Correction pit free canopy height model derived from LiDAR data for the broad leaf tropical forest

Lindah Roziani Jamru
Geosciences & Digital Earth Centre (INSTeG), Research Institute of Sustainable Environment, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru
lroziani2@live.utm.my

Abstract. The accuracy of forest inventory information, such as tree height estimation and individual tree crown delineation depending on the extraction of Canopy Height Model (CHM) that represent the absolute tree height of vegetation. However, uncertainties variation of the CHM known as a pit, negatively influence the biophysical measurements. These occur when the laser beam penetrates the part of the leaf and the branch before produces the first return to represent the surface of canopies. Despite of that, multiple laser beam from the different flight line cause a variation of the tree heights. To reduce the error of CHM, this work applied pit free algorithm CHM for broad leaf tropical forests. The results shown that, the pit free CHM produces a good correlation with the ground data with $R^2$ 0.86, meanwhile, the standard CHM with $R^2$ 0.78. For the accuracy assessment, Accuracy Index (AI) has been used to calculate the error of commission and omission for tree detection. The analysis proved that the pits free CHM improved the accuracy of tree detection with the total Accuracy Index value 69% and standard CHM recorded 52.3%.

1. Introduction

Light Detection and Range (LiDAR) has high potential to provide an accurate extraction on variables of forest structure [1]. The laser scanning consist of a million of point clouds that capable to penetrate every layer of crown produced the variables of forest structure such as tree height, tree density, canopy gap, crown area and tree crown layer [2]. It is important to gather the accurate variables of forest structure, since it becomes an input model for related ecological studies. The accuracy of the other ecological studies depends on the how accurate the extraction of the variables forest structure has been done. Commonly, all the parameter obtains from the CHM. Previous work by [3] reported 70 percent of tree crown successfully delineated based on CHM for suburban vegetation using LiDAR fullwave form data.

The computation between Digital Surface Model (DSM) and Digital Terrain Model (DTM) produced CHM that represents the tree height. However, several issues came out when discuss about the accuracy of CHM due to the influence of pits. Pits known as a hole or the uncertainties variation of the tree height that causes the error in the CHM values. This happens when the first LiDAR pulse penetrates the leaf and branch before producing the first return [4]. The other reason is first returns that produce from different flight lines that combine together show best view from the horizontal but has a variation in vertical distribution. Besides, the post processing raw point cloud also contributes to the pits; this is due to miss filtering the noise point cloud.

It is important to remove the data pits, since it can affect the extraction of the variables forest structure and tree detection from the CHM [5]. Pits will cause the difficulty in an identification separate crown, crown radius and underestimated the tree height [6]. So that previously, smoothing
technique has been applied to minimize pits on the CHM. Commonly, there are four types of filtering that have been used, such as natural neighbor (NN), interpolation of the highest point method (HPM), median filter, and mean filter. Filtering processes will smooth the raster of CHM, however, this process will change the tree height values not only for the pits but also for small trees. The values of CHM after filtering are not the original values, but according to the nearest neighbor values [7]. This will negatively affect the measurement of variables related to forest structure. In this study, the new algorithm introduced by [8] has been applied to remove data pits. However, the value of parameters has been modified for suitability of Tropical forest. The advantage of using this algorithm is to remove the noise point cloud and maintain the boundary and forms of tree crown and canopy gaps, without modifying the values. Besides, this work will provide information on CHM according to the interval 5 meter threshold. Altogether, this study will improve the extraction of biophysical forest structure of Tropical forest.

2. Materials and methods

2.1. Study area

Danum Valley is located at the south-eastern part of Borneo Island in the state of Sabah, specifically at 4°50’N-5°00’N and 117°35’E-117°45’ (Figure 1). The total area of 43,800 ha of land divided into primary forest, secondary forest, and replanting timber is under Sabah foundation management. Generally, the forest of Danum is dominated by Dipterocarps, particularly Parashorea malaanonan, P. tomentella and Shorea johorensis. Commonly, the Dipterocarps dominated the upper layer, while understory species from families such as Euphorbiaceae and Rubiaceae. Overall, 44 types of species have been recorded during this study. There are 30 plots involved in this study located randomly in the watershed; according to the flight line scanning LiDAR that overall represents the structural conditions in Danum Valley.

