Department of Primary Industries

Technical Report for Local Land Services

Site Index Mapping for Private Native Forest, North and South Coast NSW using LiDAR Data

www.dpi.nsw.gov.au

Published by the NSW Department of Primary Industries

Amrit Kathuria

Title: Site Index Mapping for Private Native Forest, North and South Coast NSW using LiDAR data

First published March 2021

ISBN number

More information

amrit.kathuria@dpird.nsw.gov.au / Forest Science, DPI Forestry, Parramatta, NSW

Acknowledgements

This project was funded by Local Land Services and DPI. Thanks to Nick Cameron for technical advice on forest related issues, Murray Webster for training data, Masoomeh Alaibakhsh for GIS support and Sam Hislop for processing of spatial services Lidar data.

This report is one of 12 deliverables arising from the NSW Local Land Services Private Native Forests (PNF) research and development program with DPI Forest Science, available at https://www.dpi.nsw.gov.au/forestry/science

www.dpi.nsw.gov.au

Your Reference number (generally CM9)

Contents

Executive Summary

This study modelled the site productivity of forest on the NSW north coast using airborne laser scanner (ALS) data. Site productivity is a quantitative estimate of the potential of a site to produce plant biomass. It is often quantified as an index, typically site class or site index. Most commonly used site indices are based on or derived from estimates of stand height at a given age. As forest height is captured really well using ALS data accurate site index can be estimated using ALS metrics. The model may be readily applied to other regions where LiDAR data is available.

The aims of this study were to:

a) Analyse the relationship between dominant height of forest and ALS metrics
b) Develop a spatially-explicit model (20m resolution) of dominant height using ALS data
c) Produce a site Index map of North Coast NSW forest
d) Produce a site Index map of South Coast NSW forest

Key findings of the study were:

- Dominant height was highly correlated with a number of ALS metrics, such as average height, 99th height percentile, 95th height percentile and 90th height percentile of first returns.

- A spatially explicit model for site index forest was developed at 20m resolution for North Coast NSW.

- The model showed a good agreement with the reference data, the estimated bias was low (-0.17 m) and low root mean square error (RMSE) of 3.5 m and relative RMSE: 12.7%

- The model was applied to the ALS data from spatial services and 20 m resolution site index map for North Coast NSW forest was created.

- The model was applied to the ALS data from spatial services and 20 m resolution site index map for South Coast NSW forest was created.

Introduction

Reliable estimates of site productivity are crucial for the sustainable management of forest resources. Site refers to a geographic location that is considered homogeneous in terms of its physical and biological environment, and Site productivity is a quantitative estimate of the potential of a site to produce plant biomass (Pretzsch, 2009). A narrow definition is generally used in forestry that refer to part of the site potential that is or is expected to be realized by the trees for wood production. In that sense, forest site productivity may be defined as the potential of a particular forest stand to produce aboveground wood volume, referring to the production unit formed by the site and the stand of trees in concert. Within this narrower context, site productivity is often quantified as an index, typically site class or site index. Most of the indices are derived from estimates of stand height at a given age. Top and dominant height, are least affected by thinning and are the most stable height-based indicators of site productivity (Næsset, 1997, Skovsgaard, 2008).

The aims of this study were to:

a) Analyse the relationship between dominant height of forest and ALS metrics
b) Develop a spatially-explicit model (20m resolution) of dominant height using ALS data
c) Produce a site Index map of North Coast NSW forest
d) Produce a site Index map of South Coast NSW forest

Data

Study Area

The study area is private native forest on the NSW North Coast that has NSW Spatial Services LiDAR coverage (as shown at Figure 1).

Data for Model development: Eden Data (FCNSW)

The data from Eden Native forest was used. The Eden CRA Region totals about 800 000 hectares and extends, in broad terms, from Bermagui and Nimmitabel in the north to Delegate and Cape Howe in the south. About two thirds of the area is forested and most of the forest is on Crown lands – either as State forest (25% of the total area) or national park (32%). Other Crown-timber lands are a minor proportion of the total area (about 2%) (Towards an Eden Regional Forest Agreement, 1998).

The data file contains 350 forest plots with mean top height (MTH) as the response variable (figure 2). Mean Dominant Height is the mean height of the tallest 40 trees per hectare or 4 trees for a 0.1 ha plot. ALS metrics were calculated for the plots and used as predictor variables.

