SCION

New Zealand Journal of Forestry Science

41 (2011) 191-205

www.scionresearch.com/nzjfs

published on-line:
09/11/2011

Determining an optimal model for processing lidar data at the plot level: results for a *Pinus radiata* plantation in New South Wales, Australia

Christine Stone[1,*], Trent Penman[2] and Russell Turner[1]

[1] *Forest Science Centre, NSW Primary Industries, P.O. Box 100, Beecroft, NSW, Australia, 2119*

[2] *Centre for Environmental Risk Management of Bushfires, Institute of Conservation Biology and Environmental Management, Wollongong, NSW, Australia, 2522*

(Received for publication 14 January 2011; accepted in revised form 21 October 2011)

*corresponding author: christine.stone@industry.nsw.gov.au

Abstract

Small-footprint, discrete return airborne laser scanning (ALS or lidar) data is increasingly being used by forest managers to assist forest inventories. In this study, airborne lidar and plot-based data were collected from a 5 000 ha study site within Green Hills State Forest, a *Pinus radiata* D.Don plantation in southern New South Wales, Australia. A series of area-based lidar metrics were extracted and modelled against four inventory attributes (mean tree height, stem density, basal area and stand volume) obtained from 63 ground plots. For all response variables, regression tree models had the best model fit compared to Random Forest and Bayesian Model Averaging modelling techniques. The best regression tree models were based on the lidar metrics: the 5th and 95th height percentiles, minimum vegetation height, density of non-ground returns and a measure of spatial variation, the rumple index. All these metrics can be easily derived from the lidar data. The best regression tree models for each inventory attribute produced the following R^2 values: for mean tree height (m), R^2 = 0.94; stocking (trees ha^{-1}), R^2 = 0.85; basal area (m^2 ha^{-1}), R^2 = 0.81 and for stand volume, R^2 = 0.81 (m^3 ha^{-1}) while the corresponding relative RMSEs were 5.8%, 23.4%, 15.5% and 22.3%, respectively. These models were then used to produce prediction maps over a 50 m grid across the 5 000 ha study site. Results from this study support the operational inclusion of airborne lidar data within *P. radiata* resource inventory systems.

Keywords: discrete-return airborne laser scanning; lidar metrics; inventory; *Pinus radiata*; regression trees.

Introduction

Systematic assessment of *Pinus radiata* D.Don plantations is essential for predicting current and future stand volumes and implementing silvicultural regimes aimed at maximising returns. Field-based inventory methods and sampling designs are well developed and accurate, if sufficient plots capture the full range of variability of the population being surveyed. However, this approach can also be time consuming and hence expensive if required across large areas. Keeping plantation database records up to date is also becoming more challenging as the commercial forestry sector consolidates its workforce, reducing staff available for this task. The retrieval of forest-stand parameters using remotely sensed data is now viewed as a viable solution to tackling these issues. Recent reviews of this subject include those by Hyyppä et al., (2008) and van Leeuwen & Nieuwenhuis (2010).

Airborne laser scanners (ALS) belong to the type of sensors commonly referred to as lidar (light detection and ranging), and are becoming a popular method for estimating stand-level forest inventory parameters (e.g. Goerndt et al., 2011; Hudak et al., 2008; Hyyppä et al., 2008; Maltamo et al., 2006; Tesfamichael et al., 2010). Small footprint, discrete-return lidar sensors operate by rapidly emitting a laser pulse toward a target, such as a forest stand and recording the time, location, and quantity of the reflected energy. The

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

sensor is mounted on an aircraft in conjunction with a highly accurate Global Positioning System (GPS) and an inertial measurement unit (IMU), which allows correction in data processing caused by the attitude (pitch, roll, and yaw) of the aircraft (Hyyppä et al., 2008; van Leeuwen & Nieuwenhuis, 2010). Unlike passive optical sensors, lidar systems can operate independently of natural sunlight, and therefore are less restricted by weather conditions and more operationally flexible by being able to function day or night, and under cloud cover. While optical sensors provide data in raster form (with pixel values), discrete return lidar systems initially provide point data (in x, y and z coordinates) which can later be converted to raster surfaces. This point data is usually in the format of text files or Log ASCII Standard binary files with LAS extension (American Society for Photogrammetry and Remote Sensing (ASPRS), 2009). Log ASCII Standard files contain millions of points and when viewed three dimensionally appear as 'point cloud' data.

Several overseas studies have shown that ALS-based techniques are able to produce highly reliable estimates for a range of inventory parameters and with levels of precision associated with mean stand height often superior to that obtained through operational, plot-based inventory (Maltamo et al., 2006; Næsset & Økland, 2002; Stone et al., 2011; Tesfamichael et al., 2010). There are two main approaches for the estimation of stand-level inventory attributes from lidar data. One is based on the detection of individual tree crowns (Heurich, 2008; Holmgren, 2004; Lindberg et al., 2010; Maltamo et al., 2004) and the other is an area-based statistical approach (ABA), also referred to as the canopy height distribution method (Hudak et al., 2008; Næsset, 2002; Yu et al., 2011). Both approaches use canopy height models (CHM) or canopy height-corrected point clouds to derive a set of features (Yu et al., 2010).

The individual tree based approach focuses on detecting and identifying individual trees and producing tree-level information such as tree height distributions (and derived diameter distributions), so that stand-level information becomes an aggregate of the tree-level attributes (Chen et al., 2007; Peuhkurinen et al., 2011; Stone et al., 2011). This approach requires the use of an algorithm to detect individual trees with the lidar data by identifying gradient changes in canopy height or by using variable window technology (Maltamo et al., 2004; Yu et al., 2011). The expectation with the individual-tree approach is that height can be determined with no, or a consistent (negative), bias such that no site-specific calibration is needed (Hyyppä et al., 2008). In dense stands, however, estimates of mean stand height and timber volume usually contain a negative bias due to interlocking crowns and that suppressed crowns become occluded by the dominant crowns, making it difficult to isolate individual trees (Falkowski et al., 2008). If individual tree crowns can be recognised accurately, then this approach tends to outperform the area-based methods (Yu et al., 2010). The detection of individual trees, however, requires higher pulse densities ($\geq$ 2 points m^{-2}) than the area-based approach, and this can incur high acquisition costs (Packalén & Maltamo, 2007; Peuhkurinen et al., 2011). Additional costs can also arise from the requirement of higher Global Positioning System (GPS) position accuracies for the individual-tree approach compared to the area-based statistical approach. Also, since lidar data comprise three-dimensional canopy information through geo-referenced height measurements, stand height is often estimated directly from the lidar metrics, whereas stand DBH, volume and biomass tend to be derived using height driven allometrics. Therefore, individual tree based methods also require good physical correspondence with stem diameter (DBH) and volume estimation (Hyyppä et al., 2008; Peuhkurinen et al., 2011). If the number of laser pulses is acquired at < 2 points m^{-2} and canopy cover is > 75% then an area-based approach should be given priority. From an economic perspective, the area-based method is more efficient both in computation and laser data acquisitions (Hyyppä et al., 2008; Yu et al., 2010). Regardless of which approach is selected as providing the best predictive models, these relationships are then used to spatially extend model predictions of the target variables across all areas of interest where lidar data were captured.

