PO Box 1127
Rotorua 3040
Ph: + 64 7 921 1883
Fax: + 64 7 921 1020
Email: info@ffr.co.nz
Web: www.ffr.co.nz

Theme: Radiata Management

Task No: F10409
Milestone Number: 4.09.12
Report No. : R070

Using LiDAR based Regression Estimation in New Zealand Forestry Inventory

Authors:
H Marshall, J Dash, B Rawley, T Adams

Research Provider:
Interpine, Scion

This document is Confidential
to FFR Members

Date: February 2012

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This paper demonstrates how through the use of regression estimation, airborne LiDAR can be successfully implemented in forest inventory. Using LiDAR in this manner means that the number of ground plots can be reduced while maintaining the same level of precision. This technique relies on knowing the population size and mean of the auxiliary variable to be used in the regression model, combined with having sufficient correlation between the variable of interest and the auxiliary variable.

In this study the variable of interest was total recoverable volume (TRV) and the auxiliary variable was a LiDAR height metric which is strongly related to the TRV. Three different LiDAR height metrics were used, the 30th height percentile, 95th height percentile and 95th-30th height percentile, to investigate the impact of the strength of the relationship on the increase in inventory precision achieved. The technique was trialled on a 210.2 hectare forest in the Eastern Bay of Plenty, New Zealand. A traditional ground survey was installed consisting of 173 plots. Airborne LiDAR was acquired across the whole estate at approximately 2 pulses per square metre.

The results of the case study indicate that by following a regression estimation approach using LiDAR metrics, forest inventory could contribute approximately $15 per hectare towards the cost of acquiring LiDAR without increasing the inventory budget. This is a substantial contribution toward the current cost of acquiring LiDAR. If the cost of the acquisition falls below this level, the difference would represent a saving in inventory costs. In addition this approach could be easily be implemented in the majority of New Zealand’s current inventory systems.

Large scale acquisition of LiDAR data is a significant investment for many forest owners in New Zealand. The decision to carry out LiDAR acquisition projects requires a strong business case to be presented. The results of this research suggest that integrating LiDAR into forestry inventory has an important role to play in the development of a positive business case for LiDAR acquisition.

INTRODUCTION

Regression and ratio sampling, like stratification, were developed to increase the precision and efficiency of sampling. This is achieved through the use of an auxiliary variable that is measured on each sample unit in addition to the variable of interest [4]. Regression or ratio sampling can be used when the population mean or total for the auxiliary variable is known without error. This is different from double sampling where the true population mean of the auxiliary variable is not known [6] and hence needs to be estimated from a sample. To utilise regression or ratio sampling successfully, a strong relationship between the auxiliary variable and the variable of interest is required. Ratio sampling can be used instead of regression estimation where this relationship has a y intercept of zero [2]. The increase in precision gained through using regression estimation is proportional to the correlation between the variable of interest and the auxiliary variable. Impressive improvements in precision are possible with regression estimation when the correlation is close to 1 [2].

The use of Light detection and ranging (LiDAR) in forestry has been studied since about 1978. Airborne LiDAR has been developed into a tool that can produce direct and indirect measurement of trees and forests. There are many examples of the use of LiDAR to derive timber volume estimates [11], [10], [13]. A number of the characteristics of LiDAR data allow generation of additional information suitable for stand and landscape-level management. Numerous studies, both international and within New Zealand, have shown that LiDAR metrics can be related to both total standing and recoverable volume [15], [11]. In most cases when aerial LiDAR is obtained across a forest, wall-to-wall total coverage is obtained. This means that total population statistics for the LiDAR metrics can be determined without sampling, making them suitable for use in a regression estimation approach.

The objective of this project is to investigate the costs and benefits of using airborne LiDAR to supplement traditional forest inventory as carried out in New Zealand. This paper focuses solely on the use of LiDAR metrics as an auxiliary variable in regression estimation sampling. A key goal of this research is to determine the cost savings that can be achieved through reducing the number of ground plots required while still achieving a desired level of precision on the total recoverable volume (TRV) estimates.

METHODOLOGY

Data Collection and Preliminary Analysis

The data for this study were obtained from a forest located in the Eastern Bay of Plenty region of New Zealand's North Island on steep, broken terrain with an elevation range of approximately 150 to 300 metres above sea level. The tree crop is first rotation *Pinus radiata* planted onto cleared native forest.

A pre-harvest inventory based on systematic random sampling was established using existing stand/harvesting boundaries as the sampling frame. In total 172 plots were installed, using a single grid with a randomised start point across all stands. These plots were used to sample a total area of 210.2 hectares. The plot size varied from population to population to achieve a target of approximately 20 trees per plot; only one plot size was used per stratum. The exact locations of plot centres were recorded using a survey grade global positioning system (GPS) (Trimble Pro XT) capable of post differential correction. Within each plot all trees were measured for diameter with a subset of heights being measured. Stem quality descriptions were also recorded using the method described in PlotSafe Overlapping Feature Cruising Forest Inventory Procedures [18] and the RAD05 feature cruising domain. Data collection took place in the early part of 2011. Table 1 gives a summary of the inventory design.

Table 1. A Summary of the Inventory Design

Population	Area (ha)	Number of Plots	Plot Size (ha)
A	57.1	45	0.09
B	28.4	24	0.06
C	26.9	26	0.09
D	22.4	17	0.09
E	51.5	38	0.07
F	23.9	23	0.08

Figure 1 shows a map of the population boundaries overlaid on a digital elevation model of the study area. Each population has only one stratum. All stands within the forest were established in 1986 in *Pinus radiata*. The red points on the map show the location of the plots within the existing stand boundaries. Several plots were removed from the analysis, as accurate GPS locations could not be obtained.

