similar to those available in literature, which demonstrate these above-mentioned texture parameters as critical for AGB estimation (Nichol and Sarker, 2011; Sarker and Nichol, 2011). For example, these studies have shown that mean, entropy, dissimilarity, correlation, homogeneity and variance texture parameters are capable of enhancing AGB estimation accuracies. In addition, the study has demonstrated that by using Landsat data, a 3 x 3 window size and an offset of [0, 1] results in improved AGB estimates, when compared to other offsets and larger window sizes. High AGB estimation accuracies in terms of R^2 and RMSE and CV-RMSE were obtained for the three different plantation forest species (ED, EG and PT) under study. Similar

performances from texture band ratios were also observed in estimating all species AGB from all three analysis stages (*i.e.* Analysis I, II and III). Improved performance and estimation accuracy from the 30-m multispectral Landsat-8 OLI sensor can be associated with the advanced image processing methodologies adopted (*e.g.* texture retrieval) and the advanced machine-learning algorithms applied in data analysis (Güneralp *et al.,* 2014). These results are in line with the findings from literature, which reported good AGB predictive accuracies obtained from the use of texture metrics, such as entropy, mean and correlation based on a 3 x 3 window size and an offset of [0, 1]. Literature further shows that machine-learning algorithms, such as stochastic gradient, have the potential to significantly improve AGB estimates. (De'ath, 2007; Moisen *et al.,* 2006). This is mainly because these algorithms utilise the combined strengths of bagging and boosting, which have the capability to improve model stability and the final predictive accuracy.

Furthermore, the results from this work have shown that texture metrics derived from medium multispectral sensors (10-100m pixel size) offers invaluable opportunities for the improvement and better understanding of forest AGB in: (i) areas with tall forest and complex canopies structure, (ii) areas associated with limited resources and (iii) limited access to high resolution satellite, such as hyperspectral, LiDAR and radar *etc*. The plausible performance and strength of texture metrics in improving AGB estimates is in line with the findings from studies done elsewhere, using other optical sensors, such as synthetic aperture radar (SAR) and Worldview-2, among others (Bastin *et al.,* 2014; Champion *et al.,* 2013; Cutler *et al.,* 2012; Eckert, 2012a; Ploton *et al.,* 2013; Sarker and Nichol, 2011; Singh *et al.,* 2014). The credible performance displayed by the use of Landsat-8 OLI derived texture metrics may be attributed to the presence of crucial and sensitive vegetation information, as well as related biophysical properties, such as tree age, leaf area index and, most importantly, biomass (Champion *et al.,* 2008; Fuchs *et al.,* 2009), a complex challenge when using spectral vegetation indices. Eckert, (2012) states that texture analysis is efficient in addressing saturation problems associated with vegetation indices, when mapping biomass, especially in dense canopies, as it correlates more with AGB and carbon, than spectral parameters. The other reason for the unique performance of texture measures in estimating AGB can be explained by their capability to detect varying forest canopy structural characteristics, as well as the more inherent sensitivity to the spatial aspects of the canopy, than spectral reflectance or band ratios (Eckert, 2012).

In addition, improvements in AGB estimation were obtained from the use of simple texture ratios derived from a 30-m Landsat-8 OLI sensor. Thus far, no study, to the best of our knowledge, has

used texture ratios derived from medium spatial resolution satellite data (*e.g.* pixel size greater than 10-m) to estimate AGB. The greater performance from optical Landsat-8 OLI image texture metrics, when compared to spectral vegetation indices, can be attributed to a number of factors and these include: (i) the simplification of complex forest canopy structures and the stronger correlation between texture measures and biophysical properties, such as leaf area index and AGB (Barbosa *et al.,* 2014; Eckert, 2012a; Fuchs *et al.,* 2009; Nichol and Sarker, 2011; Sarker and Nichol, 2011); and (ii) the capability to minimize atmospheric effects and canopy geometry, sensor view angle and sun angle by this sensor (Fan *et al.,* 2014; Franklin *et al.,* 1996; Laurent *et al.,* 2014; Mousivand *et al.,* 2014; Wulder *et al.,* 1998). In addition, the great performance from texture ratios in modelling AGB can be attributed to the fact that these ratios combine the strengths of both texture and ratios, which is critical in improving AGB estimation accuracies (Nichol and Sarker, 2011; Sarker and Nichol, 2011). In addition, the good performance of texture metrics derived from Landsat-8 OLI sensor can be attributed to its push-broom design, which is associated with an improved signal-to-noise ratio, high radiometric resolution of 12-bits, making it (i) more sensitive and robust in detecting crucial vegetation structural attributes (El-Askary *et al.,* 2014; Pahlevan and Schott, 2013); and (ii) the narrow or refined near-infrared spectral band that prevents the absorption effect of water vapour (*e.g.* at 0.825 μm), thereby permitting accurate surface spectral detection, while minimizing satellite spectral saturation problems (Lu, 2006).

One of the observations from this study is that, when forests have complex canopies (*e.g.* high biomass levels) (Basuki *et al.,* 2011), vegetation indices derived from medium spatial resolution sensors become less useful in estimating AGB, when compared to the use of texture metrics (Mutanga and Skidmore, 2004b; Thenkabail *et al.,* 2000). Despite the poor performance of vegetation indices in this study, other studies indicate their successful application in estimating AGB in simple structured forests, with less complex canopies, where AGB and LAI are low (*e.g.* less than 100% vegetation cover) (Darvishzadeh *et al.,* 2011; Englhart *et al.,* 2012; Gallardo-Cruz *et al.,* 2012; Goh *et al.,* 2014; Vashum and Jayakumar, 2012). The other possible explanation for the poor performance from the spectral vegetation indices computed in this study may be as a result of the presence of tall forests. Hence, shadow effects might have contributed immensely to the spectral reflectance captured by the sensors, resulting in their poor performance. Most of these vegetation indices have been designed, based on simple and less complex forest canopies, hence their inability to estimate AGB in medium-to-dense forests. Moderately weaker performance by raw spectral bands and vegetation indices derived from Landsat-8 OLI sensor observed in this

study. This may also be attributed to saturation problems associated with medium-to-coarse resolution multispectral sensors when estimating AGB in medium-to-high density vegetation canopies. Literature shows that vegetation indices *e.g.* normalised difference vegetation indices derived mainly using the red and near infrared bands of the electromagnetic spectrum, suffer immensely from saturation problems, as they asymptotically reach a saturation level beyond a certain density of AGB and/or leaf area index (Adam *et al.,* 2014; Gao *et al.,* 2000; Mutanga and Skidmore, 2004b; Sarker and Nichol, 2011; Thenkabail *et al.,* 2000; Thenkabail *et al.,* 2004b). These studies have clearly demonstrated that the use of spectral vegetation indices, such as the normalised difference vegetation index, have limited significance in estimating AGB in areas with high vegetation cover during the peak of periods.

Finally, improvement in AGB estimation, using the 30-m Landsat-8 OLI multispectral sensor derived texture metrics, can be linked with the strength of advanced machine learning algorithms, when compared to the use of stepwise multiple linear regression methods. Previous research has precisely demonstrated the strength and capabilities of advanced machine learning techniques (*i.e.* SGB and the most popular RF regression ensembles) in simplifying and enhancing AGB estimation accuracy, when compared to the use of simple multiple linear methods (Adam *et al.,* 2014; Carreiras *et al.,* 2012b; Gara *et al.,* 2014b; Güneralp *et al.,* 2014). One of the most important strengths of the SGB regression ensemble is its robustness in handling inaccurate training data, outliers, missing and unbalanced data sets (Elith *et al.,* 2008), as well as the capacity to handle, ascertain and choose the most crucial predictor variables from a huge amount of predictors (Carreiras *et al.,* 2012b; Dube *et al.,* 2015).

6.5 Conclusion

This research investigated the performance and strength of texture metrics calculated from the newly-launched push-broom designed Landsat-8 OLI sensor in estimating medium-density AGB for plantation forest species in the midlands region of KwaZulu-Natal, South Africa.

The findings of this work have demonstrated that:

1. texture metrics yielded more accurate AGB estimates, when compared to the use of Landsat-8 OLI sensor derived spectral vegetation indices,
2. AGB for medium-density *Eucalyptus dunii, Eucalyptus grandis, Pinus taeda* and all species data combined can be accurately estimated by using the newly-launched 30-m Landsat-8

OLI multispectral sensor derived texture ratios. Thus far, no study, to the best of our knowledge, has examined the strength of texture ratios derived from optical remote sensing sensors with a pixel size greater than 10-m for AGB estimation,

3. texture parameters derived from the newly-launched Landsat-8 OLI data set provide an important tool for the creation, mapping and continuous updating AGB maps, which is critical for well-informed land management purposes, and
4. the SGB regression ensemble has proven useful and more dependable in enhancing AGB for medium-density plantation forest species, based on medium spatial resolution multispectral Landsat-8 OLI sensor derived texture metrics.

Overall, this study presents an operational, successful and effective application of texture metrics derived from the newly-launched push-broom Landsat-8 OLI multispectral data and stochastic gradient boosting algorithm, in estimating AGB in resource constrained regions. The results of this study can possibly aid in understanding the contribution of forest ecosystems in regulating atmospheric carbon. Despite plausible AGB estimation accuracies obtained, using remotely sensed Landsat-8 OLI texture metrics, the utility of remotely sensed data is yet to be assessed in quantifying the variations of AGB across vital tree structural attributes, which remain difficult to obtain, in data-scarce sub-Saharan Africa. Therefore, future research has to evaluate the utility of models developed, using remotely sensed data, in quantifying and mapping the variability and allocation patterns of aboveground carbon stocks across different forest structural attributes.

CHAPTERS SEVEN AND EIGHT

FORESTRY APPLICATION AND SYNTHESIS

7. QUANTIFYING THE VARIABILITY AND ALLOCATION PATTERNS OF ABOVEGROUND CARBON STOCKS ACROSS PLANTATION FOREST STRUCTURAL ATTRIBUTES AND AGE

This Chapter is based on:

Dube, T, Mutanga, O. 2015. Quantifying the variability and allocation patterns of aboveground carbon stocks across plantation forest types, structural attributes and age in sub-tropical coastal region of KwaZulu Natal, South Africa using remote sensing. *Journal of Applied Geography,* **64**, 55-65.

Abstract

Quantifying the variability and allocation patterns of aboveground carbon stocks across plantation forests is central in deriving accurate and reliable knowledge and understanding of the extent to which these species contribute to the global carbon cycle and towards minimizing climate change effects. The principal objective of this study was to quantify the variability and allocation patterns of aboveground carbon stocks across *Pinus* and *Eucalyptus* plantation forests, tree-structural attributes (*i.e.* stems, barks, branches and leaves) and age groups, using models that have been developed based on remotely sensed data. The results of this study demonstrate that aboveground carbon stocks significantly ($\alpha = 0.05$) vary across different plantation forest species types, structural attributes and age. The *Pinus taeda* and *Eucalyptus grandis* species contained aboveground carbon stocks above 110 t C ha^{-1}, and *Eucalyptus dunii* had 20 t C ha^{-1}. Across plantation forest tree structural attributes, stems contained the highest aboveground carbon stocks, compared to barks, branches and leaves. Aboveground carbon stock estimates also varied significantly ($\alpha = 0.05$) with stand age. Mature plantation forest species (*i.e.* between 7 and 20 years) contained the highest aboveground carbon stock estimates of approximately 120 t C ha^{-1}, when compared to younger species (*i.e.* between 3and 6 years), which had approximately 20 t C ha^{-1}. The map of aboveground carbon stocks showed distinct spatial patterns across the entire study area. The findings of this study are important for understanding the contribution of different plantation forest species, structural attributes and age in the global carbon cycle and possible climate change moderation measures. In addition, this study demonstrates that data on vital tree structural attributes, previously difficult to obtain, can now be easily derived from cheap and readily-available satellite data for inventorying carbon stock variability.

Keywords: carbon quantification, carbon allocation, structural components, tree age, carbon variability.

7.1 Introduction

Plantation forest ecosystems account for a dominant share of terrestrial carbon (UNFCCC, 1998). The sequestered carbon is stored both in the form of biomass (stems, barks, branches and leaves, foliage, roots, *etc.*) and in the form of organic carbon in the soil (Chen *et al.,* 2015; Dalu *et al.,* 2015; Raich *et al.,* 2014). Information on forest carbon storage and allocation patterns is therefore central for effective bioenergy production, the detection of land-use change and the assessment of carbon stocks for initiatives, such as REDD management practices, planning and understanding their contribution on the global carbon cycle (Carreiras *et al.,* 2012b; Chen *et al.,* 2015; Chinembiri *et al.,* 2013a; Gara *et al.,* 2014b; IPCC, 2003b; Laurin *et al.,* 2014; UNFCCC, 1998). Moreover, understanding the spatial patterns of these carbon reserves is also important, as they significantly contribute to the Gross Domestic Product (GDP) of some countries. In South Africa, for instance, the economic value of these commercial forests is equivalent to approximately 7.3% (ZAR 20.4 billion) of national annual revenues. In addition, over 1.4 million people, directly or indirectly, depend on this industry for a living (Dube *et al.,* 2015).

So far, commercial forest resources cover approximately 3.1% (1.3 million ha) of the country's land surface (121.9m per ha) and form an essential base for timber, medicinal, pulp and paper products. Nearly 80% of these forest ecosystems are located in the south-eastern regions of the country, specifically in the Eastern Cape (11%), KwaZulu-Natal (38.9%) and Mpumalanga (40%) provinces (FSA, 2010). When these plantation forests are equated to other forests within South Africa, the dominant hardwoods are *Eucalyptus* species, covering 39% and *Acacia mearnsii,* with 17% of the total land area, whereas softwoods, mainly the *Pinus* forest species occupy the remaining 54% of the total land area (FSA, 2010). In addition, the majority of these plantation forests are managed on a rotational basis (the rotation length can range from six- to twety-five years, depending on the nature of the end product). For instance, when trees growth rate starts to diminish, they are clearfelled and another crop of trees is planted (Christie and Scholes, 1995; Schönau and Boden, 1982a). Despite the fact that these plantation forests eventually get felled, they account for the dominant share of terrestrial carbon stocks and the country's GDP.

Despite an increase in the areal extent of plantation forest species in sub-Saharan Africa and South Africa, in particular, their actual contribution to the global carbon cycle has not been fully quantified. This gap in knowledge can be attributed to the fact that most of these plantation forests are meant for commercial timber, pulp and paper production, and as a result most studies conducted

in these forests focus mainly on volume or yield estimation (Christie and Scholes, 1995; Schönau and Boden, 1982). In addition, most of these plantation forests are privately owned and are not easily accessible for biomass related studies by environmentalists, ecologists and the remote sensing communities. However, for a transition towards less global emissions and the reduction of atmospheric carbon, there is need for the timely-assessment of the role of these forests and the associated structural components, as well as age, in the global carbon cycle, especially in areas where their contribution as atmospheric carbon sinks is often ignored. A number of studies show that terrestrial ecosystems sequester large amounts of the atmospheric carbon, approximately 3 GtC per year (Patenaude *et al.,* 2004; UNFCCC, 1998b; 2001; 2011; Wei *et al.,* 2013), and normally account for approximately 80% of the earth's aboveground biomass (Giardina and Ryan, 2002; Pan *et al.,* 2011; Raich *et al.,* 2014; UNFCCC, 1998b).

Information more specifically related to plantation forests biomass or carbon sequestration is therefore central for deriving forest carbon stocks and associated carbon fluxes. Given the importance of terrestrial ecosystems in the carbon cycle, there is a need for carbon quantification at local or regional scales, to facilitate a more accurate, timely and precise assessment of the regional carbon cycle, carbon and bioenergy policies towards sustainable forest management. In African ecosystems, for instance, literature shows that there is great uncertainty in the current carbon balance and an unstable source, with a carbon sink of about 0.3 Pg Cyr^{-1}, when compared to other parts of the world (Laurin, *et al.,* 2014; Wolf, *et al.,* 2011). To some extent, this assertion holds because in sub-Saharan Africa, especially South Africa, studies concerning the carbon storage and allocation pattern among most forest ecosystems, particularly, plantation forest species, are still scarce (Dube et al. 2015). In addition, despite the vivid interest for carbon accounting in the region, no study has yet quantified the variability and allocation patterns of aboveground carbon stocks across different plantation forest species and different tree structural attributes.

The accurate and timely quantification of the variability of aboveground carbon stocks, across different plantation forest species, various tree structural attributes and age groups, is thus a critical step towards reducing the great uncertainity in the current forest carbon balance. Thus far, literature demonstrates that plant carbon allocation significantly varies between above- and below-ground components (Giardina and Ryan, 2002). For instance, the study by Giardina and Ryan (2002) shows that the total belowground carbon allocation of a *Eucalyptus saligna* plantation decreases

with stand age and this was facilitated by the nitrogen supply via regulating cytokinins and sucrose production (van der Werf *et al.,* 2006a). Chen *et al.* (2015) further showed the unequal carbon allocation in plants, by demonstrating that the root carbon storage was only between 16 and 20% of the total plant carbon storage and the proportion of different compartments to total plant carbon storage was largely influenced by tree traits. Now the main question that remains unanswered is which plantation forest species and structural components (*i.e.* stems, bark, branches and leaves) constitute the highest proportion of carbon storage. This gap in knowledge is further pointed out by Wei *et al.* (2013) who highlighted that AGB is bound to be linearly controlled by forest stand age and forest types, amongst other characteristics. The validity of this assertion is further supported by Pregitzer and Euskirchen (2004), who stated that forest carbon pools are largely affected by forest type, stand age and environmental factors. In that regard, for accurate and reliable quantification of plantation forest aboveground carbon stocks, all these variables have to be considered. The integration of numerous variables, particularly stand age characteristics, is critical, as it provides information about the growing stages at which these species substantially contribute to the global carbon cycle as sinks. Thus far, the relationship between aboveground carbon storage and tree-age has not yet been fully established. For instance, some studies have reported a decrease in tree carbon storage with age and vice-versa (Chen, *et al.,* 2015; Wei, *et al.,* 2013). It is, therefore, hypothsesized in this study that plant carbon allocation varies significantly across tree species, structural attributes and age groups.

