INFLUENCE OF LIDAR FULL-WAVEFORM DENSITY AND VOXEL SIZE ON FOREST STAND ESTIMATES

Pablo Crespo-Peremarch and Luis A. Ruiz

Geo-Environmental Cartography and Remote Sensing Group (CGAT), Universitat Politècnica de València. Camí de Vera s/n. 46022. València. Spain.

ABSTRACT

The aim of this study was to analyze the variation on accuracy and errors of aboveground biomass (AGB) and canopy base height (CBH) estimates when modifying lidar full-waveform (FW) pulse density, voxel size and regression methods. We reduced randomly pulse density from 9 to 1 pulses.m^{-2} in steps of 0.5 pulses.m^{-2} in 36 plots. Afterwards, eight FW metrics were computed for each pulse density, voxel size (i.e. 0.25, 0.5 and 1 m) and plot. These metrics were used as explanatory variables, while AGB and CBH obtained from field work were used as response variables in linear, square-root-transformed, exponential and power regression models. Results on CBH showed that power regression reduces pulse density influence. AGB was however influenced by voxel size, showing that it was more affected by pulse density changes when using smaller voxel sizes. In addition, square-root-transformed and power regressions performed better for AGB and CBH, respectively.

Index Terms— lidar full-waveform, pulse density, forest structure, aboveground biomass, canopy base height, regression methods

1. INTRODUCTION

Characterizing forest stand variables is essential for forest management and to better analyze ecosystem processes [11]. Several applications, such as harvesting, mapping carbon balance and wildlife habitats [10] require information given by forest stand variables.

These variables have traditionally been estimated on the ground or using destructive methods, which are costly and time consuming [12]. However, newer remote sensing technologies, such as light detection and ranging (lidar), have been successfully employed in the last decades instead of traditional techniques. Lidar technology can easily be collected over large areas and is capable of predicting accurately forest stand variables [2]. This makes lidar data a fundamental tool to map forest stand variables over large areas.

Lidar full-waveform (FW) is capable of registering the complete signal sent from the aircraft and intercepting the objects present on the ground. These data are especially useful in forested areas, providing more information from the different vertical layers than discrete lidar [8].

Both discrete and FW have not a homogenous pulse density over the flown area. Overlapped areas between different flight stripes have higher pulse densities than other areas. This makes that some areas are therefore more exhaustively registered by the lidar sensor than others. Consequently, these pulse density differences may affect estimate of forest stand variables.

How pulse density changes affect discrete lidar metrics and forest stand variables estimated from lidar metrics has been analyzed in several studies. [7] observed that forest stand variables model results were similar when reducing pulse density from 8 to 0.5 pulses.m^{-2} in Northwestern Spain. They also tested three different regression methods (linear, exponential and power), performing better results the exponential one. [9] obtained that prediction results of forest stand variables were similar above low pulse densities (i.e. 1 pulse.m^{-2}) in California USA. Coverage variables (i.e. canopy cover, tree density and shrub cover) were more affected by pulse density variation, while other variables (i.e. height, diameter at breast height, shrub height and basal area) did not differ much with pulse densities higher than 1. [13] analyzed how plot size and pulse density affect estimates of forest stand variables in Central Spain. They concluded that both factors affected model results, however pulse densities had a lower influence.

On the contrary, FW data has been less used and its preprocessing differs from discrete lidar, and therefore it has received less studies. An accepted preprocessing method to extract metrics is voxelization. It consists of clustering lidar return pulses into voxels (e.g. rectangular prisms), whose value is the maximum amplitude of all the return pulses contained in. Once voxelization is done, pseudo-vertical waveforms are described by the voxel values of each vertical column from the top to the ground [8]. Afterwards, FW metrics are extracted from this pseudo-vertical waveform. [5] analyzed the effect of pulse density variations on FW metrics in Oregon (USA). They observed that the lower the density was, the higher the number of voxels with null values. This made that FW metrics measured in adjacent samples with similar forest features but with different pulse density were statistically different. In order to solve this, they increased the voxel size to diminish the number of empty voxels. Apart

from FW metrics, changes on estimate of forest stand variables for different pulse densities using FW have however not been analyzed in the literature.

The aim of this paper is to analyze how aboveground biomass (AGB) and canopy base height (CBH) are affected by pulse density variations and voxel size using FW data, as well as testing four regression methods (linear, square-root-transformed, exponential and power) to estimate AGB and CBH. In order to do this, FW metrics were measured in a set of plots where pulse density was randomly diminished. In addition, FW metrics were computed for three voxel sizes (i.e. 0.25, 0.5 and 1 m).