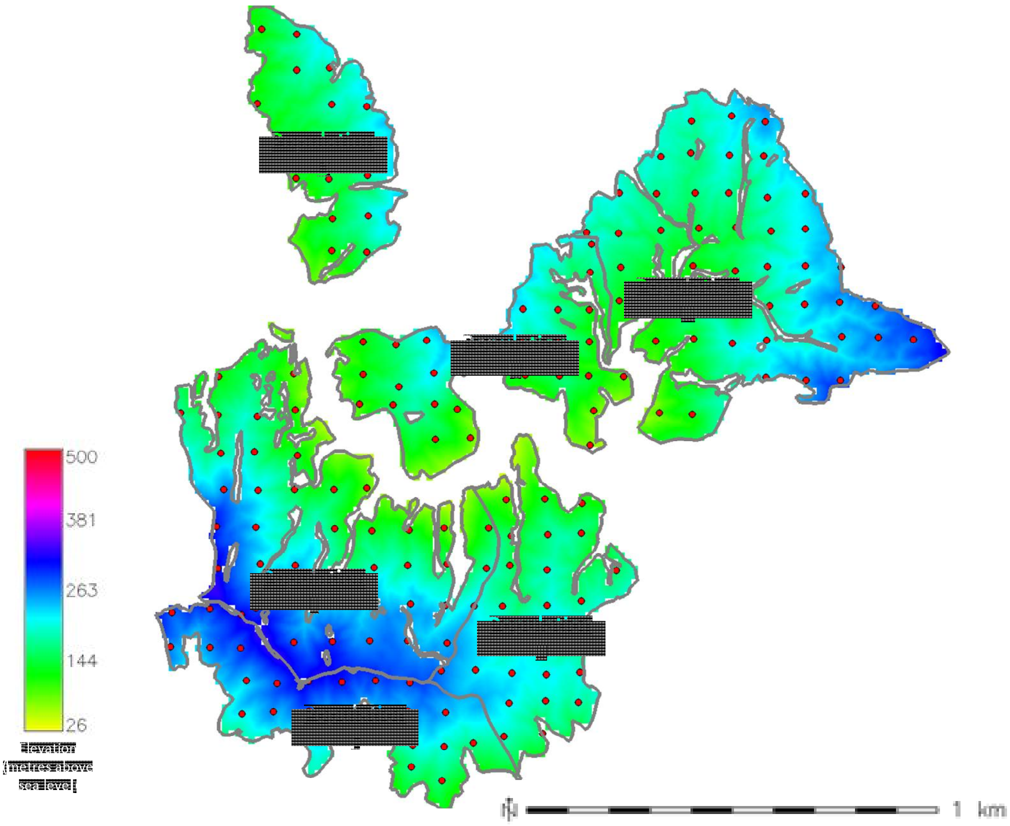

Figure 1. A DEM of the forest area showing the location of the plots.

The inventory data were collected using PlotSafe data capture tool [17] and analysed using the YTGEN yield analysis software product [16]. Volume and taper functions known as 182, developed for *Pinus radiata* managed under a direct sawlog regime in New Zealand [7], were used. A set of cutting instructions based on the Ministry of Agriculture and Forestry (MAF) generic log grades was used to determine the product mix for each plot [9]. The plot level statistics such as total standing volume, total recoverable volume, and basal area were extracted from the YTGEN output.

The LiDAR was acquired between 24 May and 1 June 2011 using an Optech ALTM 3100EA LiDAR system. The flying altitude, overlap and scanner settings were selected to create a dataset with a minimum 2 points per square metre pulse density in open ground. The LiDAR was processed using Optech LiDAR Mapping Suite, with the automated LiDAR point classification carried out in the TerraSolid software suite. Comprehensive manual editing of the LiDAR points was carried out to improve the quality of the ground point classification.

The LiDAR plot metrics used in this study are given in Table 2. These metrics were chosen as they have been shown in previous studies [20], [15] to be well correlated to total standing volume. The plot level metrics were calculated by extracting the point cloud associated with each individual plot by using the high grade GPS plot centre coordinates. The plot level metrics were calculated using the Cloudmetrics function of the FUSION 2.90 software product [8]. These metrics were also calculated over the whole landscape using the Gridmetics function. The metrics were calculated at a resolution that gave an area for each grid square that was equivalent to the circular ground plots. The FUSION user manual [8] gives a full explanation of these functions and metrics. All LiDAR returns within 0.5 m of the ground were excluded from analysis to remove the effects of under storey vegetation.

Table 2. LiDAR Metrics Used

Metric Name	Description
H30	30 percentile value for cell
H95	95 percentile value for cell
H95-H30	95 percentile value for cell - 30 percentile value for cell

Within the ground inventory, five of the plots were measured as edge plots using the mirage technique [14]. In this inventory all the mirage plots were measured as half mirages, meaning that every tree in the plot centre was placed on the forest boundary and each tree was counted twice to calculate the per hectare volume of the plot. To account for the mirage plots when calculating the LiDAR metrics, only the portion of the plot LiDAR cloud within the stand boundary was included.