Figure 1. Map of study area
2.2. Satellite and ancillary data

2.2.1 LiDAR data and pre-processing
LiDAR data used in this study was acquired on 11 October 2013 using laser Optech C200 HD ILRIS system. The total area of interest has been scanning was 20.59 km$^2$ with flying height is 500 meters. The specification of LiDAR systems illustrated in Table 1. The LiDAR system also equipped with Digital Global Positioning System (DGPS) receiver and inertial measurement component that ascertain a sub-decimeter differential position can be quantified for the aircraft during post-processing. This sensor collected four returns per pulse that emit the laser for each layer of forest structure with an average 25 points/m$^2$. The point cloud was classified to non-ground points and ground point using Terra Scan filtering algorithm.

Table 1. Specification of LiDAR systems

| Parameter               | Specification     |
|-------------------------|-------------------|
| Laser systems           | Optech- HD ILRIS discrete return |
| Laser frequency         | 75 Hz             |
| Flying speed            | 90 KPH            |
| Flying height           | 500m              |
| Laser swath width       | 693m              |
| Scan angle              | 14.2 degree       |
| Swath overlap           | 35%               |

2.2.2 In-situ measurement
There are 30 plots involves in field measurements represent the forest structure in Danum Valley. The area of study was sampled using square plots with plot size 30 m x 30 m. For the ground measurement, the parameters of forest structure has been collected is tree height, Diameter Breast Height (DBH), tree crown, tree position and species. All trees with DBH larger or equal to 10 cm within the plot area were measured. For tree positioning, the horizontal distance was measured from the northwest corner of the plot. Overall the total trees were measured is 1402. But for this paper, DV 12 has been chosen as a sample plot with 42 trees. For the verification coordinates of plot location, three DGPS has been set up at benchmark Taliwas, weather station Danum Valley and at the center of each plot. It is important to know the difference between the ranges of signals for correction positioning error. Table 2 shows the descriptive statistic of trees for the field data.

Table 2. Descriptive statistic of the ground measurements

|             | n = 42 | DBH    | Height | CD    |
|-------------|--------|--------|--------|-------|
| Minimum     | 10     | 7.4    | 1.85   |       |
| Maximum     | 68.7   | 45.3   | 10.65  |       |
| Median      | 17.85  | 15.1   | 4.0625 |       |
| Mean        | 21.8   | 16.7   | 4.6    |       |
| Standard Deviation | 14.9  | 7.8    | 1.9    |       |

2.2.3 Application of pit free algorithm
There are two phase process of generating the CHM involved in this work. First of all, the normalized LiDAR point cloud was used to generate canopy height model that represents the absolute tree height. During this process, the undesired point cloud that derived from LiDAR raw data will be eliminated. Refer to Figure 2, show the difference between the normalized point cloud data starting from zero that relative to the similar ground altitude datum. Differ with the raw data that starting with value of negative due to the influenced of the terrain. The Lasheight in the lastools have been used in this study to normalize the point cloud. The previous technique has been used to obtain the tree height based on
the computation between two raster DTM and DSM. But in this study, we use the normalized point cloud and convert it to raster to get the absolute value of the tree height.

![Figure 2](image)

**Figure 2.** Raw point cloud (a), normalized point cloud (b)

Then, the second phase is to produce the CHM from the first return pulse that relate to higher vegetation hits. In this study, we used the interval threshold 5 meters to maintain the original morphological structure of the forest. Figure 3 shows the generating of CHM pit free according to the threshold level. The first layer generated is a standard CHM using all first return. Then a second layer generated using first return with a height of 2 meter and above. Next layer produced based on the every 5 meter threshold and above. The entire layer is combined to produce pit free of CHM and the basic concept is to compute the profile of the crown at different levels of tree height. The cell size has been used to generate the CHM is 0.3.
Figure 3. Series of height and a rasterization threshold
Figure 4 shows the different between pit free CHM and standard CHM that generated from LiDAR data. There are small holes known as a pit detectable within the tree crown in standard CHM, but it is invisible in the pit free CHM. The algorithm applied using Lastool well eliminated the pits and conserved the boundaries of the tree crown. The accuracy of extraction variables of forest structure is influenced by the pits. Then, it is important to eliminate the noise to get the precise variables of forest structure and the accurate value of the tree height.