Numerous lidar metrics have been derived and used as predictor variables in models for estimating a range of forest structure attributes including mean canopy height (e.g. Næsset & Økland, 2002), basal area and mean standing volume (e.g. Holmgren, 2004; Means et al., 2000; Rooker Jensen et al., 2006; Falkowski et al., 2010), and biomass (e.g. Lim & Treitz, 2004; Ni-Meister et al., 2010; van Aardt et al., 2006). These metrics include height percentiles, mean height, maximum height, coefficient of variation of height, kurtosis, skewness and canopy cover percentiles. While a large number of lidar metrics can be derived, in general, three broad (orthogonal) categories of lidar metrics are commonly selected by the modelling process: (1) a measure of height (e.g. the 95th percentile height of first returns); (2) a variation of height (e.g. standard deviation or coefficient of variation of first returns); and (3) a measure of vegetation density (usually the proportion of first returns greater than a lower height limit) (Frazer et al., 2011; Kane et al., 2010; Lefsky et al., 2005).

Numerous statistical modelling techniques have been used to relate the lidar metrics and other auxiliary geo-referenced variables to field data in order to construct predictive models for forest attribute estimation (e.g. Goerndt et al., 2011; Straub et al., 2010). The empirical relationships between the forest inventory attributes and lidar metrics vary between and within forest types. Differences between forest types are a function of the architecture of the tree species of interest, the local environment and the way these are presented by the 'cloud' of lidar data (e.g. acquisition specifications).

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

Even within forest stands of the same species, site quality also affects the relationships between lidar data and forest attributes (Li et al., 2008; Næsset & Økland, 2002; Rombouts et al., 2010). The robustness and accuracy of these models is dependent, in part, on the representativeness of the empirical plot-data, which serve as a validation datasets, requiring sufficient plots to capture the full range of variability present in the area of lidar coverage (Chatterjee et al., 2000; Yu et al., 2011). In reality, however, there is a trade-off between the accuracy of estimates and the intensity of the accompanying field campaign. Obtaining good representative field data is not always an easy or affordable option, especially in stands established on steep terrain or having a significant understorey component (e.g. blackberries). The effectiveness of the sampling design used to acquire empirical data from actual plots, therefore, affects the appropriateness of the modelling approach (Maltamo et al., 2011). Plot-sampling design can be optimised by using either existing stand-structure information to create pre-stratification or the properties of the lidar data as *a priori* information for selection of plot locations within stands (Hawbaker et al., 2009; Maltamo et al., 2011; van Aardt et al., 2006;).

Lidar data sets are inherently large with high degree of collinearity amongst the derived lidar predictors. The prediction models have to account for high-dimensional data sets as well as limited field calibration data (Monnet et al., 2011). Original modelling approaches were based on multiple linear regression and stepwise variable selection (Næsset, 2004; Næsset & Økland, 2002). In 2005, Næsset et al. compared the accuracy of three parametric regression techniques (ordinary least squares (OLS), seemingly unrelated regression (SUR), and partial least squares (PLS)) to retrieve plot height. They showed no increase in accuracy of parameter retrieval when the more complex parametric regression techniques were used and recommended the use of ordinary least squares regression. Since then, numerous variable extraction and selection techniques have been examined, including advanced machine-learning techniques. It is now suggested that for known linear relationships between the lidar metrics and stand attributes, e.g. mean tree height, a simple estimator like the multiple linear regressor will provide results comparable to more complex nonlinear estimators but this might not always be the case for more complex relationships (e.g. stand basal area) (Dalponte et al., 2011).

Non-parametric methods that have recently received attention include: nearest-neighbour techniques (Breidenbach et al., 2010; Falkowski, et al., 2010; Latifi et al., 2010; Lindberg et al., 2010; McInerney et al., 2010; Packalén & Maltamo, 2007); tree-based ensemble classifiers such as random forest (RF) (Breiman, 2001; Hudak et al., 2008; Falkowski et al., 2010; Stojanova et al., 2010; Yu et al., 2011); and Bayesian approaches (Junttila et al., 2008, 2010). Hudak et al. (2008) compared RF classification with OLS regression and found that OLS regression resulted in strongly biased models, which was not the case for RF classification. The bias in the OLS classification was assumed to result from artefacts in the necessary logarithmic transformations of the response variable to ensure linearity. Regression estimation in lidar forest surveys can be challenged by scale-dependent, nonlinear relationships that arise between forest inventory variables and lidar metrics (Frazer et al., 2011). Both Hudak et al. (2008) and Yu et al. (2011) concluded that nonparametric estimation methods based on machine-learning algorithms such as RF classification represent a flexible and robust alternative to traditional imputation methods. Finally, these nonparametric modelling approaches need not be restricted to the use of only lidar metrics as predictors, information extracted from spectral, topographic and terrain coverages can also be incorporated into the models, as long as the coverages are of compatible resolution and geo-registered (Breidenbach et al., 2010; Hudak, et al., 2008; Ke et al., 2010; McCoombs et al., 2003).

To our knowledge only one published study has used lidar to assess *Pinus radiata* plantation inventories (Rombouts et al., 2010). In their study, linear regression was used to predict volume of plantations across a range of acquisition ‘campaigns’ for South Australian forests aged 7 – 11 years. Volume was found to be related primarily to quadratic mean height within this age class grouping. While Rombout’s study represents an important first step, *P. radiata* plantations have rotations of up to 35 years and can be subjected to several thinning regimes.

Our study of *Pinus radiata* in New South Wales uses lidar-derived metrics to predict four key inventory attributes: mean stand height; basal area (m^2 ha^{-1}), stnnd volume (m^3 ha^{-1}); and stocking (stems ha^{-1}) at the plot level in a plantation with a broader range of ages and stem densities than that studied by Rombouts et al. (2010). We compared the effectiveness of three area-based regression techniques (regression trees, RF and Bayesian Model Averaging) to find the model with the best predictive capability. Grid-based predictions are provided for the entire study area to illustrate the practical usability of the tested methods.

Materials and Methods

The 5000 ha study area is located within Green Hills State Forest (SF) (35.5°S, 148.0°E), near Batlow on the southern slopes of New South Wales (NSW), Australia and managed by Forests NSW (Figure 1). Green Hills SF is a large, commercially active *Pinus radiata* plantation with 835 compartments and a net planted area of 20400 ha. It is situated on mostly undulating topography, with a mean elevation of 750 m and annual rainfall of approximately 1200 mm.

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

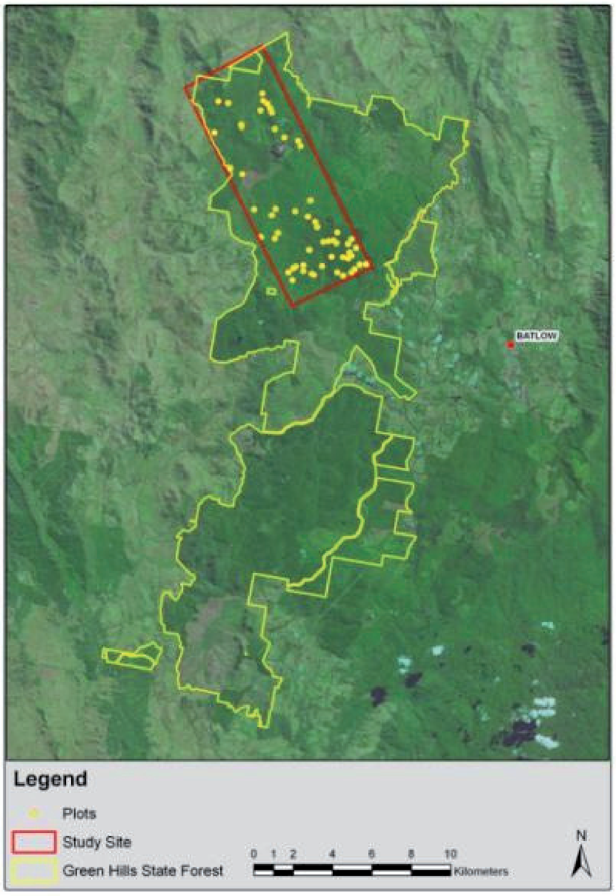

FIGURE 1: The location of the 5000 ha study area and the 63 plots within the Green Hills State Forest *Pinus radiata* planation in the Hume Region of Forests NSW. (Overlain on SPOT5 imagery). The location of the town of Batlow (-35.522, 148.144 decimal degrees (GDA94)) is also shown.