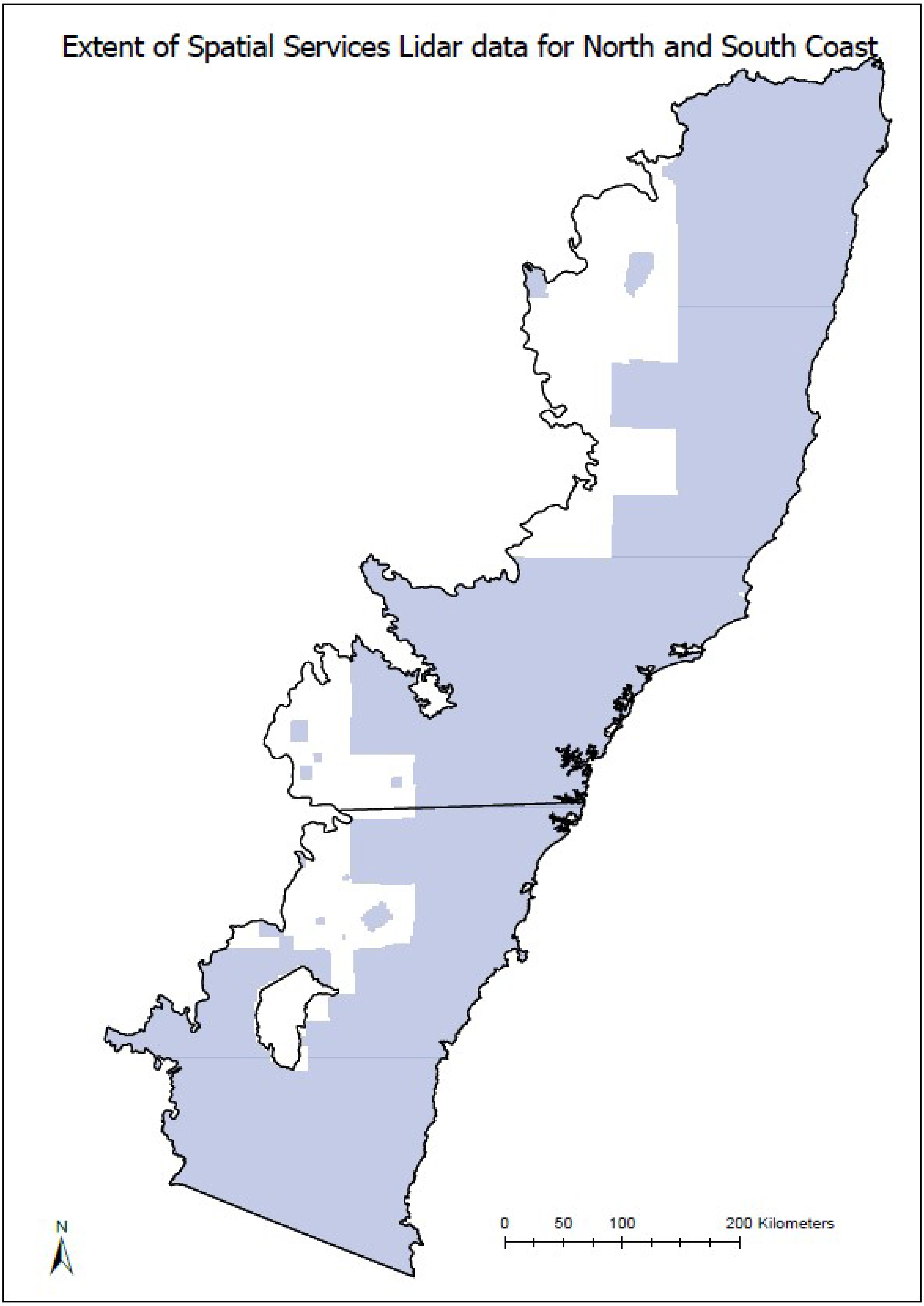

Figure 1: Spatial services LiDAR data showing the extent of the LiDAR data for North and South Coast PNF areas.

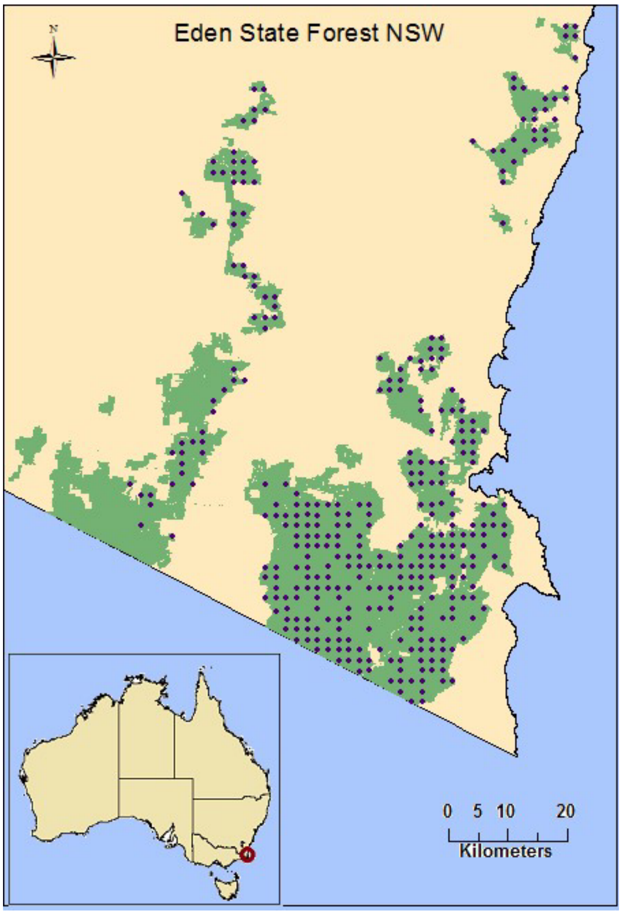

Figure 2: Location of 350 Eden study plots.

Validation data: Forest monitoring feasibility data

Data from 58 plots were collected for forest monitoring pilot study. Plot level data was collected from three regions: NSW North Coast (NNC), Brigalow Belt South (BBS) and South Eastern Highlands (SEH). There were 20 plots from NNC, 24 from BBS and 14 from SEH. The plots were spread across three tenures; National Parks (23), State forest (20) and Private property (15). The distribution of the plots by Region and by Tenure is presented in Table 1. Lidar data was available for 57 (Figure 3) plots and height information was not collected in 4 plots. So data from 52 plots was used for validation.

Table 1: IBRA and tenure of the Forest Monitoring feasibility study plot data.

	Region			
Tenure	**BBS**	**NNC**	**SEH**	**Total**
National Park	11	7	5	23
Private Property	3	8	4	15
State Forest	10	5	5	20
Total	**24**	**20**	**14**	**58**

