Modelling Terrain Roughness Using LiDAR Derived Digital Terrain Model in Eucalyptus Plantation Forests, in KwaZulu-Natal, South Africa

by

Roxanne Munsamy

Submitted in fulfilment of the academic requirements of

Master of Science in Environmental Science

Discipline of Geography

School of Agricultural, Earth, and Environmental Sciences

University of KwaZulu-Natal

Westville

South Africa

18 September 2017

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Geography, School of Agricultural, Earth and Environmental Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, South Africa. The National Research Foundation and School of Agriculture, Earth, and Environmental Sciences financially supported the research.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

Signed: Ms Roxanne Munsamy

Date: 18 September 2017

Signed: Dr Michael Gebreslasie

Date: 18 September 2017

DECLARATION 1: PLAGIARISM

I, Roxanne Munsamy, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation 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, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at research conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: Ms Roxanne Munsamy

Date: 18 September 2017

DECLARATION 2: PUBLICATIONS

My role in each paper and presentation is indicated. The * indicates corresponding author.

Chapter 2

> Munsamy, R*., Ismail, R., and Gebreslasie, T. M. (2017) Modelling the Effect of Terrain Variability in Even-aged Eucalyptus Species Volume using LiDAR-derived DTM Variables. *In preparation.*

Chapter 3

> Munsamy, R*., Ismail, R., and Gebreslasie, T. M. (2017) Modelling Terrain Roughness using a LiDAR-derived DTM and a Supervised Random Forest Approach. *In preparation.*

Signed: Ms Roxanne Munsamy	Signed: Dr Michael Gebreslasie
Date: 18 September 2017	Date: 18 September 2017

ACKNOWLEDGEMENTS

Firstly, I would like to give my deepest appreciation to SAPPI Forest Company for providing me with the data that was used in this study, without which would not have been possible.

My deepest gratitude is extended to the National Research Foundation and School of Agricultural, Earth, and Environmental Sciences for providing me with the funding for the duration of this research.

I cannot express how thankful I am to my parents, my sister Courtney and Ashlen for being my strength during this research.

I am also deeply grateful to my co-supervisor Dr Riyad Ismail, for his assistance in the technical components of the research. Your assistance in this project has no bounds.

Lastly, I would like to express my warmest thanks and gratitude to Dr Michael Gebreslasie, my supervisor, for his professional support and personal guidance throughout my postgraduate studies. His love for Geographic Information Systems and Remote Sensing Sciences has also been a reason for the development of my passion in the geospatial sciences.

ABSTRACT

South African commercial plantation forests are established primarily to meet both the local and global demands of industries that require direct raw materials such as pulpwood or timber. Consequently, the commercial forest industry in South Africa is held in high esteem as it makes up one of the largest economic forces within the country. For this reason, individuals responsible for implementing strategies pertaining to silvicultural and harvesting operations within commercial plantations require up to date and detailed multi-forest inventory datasets to ensure that optimal yields are guaranteed and that sites are well maintained.

Despite this, various drawbacks within commercial plantations exist: steep slopes, high elevations, and other forms of topographic irregularities, can affect the productivity of the site and impact mechanical silvicultural and harvesting operations. In lieu of making more informed and efficient decision-making protocols, forest researchers are often tasked with implementing and utilising alternative technologies such as remote sensing to determine if specific methodologies can be used for gathering multi-forest inventory data that also incorporate terrain information. Light Detection and Ranging (LiDAR), a recent remote sensing technology, has demonstrated that it is highly robust and can lend itself towards providing highly accurate vertical forest structural attributes and horizontal topographic derivatives.

This study employs the use of a LiDAR derived Digital Terrain Model (DTM) (1 m x 1 m spatial resolution) to create terrain indices that are representative of the horizontal features within the commercial forest sites of interest. In addition, a machine learning approach using a random forest (RF) ensemble classifier was adopted to determine how much of the variation in forest structural attributes: mean dominant height, mean height, pulpwood volumes and diameter at breast height can be attributed to terrain when using the LiDAR derived DTM terrain variables.

The overall findings presented in this study are encouraging and show that a LiDAR derived DTM can be successfully used for creating highly accurate terrain indices and can be used for predicting variability within even-aged Eucalyptus forest structural attributes within commercial plantation forests in KwaZulu-Natal, South Africa, with an acceptable level of accuracy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3D	Three Dimensional
ALS	Airborne Laser Scanning
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
ASP	Aspect
BA	Basal Area
BAM	Bayesian Model Averaging
CC	Curvature Classification
CI	Convergence Index
DBH	Diameter Breast Height
DC	Downslope Curvature
DDG	Downslope Distance Gradient
DEM	Digital Elevation Model
DTM	Digital Terrain Model
DWAF	Department of Water and Forestry
FA	Flow Accumulation
GAM	Generalised Addictive Model
GCP	Ground Control Point
GIS	Geographic Information Systems
GPS	Global Positioning Systems
Ha	Hectares
HtD	Mean dominant Height
Htm	Mean Height

kNN	K-Nearest Neighbour
LC	Local Curvature
LDC	Local Downward Curvature
LiDAR	Light Detection and Ranging
LSF	LS Factor
LUC	Local Upward Curvature
MBI	Morphometric Balance Index
MLR	Multiple Linear Regression
MPI	Morphometric Protection Index
MRN	Melton Ruggedness Number
MS	Mid Slope
MSE	Mean Square Error
MSFI	Multi Source Forestry Inventories
OA	Overall Accuracy
OOB	Out Of Bag
PA	Producers Accuracy
PC	Profile Curvature
PLS	Partial Least Squares
SD	Standard Deviation
SDELV	Standard Deviation of Elevation
SDPC	Standard Deviation of Profile Curvature
SDSLP	Standard Deviation of Slope
SL	Slope Length
SLP	Slope

SSP	Surface Specific Points
TCI	Terrain Curvature Index
TPI	Topographic Position Index
TRI	Terrain Ruggedness Index
TSC	Terrain Surface Convexity
TWI	Terrain Wetness Index
UA	Users Accuracy
UC	Upslope Curvature
UCA	Upslope Contributing Area
VD	Valley Depth
VRM	Vector Ruggedness Measure

CHAPTER ONE

GENERAL INTRODUCTION

1.1 Introduction

Over the past few years, there has been a significant increase in the expanse of areas that has been used for the establishment of eucalypt forest around the world (de Moraes Goncalves *et al*., 2004). In South Africa, approximately 1.35 million hectares (ha) (0.5%) of land is used in the commercial industry alone with eucalypt making up the largest of monoculture forests due to its high growth rate (Shackleton *et al*., 2007). Commercial forestry therefore, holds a high intrinsic value and makes up one of the largest economic sectors (Shackleton *et al*., 2007). For this reason, research pertaining to forest productivity has always been of high interest for forests that are managed for commercial ventures.

Most often, the data that is gathered for research and operational purposes take the form of multi-source forest inventory (MSFI) that include variables such as height, volume, and diameter breast height (DBH) amongst others. Traditionally, all forest structural and terrain measurements are captured manually using hand-held/ mobile equipment in the field (Hyde *et al*., 2006). They also employ expensive and time-consuming sampling methodologies that requires efficient personnel (Tesfamichael *et al*., 2010a). Whilst field assessments are often described as the most accurate method to capture data, it has many disadvantages, this is especially so when working with large field plots or large areas that spans over hundreds of hectares such as within commercial plantations (Hyde *et al*., 2006). Many a time, it limits mapping assessments to be carried out at fine scales, to which detailed information of the area of interest is not acquired. In South Africa, similar procedures are applied for gathering of the required structural and terrain measurements. However, detailed information that covers all areas of interest is particularly important for commercial plantation forests and particularly for applications that require detailed accurate representations of the terrain as South African forests often consist of highly heterogenous landscapes.

Remote sensing has been used for various application fields within the last few decades. Within the forestry sector, researchers have seen the benefits of alternative technologies such as that of remote sensing for the extraction of MSFI data for variables such as height stratification and volume yields with high success rates (Tesfamichael *et al*., 2010a; Tesfamichael *et al*., 2010b; Järnstedt *et al*., 2012). Over the past decade alone, advances in remote sensing technologies have been plentiful and remote sensing has been crucial at addressing challenges that traditional data collections methods cannot handle. For instance, it is well established that is not always feasible to gather forest inventory data through manual operations when dealing with areas that are over great spatial scales (Wulder, 1998). For this reason, remote sensing becomes advantageous as it offers spatially consistent data sets which covers larger areas, it provides high spatial detail and is more efficient for applications that require higher temporal resolutions (Wulder, 1998). Remote sensing technologies also become highly promising as it can be used to provide synoptic views of areas that are otherwise inaccessible and which would result in breaks of the spatial data collected if it cannot be gathered.

Multispectral remotely sensed data sources have provided great strides within the forestry sector and have demonstrated high capability for applications related to structure such as for the extraction of detailed forest information (Wulder, 1998). However, in the twenty first century researches pertaining to forestry and terrain have predominantly shifted to the utilisation of optical remote sensed data sources. Optical remotely sensed data sources such as the use of aerial imagery, Very High Spatial Resolution (VHR) imagery contain spatial resolutions ranging from up to 10 cm or between 2m to 3m resolutions, have been incorporated to attain MSFI data (Tuominen and Pekkarinen, 2005). Aerial imagery is highly favoured as the mixed pixels that occurs within this data is lower than that of multispectral imagery and therefore can provide relatively accurate estimates of forest stand characteristics (Tuominen and Pekkarinen, 2005). Despite this advantage noted complications does exist with this data source, for example in aerial imagery a single pixel does not represent a forest stand, as opposed to that of multispectral imagery, where stand information can be estimated from a single pixel (Tuominen and Pekkarinen, 2005). In addition, the stand information or information for a single tree must be estimated from a local neighbourhood approach or alternatively convert the pixel size into a larger spatial unit which sometimes lead to loss of detailed information (Tuominen and Pekkarinen, 2005).

Light Detection and Ranging (LiDAR) or Laser Scanning, one recent active optical remote sensing technology has gained immense popularity by forestry researchers for its robust utility for applications within the forestry sector. LiDAR provides two sets of surface models either from its first returns which is used for modelling vegetation canopy and last returns which are used for representing ground surface and which is used predominantly for terrain applications (Wilson, 2012). The advantages of using LiDAR are abundant, for example LiDAR offers high density sampling, it provides a high vertical accuracy and is used to provide highly accurate surface derivates such as a Digital Elevation Model (DEM) and a Digital Terrain Model (DTM)

(Wilson, 2012). Often in remote sensing the two terms are frequently used synonymously however, a DEM refers to a gridded raster of a bare earth ground surface, whereas the DTM is derived to represent the natural geodetic properties of the surface only, and represents ridges and valleys (Ullrich *et al.*, 2007; Wilson, 2012). In a DTM all non-natural surfaces such as buildings and vegetation is subsequently removed and therefore for this reason, DTMs become highly valuable for applications such as for the extraction of terrain roughness.

Terrain roughness can be best described as the topographic variability that exists at a given scale and is an important factor known to deter the productivity of forest (Grohmann *et al.*, 2011). For this reason, accurate information on topographical variability that exists within commercial plantation forests are of equal importance for tasks related to productivity through the associated impacts of aspect, slope, or incoming solar radiation on the plantation health and growth (Maack *et al.*, 2016). At different slopes and aspects within a forest stand, a difference in the incoming radiant energy, light, heat, and moisture contents may exist to which variances in the structural growth of the stand may be exhibited (Bale *et al.*, 1998). Therefore, these derivatives are determined, as areas with steeply sloped terrain or poor aspect are often avoided (Bale *et al.*, 1998). These factors thus have the potential to affect the site productivity of the stand and volume yields.

In addition to productivity levels, the importance of terrain roughness is twofold as it can drastically impact harvest productivity levels as topographic information is a prerequisite for optimal tree felling operations because highly rugged terrain affects the rate to which the operations are conducted when using mechanical methods (Davis and Reisinger, 1990; Visser and Spinelli, 2012). Inaccurate terrain representations can also lead to the destruction of machinery if it is deployed to areas that are dominated by steeped slopes or by terrain that is dominated by large boulders (Visser and Spinelli, 2012). The accurate detection of terrain therefore allows forest managers to have up to date spatial information that can be used to make informed decision pertaining to these operations.

1.2 Motivation for the study

Over the past few decades, remote sensing has exceeded expectations for providing accurate data sources for applications within the forestry sector and researchers have often exhausted their means to demonstrate the importance of how climate, edaphologic factors and mean annual rainfall affect productivity and volume yields within commercial plantation forests. However, thus far little research exists on how land surface parameters (i.e. slope, aspect, roughness) affects commercial plantation growth and quality when using remotely sensed data sources and techniques.

Additionally, most research to date that utilises optical remotely sensed data sources such as LiDAR has focused on using the data as a tool for characterising vertical forest structure attributes such as that of tree and stand height, volume, and biomass (Wulder *et al*., 2008). This study is therefore concerned with LiDAR, and its application for terrain roughness modelling and wishes to promote an interest on the horizontal capabilities associated with LiDAR for determining topographic variability that exist within commercial plantation forests.

The information that is provided from forest inventories are compulsory for informative and effective decision making across various spatial and temporal scales for the management of commercial plantations (Wulder *et al*., 2008). In lieu of sustainable forest management, forest inventories that are up to date and which include terrain variability are required for consistent assessments pertaining to distribution, composition, and condition of the forest resources (Wulder *et al*., 2008).

Once tested this methodology can be applied to South African commercial plantation forest to significantly improve the detection of microscale terrain roughness which would result in more complete MSFI data and therefore lead to more informed decision making and effective management protocols within plantations. One of the main advantages of including terrain information into MSFI data is that the terrain does not experience drastic changes on an annual basis and therefore terrain assessments conducted can be used for long periods or until there has been significant surface disturbance on the terrain.

1.3 Aim

Given the above discussion, the general aim of this study is to determine if LiDAR derived DTM terrain roughness indices can be used to detect microscale terrain variations within commercial Eucalypt plantation forests in KwaZulu-Natal, South Africa.

1.4 Objectives

This study contributes to the current standing of knowledge on LiDAR derived DTM modelling for terrain roughness modelling, however in this study it is applied to detect microscale variability in topography within that of commercial plantation forests in a South African context.

The objectives of this study are:

1. To model terrain roughness using high resolution LiDAR derived DTM for commercial plantation forests
2. To determine which roughness indices can accurately detect microscale terrain roughness within commercial plantation forests of South Africa
3. To evaluate if terrain roughness has significant influence on structural variables (i.e. tree height, diameter breast height and volumes) when using LiDAR derived DTM derived roughness indices for predictive modelling
4. To determine if the random forest machine learning algorithm can be used for predictive modelling of terrain variables

1.5 Study area

This study utilises two different commercial plantation forests, Riverdale and Comrie, located within Pietermaritzburg, KwaZulu-Natal, South Africa. The average altitude for the Riverdale plantation is 1190 m and the terrain is characterised by low mountains and undulating hills. The average altitude for the Comrie plantation ranges between 70 m and 650 and displays high levels of heterogeneity, as the study site is composed of highly dissected low undulating mountains, undulating hills, lowlands, and plains. The main land use in both Riverdale and Comrie is plantation forests. A visual representation of the study areas can be seen in figure 1.

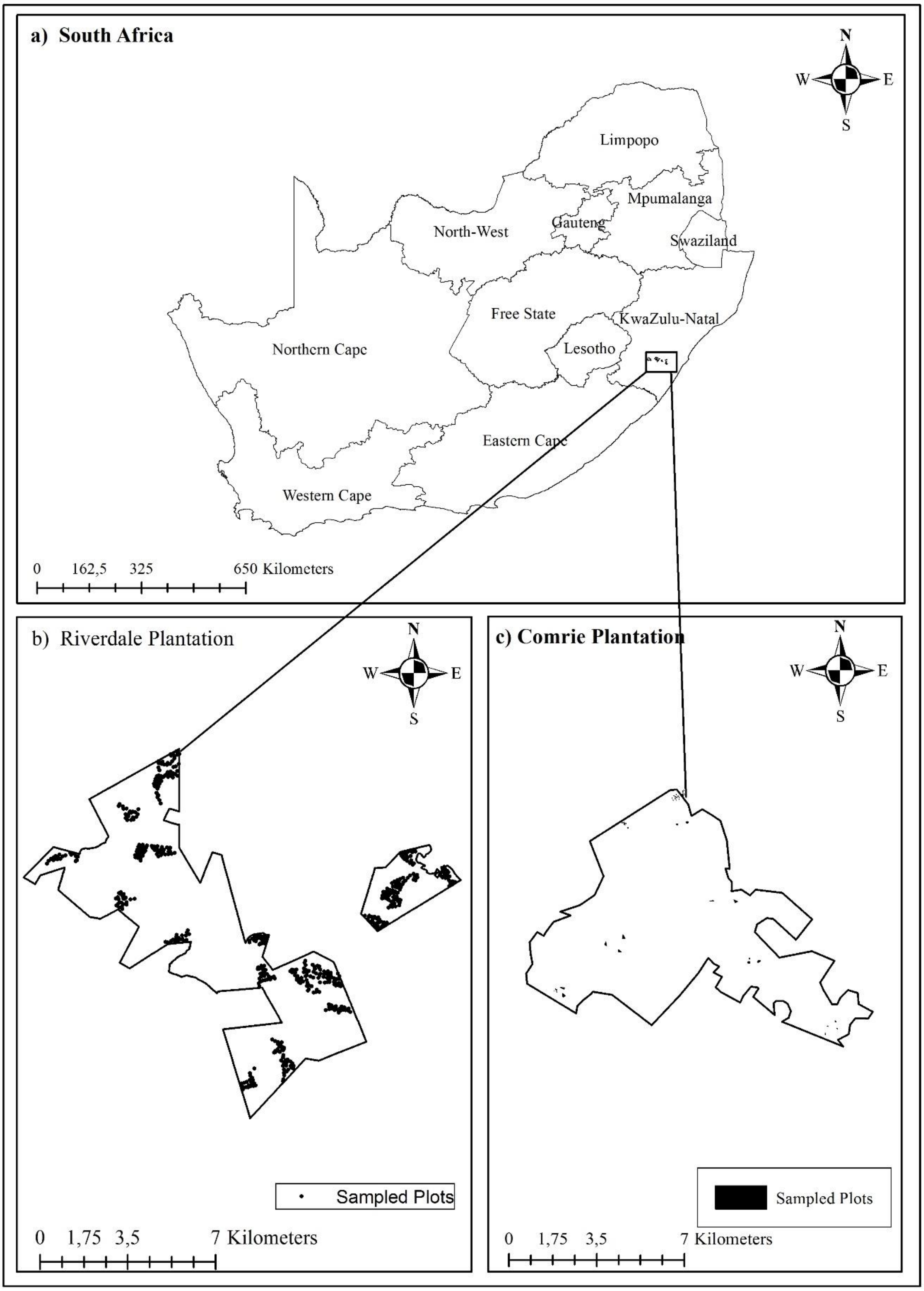


Figure 1: Location of the study areas: a) South Africa; b) Riverdale Plantation and c) Comrie Plantations

1.6 Outline of the thesis

The thesis contains four chapters, two of these chapters were prepared in a peer-reviewed publication format with the intention of submitting to peer-reviewed journals. The title of the paper is therefore mentioned at the beginning of each of the two subsequent chapters.

