A REGRESSION–BASED METHOD FOR CANOPY COVER ESTIMATION FROM CLASSIFYING LIDAR POINT BY FAMILY OF RETURNS

M.A.V. Posilero, E.C. Paringit, R.J.L. Argamosa, C.A.G. Ibañez, R.A.G. Faelga
A.A. Maralit, I.A.R. Teh, A. Palmon, G.P. Zaragosa

Phil-LiDAR 2, Project 3: Forest Resources Extraction from LiDAR Surveys, University of the Philippines-Training Center for Applied Geodesy and Photogrammetry, Diliman, Quezon City, 1001, Metro Manila, Philippines
Email: markposi@gmail.com

KEY WORDS: Light Detection and Ranging (LiDAR), Forest Canopy Cover, Digital Hemispherical Photography (DHP), LiDAR Point Return Family Classification

ABSTRACT: This paper developed and investigated a LiDAR-based canopy cover estimation model using metrics that can be derived from the distribution of points based on pulse origin. For a specific airborne system with 4 discrete returns, each point may possibly belong to ten (10) classes. Accordingly, this grouping can be used to infer other canopy biophysical characteristics like canopy cover. The method was tested over a 2-hectare forest plantation (FP) site against field measurement of canopy cover generated through Digital Hemispherical Photography (DHP) survey conducted on 15-24 February, 2015. The FP site was established in 1998 at 2m x 5m plant spacing and is characterized predominantly by mature Caribbean Pine trees with diameter at breast height ranging from 5cm-55.5cm and canopy height ranging from 13.01m-34.34m. The corresponding LiDAR point cloud data was acquired using Optech™ ALTM Gemini on August 16, 2013 at a flying height of 1979.547 masl. Field canopy cover observed in the FP site ranges from 48.87% to 72.39%. Results show that LiDAR canopy cover estimation for the FP was observed at R^2 = 0.7088 with a 2.97 percent canopy cover RMSE.

1. INTRODUCTION

Forest characterization with remote-sensing technology, specifically with the airborne light detection and ranging (LiDAR) data has provided opportunities for the availability of reliable, accurate, and up-to-date information to support decision-making in monitoring and management activities across a range of spatial and temporal scales (Wulder *et. al.*, 2008). Among the forest structural attributes that can be derived from LiDAR is the canopy cover, defined as the area of the ground obscured by the vertical projection of the tree canopy when viewed at a single point. Various field-based canopy cover measurement methods have been used across forestry research studies like sighting tubes, line-intercept sampling, densiometers, and digital hemispherical photography (DHP) (Korhonen *et. al.*, 2006; and Fiala *et. al.*, 2006). DHP is among the mature and relatively time-efficient, indirect optical field technique capable of generating reliable and accurate canopy cover measurements across forest classes, widely used as a validation or calibration technique for remotely-sensed data (Morsdorf *et. al.*, 2006; Palletto & Tosi, 2009; Smith *et. al.*, 2009; and Gonsamo *et. al.*, 2013). DHP utilizes a camera and fisheye lens system to acquire permanent records of canopy structure and geometry distributions.

With the increasing availability of airborne LiDAR data in the Philippines, this paper attempts to design and investigate an approach for profiling LiDAR point cloud distribution into groups or family of returns to measure canopy cover. Particularly, a regression modeling approach is introduced as a means to estimate canopy cover. To validate the results of the modeling exercise, LiDAR-derived canopy cover values was compared with DHP-generated canopy cover from a forest plantation (FP).

2. STUDY AREA

The study was conducted over a 2-hectare forest plantation (FP) site covered within a reforestation project of the Philippine Government in the Province of Bukidnon which commenced on February 1989 with the funding and technical assistance by New Zealand through Bukidnon Forests Incorporated (Clark, 1993). The FP site was established in 1998 at 2m x 5m plant spacing and is characterized predominantly by mature Caribbean Pine (*Pinus caribaea*) trees with diameter at breast height ranging from 5cm-55.5cm and canopy height ranging from 13.01m-34.34m (Argamosa *et. al.*, 2016). The FP site is located in Barangay Silo-O, Municipality of Malitbog, Bukidnon Province, Philippines and approximately situated at geographical coordinates 124° 59' 26.9052" N and 8° 29' 22.2504" E with elevation ranging from 970.58 m above sea level (asl) to 1009.80 masl (Figure 1).