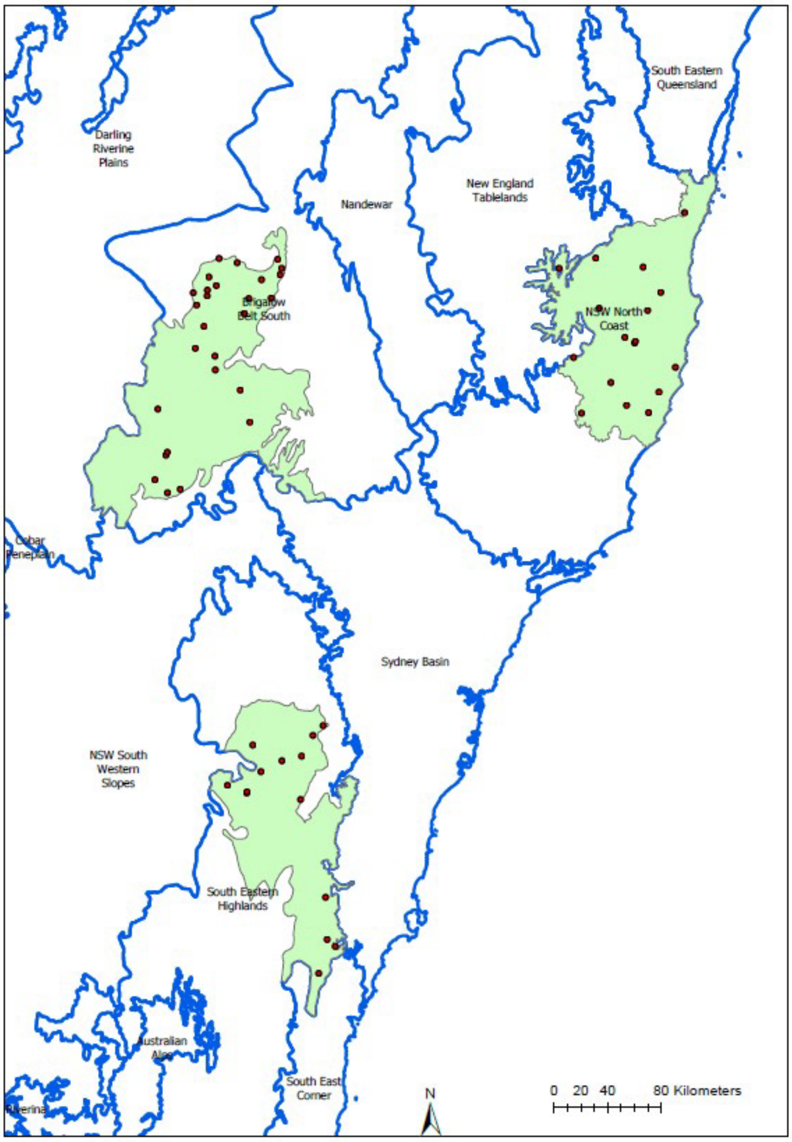

Figure 3: Location of the Forest Monitoring feasibility study plots.

Methods

LiDAR Processing

The point density for the ALS point cloud data for FCNSW was 1 to 2 points per square meter. LAStools (rapidlasso GmbH) software were used to process the ALS point cloud data. The noise points were identified using lasnoise, the data was then classified ground and non-ground. Digital terrain models were created and normalised point cloud data was created. Once the ALS data were normalized, the point data were processed to extract ALS metrics at the plot level.

A total of 50 ALS variables were initially selected. The final variables for model were selected after looking at a number of publications and identifying the variables strongly correlated with MTH.

Multiple Regression

As is often the case with ALS data analysis, there is a large number of predictor variables in the model. A number of the predictor variables were highly correlated, so a number of variable selection methods were examined to select the final set of predictor variables. Initially, a Spearman's correlation matrix was used to reduce the number of predictor variables and 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 0.9, we selected one variable and removed all others. This was done on the basis of selecting the variable from the group of highly correlated variables that made the most biological sense and were easy to interpret in terms of their effect on tree identification.

A number of techniques have been developed to reduce the number of variables such as; forward, backward and best subset selection. There are techniques which use p values, R2 values, Akaike information criterion (AIC), Bayesian information criterion (BIC) values as the selection criteria (James et al. 2014,P. Bruce and Bruce (2017)). None of these methods are fool proof and care has to be taken in their application. We used the best subset selection method with AIC to select for the predictor variables (Harrell 2014).

Model evaluation

The precision of models were compared using the root mean square error (RMSE)

$$\mathrm{RMSE}(y_i) = \sqrt{\frac{\sum_1^{\mathrm{n}}(y_{obs}-y_{imp})^{\wedge}2}{\mathrm{n}}} \quad (1)$$

and bias

$$\mathrm{bias}(y_i) = \frac{\sum_1^{\mathrm{n}}(y_{obs}-y_{imp})}{\mathrm{n}} \quad (2)$$

where n is the number of plots, y_{obs} is the observed value of the stand variable y, and y_{imp}, is the imputed value. RMSE and bias were calculated in relative terms (%RMSE and %bias), i.e. the RMSE and bias values for the stand variable y were divided by their observed mean values and multiplied by 100.

Model Assessment and Validation

Model assessment and model validation is an important step in the development of a model. Model validation, in its simplest form, is a comparison between predicted and observed values. All models were validated using apparent validation (same data used for development and testing). Apparent performance is attractive because it is easy to perform, it results however in optimistic estimates of performance. Internal validation methods were used to evaluate model performance on an underlying population (also labelled 'reproducibility'). Internal validation methods aim to provide a more accurate estimate of model performance for new data that is not used in model building. This was done by dividing the reference data into training and test data and the validation was done using the test data. Eden data was used for model development and Forest Monitoring data was used as external validation. The model was developed using Eden data and was applied to Forest Monitoring data and errors were calculated.

Map Creation

Using the results of the model assessment, site index map for North and South Coast NSW was created at 20m resolution.

Results

Variable of Interest

Dominant height of mature forest is a good indicator of site quality. For Eden data there were 350 plots but only 54 of them were mature forest plots. A number of ALS height metrics are extracted for the plots and some of them have been shown to be highly correlated with MTH. The plot of MTH vs ALS metrics; average height, P90, p95, and p99 for Eden data is presented in figure 4. The ALS height metrics are highly correlated with MTH. Avg is under estimating MTH and p99 over estimates the MTH. If only one variable is to be used then p90 is the closest to MTH. A plot of the distribution of the MTH and ALS metrics is presented in figure 5. Again similar trends are observed. The plot at the top are the ALS metrics calculated using the first returns and the one at the bottom is for all returns. Again, p90 with first returns with returns >2m, is the closest in distribution to MTH.