Chapter one deals with the general introduction for this research, the motivation for the study undertaken, and the aim and objectives of the study are also provided.

In chapter two, modelling the effect of terrain variability on even aged Eucalyptus species volume using LiDAR derived DTM variables are investigated. In this study 32 terrain variables are determined and modelled against structural variables (volume, heights, and diameter breast heights) by using machine learning and particularly the random forest algorithm.

In chapter three, the potential of eight terrain toughness indices extracted from a LiDAR derived DTM data for detecting terrain roughness using a supervised random forest classification method is then investigated. An Analysis of Variance is also conducted to determine the significance level of each terrain indices derived for detecting terrain roughness.

Chapter four, provides a conclusion that synthesises the findings of the two papers provided and of the overall research conducted.

CHAPTER TWO

Modelling the Effect of Terrain Variability in Even-aged Eucalyptus Species Volume using LiDAR-derived DTM Variables.

This chapter is based on:

Munsamy R., Ismail, R., and Gebreslasie, M. (2017) Modelling the Effect of Terrain Variability in Even-aged Eucalyptus Species Volume using LiDAR-derived DTM Variables. *In preparation.*

Abstract. Reliable and accurate multi-source forest inventory measurements that include vertical and horizontal information can assist commercial forest managers in making informed decisions pertaining to forest services such as tree species selection, estimating volume yields, harvesting planning, as well as the designing of silvicultural protocols within stands of commercial plantations. Thus, the aim of this study is to characterise the variation in forest structural attribute measurements of *Eucalyptus* species based on terrain variability derived from a Light Detection and Ranging (LiDAR) dataset. In this study 32 terrain variables at five different spatial resolution were computed from a LiDAR-derived digital terrain model (DTM). Field data were collected for 502 plots within the study area in Richmond, KwaZulu-Natal, South Africa. A Random Forest (RF) regression statistical technique was applied to model the effect of terrain on forest structural variables such as volume, dominant tree height (HtD), mean tree height (Htm), and diameter breast heights (DBH). The results from the random forest regression showed that a single resolution analysis returned statistically insignificant relationship between most of the computed spatial resolutions and forest structural attribute. Only two of the spatial resolutions i.e. 1 m and 7 m produced statistically good relationship with HtD of a young *E.dunni* and Htm of a mature *E.dunni* returned R^2 of 0.70 and 0.68 and their RMSE was 1.24 m and 2.33 m respectively. Whereas multi resolution analysis produced promising results for example for young *E. grandis* stands, the RF model predicted high significance for HtD and Htm with a R^2 value of 0.98 and 0.80, and their RMSE was 1.0023 m and 1.80 m respectively. Variable importance indicated that the incoming solar radiation terrain variable is the most significant variable for modelling forest structural variability, especially for dominant tree height (HtD). The findings from this study indicate that the terrain variable incoming solar radiation that is extracted from high resolution LiDAR-derived DTM data is particularly useful for height stratification within the plantation forests of South Africa.

Keywords: Light Detection and Ranging, Terrain variables, Random forest, regression, volume, tree height

2.1 Introduction

Terrain derived variables such as slope (Wilson and Gallant, 2000), aspect (Grohmann, 2015) and local curvature (Freeman, 1991) impact productivity levels within plantations (Maack *et al*., 2016). Therefore, by modelling terrain variables, additional information related to forest structural attributes can be acquired. Several studies have demonstrated a statistically significant relationship between the variables derived from LiDAR and forest inventory measurements (Tesfamichael *et al.*, 2010a; Tesfamichael *et al.*, 2010b; Van Leeuwen and Nieuwenhuis, 2010; Järnstedt *et al.*, 2012; Jakubowski *et al.*, 2013). However, only a limited number of studies has investigated the utility of terrain variables derived from LiDAR to predict forest inventory measurements. An increase in the detail of the knowledge and understanding of the role of terrain variability (specific to plantation sites) would lead to the derived information being used to successfully manipulate and homogenise stands (Li *et al.*, 2014), thereby increasing site productivity and decreasing the heterogeneity associated with large commercial plantations.

In a recent study Ediriweera *et al*. (2016) aimed to characterise the variation in the structural attributes of vegetation with relation to terrain by calculating the Terrain Wetness Index (TWI), potential solar insolation, slope and elevation using a LiDAR dataset for both an open canopy eucalypt forest and a closed subtropical rainforest within Australia. A general linear model approach was employed to examine the relationship of structural attributes and terrain. The results showed that maximum over story height decreased when there was an increase in potential solar radiation in the eucalypt forest (R^2=0.45) and showed that eucalypt forests were more prone to variations in terrain than subtropical rainforests (Ediriweera *et al.*, 2016).

In another study Saremi *et al*. (2014a) employed the use of a mixed linear model to investigate the relationship between terrain variables, i.e. slope and aspect, derived from LiDAR against the mean tree height (Htm) of radiate pine (*Pinus radiate. D.Don*) for two even-age (9 and 34 year-old) sites. The mixed linear model used in this study was based on one continuous dependent variable with several explanatory variables and showed that the derived height estimates were highly correlated with field heights (R^2=0.90 and RMSE =0.66) for 9-year sites and (R^2=0.87 and RMSE=1.49) for the 34-year sites (Saremi *et al.*, 2014a). The results obtained from this study also showed that taller trees were present in low slopes with southerly aspects, whilst short trees were found in steep slopes with northerly aspects.

The same authors also applied a mixed linear model to quantify the relationship between the DBH of these trees and height classes along with the terrain factors (slope and aspect) (Saremi *et al.*, 2014c). The results obtained from that study showed that a greater DBH was found in gentle slopes with southerly aspects (Saremi *et al.*, 2014c). Further investigations by Saremi *et al.*, (2014b) found that micro-scale variations of DBH and Htm could be quantified, based on incoming solar radiation. The results were reported as (R^2=0.58) for height and (R^2=0.60) for DBH for mature stands; and (R^2=0.58) and (R^2=0.60) for the young stands respectively (Saremi *et al*., 2014b). Trees that had larger DBH and which were taller were found in areas with lower incoming solar radiation, with the authors stating that variation exists within stands of the same age category (Saremi *et al*., 2014b). These studies have introduced and exampled the benefits of utilizing LiDAR and its derivatives within the forestry sector. For more information see Popescu *et al*., 2003; González *et al.*, 2008; Wulder *et al*., 2012).

Many of these studies apply statistical techniques to quantify these measurements. Regression techniques including Artificial Neural Networks (ANN) (Svetnik *et al.*, 2003); Multiple Linear Regression (MLR) and k-nearest neighbour (KNN); Partial Least Squares (PLS) (Duncanson *et al*., 2015); Support Vector Regression (SVM) (Jakubowski *et al*., 2013); Bayesian Model Averaging (BAM) (Verkerk *et al*., 2015); Generalized Addictive Model (GAM) (Maack *et al*., 2016) and Random Forest (RF) (Aertsen *et al*., 2012) are examples of statistical techniques that have been utilised to explain the relationships between forest structural attributes and remote sensing data. However, whilst ANN, MLR, KNN and many more of these methods provide relatively high prediction performance, the drawback in using these methods is that they are not able to deal with high-dimensional data without performing dimension reduction (Svetnik *et al*., 2003).

RF for regression, however, has been a method consistently favoured by the remote sensing community. For example, Yu *et al*. (2011) and Nurminen *et al*. (2013) utilised RF to model forest structural attributes using aerial imagery and Lidar data respectively. Yu *et al.* (2011) successfully used the RF algorithm to predict forest structural attributes (i.e. tree height, DBH and stem volume) within a boreal forest in Southern Finland. The independent variables included 26 tree features that were derived from the LiDAR data, whilst the dependent variables included the forest structural attributes. The results derived showed high correlations for the observed and predicted height (R=0.93 and RMSE=10.03%), DBH (R=0.79 and RMSE=21.35%) and volume (R=0.87 and RMSE=45.77%). Similarly, Nurminen *et al*. (2013) used a RF approach to predict Htm, DBH and volume for plots extracted from both LiDAR point clouds and aerial photography (Nurminen *et al*., 2013). Such results show that LiDAR-derived forest attributes are more accurate than information derived from digital aerial photography. The results also showed high correlations for mean height (R=0.98 and RMSE=0.97), DBH (R=0.94 and RMSE=2.16m) and volume (R=0.93 and RMSE=37.58m^3/ha).

Therefore, in the last decade alone RF has gained widespread acceptance by researchers involved with remote sensing analysis due to its ability to successfully analyse non-parametric data that may exhibit high levels of noise. RF is able to provide a higher prediction accuracy than other machine learning approaches with a faster computation rate (Ismail and Mutanga, 2011). As an ensemble technique, RF consist of unpruned regression trees that are created using bootstrap sampling for training data (Svetnik *et al.*, 2003), and by growing a large number of trees it is able to keep prediction bias low (Prasad *et al.*, 2006; Cutler *et al.*, 2007; Dye *et al.*, 2012). It is also consistently used as a random feature selection in tree induction and provides measures of variable importance (Svetnik *et al.*, 2003).

Based on the studies reviewed we therefore acknowledge that LiDAR can provide valuable forest structural and terrain information, and machine learning statistical techniques such as RF can be used with highly accurate results for predictive modelling. Hence, in this study RF is adopted to access the relationship between LiDAR terrain variables and forest structural attributes. To our knowledge, the question that remains is: Does the terrain have an impact on forest structural attributes within the *Eucalyptus* plantation environments of South Africa? The aim of this study is therefore to explain the variability in forest structural attributes such as height, pulpwood volumes and DBH across terrains when using LiDAR-derived terrain variables.

2.2 Materials and Methods

2.2.1 Study Area

The study was conducted in Sappi's Riverdale plantation. The extent of the plantation is 2503 hectares (ha). The Riverdale plantation is composed of two areas with compartment blocks found near the west and east of the town of Richmond, which is in the Midlands of KwaZulu-Natal, South Africa. The area of interest within the plantation spans approximately 6 ha. The average altitude for the plantation is 1190 m and the terrain is characterised by low mountains and undulating hills. The geology of the region is dominated by mudstones, sandstones, tillite, amphobilite and basalts. The average air temperature is 16. 1^{o} C. The mean annual precipitation reported for the region is 916 mm and the mean annual runoff for the plantation is 143 mm. The Riverdale plantation comprises of areas that are dominated by the Ngongeni veld of Natal (40%), Highveld Sourveld (30%) and Southern Tall Grassland (20%) veld types. The soils found in the plantation are composed mostly of sandy-clay and sandy-clay loams.

The dominant land use in the study area is commercial forestry consisting of *Eucalyptus* species which provide affordable direct raw materials for industries producing pulp for paper and packing, and timber for commercial processing and for the production of wood chips (Hassan, 1999). The *Eucalyptus* stands in the study area are aged between 2 and 10 years in the plantation, and *Eucalyptus* is the main species grown due to its growth rate being favourable in KwaZulu-Natal province (Godsmark, 2013). *Eucalyptus* species are known for their fast-growing stands and the trees are therefore planted as clones from seedlings. These stands are established at 1667 trees per ha, as per the pulpwood regime, and are harvested between 6 and 7 years (Dube *et al*., 2015). This study area consisted of *E. grandis* and *E. dunni* stands.

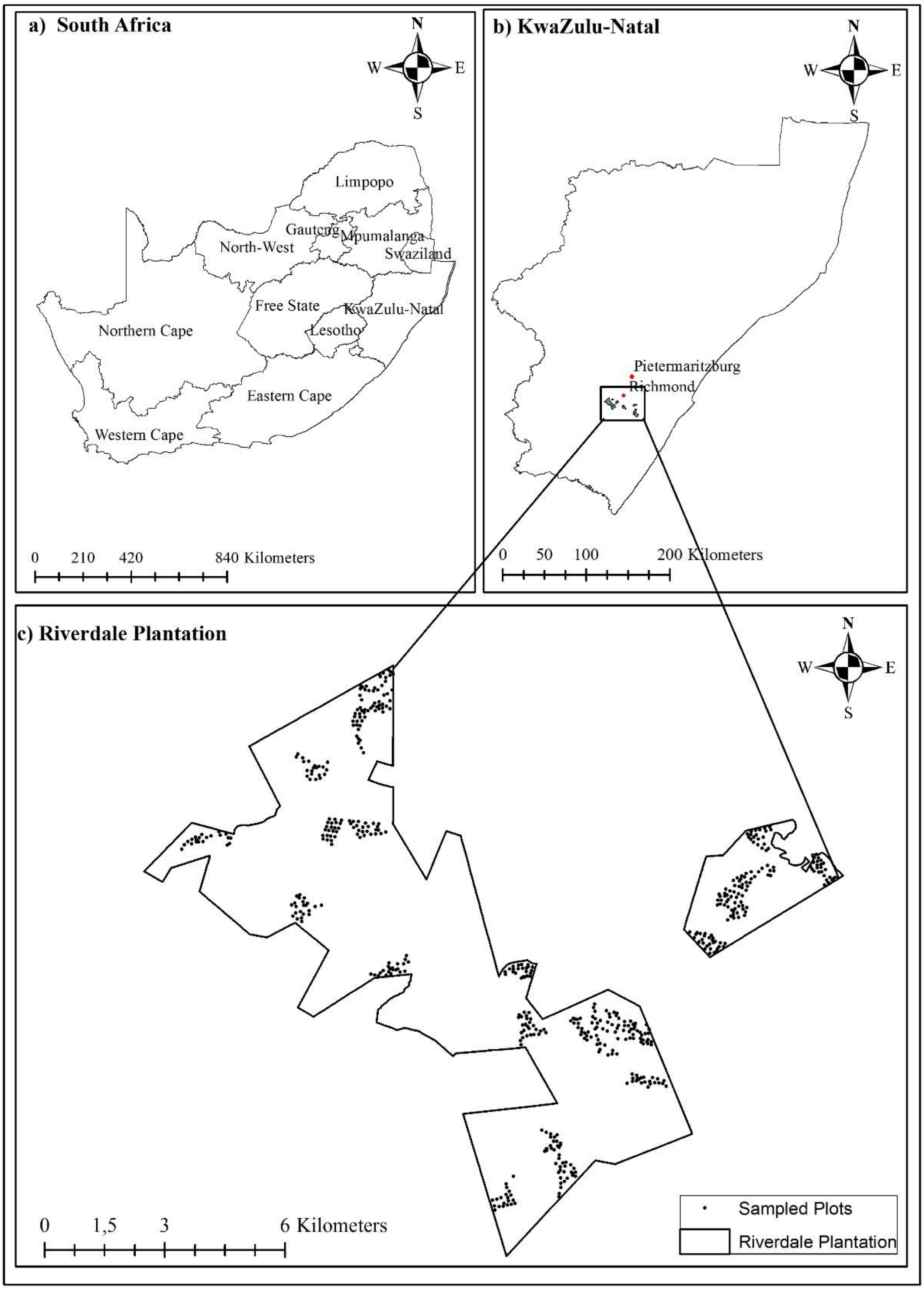


Figure 2: Location of the study area: a) South Africa; b) KwaZulu-Natal and c) Riverdale Plantation

2.2.2 LiDAR Dataset

The LiDAR data was acquired by Land Resources Institute (LRI). The LiDAR surveys were conducted between the 15th and the 22nd of March 2014 at the Riverdale plantation. The surveyed point cloud data was used to create a very high resolution DTM. The data was then projected to the Transverse Mercator with a Gauss Conformal projection. The central meridian was 31 and the datum used was Hartebeeshoek 94. The flight and sensor instrument parameters used for the collection of the LiDAR data are presented in table 1.

Table 1: LiDAR flight and sensor instrument parameter

LiDAR Survey Parameters	**Unit**	
Altitude	m AGL	800
Flight speed	kt	100
Scan angle	º	25
Scan swath width	M	324.3
Scan overlap	%	50
Scan rate	Hz	52
Laser pulse rate	Hz	128000
Laser pulse density	pulses/ m^2	4

2.2.3 Field Surveys

Field surveys were conducted from the 12th to the 22nd May 2014; a total of 502 plots spanning over 27 compartments at the Riverdale plantation were covered. The following structural attributes were included in this study from the inventory surveys: volume, mean dominant height (HtD), mean height (Htm) and diameter at breast height (DBH). A Global Positioning Device (GPS) was used in the field to survey circular plots at a 10-m radius using a grid-based systematic sampling technique. DBH was measured using a Vertex IV laser instrument and the tree heights were measured using a Haglof Digitech Calliper Methods.

Further, to be able to consider the species and age variation in this study, the plots were partitioned into datasets based on species and age. *Eucalyptus grandis* was separated into age categories of young (3-6, n = 151) and mature (7-10, n = 137), while *Eucalyptus dunnii* included young (2-5, 104) and mature (6-9, n = 110). The descriptive statistics for the age categories are provided in Table 3.