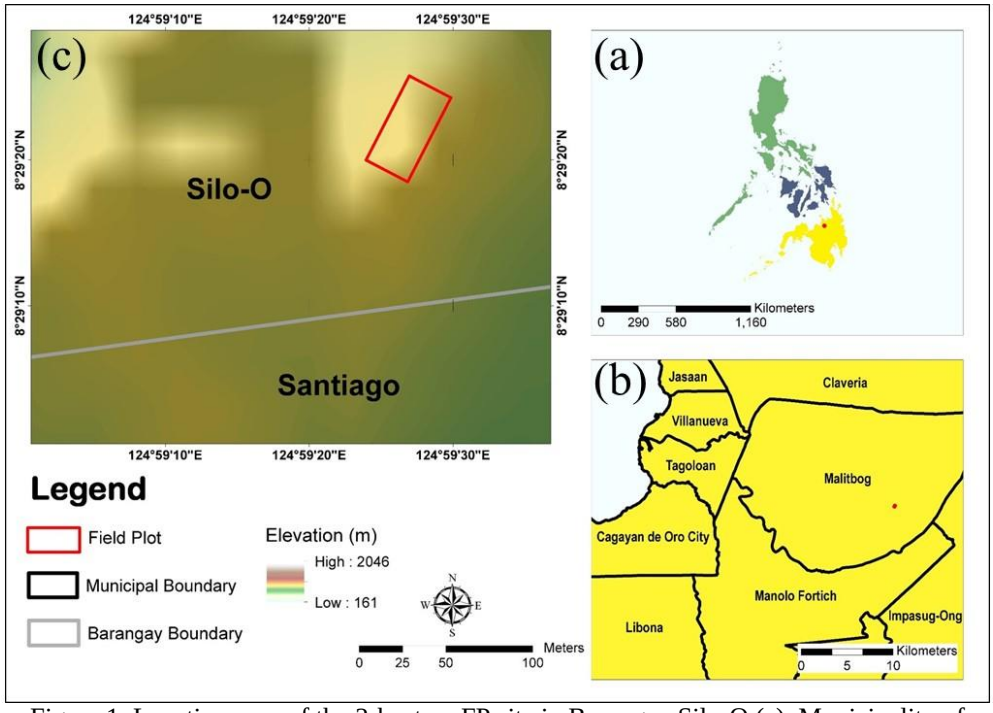

Figure 1. Location map of the 2-hectare FP site in Barangay Silo-O (c), Municipality of Malitbog (b), Province of Bukidnon, Philippines (c).

3. METHODOLOGY

3.1 Plot Design and Layout

The 2-hectare study area was first divided into 20m x 20m plots making a total of fifty (50) grids. Canopy images were then acquired from the established DHP set-up points situated approximately at the center of the grids as shown in Figure 2.

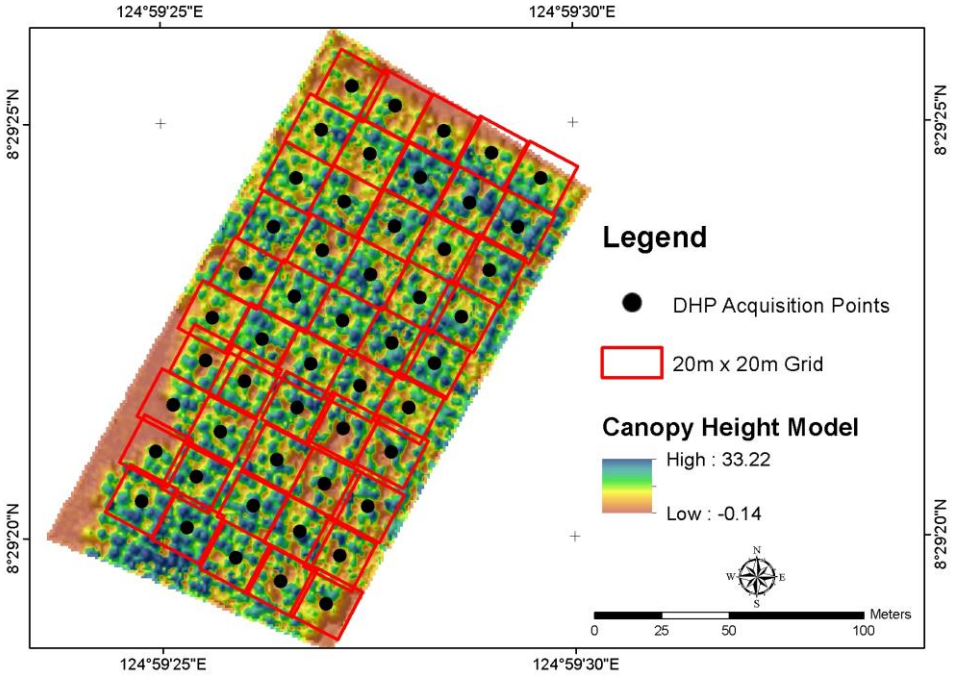

Figure 2. DHP plot layout and design in the FP site.

3.2 LiDAR Data

The airborne LiDAR data for Silo-O, Malitbog, Bukidnon was generated from Flight 388G acquired on August 16, 2013 using Optech™ ALTM Gemini at an average flying height of 1979.547 masl (UP TCAGP, 2015).

