RADIATA MANAGEMENT TECHNICAL NOTE

Number: RTN-012
Date: March 2012

Regression Estimation using LiDAR for Forest Inventory

Summary

This technical note provides a worked example of the use of regression estimation to increase the precision of the total recoverable volume estimate for a forest population through combining LiDAR metrics and field measurements. The technique requires the population mean of an auxiliary variable, in this case a LiDAR metric, and a strong relationship between the auxiliary variable and the variable of interest. Hence the technique requires LiDAR to be acquired continuously across the whole population. By applying this methodology it is possible to reduce the number of ground plots installed, compared to a conventional forest inventory, while maintaining the same level of precision. The required equations to obtain precision estimates described within this paper can easily be implemented into Microsoft Excel™**,** SAS or R etc.

Author: Hamish Marshall, Jonathan Dash – Interpine Forestry

Introduction

In New Zealand, a forest inventory is normally implemented using a network of plots located within the area of interest on a systematic grid with a random start point. Where possible the sampling population is subdivided into stratum in an attempt to increase the precision of the overall inventory. If the individual strata are more homogenous than the sampling population as a whole, then the precision of the inventory can be improved. The network of plots is either analysed as a systematic random sample or, where strata are employed, a stratified random sample.

In forest inventories the most common way to increase the precision of population estimates is to increase the number of sample plots, which is likely to reduce sampling error. There is always a trade-off between the cost of carrying out inventory and the value of the information obtained (Gordon 2005). Higher levels of precision come at an additional cost as each additional plot costs additional time and money. The increasing use of Light distance and ranging (LiDAR) technologies in forestry for forest engineered purposes creates an opportunity to increase the precision of forest inventory without increasing plot numbers. This increase in precision is achieved through using LiDAR metrics as auxiliary variables in a regression estimation approach. Alternatively the methodology can be used to reduce required plot numbers to achieve a target precision level.

The goal of this technical note is to give resource foresters a worked example of how to integrate LiDAR with forest inventory through the use of LiDAR metrics as auxiliary variables in regression estimation. The formulation and working within this technical note are largely based on the regression estimation section of Avery and Burkhart (1994).

Regression Estimation

Regression and ratio estimation, like stratification, were developed to increase the precision and efficiency of sampling. This is achieved by making use of an auxiliary variable that is measured on each sample unit in addition to the variable of interest [3]. Regression or ratio sampling can be used when the population mean or total of the auxiliary variable are known without error. This is different from double sampling where the true population mean of the auxiliary variable is not known [8] and needs to be sampled. To successfully use regression or ratio sampling, a strong relationship between the auxiliary variable and the variable of interest is required. Ratio sampling can only be used over regression sampling where this relationship has a y intercept of zero. The increase in precision gained through using these approaches is related to the correlation between the variable of interest and the auxiliary variable. Impressive gains in precision are possible through regression estimation when the correlation is close to 1 [2].

Worked Example

The following is a worked example of how to calculate the mean total recoverable volume (TRV), 95 percent confidence interval and probable limit of error (PLE) using regression estimation for a forest inventory containing one unstratified population. The example stand has a net stock area of 57.1 hectares of *Pinus radiata* established in 1986. A systematic sample of 45 plots was installed using a grid with a randomised start point when the stand was aged 25.

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

(Figure 1). The recoverable volume per hectare for each plot (Table 1) was calculated using the tree over bark diameter at breast height (dbh), tree heights, and tree volume and taper functions 182 [7].

Table 1. The volume (m^3/ha) and LiDAR metric (30 Height Percentile) for each plot.

Plot Number	y = volume (m^3/ha)	x = LiDAR metric (30 height Percentile)
1	681.40	16.1
2	582.94	17.7
3	555.76	16.4
4	677.49	19.4
5	662.83	16.9
6	508.14	14.1
7	354.01	14.4
8	755.01	19.8
9	532.22	12.3
10	538.68	16.5
11	560.70	14.6
12	537.87	18.1
13	646.20	18.1
14	346.86	12.0
15	338.70	14.7
16	571.32	16.7
17	535.58	15.3
18	270.12	8.4
19	512.58	16.8
20	408.61	19.4
21	370.39	10.1
22	483.31	15.8
23	509.06	14.1
24	640.25	22.1
25	564.98	16.5
26	555.54	8.4
27	508.36	18.7
28	553.93	19.0
29	269.77	8.3
30	507.93	23.2
31	574.72	20.3
32	348.54	7.6
33	435.70	7.8
34	490.99	16.6
35	518.91	12.8
36	485.96	15.8
37	615.54	19.2
38	634.21	21.6
39	516.58	14.3
40	518.97	14.3
41	578.79	14.1
42	640.28	17.2
43	747.84	19.3
44	415.19	13.9
45	470.30	18.6
Means	$\bar{y} = 522.96$	$\bar{x} = 15.71$
Variance	$s_{\bar{y}}^2 = 12700.46$	$s_{\bar{x}}^2 = 14.78$