It is also clear that accurate carbon accounting requires a strong understanding of the variability and allocation patterns in aboveground carbon stocks within and between different forest species, structural attributes and age groups. Ecologists, environmentalists and the remote sensing community usually attempt to quantify forest AGB and carbon stocks, using growth patterns, which minimizes bias and improves the estimation accuracy. Although great strides have been conducted on aboveground carbon stock estimation, studies that have quantified variations and allocation patterns specifically across various tree species and tree structural attributes (*i.e.* roots, stem, bark, branches and leaves) and stand age, especially in managed plantation forests, are scarce (Chen, *et al.,* 2015; Wei, *et al.,* 2013). Local measurements or field surveys are costly and there is a great interest in obtaining reliable estimates over large areas from remote sensing data. In addition, model accuracies and the spatial heterogeneity greatly increase the error of estimation obtained, using field data, and model simulation approaches generally lead to the substantial underestimation and

misunderstanding of the actual carbon sequestered by terrestial ecosystems (Baishya, *et al.,* 2009; Chen, *et al.,* 2015; Giardina and Ryan, 2002; Guo, Fang, *et al.,* 2010; Raich, *et al.,* 2014; Yang and Guan, 2008; Zhang, *et al.,* 2012).

This study seeks to evaluate plantation forests aboveground carbon stocks variability across different tree species, structural attributes (*i.e.* stems, barks, branches and leaves) and age groups, using remotely sensed data. Literature demonstrates that remote sensing data provides a better alternative or data-source critical for quantifying and providing the timely and spatial explicit variations of forest AGB, particularly over large extents and in areas where field surveys remain a challenge (Houghton, *et al.,* 2007; Yang, *et al.,* 2011; G. Zhang, *et al.,* 2014; Zhang and Kondragunta, 2006; Zhao, *et al.,* 2009). For instance, the large-scale forest biomass estimation by ground-based measurements requires a dense network of inventory plots, to reach good accuracies. In most areas, this is not practical, due to high costs and the required man power. Using remote sensing data is therefore the only practical option for the accurate and timely estimation of AGB or carbon stocks on these scales with affordable effort. The objectives of this study were therefore to quantify the variability and allocation patterns of aboveground carbon stocks across plantation forest types, structural attributes and age in KwaZulu-Natal, South Africa, using remote sensing data.

7.2 Materials and Methods

7.2.1 Location and Description of Study Area

Field surveys were carried out in the Sappi Clan forests located along the Albert Falls dam in the north-eastern part of Pietermaritzburg, KwaZulu-Natal, South Africa, at 29°.413 - 29°.296 S and 30°309 - 30°475 E. The altitude of the area varies between 644 and 1266 m a.s.l. The area is characterised by a mean annual rainfall, varying between 644 and 1266 mm, mainly from October to February, with a mean annual temperature around 21.7°C. The field samples were primarily dominated by 3- to 20-year old *Eucalyptus* and *Pine* species, namely, *Eucalyptus dunii, Eucalyptus grandis* and *Pinus taeda*. This work therefore focused on quantifying the structural variability and allocation patterns of aboveground carbon stocks of the above-mentioned plantation tree species, predominantly grown for pulp and paper production.

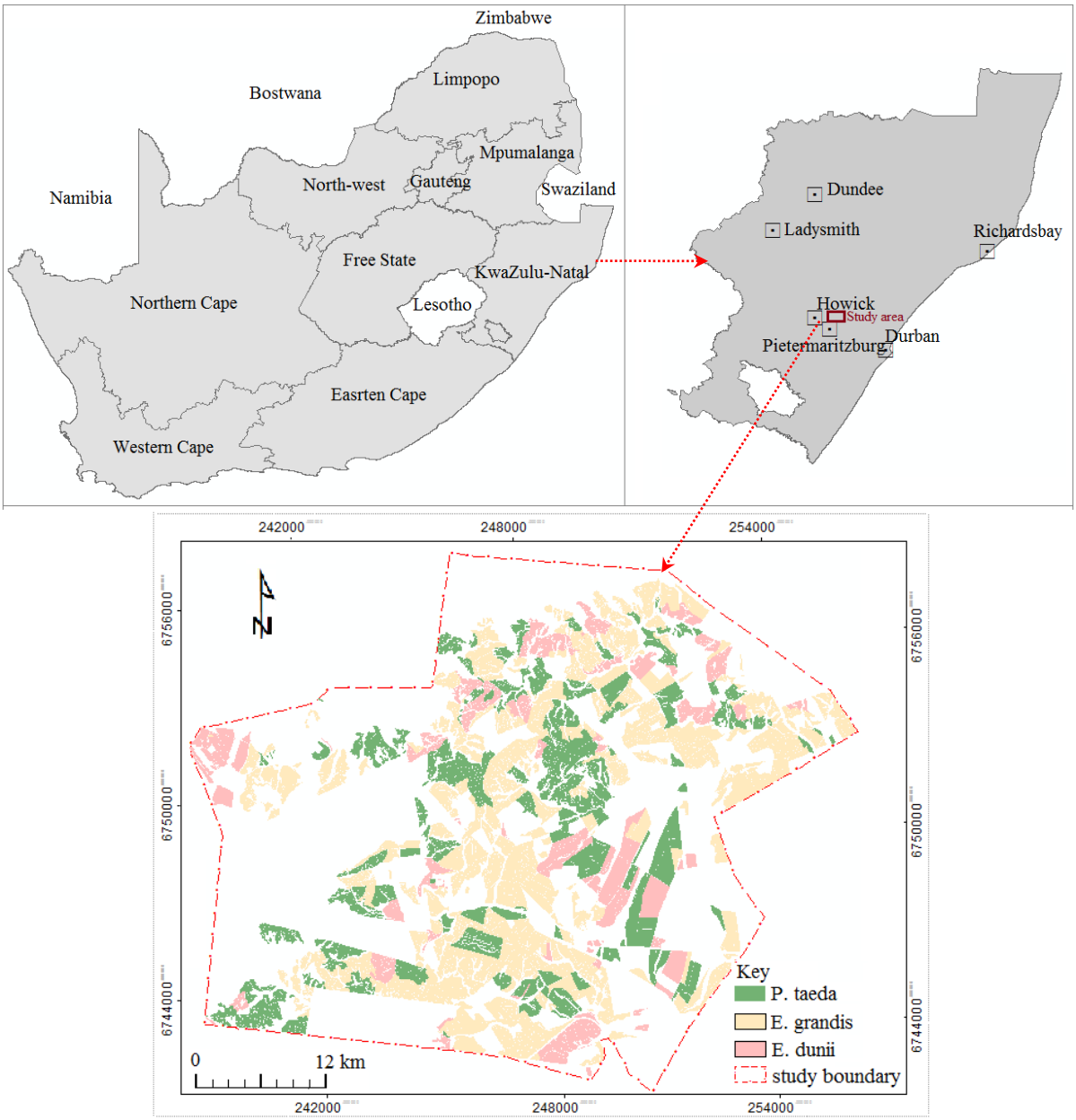


Figure 7.1 Location of the Clan Sappi Forests in the uMgeni Catchment, South Africa

7.2.2 Field Surveys and Remotely Sensed Aboveground Carbon Stocks

Field data sets were collected between the 30th of July and 22nd of August, 2013, concurrently with Sappi annual routine field data collection, using vector maps of the area, by courtesy of Sappi Forest), as well as a forest specialist to navigate to selected forest compartments and plots of interest. A grid-based systematic sampling method, with 10-m radius circular sample plots, systematically distributed (~100 m) within the stand, as detailed by Wessels and Kassier (1985), was used during field surveys. Data were collected mainly from plantation tree species (*Eucalyptus dunii*, *Eucalyptus grandis* and *Pinus taeda*), with diameter-at-breast height (DBH: at 1.3 m above

the ground) greater than, or equal to, five centimeters and 3 m in height, as recommended in literature (Gara *et al.,* 2014b; Gschwantner *et al.,* 2009). Diameter-at-breast height and tree heights were respectively measured, using the Haglof Digitech Calliper and Vertex IV laser instruments. All plot centre GPS locations for *Eucalyptus dunii* (n = 49), *Eucalyptus grandis* (n = 65) and *Pinus taeda* (n = 71) were captured at sub-meter accuracy, using the Tremble GeoXH 6000 series handheld Global Position System (GPS). Based on the field data, AGB for different tree structural attributes (*i.e.* stem, bark, branches and leaves) across different tree species were derived from the calculated volume (ha/m^3), using expansion factors (Dovey, 2009). In deriving total AGB for individual tree species (*i.e. Eucalyptus dunii, Eucalyptus grandis* and *Pinus taeda*), two methods were applied. For example, volume (m^3/ha), based on Bredenkamp (2000), together with species specific biomass expansion factors (BEFs) available in literature (Dovey, 2009), were used to derive total AGB for *Eucalyptus dunii* and *Eucalyptus grandis*. For *Pinus taeda*, DBH measurements were converted to AGB, using a general functional group equation (*i.e.* allometric equation), and subsequently summed across the entire plot (Penman, *et al.,* 2003). The allometric equation was developed in an area with rainfall of approximately 800 to 1500 mm and a temperature range of 21-34°C, comparable to those in the midlands region of KwaZulu-Natal, South Africa. It is also important to note that this equation was developed, based on DBH stretching from 0.6 to 56 cm, a characteristic almost identical to the area under study (Penman, *et al.,* 2003). The use of different methods in computing AGB in this study was driven by the prevalance of unique tree species with varying structural and taxonomical properties. Moreover, previous work indicates that for accurate AGB estimation, the use of species-specific mathematical functions is a necessity (Henry, *et al.,* 2011).

The derived AGB of the three species under study was then used to compute the biomass of each of the three tree structural components (*i.e.* stem, bark, branches and leaves), based on the already established multipliers (Dovey, 2009). The obtained biomass was then converted to aboveground carbon stocks, based on the factor 0.5. The Intergovernmental Panel on Climate Change proposes the application of standardized approaches, when evaluating the contribution of various terrestrial ecosystems in the global carbon cycle (Denman, 2007; Gil *et al.,* 2011; IPCC, 1996; MacDicken, 1997). Specifically, field measured AGB was converted to carbon, using the carbon fraction of dry biomass (*i.e.* equivalent to 0.5) (Dixon *et al.,* 1994; Houghton, 2005). In the literature, this value is demonstrated to vary slightly with species, or the AGB component under study *i.e.* understory vegetation, roots, branches, and trunk *etc*. (Dixon *et al.,* 1994; Houghton, 2005). Similarly, for this

study, 50% of the net field measured AGB was considered to be equivalent to the amount of carbon sequestered or stored in these forests (Dixon *et al.,* 1994; Houghton, 2005).

The field measured AGB data set was integrated with transformed Landsat-8 OLI image, using the SGB algorithm to map the spatial distribution of aboveground carbon stocks for stems, bark, branches and leaves within the Sappi Clan Forest area. The 30-m multispectral Landsat-8 OLI was used, because it has a reasonably improved spatial resolution and is cheap and available for vegetation mapping and biomass related applications, particularly in resource-limited environments, such as the sub-Saharan region, where costs of acquiring high spatial resolution remotely sensed data sets (*i.e.* with improved spectral and radiometric properties) remain a major challenge. Literature shows that the Landsat-8 OLI's improved radiometric resolution from 8 bits to 12 bits and signal-to-noise ratios made this sensor to be the most appropriate source of otherwise scarce data in understanding AGB variability across tree species, structural components and age groups in a mixed plantation forest (Dube and Mutanga, 2015). The technical details of the image processing approach used to map AGB, using Landsat-8 OLI, are provided in Dube and Mutanga (2015). The integration of field measured data and remotely sensed data was based on the best texture metrics derived in Dube and Mutanga (2015). The best texture metrics were derived from the Landsat-8 OLI spectral bands (i.e. band two–blue, band three – green, band four – red, and band five – near IR), using a 3 x 3 window size and an offset [0, 1] (Dube and Mutanga, 2015). These include: 1-mean, 2-variance, 3-homogeneity, 4-contrast, 5-dissimilarity, 6-entropy, 7-second moment, 8-correlation, 9-standard deviation and 10-data range. s (*i.e.* band two–blue, band three–green, band four – red, and band five – near IR). The best Landsat-8 OLI texture metrics selected, based on the SGB algorithm and used to derive aboveground carbon stocks of stems, bark, branches and leaves in this study, were the mean, entropy, dissimilarity, correlation, homogeneity and variance derived from the Landsat-8 OLI spectral band ratios.

7.2.3 Statistical Analysis

To evaluate the variability of aboveground carbon stocks across different plantation forest species, structural attributes and different age groups, statistical anaylisis was performed, using remotely sensed carbon estimates. The Analysis of Variance (ANOVA) procedure was performed to test for significant difference ($\alpha = 0.05$) in aboveground carbon stocks across different tree components (stem, bark, branches and leaves) and different age groups. The hypothesis that aboveground carbon stocks vary significantly amongst different tree species, structural components and age

groups was tested. One-way ANOVA was used with a post-hoc dunnet test for significant differences within the groups (*i.e*. three tree structural components, three age groups (*i.e.* 3-6, 7-11 and 12-20 years) and three groups of species). Statistical analysis was implemented, using SPSS version 13.0 software.

7.3 Results

7.3.1 Spatial distribution of aboveground carbon stocks in the study plantation across species, structural traits and age groups

The results in Figure 7.2 a-c show the spatial distribution of aboveground carbon stock estimates of the three plantation forest species (*i.e. Eucalyptus dunii*, *Eucalyptus grandis* and *Pinus taeda*) under study, derived from the Landsat-8 OLI remotely sensed data. It can be observed from Figures 7.2 and 7.3 that plantation forests aboveground carbon stocks vary significantly ($\alpha = 0.05$) across the area under study, with *Eucalyptus grandis* and *Pinus taeda* containing higher aboveground carbon stocks than *Eucalyptus dunii*. Furthermore, in Figure 7.3 and Table 7.1, it can be observed that plantation forest stems contain the highest carbon stocks, when compared to the bark and branches and leaves. These results clearly show that, of all three tree structural components, stems contain the highest carbon allocations followed by branches and leaves. The maps of aboveground carbon stocks for the three species also show distinct spatial patterns across the entire study area.

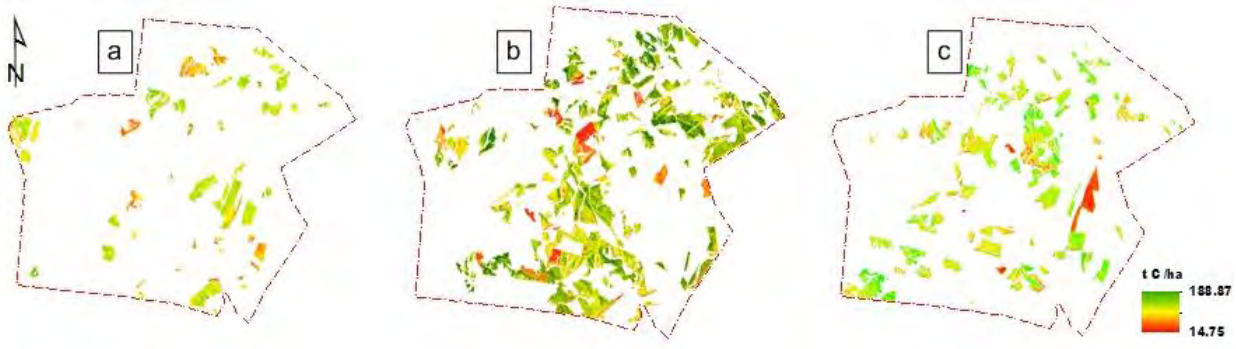


Figure 7.2 Aboveground carbon stocks distribution maps for (a) *E dunii*, (b) *E grandis* and (c) *P taeda* derived using models developed from remotely sensed data.