Table 1. Scanning characteristics of LiDAR data.

System Specifications	Data Attributes
Scanning Equipment	ALTM Gemini*
Acquisition Date	August 16, 2013*
Flying Height (masl)	1979.547*
Scan Angle (°)	-20, 8
Number of Returns per Laser Pulse	4
Number of First Returns	40684
Number of Intermediate Returns	3386
Number of Last Returns	40450
Number of Single Returns	32037
Average Point Density per m²	2.17
Average Point Spacing per m²	0.68
Average Pulse Density per m²	1.67
Average Pulse Spacing per m²	0.77

* UP-TCAGP (2015)

3.3 DHP-Generated Field Canopy Cover

A Nikon D40X digital single reflex camera and Rokinon 8mm F/3.5 aspherical, full frame fisheye lens mounted on a tribrach and tripod system was used to capture north-oriented, upward-looking canopy images of the FP site on February 15-24, 2015 at the inherent highest possible maximum resolution (3872 x 2592 pixels). Each DHP set-up was levelled using a two-axis bubble level and consequent hemispherical images were taken at instrument height of 1.3m, either early in the morning or late afternoon, or when there is no direct sunlight passing through canopy can be captured in the photographs to have a good visual discrimination of the sky and vegetation elements and minimize occurrence of mixed pixels. All DHP images were taken from the camera and fisheye lens system with uniform settings, aperture of f/8, shutter speed of 1/40,50 seconds and ISO 100 resulting to an exposure value of 11.33 and 11.67, respectively.

Nominal canopy cover (CC_N) was then then generated considering a 100° field of view angle of the hemispherical image samples, wherein a maximum likelihood classification tool embedded in ArcGIS 10.2.2. software was applied to quantify the portions of sky and vegetation. CC_N can then be related to DHP canopy cover (DHP_{CC}) by:

$$CC_{DHP} = CC_N \frac{r_1^2}{r_i^2 - r_{i-1}^2} \quad (1)$$

The area normalization factor, $\left(\frac{r_1^2}{r_i^2 - r_{i-1}^2}\right)$ based on the pixel radius (r) within a per zenith angle view expressed as $y = 0.6927x^{-1.093}$, where y is the area normalization value at a given zenith angle x, was introduced to the binarized (sky/gap and vegetation) hemispherical image samples to account for the exaggerated area of canopy cover with increasing angle of view towards the edges of the photographs (Posilero, *et. al.*, 2016).

3.4 LiDAR-Based Canopy Cover Estimation

The corresponding classified LiDAR (ground and vegetation points) data of the 2-hectare FP site (Figure 3a) was normalized (Figure 3b) using a point cloud data processing software so that all ground points will have an elevation of zero while the vegetation points which represent the trees will have relative heights above a flattened ground. Consequent processing of the normalized LiDAR data was done using an automated script presented in the workflow of classifying the family of returns (Figure 4). A 1.3m height above ground cut-off was first introduced to the normalized data similar to the DHP set-up acquisition height, wherein points falling below this threshold will be dropped, making sure that only vegetation points will be considered in the consequent processing. Point metrics, such as X, Y, Z coordinates, return number, number of returns for a given pulse and corresponding GPS time was generated from the remaining LiDAR vegetation returns. From this information, vegetation points with similar GPS time were grouped into ten (10) family return classes, namely: 1 Only (1_0), 1 of 2 (1_2), 2 of 2 (2_2), 1 of 3 (1_3), 2 of 3 (2_3), 3 of 3 (3_3), 1 of 4 (1_4), 2 of 4 (2_4), 3 of 4 (3_4), and 4 of 4 (4_4). To compare LIDAR-derived and DHP-based

canopy cover, there is a need to confine the comparison within the field of view of the DHP. A 10m circular radius buffer for each DHP set-up point was then made and used to clip the family return classes.

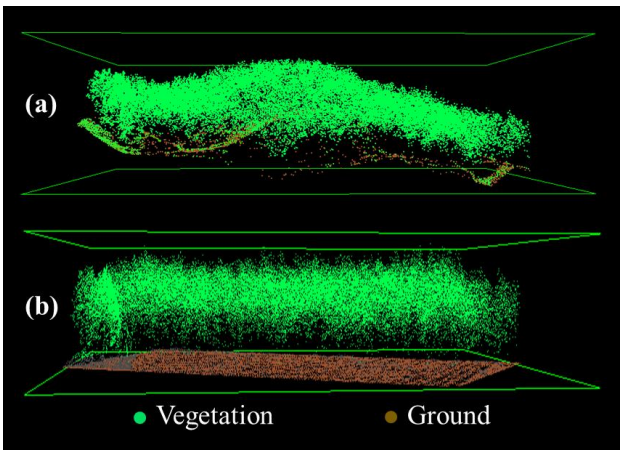

Figure 3. Height normalization of LiDAR point cloud data.