2.2.4 Accuracy assessment
For the accuracy assessment, the detected tree from pit free CHM was compared with the tree has been recorded during fieldwork. The DBH size has been used to evaluate the error of commission and omission. The nearest detected treetop within the tree crown edge was counted as a correctly detected tree. Based on the field data, if there is more than one tree detected, then the other will assumes as a commission, means false detected tree. Meanwhile, if there is no treetop detected inside the borderline, then it is known as an omission. The accuracy index (AI) by [9] has been used to calculate the accuracy of tree detection. The equation (1) is defined as below;

$$AI(\%) = \left(\frac{n - O + C}{n}\right) \times 100$$

where,  
$n$ is the number of reference trees in the plot,
$O$ is the omission error,
$C$ is the commission error

3. Results and Discussions
The underestimation of the tree height is the main issues for CHM in forest areas [10]. So, the tree height at the plot level was derived from LiDAR data compared with the field measurements for validation. Figure 5 shows the correlation between tree height of pit free CHM and ground measurements with $R^2$ 0.86. Meanwhile, Figure 6 show correlation between tree height derived from standard CHM and ground data with $R^2$ 0.78. The highest tree has been recorded from field data, pit free CHM and standard CHM is 45.3, 44.19 and 48.7 meter respectively. Whereas, the lowest tree height has been recorded is 7.4, 8.09 and 11.44 meters for ground measurement, pit free CHM and
standard CHM respectively. The previous work by [11] reported the $R^2$ between simulated data and measured data tree height for method mean filter and median filter is 0.66, while HPM is 0.71. Results shown that pit free algorithm is the best method compared to the other methods. This method need rasterization and tree height threshold to construct the CHM. It is more complicated than other methods, but the computation of the tree crown at the different level of layer able to remove the pits without modifying the values of CHM.

![Figure 5](image1.png)

**Figure 5.** Regression between ground observation of tree height with tree height derived from pit free CHM

![Figure 6](image2.png)

**Figure 6.** Regression between ground observations of tree height with tree height derived from standard CHM

Figure 7 shows the commission errors, omission errors and correctly detected trees that derived from LiDAR data. Based on the figure, the position of the tree was measured in the field show by the black square, while the star point shows the tree detected from pit free CHM and the circle represent the radius of the tree crown. Table 3 illustrates the number of correctly detected trees, commission and omission errors and the accuracy index that derived from standard CHM and pit free CHM. All the parameters were assessed based on the classification of DBH size. DBH was classified into four classes has listed in Table 3. There are 42 trees has been measured in the sample plot and DBH size from 10 to 15 were dominants in the sample plot. Typically, the rates of commission error are higher in the dense forest, especially tropical forest that has a complex ecosystem. The results show that most of the small tree with the DBH size 10 to 15cm has a higher commission error. The analysis proved
that the pits free CHM improved the accuracy of tree detection with the total Accuracy Index value 69 percent, while standard CHM recorded 52.3 percent.

**Figure 7.** Detection of individual tree, commission error and omission errors from pit free CHM

**Table 3.** Results of tree detection from standard CHM and pits free CHM

| Field measurement | Standard CHM | Pits free CHM |
|-------------------|--------------|---------------|
|                  |  |                  |
| DBH sizes        | Number of tree | Commission | Omission | Correct n (%) | AI (%) | Commission | Omission | Correct n (%) | AI (%) |
| 10 - 15          | 16 | 9 | 2 | 5 | 37.5 | 6 | 1 | 9 | 56.25 |
| 15.1 - 20        | 11 | 4 | 0 | 9 | 63.6 | 3 | 0 | 10 | 72.7 |
| 20.1 - 25        | 6 | 1 | 2 | 5 | 50.0 | 2 | 0 | 6 | 66.6 |
| 25.1 >           | 9 | 2 | 0 | 6 | 77.7 | 1 | 0 | 7 | 88.8 |
| Total            | 42 | 16 | 4 | 25 | 52.3 | 12 | 1 | 32 | 69.0 |

