Growing confidence in forestry's future Research Programme

SCION forests · products · innovation

Date: June 2016
Reference: GCFF TN-09

Technical Note

Refined models of leaf area index from LiDAR

Summary:

Estimation of LAI using remote sensing techniques offers the potential for large scale LAI assessment which may be used to identify diseased and infertile areas of forests for targeted interventions. Previous work to estimate LAI from LiDAR has shown improved results over spectral-based methods. In this tech note we summarise work done to improve LiDAR-based estimates of LAI in New Zealand's radiata pine plantations. A large number of candidate LiDAR metrics were tested across a wide range of plot parameters using two different modelling approaches.

Here we show that the use of non-standard LiDAR metrics and plot parameters greatly improved accuracy of LAI estimates. LiDAR data extracted from a variable plot radius equal to 100% or 150% of canopy height were optimal, with mean squared error (MSE) of 0.32 – 0.35 for the best models. However, a fixed radius of 20 – 26 m from plot centre also produced strong relationships between LAI and LiDAR (MSE for top models of 0.41 – 0.43).

Above a minimum level of plot radius the ratio metrics emerged as the strongest predictors of LAI. This class of metrics computes ratios of returns by type above and below a specified height threshold. The choice of height threshold strongly impacted model performance and complexity. In our study, the optimal height threshold was between 5 – 6 m above ground. Models utilising ratio metrics constructed at these height thresholds produced the most accurate models, and ratio metrics were selected and ranked well above other metrics in this region. An alternative approach using a variable height threshold showed that a height threshold equal to 20% of canopy height for individual plots was also effective in ratio metrics, and is likely to perform better where plot height is highly variable.

Elastic-net regularised regression models produced the best models of LAI from LiDAR, with reduced complexity at the optimal combination of LiDAR plot radius, and ratio metric height threshold. Models based on Random Forests did not improve on the elastic-net approach. The scoring of metric importance from the Random Forests algorithm largely agreed with metric selection in elastic-net, and highlighted the importance of ratio metrics for LiDAR-LAI estimation.

Based on the results of this study we make general recommendations for those seeking to estimate LAI from LiDAR in New Zealand's radiata forests, such as the choice of LiDAR plot parameters and candidate metrics. The success demonstrated using this approach opens up future opportunities to explore the use of LiDAR-LAI to detect gradients of fertility and disease across large areas.

Author/s: Grant Pearse, Michael Watt, Justin Morgenroth

Corresponding author: Grant Pearse: grant.pearse@scionresearch.com

supported by forestgrowers commodity levy

FOREST OWNERS ASSOCIATION

This programme is also funded by Scion and the Ministry of Business, Innovation and Employment, New Zealand.

Introduction

Leaf area index (LAI) is a measure of leaf or needle surface area per unit ground surface area[1]. The link between photosynthesis and LAI makes it a key forest physiological trait with close links to forest health, productivity, and site fertility[2]. These properties also make measurement of LAI valuable to forest managers, with potential applications for disease identification, disease control, and targeted site interventions such as aerial fertilisation. Spectral indices with theoretical links to photosynthesis, for example the normalised difference vegetation index (NDVI), have been used to successfully predict LAI across large areas of grassland. However, these techniques do not perform well in forested areas where sensor saturation (LAI > 4), understory interference, and processing difficulties limit the usefulness of these approaches[3–5].

The advent of LiDAR has provided a new set of tools for LAI estimation in forest ecosystems. In contrast to spectral methods, LiDAR-LAI estimation relies on the derivation of empirical relationships between ground based LAI measurements and LiDAR metrics, often with only tentative links to LAI-related canopy properties[6]. While this approach has often been used successfully, the lack of a solid theoretical link to LAI has resulted in a proliferation of approaches which are largely study specific. In the context of New Zealand forestry, very little work has been done to develop methods for LiDAR-LAI estimation in plantations of *Pinus radiata* (D. Don). In a previous GCFF tech note (TN008), we compared standard LiDAR metrics and a broad range of spectral indices to determine the best approach for LAI estimation in New Zealand's radiata forests. As anticipated, the LiDAR metrics showed much greater measures of association with LAI measurements taken from 21 plots across Kaingaroa forest. However, predictive models of LAI using standard LiDAR metrics from a larger sample (n = 134) of plots in Kaingaroa showed only modest agreement (R^2 = 0.61). These estimates were still considerably better than those from satellite imagery, but were weaker than some international examples of LiDAR-LAI. We attribute this result to the fact that initial efforts did not attempt to develop New Zealand specific methods for LiDAR-LAI estimation, and relied only on standard LiDAR metrics and methods.

