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

Roxanne Munsamy, Michael Gebreslasie, Kabir Peerbhay, Riyad Ismail

University of KwaZulu-Natal, School of Agricultural, Earth and Environmental Sciences, Discipline of Geography, P/Bag X01, KwaZulu-Natal, South Africa.
Peerbhaykabir@gmail.com;Tel: +27 333862314; mobile: +27 738552166

DOI: http://dx.doi.org/10.4314/sajg.v9i2.9

Abstract

Accurate multi-source forest inventory attributes are necessary for estimating productivity and timber stock in commercial forest plantations. This study aims to uncover the effects of terrain variation on the growth of even aged Eucalyptus forest species using Light Detection and Ranging (LiDAR) topographical variables. Using 32 generated variables at 5 different spatial resolutions (1m, 3m, 5m, 7m, 9m), the random forest (RF) regression successfully revealed variations for structural attributes such as volume (Vol/ha), dominant tree height (HtD), mean tree height (Htm), and diameter breast heights (DBH). Results indicate that smaller spatial resolutions performed better for younger stands while larger resolutions produced the best results for mature stands. Using the multi-resolution approach results improved with variable selection. Incoming solar radiation and slope variables were among the most important terrain variables for modelling forest structural variability. The findings from this study demonstrates the value of stratifying forest productivity across the commercial forest landscapes of South Africa.

1. Introduction

Terrain derived variables such as slope (Wilson and Gallant, 2000), aspect (Grohmann, 2015) and local curvature (Freeman, 1991) impact productivity levels within forest plantations (Maack *et al*., 2016). Modelling the variation of such variables across the plantation landscape then provides valuable information related to resource production such as forest structural attributes. Several studies have demonstrated the statistically significant relationship between variables derived from Light Detection and Ranging (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 have investigated the benefits of terrain variables derived from LiDAR to assess forest productivity inventory measurements. An increase in the detail of the knowledge and the understanding of the role of terrain variability, specific to plantation sites would lead to the opportunity to effectively manipulate and homogenise stands (Ediriweera *et al*. 2014, Li *et al*. 2014), thereby increasing site productivity and decreasing the heterogeneity associated with large commercial plantations.

Recently, Ediriweera *et al*. (2016) aimed to characterise the variation in vegetation growth in relation to terrain. They calculated the Terrain Wetness Index (TWI), potential solar insolation, slope and elevation derived from LiDAR data for both an open canopy eucalypt forest and a closed subtropical rainforest within Australia. Using a general linear model approach, 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 topographical variations in terrain than subtropical rainforests (Ediriweera *et al.*, 2016). Simlarly, Saremi *et al.* (2014a) employed the use of a mixed linear model to investigate the relationship between topographical factors (i.e. slope and aspect) derived from LiDAR against the mean tree height (Htm) of radiate pine (*Pinus radiate. D.Don*) aged at 9 and 34 years. The result was based on one continuous dependent variable with several explanatory variables and showed that the derived height estimates were highly correlated with field heights for the 9 year ($R^2 = 0.90$ and RMSE = 0.66) and for the 34-year-old site ($R^2 = 0.87$ and RMSE = 1.49). The results obtained from this study also showed that taller trees were present in low slopes with southerly aspects, whilst short trees were found on steeper slopes with northerly aspects.

In a subsequent study, they applied a mixed linear model to quantify the relationship between DBH and height classes coupled with slope and aspect variables (Saremi *et al.* 2014c). The outcome showed that greater diameter breast heights (DBH) was found in gentle slopes with southerly aspects (Saremi *et al.* 2014c) while further investigations found that micro-scale variations of DBH and mean tree height (Htm) could be quantified and based on incoming solar radiation (Saremi *et al.* 2014b). The results reported for height were $R^2 = 0.58$ and for DBH were $R^2 = 0.60$ in mature stands; and $R^2 = 0.58$ and $R^2 = 0.60$ for young stands respectively. Trees that displayed larger measurements of DBH and which were taller were shown to be in areas with lower incoming solar radiation with high soil moisture, with variation existing within stands of the same age category (Saremi *et al*. 2014b). Since the benefits of utilizing LiDAR and its derivatives such as elevation, slope and aspect, within the forestry sector have previously proven successful (Popescu *et al*. 2003, González *et al.* 2008, Wulder *et al*. 2012), the challenge of choosing an appropriate statistical algorithm to fully exploit the information linked with LiDAR datasets becomes eminent.

Regression techniques including Artificial Neural Networks (ANN) (Svetnik *et al.* 2003); Multiple Linear Regression (MLR) and k-nearest neighbour (KNN); Partial Least Squares (Duncanson *et al.* 2015); Support Vector Regression (Jakubowski *et al.* 2013); Bayesian Model Averaging (Verkerk *et al*. 2015); Generalized Addictive Model (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 remotely sensed data. While the majority of methods provide successful prediction performances, the drawback 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) utilised RF to model forest structural attributes using aerial imagery and LiDAR data respectively within a boreal forest in Southern Finland. The results demonstrated that the 26 derived tree features showed high correlations between the observed and predicted tree height (R^2=0.93 and RMSE=10.03%), DBH (R^2=0.79 and RMSE=21.35%) and volume (R^2=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). These results show that LiDAR-derived forest attributes performed better than information derived from digital aerial photography. The results also showed high correlations for tree heights (R^2=0.98 and RMSE=0.97), DBH (R^2=0.94 and RMSE=2.16m) and volume (R^2=0.93 and RMSE=37.58m^3/ ha).