7.3.2 Spatial distribution of tree structural attributes carbon stocks

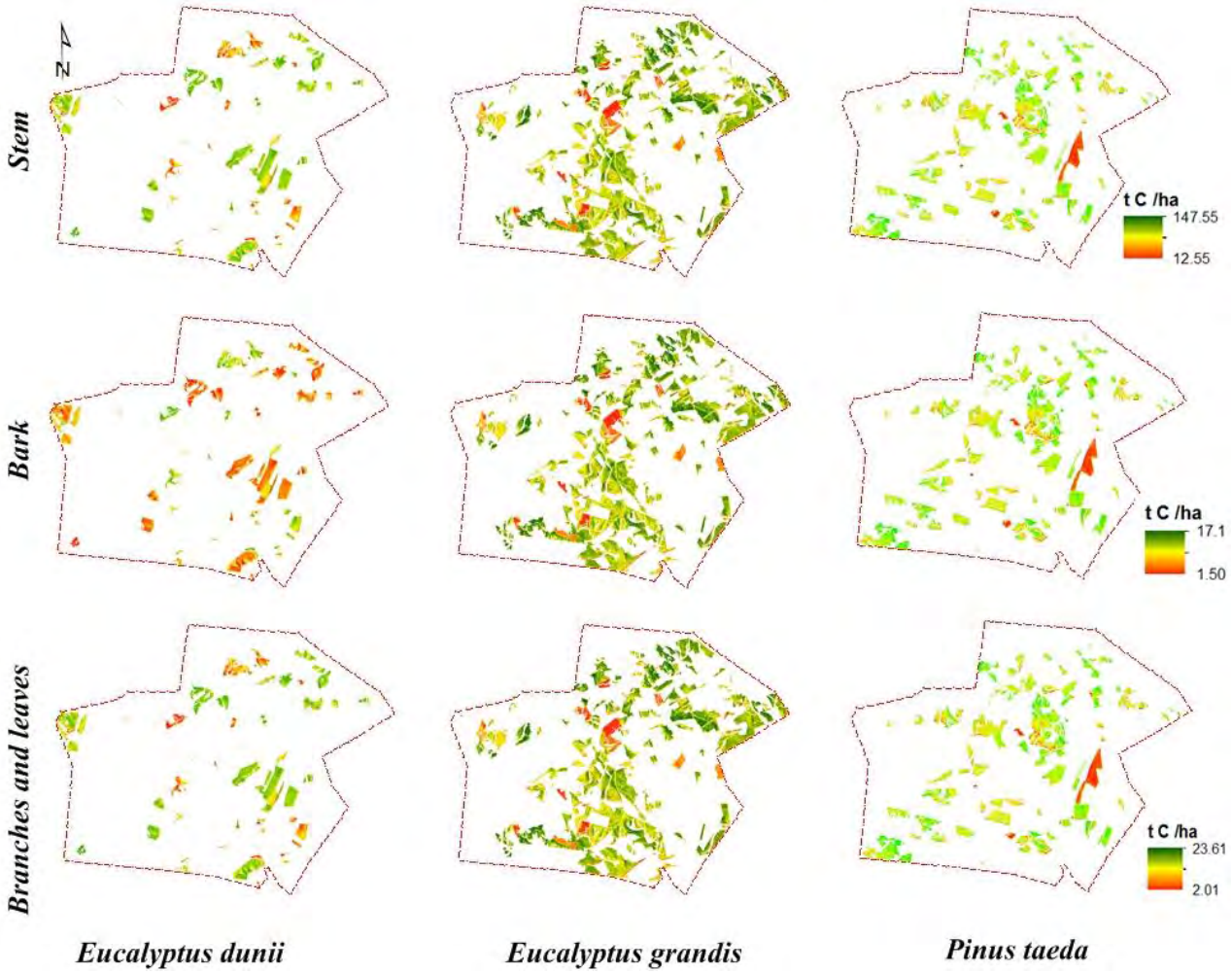


Figure 7.3 Aboveground carbon stock maps for the stem, bark, branches and leaves for the three species derived, using models developed from remotely sensed data

7.3.3 Principal Characteristics of Aboveground Carbon Stock derived using Remotely Sensed Data

Table 7.1 summarises aboveground carbon stock estimates of the three different plantation forest structural components, namely, stem, leaves and branches, as well as the bark of *Eucalyptus* and *Pinus* species, derived using remotely sensed data. It can be observed from the results that different tree structural components have varying aboveground carbon stocks (t C ha^{-1}) allocation patterns, with the stem containing a greater allocation than the bark, branches and leaves. For example, it can be observed from Table 7.1 that *Eucalyptus grandis* and *Pinus taeda* contain the highest carbon stock allocation patterns, when compared to *Eucalyptus dunii,* which had minimum and maximum carbon stocks of 12.55 and 27.36 t C ha^{-1}. It is also important to note that the bark, branches and

leaves of all three tree species had the least aboveground carbon stock allocation patterns, when compared to those of the stems.

Table 7. 1 Remotely sensed principal characteristics of aboveground carbon stock estimates across different plantation forest species

Species	Structral attributes	Min.	Max.	Mean	SD	SE
E. dunii	stem	12.55	27.36	19.16	4.20	0.60
	branch + leaves	2.01	4.38	3.07	0.67	0.10
	Bark	1.51	3.28	2.30	0.50	0.07
	Total	14.75	36.82	23.49	4.61	0.66
E. grandis	stem	50.92	113.81	80.77	16.95	2.10
	branch + leaves	8.15	18.21	12.92	2.71	0.34
	Bark	6.11	13.66	9.69	2.03	0.25
	Total	65.18	145.67	103.39	21.69	2.69
P. taeda	stem	19.05	147.55	111.87	25.52	3.03
	branch + leaves	3.05	23.61	17.90	4.08	0.48
	Bark	2.29	17.71	13.42	3.06	0.36
	Total	24.38	188.87	143.19	32.66	3.88

7.3.4 Remote Sensing of total tree structural Aboveground Carbon Stocks

The estimated plantation forest species aboveground carbon stocks for the three studied tree species and across different tree structural components (*i.e.* stem, bark, branches and leaves) are depicted on Figures 7.2a and 7.2b, respectively. The results show that *Pinus taeda* (138 t C ha^{-1}) and *Eucalyptus grandis* (110 t C ha^{-1}), contain the highest carbon stock estimates, when compared to *Eucalyptus dunii* (about 20 t C ha^{-1}). ANOVA results also portrayed significant differences (α = 0.05) between aboveground carbon stock estimates amongst *Pinus* and *Eucalyptus* tree species structural components. Overall, it can be observed that stems for all the three species under study contain the highest aboveground carbon stocks allocation patterns, when compared to those contained in the barks, branches and leaves (Figure 7.2b). The results further illustrate that there are significant differences (α = 0.05) in aboveground carbon stock allocation patterns across different structural attributes (*i.e.* stems, barks, branches and leaves).

It can also be observed that, for *Eucalyptus grandis* and *Pinus taeda*, branches and leaves yielded almost similar estimates of aboveground carbon stocks (α = 0.05). Comparable results were also observed for aboveground carbon stocks contained in the bark, as shown in Figure 7.2b. Moreover, ANOVA results in Figure 7.2b show that there are significant differences (α = 0.05) in

aboveground carbon stock estimates for barks, branches and leaves across the *Eucalyptus grandis* and *Pinus taeda* species. However, the aboveground carbon stocks estimates derived from the *Eucalyptus dunii* species for the bark, branches and leaves, were approximately 5 t C ha^{-1}. Overall, the estimated aboveground carbon stocks derived across all tree structcural compartments, namely, stems, barks, branches and leaves, were significantly lower, ranging between 5 and 20 t C ha^{-1}.

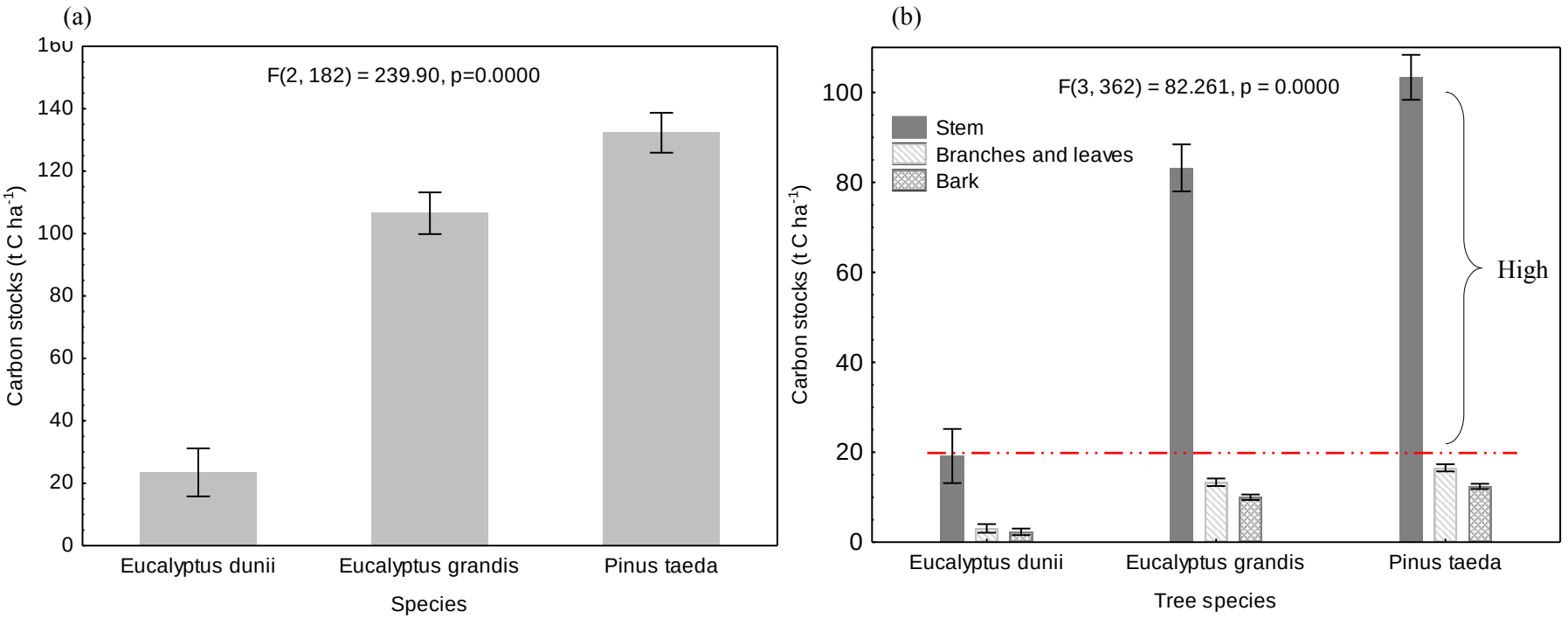


Figure 7.4 Mean total carbon stock estimates for the three studied plantation species. (a) Pooled aboveground carbon stock estimates for *E grandis*, *E dunii* and *P taeda* (b) mean aboveground carbon stock estimates derived from different tree structural compartments (*i.e.* stems, barks, branches and leaves). Bars represent the mean carbon stock estimates for *E dunii*, *E grandis* and *P taeda*. Whiskers represent the 95 % confidence interval . The red line separates lowest and highest aboveground forest carbon stocks.

7.3.5 Evaluating the Structural Variability and Allocation Patterns of estimated total Aboveground Carbon Stocks across Different Species Age groups

Figure 7.5a illustrates the total carbon stock estimates of *Eucalyptus* and *Pinus* plantation forest species across different age groups, derived using remotely sensed data. It can be observed that plantation forest species between 7 and 20 years contained high aboveground carbon stock estimates, relative to those, of 6 years and less. For example, plantation forest species between 7 and 20 years old contain aboveground carbon stocks of between 80 and 100 t C ha^{-1}, whereas for those species from 3 and 6 years old had aboveground carbon stocks ranging around 20 to 25 t C ha^{-1}. It can also be noted that there is less variability in carbon stock allocation estimates for all the three plantation forest species at the age group of 3 to 6 years, when compared to those of between 7 to 20 years old. For example, significant differences ($\alpha = 0.05$) were observed amongst the 7-, 9-, 11-, 18- and 20-year old trees. Overall, the results indicate that the oldest, or more mature plantation forest species (20 years) contain higher carbon stocks than the younger tree species.

Figure 7.5b illustrates the structural variability and allocation patterns of aboveground carbon stock estimates of plantation forest species across different age groups. It can be noted that stem carbon stock estimates vary significantly ($\alpha = 0.05$) across different species age groups, especially between the age groups of 7 and 20 years old. For example, for 7- and 20-year old tree species, the carbon stocks range between 80 and 150 t C ha^{-1}, and between 18 and 25 t C ha^{-1}, for 3- to 6-year old tree species. On the contrary, there is less variability in aboveground carbon stock estimates in branches and leaves for the 3 to 6 age groups. Carbon stock estimates derived from tree branches and leaves for the 3 to 6 tree age groups was almost similar (Figure 3b). Furthermore, for mature plantation forest species (*i.e.* at 7 and 20 years old), a slight variability in carbon stock estimates can be observed for the barks, branches and leaves. However, aboveground carbon stock estimate results for different tree stems portray high variability. Overall, all plantation forest stems contain the highest aboveground carbon stocks than the other tree structural attributes or components (*i.e.* barks, branches and leaves) under study.

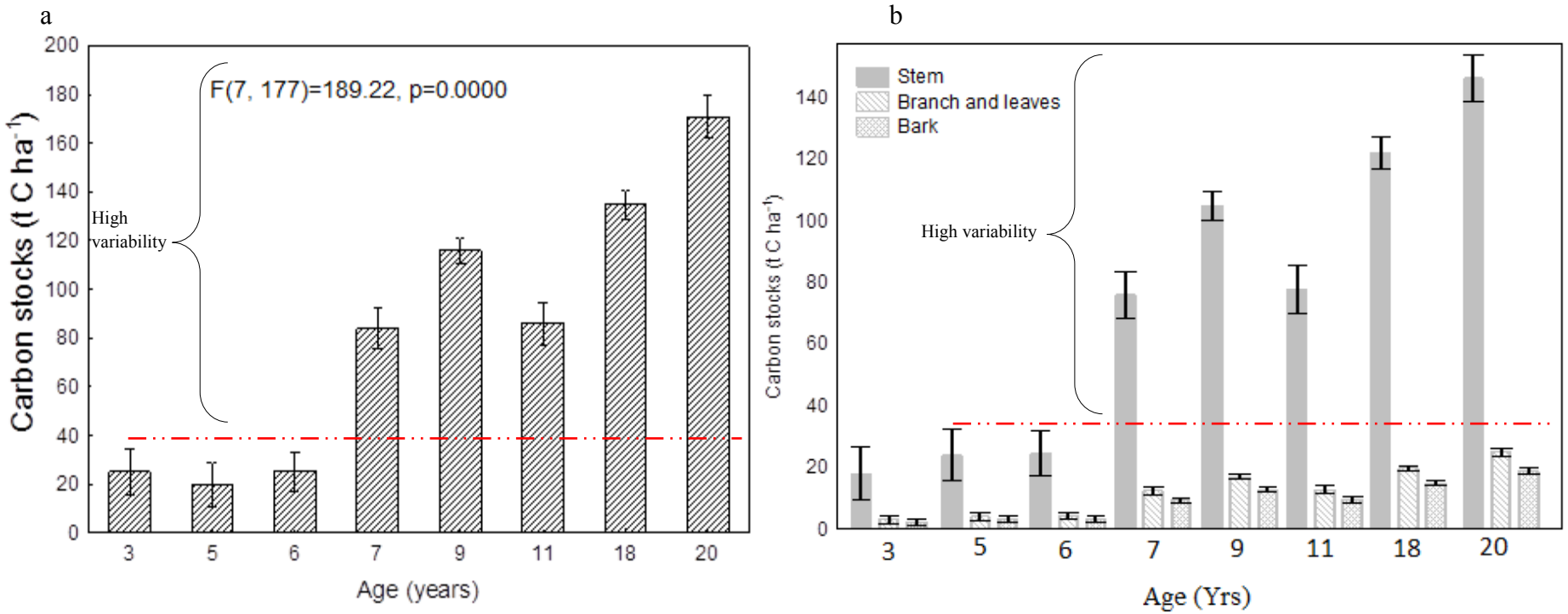
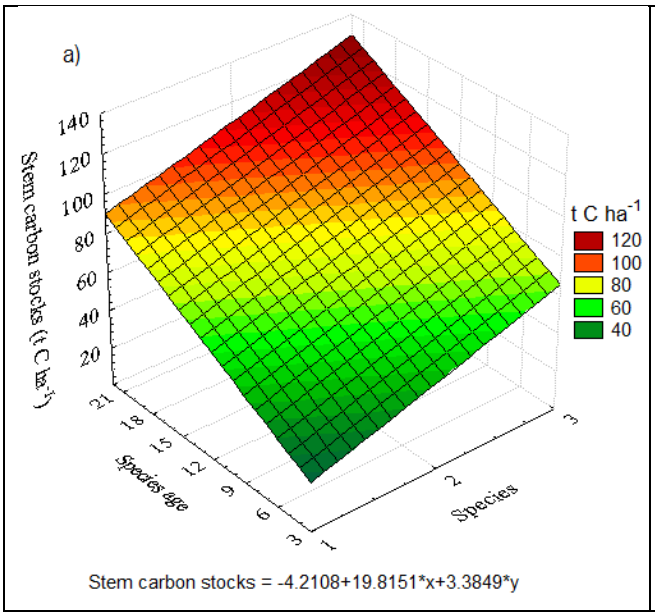
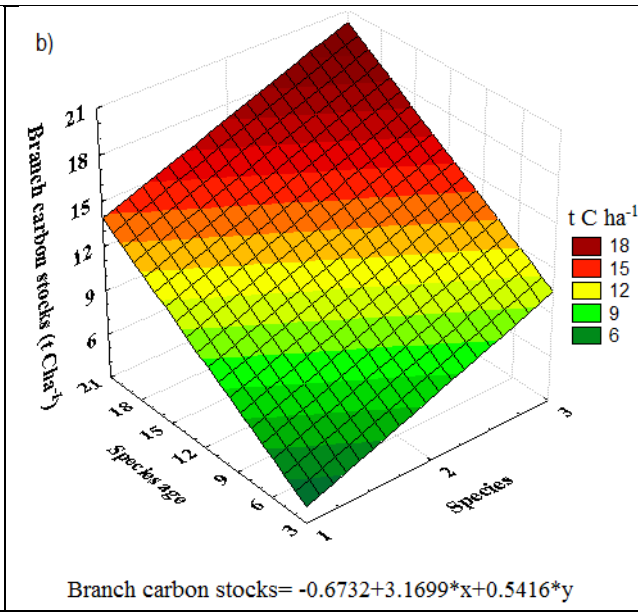


Figure 7.5 Mean total carbon stock estimates derived across different planting ages. (a) total aboveground carbon stock estimates across different plantation forest species age groups (b) is the mean carbon stock estimatess derived from different tree structural attributes or compartments for different plantation forest age groups. Bars represents the mean carbon stock estimates across different plantation forest ages. Whiskers represent the 95 % confidence interval. The red line separates lowest and highest aboveground forest carbon stocks

7.3.6 The Combined Effect of Species Type and Age on Aboveground Carbon Stock Estimates

The three dimensional (3-D) surface model results in Figures 7.6(a) – 7.6(d) show the combined effect of species type and age on aboveground carbon stock estimates. It is interesting to note that all the graphs clearly illustrate that the aboveground carbon stock estimates increase, with an increase in the age of plantation forest species (Figures 7.6a – 7.6d). It can also be observed that stems contain the highest carbon stock allocations of approximately 120 t C ha^{-1} for older species, at the age of eighteen to twenty years, when compared to the other tree structural components. Moreover, *Pinus taeda,* followed by *Eucalyptus grandis,* contain the highest stems carbon stock estimates, relative to the *Eucalyptus dunii* species. A similar pattern can be observed for total aboveground carbon stock estimates for barks, branches and leaves (Figures 7.6a-7.6d). As observed in Sections 3.2 and 3.3, the 3-D surface models further illustrate that stems contain the highest carbon stock estimates (above 120 t C ha^{-1}) than the other structural attributes (*i.e.* barks, branches and leaves) (Figures 7.6a – 7.6d).