2. METHODS

2.1. Study area

The study area is located in Panther Creek (Oregon, USA) (see Fig. 1), with a total area of 2,258 ha ranging from 100 to 700 m of altitude. The dominant species is Douglas-fir (*Pseudotsuga menziesii*), and it is occasionally mixed with other conifers. The tree height is variable due to harvesting, being sometimes higher than 60 m.

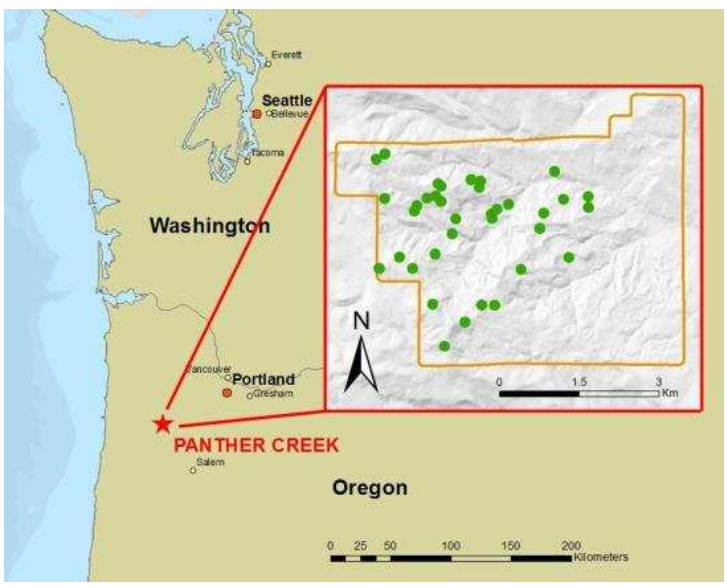

Fig. 1. Study area location and Douglas-fir dominated plots (green-colored) in the study area (orange-colored).

2.2. Lidar FW and field work data

FW data was collected in July 2010 using a Leica ALS 60 sensor. Average flight altitude was 900 m registering data at 105 kHz and a scan angle of ±15º, resulting a pulse density ≥8 pulses.m^{-2}. FW data was registered in 256 bins with a footprint size of 0.25 m and a temporal sample spacing of 2 ns (i.e. 0.3 m).

Regarding field work, a total of 84 circular plots with 16 m radius were measured. Within each plot, the dominant species and every tree with a diameter at breast height greater than 2.5 cm were registered. As a result, there were 47 plots where Douglas-fir was dominant, and 37 with mixed species. Afterwards, AGB and CBH were estimated using collected field data and allometric equations described by [15].

2.3. Lidar data processing

In order to work with plots having the highest pulse densities, we selected a representative subsample of 36 plots (Fig. 1). These plots had Douglas-fir as dominant species and its pulse density was higher than 9 pulses.m^{-2}, being then set as the initial pulse density. Pulse density was randomly reduced from 9 to 1 pulse.m^{-2} every 0.5 pulses.m^{-2} for the 36 plots from the subsample. The way we followed to select randomly the pulses for each density was computing the number of pulses required to obtain a density *d* within the plot area, and then selecting them randomly.

Once the initial dataset was modified according to pulse densities, we computed FW metrics, requiring voxelization. In this study we chose three voxel sizes in XY dimensions: 0.25, 0.5 and 1 m, being the smallest size the footprint size; and a size of 0.3 m in Z dimension according to the temporal sample spacing.

For each of the eight metrics (i.e. HOME, WD, NP, ROUGH, HTMR, VDR, RWE and FS, proposed by [6] and further described by [3]) we computed its mean within each plot. As a result, we obtained eight metrics for each plot and pulse density.

2.4. Regression models

Before generating the regression models, a selection of metrics was performed only for the highest pulse density (i.e. 9 pulses.m^{-2}) and each regression methods (i.e. linear, square-root-transformed, exponential and power). These selected metrics were used for all the pulse densities in order to better compare how it affects estimates without using different explanatory variables. The process followed for the metric selection was to compare the Akaike Information Criterion (AIC) [1] of all the possible models with a maximum of three FW metrics.

Once FW metrics were selected, we generated linear, square-root-transformed (sqrt), exponential (exp) and power regression models, as suggested by [4] and [7], for each pulse density. Finally, regression models were evaluated by comparing adjusted coefficient of determination (R^2) and root-mean-square error (RMSE), and using leave-one-out cross-validation.