Regression and Simple Sampling

The regression estimator was calculated using the method described by Avery and Burkhart [2]. In this study the LiDAR metrics given in Table 2 were used as the auxiliary variables; each of these variables was regressed against TRV in separate regression models. The population size and mean for the auxiliary variable were calculated from the LiDAR metrics raster. The population size is used only in the finite population correction factor (1-n/N). When the sampling fraction is small, which is true in this case, the finite population correction factor has very little effect and can be omitted. The variable of interest in this study was TRV. This is defined as the volume of the recoverable logs and does not include the volume of the non-recoverable products such as log-making waste and breakage volume. An estimate of the mean population value of (TRV) is calculated, as well as the standard error for this estimate, from which the 95th percentile confidence interval is calculated. In New Zealand the precision of a forest inventory is most often expressed as probable limit of error (PLE). PLE is defined as half the width of the estimated 95% confidence interval expressed as a percentage of the estimated mean [5].

As a benchmark against which to compare LiDAR-based regression estimation, the plot network was analysed as a simple random sample. Once again the estimated mean TRV for each population was calculated as well as the precision statistics.

To provide a fair comparison between the two inventory approaches for the cost benefit analysis, the sample size required to achieve a 10% PLE was determined. A PLE of 10% on TRV is often used in New Zealand forestry as a precision target to determine an appropriate sample size.

Cost Benefit Analysis

The cost benefit analysis provides a comparison of the total cost of the inventory with and without the use of LiDAR. The cost of LiDAR is variable in Australasia; costs in Australia can be as low as a couple of dollars per hectare to over $50 per hectare in New Zealand for high resolution LiDAR [1]. On average a reasonably sizeable area (greater than 500 hectares) can be scanned at an intensity of 2-3 pulse per m^3 at a cost of $16 per hectare. This cost includes acquiring and processing the LiDAR data to deliver products such as digital elevation models and classified point cloud files. It does not include the cost of calculating the LiDAR metrics for the plots and stands. For the purpose of this paper this cost has been estimated at an extra $2 per hectare over the traditional inventory processing cost.

As part of this analysis the LiDAR cost will be varied to determine a breakeven point where the total inventory cost is the same for both traditional inventory without LiDAR and inventory supplemented with LiDAR. This breakeven cost can also be thought of as the amount of money that forest inventory can contribute to the overall cost of acquiring LiDAR data.

The summarised field costs (Table 3) assume that the field crew will stay away from their base, and includes the cost of accommodation and an away allowance. The cost covers a two-man field crew, truck and the required equipment including the high grade GPS.

Table 3. Summary of field based costs

Item	Cost ($/day)
Crew Day Rate	800
Accommodation	80
Allowance	90

Another input into the cost benefit analysis is the number of plots that a crew can measure per day (plot rate). This rate is highly dependent on the type of terrain in which the field crew is working. The budget cost rate for the inventory used in this study was 6.4 plots per day. A normal plot rate can vary between 6 and 9 plots per day for a typical pre-harvest inventory.

RESULTS

The relationships between the metrics in Table 2 and TRV are shown in Figure 2 – 4. The fitted regression lines show that for H30 and H95-H30 there was a moderate correlation between the metrics and TRV with R^2 of 0.42 (p value < 0.001) and 0.39 (p value < 0.001) respectively. Figure 4 shows that the relationship between H95 is weaker with R^2 of 0.006 (p value = 0.2932). The intercepts for all the relationships were not zero, meaning that regression estimation must be used rather than ratio sampling (Avery and Burkhart 1994). The regression lines in Figure 2 – 4 are for the combined dataset. The relationships used in the analysis are fitted for each individual population separately.

Despite there being a poor correlation between H95 and TRV, H95 has been included in the analysis to demonstrate that precision cannot be improved through the use of an auxiliary variable which has a poor correlation to the variable of interest.