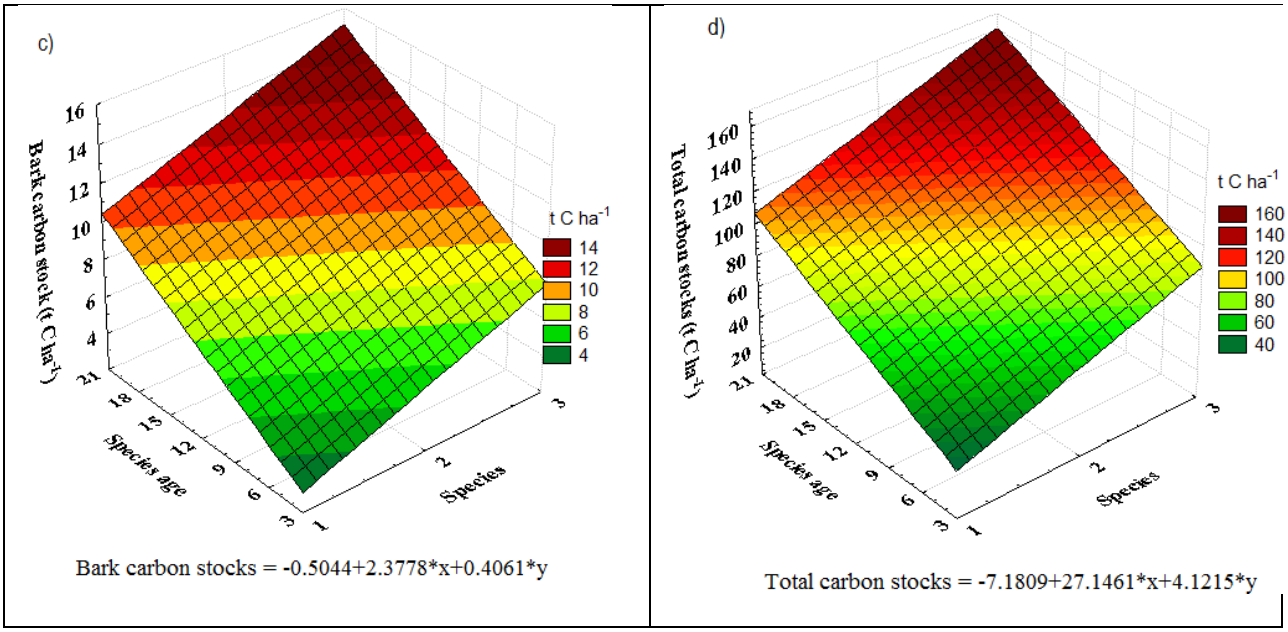


Figure 7.6 The 3-D relationship amongst species type, age and aboveground carbon stock estimates from remotely sensed data for the different tree structural compartments. (a) stems, (b) barks, (c) branches and leaves (d) tree species total aboveground carbon stocks. 1, 2 and 3 from Figure 4 a-d represent *E dunii*, *E grandis* and *P taeda* species, respectively

7.4 Discussion

The essence of this study was to quantify the variability of plantation forest aboveground carbon stocks across different tree species, structural attributes (*i.e.* stems, barks, branches and leaves) and age groups, using remotely sensed data in the sub-tropical coastal lowland region of KwaZulu-Natal, South Africa. Reliable and accurate carbon accounting and inventorying of the global climate change stocks requires a strong understanding of the variability and allocation patterns in aboveground carbon stocks across different plantation forest species, various structural compartments and age groups. The aim of this study was to map and quantify the variability in aboveground carbon stocks across the fast-growing exotic plantation forests, using the cheap and readily-available remotely sensed data. In addition, the study tested whether there were significant differences ($\alpha = 0.05$) in aboveground carbon stock estimates derived from species types, structural components and age, based on models developed, using Landsat 8 OLI remotely sensed data. The variability in plantation forest aboveground carbon stocks allocation patterns is one of the key quantitative characteristics in forest ecosystems.

The findings of this study show that remotely sensed data is capable of providing spatially explicit aboveground carbon stock estimates for different plantation forest species over a large area. For instance, the results showed that the spatial distribution of aboveground carbon stocks vary significantly over large areas and this variability can be detected, using remotely sensed data. The findings are in line with findings from literature that have shown or depicted remote sensing data as the only practical option, to accurately and timely estimate aboveground biomass or carbon stocks across various scales, with affordable effort (Chen, *et al.,* 2015; Kelsey & Neff, 2014; Laurin, *et al.,* 2014; Palace, *et al.,* 2015; Patenaude, *et al.,* 2004; Yang and Guan, 2008). Unlike ground-based aboveground biomass measurements, which require a dense network of inventory plots to reach good accuracies of aboveground carbon stocks over a wide area, which are, in some instances, unfeasible, due to high costs and being laborious, remotely sensed data offers an effective and robust primary data-source necessary for large-scale estimation, with limited, or no cost. For example, Kelsey and Neff (2014) state that the use of remote sensing in estimating AGB has two main advantages and these include: (i) the production of AGB maps, which can accurately illustrate the variations at landscape scale; and (ii) remote sensing allows changes to mapped forest AGB to be easily updated.

The results of this study have shown further that aboveground carbon stocks vary significantly ($\alpha = 0.05$) across different and fast-growing exotic plantation forest species. For instance, results showed that *Eucalyptus grandis* and *Pinus taeda* contain comparatively higher amount of aboveground carbon stocks, than the *Eucalyptus dunii*. Similarly, findings of previous studies demonstrated that these fast-growing plantation forest species contain varying amounts of aboveground carbon stocks (Chen *et al.,* 2015; Gara *et al.,* 2014; Schönau and Boden, 1982). For instance, Chen *et al.,* (2015) found that *Eucalyptus urophylla* accumulated more carbon in plant biomass (about 1.9 and 2.2 times) greater than a *Castanopsis hystrix* plantation. Schönau and Boden (1982a) also indicated that considerable proportional differences in AGB existed across seven *Eucalyptus* species (*i.e. E. deanei, E. globulus ssp. globulus, E. nitens, E smithii* and *E. viminalis)* in the Transvaal Highveld, South Africa, which is consistent with the findings of this work.

The observed variations in aboveground carbon stocks across the fast-growing plantation forest species can be attributed or explained by plant traits that exist among various plantation forest tree species, such as higher maximum net photosynthetic rate (P_{max}) per unit forest land area (P_{max} multiplied by leaf area index). This is supported by Chen *et al.,* (2015), whose study demonstrated

that the P_{max} per unit forest land area of *Eucalyptus urophylla* and *Acacia crassicarpa* was 31.8 and 36.9 $\mu mol \cdot m^{-2}\ s^{-1}$, hence the *Eucalyptus urophylla* species contains higher AGB accumulation characteristics and a higher carbon storage capacity, than the latter.

In addition, the findings of this study have shown that the variability in aboverground carbon stock estimates across different plantation forest species, structural attributes (*i.e.* stems, barks, branches and leaves) and age groups can easily be derived accurately, based on the cheap and readily-available remotely sensed data sets, such as the Landsa-8 OLI. In this study, results indicated that all plantation forest tree stems contained the highest aboveground carbon stocks than the other plant attributes, namely, the barks, branches and leaves. The results are in line with findings from literature which demonstrate that stem biomass comprises the largest fraction of aboveground carbon stocks in mature forests, and it is related to the product of basal area and height (Chen *et al.,* 2015).

Furthermore, the highest amount of aboveground carbon stock estimates contained in plantation forest stems can simply be attributed to the structural size and length of these stems, relative to the other components with tiny structural characteristics (Chen *et al.,* 2015). For instance, the plantation forest species in this study are characterized by tall and bulky stems, with slightly thin layers of the barks, as well as tiny branches and leaves. These structural differences are therefore assumed to explain the observed variability and differences in the allocation patterns of aboveground carbon stock across different plantation forest structral components (*i.e.* stem, bark, branches and leaves). This finding is reinforced by Souza *et al.* (2012), whose work indicated that tree species composition and stem size distribution influence AGB hence, affecting regional and local carbon stocks variability. Similarly, the study by Keller *et al.* (2001) indicated that large trees contain high biomass content, due to the size of their diameter and height. These studies therefore clearly demonstrate that the structure and size of a tree /or plant component plays a crucial role in governing the allocation patterns of aboveground carbon stocks.

7.4.1 Aboveground Carbon Stocks Variability as a Function of Age

The results of this study have also demonstrated significant differences ($\alpha = 0.05$) in aboveground carbon stocks across different species age groups. It was noted that older plantation forests species (between the seven to twenty age groups) yielded the highest carbon stock estimates, when compared to those that were six years old and less. This observation can be attributed to the fact

that mature plantation forest species between seven and twenty years would have accumulated high biomass and have large structural components (*i.e.* stems, barks, branches and leaves), which store more carbon stocks, when compared to those that are six years old and less. These results are consistent with those of Muvengwi et al. (2015) who demonstrated that branch biomass increased with the increase in branch diameter. Muvengwi *et al.* (2015) underscored the fact that branches with a diameter of 55 to 60 mm had a biomass of about 18 kg, whereas those with a diameter of between 5 and 15 mm had a biomass of less than 2 kg. Therefore, this is a clear indication that young trees with small branches contain low carbon stocks, when compared to older species with large branches. Similarly, older plantation forest species have bulky stems, when compared to young stems that are not fully developed. Therefore, results of the present study underscore the fact that older plantation forest species, that are about to be harvested for pulp and paper production, play a critical role in the carbon cycle, when compared to young species with low carbon stocks. In addition, the findings from this study suggest that afforestation with fast-growing tree species, such as *Eucalyptus* and *Pinus* species, can increase the amount of carbon sinks, which are critical for minimizing atmospheric carbon, as well as the related climate variability and its effects.

7.4.2 The Influence of Environmental Factors on the Variability of Aboveground Carbon Stocks

The Pearson's Correlation Analysis results in Table 7.2 show the relationships between aboveground carbon stock estimates of the fast-growing plantation forest tree species and a few selected environmental variables (*i.e.* aspect, slope, total wetness index (TWI), elevation and insolation). A negative correlation can be observed between the environmental variables and aboveground carbon stocks for the three plantation forest species (*Eucalyptus dunii*, *Eucalyptus grandis* and *Pinus taeda*), with a few exceptions. In most instances, TWI, aspect and slope had relatively weak correlations with aboveground carbon stocks, particularly for *Eucalytpus dunii*.

Table 7.2 Correlation coefficients (r) between aboveground carbon stocks for plantation forest species and selected environmental variables

	Eucalyptus dunii	*Eucalyptus grandis*	*Pinus taeda*
Aspect	0.04	0.14	0.11
Slope	0.14	0.28*	0.31*
TWI	-0.04	-0.35*	-0.14
Elevation (m)	-0.38*	-0.40*	-0.39*
Insolation	-0.46*	-0.46*	-0.25*

Significant level: $\alpha = 0.05$, TWI stands for Total wetness index.

The variability in aboveground carbon stock estimates was found to be greatly influenced by certain environmental factors. For instance, the Pearson's Correlation Analysis results indicated that there is a strong negative association between species carbon stock estimates and total wetness index, elevation, insolation and slope. These results imply that, besides the plantation forest stand age, environmental variables also govern plant AGB allocation patterns. The findings of this study show clear patterns in aboveground carbon stocks variations within the study area. For instance, the results of this study have illustrated that environmental factors, such as aspect, slope, insolation and total wetness index, which is a proxy for soil moisture, affect the distribution of aboveground carbon stocks. Elevation and insolation were found to have the highest correlation with aboveground carbon stocks, with aspect to having the least correlation. These results are in line with the findings from literature which show the importance, or the relationship, between elevation and the distribution of biomass across the landscape (de Castilho *et al.,* 2006; Luizão *et al.,* 2004; Sharma *et al.,* 2011; Singh *et al.,* 1994). Furthermore, Singh *et al*. (1994) demonstrated that live tree biomass across different forest types decreased, with an increase in altitude, in the Kumaun region in Uttarakh and Himalaya. This observation is also in agreement with the study by de Castilho *et al.* (2006) who demonstrated that a two-fold variation in the total AGB estimates could be explained by topography. A study by Asner *et al*. (2009) also confirms that an increase in elevation corresponded to a 53–84% decline in AGB. Environmental variables, such as insolation, regulate almost all plant biochemical processes, including atmospheric carbon fixation through photosynthesis (Luo, 2007) hence, this explains the observed variations in aboveground carbon stocks across different plantation forest species, structural attributes and age groups. Overall, the results of this study have demonstrated that, in addition to plant structural traits, and age, environmental variables play a critical role in controlling the spatial distribution of aboveground carbon stocks.

7.5 Conclusion

The essence of this study was to quantify the variability of plantation forest aboveground carbon stocks across different tree species, structural attributes (*i.e.* stems, barks, branches and leaves), as well as age groups, using remotely sensed data in the sub-tropical coastal lowland region of KwaZulu-Natal, South Africa.

The results of this study have shown that:

1. allocations of aboveground carbon stock estimates derived, using remotely sensed data, vary significantly ($\alpha = 0.05$) across different tree age groups, species and structural attributes (*i.e.* stem, bark, branches and leaves),

2. tree stems contained the highest carbon stocks (approx. 120 t C ha^{-1}), relative to, other tree structural attributes such as bark, branches and leaves, which had less than 25 t C ha^{-1},

3. the variability in aboveground carbon stocks across plantation forest stems, bark, branches and leaves can be successfully quantified and mapped, using remotely sensed data,

4. important forest species information (*i.e.* aboveground for different tree structural attributes namely stem, bark, branches and leaves) that used to be difficult to obtain, using field data sets, can now be easily obtained from the cheap and readily-available satellite data.

Overall, the findings of this study demonstrate the importance of the remote sensing technology as an important tool that supports and enhances the quantification of forest carbon stock variations across different tree species, structural components and age groups. This study underscores the utility of remotely sensed data in deriving data that is invaluable for carbon stock accounting.

8. OPTICAL REMOTE SENSING OF ABOVEGROUND FOREST BIOMASS AND CARBON STOCKS IN RESOURCE-CONSTRAINED AFRICAN ENVIRONMENTS: A SYNTHESIS

8.1 Introduction

Forest plantations serve as an important key driver of regional and local climate systems through biosphere-atmospheric interactions. Information on the relative amount of forest AGB and its spatial distribution across the landscape is central in determining their role in the global or regional carbon cycle and climate change (Patenaude *et al.,* 2004; Patenaude *et al.,* 2005a; van der Werf *et al.,* 2009; Woollen *et al.,* 2012a). Thus far, the amount of AGB and carbon stocks stored and sequestered by forest plantations, epitomizes one of the extreme uncertainties in comprehending their contribution to the global carbon pool (Mackey *et al.,* 2013). Therefore, the accurate and precise quantification of this uncertainty requires approaches that have the capability to measure forest carbon dynamics, and to map their spatial and temporal coverage at landscape scale. Recent studies show that earlier carbon accounting efforts repeatedly relied on the estimation of gross emissions, which involved the quantification of forest loss, with limited attention on potential biomass replacement from the regeneration of natural forests, as well as plantation forestry practices. While the AGB of most parts of the world have been systematically inventoried over the years (Houghton *et al.,* 2009), African regions suffer from operational limitations and consequent inadequate data, due to the associated costs. Quantifying and mapping forest AGB and carbon stocks in African environments is therefore essential for understanding forest contribution to the global or regional carbon cycle. However, the use of conventional methods in estimating and mapping AGB and carbon content is challenging, as it is very costly and time consuming and, besides, it lacks the spatial aspect. It is thus imperative to come up with timely techniques that would allow AGB estimation, even in resource-constrained areas. In this regard, remote sensing technologies provide quick and reliable primary data that are required for the extraction and quantification of AGB and carbon stock estimates, at local, regional and global scales, with acceptable accuracies.

Nonetheless, the major limitations associated with the use of remote sensing data sets in estimating AGB and carbon mapping include saturation problems, the occurrence of mixed pixels and the cost of high spatial resolution sensors, especially in resource-constrained areas. Furthemore, alternative sensors, such as the broadband multispectral sensors, are prone to saturation challenges, especially in dense canopy areas, resulting in poor AGB estimates. To overcome this challenge, therefore, there is need to consider the new generation of multispectral sonsors, with a large swath-width, which can be used to develop effective techniques that can

estimate and map forest plantation AGB and carbon stocks, especially in resource-constrained areas across local, regional and global scales. Hence, the objectives of this study were:

1. to predict *Eucalyptus spp.* stand volume, using the stochastic gradient boosting regression ensemble, with multisource data sets at a local scale.
2. to test the utility of high spatial resolution spaceborne multispectral RapidEye Spaceborne sensor and advanced machine-learning algorithms in estimating intra- and inter-species AGB prediction in plantation forests,
3. to evaluate the utility of the medium-spatial resolution Landsat-8 OLI multispectral sensor in quantifying AGB at a regional scale,
4. to investigate the robustness of the newly-launched Landsat-8 OLI push-broom sensor derived texture indices in estimating medium-density plantation forest species AGB, and
5. to quantify the variability in aboveground carbon stocks across plantation forest different structural attributes and age groups, using remotely sensed models.