3. RESULTS

Fig. 2 shows R^2 values obtained in AGB and CBH estimates for the different regression methods, pulse densities and voxel sizes. Regarding AGB, 1 m has the highest and less affected results by pulse density for all the regression methods. In addition, sqrt regression is also unaffected by pulse density at a lower voxel size (i.e. 0.5 m). Overall, sqrt has the highest R2 values. On the other hand, R2 values from CBH estimates stay steady until 1.5 pulses.m^{-2}, where they suddenly drop for all voxel sizes, except for 0.25 m with power and exponential regressions. For CBH estimates, the

power regression is more constant for all the voxel size than the linear, sqrt and exp regressions.

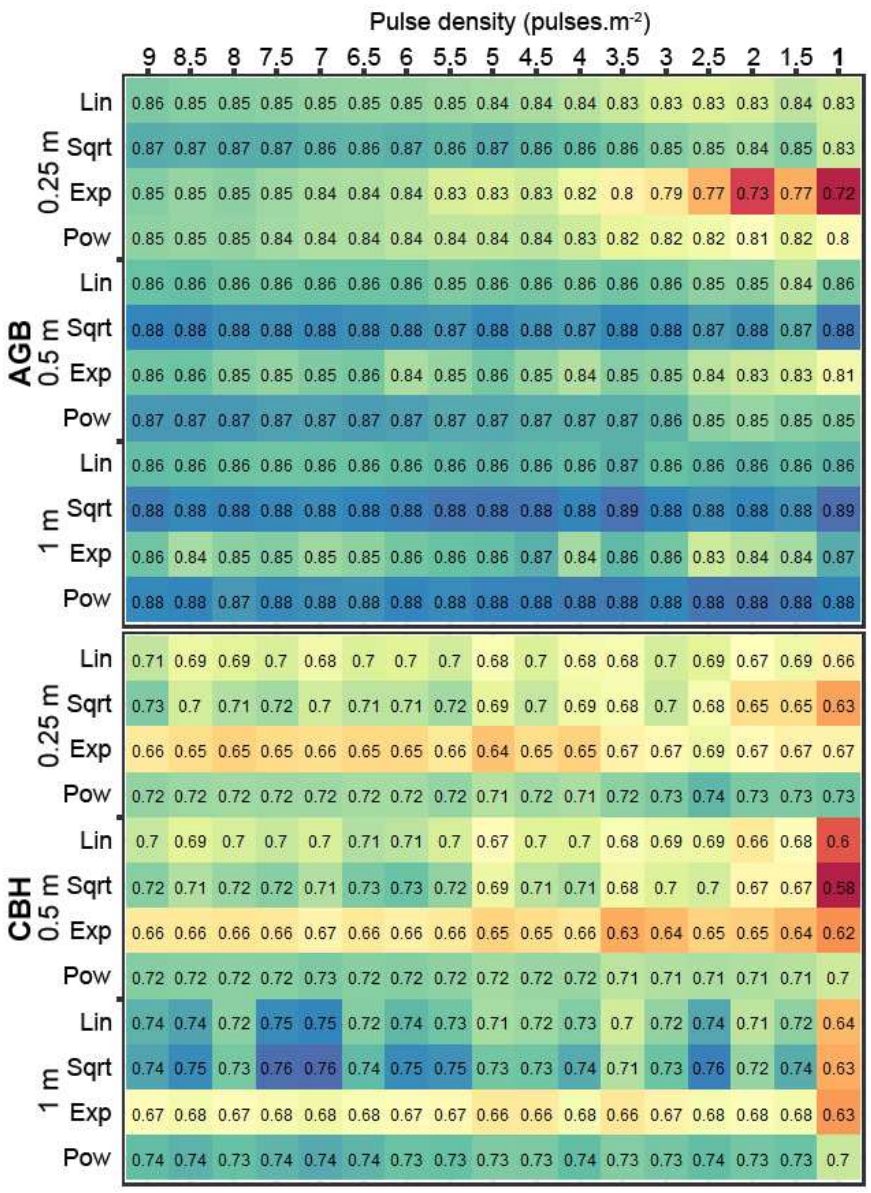

Fig. 2. Variation of R^2 values for AGB and CBH estimates for the different regression models (Lin: linear, Sqrt: square-root-transformed, Exp: exponential, Pow: power), pulse densities and voxel sizes. Red- and blue-colored cells represent the lowest and highest R2 values, respectively, for each variable.