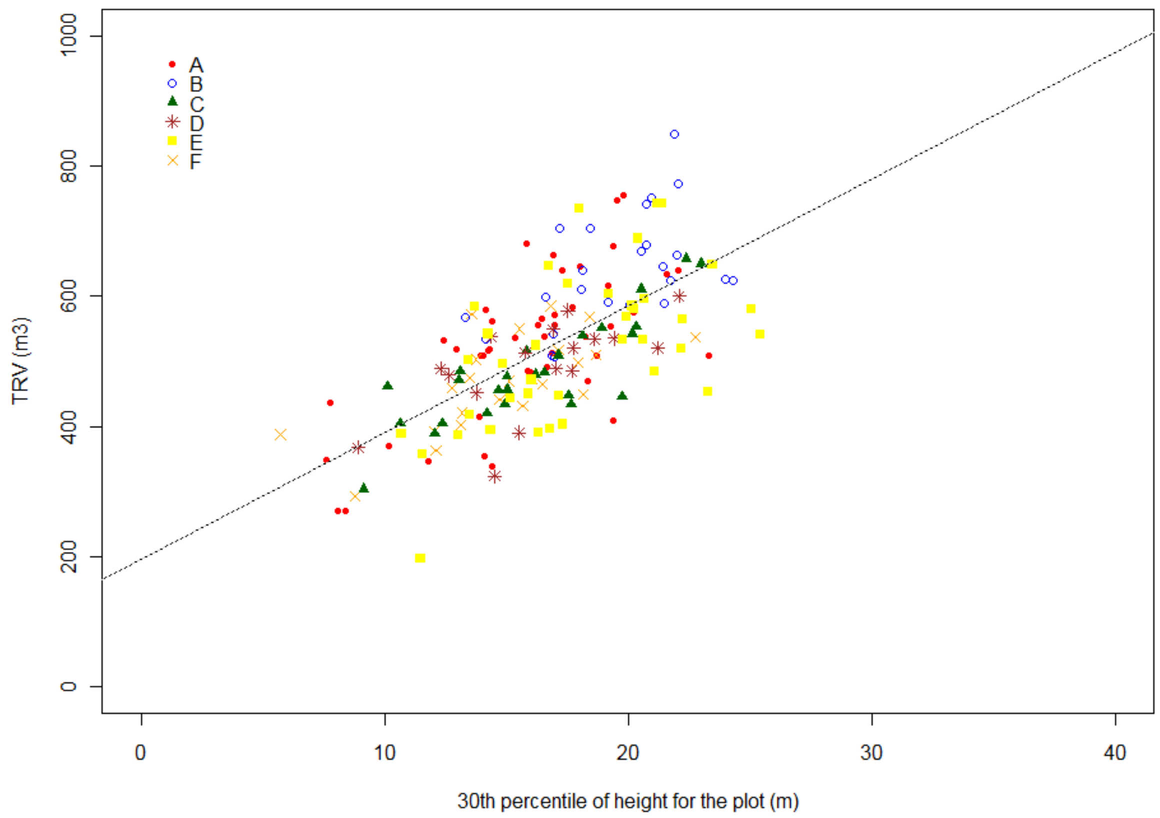

Figure 2: 30th percentile of height for the plot vs TRV

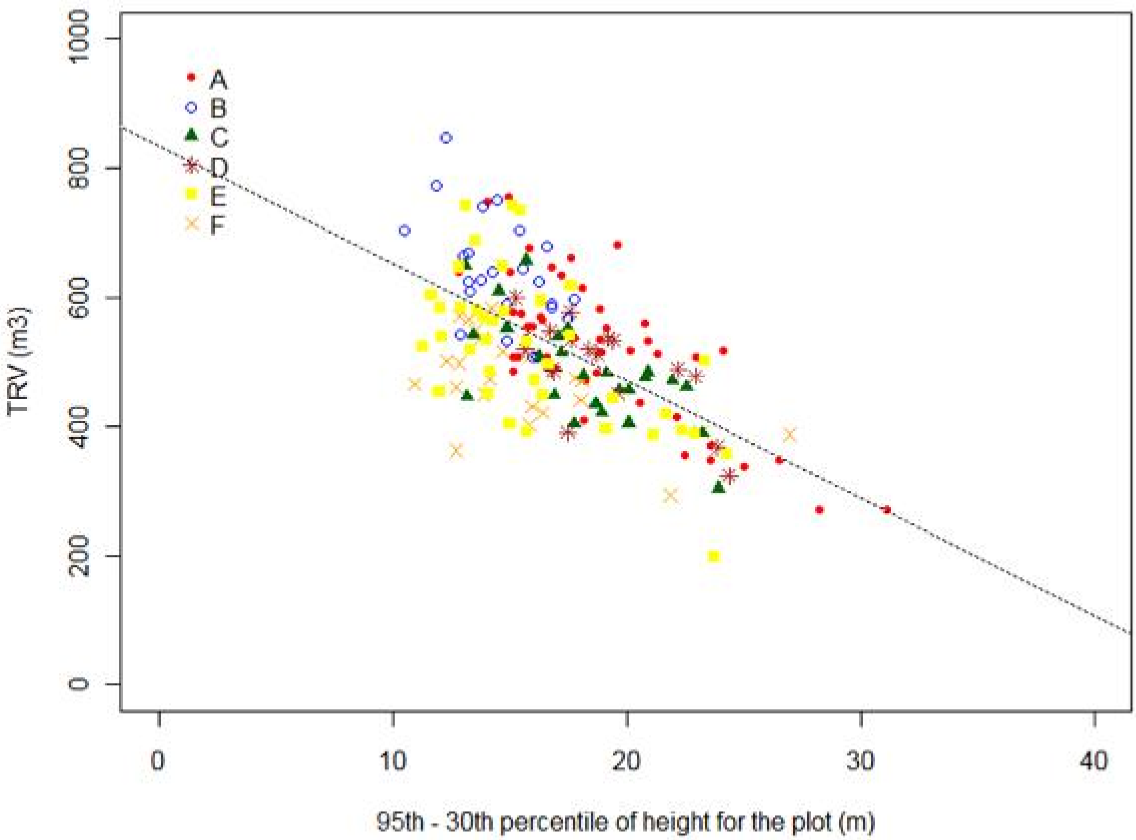

Figure 3: 95th - 30th percentile of height for the plot vs TRV

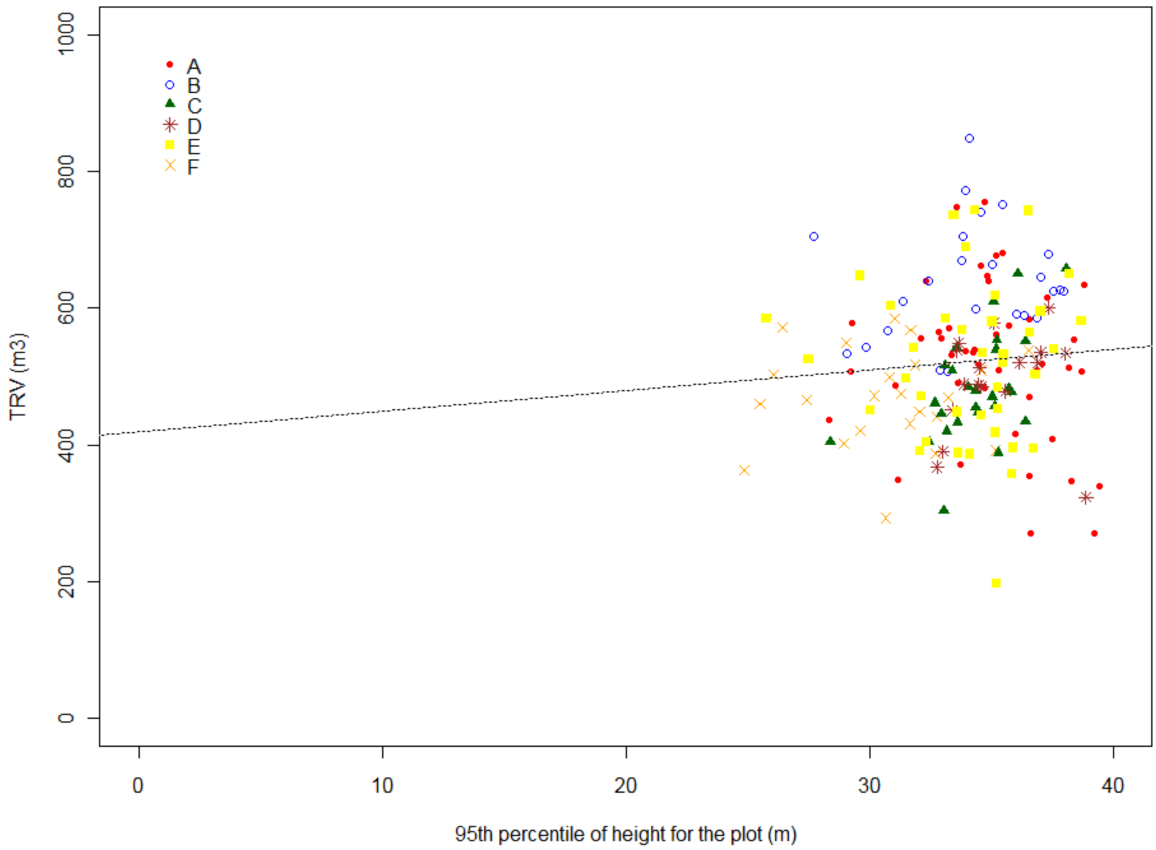

Figure 4: 95th percentile of height for the plot vs TRV

A comparison of the statistics calculated using simple random sample and regression estimation with different auxiliary variables is given in Table 4. The lowest PLE is achieved in three of the six populations using regression estimation utilising the H95-H30 LiDAR metric. For all but one population, regression estimation using either the H30 or H95-H30 provides better precision than simple random sampling. In population B the simple random sample still gives the lowest PLE.

Table 4. Summary Statistics for the different Sampling Methods

Population	Sampling Method	Mean	CI (95%)	PLE
A	Simple Random	522.96	32.95	6.30
	Regression (H30)	522.07	24.35	4.66
	Regression (H95-H30)	512.90	21.58	4.21
	Regression (H95)	519.63	33.16	6.38
B	Simple Random	638.59	35.08	5.49
	Regression (H30)	615.05	35.13	5.71
	Regression (H95-H30)	601.07	40.41	6.72
	Regression (H95)	638.91	35.46	5.55
C	Simple Random	483.52	30.73	6.35
	Regression (H30)	461.80	18.26	3.95
	Regression (H95-H30)	459.25	24.88	5.42