The plot level LiDAR metrics in Table 1 were calculated by extracting the point cloud associated with each individual plot. This is possible only when the plot centre has been located using a high grade global positioning system (GPS). The plot level metrics were calculated using the Cloudmetrics function of the FUSION 2.90 LiDAR analysis software. The command below is an example of how to use Cloudmetrics function to calculate the metrics for one plot [4].

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

```
cloudmetrics.exe /minht:0.5 plot1.las metrics.csv
```

Ignores all points below 0.05 metre to reduce impact of low ungrowth

Input LiDAR file of the plot in LAS format

Output File for LiDAR metrics

The LiDAR metric was calculated over the entire population using the Gridmetrics function of FUSION 2.90 [4]. The command below is an example of how to create an ArcGIS ascii raster file of the 30 percentile of height metric over the whole stand.

```
gridmetrics /minht:0.5 /ascii /raster:p30 C:\stand.dtm 20 C:\H30 C:\stand.
```

Ignores all points below 0.05 metre to reduce impact of low

Output results as a ArcGIS ascii grid

Metric to be calculated. In t case the 30 hei

The ground surface digital

The raster size this case 20

The raster grid

Input LiDAR file of the stand in LAS format

In this worked example a raster grid cell of 28.3 metres was used so that the area of each grid cell was the same as the ground plot area. For more information on the FUSION commands please refer to the FUSION user manual [4]. GIS tools such as ArcGIS (with the Spatial Analyst add-on) or the free software GRASS GIS (GRASS Development Team 2010) can be used to calculate the mean ($\bar{\mu}_x$) of the population using the H30 raster (Figure 1). The mean for this stand was 15.857 with a population size (number of grid cells) of 1432.

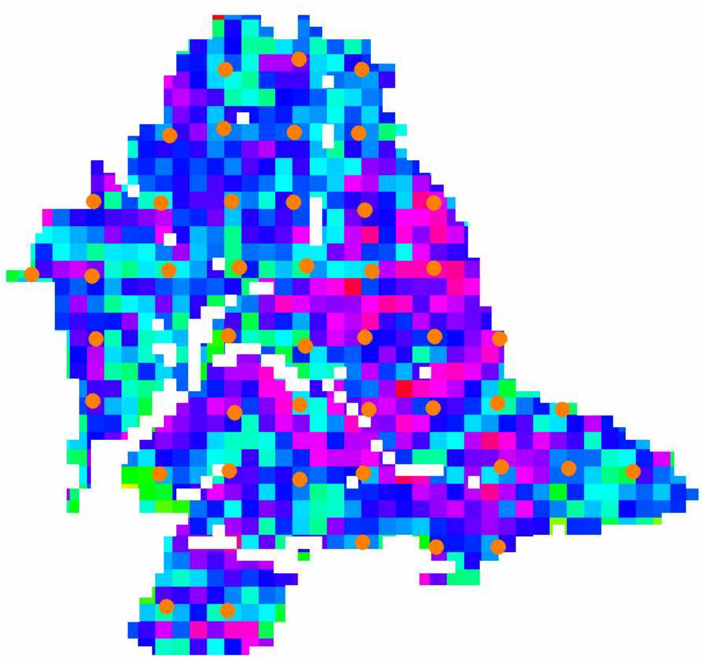

Figure 1. 28.3-metre resolution Raster of H30 Lidar Metric with the plot ground plot location marked.

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

The linear regression estimate of the population mean of $y(\bar{y}_R)$ is

$$\bar{y}_R = \bar{y} + b(\bar{\mu}_x - \bar{x})$$

where $\bar{y}$ is the mean of y and $\bar{x}$ is mean of x from the sampling population, $\bar{\mu}_x$ is the population mean of and x and b is the linear regression slope coefficient. The application of linear regression estimator requires the following assumptions to be fulfilled:

- A linear relationship between x and y must exist;
- The scatter of the y observations is roughly the same throughout the range of the x observations.

The plot of y (total standing volume) versus x (H30 LiDAR metric) (
) shows that these required assumptions are reasonable for these data.