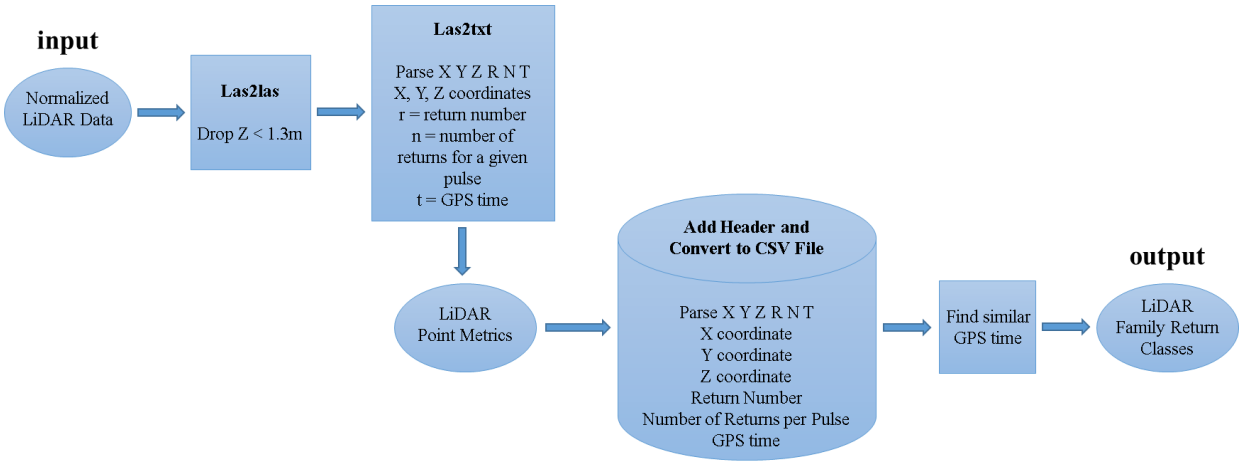

Figure 4. Workflow for classifying LiDAR points into family of return.

To account for the area coverage of the family return classes clipped within each DHP extent, beam radius (r_i) of the of the LiDAR pulses were also computed at different return heights (z_i) considering the beam divergence (θ = 0.25 mrad) and flying altitude during LiDAR data acquisition using:

$$r_i = (flying\ altitude - z_i)\left(tan\frac{\theta}{2}\right) \quad (2)$$

The computed beam radius across the 35-meter LiDAR point return height in the FP site can be expressed as y = -0.0001x + 0.2474, where y is the beam radius at a given LiDAR point return height x (Posilero *et. al.*, 2016).

Ten LiDAR point family classification coefficients were then derived using what-if analysis tool (Microsoft Office Excel Solver add-in) and a simple linear regression model was designed and tested based on the comparison of CC_{DHP} and CC_{LiDAR} in the FP site using:

$$CC_{LiDAR} = K + \sum_{i=1}^{n} F_i a_i \quad (3)$$

where CC_{LiDAR} is the LiDAR-derived canopy cover, K is the intercept, Fi is the coefficient of the *ith* family return extent.

4. RESULTS AND DISCUSSION

LiDAR family return classes generated is presented in Table 2. Majority of the vegetation points were classified under 1_0. These are the laser pulses which recorded only 1 laser return from the canopy. In general, LiDAR family return classes can be broadly categorized as family of 1 (1_0), family of 2 (1_2, 2_2), family of 3 (1_3, 2_3, 3_3) and family of 4 (1_4, 2_4, 3_4, 4_4) (Figure 5). Using these categories, returns belonging to family of 2 actually have the most number of points constituting about 39% followed by family of 3, family of 1, and family of 4, respectively. Information values from these distinctive LiDAR family classification and distribution can then be utilized as useful input for canopy cover estimation.

Table 2. Distribution of LiDAR family return classes.

Family Return Class	Number of Points	Percentage
1_0	13859	20.53
1_2	13158	19.50
2_2	13158	19.50
1_3	6885	10.20
2_3	6885	10.20
3_3	6885	10.20
1_4	1666	2.47
2_4	1666	2.47
3_4	1666	2.47
4_4	1666	2.47