In this tech note we describe an improved approach to LiDAR-LAI estimation in New Zealand's radiata forest type. This approach is based on a comprehensive evaluation of the optimal methodology for this forest type. In particular, three key research areas were addressed: (1) evaluation of a large number of LAI specific LiDAR metrics proposed in the literature, (2) determination of the optimal parameters for the construction of these metrics, (3) trialling the use of both linear and non-linear approaches to capture the true form of any relationship between LiDAR metrics and LAI.

Methods

Field data

LAI measurements were acquired using the LICOR LAI-2200C[7] in 134 plots across Kaingaroa forest. All plots were coincident with 0.06 ha inventory plots. Measured stand ages ranged from 10 – 36 years, and mean top height (MTH) ranged from 14 – 43 m. Plot LAI was determined as the average of 18 – 25 measurements taken within the boundaries of the inventory plots. Measurements targeted a 95% confidence level that the plot mean LAI was within 10% of true LAI. The LAI measurements and post-processing were conducted using the protocol described by Pearse et al.[8].

LiDAR metrics

Several classes of LiDAR metrics have been proposed for LiDAR-LAI estimation. Our analysis included an extensive assessment of these metrics, here we present only a brief outline of our approach. As a first step, the list of useful metrics for LiDAR-LAI estimation were divided into the following classes:

Standard height metrics

This class includes standard height percentiles and descriptive statistics. These metrics depend only on the choice of a plot radius from which to extract LiDAR data and compute the relevant statistics. Optical LAI instrumentation (including the LAI-2200C) will view canopy elements beyond the boundaries of a small inventory plot. As a result, the choice of LiDAR plot radius impacts the strength of LiDAR-LAI relationships[6,9]. For example, Zhao & Popescu[6] found a fixed radius of around 25 m was best. Solberg et al.[9] abandoned use of a fixed radius, and found improved LiDAR-LAI relationships when the plot radius was set as a multiple of dominant tree height within the plot.

Ratio metrics

These metrics compute the ratio of returns (often separated by type) above and below a chosen height threshold. For example, canopy cover may be estimated as the fraction of first returns above DBH (1.4 m) to the total number of first returns. This class of metrics has been widely used for LiDAR-LAI estimation[6,10,11]. However, the optimal choice of a fixed height threshold to separate returns is not well established, and varies widely in the literature.

Complex metrics

These metrics are typically computed using more complex approaches such as division of the plot into smaller sub-pixels for metric computation. Metrics of this type have been used successfully in some settings, and are often based on tentative links to LAI theory[12,13].

LiDAR parameters for metric calculation

To identify the optimum combination of LiDAR parameters the following settings were trialled when generating the aforementioned metrics:

(1) All metrics were computed using LiDAR data extracted at fixed plot radii ranging between 8 – 30 m in 2 m increments from each plot centre.

(2) The mean top height for each plot was used as the basis for variable plot radii, with LiDAR data extracted at multiples of MTH ranging from 25 – 200% of MTH in 25% increments and used to generate all metrics.

(3) For ratio metrics, which require a height threshold to be chosen, each ratio was computed using height thresholds ranging from 0.5 m – 10 m above ground in 0.5 m increments.

(4) For our study we also introduce the novel concept of constructing ratio metrics using a variable height threshold. In this approach metrics such as canopy cover are computed as the ratio of returns above and below a height threshold determined as some fraction of maximum canopy height in each plot. This approach allows the height threshold to vary with tree height, whereas a fixed height threshold would be applied to all plots, regardless of tree height. We trialled variable height thresholds equal to 10% - 60% of the maximum canopy height within plots.