In summary, LiDAR data 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 effectively uncover the effect terrain has on forest structural attributes within the *Eucalyptus* plantation environment. Thus, the aim of this study is therefore to explain the variability in forest structural attributes such as height, pulpwood volumes and DBH across commercial forestry terrains when using LiDAR-derived terrain variables.

2. Methods and Materials

2.1. Study Area

The study was conducted in Sappi's Riverdale plantation that extents over 2503 ha near the town of Richmond, in the Midlands region of KwaZulu-Natal, South Africa (Figure 1). 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°C with mean annual precipitation reported at 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. Commercial *Eucalyptus* species (*E. grandis* and *E. dunni*) provides 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 is the main species grown due to its growth rate being favourable in KwaZulu-Natal province (Godsmark 2013). These stands are established at 1667 trees per hectare as per the pulpwood regime, and are harvested on average between 6 and 7 years (Duncanson *et al*. 2015).

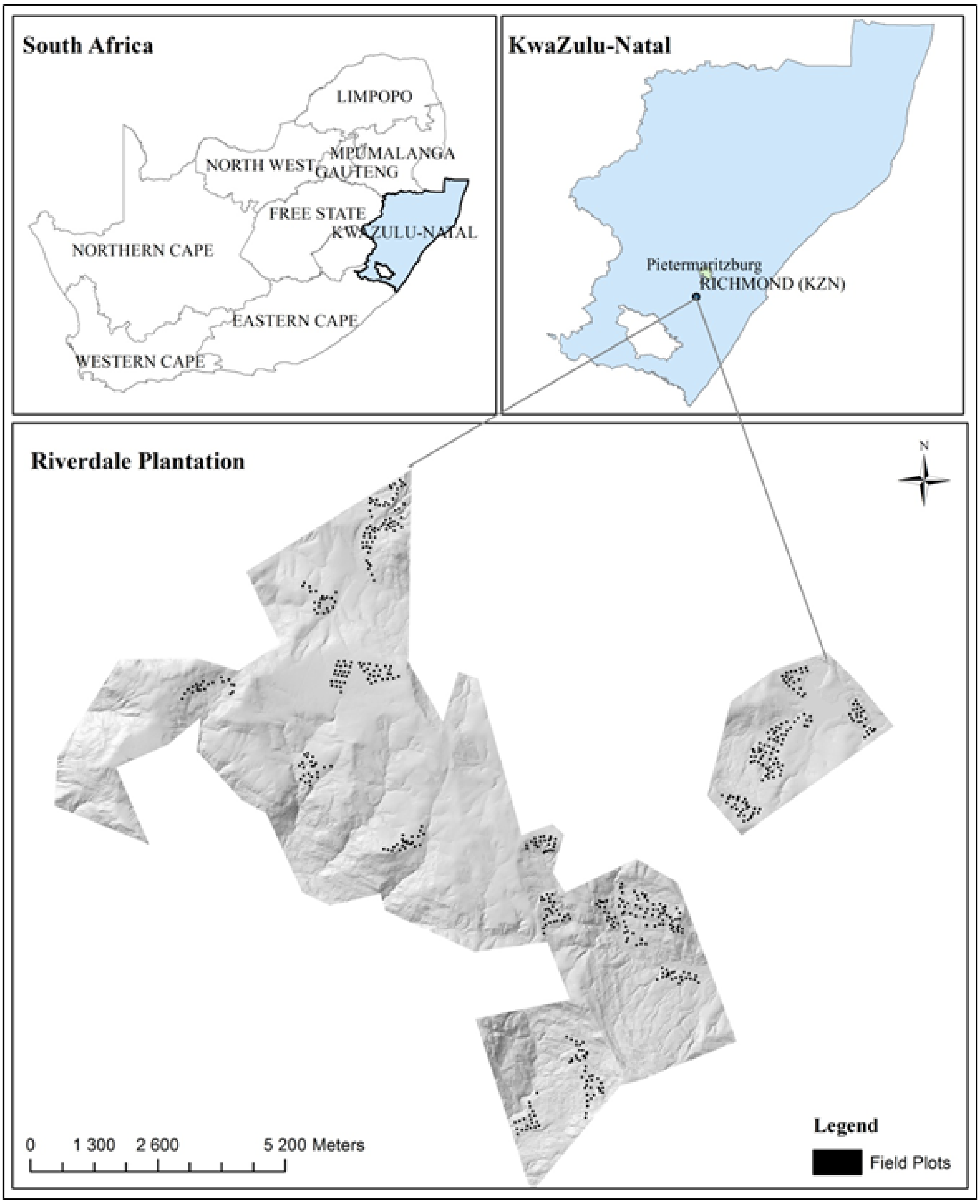

Figure 1. The location of Sappi's Riverdale plantation.