	Regression (H95)	486.41	30.73	5.43
D	Simple Random	491.70	36.36	7.40
	Regression (H30)	475.13	30.57	6.43
	Regression (H95-H30)	472.49	29.81	6.31
	Regression (H95)	491.24	37.44	7.62
E	Simple Random	519.82	37.66	7.24
	Regression (H30)	518.35	30.99	5.98
	Regression (H95-H30)	520.81	29.93	5.75
	Regression (H95)	519.93	38.25	7.36
F	Simple Random	467.67	30.10	6.44
	Regression (H30)	458.22	23.27	5.08
	Regression (H95-H30)	448.86	27.81	6.20
	Regression (H95)	468.96	32.70	6.97

Figure 5 shows the number of plots that would be required for the simple random sampling strategy to equal the PLE of the regression (H30) sampling strategy using the actual plots installed. It shows that a large number of plots are required to achieve the very low PLE achieved through the use of regression estimation. It is unlikely that a forest manager would install these plot numbers to achieve PLE of 4-5%. A total of 248 plots would be required to match the precision level achieved by the regression estimation approach, which is an additional 65 plots added to the already intensive inventory.

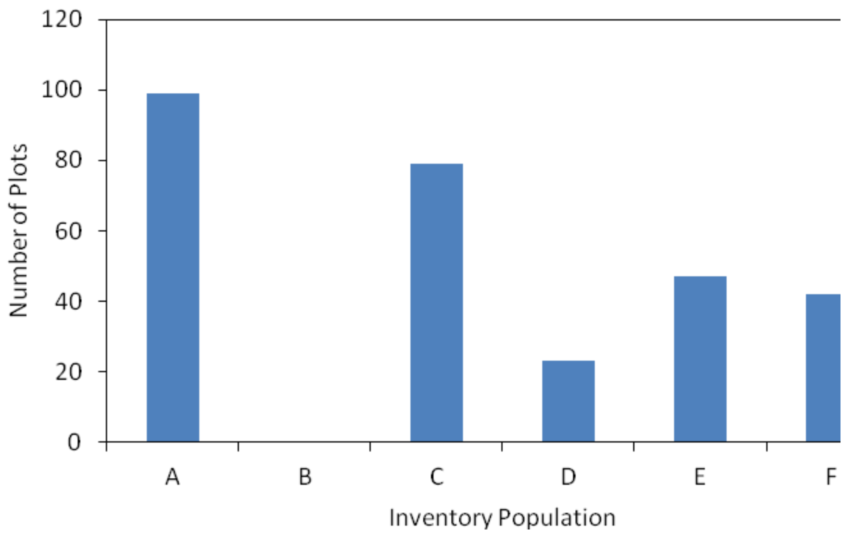

Figure 5. The number of plots by population required under a simple random sampling strategy to match the PLE of the regression estimation strategy using LiDAR.