The sampling design was that of stratified random sampling with strata defined using three age classes (10 – 20 years; 21 – 30 years; > 30 years), three slope levels (0 < 10 degrees; 10 – 20 degrees; > 20 degrees, and three thinning regimes (unthinned; first thinning; second thinning) (Table 1). Within Forests NSW, compartments are planted to approximately 1000 stems ha^{-1}, generally thinned between the ages 13 to 17 years old down to 450-500 stems ha^{-1} and then thinned again after 23 years down to 200 to 250 stems ha^{-1}. Most compartments are harvested before 35 years of age. Individually, the full range of age classes, slopes and thinning treatments are represented in the 5000 ha study area. However, of the 27 possible strata, only sixteen were represented in the study area (Table 1). In September 2008, four circular ground plots were randomly located and established in each stratum to a total of 63 plots (one plot not measured; Figure 1). Each ground plot had approximately 15 trees per plot and with radii ranging from 7 m – 20 m (Table 1). The centre location of each ground plot was accurately surveyed using a laser theodolite (Leica 2 second T1100 total station) and a Differential Global Positioning System (dGPS; Trimble Navigation Ltd., Fortitude Valley, Queensland), with the differential processing done in real time. Two reference pegs were placed on a nearby road or track with reasonable sky access for satellite coverage. These reference pegs were spatially defined to less than 50 mm. The surveyor then traversed to the peg located at the plot centre.

Empirical data for each response variable were obtained as follows: every tree in each ground plot was labelled and diameter at breast height over bark (DBHOB at 1.3 m) and tree height (m) measured. Tree height was measured twice using an ultrasonic hypsometer (Vertex III, Haglöf, Sweden). A summary of the ground-based tree measurements is presented in Table 1. Plot volume (m^3 ha^{-1}) and basal area (m^2 ha^{-1}) were calculated using in-house algorithms (H. Bi, NSW Department of Primary Industries, pers. comm.).

Lidar imagery acquisition and processing

Small-footprint discrete return lidar data was acquired using a Lite Mapper LMS-Q5600 ALS system (Riegl, Austria) mounted in a fixed-wing aircraft and supplied through Digital Mapping Australia Pty Ltd (Perth, Australia). The lidar mission was flown in July 2008 over the 5000 ha study area to coincide with winter, as this is the period when the blackberry canopies, the key understorey weed species in the region, are most transparent. In some areas, blackberry infestations can be extremely dense making the task of mapping the terrain surface with lidar more difficult. No snow was present.

The near infra-red (NIR) lidar system was configured for a pulse rate of 88000 pulses second^{-1}, mean footprint size of 60 cm diameter, maximum scan angle of 15^{o} (off vertical), mean swath width 500 m and a mean point density of 2 pulses m^{-2} (based on the non-overlap portion of the swath). The lidar data was received in LAS file format with the first and last return for each laser pulse recorded but not tagged. The return signal intensity (echo strength) values were also recorded. The laser scanning (lidar) points were processed, geo-referenced and classified by the service provider into ground and non-ground categories using their proprietary method and TerraScan software (TerraSolid, Finland) integrated within a MicroStation CAD environment (Bentley Systems, USA). Processed lidar point data was supplied on an external drive with each file representing a 1 km x 1 km area (tile). Coordinates were expressed in Map Grid Australia (MGA) zone 55 projection and Geodatic Datum of Australia 1994 (GDA94) datum.

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

TABLE 1: Mean class data of inventory attributes obtained from the 63 plots in Green Hills *Pinus radiata* plantation. (Numbers in brackets are minimum and maximum values). Four plots per stratum class[1].

Strata Class				Mean Tree Data					
Class no.	Slope[2]	Age[3]	Thinning regime[4]	Age	Number of trees ha^{-1}	Diameter at 1.3 m	height (m)	Basal area (m^2 ha^{-1})	Volume (m^3 ha^{-1})
1	0 < 10	≥ 10 to ≤ 20	UT	12.24 (10-14)	1021 (902-1150)	17.2 (15.4-18.2)	15.0 (12.2-16.8)	26.6 (21.0-31.8)	161.3 (106.5-193.4)
2	0 < 10	≥ 10 to ≤ 20	T1	16.25 (16-17)	626 (382-837)	21.5 (19.0-23.7)	20.2 (17.3-22.3)	23.6 (18.1-28.4)	176.0 (148.2-228.6)
3	0 < 10	≥ 21 to ≤ 30	UT	27.0 (26-28)	817 (791-896)	26.8 (25.5-28.8)	28.8 (27.6-31.5)	51.5 (43.8-54.9)	580.8 (452.7-760.3)
4	0 < 10	≥ 21 to ≤ 30	T1	24.0 (24-24)	403 (283-697)	33.1 (28.6-39.2)	28.7 (26.9-32.5)	34.2 (25.1-47.2)	339.8 (245.1-445.4)
5	0 < 10	≥ 21 to ≤ 30	T2	26.75 (22-29)	250 (121-478)	39.5 (33.4-46.2)	28.7 (24.2-30.7)	28.4 (20.6-43.1)	272.3 (214.2-364.2)
6	0 < 10	> 30	T2	30.0 (30-30)	265 (187-305)	41.9 (40.3-43.7)	31.6 (30.3-33.6)	37.3 (29.4-42.8)	402.3 (322.1-450.7)
7	≥ 10 to ≤ 20	≥ 10 to ≤ 20	UT	12.5 (11-13)	1124 (844-1429)	17.9 (15.8-19.7)	16.0 (13.1-18.4)	30.5 (22.1-36.6)	191.3 (118.3-252.6)
8	≥ 10 to ≤ 20	≥ 10 to ≤ 20	T1	16.5 (16-17)	775 (563-1104)	21.6 (20.7-23.1)	20.9 (19.1-23.0)	29.8 (23.3-39.8)	232.3 (177.3-336.5)
9	≥ 10 to ≤ 20	≥ 21 to ≤ 30	UT	23.5 (23-25)	994 (796-1299)	25.6 (20.1-29.5)	23.1 (20.2-26.3)	52.0 43.2-59.0)	433.1 (322.0-478.0)
10	≥ 10 to ≤ 20	≥ 21 to ≤ 30	T1	24.25 (24-25)	280 (211-395)	36.3 (31.8-41.9)	26.1 (24.0-28.3)	29.0 (25.5-33.3)	270.0 (240.7-306.7)
11	≥ 10 to ≤ 20	≥ 21 to ≤ 30	T2	26.5 (22-29)	245 (146-310)	36.6 (45.3-45.9)	25.7 (18.9-30.1)	25.4 (18.5-29.2)	234.0 (129.7-299.0)
12	≥ 10 to ≤ 20	> 30	T2	30.0 (30-30)	283 (265-305)	42.4 (41.1-43.9)	31.2 (30.0-31.8)	40.4 (35.6-43.6)	423.1 (374.3-462.9)
13	> 20	≥ 10 to ≤ 20	UT	13.25 (13-14)	983 (508-1412)	19.8 (16.9-24.5)	17.6 (14.5-20.8)	30.3 (26.5-35.5)	202.6 (147.0-231.8)
14	> 20	≥ 21 to ≤ 30	T1	21.0 (21-21)	819 (599-1053)	27.1 (24.4-28.7)	19.9 (18.1-23.1)	49.0 (38.2-65.4)	360.7 (282.7-455.6)
15	> 20	≥ 21 to ≤ 30	T2	25.0 (25-25)	139 (126-150)	44.2 (42.1-46.4)	27.5 (25.9-28.8)	21.9 (20.3-23.9)	207.6 (187.7- 228.4)
16	> 20	≥ 21 to ≤ 30	UT	22.6 (21-26)	611 (478-844)	30.2 (27.5-33.5)	22.2 (19.2-24.3)	45.8 (29.6-61.7)	377.5 (209.5-545.4)

[1] Only 3 plots were measured in Class 16.
[2] Slope levels: 0 < 10 degrees; 10 – 20 degrees; > 20 degrees.
[3] Age classes: 10 – 20 years; 21 – 30 years; > 30 years.
[4] Thinning regime: unthinned (UT); first thinning (T1); second thinning (T2).