2.2. Field Surveys

Field surveys were conducted from the 12th to the 22nd May 2014 whereby a total of 502 plots spanning over 27 compartments at the Riverdale plantation were covered. The following productivity tree attributes were included from the inventory surveys: volume (Vol/ha), mean dominant height (HtD), mean height (Htm) and diameter at breast height (DBH). A Global Positioning System (GPS) Device with sub-meter accuracy (10 cm) 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 Haglof Digitech Calliper instrument while tree heights were measured using a Vertex IV Ultrasonic Hypsometer. To consider the species and age variation in this study, the plots were partitioned into datasets based on the available species compartments and respective age. *Eucalyptus grandis* was separated into age categories of young (3 – 6 year, n = 151) and mature stands (7 – 10 years, n = 137), while *Eucalyptus dunnii* also included young (2 – 5 years, n = 104) and mature stands (6 – 9 years, n = 110).

2.3. LiDAR Data Acquisition

The LiDAR data was acquired by a competent aerial survey vendor, Land Resources International (LRI) who is based in South Africa. The LiDAR surveys were conducted between the 15th and the 22nd of March 2014 at the Riverdale plantation to coincide with the field sampling assessment. The surveyed point cloud data was regularised and filtered for noise before being released in LAS format. Subsequently, a very high resolution digital terrain model (DTM) was generated having a cell size of 1 m using LAS tools and predefined filters for generating ground and non-ground returns (ArcGIS 2016). 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. The LiDAR system used a Riegl Q560 laser with a minimum measurement range of 50 m, a ranging accuracy of 20 mm and an angle measurement resolution of 0.001^{o}.

Table 1: LiDAR flight and sensor instrument parameters

LiDAR Survey Parameters	Unit	Measurement
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	5

2.4. Extracted Terrain Variables

Terrain variables were calculated using the nearest neighbour re-sampling technique and 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 (ArcGIS 2016) were then used to extract the zonal statistics for each of the terrain variables (n = 32) at a plot level.

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 (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 its neighbouring cells	Freeman (1991)
Upslope Curvature	UC	Distance of weighted average of local curvature	Freeman (1991)
Local Upslope Curvature	LUC	Local curvature on flow direction as a sum of neighbour cells that are facing upwards	Freeman (1991)
Downslope Curvature	DC	Local curvature on flow direction as a sum of neighbour cells facing downslope	Freeman (1991)
Local Downslope Curvature	LDC	Calculates local curvature as a sum of neighbour cells	Freeman (1991)
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)

2.5. Regression Analysis

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 the 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). The coefficient of determination (R^2) was used to assess the relationship between the field data and the LiDAR-derived DTM variables under study, whereby values closer to 1 predict better results. Model assessment was based on test data which constituted 30%, while the training data comprised of the remaining 70%.

2.5.2. Multi-resolution analysis

In this section, the following was examined:

i. The terrain variables (n = 32) 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 young and mature *Eucalyptus* species, a backward feature selection approach was used to reduce the number of input terrain variables to enhance the quality of the terrain variables that best explain the variation among the attributes

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). Therefore in this study, RF variable importance is based on the permutation accuracy method.

2.5.4. Random Forest Variable Selection

Variable selection becomes important when dealing with multiple input variables for a prediction model, as many predictors may lead to a decrease in model performance. A variable selection method based on a 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 does not contribute to the predictive accuracy of the model (Granitto *et al.* 2006). The algorithm runs until all unnecessary variables are progressively dropped. The permutation measure of variable importance, determines the percentage increase in mean square error (MSE) when the OOB samples for each variable is permuted, while the others remain unchanged.

3. Results

The descriptive statistics 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

3.1. Spatial Resolution Analysis using LiDAR DTM variables

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 three selected terrain variables re-sampled to the various spatial resolutions is to be found in figure 2 below.