8.2 Predicting Eucalyptus Species Stand Volume using Multisource data sets

The application of broadband multispectral sensors in estimating tree structural attributes, such as tree volume and AGB, has faced great challenges due to a failure to improve the error of estimation, especially in high density canopies, mainly due to saturation problems and the prevalence of mixed pixels. Although the use of high spatial resolution airborne sensors, such as hyperspectral and lidar can help solve these challenges (Koch, 2010; Patenaude *et al.,* 2004; Patenaude *et al.,* 2005b), the cost and availability of such data sets in resource-constrained areas remains the major challenge. The availability of cheap, readily-available and quick operational remote sensing techniques therefore remains the most possible solution necessary for improving the prediction of critical tree structural attributes in these areas. In this thesis, substantial evidence on the potential and strength of integrating broadband multispectral sensors with ancillary data sets, to accurately predict *Eucalyptus* stand volume, is demonstrated (Chapter Three). To accomplish the above goal, a large number of predictor variables, comprising SPOT-5 bands, vegetation indices, rainfall derived variables and stand age, and multisource data sets (all variables investigated) were tested independently, using three predictive models, namely, stochastic gradient boosting, random forest and stepwise multiple regression.

The results demonstrated that the integration of broadband multispectral sensors with ancillary data sets improves the predictive accuracy of tree-structral attributes, such as *Eucalyptus* stand volume, diameter-at-breast height and tree height, when compared to the use of the multispectral or ancillary data sets in isolation. For instance, when the SPOT-5 bands were used as an independent data set in predicting stand volume, the three predictive models yielded weaker results (*i.e.* SGB model produced an R^2 value of 0.41 and an RMSE value of 54.40 m^3ha^{-1}) (Table 3.3). A similar performance was also observed when predicting other tree structural attributes like tree height and diameter-at-breast height. The study, however, illustrated that the use of multisource data sets, based on all three modelling techniques considered in this study, yielded the best results. For example, an R^2 value of 0.78 and an RMSE value of 33.16 m^3ha^{-1} was obtained using the SGB, an R^2 value of 0.76 and an RMSE value of 37.26 m^3ha^{-1} for the RF, while the SMLR produced an R^2 value of 0.65 and an RMSE value of 42.50 m^3ha^{-1} (Table 3.3). The results also showed that the SGB model outperformed the other two predictive models. The study further tested the potential of identifying most important variables and whether the use of the model selected variables could improve the prediction accuracies. In assessing the relative importance of variables, the SWIR, minR, EnumAge and covR variables were found to be the most optimal variables for predicting mean DBH, mean tree height and stand volume. The use of optimal selected variables produced an R^2 value of 0.81 and an RMSE value of 1.21 cm for mean DBH, an R^2 of 0.86 and RMSE of 1.39 m for tree height and an R^2 value of 0.83 and an RMSE value of 29.58 m^3ha^{-1} for stand volume (Table 3.5).

8.3 Intra- and inter-species biomass prediction using high resolution data sets

In order to better understand the contribution of plantation forest species in the global carbon cycle, the possibility of applying the high resolution RapidEye imagery and stochastic gradient boosting and random forest models was investigated in estimating biomass (Chapter Five). More precisely, the study examined the performance of two machine learning algorithms (*i.e.* SGB and RF) and whether algorithm variable selection could enhance the AGB estimates, using spectral information and vegetation indices derived from the RapidEye image. The use of these spectral information and vegetation indices was based on their successful application in AGB, particularly in natural ecosystems and tree species mapping and discrimination (Adam *et al.,* 2014; Adelabu *et al.,* 2013a; 2013b; Eckert, 2012a; Mutanga *et al.,* 2012).

The results have shown that both the SGB and RF regression models performed well in estimating AGB, although the SGB model outperformed the RF model. Specifically, a comparative assessment of the two predictive models demonstrated that the SGB model produced better predictive results for the intra-species data set, with an R^2 of 0.75 and an RMSE of 18.40 t ha^{-1} (10.80%) for EG R^2 of 0.77 and an RMSE of 19.43 t ha^{-1} (19.18%) for PT (Figure 4.3). Furthermore, the SGB algorithm yielded good results in predicting interspecies biomass, producing an R^2 of 0.58 and an RMSE of 46.51 t ha^{-1}, 33.25% of the mean compared to RF, which had an R^2 of 0.33 and an RMSE of 64.27 t ha^{-1}, 45.94% of the mean (Figure 4.4). The SGB and RF variable importance approaches in determining the relevance of all RapidEye derived input variables was further explored. Both models successfully managed to select the most suitable set of variables (EG: n = 4, ED: n = 7, PT: n = 6 and all species data combined: n = 19) for predicting both intra- and inter-species AGB (Figure 4.5).

8.4 Plantation Forest Aboveground Biomass Quantification using Landsat-8 OLI Sensor

The potential of using newly-launched broadband multispectral sensor with unique radiometric, spatial and spectral characteristics in estimating and mapping AGB for three plantation forest species, namely, *Eucalyptus dunii* (ED), *Eucalyptus grandis* (EG) and *Pinus taeda* (PT), was examined. The newly-launched Landsat-8 OLI, with changes in the sensor design enhances the signal to noise ratios, which is almost twice, as good as Landsat 7 Enhanced Thematic Mapper plus (ETM^+). Landsat 8 also constitute a refined near-infrared spectral range, improved radiometric resolution (8 to 12 bits), wide swath width and improved sensor dwell-time in sampling radiation, at each ground pixel (El-Askary *et al.,* 2014; Pahlevan and Schott, 2013). These sensor properties are perceived to provide the most needed solution for AGB mapping and monitoring at regional scales in data-scarce areas. Moreover, these improvements prevent water vapour absorption effect at 0.825 μm and improve radiometric resolution, which is known to enhance the spectral precision, as well as eliminating sensor spectral saturation problems, a common phenomenon to prior Landsat products (Irons *et al.,* 2012). In that regard, the improvements are perceived to be critical in enabling the characterization of vegetation biochemical and biophysical properties. Despite these promising capabilities associated with the newly-launched Landsat-8 OLI, no study, to the best of our knowledge, has tested its utility in estimating and mapping AGB at both local and regional scale. Therefore, to test the utility of this sensor, stochastic gradient boosting and random forest were applied, based on three different

variable sets which are: (i) spectral bands, (ii) vegetation indices, and (iii) all data sets combined. The results obtained were then compared to those derived from Landsat 7 ETM+ (Chapter Four). Aboveground biomass results for ED, EG and PT were evaluated in terms of the coefficient of determination (R^2), RMSE, RMSE%, percent bias and mean of the absolute forecast error (MAE). The results showed that Landsat 8 OLI spectral data set provides better biomass estimates, when compared to Landsat 7 ETM+ spectral data sets. For example, the results indicated that Landsat 8 OLI spectral information, utilizing the SGB algorithm, produced an R^2 value of 0.47, the lowest RMSE (16.35 t ha^{-1}; 30.93% of the mean), the lowest percentage bias of -1.40 and the lowest MAE (14.23) for the *E. dunii* species. A similar performance was observed for EG and PT, using Landsat-8 OLI derived spectral information. Similarly, the use of Landsat-8 OLI derived vegetation indices exhibited the best estimation accuracies, when compared to the estimates derived using Landsat 7 ETM^+. For instance, Landsat ETM^+ derived vegetation indices applying the SGB algorithm produced the RMSE of 18.66 t ha^{-1} (35.30% of the mean), a bias % of -4.30 and the MAE value of 14.65 for *E. dunii* species, whereas for *E. grandis*, an RMSE of 26.54 t ha^{-1}, 15.60% of the mean), a percentage bias of -2.26 and the MAE of 22.92 were obtained. The use of Landsat 8 OLI and 7 ETM^+ extracted spectral information and vegetation indices further improved AGB estimates for all three species, with the former producing optimal accuracies. Moreover, when the same variable groupings were used to estimate AGB for all three species data set combined, the results demonstrated that both Landsat-8 OLI and ETM+ sensors estimated combined plantation forest AGB with weaker accuracies. Overall, the combined Landsat-8 OLI sensor spectral information and vegetation indices produced better results, when compared to Landsat 7 ETM^+. Furthermore, when variable selection was implemented, based on the best performing algorithm in estimating AGB for individual species and combined data set, the SWIR bands, NDVIc, NDVI, NDII, NIR, SAVI, red band, MSI, and green band were selected. The use of these variables significantly reduced the error of estimation for both remote sensing data sets, with Landsat-8 OLI sensor producing the least errors for all three species. For instance, for PT, an R^2 of 0.69, RMSE of 21.65 t ha^{-1} (10.50% of the mean), using Landsat-8 OLI, and an R^2 of 0.65, an RMSE of 22.33 t ha^{-1}, (10.83% of the mean), using Landsat ETM+ selected variables, were observed (Figure 5.2). In addition, a comparative analysis of the AGB estimates for combined species from both sensors showed that Landsat-8 OLI sensor yielded better, but weaker estimates, producing an R^2 of 0.42 and an RMSE of 55.32 t ha^{-1} (41.13%), when compared to the Landsat 7 ETM^+ sensor, which yielded an R^2 of 0.32 and an RMSE of 64.26 t ha^{-1} (43.86%) (Figure 5.3 and Table 5.7).

8.5 Enhancing aboveground biomass estimation using Landsat-8 OLI texture metrics

The use of vegetation indices computed from multispectral sensor tends to suffer from saturation challenges in areas associated with high canopy cover and the presence of mixed pixels (Gara *et al.,* 2014b; Ingram *et al.,* 2005; Lu, 2006; Mutanga *et al.,* 2012; Mutanga and Skidmore, 2004; Nichol and Sarker, 2011). Although high resolution airborne data sets can be used to address the aforementioned problems, their cost remains one of the limiting factors, especially in resource-constrained areas (Eckert, 2012a; Koch, 2010). Hence, there is a need for techniques that could help improve AGB and carbon stock estimates from the readily-available and cheap multispectral sensors, such as the Landsat products. In this study, the utility of the newly-launched Landsat-8 derived texture metrics and texture ratios is tested, using two machine learning algorithms (*i.e.* SGB and RF) in improving AGB and carbon stock estimation in high density canopies (Chapter Four).

The results demonstrated that the mean, entropy, correlation, dissimilarity, homogeneity and variance texture parameters obtained, using a 3 x 3 window size and an offset of [0,1], are the best parameters for the accurate estimation of AGB. The results further revealed that raw Landsat-8 OLI spectral information produced weaker AGB estimates across all tree species (*i.e.* ED, EG, PT and all species data), when compared to those derived using raw texture measures derived from Landsat-8 OLI (Table 6.3), based on the stochastic gradient boosting algorithm. Furthermore, it was shown that texture-based models, derived, using the stochastic gradient boosting algorithm, enhanced accuracy in AGB estimation, relative to the use of spectral bands and spectral vegetation indices (Tables 6.3 and 6.4). While the results derived using both raw spectral band ratios and texture ratios appear to be promising, raw spectral band ratios, based on AGB estimates, generally yielded low accuracies. On the other hand, the use of Landsat-8 OLI sensor texture ratios derived from different band combinations: band two/band three, band three/band five, and band four/ band five produced stronger correlations between predicted and observed biomass, compared to other methods investigated, such as the use of raw band ratios *etc.* (Tables 6.3, 6.4 and Figure 6.4). Above all, the use of combined Landsat-8 OLI texture ratios in estimating AGB yielded the best performance, with higher R^2 values of 0.76 (RMSE = 9.55 t ha^{-1} and CV-RMSE of 0.18), 0.74 (RMSE = 12.81 t ha^{-1} and CV-RMSE of 0.08), 0.74 (RMSE = 12.67 t ha^{-1} and CV-RMSE of 0.06) and 0.53 (RMSE = 20.15 t ha^{-1} and CV-RMSE of 0.15) overall for ED, EG, PT and all species data investigated, respectively (Table 6.4).

8.6 Quantifying the Variability and Allocation Patterns of Aboveground Carbon Stocks across Plantation Forest types, Structural Attributes and Age

Information on the variability of forest carbon stocks across different tree structural attributes and groups is critical for the accurate understanding of the contribution of forest ecosystems to the global carbon cycle. Thus far, spatial knowledge on forest structural carbon allocation patterns and aboveground carbon stocks remains scarce, particularly in both managed and unmanaged plantation forest ecosystems. Notwithstanding the vivid attention on global terrestrial carbon accounting, literature shows that the bulk of AGB quantification and carbon-related studies, using both remotely sensed and traditional approaches have focussed on total aboveground component, specifically in sub-Saharan Africa (Christie and Scholes, 1995; Eckert, 2012b; le Maire *et al.,* 2011; Schönau and Boden, 1982a). However, to reduce the great uncertainty in the current carbon balance in Africa (Laurin *et al.,* 2014; Wolf *et al.,* 2011), the accurate and timely quantification of the spatial variability in aboveground carbon stocks across different plantation forest species, various tree structural attributes and age groups are fundamental. Based on these premises, the challenge was therefore to test the potential of quantifying plantation forest aboveground carbon stocks across different tree species, structural attributes (*i.e.* stems, barks, branches and leaves) and age groups, using Landsat-8 OLI remotely sensed data.

The results demonstrated that plantation forest aboveground carbon stocks vary significantly (α = 0.05) across different tree species. A comparison of carbon stocks amongst the three plantation species (*i.e. Eucalyptus dunii*, *Eucalyptus grandis* and *Pinus taeda*) shows that *Pinus taeda* contained the highest amount of aboveground carbon stocks (Figures 7.2 and 7.4). In addition, *Eucalyptus grandis* and *Pinus taeda* branches and leaves yielded almost similar estimates of aboveground carbon stocks (Figures 7.2 and 7.4). The results further illustrated that plantation forest stems contain the highest carbon stocks, when compared to other forest structural attributes, such as barks, branches and leaves. The observed findings were further confirmed by ANOVA results, which showed significant differences (α = 0.05) between aboveground carbon stock estimates amongst *Pinus* and *Eucalyptus* tree species structural components (*i.e.* stems, barks, branches and leaves). Overall, estimated aboveground carbon stocks derived across all tree structural compartments, namely, stems, barks, branches and leaves, were significantly lower, ranging between 5 and 20 t C ha^{-1} (Figures 7.3 and 7.4).

It was further established from the results that mature plantation forest species (normally between the ages of seven and twenty years) contain the highest aboveground carbon stock estimates relative to younger species (*i.e.* six years and less). Seven- and twenty- year old plantation forest species contained between 80 and 100 t C ha^{-1} of aboveground carbon stocks, whilst younger species had between 20 and 25 t C ha^{-1} (Figure 7.5). However, amongst younger species, there was less variability in carbon stock estimates across all three plantation forest species, when compared to mature species (Figure 7.5).

8.7 Conclusion

The main essence of this work was to assess the performance and strength of the new generation multispectral sensors in estimating and mapping plantation forest species AGB and carbon stocks. The results of this study have demonstrated that the use of emerging new generation multispectral sensors and image processing techniques, integrated with non-parametric algorithms, can accurately improve AGB and carbon stock estimations in high density canopies, particularly in resource-scarce areas, a previously challenging task with old traditional multispectral sensors.

Based on the findings from different objectives in this study, the following were concluded:

1. The integration of SPOT-5 derived data (*i.e.* spectral bands and vegetation indices) with ancillary stand data (*i.e.* rainfall metrics and stand age) can be used to improve the prediction of *Eucalyptus* tree mean DBH, mean height and stand volume, based on machine learning algorithms, such as SGB and RF. The findings demonstrate that the new generation multispectral sensors remain a potential primary data source for estimating important tree structural attributes, especially when integrated with ancillary data sets. Moreover, the results have demostrated the strength associated with machine learning algorithms in predicting different tree-structural attributes,
2. Intra- and inter-species AGB could be predicted accurately, using spectral information and spectral vegetation indices derived from the high spatial resolution multispectral RapidEye sensor,
3. Spectral information and vegetation indices derived from the newly-launched medium spectral resolution Landsat-8 OLI multispectral sensor, with a wide swath width of 185 km coupled with improved signal-to-noise ratios, offers an invaluable primary data source required for accurate AGB estimation, especially in data-scarce environments, when compared to the whisky-broom Landsat ETM+ sensor,

4. Similarly, the results of the study further demonstrated the importance of combining spectral information and vegetation indices in estimating AGB in high canopy areas. These findings have prompted further tests on the utility of free and readily-available Landsat-8 OLI derived texture metrics in enhancing plantation forest AGB quantification,
5. The use of newly-launched Landsat-8 OLI derived texture metrics have the potential for enhancing AGB estimates, particularly in resource-constrained areas. The sensor can therefore be viewed as the primary data source for estimating AGB in data-scarce regions, as it produces results almost comparable to those obtained using high resolution sensors like the RapidEye Spaceborne sensor, and
6. Overall, the findings of the study demonstrated the importance of remotely sensed data as an important tool that supports and enhances the quantification of forest carbon stockvariations across different tree species, structural components and age groups. This study underscores the utility of remotely sensed data providing invaluable data set for regional scale AGB and carbon stocks accounting.