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

A Digital Terrain Model (DTM) at both 0.5 m and 1.0 m pixel resolution was constructed from ground point data using a standard linear triangulation surface modelling technique in Environment for Visualizing Images (ENVI) software (Research Systems Incorporated, USA). The DTM represents (in theory) the bare terrain elevation above sea level. To get the vegetation heights for each lidar point, the DTM was substracted from the point elevation value. On the occasion where vegetation heights were negative, these values were set to zero. Sampling density error was inherent in the final DTM error but was shown to be, on average, very low (< 0.6 m) at tree bases for a sub-set of 145 trees selected within the study site (Johnson, 2010). The DTM error would have been lower in open ground. Johnson (2010) used this lidar dataset to compare four surface modelling techniques (Triangular Irregular Network (TIN), Inverse Distance Weighting (IDW), Kriging, and Spline) used to generate DTMs and concluded that each of these surface modelling techniques produced similar results at both 0.5 m and 1.0 m grid cell size resolution.

Lidar metrics

Lidar metrics were extracted for each of the 63 plots using ground and non-ground LAS files. Using the non-ground height values, we calculated the mean, median, mode, maximum, minimum, 5th and 95th percentile, variance, standard deviation, coefficient of variation, range, relative range (i.e. range divided by the mean), range of the 95 percentile (height of the 95th percentile minus the minimum height), quadratic mean canopy height (square root of the sum of the squared heights divided by the number of heights), skewness and kurtosis. For a series of height categories, we calculated the percentage of non-ground returns in each category: 0 – 3 m; 3 – 10 m; 10 – 20 m; and 20 m and greater. Regardless of heights, we calculated the density of non-ground and ground returns. A rumple index was calculated and standardised by the mean height of the plot. The rumple index is a measure of canopy structure heterogeneity and is the ratio of the surface area of the surface created by the non-ground heights divided by the area of the flat surface (Kane et al., 2010). The rumple index was calculated on both a 0.5 m and 1 m grid cell size using surface area functions within the GRASS GIS package (GRASS Development Team, 2010). The derived 1 m DTM was also used to calculate the mean slope and aspect for each plot (auxiliary predictor variables). All data extraction and analyses were conducted using the open-sourced R-statistical package v.2.11.1 (R-Development Core Team, 2007), in conjunction with freely available libraries written for the R package and the GIS package GRASS (GRASS Development Team, 2010). The open-source package spgrass6 was used to provide the interface between R and GRASS and was accessed from the cran.r-project website (Bivand, 2010).

Data were extracted for a range of subsets of the non-ground return data files and each dataset was modelled separately. Analysis was conducted on all the points within the LAS files, hereafter termed "raw" data. During the data collection, first and last returns were not identified. To account for this we created a further two sets of pseudo first return data. These datasets were created by generating either a 0.5 m or a 1 m grid over the lidar returns and taking a maximum within each grid cell, hereafter the "0.5 m" or the "1 m" data respectively. Within each of these three sets of data (0.5 m, 1 m and raw), we calculated the aforementioned variables based on all data points (hereafter "all") and on the canopy data, classified as all returns greater than 2 m (hereafter "canopy"). Finally, for all combinations we calculated the variables based on a fixed radius of 30 m around each plot centre (hereafter "30 m") and on the variable radius measured on the ground (hereafter "variable radius"). In total there were 12 sets of variables calculated.

Statistical analysis

Five stand response variables were derived from the plot measurements, i.e. maximum tree height (m), mean tree height (m), stocking (stems ha^{-1}), basal area (m^2 ha^{-1}) and volume (m^3 ha^{-1}) (Table 1). The predictor variables, being the derived lidar metrics, were then modelled against the response variables. Models were prepared for both the raw values and the log-transformed values of these variables. The log-transformed, predicted values were bias-corrected using a correction factor of 0.5 times the mean squared error before back transformation (Goerndt et al., 2010). A large number of predictor variables was available from the lidar data, however, many of these variables were highly correlated. We calculated a Spearman's correlation matrix and used the output from this to reduce the number of predictor variables and also to remove the potential for multi-collinearity in the models (Chatterjee et al., 2000). When two or more variables were found to have a correlation greater than 0.7, we selected one variable and removed all others. Preference for response-variable retention was given to proximal rather than distal variables (after Wintle et al., 2005) and to those variables that have been reported to be useful in similar studies. The resulting set of non-correlated variables were the rumple index based on the 0.5 m grid cell, mean slope, height of the 5th and 95th percentiles, minimum height, skewness and the density of ground and non-ground returns (a measure of canopy openness).

Three modelling approaches were used in the analysis: (1) regression trees; (2) RF; and (3) Bayesian Model Averaging (BMA).

1. Regression trees are a simple, but powerful, modelling approach to the analysis of complex environmental data that can allow for nonlinear relationships (De'ath & Fabricius, 2000). This

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

approach seeks to explain variation in data by repeatedly splitting data into more homogeneous groups using the predictor variables. After the initial tree was built, we used a k-fold cross-validation analysis to optimise and prune the tree (De'ath & Fabricius, 2000). All regression tree analysis used the R package called "*tree*" (Ripley, 2010).

2. The RF approach extends the regression tree approach by "growing" multiple (500) "trees" based on subsets of the data and getting the majority vote for the outcome (sometimes referred to as an ensemble method) (Breiman, 2001). Random forest analyses used the R package *randomForest* (Liaw & Wiener, 2002).

3. Bayesian Model Averaging is a linear regression modelling approach which builds on the commonly applied generalised linear modelling approach (e.g. Wintle et al., 2003). The BMA method builds linear models based on all combinations of the predictor variables and a best set of the models are chosen based on the Bayesian Information Criterion (BIC). Parameter estimates of the BMA model are averaged based on the weighting derived from the BIC (Wintle et al., 2003). Models were removed from consideration when a simpler version of a model (i.e. with a subset of the predictor variables) had a better fit, i.e. lower BIC. All BMA analyses were conducted using the R package called "*BMA*" (Raftery et al., 2010).

Comparisons between modelling techniques were based on the coefficient of determination (R^2). The R^2 value ranges from 0 to 1 and describes the proportion of variability in the dataset accounted for in the statistical model (Sokal & Rohlf, 1995). A high value of R^2 means that there is good agreement between the observed and modelled values. Differences between observed and estimated plot-level means were compared using the absolute root mean square error (RMSE) and relative RMSE, which is the RMSE expressed as a percentage of the observed plot means. For each response variable, we considered all of the derived models (regardless of the statistical approach) with an R^2 within 0.05 (or 5%) of the best model. We refer to this as the "best set" of models. We made predictions from the best set of models for each response variable by selecting only the model with the highest R^2 for each modelling approach represented in the best set. The best set of models was then applied over the whole study area in the form of a prediction map. This procedure was facilitated by the R/GRASS interface software package spgrass6 (Bivand, 2010). The predictions were made over a 50 m grid across the study area. A 50 m grid was selected as it was similar to the 30 m circular radius plots used in the analysis (2 500 m^2 for the grid versus 2827 m^2 for the 30 m plot).