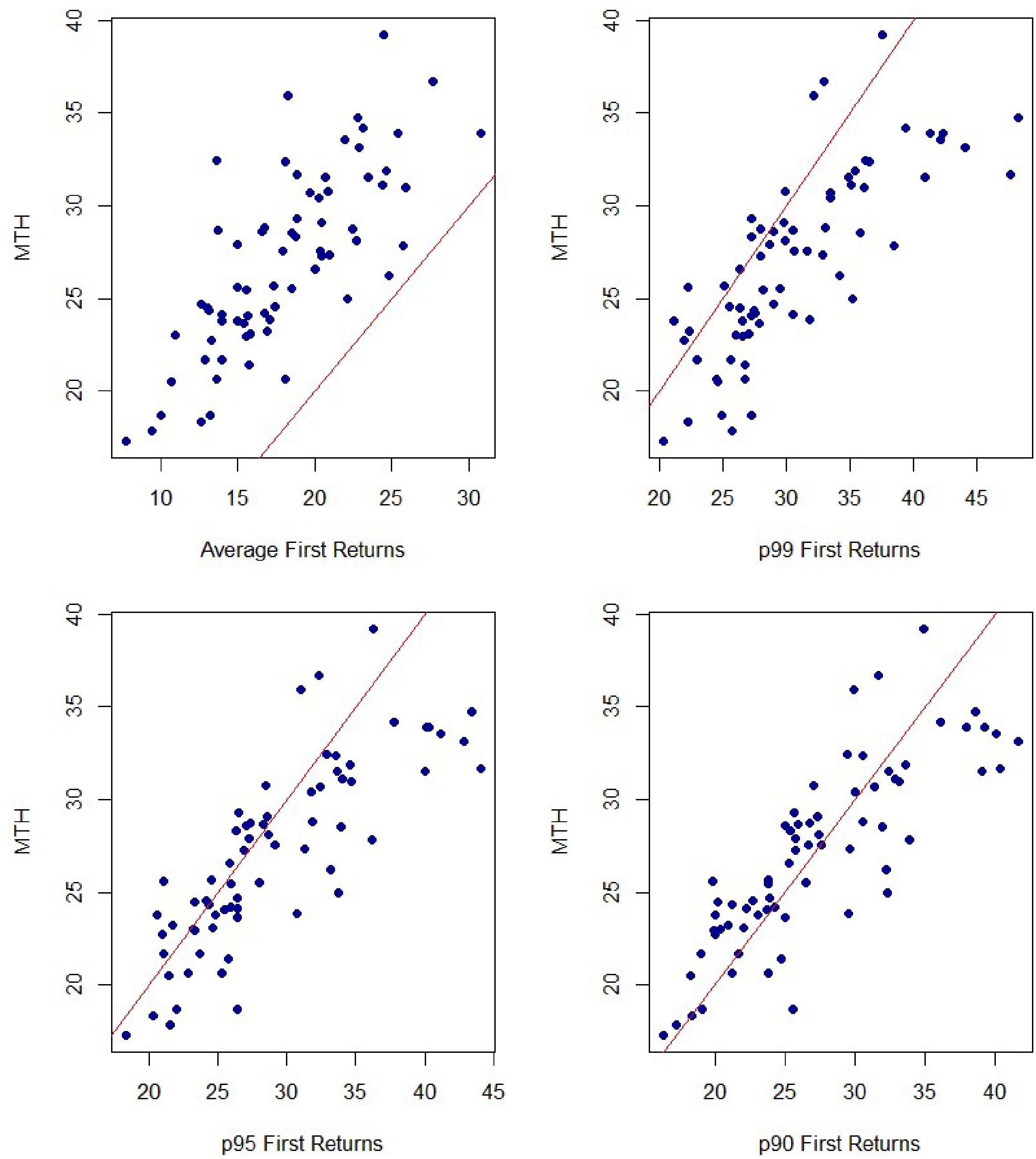

Figure 4: Plot of Mean top Height (MTH) with the height ALS metrics for mature forest plots of Eden Study area.

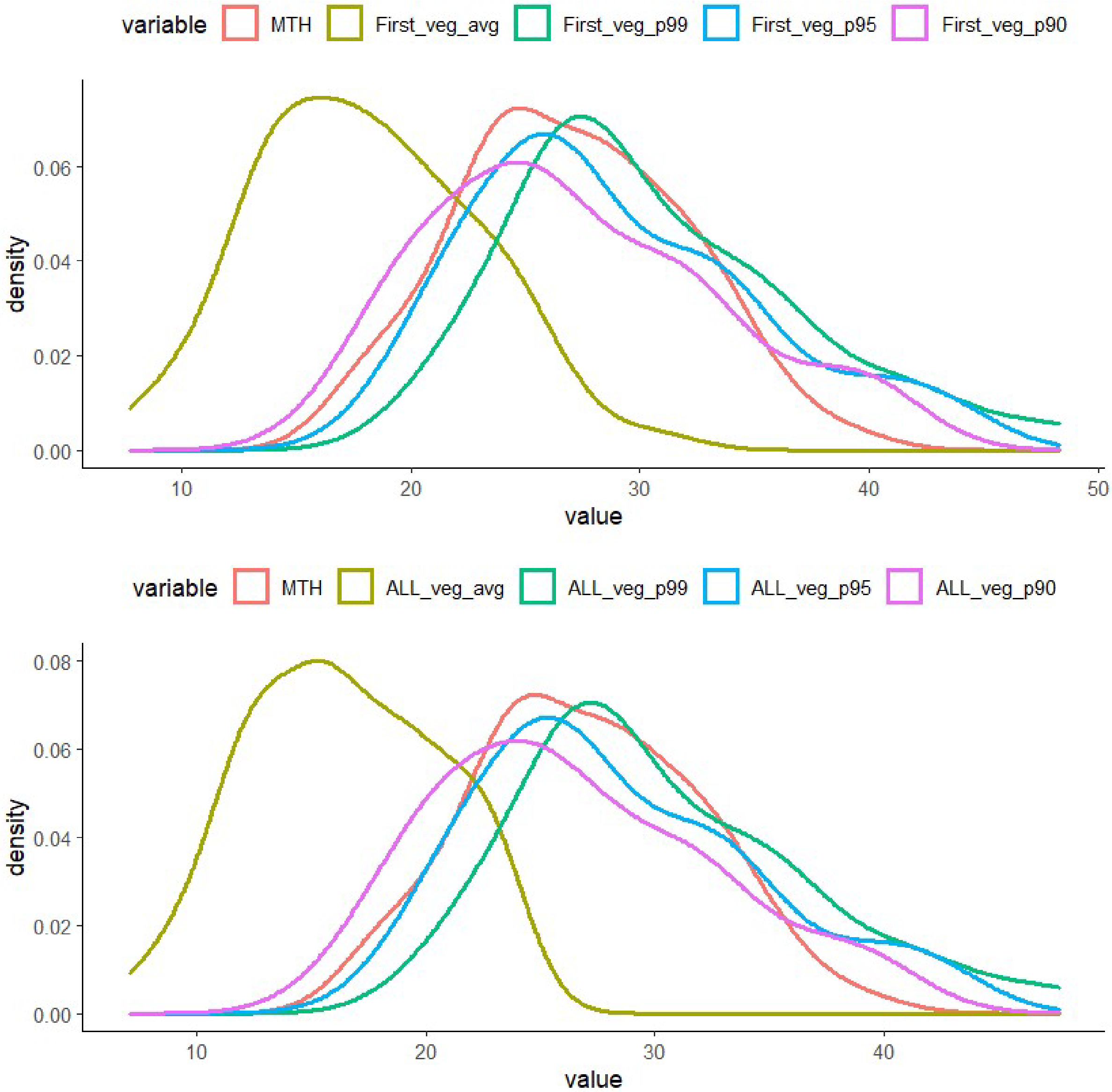

Figure 5: Density plot of Mean top Height (MTH) with the height ALS metrics for mature forest plots of Eden Study area.