8.8 The Future

Multisource and new generation mulitspectral sensors present an effective and robust primary data source required for AGB and carbon stocks mapping in resource-constrained areas. The findings of this study provide the necessary insight and motivation to the remote sensing community, ecologists and environmentalists, to shift towards embracing cheap and readily-available remote sensing sensors necessary for reliable and accurate forest AGB and carbon stocks monitoring in data-scarce environments, where the use of high resolution airborne sensors still remains a challenge, due to the associated costs. Furthermore, the results of this study have demonstrated the contribution of plantation forest species in the global carbon cycle and hence could lay the necessary basis for potential conservation strategies that will facilitate the effective and sustainable utilization of forest resources. This study therefore suggests the following recommendations for future research:

1. the present study demonstrated the strength and performance of integrating multispectral data with ancillary data sets, as well as the use of new generation multisctral sensors in estimating volume, AGB and carbon stocks in plantation forests. Therefore, future experiments on the utility of these data sets need to be tested in naturally forested areas,

particularly in sub-Saharan Africa, where severe harvesting (for firewood and farmlands) is experienced,

2. emerging new generation multispectral sensors are associated with a limited number of unique-band settings that do not require complex processing approaches and are reasonably available at no, or low, cost. It is therefore against this background that the utility of other emerging multispectral sensors, such as Sentinel 2, with improved spatial resolution and swath width, need to be tested for large-scale AGB and carbon stock estimation across African environments,
3. although the present study demonstrated accurate AGB and carbon stocks estimation from the use of new generation remote sensing sensors, it will be ideal for future research to further examine and compare the performance of these sensors against those of high resolution airborne hyperspectral and lidar data sets in estimating and mapping AGB, especially in high density canopy areas.

References

Adam, E., Mutanga, O, Abdel-Rahman, E.M., Ismail, R., 2014. Estimating standing biomass in papyrus (Cyperus papyrus L.) swamp: exploratory of *in situ* hyperspectral indices and random forest regression. *International Journal of Remote Sensing* 35, 693-714.

Adam, E.M., Mutanga, O., Rugege, D., Ismail, R., 2011. Discriminating the papyrus vegetation (*Cyperus papyrus L.*) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing,* 33, 552-569.

Adelabu, S., Dube, T., 2014a. Employing ground and satellite-based QuickBird data and random forest to discriminate five tree species in a southern African woodland. *Geocarto International*, 1-15.

Adelabu, S., Mutanga, O., Adam, E., 2014. Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing,* 95, 34-41.

Adelabu, S., Mutanga, O., Adam, E., Cho, M.A., 2013a. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. *Journal of Applied Remote Sensing*, 7, 073480.

Adelabu, S., Mutanga, O., Adam, E., Sebego, R., 2013b. Spectral discrimination of insect defoliation levels in Mopane woodland using hyperspectral data. *Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 1-11.

Adjorlolo, C., 2013. *Remote sensing of the distribution and quality of subtropical C3 and C4 grasses.* School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, South Africa, Pietermaritzburg.

Amuzu-Sefordzi, B., Huang, J., Sowa, D and Baddoo, T. D. 2015. Biomass-derived hydrogen energy potential in Africa. *Environmental Progress and Sustainable Energy*.

Anaya, J., Chuvieco, E., Palacios-Orueta, A., 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management,* 257, 1237-1246.

Asner, G., Flint Hughes, R., Varga, T., Knapp, D., Kennedy-Bowdoin, T., 2009. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems* 12, 261-278.

Atta-Boateng, J., William, M.J., 1998. A method for classifying commercial tree species of an uneven-aged mixed species tropical forest for growth and yield model construction. *Forest Ecology and Management,* 104, 89-99.

Baccini, A., Friedl, M.A., Woodcock, C.E., Warbington, R., 2004. Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, 31, L10501.

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.

Baeza-Yates, R., Ribeiro-Neto, B., 1999. Modern information retrieval. ACM Press, New York.

Baishya, R., Barik, S.K., Upadhaya, K., 2009. Distribution pattern of aboveground biomass in natural and plantation forests of humid tropics in northeast India. *Tropical Ecology*, 50, 295.

Bannari, A., Morin, D., Bonn, F., Huete, A., 1995. A review of vegetation indices. *Remote sensing reviews* 13, 95-120.

Barbosa, J., Broadbent, E., Bitencourt, M., 2014. Remote sensing of aboveground biomass in tropical secondary forests: A Review. *International Journal of Forestry Research*.

Bastin, J.-F., Barbier, N., Couteron, P., Adams, B., Shapiro, A., Bogaert, J., De Cannière, C., 2014. Aboveground biomass mapping of African forest mosaics using canopy texture analysis: towards a regional approach. *Ecological Applications*.

Basuki, T.M., Skidmore, A.K., van Laake, P.E., van Duren, I., Hussin, Y.A., 2011. The potential of spectral mixture analysis to improve the estimation accuracy of tropical forest biomass. *Geocarto International,* 27, 329-345.

Belsley, D.A., 1991. *Conditioning diagnostics*. Wiley Online Library. London, UK.
Blackburn, G.A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing*, 19, 657-675.
Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., 2014. Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing,* 87, 180-191.
Boahene, K. 1998. The challenge of deforestation in tropical Africa: reflections on its principal causes, consequences and solutions. *Land Degradation and Development*, 9: 247-258.
Boyd, D.S., Foody, G.M., Curran, P., 1998. Estimating biophysical properties of tropical forests using radiation reflected in middle infrared wavelengths (3.0 - 5.0 pm). *IEEE.*
Boyd, D.S., Foody, G.M., Curran, P., 1999. The relationship between the biomass of Cameroonian tropical forests and radiation reflected in middle infrared wavelengths (3.0-5.0 mu m). *International Journal of Remote Sensing*, 20, 1017-1023.
Bredenkamp, B., 2000. *Volume and Mass of Logs and Standing Trees*. Southern African Institute of Forestry, Menlo Park.
Breiman, L., 2001. *Random Forests*. Machine Learning 45, 5-32.
Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment,* 76, 156-172.
Brown, L., Chen, J.M., Leblanc, S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote sensing of environment*, 71, 16-25.
Buston, P.M., Elith, J., 2011. Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *Journal of Animal Ecology*, 80, 528-538.
Canadell, J. G., C. Le Que´re´, M. R. Raupacha, C. B. Fielde, E. T. Buitenhuisc, P. Ciaisf, T. J. Conwayg, N. P. Gillettc, R. A. Houghtonh, and G. Marlandi. 2007. "*Contributions to Accelerating Atmospheric CO2 Growth from Economic Activity, Carbon Intensity, and Efficiency of Natural Sinks."* In Proceedings of the National Academy of Sciences, Vol. 104, edited by W. C. Clark, 18866–18870. Cambridge, MA: Harvard University.
Canavesi, V., Ponzoni, F.J., Valeriano, M.M., 2010. Stand volumes estimate Eucalyptus spp. plantations in forests using hyperspectral and topographic data. *Revista Árvore*, 34, 539-549.
Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62, 241-252.
Carreiras, J., Vasconcelos, M., Lucas, R., 2012a. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sensing of Environment*, 121, 426-442.
Carreiras, J.M.B., Maria J. Vasconcelos, Lucas, R.M., 2012b. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sensing of Environment,* 121, 426-442.
Carreiras, J.M.B., Melo, J.B., Vasconcelos, M.J., 2013. Estimating the Above-Ground Biomass in Miombo Savanna Woodlands (Mozambique, East Africa) Using L-Band Synthetic Aperture Radar Data. *Remote Sensing,* 5, 1524-1548.
Champion, I., Da Costa, J.P., Godineau, A., Villard, L., Dubois-Fernandez, P., Le Toan, T., 2013. Canopy structure effect on SAR image texture versus forest biomass relationships. *EARSeL eProceedings* 12, 25-32.
Champion, I., Dubois-Fernandez, P., Guyon, D., Cottrel, M., 2008. Radar image texture as a function of forest stand age. *International Journal of Remote Sensing,* 29, 1795-1800.

Chan, J.C.-W., Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment,* 112, 2999-3011.

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113, 893-903.

Chaves, M., Flexas, J., Pinheiro, C., 2009. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of botany,* 103, 551-560.

Chen, Y., Liu, Z., Rao, X., Wang, X., Liang, C., Lin, Y., Zhou, L., Cai, X.-a., Fu, S., 2015. Carbon Storage and Allocation Pattern in Plant Biomass among Different Forest Plantation Stands in Guangdong, China. *Forests,* 6, 794-808.

Cheng, P., Chaapel, C., 2008. Increased image collection opportunities, Digital Globe's worldview-1 satellite, *Geoformatics online magazine*.

Chinembiri, T.S., Bronsveld, M.C., Rossiter, D.G., Dube, T., 2013a. The Precision of C Stock Estimation in the Ludhikola Watershed Using Model-Based and Design-Based Approaches. *Natural Resources Research*, 1-13.

Chinembiri, T.S., Bronsveld, M.C., Rossiter, D.G., Dube, T., 2013b. The Precision of C Stock Estimation in the Ludhikola Watershed Using Model-Based and Design-Based Approaches. *Nat Resour Res,* 22, 297-309.

Cho, M.A., Skidmore, A., Corsi, F., van Wieren, S.E., Sobhan, I., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, 9, 414-424.

Choudhury, B.J., 1987. Relationships between vegetation indices, radiation absorption, and net photosynthesis evaluated by a sensitivity analysis. *Remote Sensing of Environment* 22, 209-233.

Christie, S.I., Scholes, R.J., 1995. Carbon storage in eucalyptus and pine plantations in South Africa. *Environmental Monitoring and Assessment,* 38, 231-241.

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

Clevers, J., Van Der Heijden, G., Verzakov, S., Schaepman, M., 2007. Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering and Remote Sensing*, 73, 1141.

Cohen, W.B., Maiersperger, T.K., Gower, S.T., Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment,* 84, 561-571.

Colgan, M.S., Asner, G.P., Swemmer, T., 2012. Harvesting tree biomass at the stand level to assess the accuracy of field and airborne biomass estimation in savannas. *Ecological Applications,* 23, 1170-1184.

Colgan, S., ASNER, G., Swemmer, T., 2013. Harvesting tree biomass at the stand level to assess the accuracy of field and airborne biomass estimation in savannas. *Ecological applications,* 23, 1170-1184.

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

Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., Brewer, C.K., 2012. Modeling Percent Tree Canopy Cover: A Pilot Study. *Photogrammetric Engineering and Remote Sensing,* 78, 715-727.

Crippen, R.E., 1990. Calculating the vegetation index faster. *Remote Sensing of Environment,* 34, 71-73.

Cryus, S., Tanja, K., 2004. Biomass estimation using Landsat-TM and -ETM+. Towards a regional model for Southern Africa? *GeoJournal,* 59, 177-187.

CSIR, 2000. *National Land-cover database 2000*. CSIR and Agricultural Research Council, Pretoria.

Cutler, M., Boyd, D., Foody, G., Vetrivel, A., 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing,* 70, 66-77.

DAFF, 2008. *Report on Commercial Timber Resources and Primary Round Wood Processing in South Africa*.

Dalu, T., Dube, T., Froneman, P.W., Sachikonye, M.T., Clegg, B.W., Nhiwatiwa, T., 2015. An assessment of chlorophyll-a concentration spatio-temporal variation using Landsat satellite data, in a small tropical reservoir. *Geocarto International*, 1-14.

Darvishzadeh, R., MATKAN, A., Eskandari, N., 2011. Evaluation of ALOS-AVNIR2 Spectral Indices for Prediction of Rice Biomass. *Journal Of Geographical Landscape*.

De'ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88, 243-251.

de Castilho, C.V., Magnusson, W.E., de Araújo, R.N.O., Luizao, R.C., Luizao, F.J., Lima, A.P., Higuchi, N., 2006. Variation in aboveground tree live biomass in a central Amazonian Forest: Effects of soil and topography. *Forest ecology and management* 234, 85-96.

de Coning, E., 2013. Optimizing Satellite-Based Precipitation Estimation for Nowcasting of Rainfall and Flash Flood Events over the South African Domain. *Remote Sensing* 5, 5702-5724.

de Coning, E., Poolman, E.R., 2010. South African Weather Service operational satellite based precipitation estimation technique: applications and improvements. Hydrol. *Earth Syst. Sci. Discuss*. 7, 8837-8871.

De Jong, S.M., Pebesma, E.J., Lacaze, B., 2003. Aboveground biomass assessment of Mediterranean forests using airborne imaging spectrometry: the DAIS Peyne experiment. *International Journal of Remote Sensing* 24, 1505-1520.

de Sousa, C.H.R., Souza, C.G., Zanella, L., de Carvalho, L.M.T., 2012. *Analysis of Rapideye's red edge band for image segmentation and classification*, GEOBIA, Rio de Janeiro - Brazil, p. 518.

Dekker, R.J., 2003. Texture analysis and classification of ERS SAR images for map updating of urban areas in The Netherlands. *Geoscience and Remote Sensing,* IEEE Transactions on 41, 1950-1958.

Denman, K.L., 2007. IPCC *Climate Change 2007*, in: S, S. (Ed.), The Physical Science Basis. Cambridge University Press, pp. 499-587.

Dixon, R.K., Solomon, A., Brown, S., Houghton, R., Trexier, M., Wisniewski, J., 1994. Carbon pools and flux of global forest ecosystems. *Science* 263, 185-190.

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.

Dorren, L.K., Maier, B., Seijmonsbergen, A.C., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management*, 183, 31-46.

Dovey, S.B., 2009. Estimating biomass and macronutrient content of some commercially important plantation species in South Africa. *Southern Forests*, 71, 245-251.

Dube, T., 2012. *Primary productivity of intertidal mudflats in the Wadden Sea: A remote sensing method,* Water Resources. ITC Faculty of Geo-Information Science, University, Enschede.

Dube, T., 2013. Evaluating the applicability of remote sensing techniques in quantifying microphytobenthic biomass in estuarine ecosystems. *Journal of Environmental Science and Water Resources* 2, 172 - 180.

Dube, T., Gara, T.W., Gumindoga, W., Chivhenge, E., Chinembiri, T.S., 2013. Characterization of the top sediment layer in coastal intertidal mudflats from medium-to-coarse resolution satellite imagery and field measurements. *International Journal of Water Resources and Environmental Engineering*, 5, 676-686.

Dube, T., Gumindoga, W., Chawira, M., 2014a. Detection of land cover changes around Lake Mutirikwi, Zimbabwe, based on traditional remote sensing image classification techniques. *African Journal of Aquatic Science* 39, 89-95.

Dube, T., Mutanga, O., 2014. *Estimating aboveground biomass in the uMgeni catchment, South Africa using medium-spatial resolution Landsat-8 OLI multispectral sensor*, 2nd National Conference on Global Change. Nelson Mandela Metropolitan University, p. 15.

Dube, T., Mutanga, O., 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing* 101, 36-46.

Dube, T., Mutanga, O., Abdel-Rahman, E., Ismail, R., 2014b. *Eucalyptus stand volume prediction in Zululand, South Africa, AARSE 2014*. AARSE2014, University of Johannesburg, South Africa, pp. 1-18.

Dube, T., Mutanga, O., Abdel-Rahman, EM., Ismail, R & Slotow R., (2015) Predicting Eucalyptus spp. stand volume in Zululand, South Africa: an analysis using a stochastic gradient boosting regression ensemble with multi-source data sets, *International Journal of Remote Sensing,* 36:14, 3751-3772, DOI: 10.1080/01431161.2015.1070316.

Dye, M., Mutanga, O., Ismail, R., 2011. Examining the utility of random forest and AISA Eagle hyperspectral image data to predict Pinus patula age in KwaZulu-Natal, South Africa. *Geocarto International* 26, 275-289.

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

El-Askary, H., Abd El-Mawla, S.H., Li, J., El-Hattab, M.M., El-Raey, M., 2014. Change detection of coral reef habitat using Landsat-5 TM, Landsat 7 ETM+ and Landsat 8 OLI data in the Red Sea (Hurghada, Egypt). *International Journal of Remote Sensing* 35, 2327-2346.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802-813.

Elvidge, C.D., Chen, Z., 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment* 54, 38-48.

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

ENVI, 2009. *Environment for Visualising Images*. USA: ITT industries, Inc.

Fan, Y., Koukal, T., Weisberg, P.J., 2014. A sun–crown–sensor model and adapted C-correction logic for topographic correction of high resolution forest imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96, 94-105.

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.

Ferwerda, J.G., Skidmore, A.K., Mutanga, O., 2005. Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *International Journal of Remote Sensing* 26, 4083-4095.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* 281, 237-240.
Foody, G.M., Boyd, D.S., 2002. Sharpened mapping of tropical forest biophysical properties from coarse spatial resolution satellite sensor data. *Neural computing and applications* 11, 62-70.
Foody, G.M., Boyd, D.S., Cutler, M.E., 2003a. 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., Boyd, D.S., Cutler, M.E.J., 2003b. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment* 85, 463-474.
Franklin, S.E., Wulder, M.A., Lavigne, M.B., 1996. Automated derivation of geographic window sizes for use in remote sensing digital image texture analysis. *Computers & Geosciences* 22, 665-673.
Fransson, J.E.S., Walter, F., Ulander, L.M.H., 2000. Estimation of forest parameters using CARABAS-II VHF SAR data. *Geoscience and Remote Sensing, IEEE Transactions* 38, 720-727.
Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* 55, 119-139.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *The Annals Statistics* 29, 1189-1232.
Friedman, J.H., 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38, 367-378.
Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. *Statistics in medicine* 22, 1365-1381.
Froeschke, J.T., Froeschke, B.F., 2011. Spatio-temporal predictive model based on environmental factors for juvenile spotted seatrout in Texas estuaries using boosted regression trees. *Fisheries Research* 111, 131-138.
FSA, 2010. *South African Government Information: Forestry*, in: Louise van Niekerk (Ed.), Government Communication and Information System GCIS.
Fuchs, H., Magdon, P., Kleinn, C., Flessa, H., 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sensing of Environment* 113, 518-531.
Gallardo-Cruz, J.A., Meave, J.A., González, E.J., Lebrija-Trejos, E.E., Romero-Romero, M.A., Pérez-García, Gallardo-Cruz, R., Hernández-Stefanoni, J.L., Martorell, C., 2012. Predicting tropical dry forest successional attributes from space: is the key hidden in image texture? *PloS one* 7, e30506.
Ganjisaffar, Y., Caruana, R., Lopes, C.V., 2011. Bagging gradient-boosted trees for high precision, low variance ranking models. , . *34th Annual Association for Computer Machinery (ACM) Special Interest Group of Information Retrieval (SIGIR),* Beijing, China.
Gao, X., Huete, A.R., Ni, W., Miura, T., 2000. Optical–Biophysical Relationships of Vegetation Spectra without Background Contamination. *Remote Sensing of Environment* 74, 609-620.
Gara, T., Murwira, A., Chivhenge, E., Dube, T., Bangira, T., 2014a. Estimating wood volume from canopy area in deciduous woodlands of Zimbabwe. Southern Forests: *Journal of Forest Science* 76, 237-244.
Gara, T.W., Murwira, A., Chivhenge, E., Dube, T., Bangira, T., 2014b. Estimating wood volume from canopy area in deciduous woodlands of Zimbabwe. Southern Forests: *Journal of Forest Science* 76, 237-244.