Results

For each response variable, regression trees model had the best fit compared to the other two statistical approaches. Values of R^2 for the best regression tree models ranged from 0.95 – 0.93 for the two height response variables, i.e. maximum height and mean tree height. Values of R^2 for the best regression tree models ranged from 0.85 – 0.81 for the other three variables, i.e. stocking, derived basal area and derived stand volume (Table 2). Similarly, relative RMSE values from the best regression tree models were the lowest for the two height variables, 4.8% for maximum height and 5.8% for mean tree height (Table 2). Bayesian Model Averaging models had similar R^2 values to the corresponding regression trees for the two height variables but lower values for the other three variables (i.e. basal area, stand volume and stocking, Table 2). The RMSE values were also correspondingly lower for all the best BMA models compared to the regression tree models. Model fit for any of the RF models tested for all response variables was significantly lower than for either the regression tree or BMA models, and are not considered further. When the best model from each lidar data extraction technique was selected, all but one model was based on the 1.0 m filter, with the other being from the 0.5 m filter for stocking. Of the best models, all but two models were based on the lidar data taken from the 30 m plot radius (Table 2).

An examination of the most commonly selected and influential lidar metrics revealed that the best set of models for mean tree height from both the regression tree and BMA techniques all included the 95th percentile height (h_{95}) metric. The best regression tree model was almost entirely derived from the height of the 95th percentile, with minimum height included on a lower split. Minimum height and the density of non-ground returns featured in 37.5 and 25% of models respectively. Ground-return density, skewness and 5th percentile height (h_5) all appeared in less than 10% of models. The rumple index and slope did not occur in any of models. The best models using the regression tree technique were based on the log-transformed values of mean tree height.

Only three models fell in the best set for the response variable of stocking and these were all regression tree models. Density of ground returns, minimum height and h_{95} occurred in all three models, with slope, h_5 and density of non-ground returns each occurring in two of the three models. While height of the 95th percentile and minimum vegetation height had the strongest influence in the best regression tree model, the models also included slope, height of the 5th percentile, and the density of non-ground and ground returns.

The response variable basal area was log transformed in the three best models, all of which were regression tree models. Slope, h_5, h_{95}, and the density of ground

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

TABLE 2: R^2 values (coefficient of determination) and RMSE (root mean square error) values of the best model for each response variable obtained from the regression tree (Reg Tree) and Bayesian Model Averaging (BMA) modelling approaches.

Modelling method	Response Variable	Lidar data point heights	Grid size filter (m)	Lidar data radius	Model R^2	RMSE	Observed value	Relative RMSE[1]
Reg. tree	Max. tree height (m)	all	1.0	30 m	0.95	1.31	27.3	4.8
BMA	Max. tree height (m)	all	1.0	30 m	0.94	1.49	27.3	5.5
Reg. tree	Mean tree height (log) (m)	canopy	1.0	30 m	0.94	1.40	24.0	5.8
BMA	Mean tree height (log) (m)	canopy	1.0	30 m	0.93	1.59	24.0	6.6
Reg. tree	Basal area (log) (m^2 ha^{-1})	all	1.0	30 m	0.81	5.36	34.5	15.5
BMA	Basal area (log) (m^2 ha^{-1})	all	1.0	30 m	0.55	7.67	34.5	22.2
Reg. tree	Stocking (trees ha^{-1})	all	0.5	30 m	0.85	140.8	602.5	23.4
BMA	Stocking (trees ha^{-1})	all	1.0	plot	0.71	230.8	602.5	38.3
Reg. tree	Volume (log) (m^3 ha^{-1})	all	1.0	plot	0.81	67.6	302.8	22.3
BMA	Volume (log) (m^3 ha^{-1})	all	1.0	30 m	0.71	330.0	302.8	108.9

[1] Relative RMSE = (RMSE/Observed value) x 100.

and non-ground returns occurred in all three models, with the rumple index occurring in one of the models. In the best regression-tree model, basal area was predicted from the rumple index, slope, height of the 5th and 95th percentiles, minimum height and the density of non-ground and ground returns with height of the 5th and 95th percentiles and slope having the greatest influence on the model. Eleven models formed the best set of models predicting the response variable of stand volume. All eleven of these models were regression-tree models and the top nine all used the log transformed values for stand volume.

Overall, the best regression tree models for predicting the response variable of height had simpler structures with fewer lidar metrics than the regression-tree models predicting stocking, basal area and stand volume. Slope and h_{95} appeared in all models and the density of non-ground returns appeared in seven models. The remaining explanatory variables appeared in four or five of the best set of models. The best model had the following explanatory variables: the 5th and 95th percentile, slope, skewness and density of non-ground returns. As with basal area, h_5 and h_{95} had the greatest influence on the estimate of stand volume.

Predictions

The best regression tree model and the best BMA model selected for each of the five response variables were used to predict these variables across the study region (Table 3). These predicted values were then compared with empirical data from the 63 plots. The best regression tree models closely predicted observed mean and standard deviations for all response variables, except for the stocking, which predicted a mean of 74 trees ha^{-1} higher than the measured values (Table 3). The best BMA models were more variable in

TABLE 3: Comparisons of predicted values (mean & standard deviation) for each response variable from the best regression tree (Reg. tree) and BMA models and measured values (mean & standard deviation) for the 63 plots.

Response Variable	Measured mean (m)	Measured SD	Reg. tree mean (m)	Reg. tree SD	BMA mean (m)	BMA SD
Maximum tree height	27.3	6.0	27.1	5.8	31.1	5.4
Mean tree height	24.0	5.7	24.2	5.5	24.4	5.7
Stocking	602.5	367.3	676.1	370.6	680.9	295.5
Basal area	34.5	11.6	34.9	11.2	30.9	3.8
Stand volume	302.8	132.3	305.3	109.2	292.1	102.6

SD = standard deviation

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

performance with mean values for maximum height and stocking being higher than the measured values and basal area and stand volume being underestimated. Similarly, the predicted standard deviations from the BMA models were lower for the stocking, basal area and to a lesser extent stand volume. Prediction maps over a 50 m grid across the 5 000 ha study site were produced by applying the best regression tree models for mean stand height (Figure 2) and mean stand volume (Figure 3).

Overall, with the data obtained in this study, regression tree models consistently outperformed BMA and RF. Performance of the models was highest for those stand variables that were linearly related to the lidar metrics, e.g. mean height.

Discussion

Plantation managers are constantly seeking ways to reduce field inventory costs but also maintain the timely assessment of their stands for evaluation and planning. Lidar technology is proving to be a viable option to fulfil these goals. We have demonstrated that the area-based extraction of lidar metrics can be modelled to accurately predict stand height, basal area and volume across a broad range of ages and stem densities in a *P. radiata* plantation using a regression tree approach. These predictive models can be spatially extrapolated, producing high resolution maps that visually identify variation between and within compartments across the study area.

Stand height was most strongly influenced by the height of the 95th percentile (h_{95}). Tesfamichael et al. (2010) investigating the impact of discrete-return lidar point density on estimations of mean and dominant plot-level tree height in *Eucalyptus grandis* Hill ex Maiden plantations, reported that all their models comprised of higher order percentiles, with the 95th percentile being the most prevalent. Height of the 95th percentile is a better predictor than maximum height as it removes the influence of outliers making it a more reliable height estimate (Næsset & Økland, 2002; Kane et al., 2010). Heights of the 5th and 95th percentile also had a strong influence on the other models predicting basal area, volume and stocking. Slope also appears in these models and for trees of similar size, there are greater basal area and volumes on the steeper slopes. This is probably a reflection of delays in thinning schedules on steeper slopes within the study area. Alternatively it could be related to the variability in horizontal distance between tree rows on flat versus steep ground.