2.2.4 DTM developing

2.2.4.1 Extracted Terrain Variables

The various terrain variables were calculated, using the nearest neighbour re-sampling technique, at the following spatial resolutions: 3 m x 3 m, 5 m x 5 m, 7 m x 7 m and 9 m by 9 m. A complete list of terrain variables that were calculated is provided in table 2 below. Spatial analysis and map algebra tools were then used to extract the zonal statistics for each of the terrain variables (n = 32) at a plot level. The final datasets were composed of the various terrain variables at the five re-sampled spatial resolutions.

Table 2: Terrain variables modelled in this study

Variable		**Description**	**Reference**
Direct Insulation	DIRECT	Direct solar radiation received	Lukovic *et al*. (2015)
Diffuse Insulation	DIFFUSE	Solar radiation received after scattering	Saremi *et al.* (2014b)
Curvature Classification	CC	Planimetric curvature ratio	Drăguţ and Blaschke (2006)
Convergence Index	CI	Uses aspect to determine flow convergence and divergence	Wilson and Gallant (2000)
Down Slope Distance Gradient	DDG	Quantifies local drainage patterns on topography	Hjerdt *et al.* (2004)
Flow Accumulation	FA	Measures upstream catchment area for a cell	Navarro-Cerrillo *et al*. (2014)
LS Factor	LSF	Determines slope length based on the Universal Soil Loss Equation	Boehner (2006)
Mass Balance Index	MBI	Measures geomorphographic relief	Möller *et al.* (2008)
Melton Ruggedness Number	MRN	Measures basin relief	Melton (1965)
Slope Length	SL	Determines effects of erosion on slope	Navarro-Cerrillo *et al.* (2014)
Slope Variability	SV	Measures difference in relief	(Popit and Verbovšek, 2013)
Slope	SLP	Measure of steepness	Wilson and Gallant (2000)
Aspect	ASP	Direction of slope	Grohmann (2015)
Profile Curvature	PC	Rate at which slope changes	Wilson and Gallant (2000)

Surface Specific Points	SSP	Detects specific points from parallel processing of elevation	Hutchinson (1989)
Standard Deviation of Elevation	SDELV	Standard deviation of elevation from the mean	Grohmann *et al*. (2011)
Standard Deviation of Slope	SDSLP	Standard deviation of slope from the mean	Grohmann *et al*. (2011)
Terrain Surface Convexity	TSC	Measures cells having positive convexity	Iwahashi and Pike (2007)
Morphometric Protection Index	MPI	Determines immediate surrounding and how relief is protected	Olaya and Conrad (2009)
Real Surface Area	RSA	Calculates real area of slope	Olaya and Conrad (2009)
Topographic Position Index	TPI	Measures relative topographic slope position	Guisan *et al.* (1999)
Terrain Ruggedness Index	TRI	Represents a change in the sum of elevation	Riley *et al.* (1999)
Topographic Wetness Index	TWI	Measures hydrological conditions within a site relatively	Sørensen and Seibert (2007)
Local Curvature	LC	Calculates sum of the gradients to	Freeman (1991)
Upslope Curvature	UC	its neighbouring cells	Freeman (1991)
Local Upslope Curvature	LUC	Distance of weighted average of local curvature	Freeman (1991)
Downslope Curvature	DC	Calculates the local curvature on flow direction as a sum of neighbour cells that are facing upwards	Freeman (1991) Freeman (1991)
Local Downslope Curvature	LDC	Calculates the local curvature on flow direction as a sum of neighbour cells that are facing downslope Calculates local curvature as a sum of neighbour cells	
Vector Ruggedness Measure	VRM	Measures roughness around a neighbourhood	Sappington *et al.* (2007)
Midslope	MS	Position of slope	Florinsky *et al.* (2002)
Valley Depth	VD	Vertical distance to channel base	Schmidt and Hewitt (2004)
Terrain Curvature Index	TCI	Measure of terrain shape	Park *et al.* (2001)

2.2.5 Regression Analysis

2.2.5.1 Random Forest

RF was implemented in this study using libraries found in R statistical software (R Development Core Team, 2008). RF has been described as a method that is easy to implement as the user is required to input only the number of trees to be split (*ntree*) and the number of variables (*mtry*) to be used in the process. Each decision tree in the algorithm is then responsible for casting a unit vote for class that is the most popular at unit **x** (Breiman, 2001). In order to increase the diversification of decision, trees random forest makes use of a bootstrap aggregating method using one third of the data to ensure the trees grow from different subsets within the training data (Rodriguez-Galiano *et al.*, 2012). These bootstrap samples are referred to as out-of-bag (OOB) samples. The OOB data that were not used during the training process is then used for prediction, as it provides an unbiased assessment of accuracy as outlined by Breiman (2001), Rodriguez-Galiano *et al.* (2012) and Kulkarni and Sinha (2013).

2.2.5.2 Multi-resolution analysis

In this section, the following was discovered:

i. The terrain variables that were calculated at the various spatial resolutions were aggregated and used in 8 RF models to predict the forest structural attributes of the two Eucalypt species.
ii. To determine the optimal set of variables that could best explain forest structural attributes for the young and mature Eucalyptus species, a backward feature selection approach was used to reduce the number of input terrain variables that could best explain the variation in the various forest structural attributes.

2.2.5.3 Random Forest Variable Importance

Variable importance can be described as a measurement used to decide how much of an influence a variable has on the predictive accuracy of a model (Treeratpituk and Giles, 2009). In RF two types of variable importance measures are often used, a Gini importance and a permutation importance (Treeratpituk and Giles, 2009). According to Grömping (2012) the Gini importance method may result in bias due to the average impurity reduction associated with this technique for regression trees. Breiman (2001) suggests the permutation method, which has been widely adopted. In this method, for each tree *t* in the RF, the OOB mean squared error (MSE) is computed by averaging the squared deviations of the OOB responses for the predictor variables (Breiman, 2001; Grömping, 2012). In this study, RF variable importance is based on the permutation accuracy method.

2.2.5.4 Random Forest Variable Selection

Variable selection becomes important when dealing with multiple predictor variables for prediction, as many predictors may lead to a decrease in model performance. A variable selection method based on RF-recursive feature elimination was adopted. In this method, variables are selected based on their variable importance ranking. All variables are first iterated through the algorithm. The algorithm then drops any variables that do not contribute to the predictive accuracy of the model (Granitto *et al.*, 2006). The algorithm runs until all unnecessary variables are progressively dropped.

2.3 Results

The descriptive statistics (the mean, Standard Deviation (SD), minimum and maximum values) of the plots categorised by species and age from the field inventory assessment are illustrated in table 3 below.

Table 3: Descriptive statistics for the field inventory assessment, sample size n=502

		HtD (m/ha)	Htm (m)	Vol (m^3/ ha)	DBH (cm/ ha)
Young E. Grandis	**Mean**	24.15	20.14	228.16	14.42
	Standard Deviation	3.73	3.02	83.56	2.29
	Minimum	17.46	14.59	77.42	9.55
	Maximum	29.07	25.20	366.49	18.80
Mature E. Grandis	**Mean**	30.95	25.14	321.02	17.55
	Standard Deviation	3.33	2.40	76.35	2.19
	Minimum	26.9	20.27	221.73	13.50
	Maximum	40.03	31.37	590.52	23.02
Young E. Dunni	**Mean**	14.65	13.27	74.93	10.49
	Standard Deviation	3.01	2.70	32.35	1062
	Minimum	8.18	6.42	13.05	4.24
	Maximum	19.23	17.69	141.69	14.3
Mature E. Dunni	**Mean**	23.23	19.37	195.62	13.70
	Standard Deviation	4.74	3.40	71.14	1.96
	Minimum	15.53	13.54	54.91	9.30
	Maximum	32.78	27.17	344.47	19.28

2.3.1 Individual Spatial Resolution Analysis

In this section, the results of the various spatial resolutions are reported separately to determine if a specific spatial resolution (i.e. 1 m x 1 m, 3 m x 3 m, 5 m x 5 m, 7 m x 7 m or 9 m x 9 m) could best explain the variation in the forest structural attributes for the young and mature Eucalyptus species. In total, 80 RF models were developed for the two Eucalypt species that were considered in this study. A graphical representation of certain terrain variables re-sampled to the various spatial resolutions is to be found in figure 3 below.

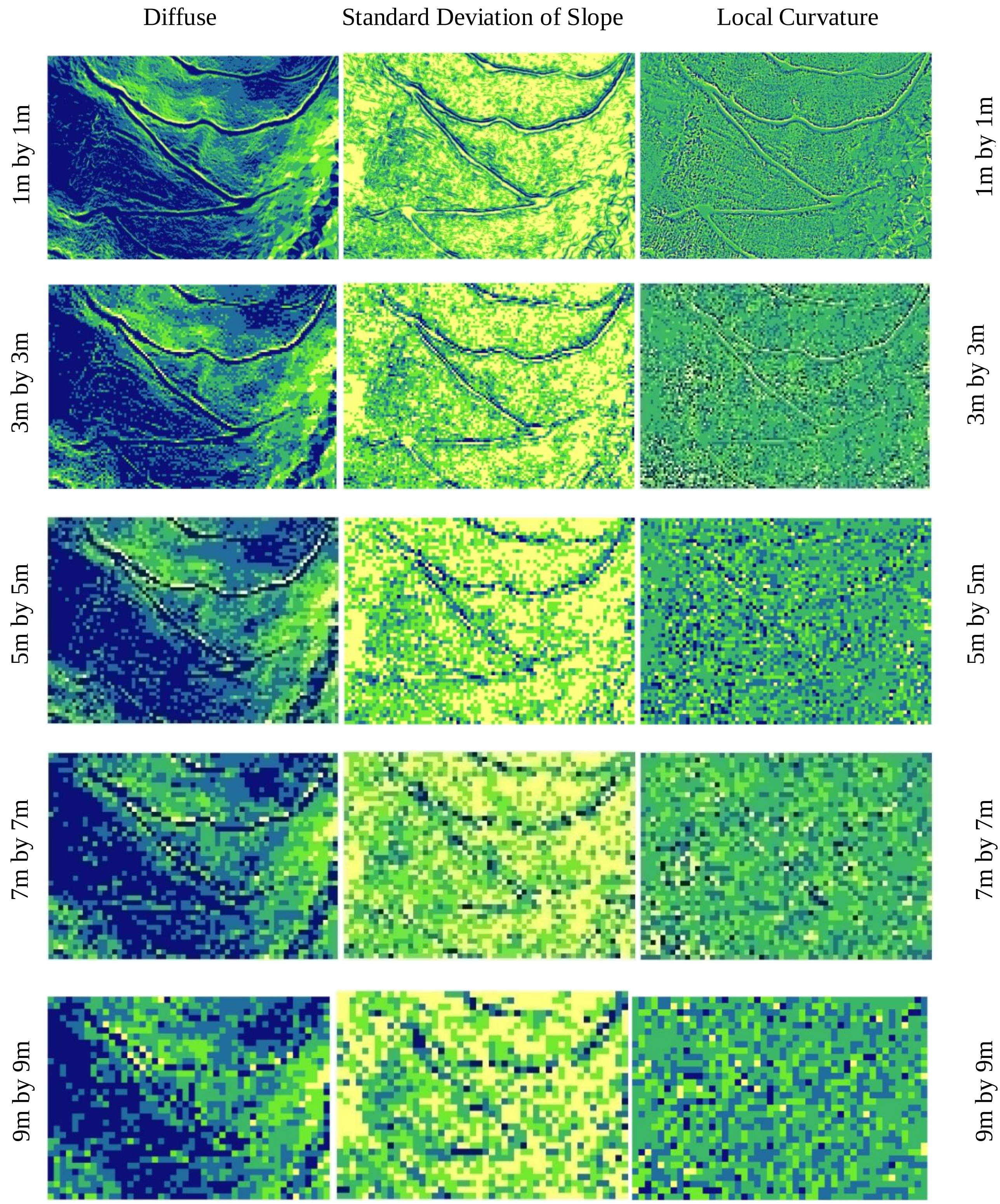


Figure 3: The effect of different spatial resolutions on diffuse, standard deviation of slope, and local curvature.

2.3.1.1 E. grandis

The results for the young (*a*) and mature (*b*) *E. grandis* stands are depicted in figure 4. For the young *E. grandis* stands, the highest predictive accuracy for the HtD response variable was obtained at a 9 m spatial resolution with a R^2 value of 0.58 and a RMSE value of 1.24 m. The highest predictive accuracy obtained for the Htm response variable was obtained at a 9 m spatial resolution with a R^2 value of 0.42 and a RMSE of 0.99 m. For the DBH response variable, the highest predictive accuracy was obtained at the 5 m spatial resolution with a R^2 value of 0.55 and a RMSE value of 0.81 cm. With regards to the volume response variable, the highest predictive accuracy was obtained at a 9 m spatial resolution with a R^2 value of 0.42 and a RMSE of 37.37 m^3/ha.

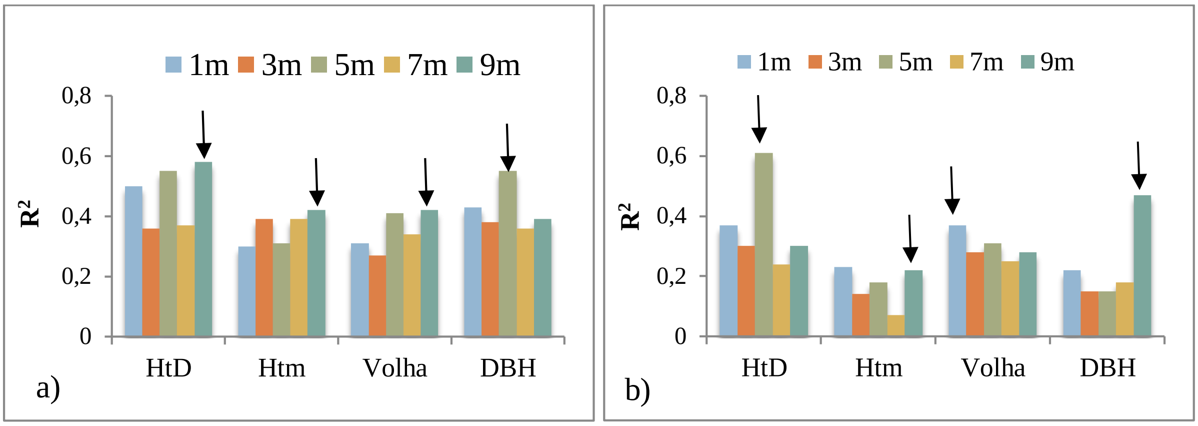


Figure 4: Random Forest predictive accuracies (R^2) obtained for (a) young and (b) mature *E. grandis* stands. The black arrows show the models with the highest predictive accuracies.

The mature *E. grandis* stands produced varied results. For the HtD response variable, the highest R^2 value was reported as 0.31 at a 5 m spatial resolution and a RMSE value of 2.18 m. The Htm response variable yielded the highest accuracy at a 1 m spatial resolution with a R^2 value of 0.23 and a RMSE value of 1.28 m. The DBH model that yielded the highest R^2 value was at a 9 m spatial resolution with a value of 0.47 and a RMSE value of 0.77 cm. Volume yielded the highest predictive accuracy at a 1 m spatial resolution with a reported R^2 value of 0.37 and a RMSE value of 29.16 m^3/ha.

2.3.1.2 E. dunnii

The results for the young (*a*) and mature (*b*) *E.dunnii* stands are depicted in figure 5. For young *E.dunnii* stands the highest predictive accuracy for the HtD response variable was obtained at a 1 m spatial resolution with a R^2 value of 0.7 and a RMSE value of 1.24 m. The highest predictive accuracy obtained for the Htm response variable was obtained at a 7 m spatial resolution with a R^2 value of 0.68 and a RMSE of 1.15 m. For the DBH response variable, the highest predictive accuracy was obtained at the 1 m spatial resolution with a R^2 value of 0.18

and a RMSE value of 1.88 cm. The highest predictive accuracy was obtained at a 1 m spatial resolution with a R^2 value of 0.29 and a RMSE of 18.53 m^3/ha for the volume response variable.

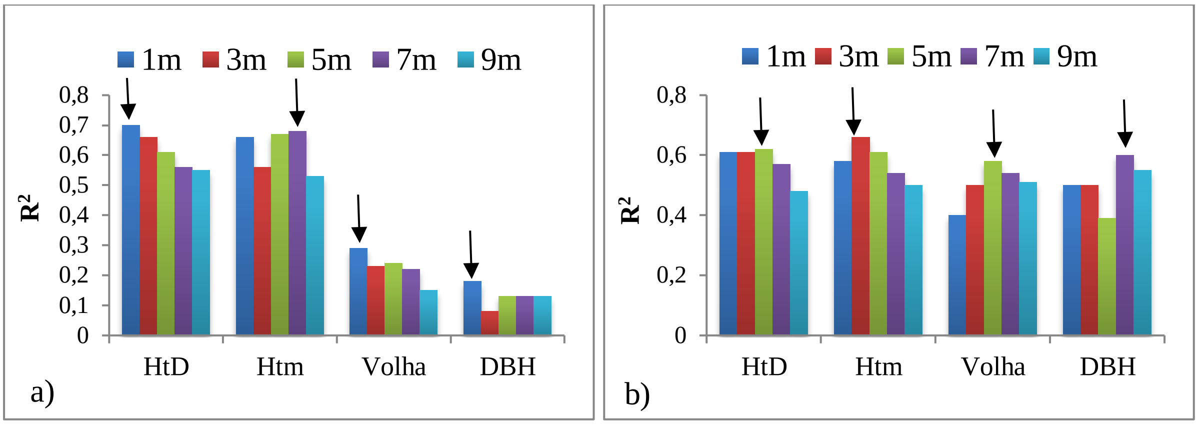


Figure 5: Random Forest predictive accuracies (R2) obtained for (a) young and (b) mature *E. dunni* stands. The black arrows show the models with the highest predictive accuracies.