The number of plots required to achieve a 10% PLE benchmark is given in Figure 6. To carry out a complete inventory of this example forest would require 91 plots to achieve at least a 10% PLE on each inventory population. This compares to only 66 plots when using regression estimation using the H30 metrics. That equates to a 27% reduction in plots. The importance of the correlation between the auxiliary variable and variable of interest on the number of plots required is evident from the H95 regression estimation approach which would require 90 plots, only one less than the simple random sample.

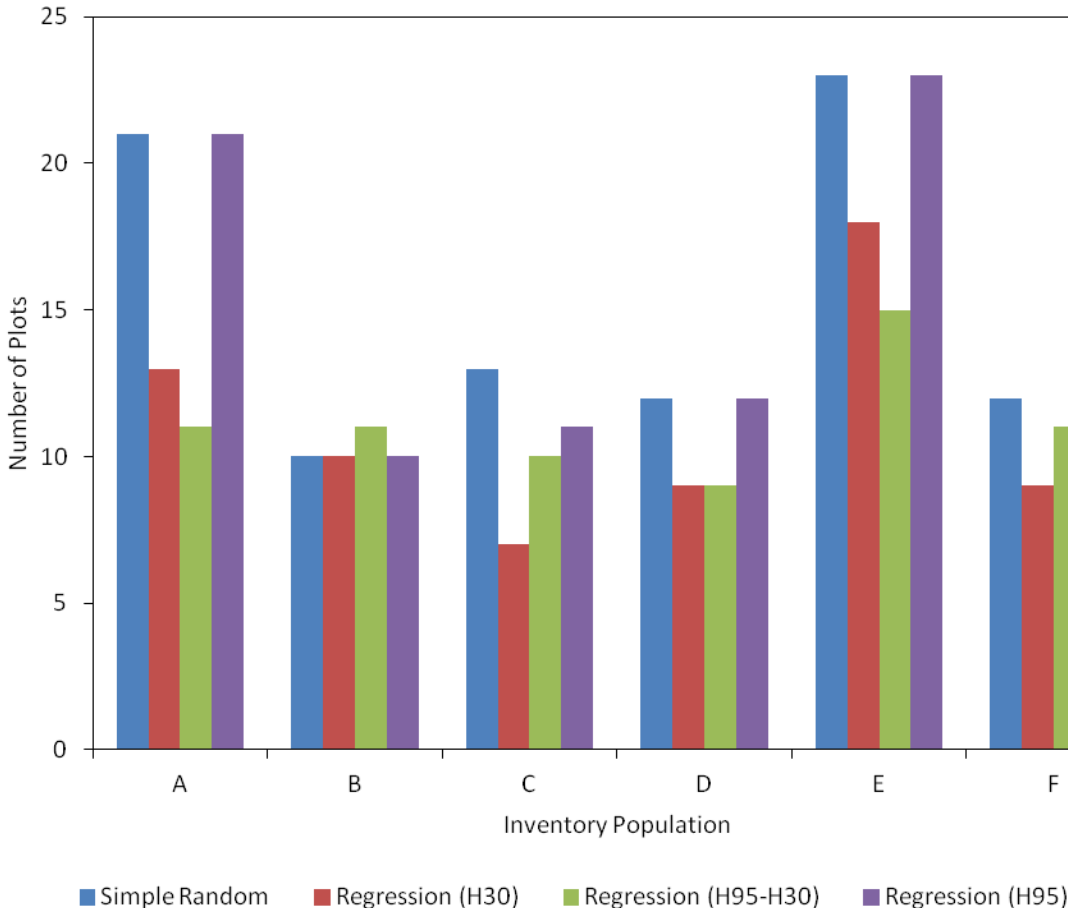

Figure 6. Number of Plots required to achieve 10% PLE under the different Sampling Strategies

Figure 7 shows the total inventory cost curves for the four different sampling strategies to achieve a 10% PLE. Where these cost curves intersect with the line representing the cost of the simple random sampling approach can be thought of as the breakeven point. Beyond this point the regression estimation approaches become more expensive than the simple random sampling approach. The LiDAR cost in Figure 7 is the cost of acquiring the LiDAR from a LiDAR provider, and does not include any additional inventory analysis cost that might be occurred by utilising LiDAR in the manner described in this report. The potential contribution diminishes as the number of plots that can be achieved by a crew per day increases. Increasing the daily plot rate to 9 would reduce the breakeven to $11 per hectare.