**4. Conclusions**

The pit free algorithms applied to remove the pits able to increase the accuracy of extraction variables of forest structure through treetop detection and reduced the underestimation of the tree height at the plot level. These methods overcome the weakness of previous methods such as Natural Neighbour (NN), interpolation of the Highest Point Method (HPM), median filter and mean filter. Besides, it is capable to remove the pits without modifying the value, forms and the structure of the tree crown and canopy gaps. For further work, this method should be compared with the latest methods has been developed by [11] using three dimensional point cloud rather than the raster CHM, that is robust locally weighted regression to perform interpolations and robust Z scores to detect data pits. The difference between these two methods is about how they remove the pits and construct the CHM.
Overall, the results show the application of pit free algorithm successfully improves the extraction of variables forest structure.

Acknowledgements
This paper was a part of the project Q.J130000.3009.00M10 “Airborne LiDAR sensing of tropical forest structures for assessment of water resource in Sabah”. This is a collaborative between Universiti Teknologi Malaysia and Universiti Malaysia Sabah under scheme “Research Acculturation Collaborative Effort (RACE) that funded by Ministry of Higher Education.

References
[1] Means J E, Acker S A, Brandon J F, Renslo w M, Emerson L and Hendrix C J 2000 Predicting forest stand characteristics with airborne scanning lidar. *Photogramm. Eng. Remote Sens.* 66, 1367–1371
[2] Maltamo M, Packalén P, Yu X, Eerikäinen K, Hyyppä J and Pitkänen J 2005 Identifying and quantifying structural characteristics of heterogenous boreal forests using laser scanner data. *For Ecol. Manage.* 216, 41–50
[3] Jamru, L.R., Hashim, M. 2015 Modeling LiDAR cloud point to estimate tree height and delineation of tree crown for selected vegetation in suburban Kluang, Johor, 36th Asian Conference on Remote Sensing: Fostering Resilient Growth in Asia
[4] Persson A, J Holmgren and U Soderman 2002 Detecting and measuring individual trees using an airborne laser scanner, *Photogrammetric Engineering & Remote Sensing*, 68 (9): 925–932
[5] Shamsoddini A, R Turner and J C Trinder 2013 Improving lidar based forest structure mapping with crown-level pit removal, *Journal of Spatial Science*, 58 (1): 29–51
[6] Ben-Arie J R, Hay G J, Powers R P, Castilla G and St-Onge B 2009 Development of a pit Filling algorithm for LiDAR canopy height models. *Computers & Geosciences*, vol. 35, no. 9, pp. 1940–1949
[7] Zhao D, Y Pang, Z Li and G Sun 2013 Filling invalid values in a lidar-derived canopy height model with morphological crown control, *International Journal of Remote Sensing*, 34 (13): 4636–4654
[8] Pouliot D A, D J King, F W Bell and D G Pitt 2002 Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration, *Remote Sensing of Environment*, 82 (2–3): 322–334
[9] Khosravipour, A, Skidmore, A.K, Isenburg, M, Wang, T and Hussin, Y.A 2014 Generating pit free canopy height models from airborne lidar. *Photogrammetric Engineering & Remote Sensing* Vol.80, No. 9, pp. 863–872. 0099-1112/14/8009–863
[10] Liu, H.; Dong, P. A 2014 new method for generating canopy height models from discrete-return lidar point clouds. *Remote Sens. Lett.* 5, 575–582
[11] Chuanfa, C, Yifu, W, Yanyan Li, Tianxiang Y, Xin W 2017 Robust and parameter-free algorithm for constructing pit-free canopy height models. *ISPRS Int. J. Geo-Inf.*, 6, 219; doi:10.3390/ijgi6070219