For mature *E. dunnii* stands, the highest R^2 value for the HtD response variables was reported as 0.62 and a RMSE value of 2.73 m at a 5 m spatial resolution. The Htm response variable yielded the highest accuracy at a 3 m spatial resolution with a R^2 value of 0.66 and a RMSE value of 1.48 m. The DBH model that yielded the highest R^2 value of 0.60 and a RMSE value of 0.71 cm occurred at a 7 m spatial resolution. The volume model yielded the highest predictive accuracy at a 5 m spatial resolution with a reported R^2 value of 0.58 and a RMSE value of 41.08 m^3/ha.

2.3.2 Multi-resolution analysis

2.3.2.1 E. grandis

For young *E. grandis* stands (table 5a), the RF model predicted high significance for HtD and Htm with a R^2 value of 0.98 and 0.80, and their RMSE was 1.0023 m and 1.80 m respectively. The RF model predicted moderate significance for volume with a R^2 value of 0.56 and a RMSE of 67.80 m^3/ha, while DBH was predicted with the lowest R^2 value of 0.48 and a RMSE value of 1.93 cm using the RF model.

In most cases, RF models for mature *E. grandis* stands (table 5a) showed poorer results than for the younger stands. The highest predictive accuracy was a R^2 value of 0.29 and a RMSE of 2.42 m/ha for HtD. Volume yielded a R^2 value of 0.27 and a RMSE of 55.67 cm^3/ha. DBH gave rise to a R^2 value of 0.22 and a RMSE of 1.97 cm. The lowest accuracy reported was for the Htm model, as a R^2 value of 0.13 with a RMSE of 2.09 m.

2.3.2.2 *E. dunni*

For young *E. dunni* stands (table 5b), the RF model also predicted high significance for HtD and Htm with a R^2 value of 0.66 and 0.61, and their RMSE was 1.85 m/ha and 1.66 m respectively. The RF predicted moderate significance for volume with a R^2 value of 0.27 and RMSE of 27.78 cm^3/ha, whilst DBH was predicted with the lowest R^2 value of 0.14 and a RMSE of 1.93 cm. In the case of mature *E. dunnii* (table 5b), like young *E. dunnii*, the RF model predicted high significance for HtD and Htm, with a R^2 value of 0.61 and 0.61, and their RMSE was 2.78 m/ha and 2.04 m respectively. The RF predicted moderate significance for DBH with a R^2 value of 0.55 and a RMSE of 1.24 m/ha, while volume was predicted with the lowest R^2 value of 0.44 with a RMSE of 49.57 cm^3/ha.

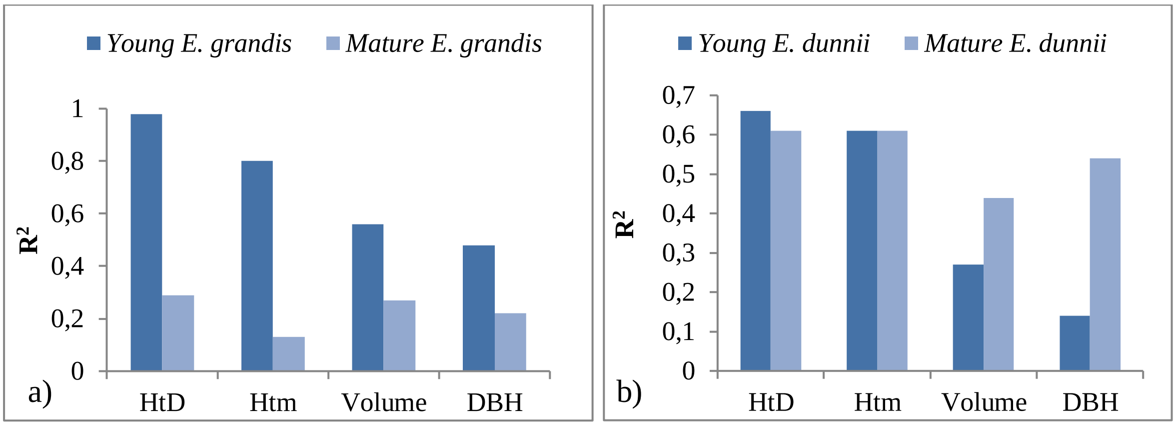


Figure 6: Coefficient of determination (R^2) for (a) young and mature *E. grandis* and (b) young and mature *E. dunni*

2.3.3 Variable Importance

There was no consistent pattern for variables that were ranked important for both young and mature eucalypt species. For young *E. grandis*, the terrain variable diffuse solar radiation was frequently returned as the most important variable, as it is placed in the first and/or second rank in most forest structural attribute predictions. For young and mature *E. grandis* species, direct solar radiation appeared highly significant, as it appeared in the top five ranking predictor variables for HtD, Htm and Volume, to which DBH is the exception. For young and mature *E. dunnii*, the terrain variable diffuse solar radiation was the most frequently returned variable, as it provided the highest-ranking variable for all structural models in these categories. There was also no consistent ranking of important variables after the terrain variable diffuse solar radiation, as the variables ranked important fluctuated for each structural attribute modelled.

The varied importance for the structural variables with the highest R^2 value for each structural attribute for each species is provided in figure 7.

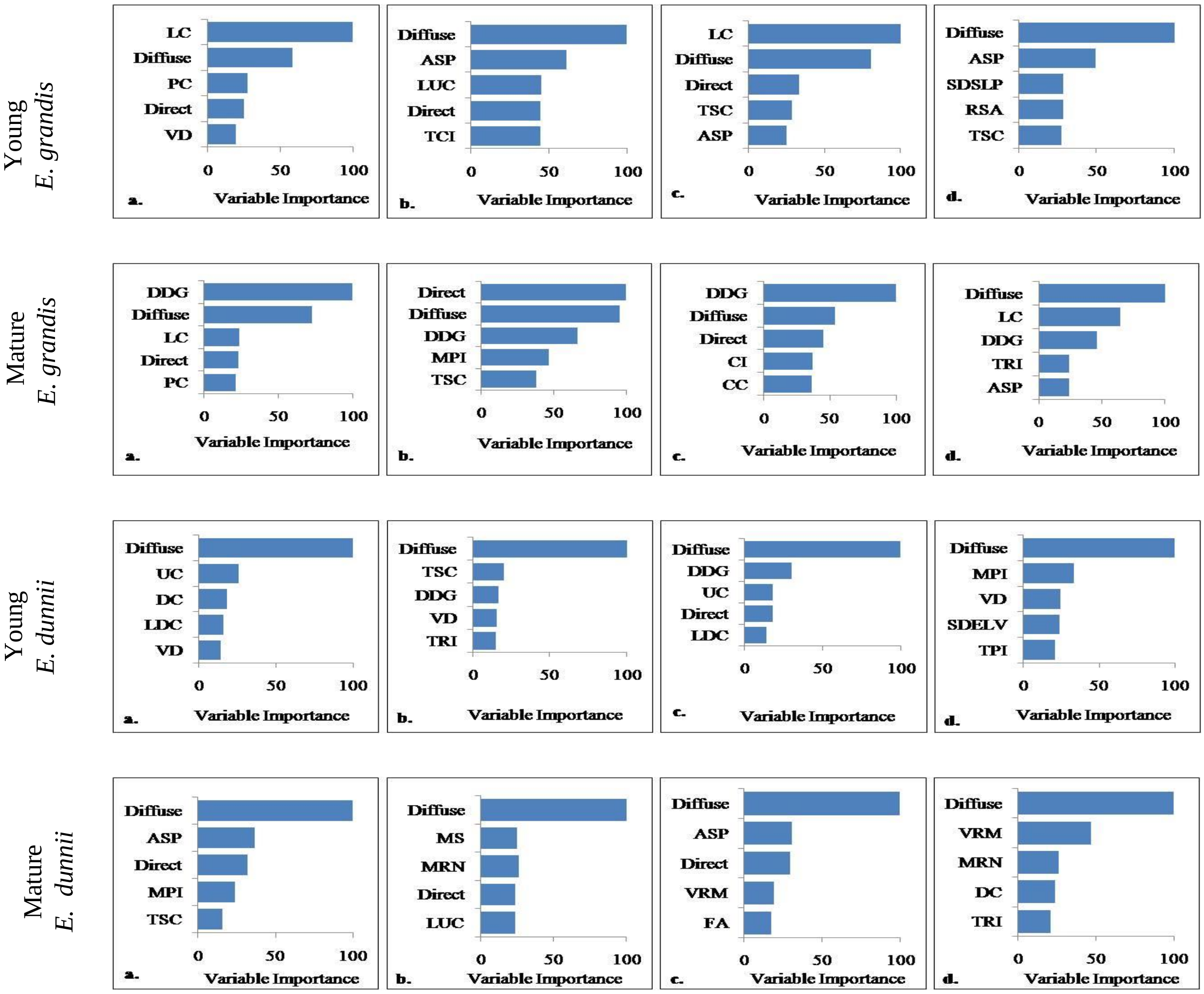


Figure 7: Varied Importance for young and mature *E. grandis* and *E. dunnii* for (a) HtD (b) Htm (c) Volume and (d) DBH

2.3.4 Random Forest Variable Selection

In efforts to improve the predictive accuracy of the structural attributes modelled, a variable selection based on RF-recursive feature elimination was implemented. For young *E. grandis* there was a decrease in R^2 values for HtD models at all spatial resolutions. For Htm there was an increase in predictive accuracy for models at the 1 m and 3 m spatial resolutions. For volume, there was a decrease in all models except at the 1 m spatial resolution. For DBH, only models at the 1 m and 3 m spatial resolution increased in predictive accuracy. However, when modelling mature *E. grandis* all spatial resolutions displayed higher R^2 values, to which the exception was for the DBH model at a 9 m spatial resolution.

For young *E. dunnii*, HtD models showed increases in R^2 values for all spatial resolutions. However, for Htm, Volume and DBH the R^2 values either decreased or remained the same. When modelling mature *E. dunnii* all R^2 values increased for HtD. The models also showed increases in R^2 values for Htm and Volume, whereas the models for DBH showed decreases in the R^2 values obtained. The results obtained for RF variable selection are provided in table 4. R^2 values with an * indicate that there was an increase in predictive accuracy after variable selection.

For young *E. grandis* the highest predictive accuracy was maintained as a R^2 value of 0.98 was reported whilst decreasing the predictor variables to three (figure 6a and table 4). The lowest result obtained in this age category was for DBH, but post-variable selection increased the predictive accuracy to a R^2 value of 0.63 by using only 12 predictor variables. For the mature age category, the highest accuracy was also maintained as a R^2 value 0.29 was reported; the algorithm decreased the predictor variables to nine. Lastly, for Htm, which produced the lowest accuracy in this age group, increased post-variable selection as a R^2 value of 0.30 was reported using 13 predictor variables.

For young *E. dunnii* the predictive accuracy for HtD improved by 1.51% with a R^2 value of 0.67 using only 32 predictor variables. For DBH the results were reported as a R^2 value of 0.20 post-variable selection using 60 predictors. For mature *E.dunnii*, HtD yielded a R^2 value of 0.71, which was an overall 8.20% increase. The algorithm decreased the predictors from 160 to 108. The lowest value obtained for volume was subsequently increased to a R^2 value of 0.48 by reducing the predictors to only 12.

Table 4: Coefficient of determination (R^2) for young and mature *E. grandis* and *E. dunnii* from variable selection

		E. grandis					***E. dunnii***				
		1m	3m	5m	7m	9m	1m	3m	5m	7m	9m
HtD	Young	0.98	0.27	0.53	0.28	0.48	0.70	0.69*	0.68*	0.53	0.59*
	Mature	0.35	0.36*	0.31	0.44*	0.45*	0.74*	0.67*	0.68*	0.69*	0.55*
Htm	Young	0.33*	0.36	0.33*	0.33	0.17	0.59	0.58*	0.68*	0.62	0.52
	Mature	0.25	0.23*	0.18	0.15*	0.22	0.62*	0.67*	0.65*	0.58*	0.62*
Vol	Young	0.34*	0.10	0.30	0.31	0.22	0.31*	0.25*	0.22	0.06	0.15
	Mature	0.40*	0.50*	0.34*	0.45*	0.39*	0.49*	0.53*	0.55	0.57*	0.51
DBH	Young	0.46*	0.35	0.55	0.35	0.38	0.14	0.09*	0.13	0.08	0.14*
	Mature	0.27*	0.28*	0.20*	0.31*	0.30	0.47	0.38	0.36	0.43	0.41

2.4 Discussion

This study investigated the variation in forest structural attributes within young and mature commercial eucalypt plantation sites. Pulpwood volume, HtD, Htm and DBH were modelled in relation to variations in topography by using a machine-learning RF statistical technique. The results obtained from this study indicate that the RF ensemble technique is useful for explaining the difference that exists between explanatory forest structural attributes and terrain-based predictor variables. It was also evident that variations in structural attributes can be primarily attributed to response variables that are associated with solar radiation.

2.4.1 Individual spatial resolution versus multi-resolution analysis

This research applied regression modelling to both individual and multi-resolution spatial approaches to determine if there was an optimal spatial resolution for each structural variable modelled. Furthermore, Navarro-Cerrillo *et al.* (2014) stated that micro-scale terrain variability cannot be accounted for at the micro scales within plantations, since terrain heterogeneity is dominated by both spatial and temporal scales. The results indicated that young and mature *E. grandis* requires a 9 m spatial resolution for predicting variations in terrain for HtD, whereas a 1 m spatial resolution provides the highest predictive accuracies for young and mature *E. dunnii* species.

Individual spatial resolution indicated that there was no spatial resolution that was consistent for predicating pulpwood volume for both eucalypt species. In addition to these trends, it was found that a multi-resolution approach did not improve the predictive accuracy of the HtD, pulpwood volume or DBH structural variables modelled, and even resulted in errors obtained for the HtD model, as it overestimated the model performance for young *E. grandis* (R^2= 0.98 and RMSE of 100.23 m). However, in that regard it was also found that Htm can be modelled using a multi-resolution approach, as the model yielded more significant results (R^2= 0.80 and RMSE of 1.80m) than did an individual spatial resolution approach for young *E. grandis*. The results obtained can be explained, as terrain details are often lost or refined when DTMs or DEMs are coarsened into larger spatial scale. Site-specific regions such as commercial plantation forests, where variations in terrain can be affected by spatial variability such as slope and aspect, may therefore require multi-resolution analysis for Htm modelling.

2.4.2 Variable Selection versus using all the variables

To improve the predictive accuracy of the structural attributes modelled, a RF-backward feature selection was implemented to reduce the number of variables that did not contribute to the predictive accuracy of the algorithm. The results obtained were surprising, as the predictive accuracy of the models did not improve, and in many cases even decreased the predictive accuracy of the explanatory structural variables. This was unexpected, as various other studies have shown improvements in the predictive accuracies of the attributes modelled when using variable selection methods (Treeratpituk and Giles, 2009; Ismail *et al.*, 2010; Genuer *et al.*, 2010). In addition to the unusual trends associated with the predictive accuracy of the structural variables modelled, variable importance analysis showed that that not all predictor variables had significant weights for the RF algorithm and therefore did not contribute to the predictive accuracy of the explanatory structural attributes modelled. Significantly high weights were placed on the diffuse and direct incoming solar radiation predictor variables. This trend was especially evident for regression models pertaining to height stratification and HtD for both eucalypt species and is evident in the literature, as solar radiation or competition for light has been found to impact tree growth. Larger diameters and height and thus pulpwood volume yields are found in areas with lower intensities of radiation (Saremi *et al.*, 2014b). It has therefore been shown that solar radiation can impact the variation in tree heights. Further variation can be seen between young and mature eucalypt species.

2.4.3 Young versus mature stands

This research has also shown that age categories (young and mature), where characteristically open canopies are associated with younger age groups and closed canopies with mature age groups, display different degrees of variability. Younger stands of *E. grandis* displayed more variation in height, pulpwood volume and DBH than mature *E. grandis* stands in the plantation. This result is comparable with Saremi *et al*. (2014a) in which mixed linear modelling explained a 90% variation in height stratification for young sites and only a 87% variation for mature radiata pine plantations. Edirirweera *et al.* (2014) also produced similar results, as an 80% variation was found for young open canopy *E. propinqua* and *E. siderophloia* forest as compared with a 60% variation in closed canopy conditions.

Different results were found for *E. dunnii*, higher variations were found for mature stands than for young stands. These results indicated that subspecies variation occurs in eucalypt plantation forest for terrain. According to Arnold *et al.* (2004) *E. dunni* is considered as an excellent alternative to *E. grandis* as it is more adaptable to dry or frost-prone areas and demonstrates faster growth rates. The results from this study therefore support those in the literature and show that *E. dunni* can adapt better to variations in terrain than *E. grandis*. In commercial plantation forests, the heterogeneity of height and volume within stands is important to ascertain, as these characteristics affect the overall profitability and quality of the deliverables. This study therefore shows that whilst plantations contain the same climatic, topographic, soil, precipitation and silvicultural regimes, micro-scale variations within stands of even-aged plantations still exist. The special factors that cause the differences in the growth rates of *E. grandis* and *E. dunnii* species would require further analysis, which is not within the scope of this study.

2.4.4 Other considerations and limitations noted in this study

Of notable interest to terrain applications is the accuracy of the LiDAR derived DTM, which can be compromised if is not captured and handled at optimal conditions, especially for studies at such as fine scale as this. According to Hawbaker *et al.* (2010) LiDAR can be captured either for on-leaf or off-leaf conditions within the forest. When using low-density LiDAR, which is often used for elevation mapping, the data should be captured during off-leaf conditions so that the pulses can reach the ground (Andersen *et al.*, 2005; Reutebuch *et al.*, 2005). However, the LiDAR that is captured for forestry applications are captured during on-leaf conditions as per Hawbaker *et al*., (2010). The impact of using on-leaf and off-leaf LiDAR is demonstrated by Bouvier *et al*. (2015), where the results indicate that obtaining LiDAR for volume estimates can be improved by collecting it during off-leaf conditions. This can therefore have substantial impacts on the DTM that may be created from the LiDAR system for terrain applications in forestry, as tree crowns intercept the pulses. The accuracy of many 3D-acquired remotely-sensed data including LiDAR is affected by topography, with many of the errors occurring in areas where the slope faces away from the sun or has a low solar angle (Rahlf *et al.*, 2014). These characteristics could therefore inadvertently have impacted the accuracy of this research.