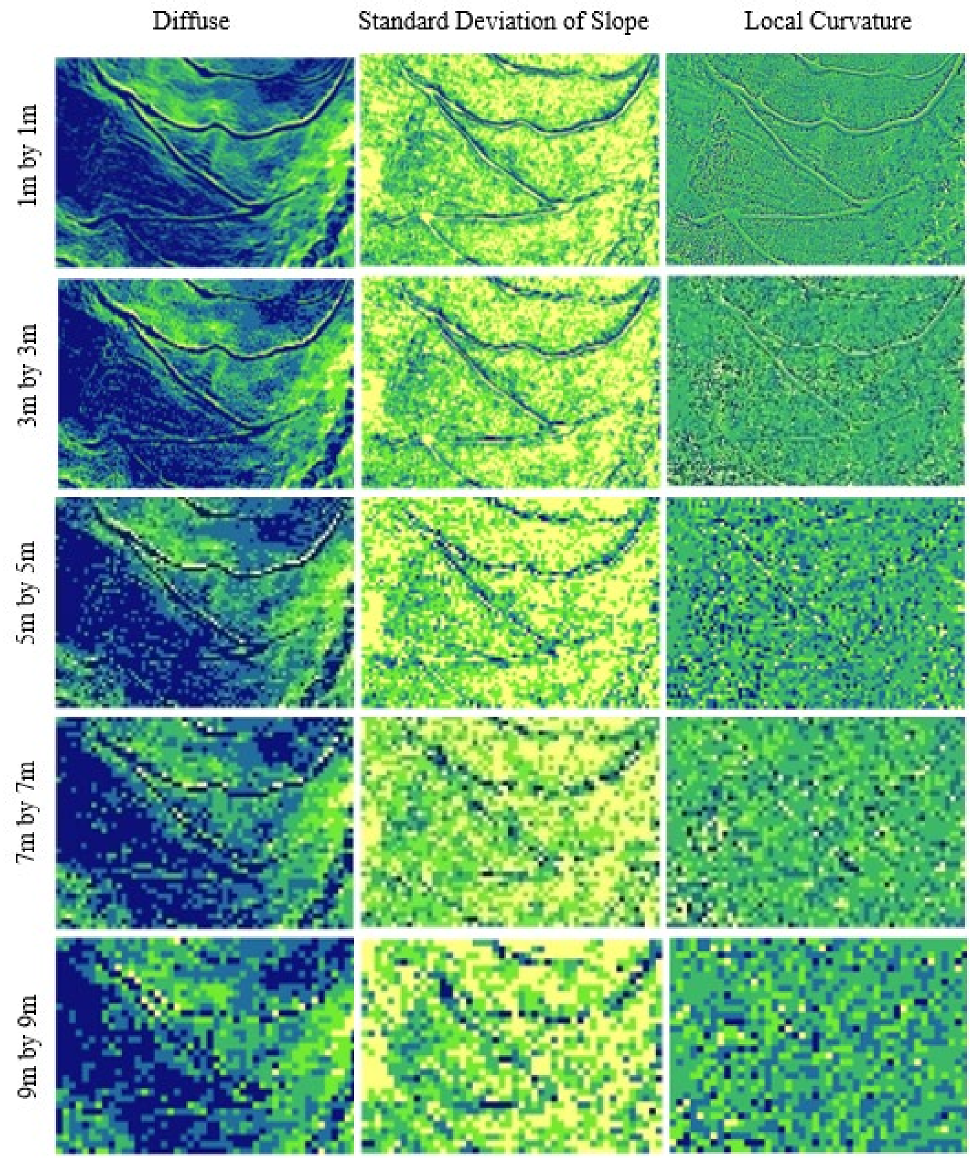

Figure 2: The effect of different spatial resolutions for diffuse, standard deviation of slope, and local curvature terrain variables.

3.2. Predicating *Eucalyptus grandis* attributes using LiDAR terrain variables

Results indicate successful regressions for the young (*a*) and mature (*b*) *E. grandis* stands (figure 3). Noticeable for the young stands, is that the best predictive accuracy was produced at 1 m spatial resolution with a R^2 value of 0.68 and a RMSE of 1.22 m for HtD. For Htm, a 3 m spatial resolution produced the best accuracy (R^2 = 0.70, RMSE of 0.97 m) while for DBH, the best accuracy was obtained at 5 m spatial resolution with a R^2 value of 0.71 and a RMSE of 0.81 cm. With regards to Vol/ha, a 9 m spatial resolution produced the best R^2 value of 0.60 with a RMSE of 32.11 m^3/ha.

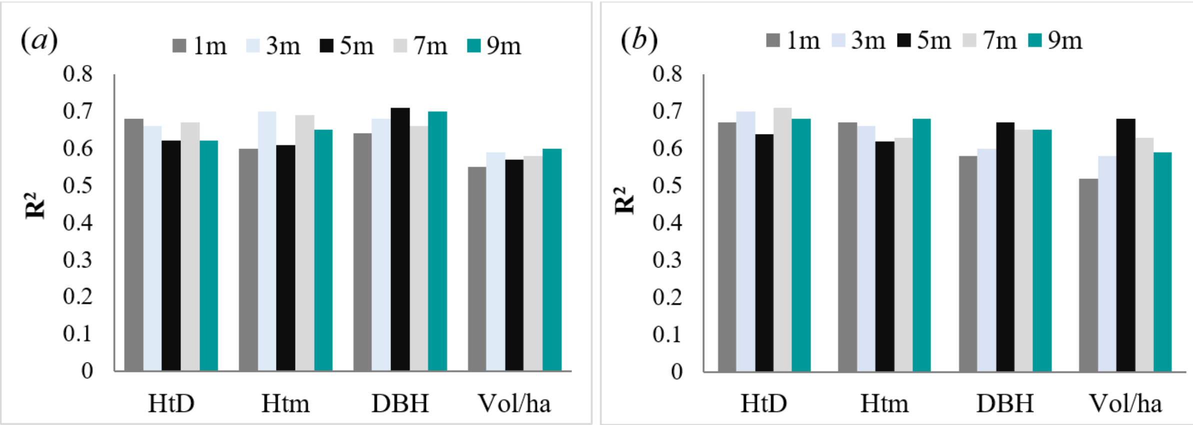

Figure 3: Random forest predictive accuracies (R^2) obtained for (a) young and (b) mature *E. grandis* stands.

When assessing the results for mature *E. grandis* stands, a 7 m spatial resolution performed best for HtD and produced a R^2 of 0.71 with a RMSE of 1.04 m. For Htm, a 9 m spatial resolution yielded the highest accuracy at with a R^2 value of 0.68 and a RMSE of 1.15 m. The DBH model yielded the highest results using a 5 m spatial resolution with a R^2 value of 0.67 and a RMSE of 0.82 cm. Vol/ha yielded the best accuracy also at a 5 m spatial resolution with a reported R^2 value of 0.68 and a RMSE of 22.56 m^3/ha.