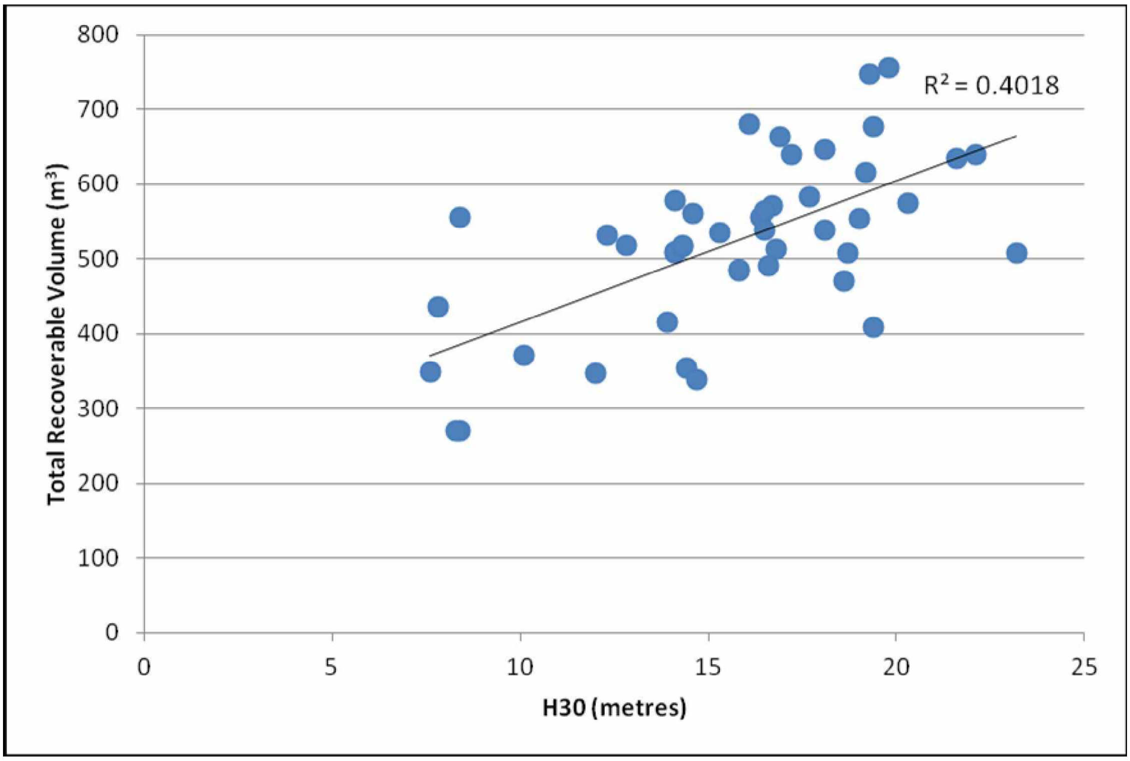

Figure 2. Relationship between H30 (30th Height Percentile) and tree volume for 45 plots in the example population

The first step is to calculate the linear regression slope coefficient which is estimated as:

$$b = \frac{SP_{xy}}{SS_x} = \frac{12700.46}{605.155}$$

$$b = 20.987$$

where SP_{xy} is the corrected sum of cross products of *x* and *y* and SS_x is the corrected sum of squares of *x*.

$$SS_x = \sum^{n} X^2 - \frac{(\Sigma^n X)^2}{n}$$

$$SS_{xy} = \sum^{n} (XY) - \frac{(\Sigma^n X)(\Sigma^n Y)}{n}$$

Substituting the computed slope value the sample means of y and x into the estimating equation for the population means of y gives:

$$\bar{y}_R = 522.9565 - 20.987 * (15.8575 - 15.900)$$
$$\bar{y}_R = 522.067$$

The standard error of $\bar{y}_R$ can be estimated as

$$S_{\bar{y}R} = \sqrt{S^2_{y.x}\left[\frac{1}{n} + \frac{(\bar{\mu} - \bar{x})^2}{SS_x}\right]\left(1 - \frac{n}{N}\right)}$$

This formulation for the standard error of $\bar{y}_R$ is documented by both Freese (1962) and Avery and Burkhart (1994). A number of publications use a slightly different equation for calculating the standard error of $\bar{y}_R$. Shivers and Borders [5] and Lohr [6] drop the $\frac{(\bar{\mu} - \bar{x})^2}{SS_x}$ term. None of the above publications give any rationale for the addition or exclusion of this term. This worked example uses the above formulation which includes the additional term as it gives more conservative results.