A multiple regression model was fitted to Eden data using a best subset selection method. Average of first returns and p99 of first returns were the two variables that were significant after best subset selection. The modelled values had a 0 bias and RMSE of 2.6 m, relative RMSE 9.7%. This is smaller than the bias when only p90 was used as predictor, bias was low (-0.17 m) and the RMSE of 3.5 m and relative RMSE, 12.7%. Two maps were created one with the multiple regression model and the other with p90 as the variable. A focal smoothing is applied using a 3 by 3 cell window. A focal operation applies an aggregation function to all cells within the specified neighbourhood, uses the corresponding output as the new value for the central cell, and moves on to the next central cell. A mean value was used an aggregation method. A visual inspection of the map produced using both the multiple regression and the p90 approaches with the canopy height model (CHM) suggest that p90 is closer when the top height is >40m. The range of MTH for the Eden study data is 17.3m to 39.2m. So in absence of data covering the full range of MTH values p90 was selected for the final map. Figure 6 is the map of the CHM and the final map produced for three PNF properties. There appears to be a very good match of the top heights and the dominant height produced using ALS data (figure 6).

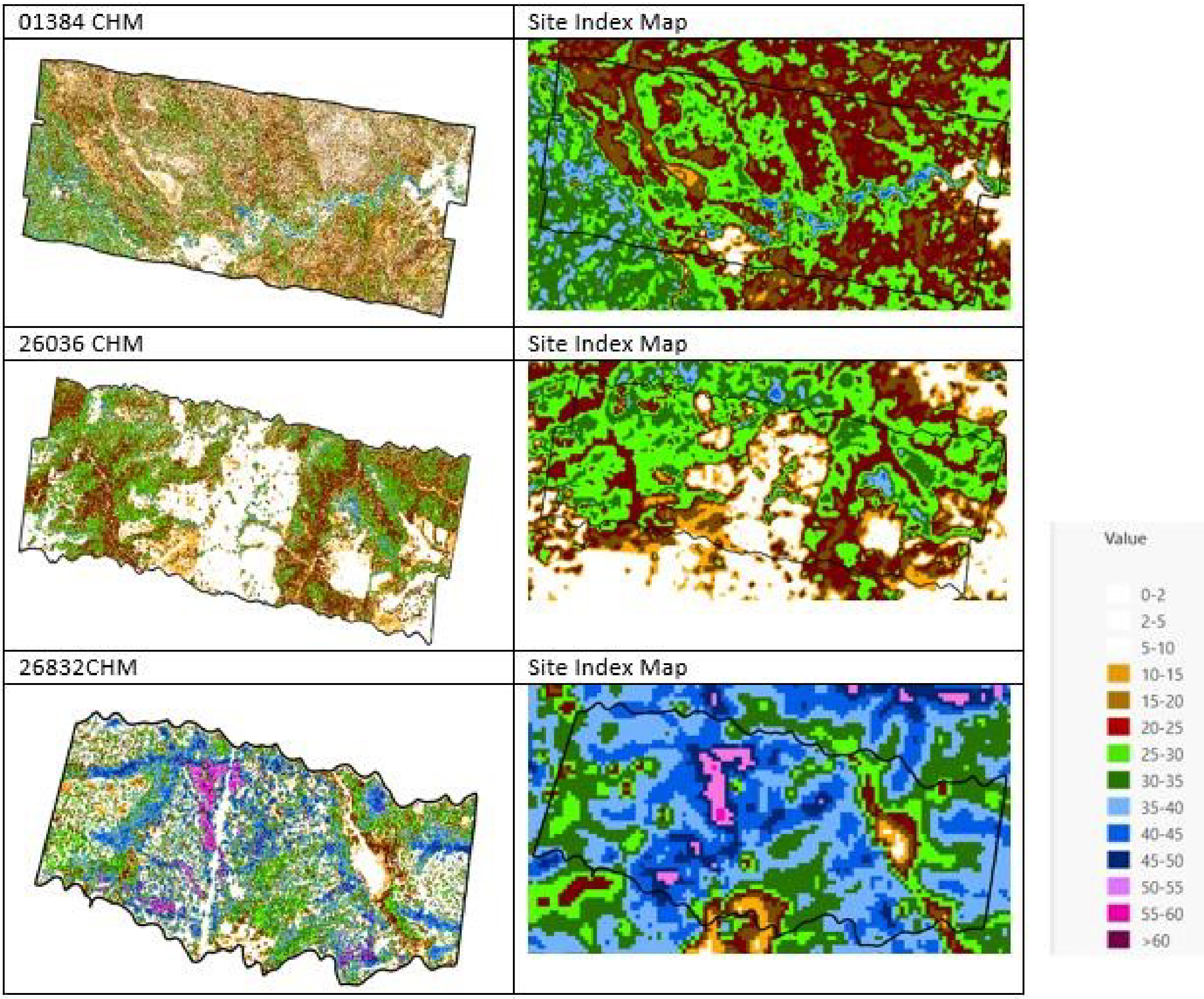

Figure 6: Canopy height model and the corresponding site index map for three PNF properties.

The model was applied to the data from Spatial Services for north and South coast and a site index map for North Coast forest is produced and is presented in Figure7 and Figure 8. Figure 9 presents the combined map for the PNF area in the north and south coast of NSW. Histogram plots of site index is presented in figures 10 and 11 and the table 2 has the area and percentages in the various size classes.