3.3. Predicting *Eucalyptus dunnii* attributes using LiDAR terrain variables

Figure 4 depicts the results for young (*a*) and mature (*b*) *E.dunnii* stands. For young stands, a 3 m spatial resolution obtained the best accuracy for HtD with a R^2 value of 0.72 and a RMSE of 1.33 m. A 1 m spatial resolution produced the best accuracy for Htm with a R^2 of 0.69 and a RMSE of 1.15 m. For DBH, the best R^2 of 0.70 and a RMSE value of 1.88 cm using a 7 m resolution. Finally, the highest accuracy for volume was obtained using a 5 m resolution with an R^2 value of 0.61 and a RMSE of 19.36 m^3/ha.

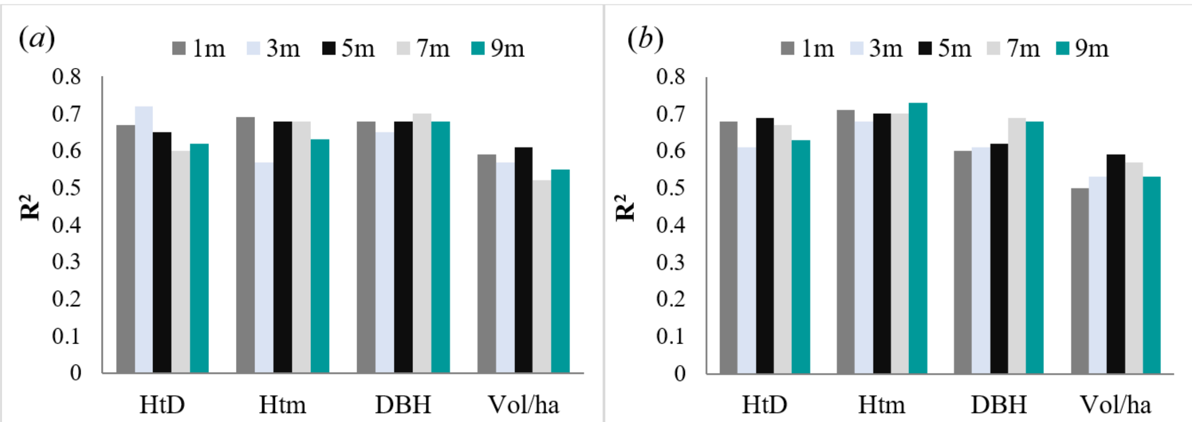

Figure 4: Random forest predictive accuracies (R^2) obtained for (a) young and (b) mature *E.dunnii* stands.

When investigating the results for mature *E.dunnii* stands using the LiDAR terrain variables, HtD was predicted with the best R^2 of 0.69 with a RMSE value of 2.73 m using a 5 m spatial resolution. Htm yielded a high R^2 of 0.73 with a RMSE of 1.48 m and only when using a 9 m resolution. The DBH model yielded the best R^2 value of 0.69 and a RMSE of 0.88 cm and occurred at a 7 m spatial resolution. For volume, the best model was produced at a 5 m spatial resolution with a reported R^2 of 0.59 and a RMSE of 41.08 m^3/ha.

3.4. Predicting *Eucalyptus* structural attributes using multi-resolution LiDAR

When combining the resolutions for each LiDAR variable, RF results were successful for young *E. grandis* stands (Figure 5a) when assessing HtD (R^2 = 0.71, RMSE = 0.96 m) and Htm (R^2 = 0.70, RMSE = 0.71 m) respectively. The RF model yielded an R^2 of 0.69 and a RMSE of 1.93 cm for DBH, while Vol/ha was predicted with an R^2 of 0.64 and a RMSE value of 47.80 m^3/ha. Nonetheless, when dealing with mature *E. grandis* stands, RF models produced comparative results. The highest predictive accuracy was produced for HtD (R^2 = 0.74, RMSE = 1.22 m) followed by Htm (R^2 = 0.72, RMSE of 0.99 m) and Vol/ha (R^2 value of 0.68, RMSE = 50.97 m^3/ha. DBH displayed reasonable accuracies (R^2 = 0.66, RMSE = 1.32 cm).

For young *E. dunni* stands (Figure 5b), the multi-resolution RF model performed well for HtD (R^2 = 0.72, RMSE = 0.89 m) and Htm (R^2 = 0.71, RMSE = 0.66 m). DBH (R^2 = 0.73, RMSE = 1.93 cm) and Vol/ha (R^2 = 0.61, RMSE = 27.78 m^3/ha) also produced relative accuracies. In the case of mature *E. dunnii* stands, the RF model produced successful for HtD (R^2 = 0.70, RMSE = 2.03 m) and Htm (R^2 = 0.70, RMSE = 1.17 m) with reasonable accuracies obtained for DBH (R^2 = 0.69, RMSE = 1.39 cm) and Vol/ha (R^2 = 0.62, RMSE = 55.67 cm^3/ha).