LiDAR data

Aerial LiDAR data were collected in early 2014 using an Optech Pegasus scanner. A maximum of four returns per pulse were captured. Calculation of LiDAR metrics was accomplished using a custom set of processing tools developed for the task. These tools extracted and processed LiDAR data at all fixed and variable radii specified, and for ratio metrics all fixed and variable height thresholds. These data were subset according to radius and height threshold and used to develop LiDAR-LAI models.

Modelling LAI from LiDAR

Much of the previous research on LiDAR-LAI estimation has relied on ordinary least squares regression (OLS)[12,14]. However, given the large number of metrics we trialled, and the large number of parameters used to generate these metrics, OLS was unsuitable. To overcome this, we chose two approaches to estimate LAI from the candidate LiDAR metrics. First, we fitted models based on elastic-net regression[15]. Elastic-net is a form of regularised linear regression that selects important variables while offering built-in controls for over-fitting. The glmnet package[16] in R [17] was used to fit elastic-net. Models were scored by mean squared error determined from 10 fold cross-validation.

For our second approach we used Random Forests[18] to model LAI from LiDAR. Random Forests is capable of capturing both linear and non-linear relationships. Importantly, while Random Forests does not provide variable selection, the algorithm provides a useful importance score for each of the variables used in the model (LiDAR metrics in our case). Models were fitted using the Party package for R[19]. MSE was calculated from the out-of-bag samples, and importance scores were computed using the conditional permutation importance measures described by[20]. By examining the variables selected in the best elastic-net models in conjunction with the Random Forests importance score, we were able to shortlist the best candidate metrics for LiDAR-LAI estimation in New Zealand's radiata pine forests. In addition, because models were fitted at every combination of height threshold and plot radius, the optimum parameters could be identified as those which produced the highest model accuracy.

Results

Optimum radius and height threshold

Results from models developed using elastic-net regression with metrics constructed from fixed radius LiDAR plots are shown in Figure 1. LiDAR data extracted from plots with a radius less than 18 m generally produced weaker models. The optimum range for plot radius was between 20 – 26 m. However, above 18 m plot radius the choice of fixed height threshold became relatively more important to model accuracy, with the optimum range occurring between 5 – 6 m. Use of a variable height threshold produced similar results to the fixed height threshold models, and accuracy was similarly impacted by plot radius.

The top 5% of models (outlined, Figure 1) were nearly tied, with a MSE of 0.41 - 0.43. Closer examination of these models showed large differences in complexity, and metric selection. As height threshold increased towards the optimum value of 5 – 6 m the number of ratio metrics present in the models showed a large increase, while overall model complexity decreased. This pattern was not true for the variable height threshold models, where the top models were generally more complex.

Use of a variable radius to extract LiDAR data (Figure 2) improved on the accuracy of fixed radius LiDAR-LAI models. The top 5% of models had a MSE of between 0.32 – 0.35. Results were more localised, with the best models produced when the radius was set at 100% or 150% of maximum canopy height. Within these bands, height threshold was again more important, with the optimum values once again between 5 – 6 m. The use of a variable plot radius set at 100% of MTH, in conjunction with a variable height threshold set at 20%

of max canopy height produced the best model of all those trialled with MSE of 0.32.

Repeating the model fitting process using Random Forests instead of elastic-net did not produce improved models of LAI. The best model for the fixed radius plots had identical MSE of 0.41. In contrast to elastic-net, variable radius plots did not produce better models, with the lowest MSE = 0.41. Random Forests models followed broadly similar trends in terms of the optimum radius and height threshold for fixed radius plots, confirming that larger radii of between 20 – 26 m, and height thresholds of 5 – 6 m produce the strongest agreement between LiDAR metrics and LAI. The weaker model performance in variable radius plots made trends in radius and height threshold less useful for comparison.