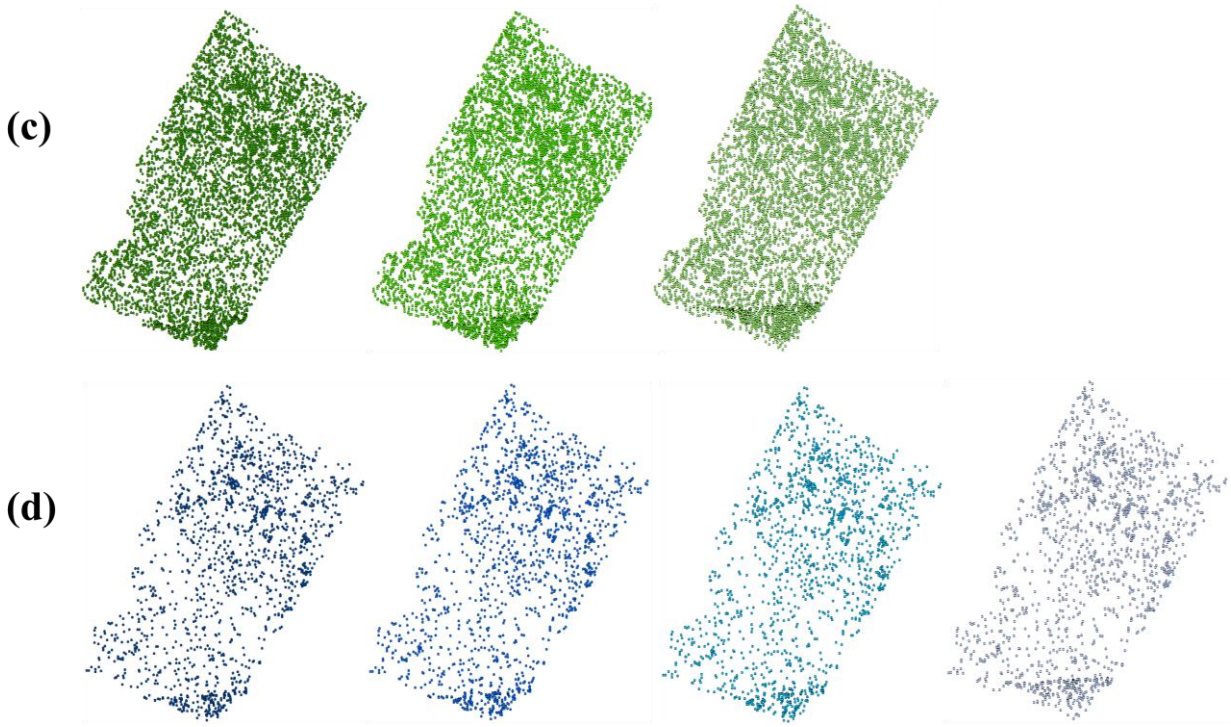

Figure 5. Plots of the LiDAR return classes aligned as: (a) family of 1, (b) family of 2, (c) family of 3, and (d) family of 4.

Field canopy cover generated from the hemispherical image samples and corresponding LiDAR canopy cover estimates of the FP site is presented in Table 3. Based on the results, CC_{DHP} and CC_{LiDAR} lies on the same magnitude having ±3.22% standard deviation with canopy cover ranges of 48.87 to 72.39% at an average of 61.14% and 53.70-70.71% at an average of 61.13%, respectively. A minimal absolute average difference of 2.69% between the field and estimated canopy cover values was also observed and based on the international definition and classification of forests (FAO, 2001), the FP site can be categorized as a closed Caribbean Pine forest plantation.

Table 3. CC_{DHP} and CC_{LiDAR} generated from the FP site.

DHP Number	*CCDHP*	*CCLiDAR*	Difference		DHP Number	*CCDHP*	*CCLiDAR*	Difference
2	49.86	55.01	-5.15		27	56.77	61.36	-4.59
3	61.51	60.26	1.25		28	60.43	62.75	-2.32
4	61.57	56.37	5.19		29	59.05	62.38	-3.33
5	53.66	56.97	-3.31		30	57.63	61.94	-4.31
6	61.75	61.94	-0.19		31	54.30	55.66	-1.37
7	63.96	65.94	-1.99		32	57.19	58.79	-1.61
9	59.25	60.74	-1.50		34	56.22	57.61	-1.39
10	54.49	59.88	-5.38		35	59.58	56.90	2.68
11	65.93	62.43	3.50		36	52.77	53.77	-1.00
12	66.09	66.51	-0.42		37	58.73	56.15	2.58
14	69.35	69.08	0.28		38	61.55	61.61	-0.06
15	62.88	62.18	0.70		39	61.46	63.20	-1.74
16	66.97	63.65	3.32		41	48.87	53.70	-4.84
17	61.43	65.58	-4.15		42	60.16	58.14	2.02
18	72.39	67.36	5.03		43	56.61	59.35	-2.74
20	66.19	66.11	0.08		44	60.13	55.92	4.20
21	65.12	62.87	2.25		45	64.83	70.71	-5.88
22	62.10	60.28	1.82		46	63.14	59.27	3.87
23	65.26	60.61	4.65		47	70.16	63.87	6.28
24	65.58	64.50	1.07		48	69.90	67.48	2.42
25	59.32	60.06	-0.74		49	66.94	63.97	2.97
26	60.99	57.81	3.17		50	58.20	59.02	-0.82