Table 1 shows RMSE mean and standard deviation obtained from AGB and CBH estimates. Analyzing AGB results, the smaller the voxel size, the bigger the RMSE and the more variability. In addition, sqrt regression has the lowest RMSE. Conversely, CBH has lower RMSE differences between voxel sizes and less variable. Hence, CBH results do not depend on voxel size. In this case, power regression has the lowest RMSE. In general, the behavior of RMSE is coincident with that of the R^2 values.

4. DISCUSSION

Interpreting the results obtained, the influence of pulse density on AGB estimates depends on voxel size when pulse density diminishes. The larger the voxel, the more stable R^2 values are and the lower the RMSE. This is due to the fact that it is more likely that larger voxels contain more

Table 1. Mean and standard deviation from RMSE for the different test sets.

Methods	AGB (Mg.ha^{-1})			CBH (m)		
	0.25m	0.5m	1m	0.25m	0.5m	1m
Lineal	92.6 ±2.9	88.4 ±1.6	86.6 ±0.7	4.8 ±0.1	4.8 ±0.2	4.5 ±0.2
Sqrt	87.5 ±3.2	82.1 ±1.2	80.6 ±1.2	5.0 ±0.3	4.9 ±0.3	4.6 ±0.2
Exp	102.1 ±11.1	93.0 ±3.7	91.1 ±3.2	5.1 ±0.1	5.2 ±0.1	5.0 ±0.1
Power	97.4 ±4.0	87.2 ±2.8	82.4 ±1.1	4.6 ±0.1	4.7 ±0.1	4.5 ±0.1

waveforms when pulse density is lower. However, small voxels have more null values, varying FW metrics, and therefore making AGB estimates less accurate. Although the influence of pulse density related to regression methods is similar, sqrt regression performs better results.

On the other hand, CBH estimate variation is more affected by the regression methods, being the power and exp regressions more constant for all the pulse densities (except exp regression at 1 m). However, exp has the lowest results. Linear and sqrt regressions are constant with fluctuations as well, but for pulse densities higher than 1.5 pulse.m-2, where values suddenly drop. In this case, there is a lack of pulses to achieve an accurate estimate even for larger voxels. Differences between voxel sizes are lower for CBH, as shown in RMSE results, however, the voxel size of 1 m has slightly higher R^2 values again due to the fact that sufficient pulse density within the voxels is ensured.

AGB and CBH estimates are then differently affected by regression methods, voxel size and pulse density. Forest stand variables related to height are influenced by regression methods, but less influenced by voxel size and pulse density.

Regarding pulse density variations, our results are coherent with those reported by [7] and [9] with discrete lidar. [7] observed a lower variation in precision for variables related to height (i.e. mean and dominant height, ΔR^2 = 1.9-2.7%) than for AGB (ΔR^2 = 5.8%) when pulse density was reduced from 8 to 0.5 pulses.m^{-2}. [9] also concluded that height variables were less affected than cover-related variables.

Regarding regression methods, [4] also found that sqrt performed better AGB estimates. On the contrary, [7] had higher R2 values estimating the AGB with the exponential regression. In our study, however, the exponential regression reached the lowest results.

On the other hand, our results agree with those from [14] using FW, in the sense that variables related to height are not significantly influenced by the voxel size. Nevertheless, analyzing the influence of both parameters together, voxel size and pulse density, on forest stand estimates using FW was not addressed previously.

In general, using FW, AGB is affected by a reduction of pulse density in small voxels. On the contrary, CBH estimates are not significantly affected by pulse densities higher than 1.5 pulses.m^{-2}, or when power regression is used. Therefore, results show that pulse densities are not crucial to accurately estimate CBH. However, voxel size must be adequately increased to keep AGB precision when low pulse densities are used.

5. CONCLUSIONS

The present study has shown how AGB and CBH estimates using lidar FW vary when pulse density and voxel size are modified. The results suggest that AGB predictions are more affected by pulse density than CBH. However, R^2 and RMSE values for AGB remain the same when larger voxel size are used, reducing the number of empty voxels and balancing the lower pulse densities. For AGB, the R^2 and RMSE values are constant for large voxel sizes (>1 m), and for CBH when the power regression is used or pulse densities are higher than 1.5 pulses.m^{-2}. In addition, sqrt and power performed slightly better results than the linear and exponential regressions for AGB and CBH, respectively. This means that, in practical terms, lower densities can be used, reducing the cost on lidar data acquisition and allowing for a wider area coverage. Further studies will be focused on analyzing the effect of pulse density variation on FW metrics, in order to reduce the side-lap effect [5], which is not reflected on estimate results.