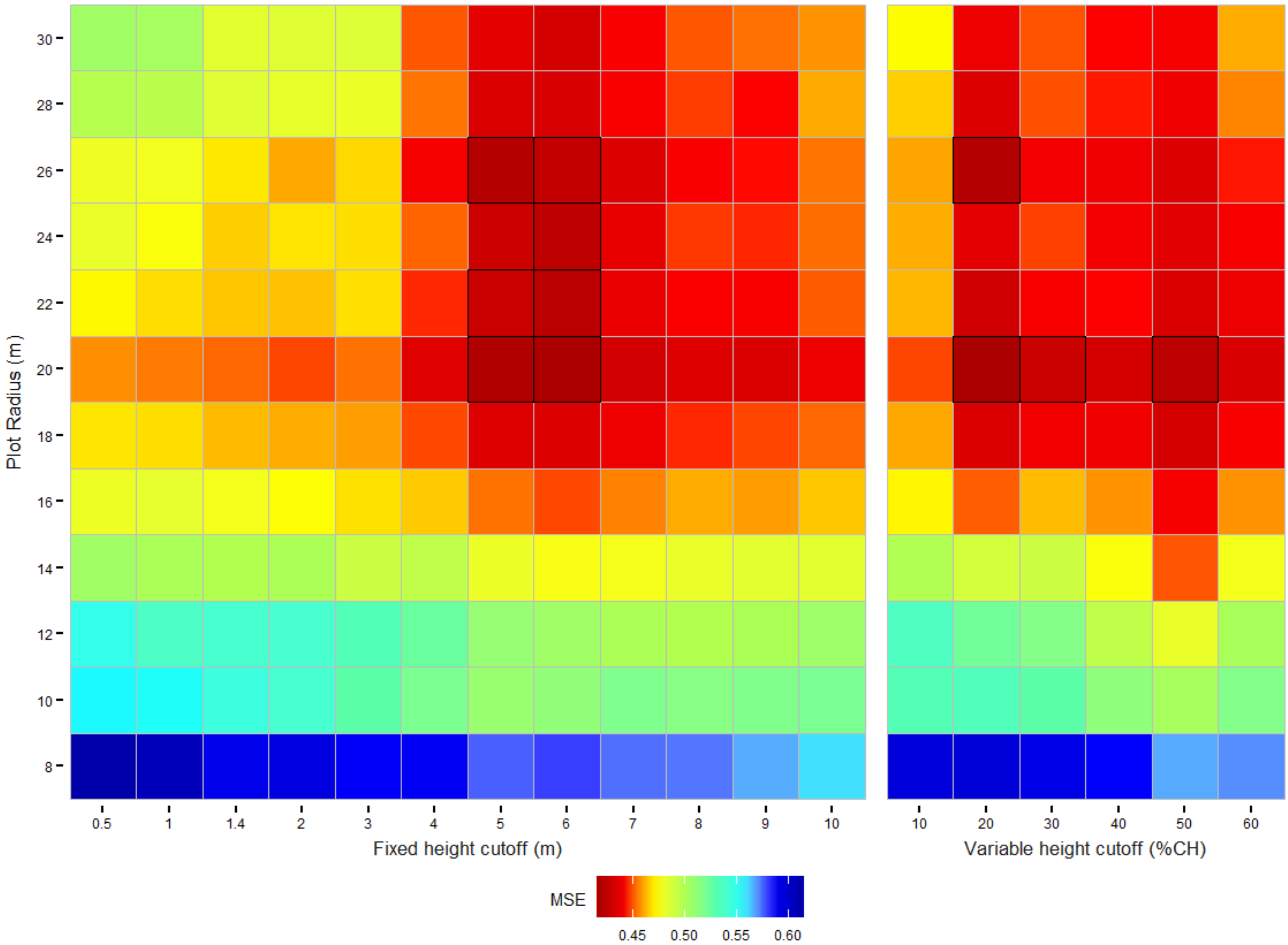

Figure 1: Results from elastic-net regression models of LAI from LiDAR metrics constructed using fixed plot radii. Ratio metrics were constructed using a range of fixed and variable height thresholds calculated as a percentage of maximum canopy height (CH). The top 5% of models by mean squared error are outlined in black and were nearly tied (MSE 0.41 – 0.43).