Good relationships were also found for basal area and stand volume when using a regression tree approach although these two variables have a more complex relationship with the lidar height metrics (Bi et al., 2010). The models for stocking had the lowest coefficient of determination values (R^2) and hence would be the least reliable. This has been reported for lidar studies in other forest systems (e.g. Magnussen et al., 2009). For example, while Peuhkurinen et al. (2011) reported relative RMSE values of 2.3% for mean height, 13.5% for stand volume and 15.0% for basal area in a *Pinus sylvestris* (L.) stand using an area-based methodology, the RMSE% for stocking, in

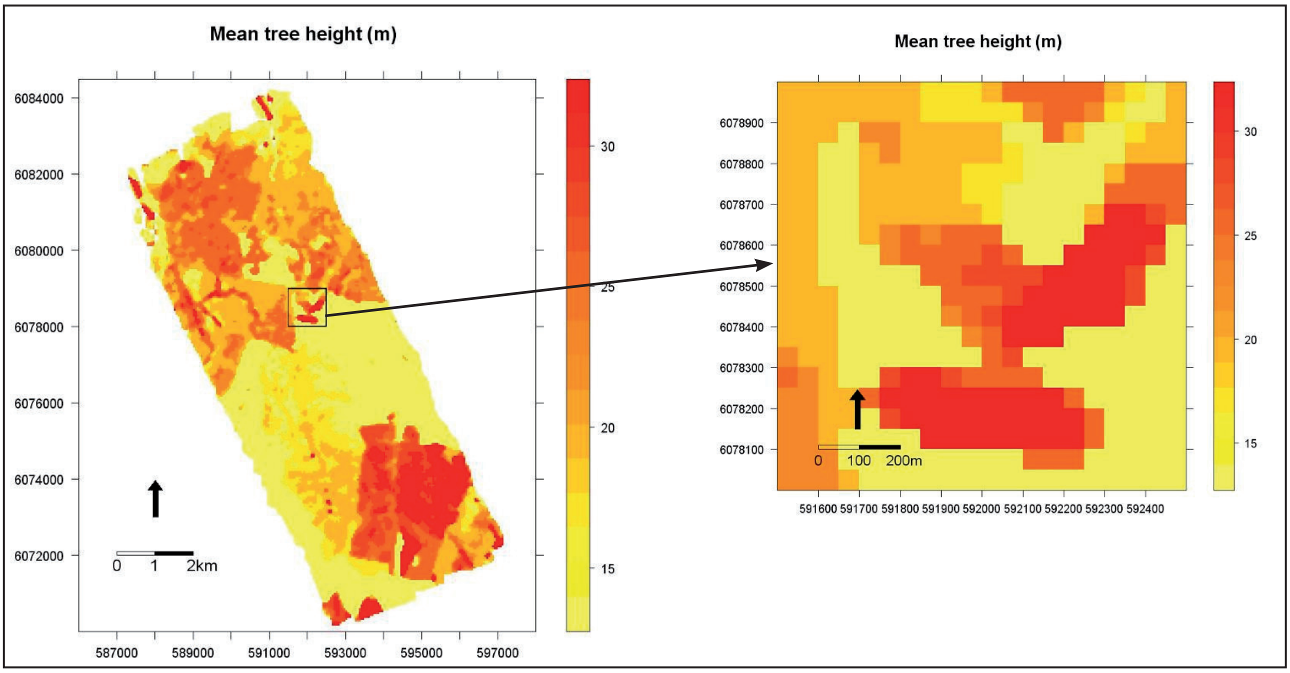

FIGURE 2: Prediction map for mean tree height over a 50 m grid covering the 5 000 ha study site within the Green Hills State Forest *Pinus radiata* plantation.

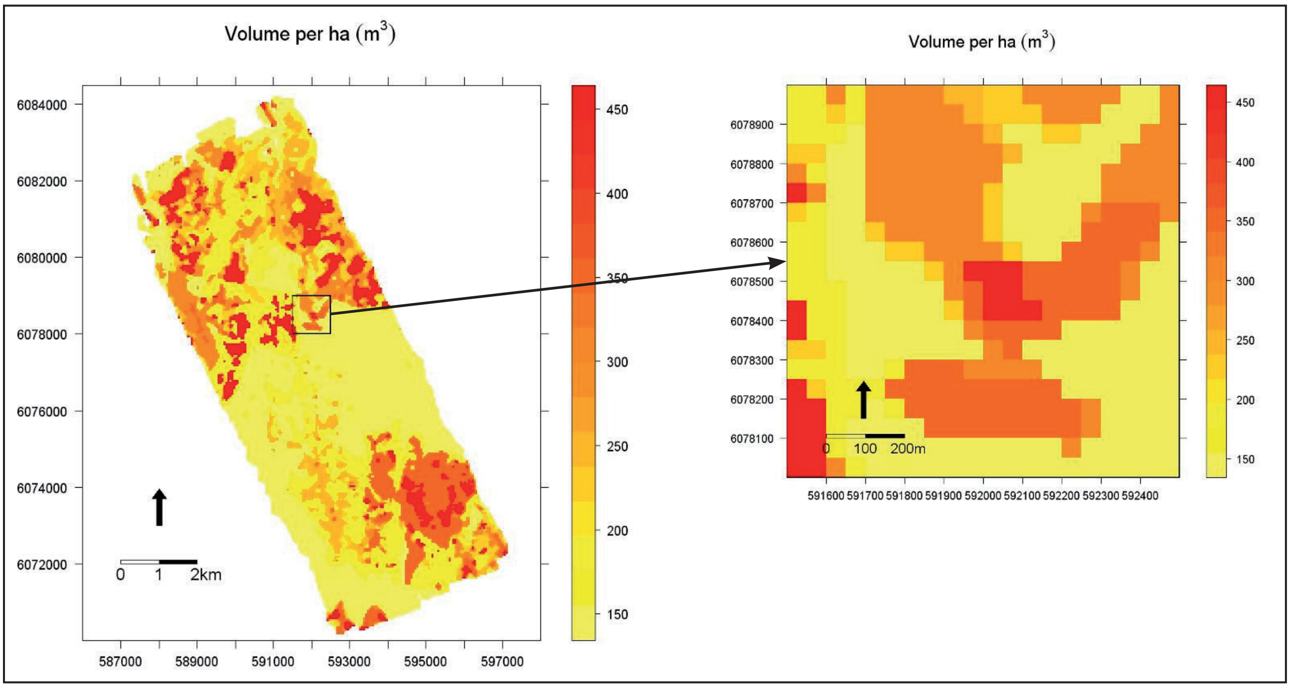

FIGURE 3: Prediction map for mean stand volume over a 50 m grid covering the 5 000 ha study site within the Green Hills State Forest *Pinus radiata* plantation.

contrast, increased to 32.5%. Detection of suppressed trees or individual tree crowns in dense canopies from a height model based on lidar data is difficult, although Maltamo et al. (2004) demonstrated that it is possible to predict stem density by using theoretical distribution functions.