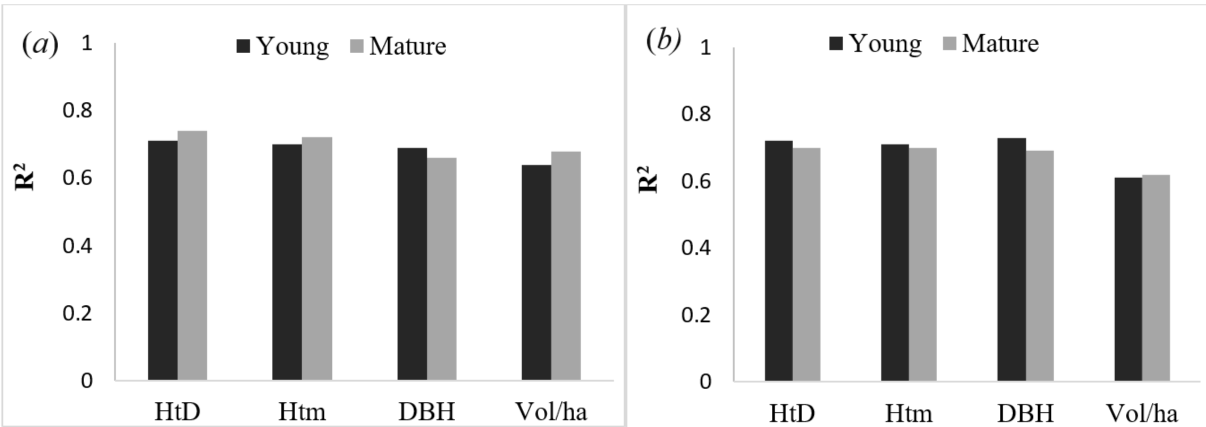

Figure 5: Coefficient of determination (R^2) for (a) young and mature *E. grandis* and (b) young and mature *E. dunnii*.

3.5. RF Variable Importance of Terrain Variables

In young *E. grandis*, diffuse solar radiation was returned as one of the most important terrain variable for estimating forest structural attributes. Direct solar radiation was highly important for both, young and mature *E. grandis* stands, 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* stands, diffuse solar radiation was also one of the most important terrain variables, as it provided the highest-ranking variable for all structural models in these categories with certain slope variables also contributing towards each model (Figure 6).