Issues associated with LiDAR pulse densities and spatial resolution are important, as errors associated with plot data can be manifested when using field inventory data. Often such studies will make use of forest inventories that have been previously collected, as this study has. The error associated with this data lies in the Global Positioning System (GPS) that is used to collect plot data, where location error by the system can be up to ten metres (Hill *et al.*, 2014). This becomes of interest when working with DTM of very high spatial resolution i.e. 1-5 metres. According to Fuller (1987), when the exact location or centre of the plot location is given, the R^2 value obtained for modelling will be higher. In addition, the area studied in this research was dominated by a heterogeneous landscape with variations in the plots sampled, different age groups ranging from two to ten years, and different heights, DBH and observed volumes. Therefore according to Straub *et al.* (2013b), should a more homogenous or a larger site be sampled there is a possibility that the RMSE could be decreased.

Despite the high cost of LiDAR and LiDAR-derived DTM, which remains a barrier to many researchers, this study has shown that a 1 x 1 m spatial resolution is required for modelling terrain variables in site-specific localities such as plantation forests in which plot level information is required. Nevertheless, various successful research projects have shown that there is no requirement for high density LiDAR within forestry applications (Hawbaker *et al.*, 2010). This study is different from those in that it suggests that there may be a need for such data when modelling terrain and structural variables at the plot level within a plantation forest. Whilst the cost associated with LiDAR and LiDAR-derived DTM is high, should one high quality DTM be purchased and readily available it could be used in many analyses, as it is expected that the terrain will remain unchanged, and the equipment could therefore be used for future inventories within forest (Järnstedt *et al.*, 2012).

2.5 Conclusion

For many decades forest managers have known the intrinsic value of attaining accurate height metrics for volume estimation at the stand level in commercial plantation forest. The main aim of this study was to examine the relationships between forest structural attributes such as height (Htm and HtD), volume and DBH across terrain variables derived from LiDAR data. This study has been a first attempt to determine how much of the variation in tree structure can be attributed to terrain, using LiDAR-derived DTM. Whilst the results have not produced the level of accuracy necessary for operational use, they do indicate that there is a great potential for LiDAR-derived DTM as a tool to determine the impacts of terrain on volume and tree structure estimates, especially on height metrics which show greatest variation in stands associated with different terrains. For this reason, this study provides a framework for use as a tool in forest inventory decision making by forest managers. Given accurate forest inventories and spatial datasets, forest managers would be able to make informed decisions to regularise stands due to their knowledge of the variations that exist within stands of even-aged species.

CHAPTER THREE

Modelling Terrain Roughness using a LiDAR-derived DTM and a Supervised Random Forest Approach

This chapter is based on:

Munsamy, R., Ismail., R and Gebreslasie, M. (2017) Modelling Terrain Roughness using a LiDAR-derived DTM and a Supervised Random Forest Approach. *In preparation.*

Abstract. The accurate detection of terrain roughness is imperative for both silvicultural and operational functions within commercial plantation forests. This study investigates the potential of Terrain Roughness Indices extracted from LiDAR data for detecting terrain roughness in a commercial plantation forest using a supervised random forest classification method. All studied Terrain Roughness indices showed strength in discriminating between roughness classes (rough, intermediate, and smooth). The results indicated that all eight terrain roughness indices could discriminate all or some classes and were significant at ($p<0.05$), whilst the Terrain Ruggedness Index (TRI), the Standard Deviation of Slope (SDSLP), the Melton Ruggedness Number (MRN) and the Slope Variability (SV) were significant at ($p<0.001$). When using all eight terrain indices the random forest supervised methodology that was adopted in this approach yielded a classification accuracy of 90% with a Cohen's kappa statistic of 0.85. However, the random forest variable importance result showed that TRI and SDSLP provide the most accurate result for the detection of all terrain roughness classes in a commercial plantation forest in KwaZulu-Natal, South Africa. A supervised random forest approach proved to be a robust technique for classifying terrain roughness.

Keywords: *LiDAR, DTM, terrain roughness, random forest, forestry*

3.1 Introduction

Airborne Laser Scanning (ALS) or Light Detection and Ranging (LiDAR) is a recent active remote sensing technology that has proven its capabilities in the forestry industry (Hudak *et al.,* 2002; Antonarakis *et al.*, 2008; Asner, 2009; Edson and Wing, 2011). LiDAR has significant advantages with regard to resolution, systems automation and cost efficiency, and provides terrain information for densely vegetated regions that is more accurate than that provided by more traditional passive remote sensing technologies (Baltsavias, 1999; Hollaus and Höfle, 2010). In addition, the technology has the capability to accurately capture the vertical and horizontal plantation forest surface structure and thereby capture terrain structure at a higher precision by extending measurements into the third dimension (3D) (Lefsky *et al.*, 2002; Van Leeuwen and Nieuwenhuis, 2010; Wulder *et al.*, 2012).

By means of utilizing remote sensing data and Geographic Information System (GIS) techniques, precise terrain information for inaccessible, densely forested regions can be provided (Baltsavias, 1999; Hollaus and Höfle, 2010). For this reason, LiDAR is rapidly becoming the sole method of gaining accurate forest data, and has already replaced traditional data collection in certain localities of the world, especially for terrain mapping within heterogeneous and inaccessible environments. Thus, the advent of high spatial resolution LiDAR data with overall accuracies (OA) ranging between 10-15 centimetres (cm) has revolutionized the way in which elevation data is captured and sourced for terrain applications (Baltsavias, 1999; White *et al*., 2010).

Most terrain-related applications make use of topographical mapping that relies on the use of Digital Elevation Models (DEM) or Digital Terrain Models (DTM) to represent surface characteristics (Wilson, 2012). From LiDAR a high-resolution DEM or DTM is derived from the last pulse of the LiDAR data, which is achieved by filtering a point cloud into off-terrain (non-ground) points and terrain points (Ullrich *et al.*, 2007; Cavalli and Marchi, 2008). A DTM represents the bare ground surface, with all non-ground surfaces or objects such as vegetation and buildings removed (Sterenczak *et al.*, 2013). DEM is often used interchangeably with DTM as an umbrella term in various research projects (Hoechstetter *et al.*, 2008). However, Podobnikar *et al*. (2000) state that the characteristic difference between the two models is that a DEM is referred to as grid data and therefore contains only elevation attributes, whilst the term DTM refers to a modelled structure of the surface that may contain additional terrain data including peak points and ridgelines.

In forestry the accuracy of the DTM is an important consideration (Sterenczak *et al*., 2013). The main advantage of DTM data is that it provides an efficient means to extract topographic information and it allows for surface processes to be easily modelled, whilst still containing a high spatial resolution and accuracy. Researchers have noted that the infrared illumination emitted by the LiDAR system has the capability to penetrate gaps in the canopy of densely forested areas and due to the high precision of the data it becomes possible to accurately identify and monitor objects of interest that are located within the forestry understory, producing accurate surface measurements and representations (Kasai *et al*., 2009; White *et al.*, 2010). For that reason, LiDAR overcomes the limitations of traditional passive remote sensing systems, thereby becoming particularly valuable in terrain mapping in the forestry sector. Primary analysis can be based on the DTM, which provides data pertaining to aspect, elevation, slope, plan and profile curvature and terrain roughness, for instance (Zomer *et al*., 2002; Mulder *et al*., 2011).

One of the most valuable topographic attributes derived from a DTM is terrain roughness, which is described as the variability or irregularity that is experienced by a topographic surface at a given scale (Grohmann *et al*., 2011). Terrain roughness consists of three categories dependent on scale and can be based on a micro-level (mm/ cm), the meso-level (dm/ m), where objects such as shrubs and boulders are depicted, or the macro-level (m/ km), the choice being determined by topographical features (Hollaus *et al.*, 2011). Differing levels of terrain roughness are therefore required for distinct applications, and information derived from terrain roughness modelling has been utilized in applications within geomorphology (Sankey *et al.*, 2010; Rodríguez-Caballero *et al*., 2012); geology (Morris *et al*., 2008; Berti *et al.*, 2013) ecology (Sanson *et al.*, 1995; May *et al.*, 2008) and forestry (Ferry *et al.*, 2010; Saremi *et al.*, 2014).

To date, a range of approaches has been put forward to quantitatively describe the roughness of terrains using LiDAR-derived data for various applications (Sankey *et al.*, 2010). Whilst numerous techniques have been employed for calculating terrain roughness in various landscapes across multiple scales, the methodological approaches utilized for modelling terrain roughness are either based on a statistical nature or are more traditional approaches which involve the physical collection of surface information, which is often recognised as a cumbersome task to perform (Shepard *et al.*, 2001; Sankey *et al*., 2010). In environments where traditional approaches are the main method of determining roughness, morphometric indicators of terrain roughness can also provide an excellent complementary method for validation purposes (Popit and Verbovšek, 2013).

An example of modelling terrain roughness is provided by Cavalli and Marchi (2008). The authors made use of a LiDAR-derived DTM with a resolution of 2 m x 2 m to quantitatively analyse the local topography of alluvial fans in a small alpine stream located in Moscardo Torrent in the Eastern Italian Alps. A topographic roughness index based on $SD_{restopo,}$ plan curvature and a shaded relief map were used to determine the roughness of the alluvial fans. The results demonstrated that the DTM was a valuable source of data for modelling surface morphology. Similarly, Grohmann *et al.* (2011), who conducted research in the Midland Valley, Scotland, utilised six common techniques for determining terrain roughness. Methods were compared for five different resolutions ranging from 5 m x 5 m, 10 m x 10 m, 25 m x 25 m, 50 m x 50 m and 100 m x 100 m, and across twelve different moving window sizes. The methods used to determine terrain roughness included the area-ratio, vector-dispersion, SDELV, $SD_{restopo}$, SDSLP, and SDPC. The outcome of the study showed that the SDSLP, the SDPC and the vector-dispersion methods yielded good results whilst the SDELV and $SD_{restopo}$ produced intermediate results and the area-ratio method failed to yield meaningful results

(Grohmann *et al.*, 2011). In a more recent study conducted by Popit and Verbovšek (2013) in the Rebrnice area, Slovenia, DEMs with a resolution of 3 m x 3 m were utilised to determine terrain roughness for landslides. In this study, the slope variability (SV) and the Terrain Ruggedness Index (TRI) were used. The results obtained from this study showed that the TRI method was preferred in this region as it yielded more defined differences in relief and visually produced a larger range of data. Other noteworthy studies include those done by Sappington *et al*. (2007), Streutker *et al*. (2011), Bretar *et al*. (2013) and De Reu *et al*. (2013). From the aforementioned studies, LiDAR-derived DTMs and DEMs have proved to be useful tools for characterizing terrain roughness across multiple scales (González *et al.*, 2008).

Whilst numerous techniques have been employed for calculating terrain roughness, a standard methodology has yet to be developed for quantitatively measuring the accuracy of terrain roughness algorithms and indices for complex environments such as that of a plantation forest. Random Forest, which is a common machine-learning classifier, has been used in many fields for classification. A Random Forest classification provides more accurate results than most commonly used classifiers (Ismail and Mutanga, 2011). It is highly favoured due to the fact that it can account for high levels of noise and is found to be more suited to dealing with any outliers that may exist within the data (Breiman, 2001; Rodriguez-Galiano *et al*., 2012). Application fields that have utilized Random Forest for classification with high accuracy include forestry, geomorphology and ecology. In addition, Random Forest is highly favoured as it can provide an indication of the predictor variables that are most important for classification accuracy. Presently the utilization of a supervised Random Forest for terrain roughness analysis is rather new, with little to no literature documenting any findings. In one study conducted by Baldwin *et al*. (2017) a supervised Random Forest algorithm was adopted to classify terrain variables for the stratification scheme for the Shale Hill's Catchment Area. Terrain variables were combined with soil information and classified. The Random Forest algorithm showed that the most important variables for detecting soil properties included depth of bedrock, Upslope Contributing Area (UCA), Valley Depth (VD) and Local Slope (LS), thereby providing an indication of the best variables to use for stratification (Baldwin *et al.*, 2017). In a contrasting research project, Zhang *et al*. (2016) utilised a Random Forest classifier to evaluate the transversability of the terrain for mobile robot navigation. The results indicated that the Random Forest classifier could effectively determine the subset of highly important variables for classification. The author was then able to use the best set of variables to attain a higher classification accuracy of 91.75%.

It is important to note that despite the popular use of remote sensing and particularly LiDAR technology in various application fields in South Africa, there is no significant research that utilizes a high spatial resolution LiDAR-derived data set for the application of modelling terrain roughness in local environments of South African forests, which indicates considerable gaps in the research in this field, and more importantly in large-scale South African agricultural plantation environments. The importance of terrain roughness detection within forestry lies in its ability to affect productivity in commercial forestry, particularly for operations that make use of mechanized systems and that may be slowed due to rough or very steep slopes (Alam *et al.*, 2012). By utilising terrain roughness algorithms and indices, primary factors within sites such as roughness or the steepness of slopes can be determined, thus allowing for a more optimal selection of the required equipment before field operations are conducted (Alam *et al.*, 2012). Thus, by evaluating the use of terrain indices from LiDAR-derived DTMs for detecting terrain roughness in plantation forests, sustainable forest management goals and up-to-date forest inventory systems with accurate spatial information can be derived (Reutebuch *et al.*, 2005; Wulder *et al.*, 2008). It is within this context that this study seeks to determine if the terrain indices extracted from high resolution LiDAR-derived DTM can be used to accurately detect terrain roughness in a commercial plantation when using a supervised Random Forest technique.

3.2 Materials and Methods

3.2.1 Study Area

The study was conducted in Sappi's Comrie forest plantation. Figure 8 represents the study site. The main land uses in the area are forestry and sugar cane cultivation. The altitude for the site ranges between 70 m and 650 m. The mean annual temperature in the plantation ranges between 19^{o} and 20^{o}. Precipitation in the site ranges between 890 mm and 940 mm per year. The terrain of the plantation displays high levels of heterogeneity, as the study site is composed of highly dissected low undulating mountains, undulating hills, lowlands, and plains. The geology of the site is composed of the Natal Group and Red-Brown Sandstones.

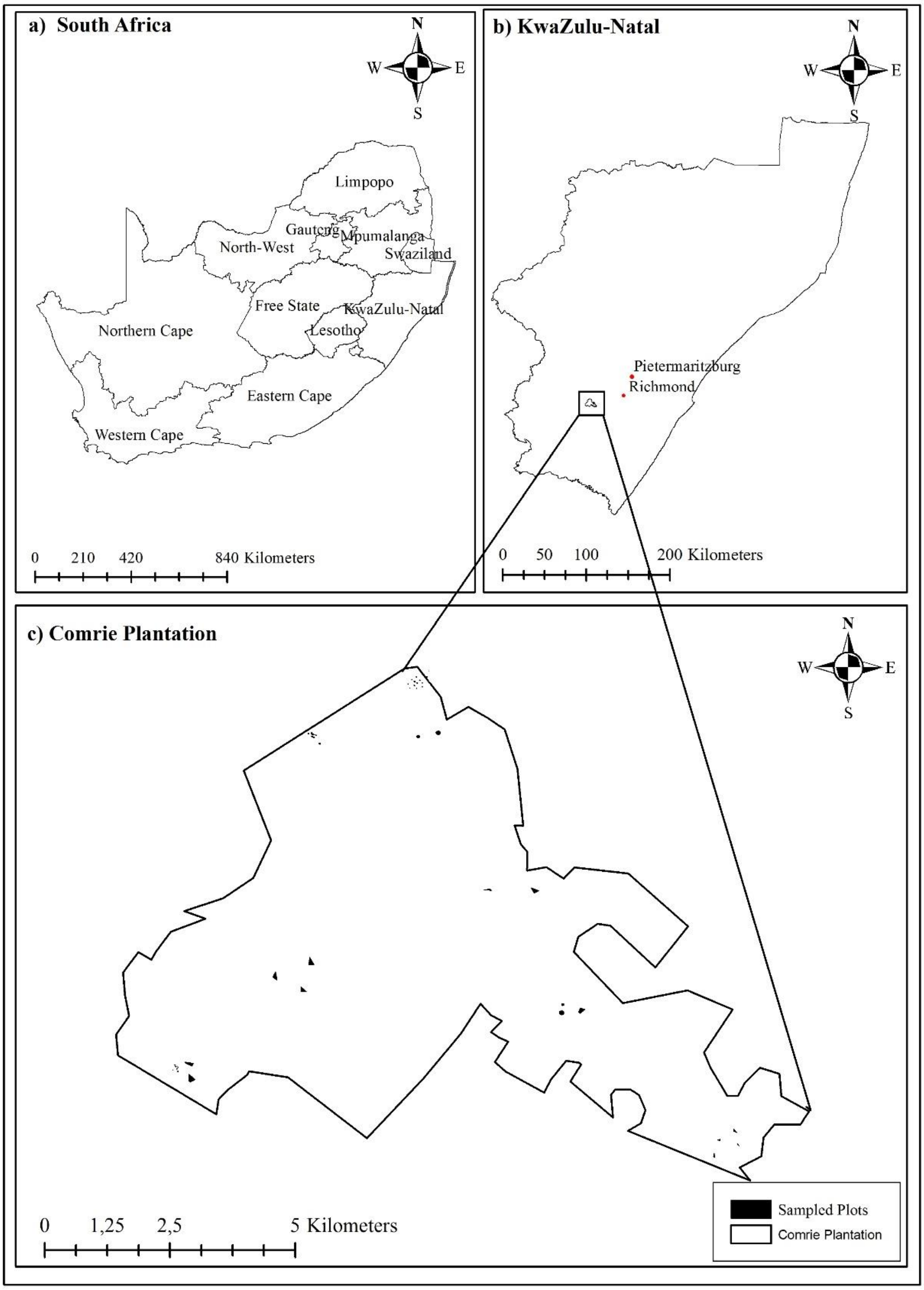


Figure 8: Location of the study area: a) South Africa, b) KwaZulu-Natal and c) Comrie Plantation

3.2.2 Dataset

3.2.2.1 LiDAR

The airborne LiDAR point data was collected over the Sappi Comrie forest plantation. The Land Resource International collected the data. The LiDAR points were classified for each flight line into either ground or non-ground points. In order to inspect the integrity of the LiDAR data set, quality assurance was conducted by using 16 Real-Time Ground Control Points (GCP) collected using a TrigNET system (Land Resources International, 2015). The RMSE for the LiDAR was reported as 0.08 m. Other residuals achieved in the accuracy assessment produced minimum residual values ranging from -0.09 m to 0.17 m (Land Resources International, 2015). By utilizing a vendor-specific proprietary technique the filtered LiDAR ground points were then thinned and used to create a Triangular Irregular Network (TIN) (Brubaker *et al.*, 2013). The resultant TIN was then utilized to create the high-resolution DTMs with a grid elevation of 1 m x 1 m based on the Hartebeesthoek 94 datum with a Transverse Mercator Lo31 projection system.