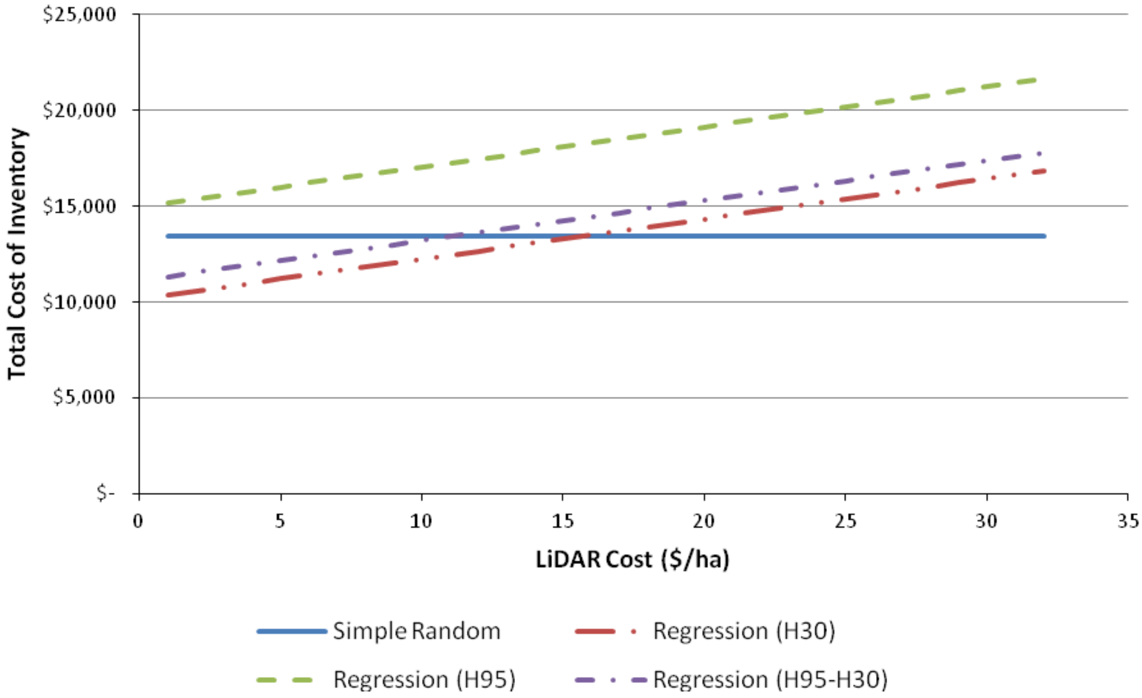

Figure 7. Total Inventory Cost Curves for the different inventory strategies.

The breakeven analysis shows that for this case study, using the H30 metric that utilising LiDAR for forest inventory could contribute between $15 and $16 per hectare towards the cost of acquiring LiDAR. The breakeven cost for the H30 metric is the point where its cost line (red) crosses the cost line for the simple random sample; at that LiDAR cost the total inventory cost under both strategies is equal. If the LiDAR can be acquired for any cost less than $15 per hectare, then using regression estimation using LiDAR metrics will save forest manager money compared to the traditional simple random (systematic) sampling approach, as fewer ground plots will be required.

DISCUSSION

In New Zealand forestry the acquisition of LiDAR is often viewed as being prohibitively expensive. This is despite LiDAR-derived products such as digital terrain models being of superior quality to those currently being used. One reason for the lack of uptake of LiDAR as an operational tool may be that none of the many potential uses is seen as justifying the cost of LiDAR alone. This study presents a strong case for the use of LiDAR as an auxiliary variable in forest inventory. In fact this case study suggests that this use could almost cover the cost of acquiring LiDAR through reducing the required number of ground plots. This is an interesting result as using LiDAR, particularly low point density as in this study, for forest inventory is often viewed as a secondary application of the technology, the primary purpose being the development of accurate digital terrain models for forest engineering purposes. The study suggests that inventory funding could contribute around $15 per hectare on LiDAR without increasing the costs of a traditional forest inventory. The figure of $15 per hectare relates specifically to this case study and caution should be used when applying these findings to other situations.

Integrating LiDAR data with forest inventory in this manner also has the advantage that the overall effect on the way forest inventory is implemented both in the field and office would be minimal. In addition to standard inventory practice, the implementation of the methodology descried in this paper would require only LiDAR acquisition, high grade GPS fixes for plot centres, and some relatively straightforward LiDAR analysis.

Due to the steep terrain of the case study forest, carrying out ground based inventory is expensive. It can also be challenging for inventory providers with inventory staff often being less willing to work on very steep slopes. A plot rate of 6.4 as used in this study is quite low compared to a rate of 6-9 plots per day that can be achieved on flatter terrain. However this study has shown that increasing the plot rate does not have a large impact on the breakeven LiDAR cost. It can also be inferred that the same will be true for reducing the daily cost of an inventory crew.

The improvement in precision achieved through using regression estimation is connected to the strength of the correlation between the auxiliary variable and variable of interest. This study investigated a relatively small set of LiDAR metrics that have been shown in past studies to be well correlated with volume [15]. In this paper only a single auxiliary variable has been used. It is likely that a multiple regression model that includes more than one variable would give a better result. In this study LiDAR was collected at a relatively low point density. Collecting LiDAR data at a higher point density may improve the correlation between the LiDAR metrics and total recoverable volume. Any improvement in this correlation may in turn increase the precision of the estimate, meaning that a lower number of plots maybe required to achieve the desired precision. As we have not accounted for any of these factors, the result presented should be treated as a conservative estimate of the value of LiDAR to forest inventory.