$$S^2_{y.x} = \frac{SS_y - (SP_{xy})^2 / SS_x}{n - 2}$$

In the standard error formula, $S^2_{y.x}$ is the estimate of variability of the individual y values about the regression of y on x:

$$S^2_{y.x} = \frac{570667.55 - (12700.47)^2 / 605.15}{45 - 2}$$

$$S^2_{y.x} = 7072.55$$

$$S_{\bar{y}R} = \sqrt{14.90\left[\frac{1}{45} + \frac{(15.8575 - 15.900)^2}{605.155}\right]\left(1 - \frac{45}{1432}\right)}$$

$$S_{\bar{y}R} = 12.34$$

Note that for this small sampling fraction, the finite population correction factor (1-n/N) could be omitted without any significant impact on the results. Confidence intervals on the mean can be computed as:

$$\bar{y} \pm tS_{\bar{y}R}$$

$$522.067 \pm (2.021)12.34$$

$$522.067 \pm 24.939$$

Future Forests Research Ltd, PO Box 1127, Rotorua. Ph: 07 921 1883 Email: info@ffr.co.nz Web: www.ffr.co.nz

The probable limit of error (PLE) is a common statistic used to describe the precision of forest inventory in New Zealand. It refers to the confidence limits expressed as a percentage of the estimated mean.

$$PLE = \frac{CI}{\bar{y}_R} \times 100$$

$$PLE = \frac{24.939}{522.067} \times 100$$

$$PLE = 4.77\%$$

If the same 45 plots were used without the use of the LiDAR metric as an auxiliary variable, how would the estimates compare with those above? The mean of total recoverable (y) would be estimated as:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\bar{y} = \frac{23533.04}{45}$$

$$\bar{y} = 522.96$$

as compared with 522.07 m^3/ha for the regression estimator. The estimate of the standard error of the mean for this simple random sample would be:

$$S_{\bar{y}} = \sqrt{\frac{S_y^2}{n}(1 - \frac{n}{N})}$$

$$S_{\bar{y}} = \sqrt{\frac{12969.7}{45}(1 - \frac{45}{1432})}$$

$$S_{\bar{y}} = 16.531$$

as compared with 12.34 for the regression estimate. Confidence intervals for the simple random sample are computed as:

$$\bar{y} \pm tS_{\bar{y}}$$

where t has n-1 degrees of freedom. Substituting in the appropriate values gives

$$522.9565 \pm 2.021(16.531)$$

$$522.9565 \pm 33.409$$

The standard error is much bigger and the confidence internal much wider for the simple random sample.

$$PLE = \frac{CI}{\bar{y}_R} \times 100$$

$$PLE = \frac{33.409}{522.9565} \times 100$$

$$PLE = 6.38\%$$

Through the use of LiDAR and regression sampling, the PLE for this inventory has been reduced by 1.6%. Table 2 gives a summary of the key statistics under the different sampling strategies.

Table 2. Comparison of Summary Statistics

Statistic	Simple Random Sampling	Regression Estimation
Mean Volume (m^3/ha)	522.96	522.07
Standard Error (m^3/ha)	16.53	12.34
CI (m^3/ha)	33.41	24.94
PLE (%)	6.38 %	4.77 %

It is possible to use muliptle linear regression estimators, but this is beyond the scope of this technical note. Readers interested in more in-depth discussions of regression estimators should refer to Cochran (1977).

Conclusion

Regression estimators can provide significant increases in precision. However the use of regression estimation in forestry inventory has been limited as the true population mean of the auxiliary variable must be known. The use of LiDAR data has changed this, as the true population mean of the LIDAR metric can be determined. This means that forest owners who either possess or are willing to acquire LiDAR over a forest of interest can use LiDAR in a regression estimation approach to either increase the precision of inventory estimates or reduce the number of plots required to achieve a specified target precision level. The techniques outlined here potentially provide significant benefits and cost savings to the forest owner which are available immediately. This lends further strength to the argument for collecting LiDAR for forest information purposes.

References

1. GRASS Devlopment Team (2010) Geographic Resources Analysis Support System (GRASS) Software Version 6.4.0. Open Source Geospatial Foundation http://grass.osgeo.org

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

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

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

5. Shivers, B.D. and Borders, B.E. (1996). Sampling techiques for forest resource invneotry. John Wiley and Sons Inc. New York, USA. 356 pp.

6. Lohr, S. (1999). Sampling: Design and Analysis, Duxbury Press, Pacific Grove. pp 608

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. Husch, B., T.W. Beers, and J.A. Kershaw, Jr. 2003. Forest Mensuration (4th Edition). John Wiley & Sons, Hoboken, NJ.