3.2.3 Methods

3.2.3.1 DTM Developing

3.2.3.1.1 Extracted Terrain Roughness

Considering the scale and local topography of the study area, the following terrain roughness methods were chosen: TRI, VRM, TPI, SDSLP, SDELV, SDPC, MRN and Slope Variability. The input elevation data was the DTM data with a 1 m x 1 m resolution. To determine an optimal cell size for this study, a preliminary test was conducted for a single terrain roughness method against two commonly used moving-window sizes, 3 m x 3 m and 10 m x 10 m. A basic comparison showed that the 10 m x 10 m moving-window visually displayed coarser detail than the 3 m x 3 m moving-window, which displayed finer detail in roughness over a greater area, as is required for micro-scale terrain roughness mapping. Even though larger relief objects could not be depicted at relatively small window sizes, it was important in this study to determine micro-scale terrain roughness which could otherwise have become obscured by using larger window sizes. As the window size becomes larger, it is impossible to determine the response of the objects of interest, as the scale becomes larger than that of the feature of interest (Grohmann *et al.*, 2011). A 3 m x 3 m cell size was therefore chosen for this study. A visual representation of the different terrain roughness variables modelled is presented in table

5. The cast of colours represented in the subset maps is based on the code rough (blue), intermediate (dark green) and smooth (light green).

The terrain roughness modelling was conducted using an open-source System for Automated Geoscientific Analysis (SAGA) GIS Software v. 2.1.4. and the Environmental Systems Research Institutes (ESRI) ArcGIS v. 10.1 software package. The following morphometric indicators of terrain analysis in the SAGA GIS package were utilized to create the TRI VRM; Slope, Aspect, and Curvature and the MRN information layers. ESRI's ArcGIS Model Builder in ArcCatalog was utilized. The Focal Statistics under Spatial Analysis Tools were added to the model to create the following information layers for the TPI method: minimum elevation, maximum elevation and a smoothed DTM. The resultant data were then imported into ArcMap, where the raster calculator in the Spatial Analysis Tools was then utilized to calculate the final roughness information layer. A similar process was followed for SDELV, Slope Variability and SDSLP methods. A detailed description of the multiple roughness methods used in this study is provided in table 5. A summary of the framework that was employed in this study is presented in figure 9.

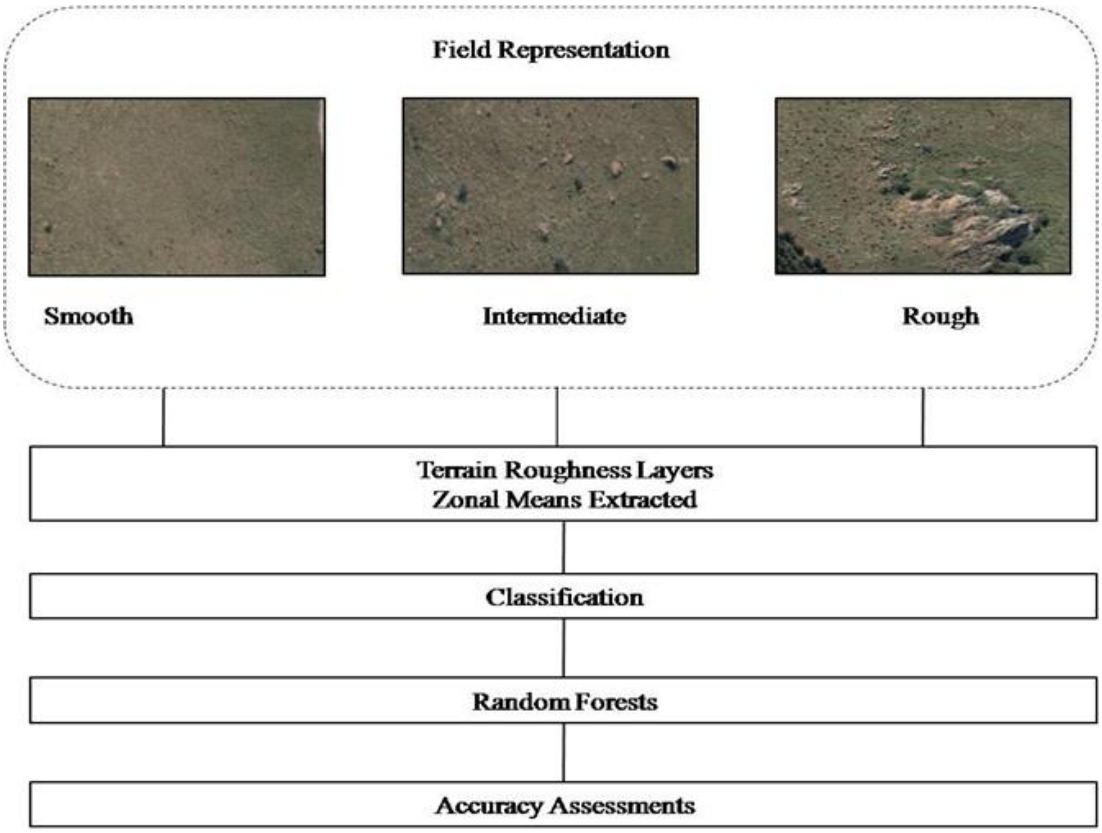
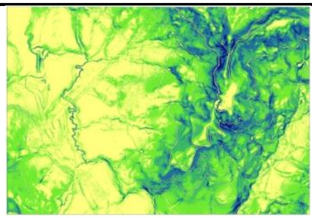
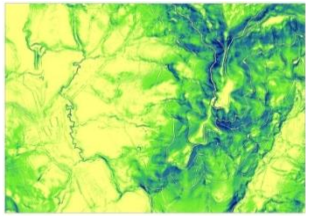
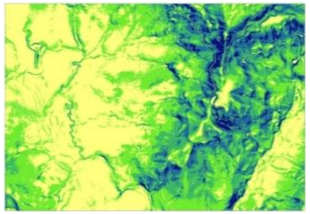
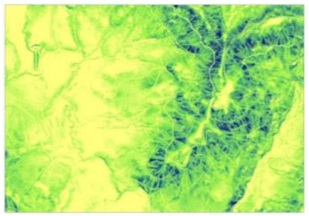
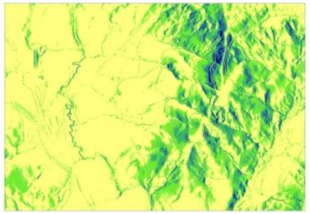
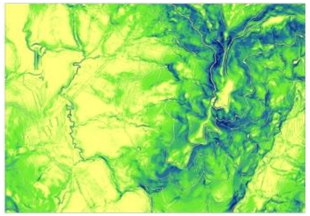
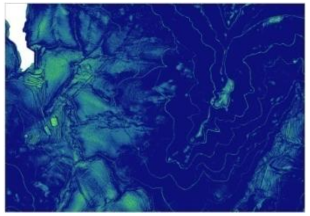
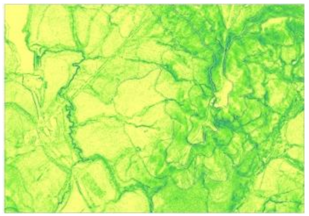


Figure 9: Framework for this study

Table 5: Terrain indices used in this study

Method Name			Equation/ Index	Description	Reference
1. Terrain Ruggedness Index	TRI		TRI= $\sqrt{\lvert(\text{maxDEM})^2 - (\text{minDEM})^2\rvert}$ maxDEM= maximum raster minDEM= minimum raster	Change in elevation between the central grid and the mean of an 8-cell neighbourhood of surrounding cells	(Riley *et al.*, 1999)
2. Slope Variability	SV		$SV = S_{max} - S_{min}$ *SV* = Slope Variability output raster; S_{max}= maximum raster; S_{min}= minimum raster	Difference in relief between the minimum and maximum elevations of a landscape	(Popit and Verbovšek, 2013)
3.Topographic Position Index	TPI		TPI= Z_0-$\bar{Z}$ Z_0 = elevation at central point $\bar{Z}$= average elevation	Measures the relative topographic slope position of the central point using the difference between the elevation of the first point and the mean elevation of a predetermined neighbourhood	(Guisan *et al.*, 1999)
4.Melton's Ruggedness Number	MRN		MRN= (Zmax-Zmin)/ Sqrt(A)	Determines basin relief divided by the square root of the basin area	(Melton, 1965)

5. Vector Ruggedness Measure	VRM		$\mid r \mid = \sqrt{}(\sum_x)2 + (\sum_y)2 + (\sum_z)2$	Roughness is measured as the variation of a three-dimensional orientation of grid cells within a neighbourhood	(Sappington *et al.*, 2007)
6. Standard Deviation of Slope	SDSLP		$SD_{slope}=\sqrt{\frac{1}{n_R-1}\sum i = 1(z_i - \overline{z})^2}$ $z_{i=}$ height and $\overline{z}$= average height	Determines roughness as a factor of standard deviation	(Grohmann *et al.*, 2011)
7. Standard Deviation of Elevation	SDELV		("meanDEM"-"DEM")/ "rangeDEM" meanDEM = mean raster, rangeDEM= raster with range of elevation values	Measures topographic roughness when the mean of the DEM is subtracted from the original DEM value	(Grohmann *et al.*, 2011)
8. Standard Deviation of Profile Curvature	SDPC		$SD_{profc}=\sqrt{\frac{1}{n_R-1}\sum i = 1(z_i - \overline{z})^2}$ $z_{i=}$ height and $\overline{z}$= average height	Measures downslope curvature and identifies breaks in slopes	(Grohmann *et al.*, 2011)

3.2.4 Statistical analysis

In this study we tested the hypothesis that the terrain roughness indices extracted from a high-resolution LiDAR-derived DTM would be able to discriminate amongst the differing roughness classes (i.e. rough, intermediate and smooth). An Analysis of Variance (ANOVA) with a Tukey's HSD post-hoc test was conducted to determine the significance level. The ANOVA and Tukey's HSD post-hoc test were conducted using SPSS.

3.2.5 Classification

3.2.5.1 Random Forests

Random Forest was implemented using the Random Forest library available in R statistical software (R Development Core Team, 2008). Random Forest is an ensemble classifier that makes use of multiple decision trees, in which each tree is selected on a bootstrap method of the original data set (Breiman, 2001). Each of the subsets of data drawn is different for each decision tree (Halmy *et al.*, 2015). This method also allows for minimal error estimation from the test data set, and is referred to as the Out–of-Bag (OOB) error. The remaining test set is predicted from the bootstrap samples in which the OBB predictions are calculated based on all trees and is referred to as the Mean Square Error (MSE_{OOE}) (Liaw and Wiener, 2012). The Random Forest algorithm requires the user to determine two parameters: the number of decision trees to be grown and the number of predictor variables that should be split for each of the decision trees (Rodriguez-Galiano *et al.,* 2012). In this study default parameters were used. An illustration of the Random Forest classifier for classification is provided in figure 10.

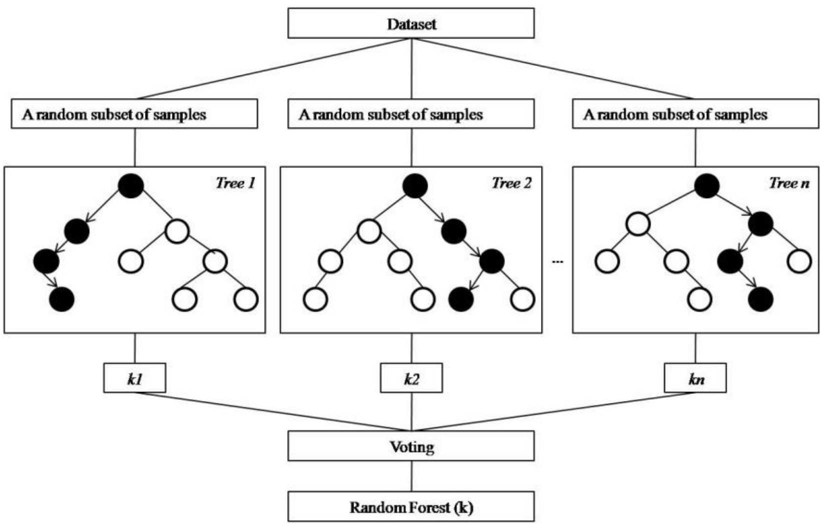


Figure 10: Illustration of the Random Forest Classifier

The most commonly used method of reporting the accuracy of classification accuracy is through the generation of a Confusion matrix (Asmala, 2012). The confusion matrix relates the accuracy of the classification on a class-to-class basis and provides the accuracy of the classification compared with the reference image and the classification data (Patil *et al.*, 2012). As this study employed a supervised Random Forest technique to train the model, the User's Accuracy (UA), which refers to the probability that a given pixel on the classified image matches the land-cover that is being classified, and the Producer's Accuracy (PA), which relates to the probability of the reference pixel's being accurately classified, are calculated. The Cohen's Kappa statistic, which includes some elements of the error matrices (classification errors), is calculated by the Random Forest classifier (Yuan *et al.*, 2005).

3.2.5.2 Variable Importance

Variable importance is described as the measurement of influence on accuracy (Treeratpituk and Giles, 2009). The amount of influence that a variable has on classification accuracy can be determined by the Random Forest classifier during the training process (Zhang *et al.*, 2016). In a Random Forest classifier, variable importance is determined when a critical variable is permutated, and other variables remain unchanged whilst the out-of-bag (OOB) mean square error accuracy shows a decrease (Zhang *et al.*, 2016). The OOB mean squared error (MSE) is determined by averaging the squared deviations of the OOB response for all the predictor variables (Breiman, 2001; Grömping, 2012).

3.4 Results

3.4.1 ANOVA and Post-Hoc Tukey Test

In this study, we tested the hypothesis that the terrain roughness indices would be able to discriminate amongst the differing roughness classes by conducting a one-way ANOVA. The ANOVA was conducted at a 0.05 significance level. The results indicated that all eight terrain roughness indices could discriminate all or some classes and were significant ($p<0.05$). A one-way ANOVA can tell that there is a significant difference between the terrain roughness classes, but it is unable to tell where the difference in the classes lies (Ismail *et al.*, 2007). A Tukey's HSD post-hoc test is therefore required to determine where the difference lies in each of the terrain roughness classes (Ismail *et al.*, 2007). The results obtained from Tukey's HSD post-hoc test are provided in table 6.

From the statistical analysis, it was found that all terrain roughness indices can discriminate between some or all classes. Post-hoc analysis also showed that all terrain roughness variables could discriminate either between all or some classes at a higher significance level ($p<0.001$). VRM was significant at ($p<0.05$), but was unable to show significance for classes intermediate and smooth at ($p<0.001$). MRN and TRI could discriminate classes at both the ($p<0.05$) and ($p<0.001$). SDELV was the only index that was unable to discriminate between classes rough and intermediate, but could discriminate between classes rough and smooth and intermediate and smooth at the ($p<0.05$). SDELV could discriminate between classes rough and smooth and intermediate and smooth at ($p<0.001$). TPI was significant ($p<0.05$) for all classes, but it demonstrated significance for classes rough and smooth and intermediate and smooth only at ($p<0.001$). SDPC showed significance for all classes at ($p<0.05$), but could discriminate between classes rough and smooth only at ($p<0.001$). SV and SDSLP showed significance for both ($p<0.05$) and ($p<0.001$) for all classes. The most significant separation of results was for the terrain variables TRI, SDSLP, MRN and SV, as these indices demonstrated the ability to distinguish at all classes at the significance level ($p< 0.001$).

Table 6: Analysis of variance results with a Tukey's Post-Hoc HSD test. Class rough, intermediate and smooth. (P<0.05= * not significant, P<0.05= ** significant, P<0.001= † not significant, P<0.001 ††= significant)

VRM	Rough	Intermediate	Smooth	MRN	Rough	Intermediate	Smooth
Rough	..	** ††	** ††	Rough	..	** ††	** ††
Intermediate	** ††	..	* †	Intermediate	** ††	..	** ††
Smooth	** ††	* †	..	Smooth	** ††	** ††	..
TRI	Rough	Intermediate	Smooth	SDELV	Rough	Intermediate	Smooth
Rough	..	** ††	** ††	Rough	..	* †	** ††
Intermediate	** ††	..	** ††	Intermediate	* †	..	** ††
Smooth	** ††	** ††	..	Smooth	** ††	** ††	..
TPI	Rough	Intermediate	Smooth	SDPC	Rough	Intermediate	Smooth
Rough	..	* †	** ††	Rough	..	* †	** ††
Intermediate	* †	..	** ††	Intermediate	* †	..	* †
Smooth	** ††	** ††	..	Smooth	** ††	* †	..
SV	Rough	Intermediate	Smooth	SDSLP	Rough	Intermediate	Smooth
Rough	..	** ††	** ††	Rough	..	** ††	** ††
Intermediate	** ††	..	** ††	Intermediate	** ††	..	** ††
Smooth	** ††	** ††	..	Smooth	** ††	** ††	..