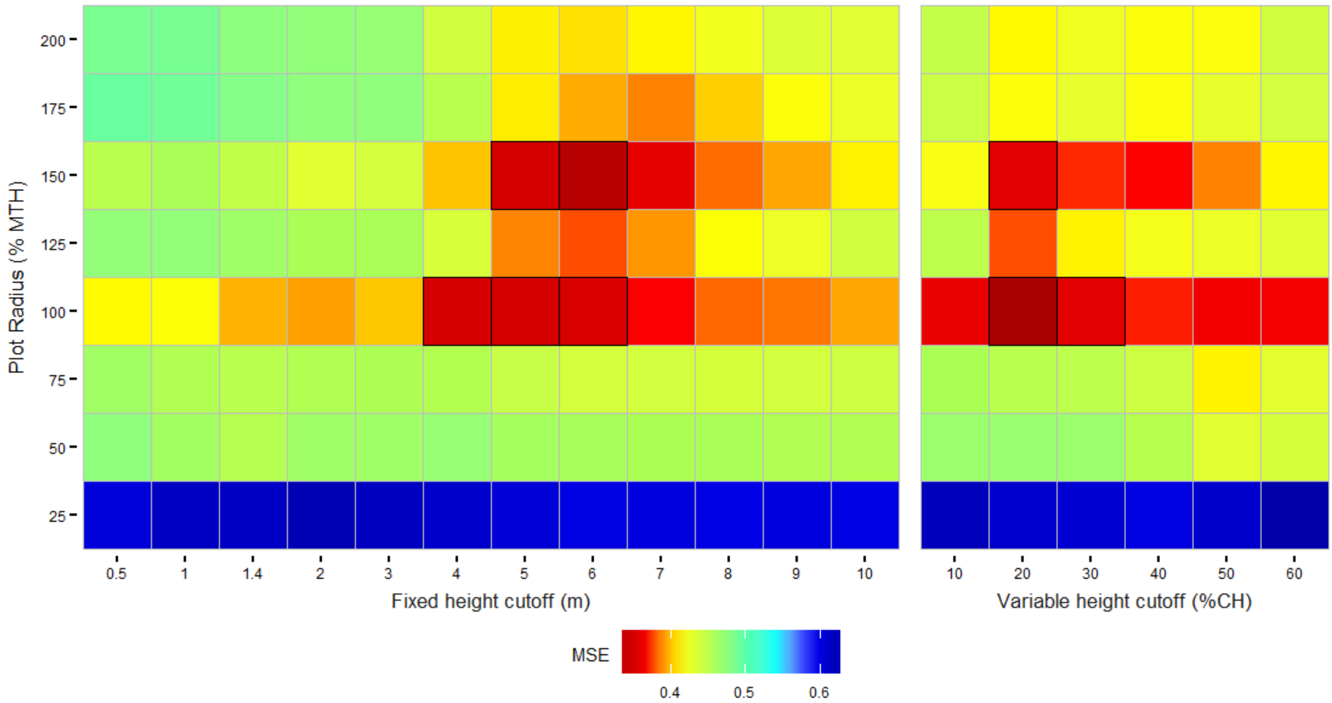

Figure 2: Results from elastic-net regression models of LAI from LiDAR metrics constructed using plot radii based on mean top height (MTH) within plots. Ratio metrics were constructed using a range of fixed heights, and variable height thresholds calculated as a percentage of maximum canopy height (CH). Models outlined in black are with 5% of the top model, with MSE ranging from 0.32 – 0.35.

Metric importance

Although Random Forests did not provide improved model performance, the variable importance scores provided additional insight into the relative importance of candidate metrics. The metrics with the highest importance scores broadly agreed with the metrics included in the best elastic-net regression models. Table 1 shows a summary of the top five metrics which were important in the best fixed and variable radius models, and which were present in nearly all top 5% models. It is noteworthy that four of the five metrics are classed as ratio metrics, and the only height metric (skew) measured return height distributional properties. Indeed, several other distributional metrics such as L-moments and measures of kurtosis and variance were occasionally present in some of the better models. Complex metrics were largely absent from the top models.