Gebreslasie, M.T., Ahmed, F.B., van Aardt, J.A.N., 2010. Predicting forest structural attributes using ancillary data and ASTER satellite data. *International Journal of Applied Earth Observation and Geoinformation* 12, Supplement 1, S23-S26.

Gebreslasie, M.T., Ahmed, F.B., van Aardt, J.A.N., 2011. Extracting structural attributes from IKONOS imagery for Eucalyptus plantation forests in KwaZulu-Natal, South Africa, using image texture analysis and artificial neural networks. *International Journal of Remote Sensing* 32:22, 7677-7701, DOI: 10.1080/01431161.2010.527392.

Geldenhuys, C.J., 2000. *Classification and distribution of forest and woodlands in South Africa* in: Owen, D.L., Vermeulen, W.J. (Eds.), South African Forestry Handbook fourth. South African Institute of Forestry, Pretoria.

Giardina, C.P., Ryan, M.G., 2002. Total belowground carbon allocation in a fast-growing Eucalyptus plantation estimated using a carbon balance approach. *Ecosystems* 5, 487-499.

Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research letters* 2.

Gil, M.V., Blanco, D., Carballo, M.T., Calvo, L.F., 2011. Carbon stock estimates for forests in the Castilla y León region, Spain. A GIS based method for evaluating spatial distribution of residual biomass for bio-energy. *Biomass and Bioenergy* 35, 243-252.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. *Pattern Recognition Letters* 27, 294-300.

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* 58, 289-298.

Godsmark, R. 2010. *The South African Forestry and Forest Products Industry 2009*. Pietermaritzburg: Forestry South Africa.

Goh, J., Miettinen, J., Chia, A.S., Chew, P.T., Liew, S.C., 2014. Biomass estimation in humid tropical forest using a combination of ALOS PALSAR and SPOT 5 satellite imagery. *Asian Journal of Geoinformatics* 13:1-10.

Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., Palace, M., 2010. Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. *Remote Sensing of Environment* 114, 1561-1575.

Gracia, C., Lasanta Martínez, T., Vicente Serrano, S.M., 2010. Aridification determines changes in forest growth in Pinus halepensis forests under semiarid Mediterranean climate conditions. *Agricultural and Forest Meteorology* 150, 614-628.

Grömping, U., 2009. Variable importance assessment in regression: linear regression versus random forest. *The American Statistician* 63.

Gschwantner, T., Schadauer, K., Vidal, C., Lanz, A., Tomppo, E., di Cosmo, L., Robert, N., Englert Duursma, Lawrence, M., . 2009. Common tree definitions for national forest inventories in Europe. *Silva Fennica* 43, 303-321.

Gumindoga, W., Rientjes, T., Haile, A.T., Dube, T., 2015. Predicting streamflow for land cover changes in the Upper Gilgel Abay River Basin, Ethiopia: A TOPMODEL based approach. *Physics and Chemistry of the Earth*, Parts A/B/C.

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

Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 56-66.

Guo, Z., Fang, J., Pan, Y., Birdsey, R., 2010. Inventory-based estimates of forest biomass carbon stocks in China: A comparison of three methods. *Forest Ecology and Management* 259, 1225-1231.

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

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.

Haralick, R.M., 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE* 67, 786-804.

Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural Features for Image Classification. Systems, Man and Cybernetics, *IEEE Transactions on SMC*-3, 610-621.

Hardisky, M., Klemas, V., Smart, M., 1983. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of. *Spartina alterniflora*, 77-83.

Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, 2 ed. Springer, New York.

He, J.-Z., Ge, Y., Xu, Z., Chen, C., 2009. Linking soil bacterial diversity to ecosystem multifunctionality using backward-elimination boosted trees analysis. *J Soils Sediments* 9, 547-554.

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

Henry, M., Picard, N., Trotta, C., Manlay, R.J., Valentini, R., Bernoux, M., Saint-André, L., 2011. Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. *Silva Fennica* 45, 477-569.

Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. *International Journal of Remote Sensing* 4, 273-288.

Horning, N., 2010. Random Forests : An algorithm for image classification and generation of continuous fields data sets, *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, Hanoi, Vietnam.

Houghton, J.T., Meira Filho, L.G., Lim, B., Treanton, K., Mamaty, I., Bonduki, Y., Griggs, D.J., Callender, B.A., 1996. *Greenhouse Gas Inventory Reference Manual, IPCC Guidelines for National Greenhouse Gas Inventories*. IPCC/UK Meteorological Office, Bracknell, UK.

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

Hudak, A.T., Lefsky, M.A., Cohen, W.B., Berterretche, M., 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment* 82, 397-416.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25, 295-309.

Hyyppä, H., Hyyppä, J., 2001. Effects of stand size on the accuracy of remote sensing-based forest inventory. *Geoscience and Remote Sensing, IEEE Transactions* 39, 2613-2621.

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

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

IPCC, 1996. Climate change 1995. *Impacts, adaptations and mitigation of climate change: scientific-technical analyses* in: Watson, R.T., Zinyowera, M.C., Moss, R.H., Dokken, D.J.

(Eds.), Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University, Cambridge, UK.
IPCC, 2003a. *Good Practice guidance for land use, land-use change and forestry,* IPCC national greenhouse gas inventories Programme. Cambridge University, Cambridge.
IPCC, 2003b. *Good Practice Guidance For Land Use, Land-Use Change and Forestry*. Institute for Global Environmental Strategies (IGES), Hayama.
IPCC, 2007. *Climate Change, The Physical Science Basis. Summary for Policymakers. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, in: Isaaks, E.H., Srivastava, R.M. (Eds.), An introduction to applied geostatistics. Oxford University Press, New York.
Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment* 122, 11-21.
Ismail, R., Mutanga, O., 2010. A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation* 12, Supplement 1, S45-S51.
Jain, N., Ray, S.S., Singh, J., Panigrahy, S., 2007. Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop. *Precision Agriculture* 8, 225-239.
Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., Li, B., 2014. Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto International*, 29:8, 941-951, DOI: 10.1080/10106049.2014.894586.
Jones, L.A., Ricciardi, A., 2005. Influence of physicochemical factors on the distribution and biomass of invasive mussels (Dreissena polymorpha and Dreissena bugensis) in the St. Lawrence River. *Canadian Journal of Fisheries and Aquatic Sciences* 62, 1953-1962.
Jordan, C.F., 1969a. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* 50, 663-666.
Jordan, C.F., 1969b. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 663-666.
Kassier, H. W. 2012. "*Stand Volume Estimated by Counting Trees*." In South African Forestry Handbook, edited by H. W. Bredenkamp and S. J. Upfold. Pinetown: South African forestry institute.
Kajisa, T., Murakami, T., Mizoue, N., Top, N., Yoshida, S., 2009. Object-based forest biomass estimation using Landsat ETM+ in Kampong Thom Province, Cambodia. *Journal of Forest Research* 14, 203-211.
Keller, M., Palace, M., Hurtt, G., 2001. Biomass estimation in the Tapajos National Forest, Brazil: examination of sampling and allometric uncertainties. *Forest Ecology and Management* 154, 371-382.
Kelsey, K.C., Neff, J.C., 2014. Estimates of aboveground biomass from texture analysis of landsat imagery. *Remote Sens.* 6, 6407–6422.
Kimes, D., Markham, B., Tucker, C., McMurtrey III, J., 1981. Temporal relationships between spectral response and agronomic variables of a corn canopy. *Remote Sensing of Environment* 11, 401-411.
Kint, V., Vansteenkiste, D., Aertsen, W., Vos, B., Bequet, R., Acker, J., Muys, B., 2012. Forest structure and soil fertility determine internal stem morphology of Pedunculate oak: a modelling approach using boosted regression trees. *European Journal of Forest Research* 131, 609-622.
Knox, N.M., Skidmore, A.K., Prins, H.H., Asner, G.P., van der Werff, H., de Boer, W.F., van der Waal, C., de Knegt, H.J., Kohi, E.M., Slotow, R., 2011. Dry season mapping of savanna

forage quality, using the hyperspectral Carnegie Airborne Observatory sensor. *Remote Sensing of Environment* 115, 1478-1488.

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.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, *International Joint Conference on Articial Intelligence (IJCAI),* pp. 1137-1145.

Kokaly, R., Clark, R., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment* 67, 267-287.

Krahwinkler, P., Rossman, J., 2011. Using decision tree based multiclass support vector machines for forest mapping, *IEEE Int. Geoscience and Remote Sensing Symp.* Vancouver, Canada.

Kraus, T., Samimi, C., 2002. Biomass Estimation for Land Use Management and Fire Management Using Landsat-TM and-ETM+ (Biomassebestimmung mit Hilfe von Landsat-TM und-ETM+ für Landnutzungsplanung und für Feuermanagement). *Erdkunde*, 130-143.

Kross, A., McNairn, H., Lapen, D., Sunohara, M., Champagne, C., 2015. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation* 34, 235-248.

Kuplich, T., Curran, P., Atkinson, P., 2005. Relating SAR image texture to the biomass of regenerating tropical forests. *International Journal of Remote Sensing* 26, 4829-4854.

Kurz, W.A., Apps, M.J., 2006. *Developing Canada's National Forest Carbon Monitoring, Accounting and Reporting System to Meet the Reporting Requirements of the Kyoto Protocol*. Mitigation and Adaptation Strategies for Global Change 11: 33-43.

Laurent, V.C., Schaepman, M.E., Verhoef, W., Weyermann, J., Chávez, R.O., 2014. Bayesian object-based estimation of LAI and chlorophyll from a simulated Sentinel-2 top-of-atmosphere radiance image. *Remote Sensing of Environment* 140, 318-329.

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

Law, M., Collins, A., 2013. *Getting to know ArcGIS for desktop*. Esri Press.

Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote sensing of environment* 90, 331-336.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment* 100, 356-362.

le Maire, G., Marsden, C., Nouvellon, Y., Grinand, C., Hakamada, R., Stape, J.-L., Laclau, J.-P., 2011. MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass. *Remote Sensing of Environment* 115, 2613-2625.

Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T., Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* 321, 267-281.

Lee, K.-S., Cohen, W.B., Kennedy, R.E., Maiersperger, T.K., Gower, S.T., 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment* 91, 508-520.

Lewin, W.C., Mehner, T., Ritterbusch, D., Brämick, U., 2014. The influence of anthropogenic shoreline changes on the littoral abundance of fish species in German lowland lakes varying in depth as determined by boosted regression trees. *Hydrobiologia* 724, 293-306.

Liaw, A., Wiener, M., 2002. Classification and regression by random Forest. R News 2, 18-22.
Lima, K., Treitz, P., M., W., M., F., K., D.V., 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography* 27, 88-106.
Little, K.M., Toit, B.D., 2003. Management of Eucalyptus grandis coppice regeneration of seedling parent stock in Zululand, South Africa. *Australian Forestry* 66, 108-112.
Lu, D., 2005a. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *International Journal of Remote Sensing* 26, 2509-2525.
Lu, D., 2005b. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing* 27, 1297-1328.
Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing* 27, 1297-1328.
Luizão, R.C., Luizão, F.J., Paiva, R.Q., Monteiro, T.F., Sousa, L.S., Kruijt, B., 2004. Variation of carbon and nitrogen cycling processes along a topographic gradient in a central Amazonian forest. *Global Change Biology* 10, 592-600.
Luo, Y., 2007. Terrestrial carbon-cycle feedback to climate warming. *Annual Review of Ecology, Evolution, and Systematics*, 683-712.
Luther, J.E., Fournier, R.A., Piercey, D.E., Guindon, L., Hall, R.J., 2006. Biomass mapping using forest type and structure derived from Landsat TM imagery. *International Journal of Applied Earth Observation and Geoinformation* 8, 173-187.
MacDicken, K.G., 1997. *A guide to monitoring carbon storage in forestry and agroforestry projects.* Winrock International Institute for Agricultural Development USA.
Machwitz, M., Giustarini, L., Bossung, C., Frantz, D., Schlerf, M., Lilienthal, H., Wandera, L., Matgen, P., Hoffmann, L., Udelhoven, T., 2014. Enhanced biomass prediction by assimilating satellite data into a crop growth model. *Environmental Modelling & Software* 62, 437-453.
Mackey, B., Prentice, I.C., Steffen, W., House, J.I., Lindenmayer, D., Keith, H., Berry, S., 2013. Untangling the confusion around land carbon science and climate change mitigation policy. *Nature climate change* 3, 552-557.
Main-Knorn, M., Moisen, G.G., Healey, S.P., Keeton, W.S., Freeman, E.A., Hostert, P., 2011. Evaluating the Remote Sensing and Inventory-Based Estimation of Biomass in the Western Carpathians. *Remote Sensing* 3, 1427-1446.
Main-Knorn, M., W. B. Cohen, R. E. Kennedy, W. Grodzki, D. Pflugmacher, P. Griffiths, and P. Hostert. 2013. "Monitoring Coniferous Forest Biomass Change Using a Landsat TrajectoryBased Approach." *Remote Sensing of Environment* 139: 277–290. doi:10.1016/j. rse.2013.08.010.
Major, D., Baret, F., Guyot, G., 1990. A ratio vegetation index adjusted for soil brightness. *International Journal of Remote Sensing* 11, 727-740.
Martinic, I., Sporcic, M., 2005. *Integrating global climate changes into core forestry activities. Climate Change–Forest Ecosystems & Landscape*, 44.
Mathieu, R., Naidoo, L., Cho, M.A., Leblon, B., Main, R., Wessels, K., Asner, G.P., Buckley, J., Van Aardt, J., Erasmus, B.F.N., Smit, I.P.J., 2013. Toward structural assessment of semi-arid African savannahs and woodlands: The potential of multitemporal polarimetric RADARSAT-2 fine beam images. *Remote Sensing of Environment* 138, 215-231.
Matthew, M.W., Adler-Golden, S.M., Berk, A., Richtsmeier, S.C., Levine, R.Y., Bernstein, L.S., Acharya, P.K., Anderson, G.P., Felde, G.W., Hoke, M.L., 2000. Status of atmospheric correction using a MODTRAN4-based algorithm, AeroSense 2000. *International Society for Optics and Photonics*, pp. 199-207.
Mitchard, E.T.A., Meir, P., Ryan, C.M., Woollen, E.S., Williams, M., Goodman, L.E., Mucavele, J.A., Watts, P., Woodhouse, I.H., Saatchi, S.S., 2012. A novel application of satellite radar

data: measuring carbon sequestration and detecting degradation in a community forestry project in Mozambique. *Plant Ecology & Diversity* 6, 159-170.
Mitchard, E.T.A., Saatchi, S.S., Lewis, S.L., Feldpausch, T.R., Woodhouse, I.H., Sonké, B., Rowland, C., Meir, P., 2011. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sensing of Environment* 115, 2861-2873.
Mitchard, E.T.A., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N.S., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T.R., Meir, P., 2009. Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters* 36, L23401.
Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards Jr, T.C., 2006. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling* 199, 176-187.
Montesano, P.M., Nelson, R.F., Dubayah, R.O., Sun, G., Cook, B.D., Ranson, K.J.R., Næsset, E., Kharuk, V., 2014. The uncertainty of biomass estimates from LiDAR and SAR across a boreal forest structure gradient. *Remote Sensing of Environment* 154, 398-407.
Mousivand, A., Menenti, M., Gorte, B., Verhoef, W., 2014. Global sensitivity analysis of the spectral radiance of a soil–vegetation system. *Remote Sensing of Environment* 145, 131-144.
Mutanga, O., Adam, E., Adjorlolo, C., Abdel-Rahman, E.M., 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. *International Journal of Applied Earth Observation and Geoinformation* 34, 178-187.
Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation* 18, 399-406.
Mutanga, O., Skidmore, A., 2004a. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing* 25, 3999-4014.
Mutanga, O., Skidmore, A.K., 2004b. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing* 25, 3999-4014.
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.
Muvengwi, J., Mbiba, M., Nyenda, T., 2015. Using branch diameter to estimate fresh biomass removal by elephants: comparison of linear and quadratic models. *African Journal of Ecology* 53, 126-129.
Næsset, E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research* 22, 433-442.
Naughton, D., Brunn, A., Czapla-Myers, J., Douglass, S., Thiele, M., Weichelt, H., Oxfort, M., 2011. Absolute radiometric calibration of the RapidEye multispectral imager using the reflectance-based vicarious calibration method. *Journal of Applied Remote Sensing* 5, 053544.
Nelson, R., Valenti, M.A., Short, A., Keller, C., 2003. A Multiple Resource Inventory of Delaware Using Airborne Laser Data. *BioScience* 53, 981-992.