3.4.2 Classification

In this study the OOB rate of error was reported as 12.5%. The results from the Random Forest classifier indicated that the classifier produced an overall accuracy of 90% with a Cohen's kappa statistic of 0.85 when using a supervised classification techique (table 3). The PA and UA were reported as 92.10% and 90.00% respectively (table 7). The Random Forest classifier also provides an indication of the optimal tuning parameter (*m*try). Several different *m*try values from two to eight were tested by the classifier. An *m*try value of two provided the lowest accuracy, whilst *m*try values three to seven demonstrated high accuracies at a stable trend and

Cohen's kappa statistics value of 82.50%. The results indicated that in this study the optimal *m*try needed to acquire the highest classification accuracy was eight (figure 11).

Table 7: Confusion matrix showing the predicted accuracy of terrain roughness when using a 3-level classification system: Rough, Intermediate, and Smooth

Class	Rough	Intermediate	Smooth	UA
Rough	35	3	2	87.50%
Intermediate	2	35	3	87.50%
Smooth	1	4	35	87.50%
PA	92.10%	83.33%	87.70%	90.00%

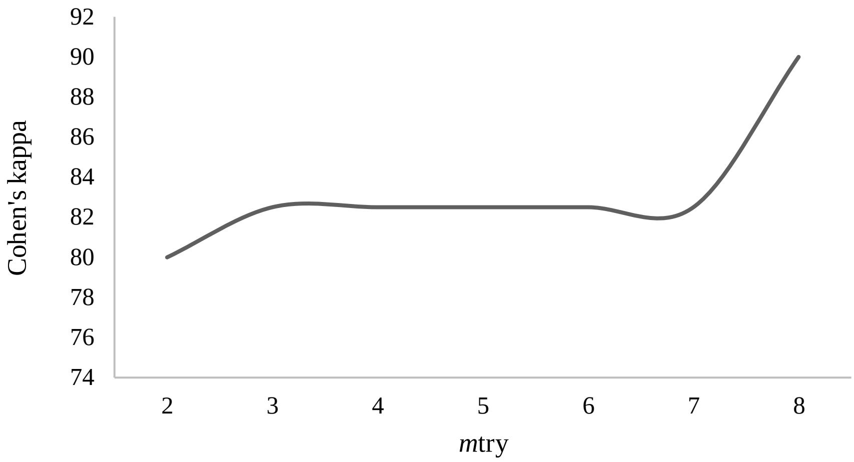


Figure 11: mtry (horizontal axis) versus the rate of error (Cohen's kappa in percent (%) (vertical axis)

3.4.3 Random Forest Variable Importance

The most important terrain variables in the Random Forest supervised classification process are presented in figure 12. The variables are sorted in accordance with their importance across the classes. In this study, VRM, TRI and SDSLP are the top three terrain roughness variables as identifiedby the classifier. SV and TPI were ranked 4th and 5th. SD_{elv} was ranked 6th, whilst MRN and SDPC were ranked 7th and 8th respectively.

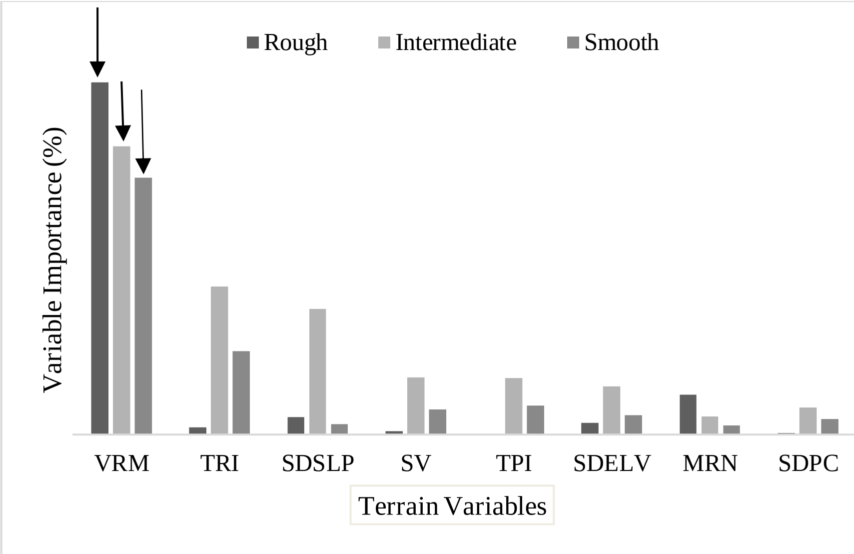


Figure 12: Variable importance determined from the RF classifier.

The most important terrain variables for the supervised classification variables are sorted according to their importance across the classes. The horizontal axis represents the terrain variables and the vertical axis represents the importance of each variable. Arrow indicated the variables with the greatest importance.

3.5 Discussion

In this study eight terrain indices were extracted from a high-resolution LiDAR-derived DTM to determine if it could be accurately utilised to detect terrain roughness within a commercial plantation forest when using a supervised Random Forest technique. The results obtained from this study indicate that high-resolution LiDAR-derived DTM and that the supervised Random Forest technique provide a robust method for accurately detecting and classifying terrain roughness within a commercial plantation forest dominated by highly heterogonous landscapes.

3.5.1 ANOVA

According to Sappington *et al.* (2007) a perfect terrain roughness measure should include both aspect and slope gradient, so that the result is a multivariate representation of the topography. However, many of the commonly used terrain roughness methods fail to provide such a representation. In this study the ANOVA indicated that all eight terrain roughness indices can significantly discriminate between the classes of interest at ($p < 0.05$) whilst four terrain indices (TRI, SDSLP, SV and MRN) could significantly discriminate between the classes at the ($p<0.001$) level.

3.5.2 Classification

In this study the OOB error rate was reported as 12.5%, but various studies have suggested that discrepancies in the OOB rate may make it an unreliable measure of class error for classification. The Cohen's kappa statistic is a more accurate measure of error (Heung *et al.*, 2014) . The classification results showed that the supervised Random Forest classifier is a robust method and provided a high classification accuracy of 90% with a Cohen's kappa statistic of 0.85. Whilst visual interpretation indicated various discrepancies, for instance that the following indices (TPI and SDELV) had overestimated roughness, certain indices appeared also to have underestimated roughness (MRN and SDPC). Furthermore, by way of visual analysis TRI, SV, VRM and SDSLP showed similarities in interpretability and produced results that were visually good.

3.5.3 Variables and their importance

Variable importance indicated that VRM, TRI and SDSLP were the top three variables for the classification. However, whilst VRM had displayed the highest variable importance for the Random Forest classification procedure and was able to discriminate between all classes at the significance level ($p<0.05$), it did not display significance for all classes at ($p<0.001$) as it was able to discriminate only between the classes intermediate and smooth. TRI and SDSLP were ranked two and three by the Random Forest classifier and were able to display discrimination between all classes at both significance levels. Furthermore, these indices were the only two of the eight indices to show significance ($p<0.001$) for all classes. Similar results were achieved in a study conducted by Grohmann *et al.* (2011) where SDSLP, SDPC and Vector Dispersion were found to have produced the best results, whilst SDELV and SD residual topography were able to provide only intermediate results.

The results obtained from this study are in accordance with those obtained in the study conducted by Grohmann *et al.* (2011), where it was found that SDSLP was one of the top terrain indices that yielded better performance than the other indices that were tested. SDSLP could correctly identify steep or smooth slopes, including areas that displayed high surface clutter such as forest stands (Grohmann *et al.*, 2011). SDSLP was also able to identify breaks of slope across multiple scales. In this study, SDPC and SDELV did not produce visually good results and seemed to either overestimate or underestimate terrain roughness. In the study conducted by Grohmann *et al.* (2011), whilst SDELV was not able to detect local terrain roughness, at larger window sizes the indices demonstrated the ability to detect breaks in slope.

The literature suggests that SDELV would be more applicable for the regional detection of terrain roughness than the microscale detection of terrain roughness.

However, TRI is closely related to slope and includes into its calculation roughness that occurs perpendicular to slope. In this study, TRI was the third most important variable as illustrated by the Random Forest classifier. Similar results were demonstrated in the study conducted by Scarpone *et al.* (2017), where the Random Forest classifier found that TRI was one of the top 3 terrain variables out of 43 variables for the detection of exposed bedrock (EB). According to Popit and Verbovšek (2013), TRI produces a larger range of data and is the preferred terrain index as it can highlight pronounced differences in the roughness that other indices may not be able to detect.

Comparable results were achieved in the study conducted by Popit and Verbovšek (2013) where TRI and SV were calculated using a LiDAR derived DTM. In this study, SV was modified to determine change in elevations of height as compared to slope. In addition, it was also found that the indices had overestimated terrain roughness for areas with relatively low terrain (Popit and Verbovšek, 2013). Similar trends were seen in this study, however, despite the overestimation of roughness, in general a satisfactory result was obtained from this indice.

In this study, TPI produced intermediate results and was ranked as only the fifth most important variable. The shortcomings of TPI were shown, as the index was unable to accurately represent terrain roughness, as it overestimated roughness. Similar results demonstrating the shortcomings of TPI were produced in the study conducted by De Reu *et al.* (2013) in north western Belgium, on a highly heterogenous landscape. According to De Reu *et al*. (2013) TPI is influenced predominantly by roughness in terrain, where height differences may occur at several metres or more, and can therefore lead to the incorrect classification of slope positions and landforms within heterogenous environments, therefore indicating the disadvantages of TPI within complex environments. Whilst this terrain roughness index did not produce satisfactory results in this study, other studies have employed TPI with higher success rates. For example, in a study conducted by Weiss (2011) in Mt Hood in Oregon, USA, TPI proved to be a very useful tool for classifying features of roughness such as slope positions and related landforms. It is worth noting that the study area used in Weiss (2001) was more homogenous than the area that was used in this study or in the study conducted by De Reu *et al*. (2013).

3.5.4 Other considerations and limitation in this study

A supervised Random Forest technique was employed in this study, which means that the user was responsible for discriminating between the various classes of interest prior to the classification process. Whilst this approach is commonly utilised with great accuracy, a high level of uncertainty and error may be introduced into the training process when a user is responsible for training the classifier. For this reason, further research should focus on unsupervised methodologies. For example, Peerbhay *et al.* (2016) employed the Random Forest classifier in an unsupervised approach for the detection of *Solanum mauritianum* (Bugweed) plant invasions in forest margins, open areas and riparian areas. Accuracies of 91.33%, 85.08%, and 67.90% were obtained respectively (Peerbhay *et al.*, 2016).

In addition, whilst this study used the Random Forest technique to determine the most important variables, no variable reduction technique was employed. Future research efforts should employ the use of variable reduction techniques with the Random Forest classifier to determine the optimal number of variables that are required to achieve the highest possible accuracy. In a study conducted by Heung *et al.* (2014) a Random Forest variable reduction technique was employed. The results indicated that the *m*try value had remained fairly consistent until the number of variables was reduced to 9 from the original 27 to acquire a kappa index of 89.6% (Heung, 2014). Other studies that have employed Random Forest variable selection with high success rates include those done by Dye *et al*. (2012) and Ismail and Mutanga (2011), amongst others.

Lastly, whilst the indices extracted from LiDAR provide a great source of roughness information that would be highly valuable for many application fields, especially in commercial plantation forests, the results obtained from this study indicate that not all of the indices extracted from high-spatial resolution LiDAR provide consistently accurate representations of the roughness. In this study TRI and SDSLP, whose indices have calculations based around slope, were found to be the most significant indices for detecting terrain roughness in complex environments such as commercial plantations. Should further research focus on the gaps that may be seen in this study, it might be possible for the accuracy level to be increased. Nevertheless, the results obtained from this research are promising and provide a new methodology to determine terrain roughness in complex landscapes.

3.5 Conclusion

Accurate terrain roughness calculations are invaluable for many operational systems in commercial plantation forests, especially for planning harvesting operations. This study has applied a new methodology using high-resolution LiDAR-derived terrain variables and a machine-learning technique to successfully detect terrain roughness. The terrain indices derived from high-resolution LiDAR provide a practical tool to map terrain roughness in commercial plantation forests as they can accurately discriminate among roughness classes of interest. The results obtained from this study show that TRI and SDSLP can be used to accurately detect roughness in commercial plantation forests.

CHAPTER FOUR

CONCLUSIONS AND RECOMMENDATIONS

4.1 Conclusions

In this thesis, the aim was to utilise LiDAR derived DTM for determining variability in terrain within commercial eucalypt forests in KwaZulu-Natal, South Africa.

The first study examined the use of LiDAR derived DTM for the application of determining variability in terrain within a dense commercial eucalypt forests within the Riverdale plantation site in KwaZulu-Natal, South Africa. More specifically this study focused on determining how much of the variability in forest structural attributes of eucalypt forests can be attributed to terrain. For this, HtD, Htm, pulpwood volumes and DBH were modelled against the 32 terrain variables that were derived from LiDAR derived DTM for even aged *E. grandis* and *E. dunni* respectively.

- The results obtained from this study indicated that from all 32 derived terrain variables, only terrain variables that was associated with solar radiation had displayed importance for determining variability of forest structural attributes associated with terrain.
- It was further found that direct incoming solar radiation influenced the forest structural attribute HtD as it demonstrated the highest variability to terrain and consistently modelled highly significant.
- It was further found that a spatial resolution of 1 m x 1 m consistently provided the highest coefficient of determination for structural variables modelled and indicated that this spatial scale is required to determine variations in structural variables that are associated with micro-scale terrain. This result indicates that a high-resolution LiDAR derived DTM is required for application fields that require modelling of forest structural attributes in relation to terrain.
- In this study, the RF as a machine learning regression technique demonstrated its applicability and indicated that it is a robust technique that can be used to attain accurate results for terrain modelling. In addition, this study applied a RF variable selection method, to which demonstrated results subpar as compared to that identified in literature.
- Lastly, in this study, it was found that *E. grandis* and *E. dunni* demonstrated different patterns of variation, and that the latter species can be successfully modelled using this proposed approach. This pattern where *E. dunni* is described as more adaptable to heterogenous environments is widely known to researchers and individuals in the forestry

sector. For this reason, one can suggest based on the results acquired from this research that *E. dunni* is more suited to plantations that are dominated by rugged terrains, different amounts of incoming solar radiation and steep slopes.

In the next study, a LiDAR derived DTM was used to investigate the potential of terrain indices for detecting terrain roughness in a commercial plantation forest using a supervised random forest classification method in the Comrie plantation in KwaZulu-Natal, South Africa. A supervised methodology was applied to detect three different roughness classes i.e. rough, intermediate, and smooth, respectively.

- This study found that whilst all terrain indices displayed significance for all or some classes and were significant at ($p<0.05$), four indices were significant at ($p<0.001$). This result indicated that whilst all terrain indices can be used for terrain detection, that TRI, MRN, SDSLP, and Slope Variability can be used for the detection of terrain roughness with higher accuracies.
- Lastly, this study indicated that the RF is a robust classifier that provides a high classification accuracy even for studies that make use of small datasets for supervised learning.

4.2 Recommendations for future research

Due to the high value that forests have on the local economy and environment, the past two decades has seen consorted efforts placed on sustainable forest management within commercial plantation. Considering this, abundant research using alternative technologies such as remote sensing has been conducted to better understand the process that affects plantation structure and growth with high success rates. LiDAR, which can overcome the two-dimensional spatial disadvantage, has demonstrated its ability to remotely measure complex vertical forest structures. In this research, it has demonstrated its ability to measure complex horizontal structure within forest.

Despite this, this research indicated that more investigations are still required to develop detailed methodologies that can be easily transferred from one researcher to the next. Further, research that utilises LiDAR derived DTM terrain variables should be conducted within a different plantation site and on a different species to determine if there is a trend in the variability that exist with terrain and with the species that were used as in this study. Further to this, whilst many researches have already focused on determining the optimal point densities of LiDAR for structural variables, no research highlights this for terrain studies; therefore, further research should focus on determining the optimal point densities that will be required for creating highly accurate DTMs.

It is further hoped; that South African researchers employ and develop methodologies that utilise LiDAR on a regular basis. Coupled with its many advantages and applicability to machine learning techniques such as the RF, it can be said that LiDAR really does offer many advantages for forestry research. Further research in this application field will be beneficial and help local foresters, forest managers and researchers to better understand the complexities and advantageous associated with utilizing LiDAR technology for forestry. However, from the recent trends seen thus far within the remote sensing community, one can postulate that developments in this application field are only expected to escalate, which can then provide even greater frontiers to a host of many other valuable research applications within the forestry sector.

References

Aertsen, W., Kint, V., De Vos, B., Deckers, J., Van Orshoven, J., Muys, B. (2012) Predicting forest site productivity in temperate lowland from forest floor, soil and litterfall characteristics using boosted regression trees. Plant and soil 354, 157-172.

Alam, M.M., Strandgard, M.N., Brown, M.W., Fox, J.C. (2012) Improving the productivity of mechanised harvesting systems using remote sensing. Australian Forestry 75, 238-245.

Andersen, H.-E., McGaughey, R.J., Reutebuch, S.E. (2005) Estimating forest canopy fuel parameters using LIDAR data. Remote Sensing of Environment 94, 441-449.

Antonarakis, A., Richards, K.S., Brasington, J. (2008) Object-based land cover classification using airborne LiDAR. Remote Sensing of Environment 112, 2988-2998.

Arnold, R. J., Johnson, I. G., & Owen, J. V. (2004) Genetic variation in growth, stem straightness and wood properties in Eucalyptus dunnii trials in northern New South Wales. Forest Genetics 11 (1), 1-12.

Asmala, A. (2012) Analysis of maximum likelihood classification on multispectral data. Applied Mathematical Sciences 6, 6425-6436.

Asner, G.P. (2009) Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. Environmental Research Letters 4, 034009.

Baldwin, D., Naithani, K.J., Lin, H. (2017) Combined soil-terrain stratification for characterizing catchment-scale soil moisture variation. Geoderma 285, 260-269.

Bale, C., Williams, J., Charley, J. (1998) The impact of aspect on forest structure and floristics in some Eastern Australian sites. Forest Ecology and Management 110, 363-377.

Baltsavias, E.P. (1999) A comparison between photogrammetry and laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing 54, 83-94.

Berti, M., Corsini, A., Daehne, A. (2013) Comparative analysis of surface roughness algorithms for the identification of active landslides. Geomorphology 182, 1-18.