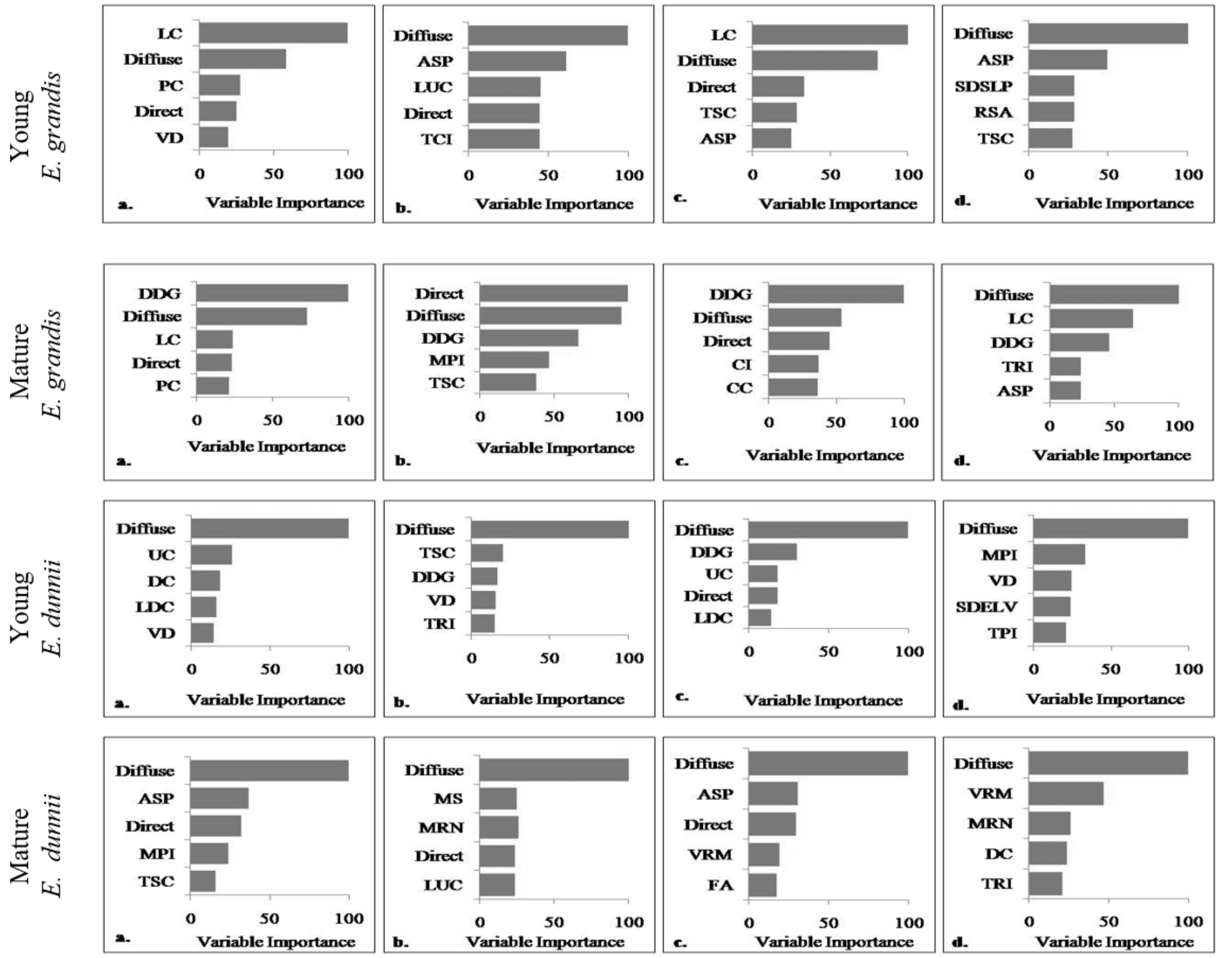

Figure 6: Variable Importance for young and mature *E. grandis* and *E. dunnii* for (a) HtD (b) Htm (c) Volume and (d) DBH.

3.6. Random Forest Variable Selection

In efforts to maximise the information provided when utilising multi-resolution LiDAR, a variable selection based on a RF-recursive feature elimination was implemented. In young *E. grandis* stands using variable selection improved the R^2 results for HtD, DBH and Vol/ha models at certain spatial resolutions (Table 4). For Htm, improvements in accuracy were evident, however no resolution produced better results that the optimal spatial resolution. Nonetheless, when modelling mature *E. grandis,* the majority of spatial resolutions displayed better R^2 values, to which the exception was for the DBH and Vol/ha. When assessing the performance of variable selection on young *E. dunnii* stands, all structural attributes produced some improvement in R^2 values and at different spatial resolutions. The same can be noticed when assessing accuracies for mature *E. dunnii* stands after variable selection. Results values that do not indicate an improved result after variable selection either decreased or remained the same.

Table 4: Coefficient of determination (R^2) for young and mature *E. grandis* and *E. dunnii* using RF variable selection. Values in bold indicate an increase in predictive accuracy after variable selection.

		E. grandis					*E. dunnii*				
		1m	3m	5m	7m	9m	1m	3m	5m	7m	9m
HtD	Young	**0.69**	**0.67**	0.60	0.65	0.58	**0.68**	0.7	**0.67**	0.58	0.6
	Mature	0.62	0.63	**0.66**	**0.73**	**0.70**	0.64	**0.63**	**0.73**	0.65	**0.64**
Htm	Young	0.56	0.64	0.58	0.6	0.63	0.65	**0.59**	0.61	0.63	0.63
	Mature	0.65	**0.63**	0.59	0.62	0.66	0.7	0.67	0.67	0.66	0.71
DBH	Young	0.63	0.67	**0.72**	0.63	0.68	0.65	**0.68**	0.64	0.7	0.69
	Mature	0.54	0.62	0.64	0.51	0.62	**0.63**	**0.63**	0.6	0.68	0.67
Vol	Young	0.5	0.55	**0.58**	0.57	**0.62**	0.53	0.52	**0.62**	**0.55**	**0.58**
	Mature	**0.56**	0.55	0.64	0.62	0.58	**0.53**	**0.56**	**0.61**	0.57	0.51

4. Discussion

This study investigated the effects of terrain variability on forest structural attributes within young and mature commercial eucalypt plantation species using LiDAR derives variables. 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 and slope.

4.1. Individual spatial resolution versus multi-resolution analysis

The results indicated that young and mature *E.grandis* stands may require different spatial resolutions for accurately predicting variations in terrain for forest structural attributes. More specifically, better accuracies were produced for HtD and Htm when using smaller (1 m and 3 m, respectively) spatial resolutions for young stands while in mature stands larger resolutions (7 m and 9 m, respectively) yielded the best results. Similar results were shown for DBH in younger stands (5 m) compared to mature stands (7 m) for producing the best predictive accuracies. While a larger spatial resolution (9 m) worked best for predicting volume for younger stands compared to mature stands (5 m) reasonable results were produced with accuracies greater than an R^2 of 0.60. For *E.dunnii,* HtD and Htm in younger stands were predicted most accurately using 3 m and 1 m spatial resolutions respectively with an optimal resolution of 7 m for DBH and 5 m for Vol/ha. In mature stands, 5 m and 9 m resolutions were ideal for HtD and Htm while 7 m and 5 m worked best for DBH and Vol/ha. These results confirm that micro-scale terrain variability does occur within commercial

Eucalypt plantations and that terrain heterogeneity can be dominated by different spatial scales (Navarro-Cerrillo *et al*. 2014).

Overall individual spatial resolution analysis indicated that there was no one spatial resolution that was consistent and the most accurate for predicating all forest attributes successfully in both eucalypt species. In addition, Vol/ha was the least accurate forest attribute predicted in both young and mature forest species. When assessing the utility of a multi-resolution analysis, results improved with greater consistency among all variables, besides for DBH which slightly reduced for *E. grandis* stands. In young stands, the multi-resolution approached provided improved results for HtD (3%) and Vol/ha (4%) while Htm remained the same (R^2 = 0.70). In mature stands, better results were obtained for HtD (3%) and Htm (4%), while Vol/ha showed no improvement. In young *E. dunnii* stands, HtD (2%), Htm (3%) and DBH (3%) produced better accuracies while Vol/ha remained the same. HtD (1%) and Vol/ha (3%) were the only two attributes with improved predictive accuracies for mature *E. dunnii* stands while DBH showed no improvement and Htm reduced in accuracy from an R^2 of 0.73 to 0.70. Overall, using a multi-resolution approach yielded successful results than an individual spatial resolution approach. The results obtained can be explained, since terrain details are often refined when DTMs are coarsened into larger spatial scales. Therefore, site-specific regions such as in commercial plantation forests, where variations in terrain can be affected by spatial variability in topography, may therefore require a multi-resolution analysis for forest structural attribute modelling.