In general, the regression model tends to overestimate CC_{DHP} in the FP site, particularly on the denser canopies. The overestimation may have been introduced from the high porosity of the pine canopy elements wherein quantification of the area covered by the vegetation points were partially based. The overestimation could also be attributed to the temporal discrepancy in the acquisition of LiDAR data and field DHP survey since more than 18 months have passed since the LIDAR data has been acquired. Changes in canopy elements could have occurred in between the period, especially due to plant growth where the pine trees could have shed-off its lower needle-leaves, cones, and branches. Furthermore, seasonality could have also affected the assessment of the CC_{DHP}. It can also be observed from the LiDAR family return class weights (Table 4) that F_{3_4}, F_{2_4}, and F_{1_3} are the top predictor variables of canopy cover in the FP site followed by F_{2_3}, F_{3_3}, and F_{1_4}, respectively. This observation can be inferred from the overlapping conical form of canopy growth characteristic to pine tree species in the FP site, and in the fact that majority of these LiDAR family returns generally lie on the upper part of the canopy height profile.

Table 4. Family return class weights in the FP site.

	Coefficients, *F*
K	60.59026765
1_0	-0.947134511
1_2	7.256800886
2_2	-6.969782444
1_3	-16.77116648
2_3	9.556533989
3_3	9.290593158
1_4	9.249854118
2_4	19.55162316
3_4	-30.12695338
4_4	0.699811279

Looking further on the residual plot of the FP site (Figure 6), the regression model exhibited a random and unpredictable (stochastic) error. In other words, the model is correct on average for all the predicted values.

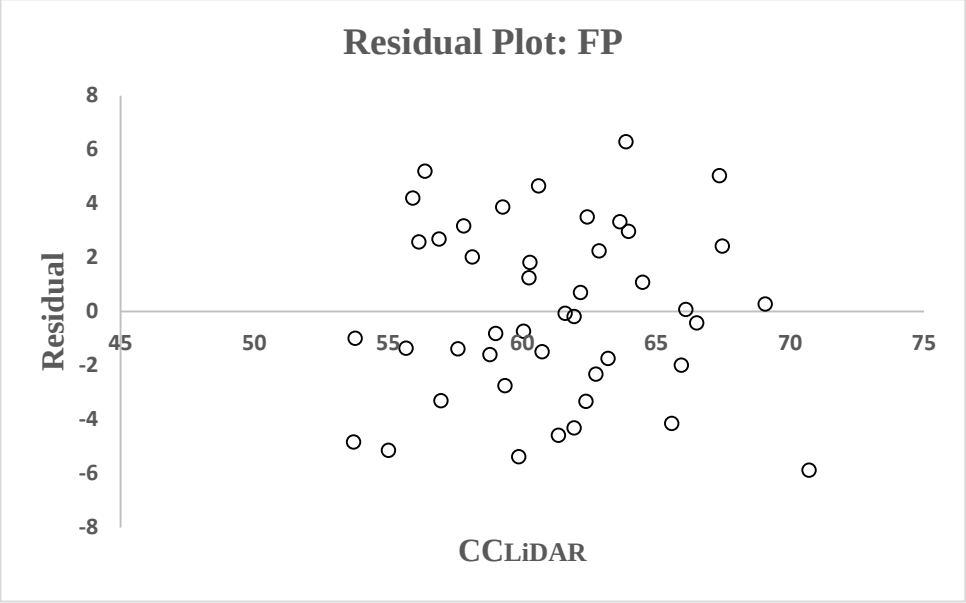

Figure 6. Residual plot of the FP site (CC_{LiDAR}).

After the outliers have been detected and removed from the analysis using the mean plus or minus two standard deviations of the residuals, the regression model was able to perform at R^2 = 0.6256 (Figure 7) and at 3.044 percent canopy cover RMSE.