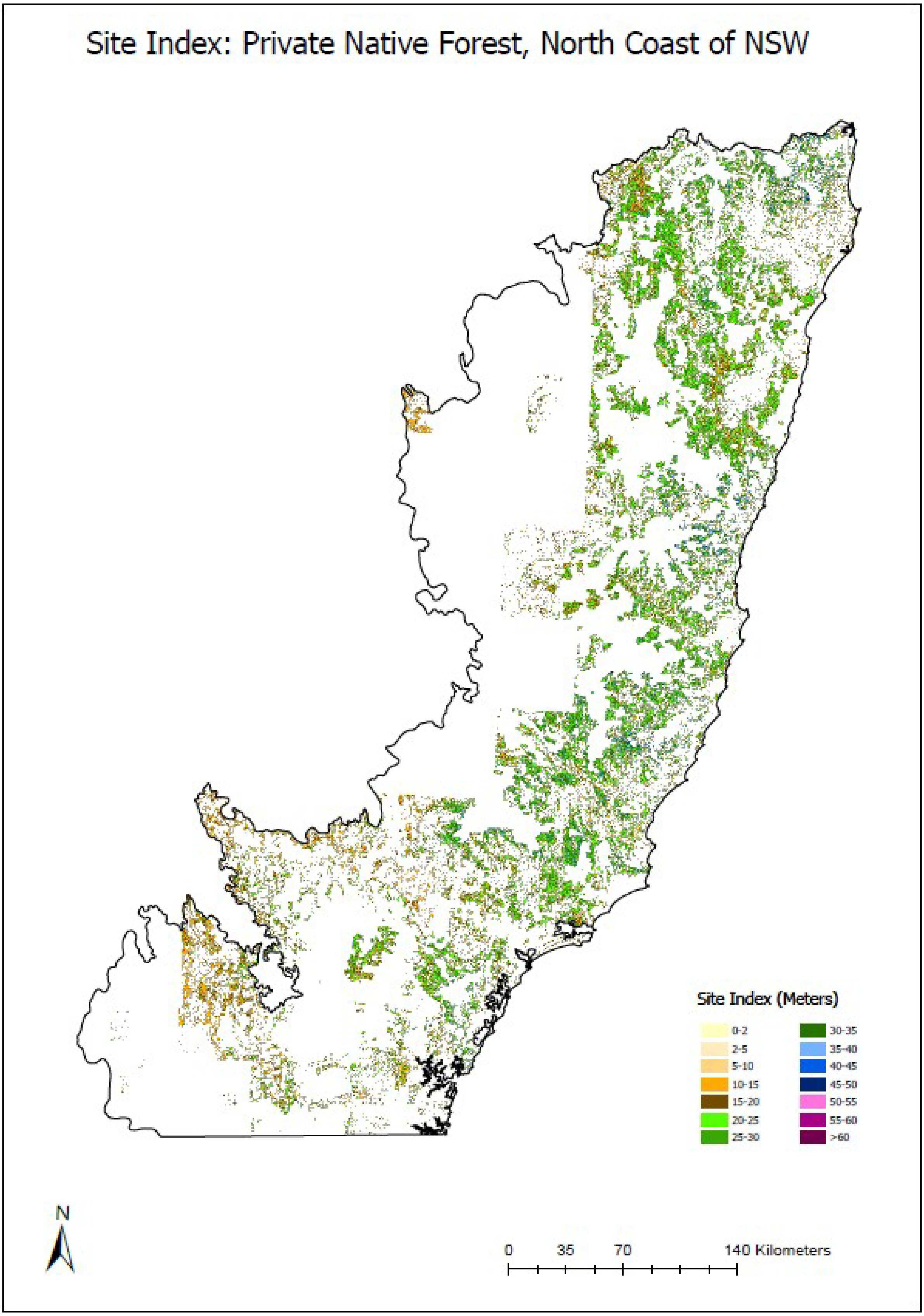

Figure 7: Site Index map of forest for North Coast NSW within the Northern NSW PNF Code region.

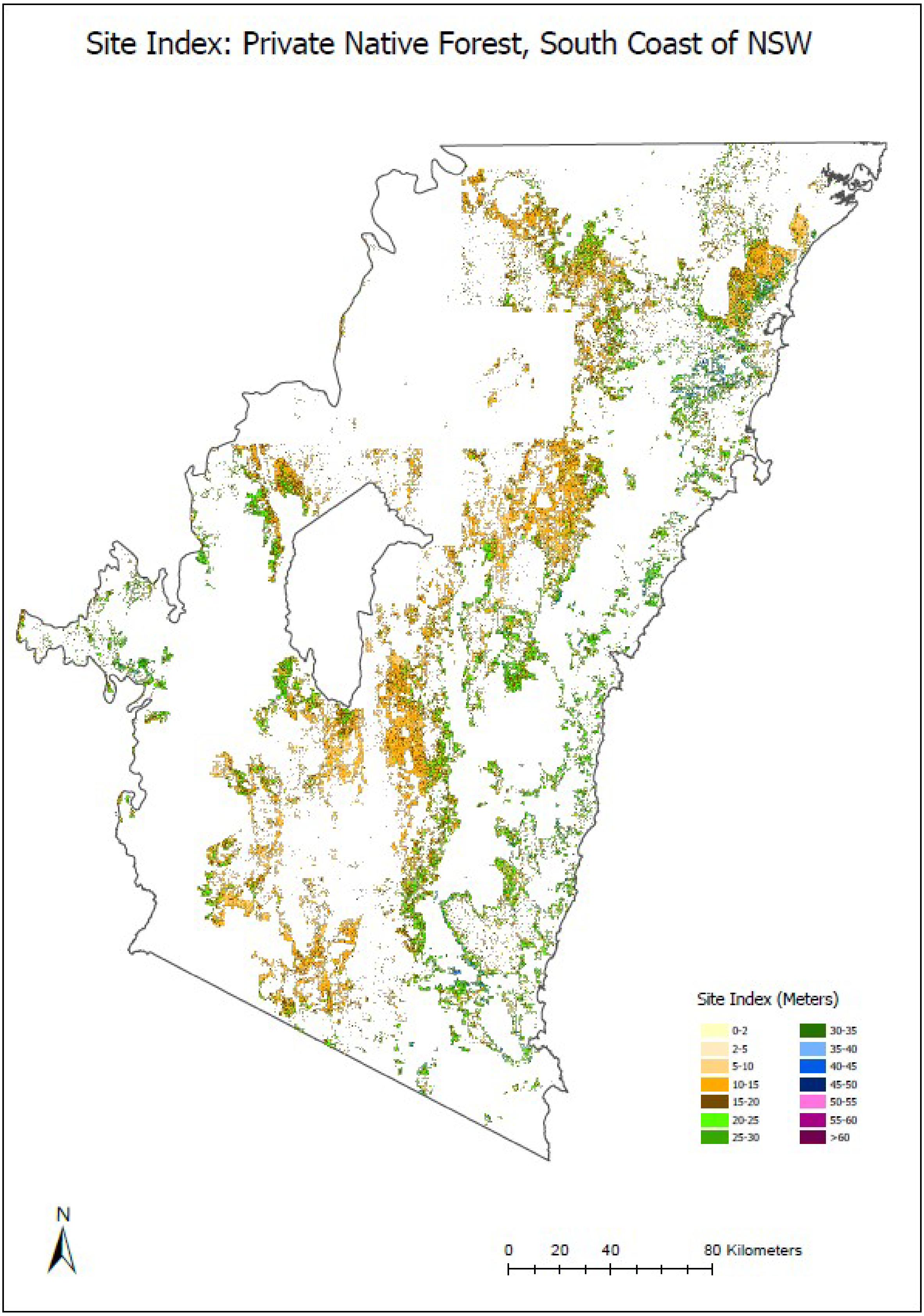

Figure 8: Site Index map of forest for South Coast NSW within the Southern NSW PNF Code region.