The standard formula for the variance of the regression estimator applies to large samples. Empirical studies have shown a tendency to under-estimate variance for small samples [3]. Regression estimator is also not design-unbiased. The design bias decreases as the sample size increases, and is normally much smaller than the sampling error so would not normally be an issue. At some point as the sample size reduces through the use of regression estimator sampling, the design bias may start to become a problem. The above two issues are important and must be accounted for when implementing regression estimation. One must take care in implementing the techniques described in this paper, particular in cases where the sample size calculation suggests that very small samples would be satisfactory.

This paper investigates only how the precision of total recoverable volume estimates are affected by the use of LiDAR as an auxiliary variable. However, obtaining a desired level of precision on individual log grades volume can also often be desirable. There is no reason to suggest that this methodology could not be applied at the log product level. The key would be to find strong relationships between a LiDAR metric or combination of metrics and individual log grade volumes. The use of LiDAR as an auxiliary variable to increase the precision volume estimate at a log grade level is a topic for future research.

The results obtained in this case study use LiDAR data that were collected at a time that was very close to the time of plot measurement.   Although the use of a regression estimator is not dependent on contemporaneous measurement, the correlation between the auxiliary variable derived from LiDAR and the plot measurement is expected to decline as the time between LiDAR acquisition and plot measurement increases. The cost of acquiring LiDAR data is very dependent on economies of scale. In forests with mixed stand ages where the requirement is to measure each stand immediately prior to harvest, a choice must be made between the acquisition of LiDAR data across large areas of mixed age in a single flight, to reduce the cost of acquisition, and acquiring LiDAR data for each stand immediately prior to harvest, to maximise the correlation. The reader should be cautious about assuming that the $15 per hectare contribution of the LiDAR data is still valid in the first situation or that the cost of LiDAR can be minimised in the second. Although results from this paper present only a single case study, there is no reason to suggest that the results obtained would not be applicable to any mid-rotation or pre-harvest inventory.

CONCLUSION

This case study has shown how integrating LiDAR as an auxiliary variable to supplement traditional ground based inventory can contribute up to $15 per hectare towards that cost of flying LiDAR without increasing inventory costs. This contribution towards the LiDAR costs is obtained by allowing the number of plots to be reduced while still maintaining the same level of precision. This contribution is not insignificant, as the current cost of LiDAR at an intensity of 2 pulses per square metre is around $16 per hectare.

Large scale flying of LiDAR is a significant new investment for many forest owners in the New Zealand industry. The decision to carry out LiDAR acquisition projects requires a strong business case to be presented. The results of this research would suggest that integrating LiDAR into forest inventory is cost effective as a stand alone proposition.

REFERENCES

1. Adams T, Brack C, Farrier T, Pont D, Brownlie R. 2011. So you want to use LiDAR? – A guide on how to use LiDAR in forestry. New Zealand Journal of Forestry, 55, 19-23

2. Avery T.E. and Burkhart H.E. (1994) Forest Measurements. 4th Edition. McGraw Hill, Boston. pp 408

3. Cochran, W. G. (1977). Sampling techniques (3rd ed.) New York: John Wiley & Sons.

4. *Freese*, F. **(***1962***).** Elementary Forest Sampling. US Dept. Agriculture Handbook No. 232.

5. Goulding C, Lawrence M (1992) Inventory Practice for managed forests. FRI Bulletin 171. pp 52.

6. Husch B., Beers T.W, Kershaw J.A. Forest Mensuration, 4th Edition, John Wiley & Sons, Inc, New Jersey. pp 443.

7. Katz, A.G., Dunningham, A., Gordon, A. (1980) A compatible volume and taper equation for New Zealand *Pinus radiata* D. Don grown under the direct sawlog regime. FRI Bulletin 67. Forest Research Institute

8. MacGaughy R.J. (2010) FUSION/LDV: Software for LiDAR Data Analysis and Visualization. Pacific Northwest Research Station. Forest Serice. USA. pp 150.

9. Ministry of Agriculture and Forestry (MAF), (2011). Log Grade Specification. http://www.maf.govt.nz/news-resources/statistics-forecasting/forestry/log-grade-specification.aspx. Accessed on 12/12/2011

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

11. Naesset, E. (1997). Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing Of Environment, 61, 246-253

12. Naesset, E. (2004) Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. Scandinavian Journal of Forest Research. 19, 554-557.

13. Parker, R.C.; Evans, D.L. (2004) An application of LIDAR in a double-sample forest inventory. Western Journal of Applied Forestry. 19, 95-101.

14. Schmid-Haas, P. 1982 Stichproben am waldrand. Mitt. Schweiz Anst. Forstl. Versuchswes. 45, 234 -303

15. Watt M.S., Adam T., Marshall H., Pont D., Lee L., Crawley D., Watt P. 2011, Modelling variation in conifer stem volume and velocity from LiDAR metrics. Remote Sensing of Enviroment. Submitted

16. Silmetra Limited, (2011), YTGen User Manual, version 2.9.0

17. Silmetra Limited, (2011), Plotsafe User Manual, version 1.5.0

18. YTGEN User Group, (2007). PlotSafe Overlapping Feature Cruising Forest Inventory Procedures. Interpine Forestry Ltd.

19. https://www.interpine.co.nz/PlotSafe%20and%20YTGEN%20Download%20Files/PlotSafe%20Cruising%20Manuals%20and%20Information/PLOTSAFE%20Overlapping%20Feature%20Cruising%20Forest%20Inventory%20Procedures%20-%20Feb%202007.pdf

20. Kimberley, M O; Beets, P N; 2007. National volume function for estimating stem total volume of radiata pine stands established on fertile sites in New Zealand. New Zealand Journal of Forestry Science 37(3): 355-371