4.2. Variable Selection of key LiDAR terrain variables

Reducing the variables within each multi-resolution model for both young and mature *E. grandis* and *E. dunnii* species, produced the highest predictive accuracies. This in particular for young and mature *E. grandis* stands for HtD (1% and 2% respectively) when using 11 and 16 terrain variables. Vol/ha showed best accuracies for young *E. grandis* (2%) using only 23 respectively while DBH (1%) also only improved for the younger stands using 15 variables. Htm produced a slight reduction in results for young (6%) and mature stands (2%) when using 28 variables only.

After performing variable selection on *E. dunnii* stands, the predictive accuracy for HtD reduced by 2% (R^2 = 0.70) using only 19 variables for young stands while improved for mature stand by 4% using 21 predictors. For DBH results were maintained (R^2 = 0.70) with a reduction in variables (22) in young stands while in mature stands an accuracy reduction of 1% was reported using 13 variables. Vol/ha improved by 1% and 2% in young and mature stands respectively using 24 and 27 variables. Those structural attributes that have improved or maintained accuracy with a reduced number if variables coincide with other studies that show improvements in the predictive accuracies obtained for forest structural attributes using variable selection methods (Treeratpituk and Giles 2009, Ismail *et al*. 2010, Genuer *et al*. 2010). In this regard, variable importance analysis showed that that not all predictor variables were important in each RF analysis. The diffuse and direct incoming solar radiation predictor variables displayed evidence pertaining to the height, DBH and Vol/ha stratification for both eucalypt species since the competition for light has been found to impact tree growth (Saremi *et al*. 2014b). For example, larger diameters, height and 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. The results also show that slope predictor variables had some contribution of importance in each predictive model and successfully captured the variations within the forest structural attributes under study. Further research is therefore required to quantify the effect and interaction between tree growth and variation in terrain.

4.3. Young and mature stand effects on terrain variation

This research has also shown that age categories between young, open canopies and mature closed canopies eucalyptus trees, display different degrees of structural 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 that obtained in the study conducted by 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, with an 80% variation found for younger (more open canopies) *E. propinqua* and *E. siderophloia* forest stands as compared with a 60% variation in mature (more closed canopies) conditions.

Different results were found for *E. dunnii* in this study, as higher variations were found for mature stands than for young stands. These results indicated that subspecies variations do occur in eucalypt plantation forests and for different terrain conditions. 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* is able to 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 target deliverables. This study therefore shows that whilst plantations contain similar climatic, topographic, soil, precipitation and silvicultural regimes, micro-scale variations within stands of even-aged plantations still do exist. Nonetheless, the accuracies of this study between structural attributes could have been influenced by anomalies within the LiDAR dataset since errors in topography may be propagated due to a lower solar angle when slopes face away from the sun (Rahlf *et al.* 2014). Additionally, mature stands are bigger and a considerable part of the Laser beam is blocked by the canopy before reaching the ground surface. On the contrary, with younger stands, a better penetration of the Laser Beam is allowed and therefore the DTM resolution may be higher. This inconsistency could affect the accuracy of the LiDAR derived datasets for younger and older stands and subsequently the results of this study. Future work may investigate the impact of forest canopy age on the accuracy and resolutions of DTM's. Whilst the cost associated with LiDAR is becoming feasible for forest industries to practice, DTM's could readily be utilised and used for future inventories within the forest, as it is expected that the terrain will remain relatively unchanged (Järnstedt *et al.* 2012).

5. Conclusion

For many decades forest managers have known the intrinsic value of attaining accurate forest productivity attributes at stand level in commercial plantation forests. The main aim of this study was to examine the topographical effect on forest structural attributes such as height (Htm and HtD), volume and DBH using variables derived from LiDAR data. Whilst the results have produced the level of accuracy necessary for operational use, they also 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 the variations that exist within stands of even-aged species.

6. Acknowledgments

We thank the National Research Foundation of South Africa for support (Grant Number 114898) and SAPPI forests-SA for allowing us to carry out our study under excellent working conditions.

7. 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.

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, 1-12.

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

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.

Ediriweera, S., Pathirana, S., Danaher, T., & Nichols, D. 2014. LiDAR remote sensing of structural properties of subtropical rainforest and eucalypt forest in complex terrain in North-eastern Australia. *Journal of Tropical Forest Science*, 397-408.

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.

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

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. 8th international conference on LiDAR applications in forest assessment and inventory, Edinburgh, UK.

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.

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

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.

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.

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.

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

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.

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.

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.

Peerbhay, K.Y., Mutanga, O., Ismail, R. 2015. Random forests unsupervised classification: The detection and mapping of solanum mauritianum infestations in plantation forestry using hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 8, 3107-3122.

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.

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.

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.

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.

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.

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.

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

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.