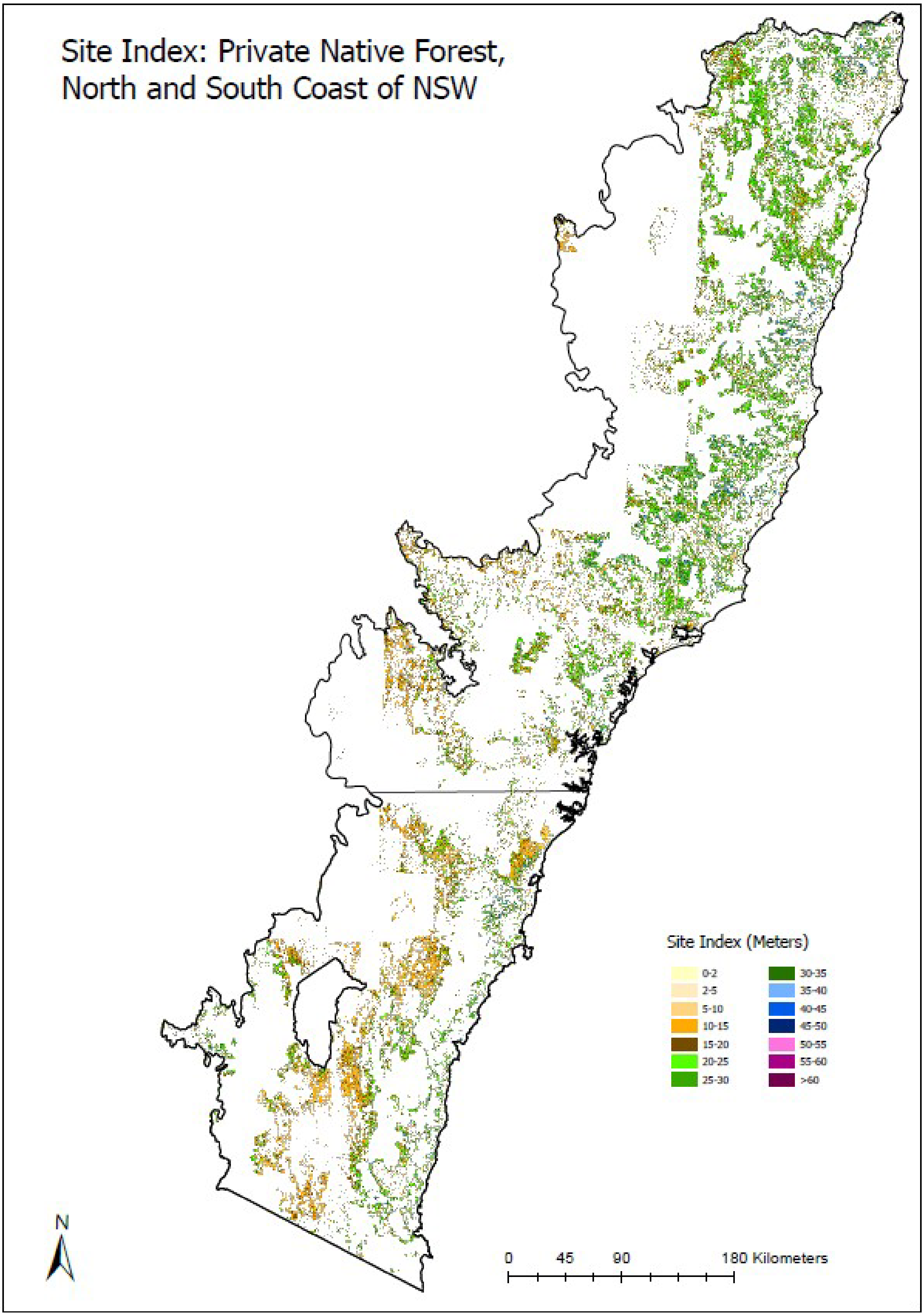

Figure 9: Site Index map of forest for North and South Coast NSW within the Northern and Southern NSW PNF Code region.

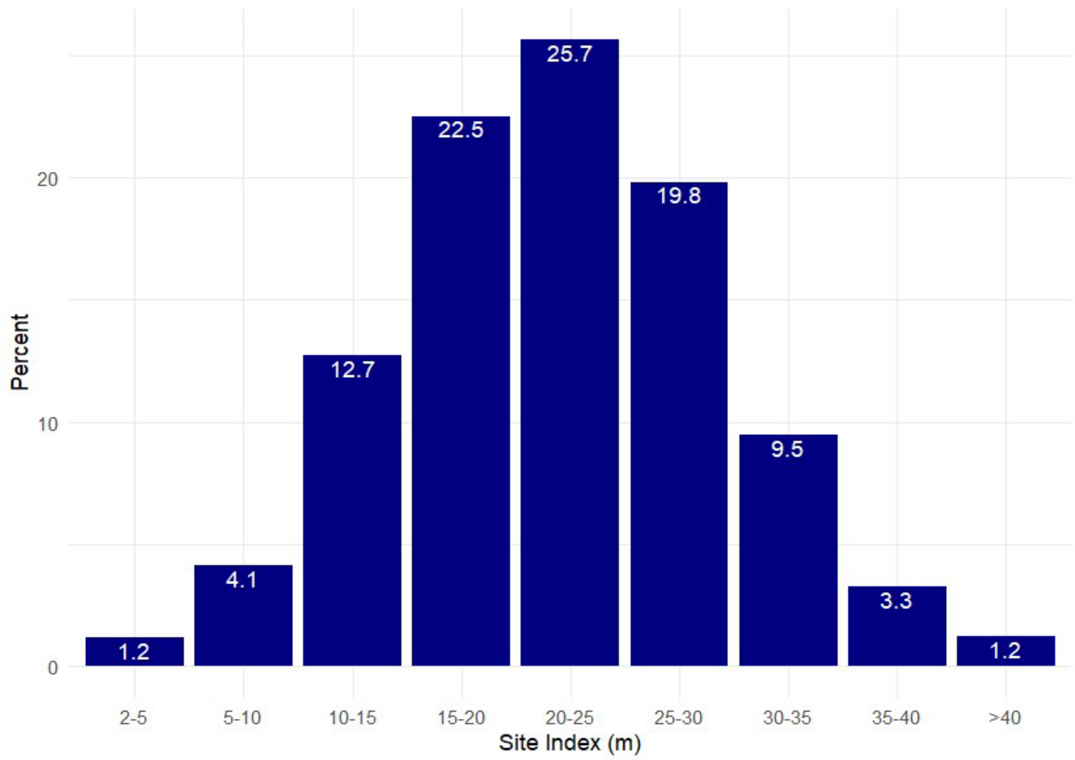

Figure 10: Percentage of Private property by site index class for Northern PNF region.