Table 1: Key LiDAR metrics selected by elastic-net, and scored highly by Random Forests importance measures, in the top 5% of fixed and variable radius models. FIRST and LAST denote return types. First and last of many are returns marked as having additional echoes from the same pulse. Subscripts A and B denote returns above or below a specified height threshold. *AVM denotes returns above the vegetation *mean, where the vegetation mean was computed as the mean height of all returns above the specified height threshold.*

Metric	Class	Source	Definition
% First above veg mean	Ratio metric	Novel	$\frac{FIRST_{AVM^*}}{FIRST}$
Last cover index	Ratio metric	Korhonen et al.[10]	$\frac{LAST_A}{LAST}$
Solberg's cover index	Ratio metric	Solberg et al[9,10]	$\frac{SINGLE_B + 0.5(FIRST\ of\ many_B + LAST\ of\ many_B)}{SINGLE_A +\ 0.5\ (FIRST\ of\ many_A + LAST\ of\ many_A)}$
Morsdorf's LAI proxy	Ratio metric	Morsdorf et al.[11]	$\frac{FIRST_A}{LAST_A}$
Skew	Height metric	N/A	Skewness of all return heights

Metrics selection and importance were strongly impacted by the choice of LiDAR plot radius and height threshold. At smaller fixed and variable plot radii metrics were all fairly uniformly poor predictors. However, as the combination of radius and height threshold approached the optimum values shown in Figures 1-2 a clear pattern emerged. In these regions ratio metrics came to dominate the list of metrics selected, and achieved the highest variable importance scores. In many cases, the height metrics included in the top 5% of models were of marginal importance and model performance would have been largely unaffected with these metrics excluded.

Discussion and recommendations

Our findings show that LiDAR based estimates of LAI can be achieved in New Zealand's radiata pine plantation forests. The error levels from the best models were competitive with many overseas examples[6,11,14]. The 'default' approach for our study would have been to construct ratio metrics using a height threshold equal to instrument height (1.4 m in our case), to extract LiDAR data at the inventory radius (14 m), and to include only the standard set of (predominantly) height metrics. Our results showed that these settings were poorly suited to our forest type, and these choices produced some of the poorest models examined.

Our results broadly agreed with proposed theoretical links between optical LAI measurements and Airborne LiDAR. A variable radius may be desirable because optical theory relates the instrument's view distance to maximum canopy height, and Solberg at al.[9] found this approach to produce superior estimates of LAI from LiDAR. The view distance of the LAI-2200C is approximately equal to canopy height, depending on canopy density[7]. Interestingly, many of the best variable radius models were found at 100% of canopy height. Zhao and Popescu[6] reasoned that ratio metrics that reflect the penetration of laser pulses through the canopy should be strong predictors of LAI, which is in turn strongly related to light extinction through the canopy. Our findings support this link, with ratio metrics forming the majority of important metrics. While there is no established theory behind the optimum height threshold for ratio metrics, our results agree with those of Zhao and Popescu[6] who found values well above instrument height to be best. Higher thresholds may strengthen the link between canopy LAI and LiDAR metrics by more accurately excluding lower canopy and understory returns.

The distribution of canopy heights in our data were concentrated between 25 – 35 m. For the majority of our plots 20% of canopy height would be approximately equal to 4 – 5 m. Indeed, 20% of the mean and median heights were equal to 5.8 and 5.9 m respectively. This suggests that where the height distribution is reasonably consistent a fixed height threshold set at approximately 20% of canopy height would suffice. A variable height threshold, while more complex to compute, offers the attractive property of producing useful ratio metrics regardless of the plot height distribution.

Recommendations for LiDAR-LAI

The observed theoretical links do not provide a method for direct LiDAR-LAI estimation, and models developed for radiata pine within New Zealand, as for most forest types, are likely to be campaign specific. Based on our study results, we propose the following set of guidelines for developing models of LAI from LiDAR data:

(1) LiDAR data extracted from plot radii coincident with instrument view distance are likely to produce better estimates of LAI.
(2) Where a variable plot radius is not practical, we recommend trialling data extracted using a fixed radius of between 20 – 26 m.
(3) Construction and inclusion of ratio metrics is strongly recommended. Distributional height metrics (e.g. skew, kurtosis, L-moments) should be preferred when selecting from standard height metrics available from many software tools.
(4) Determining the optimum height threshold for ratio metrics greatly improves the predictive power of these metrics. If technically feasible, a variable height threshold of 20% is recommended. Otherwise, we recommend trialling fixed height thresholds a few metres either side of 20% * mean or median canopy height in the plots.
(5) Modelling efforts should include controls for overfitting. We also observed strong correlations between candidate metrics, and this should be accounted for when developing LiDAR-LAI models.