Nemani, R., Pierce, L., Running, S., Band, L., 1993. Forest ecosystem processes at the watershed scale: sensitivity to remotely-sensed Leaf Area Index estimates. *International Journal of Remote Sensing* 14, 2519-2534.

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

Özçift, A., 2011. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. *Computers in Biology and Medicine* 41, 265-271.

Ozdemir, I., Karnieli, A., 2011. Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *International Journal of Applied Earth Observation and Geoinformation* 13, 701-710.

Pahlevan, N., Schott, J.R., 2013. Leveraging EO-1 to Evaluate Capability of New Generation of Landsat Sensors for Coastal/Inland Water Studies. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE* 6, 360-374.

Pal, M., 2003. Random forests for land cover classification, Geoscience and Remote Sensing Symposium, 2003. *IGARSS '03. Proceedings. 2003 IEEE International*, pp. 3510-3512 vol.3516.

Palace, M.W., Sullivan, F.B., Ducey, M.J., Treuhaft, R.N., Herrick, C., Shimbo, J.Z., Mota-E-Silva, J., 2015. Estimating forest structure in a tropical forest using field measurements, a synthetic model and discrete return lidar data. *Remote Sensing of Environment* 161, 1-11.

Palmer, D.S., O'Boyle, N.M., Glen, R.C., Mitchell, J.B., 2007. Random forest models to predict aqueous solubility. *Journal of chemical information and modeling* 47, 150-158.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A Large and Persistent Carbon Sink in the World's Forests. *Science* 333, 988-993.

Pandey, U., Kushwaha, S., Kachhwaha, T., Kunwar, P., Dadhwal, V., 2010. Potential of Envisat ASAR data for woody biomass assessment. *Tropical Ecology* 51, 117.

Patenaude, G., Hill, R.A., Milne, R., Gaveau, D.L.A., Briggs, B.B.J., Dawson, T.P., 2004. Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment* 93, 368-380.

Patenaude, G., Milne, R., Dawson, T.P., 2005a. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science and Policy* 8, 161-178.

Patenaude, G., Ronald, M., Terence, P.D., 2005b. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science and Policy* 8 161-178.

Perkins, T., Adler-Golden, S., Matthew, M., Berk, A., Anderson, G., Gardner, J., Felde, G., 2005. Retrieval of atmospheric properties from hyper and multispectral imagery with the FLAASH atmospheric correction algorithm, Remote Sensing. *International Society for Optics and Photonics*, pp. 59790E-59790E-59711.

Pierce, A.D., Farris, C.A., Taylor, A.H., 2012. Use of random forests for modeling and mapping forest canopy fuels for fire behavior analysis in Lassen Volcanic National Park, California, USA. *Forest Ecology and Management* 279, 77-89.

Pinkerton, M.H., Smith, A.N.H., Raymond, B., Hosie, G.W., Sharp, B., Leathwick, J.R., Bradford-Grieve, J.M., 2010. Spatial and seasonal distribution of adult Oithona similis in the Southern Ocean: Predictions using boosted regression trees. Deep Sea Research Part I: *Oceanographic Research Papers* 57, 469-485.

Pinto, N., Simard, M., Dubayah, R., 2012. Using InSAR coherence to map stand age in a boreal forest. *Remote Sensing* 5, 42-56.
Pinty, B., Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetation* 101, 15-20.
Ploton, P., Pélissier, R., Barbier, N., Proisy, C., Ramesh, B., Couteron, P., 2013. *Canopy texture analysis for large-scale assessments of tropical forest stand structure and biomass, Treetops at Risk*. Springer, pp. 237-245.
Podest, E., Saatchi, S., 2002. Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. *International Journal of Remote Sensing* 23, 1487-1506.
Popescu, S.C., Zhao, K., Neuenschwander, A., Lin, C., 2011. Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environm*ent 115, 2786-2797.
Poulain, M., Peña, M., Schmidt, A., Schmidt, H., Schulte, A., 2011. Aboveground biomass estimation in intervened and non-intervened Nothofagus pumilio forests using remotely sensed data. *International Journal of Remote Sensing* 33, 3816-3833.
Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 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.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 9, 181-199.
Pregitzer, K.S., Euskirchen, E.S., 2004. Carbon cycling and storage in world forests: biome patterns related to forest age. *Global Change Biology* 10, 2052-2077.
Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. *Remote Sensing of Environment* 48, 119-126.
R Development Core Team, 2008. *R: A language and environment for statistical computing*, Vienna, Austria.
Raich, J.W., Clark, D.A., Schwendenmann, L., Wood, T.E., 2014. Aboveground Tree Growth Varies with Belowground Carbon Allocation in a Tropical Rainforest Environment. *PloS one* 9, e100275.
Rana, P., Tokola, T., Korhonen, L., Xu, Q., Kumpula, T., Vihervaara, P., Mononen, L., 2013. Training area concept in a two-phase biomass inventory using airborne laser scanning and RapidEye satellite data. *Remote Sensing* 6, 285-309.
RapidEye, 2011. *Satellite imagery product specifications*. [online] Available at: http://www.RapidEye.de/upload/-RE_Product_Specifications_ENG.pdf [Accessed 01 Nov. 2011].
Ren-dong, L., Ji-yuan, L., 2002. Wetland vegetation biomass estimation and mapping from Landsat ETM data: a case study of Poyang Lake. *Journal of Geographical Sciences* 12, 35-41.
Ridgeway, G., 2006. Generalized boosted models, A guide to the gbm package. Gbm libraryforR.
Riggins, J.J., Tullis, J.A., Stephen, F.M., 2009. Per-segment aboveground forest biomass estimation using LIDAR-derived height percentile statistics. *GIScience & Remote Sensing* 46, 232-248.
Rock, B., Vogelmann, J., Williams, D., Vogelmann, A., Hoshizaki, T., 1986. Remote detection of forest damage. *BioScience,* 439-445.
Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment* 55, 95-107.

Roujean, J.-L., Breon, F.-M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* 51, 375-384.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. *Monitoring the Vernal Advancements and Retrogradation (Greenwave Effect) of Nature Vegetation*; NASA/GSFC Final Report. NASA, Greenbelt, MD, USA.

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 - PNAS* 108, 9899-9904.

Sagi, 2012. *South African Forestry*, in: Niekerk, L.v. (Ed.), Government Communication and Information System. Sagi, Pretoria, South Africa.

Santoro, M., Eriksson, L., Askne, J., Schmullius, C., 2006. Assessment of stand-wise stem volume retrieval in boreal forest from JERS-1 L-band SAR backscatter. *International Journal of Remote Sensing* 27, 3425-3454.

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

Sappi, 1993. *Forest Land Types of the Natal Region Sappi Forests Research*, Howick, KwaZulu Natal, South Africa.

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

Sarker, M.L.R., Nichol, J., Iz, H.B., Ahmad, B.B., Rahman, A.A., 2013. Forest biomass estimation using texture measurements of high-resolution dual-polarization C-band SAR data. *Geoscience and Remote Sensing, IEEE Transactions* 51, 3371-3384.

Schapire, R., 2003. *The boosting approach to machine learning – an overview*. Springer, New York.

Schönau, A., Boden, D., 1982a. Preliminary Biomass Studies in Young Eucalypts. *South African Forestry Journal* 120, 1-24.

Schönau, A.P.G., Boden, D.I., 1982b. Preliminary Biomass Studies in Young Eucalypts. *South African Forestry Journal* 120, 24-28.

Schuster, C., Forster, M., Kleinschmit, B., 2012. Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing* 33, 5583-5599.

Scott, D.F., Lesch, W., 1997. Streamflow responses to afforestation with Eucalyptus grandis and Pinus patula and to felling in the Mokobulaan experimental catchments, South Africa. *Journal of Hydrology* 199, 360-377.

Shamsoddini, A., Trinder, J.C., Turner, R., 2013. Pine plantation structure mapping using WorldView-2 multispectral image. *International Journal of Remote Sensing* 34, 3986-4007.

Sharma, C., Gairola, S., Baduni, N., Ghildiyal, S., Suyal, S., 2011. Variation in carbon stocks on different slope aspects in seven major forest types of temperate region of Garhwal Himalaya, India. *Journal of biosciences* 36, 701-708.

Shataee, S., W. Holger, and B. Manoucher. 2011. "Plot-level forest volume estimation using airborne laser scanner and tm data, comparison of boosting and random forest tree regression algorithms." *Procedia Environmental Sciences* 7: 68–73. doi:10.1016/j. proenv.2011.07.013..

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., Lucas, R., 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment.*

Singh, M., Malhi, Y., Bhagwat, S., 2014. Biomass estimation of mixed forest landscape using a Fourier transform texture-based approach on very-high-resolution optical satellite imagery. *International Journal of Remote Sensing* 35, 3331-3349.

Singh, S.P., Adhikari, B.S., Zobel, D.B., 1994. Biomass, productivity, leaf longevity, and forest structure in the central Himalaya. *Ecological Monographs*, 401-421.

Sivanpillai, R., Smith, C.T., Srinivasan, R., Messina, M.G., Benwu, X., 2006. Estimation of managed loblolly pine stands age and density with Landsat ETMþ data. *Forest Ecology and Management* 223, 247-254.

Somogyi, Z., Teobaldelli, M., Federici, S., Matteucci, G., Pagliari, V., Grassi, G., Seufert, G., 2008. Allometric biomass and carbon factors database. iForest - *Biogeosciences and Forestry* 1, 107-113.

Souza, A.F., Cortez, L.S.R., Longhi, S.J., 2012. Native forest management in subtropical South America: long-term effects of logging and multiple-use on forest structure and diversity. *Biodiversity and Conservation* 21, 1953-1969.

Soykan, C.U., Eguchi, T., Kohin, S., Dewar, H., 2014. Prediction of fishing effort distributions using boosted regression trees. *Ecological Applications* 24, 71-83.

St-Onge, B., Hu, Y., Vega, C., 2008. Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images. *International Journal of Remote Sensing* 29, 1277-1294.

Storey, J., Scaramuzza, P., Schmidt, G., Barsi, J., 2005. Landsat 7 scan line corrector-off gap filled product development, *Proceedings of Pecora*, pp. 23-27.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9, 307.

Stumpf, A., Kerle, N., 2011. Combining Random Forests and object-oriented analysis for landslide mapping from very high resolution imagery. *Procedia Environmental Sciences* 3, 123-129.

Suen, Y.L., Melville, P., Mooney, R.J., 2005. Combining bias and variance reduction techniques for regression trees, *16th European Conference on machine Learning (ECML*), Porto, Portugal.

Sun, G., Ranson, K.J., Guo, Z., Zhang, Z., Montesano, P., Kimes, D., 2011. Forest biomass mapping from lidar and radar synergies. *Remote Sensing of Environment* 115, 2906-2916.

Swatantran, A., Dubayah, R., Roberts, D., Hofton, M., Blair, J.B., 2011. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sensing of Environment* 115, 2917-2930.

Tarpley, J., Schneider, S., Money, R.L., 1984. Global vegetation indices from the NOAA-7 meteorological satellite. *Journal of Climate and Applied Meteorology* 23, 491-494.

Teillet, P., Staenz, K., William, D., 1997. Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. *Remote Sensing of Environment* 61, 139-149.

Tesfamichael, S.G., Ahmed, F.B., Van Aardt, J.A.N., 2010. Investigating the impact of discrete-return lidar point density on estimations of mean and dominant plot-level tree height in Eucalyptus grandis plantations. *International Journal of Remote Sensing* 31, 2925-2940.

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

Thenkabail, P.S., Smith, R.B., De Pauw, E., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment* 71, 158-182.

Thenkabail, P.S., Stucky, N., Griscom, B.W., Ashton, M.S., Diels, J., van der Meer, B., Enclona, E., 2004b. Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing* 25, 5447-5472.

Tigges, J., Lakes, T., Hostert, P., 2013. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. *Remote Sensing of environment* 136, 66-75.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8, 127-150.

UNFCCC, 1998a. *Kyoto Protocol to the United Nations framework convention on climate change, UNFCCC*, Bonn.

UNFCCC, 1998b. *United Nations Framework Convention on Climate Change*.

UNFCCC, 2001, 2011. *The Marrakesh Accords and the Marrakesh Declaration, Seventh Conference of the Parties (COP 7). UNFCCC*, Marrakesh, Morocco.

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

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

van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Kasibhatla, P.S., Arellano Jr, A.F., 2006. Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics* 6, 3423-3441.

Vashum, K.T., Jayakumar, S., 2012. Methods to Estimate Above-Ground Biomass and Carbon Stock in Natural Forests-A Review. *Journal of Ecosystem and Ecography* 2, 1-7.

Verbyla, D.L., 1995. *Satellite remote sensing of natural resources*. CRC Press.

Viero, P.W.M., 2002. *The effect of a soil-amending hydrogel on Eucalyptus grandis establishment practices in the Zululand forestry region,* Faculty of Forestry. Port Elizabeth Technikon.

Walawender, J.P., Hajto, M.J., Iwaniuk, P., 2012. A new ArcGIS toolset for automated mapping of land surface temperature with the use of Landsat satellite data, *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pp. 4371-4374.

Wang, C., Menenti, M., Stoll, M.-P., Belluco, E., Marani, M., 2007. Mapping mixed vegetation communities in salt marshes using airborne spectral data. *Remote Sensing of Environment* 107, 559-570.

Wang, Q., Adiku, S., Tenhunen, J., Granier, A., 2005. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sensing of Environment* 94, 244-255.

Wei, Y., Li, M., Chen, H., Lewis, B.J., Yu, D., Zhou, L., Zhou, W., Fang, X., Zhao, W., Dai, L., 2013. Variation in carbon storage and its distribution by stand age and forest type in boreal and temperate forests in Northeastern China. *PloS one* 8, e72201.

Wessels, N.O., Kassier, H.W., 1985. A Computerised System for Forest Management and Silvicultural Planning and Control in Even-aged Plantation Forestry. *South African Forestry Journal* 132, 62-64.

Wiegand, C., Richardson, A., Escobar, D., Gerbermann, A., 1991. Vegetation indices in crop assessments. *Remote Sensing of Environment* 35, 105-119.

Wolf, A., Ciais, P., Bellassen, V., Delbart, N., Field, C.B., Berry, J.A., 2011. Forest biomass allometry in global land surface models. *Global Biogeochemical Cycles* 25.

Woollen, E., Ryan, C., Williams, M., 2012a. Carbon stocks in an African woodland landscape: Spatial distributions and scales of variation. *Ecosystems* 15, 804-818.

Woollen, E., Ryan, C.M., Williams, M., 2012b. Carbon stocks in an African woodland landscape: spatial distributions and scales of variation. *Ecosystems* 15, 804-818.

Wu, W., Pauwa, E., Helldén, U., 2013. Assessing woody biomass in African tropical savannahs by multiscale remote sensing. *International Journal of Remote Sensing* 34, 4525-4549.

Wulder, M.A., LeDrew, E.F., Franklin, S.E., Lavigne, M.B., 1998. Aerial image texture information in the estimation of northern deciduous and mixed wood forest leaf area index (LAI). *Remote Sensing of Environment* 64, 64-76.

Wulder, M.A., Ortlepp, S.M., White, J.C., Maxwell, S., 2008a. Evaluation of Landsat-7 SLC-off image products for forests change detection. *Canadian journal of remote sensing* 34, 93-99.

Wulder, M.A., White, J.C., Goward, S.N., Masek, J.G., Irons, J.R., Herold, M., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2008b. Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment* 112, 955-969.

Xu, X., Du, H., Zhou, G., Ge, H., Shi, Y., Zhou, Y., Fan, W., Fan, W., 2011. Estimation of aboveground carbon stock of Moso bamboo (Phyllostachys heterocycla var. pubescens) forest with a Landsat Thematic Mapper image. *International Journal of Remote Sensing* 32, 1431-1448.

Yang, C., Huang, H., Wang, S., 2011. Estimation of tropical forest biomass using Landsat TM imagery and permanent plot data in Xishuangbanna, China. *International Journal of Remote Sensing* 32, 5741-5756.

Yang, K., Guan, D., 2008. Changes in forest biomass carbon stock in the Pearl River Delta between 1989 and 2003. *Journal of Environmental Sciences* 20, 1439-1444.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* 181, 23-48.

Zandler, H., Brenning, A., Samimi, C., 2015. Quantifying dwarf shrub biomass in an arid environment: comparing empirical methods in a high dimensional setting. *Remote Sensing of Environment* 158, 140-155.

Zarco-Tejada, P.J., Berjón, A., López-Lozano, R., Miller, J.R., Martín, P., Cachorro, V., González, M.R., de Frutos, A., 2005. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment* 99, 271-287.

Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y., Myneni, R.B., 2014. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sensing of Environment* 151, 44-56.

Zhang, H., Guan, D., Song, M., 2012. Biomass and carbon storage of Eucalyptus and Acacia plantations in the Pearl River Delta, South China. *Forest Ecology and Management* 277, 90-97.

Zhang, X., Kondragunta, S., 2006. Estimating forest biomass in the USA using generalized allometric models and MODIS land products. *Geophysical Research Letters* 33, L09402.

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

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., Ryu, S.-R., 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment* 93, 402-411.

Zheng, G., J. M. Chen, Q. J. Tian, W. M. Ju, and X. Q. Xia. 2007. "Combining Remote Sensing Imagery and Forest Age Inventory for Biomass Mapping." *Journal of Environmental Management* 85: 616–623. doi:10.1016/j.jenvman.2006.07.015.