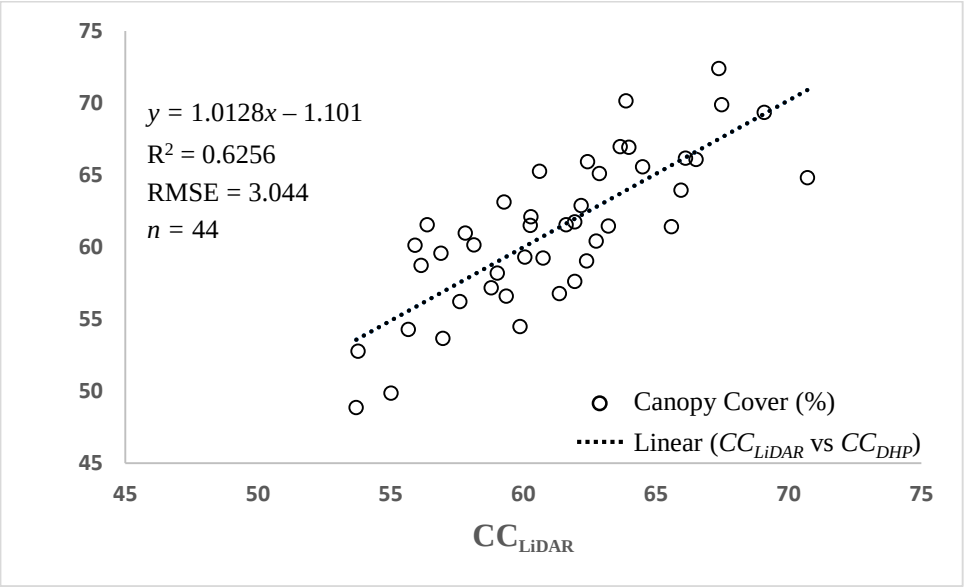

Figure 7. Regression plot of the FP site (CC_{LiDAR} vs CC_{DHP}) with outliers removed.

In addition, the regression model was also able to outperform the return ratio method in deriving canopy cover (CC_{RR}). This is done merely by dividing the number of first returns above some height threshold over all returns for a specific grid size (Hopkinson & Chasmer, 2009), which was observed at R^2 = 0.1826 and RMSE = 32.1778 (Figure 8). Residual plot of CC_{RR} also suggest that the model tend to exhibit clear trend, overestimating canopy cover values (Figure 9).

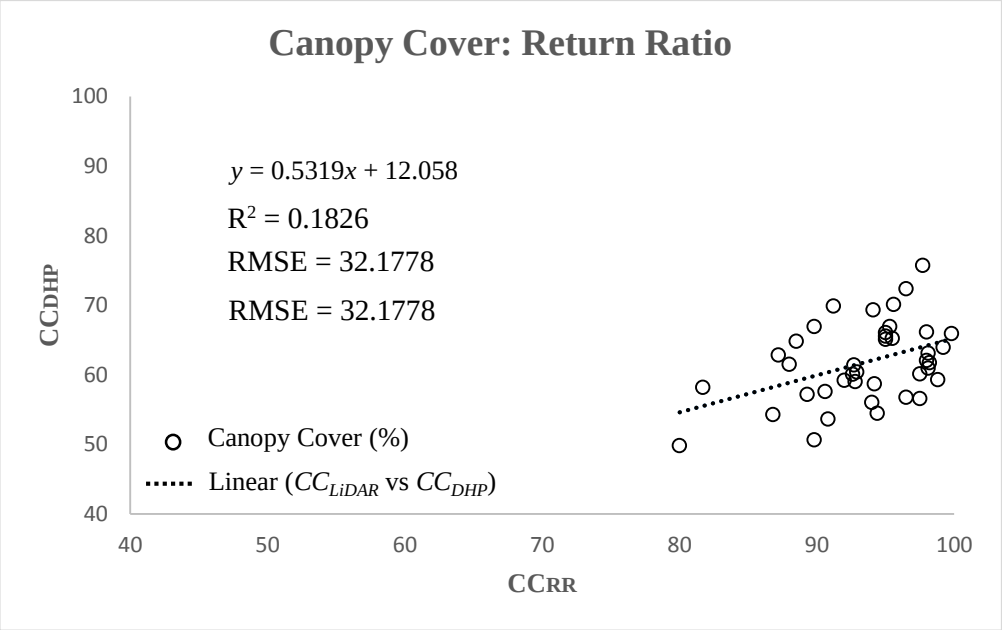

Figure 8. Regression plot of the FP site (CC_{RR} vs CC_{DHP}).