For our datasets, the regression tree models consistently outperformed BMA and RF models. Regression tree models were consistently the modelling technique with the highest variance explained for all five of the response variables tested. Other studies have had similar success with classification and regression trees (e.g. Coops et al., 2006). Linear regression using BMA provided models with strong support for the height metrics, but did not perform well for the derived metrics. The strong relationship occurred due to the linear relationships between stand height and the lidar metrics. Relationships between the lidar metrics and stocking, basal area or volume are expected to be non-linear (e.g. Bi et al., 2010). While Rombouts et al. (2010) found strong linear relationships between lidar metrics and volume, their study only considered plantations aged 7 to 11 years and within a single thinning regime. Across a greater range of ages and thinning treatments these simple (linear) relationships are unlikely to hold.

More derived regression tree approaches such as RF models (Breiman, 2001; Yu, et al., 2011), Adaboost (Freund & Schapire, 1996) and boosted regression tree models (Elith et al., 2008) would be expected to further improve the predictive ability of regression trees. Random forests is a tree-based ensemble classifier that has been shown to be well suited to the high dimensionality associated with remotely sensed datasets (e.g. Stojanova, et al., 2010). However, in our study, the fit of RF models was lower than those of regression trees models. Although not presented here, similar results were also obtained when data were analysed using boosted regression trees. The poor performance of these advanced techniques is probably related to the relatively small number (n = 63) of sample plots that were used relative to the variation in the data. Random forest models use a sub-sampling procedure to build each of the 500 “trees” within the forest (Breiman, 2001). The assessment plots had relatively low replication of various combinations of height, age and thinning treatment (i.e. 4). Sub-sampling has the potential to entirely exclude some combinations resulting in poor predictions. This may be overcome by a larger number (>100) of plots. The initial investment in well replicated reference field data for lidar model development is strongly advocated as it enables modellers to take advantage of state-of-the-art machine learning techniques for developing reliable models with high precision and accuracy that can be applied broadly over a plantation estate (Stojanova et al., 2010). These models are not static and can be improved through a routine validation process based on an optimised (plot or single tree) sampling design within an inventory program (Hawbaker et al., 2009; Maltamo et al., 2011; Parker & Evans, 2009).

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

It can be difficult to compare studies due to differences in forest type; acquisition specifications and modelling techniques. Overall, however, our results based on a large, intensively managed *Pinus radiata* plantation, concur with the findings from other studies, although derived from a different forest type; acquisition specifications and modelling techniques (e.g. Goerndt et al., 2010; Peuhkurinen et al., 2011; Yu et al., 2011). In particular, pioneering work by Magnussen & Boudewyn (1998) reported a strong correlation (R^2 = 0.8, SD = 2.2 m) between lidar metrics and field estimates for mean stand height in a Douglas-fir (*Pseudotsuga menziesii* var. *menziesii* [Mirb.] Franco) stand whereas our best models for maximum tree height and mean tree height produced R^2 values of 0.95 and 0.94 and RMSE values of 1.31 m and 1.40 m respectively. Computing capacity, data processing and modelling methodologies have all improved over the past ten years, resulting in R^2 values > 0.9 now routinely reported (e.g. Næsset & Økland, 2002; Næsset, 2004; Goerndt et al., 2010).

Further Research

Both area-based and individual-tree based methods can perform poorly when estimating stem density. Prediction accuracy for stocking for some stands (such as homogenously thinned stands) can be improved through the detection of individual tree crowns (e.g. Goerndt, et al., 2010; Yu, et al., 2010). However, success in the detection of individual tree crowns in dense (unthinned) stands is dependent on the pulse density of lidar data as it directly influences the performance of the tree crown detection process. A higher density of lidar pulses ensures improved crown detection but this comes at a higher cost of acquisition. Further research is required to improve the accuracy and efficiency of individual tree counts using lidar data. This will involve identifying *Pinus radiata* stand characteristics best suited for either the application of the area-based or individual tree approaches, including the optimisation of maxima selection rules based on multi-scale focal statistics. Secondly, stratified sampling designs based on silvicultural history have been commonly applied but lidar data could be used to optimise sampling designs in stands requiring additional information related to stem quality and product assortment. Finally, model validation with ground-based data can be significantly hampered by positional errors associated with both the field measurements and the remotely sensed data (Næsset & Økland, 2002). Work into minimising these co-registration errors is required. For example, a network of permanent reference marks, built to surveying specifications, could be established in larger plantations.

Conclusions

Increasing costs of field surveys, coupled with ever-increasing demands for collection of both timely and more detailed information, are directing forest managers to consider alternatives approaches to forest and plantation assessment. Our study supports the application of lidar (small-footprint airborne laser scanning) as a method for estimating several key inventory attributes in *Pinus radiata* using an area-based modelling approach. These models can be represented spatially on a grid basis across plantations to provide a snap-shot of the standing plantation resource as well as potentially useful inputs into future yield modelling. At present, the collection of lidar is viewed as relatively expensive but costs per product decrease if multiple products can be derived from the same lidar data, both within companies and through collaborative lidar acquisition missions.

Acknowledgements

The authors thank the following people who assisted us with this study: Duncan Watt and Bob Cooper, (Forests NSW); and Catherine Carney and Matt Nagel (Industry & Investment NSW). The manuscript was improved significantly by comments from two anonymous reviewers and the Editors of the *New Zealand Journal of Forestry Science* during the review process. This study was funded by Forest & Wood Products Australia, Forests NSW and Science & Research, Industry & Investment NSW. Results presented here were obtained as part of the FWPA project PNC058-0809 and provide a more complete analysis than that presented in the FWPA Project Final Report.

References

American Society for Photogrammetry and Remote Sensing. (APRS). (2009). *Common lidar data exchange format - .LAS industry initiative.* Retrieved 23 May 2011 from http://www.asprs.org/a/society/committees/lidar/lidar_format.html.

Bi, H., Long, Y., Turner, J., Lei, Y., Snowdon, P., Li, Y., Harper, R., Zerihun, A., & Ximenes, F. (2010). Additive prediction of aboveground biomass for *Pinus radiata* (D. Don) plantations. *Forest Ecology Management, 259,* 2301-2314.

Bivand, R. (2010). *spgrass6: Interface between GRASS 6 and R. R package version 0.6-16* [Type of Medium]. Retrieved October 2010, from http://cran.r-project.org/web/packages/spgrass6/index.html).

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment, 114,* 911-924.

Breiman, L. (2001). Random Forests. *Machine Learning, 45,* 5-32.

Chatterjee, S., Hadi, A. S., & Price, B. (2000). *Regression Analysis by Example*. New York, USA: John Wiley and Sons.

Chen, Q., Gong, P., Baldocchi, D., & Tian, Y. Q. (2007). Estimating basal area and stem volume for individual trees from lidar data. *Photogrammetric Engineering & Remote Sensing, 73,* 1355-1365.

Coops, N. C., Wulder, M. A., & White, J. C. (2006). Integrating remotely sensed and ancillary data sources to characterize a mountain pine beetle infestation. *Remote Sensing of Environment, 105,* 83-97.

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

De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology, 81,* 3178-3192.

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

Falkowski, M. J., Smith, A. M. S., Gesssler, P. E., Hudak, A. T., Vierling, L. A., & Evans, J. S. (2008). The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data. *Canadian Journal of Remote Sensing, 34,* S338-S350.

Falkowski, M. J., Hudak, A. T., Crookston, N. L., Gessler, P. E., Uebler, E. H., & Smith, A. M. S. (2010). Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbour imputation approach incorporating LiDAR data. *Canadian Journal of Forest Research, 40*, 184-199.

Frazer, G. W., Magnussen, S., Wulder, M. A., & Niemann, K. O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainity of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment, 115*, 636-649.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. (1996). *Machine Learning*: Proceedings of the Thirteenth International Conference (1996) Bari Italy (pp. 148-156). San Francisco, CA, USA: Morgan Kaufman.

Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2010). Relating forest attributes with area- and tree-based light detection and ranging metrics for western Oregon. *Western Journal of Applied Forestry, 25*, 105-111.

Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2011). A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. *Canadian Journal of Forest Research, 41,* 1189-1201.

GRASS Development Team (2010). *Geographic Resources Analysis Support System (GRASS) Programmer's Manual*. Open Source Geospatial Foundation. Retrieved October 2010 from http://grass.osgeo.org/programming6/.

Hawbaker, T. J., Keuler, N. S., Lesak, A. A., Gobakken, T., Contrucci, K., & Radeloff, V. C. (2009). Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *Journal of Geophysical Research 114*, G00E04. doi: 10.1029/2008JG000870.

Heurich, M. (2008). Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management, 255,* 2416-2433.

Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research, 19*, 543-553.

Hudak, A. T., Crookston, N. C., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbour imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment, 112,* 2232-2245.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small foot print airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing, 29,* 1339-1366.

Johnson, S. (2010). *Assessment of surface modelling techniques for the estimation of tree heights using airborne lidar*. Undergraduate Thesis (GMAT 4015). University of New South Wales, Sydney.

Junttila, V., Kauranne, T., & Leppänen, V. (2010). Estimation of forest stand parameters from airborne laser scanning using calibrated plot databases. *Forest Science, 56,* 257-270.

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

Junttila, V., Maltamo, M., & Kauranne, T. (2008). Sparse Bayesian estimation of forest stand characteristics from airborne laser scanning. *Forest Science, 54,* 543-552.

Kane, V. R., McGaughey, R. J., Bakker, J. D., Gersonde, R. F., Lutz, J. A., & Franklin, J. F. (2010). Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research, 40,* 761-773.

Ke, Y., Quackenbush, L. J., & Im, J. (2010). Synergistic use of Quickbird multispectral imagery and LiDAR data for object-based forest species classification. *Remote Sensing of Environment, 114,* 1141-1154.

Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-dervied predictors. *Forestry, 83,* 395-406.

Lefsky, M. A., Hudak, A. T., Cohen, W. B., & Acker, S. A. (2005). Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment, 95,* 517-531.

Li, Y. Andersen, H-E., & McGaughey, R. (2008). A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *Western Journal of Applied Forestry, 23,* 223-231.

Lim, K., & Treitz, P. (2004) Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research, 19*, 558-570.

Liaw, A., & Wiener, M. (2002). Classification and Regression by RandomForest. *R News, 2,* 18-22.

Lindberg, E., Holgren, J., Olofsson, K., Wallerman, J., & Olsson, H. (2010). Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. *International Journal of Remote Sensing, 31,* 1175-1192.

Magnussen, S., & Boudewyn, P. (1998). Derivations of stand height from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research, 28*(7),1016-1031.

Magnussen, S., Næsset, E., & Gobakken, T. (2009). Reliability of LiDAR derived predictors of forest inventory attributes: A case study with Norway spruce. *Remote Sensing of Environment, 114,* 700-712.

Maltamo, M., Bollandsås, O. M., Næsset, E., Gobakken, T., & Packalén, P. (2011). Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry, 84,* 23-31.

Maltamo, M., Eerikäinen, K., Packalén, P., & Hyyppä, J. (2006). Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry,* 79, 217-229.

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J., & Vehmas, M. (2004). Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment, 90,* 319-330.

McCombs, J. W., Roberts, S. D., & Evans, D. L. (2003). Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *Forest Science, 49,* 457-466.

McInerney, D. O., Suarez-Minguez, J., Valbuena, R., & Nieuwenhuis, M. (2010). Forest canopy height retrieval using LiDAR data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland. *Forestry, 83,* 195-206.

Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L., & Hendrix, C. J. (2000) Stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing, 66*, 1367-1371.

Monnet, J.-M., Chanussot, J., & Berger, F. (2011). Support vector regression for the estimation of forest stand parameters using airborne laser scanning. *IEEE Geoscience and Remote Sensing Letters, 8,* 580-584.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment, 80*, 88-99.

Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research, 19,* 164-179.

Næsset, E., Bollandsås, O. M., & Gobakken, T. (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment, 94,* 541-553.

Næsset, E., & Økland, T. (2002). Estimating tree height and tree crown properties using airborne

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

scanning laser in a boreal nature reserve. *Remote Sensing Environment, 79,* 105-115.

Ni-Meister, W., Lee, S., Strahler, A. H., Woodcock, C. E., Schaaf, C., Yao, T., Ranson, K. J., Sun, G., & Blair, J. B. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research, 115*, G00E11.

Packalén, P., & Maltamo, M. (2007). The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment, 109,* 328-341.

Parker, R. C., & Evans, D. L. (2009). LiDAR forest inventory with single-tree, double-, and single-phase procedures. *International Journal of Forestry Research,* Article ID 864108; doi:10.1155/2009/864108.

Peuhkurinen, J., Mehtätalo, L., & Maltamo, M. (2011). Comparing individual tree detection and the area-based statistical approach for the retrieval of forest stand characteristics using airborne laser scanning in Scots pine stands. *Canadian Journal of Forest Research, 41,* 583-598.

R-Development Core Team. (2007). *R: A language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. (2010). *BMA: Bayesian Model Averaging.* R package version 3.13. Retrieved September 2010 from http://CRAN.R-project.org/package=BMA

Ripley, B. (2010). *Classification and regression trees*. R package version 1.0-28 (2010). Retrieved September 2010 from http://CRAN.R-project.org/package=tree.

Rombouts, J., Ferguson, I. S., & Leech, J. W. (2010). Campaign and site effects in LiDAR prediction models for site-quality assessment of radiata pine plantations in South Australia. *International Journal of Remote Sensing, 31,* 1155-1173.

Rooker Jensen, J. L., Humes, K. S., Conner, T., Williams, C. J., & DeGroot, J. (2006). Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint lidar. *Canadian Journal of Forest Research, 36,* 1129-1138.

Sokal, R. R., & Rohlf, F. J. (1995). *Biometry, Third Edition*. New York, USA: W.H. Freeman and Company.

Stojanova, D., Panov, P., Gjorgjioski, V., Kohler, A., & Dzeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics, 5,* 256-266.

Stone, C., Turner, R., Kathuria, A., Carney, C., Worsley, P., Penman, T., Bi, H.-Q., Fox, J., & Watt, D. (2011). *Adoption of new airborne technologies for improving efficiencies and accuracy of estimating standing volume and yield modelling in* Pinus radiata *plantations.* (Final Report for Project PNC0-0809). Melbourne, Australia: Forest & Wood Products Australia Ltd. Available at www.fwpa.com.au

Straub, C., Weinacker, H., & Koch, B. (2010). A comparison of different methods for forest resource estimation using information from airborne laser scanning and CIR orthophotos. *European Journal of Forest Research, 129,* 1069-1080.

Tesfamichael, S. G., van Aardt, J. A. N., & Ahmed, F. (2010). 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.

van Aardt, J. A. N., Wynne, R. H., & Oderwald, R. G. (2006). Forest volume and biomass estimation using small-footprint lidar-distributional parameters on a per-segmented basis. *Forest Science, 52,* 636-649.

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

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., & Kavanagh, R. P. (2003). The use of bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology, 17,* 1579-1590.

Wintle, B. A., Elith, J., & Potts, J. M. (2005). Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology, 30,* 719-738.

Yu, X., Hyyppä, J., Holopainen, M., & Vastaranta, M. (2010). Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. *Remote Sensing, 2,* 1481-1495.

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)

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.

ISSN 0048-0134 (print)
ISSN 1179-5395 (on-line)