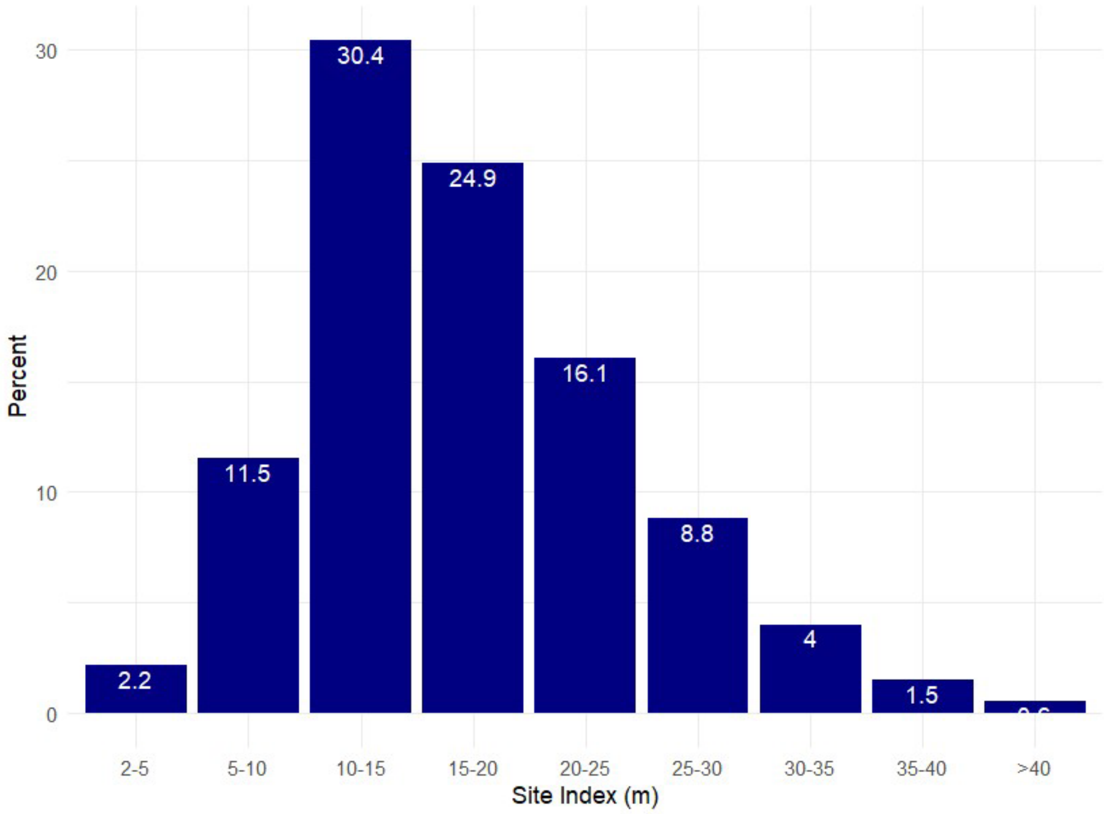

Figure 11: Percentage of Private property by site index class for Southern PNF region.

Table 2: Area and percentage of private property by site index class for forest area.

Site Index Class (m)	Northern PNF		Southern PNF	
	Percentage Area	Area (ha)	Percentage Area	Area (ha)
2-5	1.2%	28820	2.2%	22660
5-10	4.1%	100288	11.5%	119704
10-15	12.7%	308161	30.4%	315205
15-20	22.5%	545951	24.9%	257734
20-25	25.7%	622219	16.1%	166388
25-30	19.8%	480588	8.8%	91498
30-35	9.5%	229487	4.0%	41767
35-40	3.3%	79288	1.5%	15729
>40	1.2%	30286	0.6%	5789

Limitations of the Study

Coverage

Spatial Services ALS layer does not cover all of the PNF Plans in the Northern and Southern NSW PNF (Figure1), so for some PNF properties site index map is not available. But the number of such properties is small.

LiDAR

The LiDAR data used to produce the map is Spatial Services data, the point density for this data is much less than data collected in 2020 for LLS project but it has been shown that the height metrics for the two point densities are not very different (Kathuria, A. 2020). The data was collected over a long period of time, mostly from year 2001 to 2015. Any changes that happened, such as harvesting is not captured here. For area with harvesting and regrowth, site index would be lower than the actual value, but regrowth was estimated to be 1.57% using reference data from six largest properties .

Future Work

The Eden data and the FM data did not have MTH values greater than 45m. So, look for other data sources so that the multiple regression model could cover the whole range of MTH values for the north coast. Also, an attempt was made to classify regrowth, more work needs to be done on that. The reference data from six largest properties indicates that there was 1.57% regrowth.

References

Bruce, Peter, and Andrew Bruce. 2017. Practical Statistics for Data Scientists. O'Reilly Media.

Chatterjee, S., Hadi, A.S. and Price, B (2000) Regression analysis by example. John Wiley & Sons, New York.

Harrell, F.E. (2014) Regression Modelling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis. Springer, New York.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated.

Kathuria, A. (2020). Technical Report for Local Land Services: Commercial and NonCommercial Forest Mapping for North Coast NSW using LiDAR Data. DPI Ref: RDOC20/87

NÆSSET, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS J. Photogram. Remote Sens. 52:49–56.

Pretzsch, H. 2009 Forest Dynamics, Growth and Yield. Springer, Berlin, Heidelberg. XIX + 664 pp.

R Development Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Skovsgaard, J.P. and Vanclay, J.K. 2008 Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry 81, 13–31.

Towards an Eden Region Forest Agreement, New South Wales Government Sydney Commonwealth Government Canberra (1998).