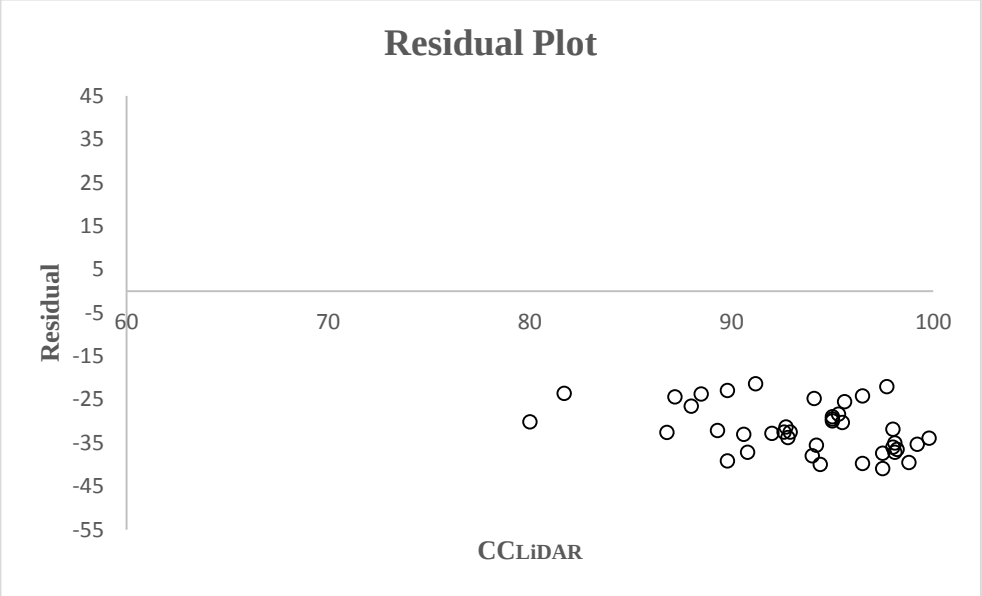

Figure 9. Residual plot of the FP site (CC_{RR}).

5. CONCLUSION

The designed and tested regression model making use of LIDAR family of returns was capable of producing accurate canopy cover estimates in the FP site. Furthermore, the analysis of LiDAR point cloud data based on family of returns provided a new insight in viewing the distribution of the vegetation points. With the application of area normalization factor to fully utilize the maximum field of view angle in hemispherical images and likewise, using the computed beam radius used to quantify area coverage of the LiDAR vegetation points could also be a reasonable approach in experimenting and identifying relationship towards an improved accuracy of LiDAR-derived canopy cover.

6. ACKNOWLEDGEMENTS

This research was done under Project 3. Forest Resources Extraction from LiDAR Surveys of the Phil-LiDAR 2 Program "Nationwide Detailed Resources Assessment Using LiDAR". The LiDAR point cloud data used were provided by Phil-LiDAR 1 Program "Flood Hazard Mapping of the Philippines Using LiDAR". Both programs are funded by the Department of Science and Technology- Grant-in-Aid (DOST-GIA). Credits are likewise due to the Central Mindanao University (CMU) FRExLS for jointly conducting the field validation and Bukidnon Forests Incorporated (BFI) for granting access to their plot sites and assisting in the field.

7. REFERENCES

Argamosa, R.J.L., Paringit, E.C., Quinton, K.R., Tandoc, F.A.M., Faelga, R.A.G., Ibañez, C.A.G., Posilero, M.A.V., & Zaragosa, G.P. 2016. Fully-Automated GIS-Based Individual Tree Crown Delineation Based on Curvature Values from a LiDAR-Derived Canopy Height Model in a Coniferous Plantation. Manuscript submitted for publication.

Clark, P.D. 1993. Philippines Forestry – Bukidnon Forests Inc. New Zealand Journal of Forestry, pp. 31-33.

FAO. 2001. Global Forest Resources Assessment 2000 (140). Rome, Italy.

Fiala, A.C.S., Garman, S.L., & Gray, A.N., 2006. Comparison of five canopy cover estimation techniques in the western Oregon Cascades. Forest Ecology and Management, 232, pp. 188-197.

Gonsamo, A., D'Odorico 2013. Measuring fractional forest canopy element cover and openness – definitions and methodologies revisited. Oikos Journal, 122, pp. 1283-1291.

Hopkinson, C. & Chasmer, L. 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. Remote Sensing of Environment, 113, pp. 275-288.

Korhonen, L., Korhonen, T.K., Rautiainen, M., & Stenberg, S., 2006. Estimation of Forest Canopy Cover: a Comparison of Field Measurement Techniques. Silva Fennica, 40(4), pp. 577-588.

Morsdorf, F., Kotz, B, Meier, E., Itten, K.I., & Britta, Allgower, 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. Remote Sensing of Environment, 104, pp. 50-61.

Paletto, A. & Tosi, V. 2009. Forest canopy cover and canopy closure: comparison of assessment techniques. European Journal of Forest Research, 128, pp. 265-272.

Posilero, M.A.V., Paringit, E.C., Argamosa, R.J.L, Faelga, R.A,G., Ibañez, C.A.G., Zaragosa, G.P. 2016. LiDAR-Based Canopy Cover Estimation Using Linear Regression Techniques. Manuscript submitted for publication.

Smith, A.M.S., Falkowski, M.J., Hudak, A.T., Evans, J.S., & Robinson, A.P., 2009. A cross-comparison of field, spectral, and lidar estimates of forest canopy cover. Canadian Journal of Remote Sensing, 35(5), pp. 447-459.

UP-TCAGP. 2015. DREAM Ground Survey for Cagayan De Oro River, Disaster Risk and Exposure Assessment for Mitigation (DREAM) Program, DOST Grants-In-Aid Program, pp. 54.

Wulder, M.A., Bater, C.W., Coops, N.C, Hilker, T., & White, J.C., 2008. The role of LiDAR in sustainable forest management. The Forestry Chronicle, 84(6), pp. 807-826.