6. ACKNOWLEDGMENTS

This research has been funded by the Spanish Ministerio de Economía y Competitividad and FEDER, in the framework of the project CGL2016-80705-R. The authors also thank the Bureau of Land Management and the Panther Creek Remote Sensing and Research Cooperative Program for the data provided.

7. REFERENCES

[1] H. Akaike, "Information theory and an extension of the maximum likelihood principle", *Proceedings of the 2nd International Symposium on Information Theory*, Budapest (Hungary), pp. 267-281, 1973.

[2] H.E. Andersen, R.J. McGaughey, and S.E. Reutebuch, "Estimating forest canopy fuel parameters using LIDAR data", *Remote Sens. Environ.*, vol. 94, no. 4, pp. 441-449, 2005.

[3] L. Cao, N.C. Coop, T. Hermosilla, J. Innes, J. Dai, and G. She, "Using small-footprint discrete and full-waveform airborne LiDAR metrics to estimate total biomass and biomass components in subtropical forests", *Remote Sensing*, vol. 6, no. 8, pp. 7110-7135, 2014.

[4] P. Crespo-Peremarch, L.A. Ruiz, and A. Balaguer-Beser, "A comparative study of regression methods to predict forest structure and canopy fuel variables from LiDAR full-waveform data", *Revista de Teledetección*, no. 45, pp. 27-40, 2016.

[5] P. Crespo-Peremarch, L.A. Ruiz, A. Balaguer-Beser, and J. Estornell, "Analysis of the side-lap effect on full-waveform lidar data acquisition for the estimation of forest structure variables" *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Prague (Czech Republic), vol. 41, pp. 603-610, 2016.

[6] V.H. Duong, "Processing and application of ICESat large footprint full waveform laser range data", Doctoral dissertation, TU Delft, Delft University of Technology (Netherlands), 2010.

[7] E. González-Ferreiro, U. Diéguez-Arand, and D. Miranda, "Estimation of stand variables in Pinus radiate D. Don plantations using different LiDAR pulse densities", *Forestry*, vol. 85, no. 2, pp. 281-292, 2012.

[8] T. Hermosilla, N.C. Coops, L.A. Ruiz, and L.M. Moskal, "Deriving pseudo-vertical waveforms from small-footprint full-waveform LiDAR data", *Remote Sens. Lett.*, vol. 5, no. 4, pp. 332-341, 2014.

[9] M.K. Jakubowski, Q. Guo, and M. Kelly, "Tradeoffs between lidar pulse density and forest measurement accuracy", *Remote Sens. Environ.*, vol. 130, pp. 245-253, 2013.

[10] M.A. Lefsky, W.B. Cohen, S.A. Acker, G.G. Parker, T.A. Spies, and D. Harding, "Lidar remotes sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests", *Remote Sens. Environ.*, vol. 70, no. 3, pp. 339-361, 1999.

[11] J.E. Means, S.A. Acker, D.J. Harding, J.B. Blair, M.A. Lefsky, W.B. Cohen, M.E. Harmon, and W.A. McKee, "Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of Oregon", *Remote Sens. Environ.*, vol. 67, no. 3, pp. 298-308, 1999.

[12] N. Nordh, and T. Verwijst, "Above-ground biomass assessments and first cutting cycle production in willow (Salix sp.) coppice – a comparison between destructive and non-destructive methods", *Biomass Bioenerg.*, vol. 27, no.1, pp. 1-8, 2004.

[13] L.A. Ruiz, T. Hermosilla, F. Mauro, and M. Godino, "Analysis of the influence of plot size and LiDAR density on forest structure attribute estimates", *Forests*, vol. 5, no. 5, pp. 936-951, 2014.

[14] L.A. Ruiz, P. Crespo-Peremarch, J. Estornell, and A. Balaguer-Beser, "Analyzing some factors affecting the extraction of full-waveform LiDAR metrics and their effect in forest structure variable estimates", *Proceedings Book 7th edition of the International Scientific Conference ForestSAT*, Santiago de Chile (Chile), pp. 24-26, 2016.

[15] J.T. Standish, G.H. Manning, and J.P. Demaerschalk, "Development of biomass equations for British Columbia tree species", *Canadian Forestry Service, Pacific Forest Research Center, Information Report BC-X-264*, Victoria (BC, Canada), 1985.

View publication stats