Conclusion

LiDAR based estimation of LAI can be accomplished in New Zealand's radiata pine forests with acceptable levels of prediction error. However, the optimum parameters for metric construction differ from those used elsewhere, and this must be accounted for. Future work will examine the ability of predictive models to detect fertility gradients and deficiencies after controlling for environmental factors. There is also scope to assess the use of cost-effective equipment such as digital hemispherical canopy photography for obtaining the ground LAI measurements required to calibrate LiDAR-LAI models.

Acknowledgements

This research was supported by the ‘Growing Confidence in Forestry’s Future’ research programme (C04X1306) which is jointly funded by the New Zealand Ministry of Business, Innovation and Employment and the New Zealand Forest Growers Levy Trust. Additional support came from the Canterbury Doctoral Scholarship administered by the University of Canterbury. We are grateful to Timberlands Ltd. for supplying the plot information and LiDAR data. The authors also gratefully acknowledge the diligent and professional contribution of all field teams involved.

References

1. Chen, J. M. Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agric. For. Meteorol.* **80,** 135–163 (1996).
2. Jonckheere, I. *et al.* Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agric. For. Meteorol.* **121,** 19–35 (2004).
3. Bréda, N. J. J. in *Encyclopedia of Ecology* (eds. Editors-in-Chief: Sven Erik Jorgensen & Brian Fath) 2148–2154 (Academic Press, 2008).
4. Stenberg, P., Rautiainen, M., Manninen, T., Voipio, P. & Smolander, H. Reduced simple ratio better than NDVI for estimating LAI in Finnish pine and spruce stands. *Silva Fenn.* **38,** 3–14 (2004).
5. Kross, A., McNairn, H., Lapen, D., Sunohara, M. & Champagne, C. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *Int. J. Appl. Earth Obs. Geoinformation* **34,** 235–248 (2015).
6. Zhao, K. & Popescu, S. Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. *Remote Sens. Environ.* **113,** 1628–1645 (2009).
7. LI-COR. LAI-2200C Plant Canopy Analyzer: Instruction Manual. (2015).
8. Pearse, G. D., Watt, M. S. & Morgenroth, J. Comparison of optical LAI measurements under diffuse and clear skies after correcting for scattered radiation. *Agric. For. Meteorol.* **221,** 61–70 (2016).
9. Solberg, S. *et al.* Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sens. Environ.* **113,** 2317–2327 (2009).
10. Korhonen, L., Korpela, I., Heiskanen, J. & Maltamo, M. Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sens. Environ.* **115,** 1065–1080 (2011).
11. Morsdorf, F., Kötz, B., Meier, E., Itten, K. I. & Allgöwer, B. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sens. Environ.* **104,** 50–61 (2006).
12. Pope, G. & Treitz, P. Leaf Area Index (LAI) Estimation in Boreal Mixedwood Forest of Ontario, Canada Using Light Detection and Ranging (LiDAR) and WorldView-2 Imagery. *Remote Sens.* **5,** 5040–5063 (2013).
13. Van Ewijk, K. Estimating Forest Structure from LiDAR and High Spatial Resolution Imagery for the Prediction of Succession and Species Composition. (2015).
14. Peduzzi, A., Wynne, R. H., Fox, T. R., Nelson, R. F. & Thomas, V. A. Estimating leaf area index in intensively managed pine plantations using airborne laser scanner data. *For. Ecol. Manag.* **270,** 54–65 (2012).
15. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67,** 301–320 (2005).
16. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33,** 1–22 (2010).
17. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2015).
18. Breiman, L. Random Forests. *Mach. Learn.* **45,** 5–32 (2001).
19. Bias in random forest variable importance measures: Illustrations, sources and a solution | BMC Bioinformatics | Full Text. Available at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-25. (Accessed: 12th June 2016)
20. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinformatics* **9,** 307 (2008).