Boehner, J., Selige, T., (2006) Spatial Prediciton of Soil Atttributes Using Terrain Analysis and Climate Regionalisation.

Bouvier, M., Durrieu, S., Fournier, R.A., Renaud, J.-P. (2015) Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sensing of Environment 156, 322-334.

Breiman, L. (2001) Random forests. Machine learning 45, 5-32.

Bretar, F., Arab-Sedze, M., Champion, J., Pierrot-Deseilligny, M., Heggy, E., Jacquemoud, S. (2013) An advanced photogrammetric method to measure surface roughness: Application to volcanic terrains in the Piton de la Fournaise, Reunion Island. Remote Sensing of Environment 135, 1-11.

Brubaker, K.M., Myers, W.L., Drohan, P.J., Miller, D.A., Boyer, E.W. (2013) The use of LiDAR terrain data in characterizing surface roughness and microtopography. Applied and Environmental Soil Science 2013.

Cavalli, M., Marchi, L. (2008) Characterisation of the surface morphology of an alpine alluvial fan using airborne LiDAR. Nat. Hazards Earth Syst. Sci 8, 323-333.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J. (2007) Random forests for classification in ecology. Ecology 88, 2783-2792.

Davis, C., Reisinger, T. (1990) Evaluating terrain for harvesting equipment selection. Journal of forest Engineering 2, 9-16.

de Moraes Goncalves, J.L., Stape, J.L., Laclau, J.-P., Smethurst, P., Gava, J.L. (2004) Silvicultural effects on the productivity and wood quality of eucalypt plantations. Forest Ecology and Management 193, 45-61.

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P. (2013) Application of the topographic position index to heterogeneous landscapes. Geomorphology 186, 39-49.

Drăguţ, L., Blaschke, T. (2006) Automated classification of landform elements using object-based image analysis. Geomorphology 81, 330-344.

Dube, T., Mutanga, O., Abdel-Rahman, E.M., 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, 3751-3772.

Duncanson, L., Dubayah, R., Cook, B., Rosette, J., Parker, G. (2015) The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation. Remote Sensing of Environment 168, 102-112.

Dye, M., Mutanga, O., Ismail, R. (2012) Combining spectral and textural remote sensing variables using random forests: Predicting the age of Pinus patula forests in KwaZulu-Natal, South Africa. Journal of Spatial Science 57, 193-211.

Ediriweera, S., Danaher, T., Pathirana, S. (2016) The influence of topographic variation on forest structure in two woody plant communities: A remote sensing approach. Forest Systems 25, 049.

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

Ferry, B., Morneau, F., Bontemps, J.D., Blanc, L., Freycon, V. (2010) Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest. Journal of ecology 98, 106-116.

Florinsky, I.V., Eilers, R.G., Manning, G., Fuller, L. (2002) Prediction of soil properties by digital terrain modelling. Environmental Modelling & Software 17, 295-311.

Freeman, T.G. (1991) Calculating catchment area with divergent flow based on a regular grid. Computers & Geosciences 17, 413-422.

Fuller, W.A. (1987) Measurement Error Models. Wiley: New York, NY, USA.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C. (2010) Variable selection using random forests. Pattern Recognition Letters 31, 2225-2236.

Godsmark, R., (2013) The South African forestry and forest products industry 2011, Pretoria: Forestry South Africa.

González, D., Becker, J., Torres, E., Albistur, J., Escudero, M., Fuentes, R., Hinostroza, H., Donoso, F., SA, F.M., (2008) Using LiDAR technology in forestry harvest planning, Proceedings of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory, Heriot-Watt University, Edinburgh, UK, 17-19 September, 2008. SilviLaser 2008 Organizing Committee, pp. 437-445.

Granitto, P.M., Furlanello, C., Biasioli, F., Gasperi, F. (2006) Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemometrics and intelligent laboratory systems 83, 83-90.

Grohmann, C.H. (2015) Effects of spatial resolution on slope and aspect derivation for regional-scale analysis. Computers & Geosciences 77, 111-117.

Grohmann, C.H., Smith, M.J., Riccomini, C. (2011) Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland. Geoscience and Remote Sensing, IEEE Transactions on 49, 1200-1213.

Grömping, U. (2012) Variable importance assessment in regression: linear regression versus random forest. The American Statistician.

Guisan, A., Weiss, S.B., Weiss, A.D. (1999) GLM versus CCA spatial modeling of plant species distribution. Plant Ecology 143, 107-122.

Halmy, M.W.A., Gessler, P.E., Hicke, J.A., Salem, B.B. (2015) Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. Applied Geography 63, 101-112.

Hassan, R.M. (1999) Correcting measures of national income for environmental values: The case of water abstraction and carbon sink externalities of industrial plantations in South Africa. Southern African Forestry Journal 184, 1-11.

Hawbaker, T.J., Gobakken, T., Lesak, A., Trømborg, E., Contrucci, K., Radeloff, V. (2010) Light detection and ranging-based measures of mixed hardwood forest structure. Forest Science 56, 313-326.

Heung, B., Bulmer, C.E., Schmidt, M.G. (2014) Predictive soil parent material mapping at a regional-scale: a random forest approach. Geoderma 214, 141-154.

Hill, A., Breschan, J., Mandallaz, D. (2014) Accuracy assessment of timber volume maps using forest inventory data and LiDAR canopy height models. Forests 5, 2253-2275.

Hjerdt, K.N., McDonnell, J.J., Seibert, J., Rodhe, A., (2004) A new topographic index to quantify downslope controls on local drainage Water Resources Research

Hoechstetter, S., Walz, U., Dang, L., Thinh, N. (2008) Effects of topography and surface roughness in analyses of landscape structure—a proposal to modify the existing set of landscape metrics. Landscape Online 3, 1-14.

Hollaus, M., Aubrecht, C., Höfle, B., Steinnocher, K., Wagner, W. (2011) Roughness mapping on various vertical scales based on full-waveform airborne laser scanning data. Remote Sensing 3, 503-523.

Hollaus, M., Höfle, B. (2010) Terrain roughness parameters from full-waveform airborne LIDAR data. na.

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.

Hutchinson, M. (1989) A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. Journal of Hydrology 106, 211-232.

Hyde, P., Dubayah, R., Walker, W., Blair, J.B., Hofton, M., Hunsaker, C. (2006) Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sensing of Environment 102, 63-73.

International, L.R. (2015) SAPPI Southern Africa Ltd, Forests Survey Report

Ismail, R., Mutanga, O. (2011) Discriminating the early stages of Sirex noctilio infestation using classification tree ensembles and shortwave infrared bands. International Journal of Remote Sensing 32, 4249-4266.

Ismail, R., Mutanga, O., Bob, U. (2007) Forest health and vitality: the detection and monitoring of Pinus patula trees infected by Sirex noctilio using digital multispectral imagery. Southern Hemisphere Forestry Journal 69, 39-47.

Ismail, R., Mutanga, O., Kumar, L. (2010) Modeling the potential distribution of pine forests susceptible to sirex noctilio infestations in Mpumalanga, South Africa. Transactions in GIS 14, 709-726.

Iwahashi, J., Pike, R.J. (2007) Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86, 409-440.

Jakubowski, M.K., Guo, Q., Kelly, M. (2013) Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sensing of Environment 130, 245-253.

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

Kasai, M., Ikeda, M., Asahina, T., Fujisawa, K. (2009) LiDAR-derived DEM evaluation of deep-seated landslides in a steep and rocky region of Japan. Geomorphology 113, 57-69.

Kulkarni, V.Y., Sinha, P.K. (2013) Random forest classifiers: a survey and future research directions. International Journal of Advanced Computing 36, 1144-1153.

Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J. (2002) Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. BioScience 52, 19-30.

Li, Y., Härdtle, W., Bruelheide, H., Nadrowski, K., Scholten, T., von Wehrden, H., von Oheimb, G. (2014) Site and neighborhood effects on growth of tree saplings in subtropical plantations (China). Forest Ecology and Management 327, 118-127.

Liaw, A., Wiener, M., (2012) Random Forest: Breiman and Cutler's Random Forests for Classification and Regression. R Package Version 4.6-7.

Lukovic, J.B., Bajat, B.J., Kilibarda, M.S., Filipovic, D.J. (2015) High resolution grid of potential incoming solar radiation for Serbia. Thermal Science 19, S427-S435.

Maack, J., Lingenfelder, M., Weinacker, H., Koch, B. (2016) Modelling the standing timber volume of Baden-Württemberg—A large-scale approach using a fusion of Landsat, airborne LiDAR and National Forest Inventory data. International Journal of Applied Earth Observation and Geoinformation 49, 107-116.

May, R., Van Dijk, J., Wabakken, P., Swenson, J.E., Linnell, J.D., Zimmermann, B., Odden, J., Pedersen, H.C., Andersen, R., Landa, A. (2008) Habitat differentiation within the large-carnivore community of Norway's multiple-use landscapes. Journal of Applied Ecology 45, 1382-1391.

Melton, M.A. (1965) The geomorphic and paleoclimatic significance of alluvial deposits in southern Arizona. The Journal of Geology, 1-38.

Möller, M., Volk, M., Friedrich, K., Lymburner, L. (2008) Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. Journal of Plant Nutrition and Soil Science 171, 419-430.

Morris, A.R., Anderson, F.S., Mouginis-Mark, P.J., Haldemann, A.F., Brooks, B.A., Foster, J. (2008) Roughness of Hawaiian volcanic terrains. Journal of Geophysical Research: Planets (1991–2012) 113.

Mulder, V., De Bruin, S., Schaepman, M., Mayr, T. (2011) The use of remote sensing in soil and terrain mapping—A review. Geoderma 162, 1-19.

Navarro-Cerrillo, R.M., del Campo, A.D., Ceacero, C.J., Quero, J.L., de Mena, J.H. (2014) On the importance of topography, site quality, stock quality and planting date in a semiarid plantation: Feasibility of using low-density LiDAR. Ecological Engineering 67, 25-38.

Nurminen, K., Karjalainen, M., Yu, X., Hyyppä, J., Honkavaara, E. (2013) Performance of dense digital surface models based on image matching in the estimation of plot-level forest variables. ISPRS Journal of Photogrammetry and Remote Sensing 83, 104-115.

Olaya, V., Conrad, O. (2009) Geomorphometry in SAGA. Developments in Soil Science 33, 293-308.

Park, S., McSweeney, K., Lowery, B. (2001) Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103, 249-272.

Patil, M.B., Desai, C.G., Umrikar, B.N. (2012) Image classification tool for land use/land cover analysis: A comparative study of maximum likelihood and minimum distance method. International Journal of Geology Earth Environmental Sciience 2, 189-196.

Peerbhay, K., Mutanga, O., Lottering, R., Ismail, R. (2016) Mapping Solanum mauritianum plant invasions using WorldView-2 imagery and unsupervised random forests. Remote Sensing of Environment 182, 39-48.

Podobnikar, T., Stancic, Z., Oštir, K., (2000) Data integration for the DTM production, Kosmatin Fras M., Mussio L., Crosilla F.(ur.): ISPRS, Proceeedings of the Workshop: International cooperation and technology transfer, Ljubljana.

Popescu, S.C., Wynne, R.H., Nelson, R.F. (2003) Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. Canadian Journal of Remote Sensing 29, 564-577.

Popit, T., Verbovšek, T. (2013) Analysis of surface roughness in the Sveta Magdalena paleo-landslide in the Rebrnice area. RMZ–M&G 60, 197-204.

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.

R Development Core Team (2008) R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., Astrup, R. (2014) Comparison of four types of 3D data for timber volume estimation. Remote Sensing of Environment 155, 325-333.

Reutebuch, S.E., Andersen, H.-E., McGaughey, R.J. (2005) Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. Journal of Forestry 103, 286-292.

Riley, S.J., DeGloria, S.D., Elliot, R. (1999) A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain Journal of sciences 5, 23-27.

Rodríguez-Caballero, E., Cantón, Y., Chamizo, S., Afana, A., Solé-Benet, A. (2012) Effects of biological soil crusts on surface roughness and implications for runoff and erosion. Geomorphology 145, 81-89.

Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C. (2012) Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sensing of Environment 121, 93-107.

Sankey, J.B., Glenn, N.F., Germino, M.J., Gironella, A.I.N., Thackray, G.D. (2010) Relationships of aeolian erosion and deposition with LiDAR-derived landscape surface roughness following wildfire. Geomorphology 119, 135-145.

Sanson, G., Stolk, R., Downes, B. (1995) A new method for characterizing surface roughness and available space in biological systems. Functional Ecology, 127-135.

Sappington, J., Longshore, K.M., Thompson, D.B. (2007) Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. The Journal of Wildlife Management 71, 1419-1426.

Saremi, H., Kumar, L., Turner, R., Stone, C. (2014a) Airborne LiDAR derived canopy height model reveals a significant difference in radiata pine (Pinus radiata D. Don) heights based on slope and aspect of sites. Trees 28, 733-744.

Saremi, H., Kumar, L., Turner, R., Stone, C., Melville, G. (2014b) DBH and height show significant correlation with incoming solar radiation: a case study of a radiata pine (Pinus radiata D. Don) plantation in New South Wales, Australia. GIScience & Remote Sensing 51, 427-444.

Saremi, H., Kumar, L., Turner, R., Stone, C., Melville, G. (2014c) Impact of local slope and aspect assessed from LiDAR records on tree diameter in radiata pine (Pinus radiata D. Don) plantations. Annals of Forest Science 71, 771-780.

Scarpone, C., Schmidt, M.G., Bulmer, C.E., Knudby, A. (2017) Semi-automated classification of exposed bedrock cover in British Columbia's Southern Mountains using a Random Forest approach. Geomorphology 285, 214-224.

Schmidt, J., Hewitt, A. (2004) Fuzzy land element classification from DTMs based on geometry and terrain position. Geoderma 121, 243-256.

Shackleton, C.M., Shackleton, S.E., Buiten, E., Bird, N. (2007) The importance of dry woodlands and forests in rural livelihoods and poverty alleviation in South Africa. Forest policy and economics 9, 558-577.

Shepard, M.K., Campbell, B.A., Bulmer, M.H., Farr, T.G., Gaddis, L.R., Plaut, J.J. (2001) The roughness of natural terrain: A planetary and remote sensing perspective. Journal of Geophysical Research: Planets (1991–2012) 106, 32777-32795.

Sørensen, R., Seibert, J. (2007) Effects of DEM resolution on the calculation of topographical indices: TWI and its components. Journal of Hydrology 347, 79-89.

Sterenczak, K., Zasada, M., Brach, M. (2013) The accuracy assessment of DTM generated from LIDAR data for forest area–a case study for scots pine stands in Poland. Baltic Forestry 19, 252-262.

Straub, C., Tian, J., Seitz, R., Reinartz, P. (2013) Assessment of Cartosat-1 and WorldView-2 stereo imagery in combination with a LiDAR-DTM for timber volume estimation in a highly structured forest in Germany. Forestry 86, 463-473.

Streutker, D.R., Glenn, N.F., Shrestha, R. (2011) A slope-based method for matching elevation surfaces. Photogrammetric Engineering & Remote Sensing 77, 743-750.

Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P. (2003) Random forest: a classification and regression tool for compound classification and QSAR modeling. Journal of chemical information and computer sciences 43, 1947-1958.

Tesfamichael, S., Ahmed, F., Van Aardt, J. (2010a) 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.

Tesfamichael, S., Van Aardt, J., Ahmed, F. (2010b) Estimating plot-level tree height and volume of Eucalyptus grandis plantations using small-footprint, discrete return lidar data. Progress in physical Geography 34, 515-540.

Treeratpituk, P., Giles, C.L., (2009) Disambiguating authors in academic publications using random forests, Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. ACM, pp. 39-48.

Tuominen, S., Pekkarinen, A. (2005) Performance of different spectral and textural aerial photograph features in multi-source forest inventory. Remote Sensing of Environment 94, 256-268.

Ullrich, A., Studnicka, N., Hollaus, M., Briese, C., Wagner, W., Doneus, M., Mücke, W., (2007) Improvements in DTM generation by using full-waveform airborne laser scanning data, 7Th International Conference On" Laser Scanning and Digital Aerial Photography. Today And Tomorrow", Moscow, Russia, pp. 2007-2007.2012.

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

Verkerk, P.J., Levers, C., Kuemmerle, T., Lindner, M., Valbuena, R., Verburg, P.H., Zudin, S. (2015) Mapping wood production in European forests. Forest Ecology and Management 357, 228-238.

Visser, R., Spinelli, R. (2012) Determining the shape of the productivity function for mechanized felling and felling-processing. Journal of Forest Research 17, 397-402.

Weiss, A., (2001) Topographic position and landforms analysis, Poster presentation, ESRI user conference, San Diego, CA.

White, R.A., Dietterick, B.C., Mastin, T., Strohman, R. (2010) Forest roads mapped using LiDAR in steep forested terrain. Remote Sensing 2, 1120-1141.

Wilson, J.P. (2012) Digital terrain modeling. Geomorphology 137, 107-121.

Wilson, J.P., Gallant, J.C. (2000) Digital terrain analysis. Terrain analysis: Principles and applications, 1-27.

Wulder, M. (1998) Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Progress in physical Geography 22, 449-476.

Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., White, J.C. (2008) The role of LiDAR in sustainable forest management. The Forestry Chronicle 84, 807-826.

Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T. (2012) Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment 121, 196-209.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R. (2011) Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. ISPRS Journal of Photogrammetry and Remote Sensing 66, 28-37.

Yuan, F., Sawaya, K.E., Loeffelholz, B.C., Bauer, M.E. (2005) Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. Remote Sensing of Environment 98, 317-328.

Zhang, H., Dai, X., Sun, F., Yuan, J., (2016) Terrain classification in field environment based on Random Forest for the mobile robot, Control Conference (CCC), 2016 35th Chinese. IEEE, pp. 6074-6079.

Zomer, R., Ustin, S., Ives, J. (2002) Using satellite remote sensing for DEM extraction in complex mountainous terrain: landscape analysis of the Makalu Barun National Park of eastern Nepal. International Journal of Remote Sensing 23, 125-143.