IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Exploring intensity metrics in raw LiDAR data processing for tropical forests

To cite this article: Lindah Roziani Jamru *et al* 2024 *IOP Conf. Ser.: Earth Environ. Sci.* **1412** 012005

Exploring intensity metrics in raw LiDAR data processing for tropical forests

Lindah Roziani Jamru [1*], Mazlan Hashim [2], Mui How Phua [3], Adi Jafar [1], Nordin Sakke [1], Oliver Valentine Eboy [1], Ubong Imang [1], Maslina Natar [2], Azizul Ahmad[4] and Sumayyah Aimi Mohd Najib [5]

[1] Faculty of Social Science and Humanities, Universiti Malaysia Sabah 88400 Kota Kinabalu, Sabah
[2] Institute of Geoscience and Digital Earth Centre ,Universiti Teknologi Malaysia 81310 Skudai, Johor
[3] Faculty of International Tropical Forestry, Universiti Malaysia Sabah 88400 Kota Kinabalu, Sabah
[4] Centre for Spatially Integrated Digital Humanities (CSIDH), Faculty of Social Sciences & Humanities, Universiti Malaysia Sarawak 94300 Kota Samarahan, Sarawak
[5] Faculty of Human Sciences, Jalan Upsi 1, Perak, 35900 Tanjong Malim, Universiti Pendidikan Sultan Idris

*E-mail: lindahroziani@ums.edu.my

Abstract. LiDAR sensing is an active sensor that can produce three-dimensional point clouds. This sensor offers the 3D acquisition and analysis of forest data, providing details on the vertical structures of the forest. This study delved into the processing of raw LiDAR data obtained through laser scanning, employing software tools such as Justin Javad, Pospac MMS, LMS, Terrascan and TerraMatch. The processes involved are mission planning, LiDAR data scanning, trajectory processing and data calibration. This is the crucial part of processing that defines the quality of the raw LiDAR data. The results showed that the standard error recorded for intensity metric ranged from 0.14 to 0.68. It is important to characterize the intensity metrics that provide useful information for identifying specific objects in a LiDAR point cloud. Foresters can leverage this information to interpret both the forest canopy and terrain, aiding in effective forest management. The precision achieved in intensity metrics enhances the utility of LiDAR technology in providing actionable data for forestry applications. This study has resulted in a data processing tool designed to optimize the advantages of utilizing intensity data for object recognition. This tool holds significant importance for users of LiDAR data.

Keywords: LiDAR, three-dimensional, point clouds, intensity metric, tropical forest.

1. Introduction

The traditional field-based approach for measuring forests is increasingly being replaced by methods that integrate on-site observations with remote sensing (RS) data. Remote sensing data are highly valuable across various fields, including environmental studies [1] [2], agriculture [3] and land use change detection [4]. A significant advancement in this evolution was the introduction of airborne laser scanning (ALS) [5] [6]. ALS enables precise examination of both vegetation and terrain, a capability challenging to achieve with passive remote sensing. Private companies in Scandinavia have primarily employed ALS for area-based forest inventory and single-tree remote sensing (STRS), as noted in studies by [7] and [8]. Additionally, ALS has found application in image based or photogrammetric STRS, owing to its user-friendly nature in three-dimensional (3D) reconstruction of intricate canopies, as demonstrated by research such as [9].

Currently, there are lot previous studies to extract the individual trees [10] [11] [12] [13]. Many new algorithms and filtering methods have been developed to minimize the error of LiDAR data and achieve the objective of the research. For instance, correction pits free algorithm, natural neighbour (NN), interpolation of the highest point method (HPM), median, and mean filter. All the filtering approach are used to minimize the error and improve the accuracy of LiDAR data.

LiDAR, or Light Detection and Ranging, is a remote sensing technology that uses laser light to measure distances and create detailed, three-dimensional maps of the terrain. In the context of tropical forests, which are often dense and complex, the intensity metrics refer to the strength of the return signal from the laser beams. The backbone of a LiDAR sensor includes its key components such as the laser source, scanning mechanism, and detectors. Different LiDAR sensors may have varying specifications, including pulse repetition rate, pulse energy, beam divergence, and detector sensitivity. These factors collectively contribute to the intensity values recorded by the sensor.

In tropical forests, the dense vegetation can affect LiDAR signals in several ways. The interaction between the laser pulses and the forest canopy leads to variations in the intensity of the return signals. Understanding how the LiDAR sensor's backbone interacts with the tropical forest environment is crucial for accurate and meaningful data interpretation. Researchers typically conduct calibration and validation exercises to characterize the sensor's response in tropical forest conditions. This involves collecting ground truth data and comparing it with the LiDAR-derived metrics. The goal is to identify any biases or limitations in the intensity values and to develop correction algorithms if necessary.

Moreover, the choice of intensity metrics matters. Some LiDAR sensors provide multiple intensity returns per laser pulse, allowing for a more detailed analysis of vegetation structure and composition. Researchers may explore metrics such as peak intensity, mean intensity, or waveform analysis to extract valuable information about the tropical forest canopy.

The lack of studies using LiDAR metrics in Tropical forests and their potential for estimate variables of forests structure was the motivation for this study. The objective of this study to process the raw data of LiDAR data in intensity metrics of tropical forests. The raw processing LiDAR sensor's in intensity metrics of tropical forests involves understanding the interplay between the sensor's design, the dense vegetation, and the resulting intensity values. This knowledge is essential for accurate and reliable remote sensing applications in tropical forest ecosystems.

2. Materials and methods

2.1. Study area

Danum Valley is situated in the southeastern part of Borneo Island, within the state of Sabah, precisely located at 4°50′N-5°00N and 117°35′E-117°45′E (refer to Figure 1). Managed by the Sabah Foundation, this region spans a total area of 43,800 hectares, subdivided into primary forest, secondary forest, and replanting timber areas. The forest in Danum is predominantly characterized by Dipterocarps, with notable species including *Parashorea malaanonan*, *Shorea johorensis* and *P. tomentella*. *Dipterocarps* dominate the upper layer, complemented by understory species from families such as *Euphorbiaceae* and *Rubiaceae*. The study conducted in this area recorded a total of 44 different species.

Figure 1. Map of study area.

IOP Conf. Series: Earth and Environmental Science **1412** (2024) 012005 doi:10.1088/1755-1315/1412/1/012005

2.2. LiDAR data
In this study, LiDAR discrete return data were used to characterise forest structures and canopies. This LiDAR system has millions of point clouds capable of penetrating every layer of the forest canopy and characterising the variables of forest structures on the horizontal and vertical axis. The system has extensive spatial coverage to collect data at point densities of one to several laser returns per square meter and is sufficient to accurately identify individual tree crowns.

2.3. Differential global positioning systems
During the flight scan, the DGPS equipment was set up at the GPS control point (M415) at Taliwas, the nearest GPS control point from the study site. DGPS data collection was set up at 9.11 am with a height of 1.403m. The GPS static data was collected at one-second intervals for time lengths of eight hours using Global Navigation Satellite System (GNSS) receivers. The antenna heights were measured and recorded from the base of the antennas and reduced to the phase centre during post-processing. Figure 2 shows DGPS setup at the JUPEM GPS control points and at the centre of the plots. This allowed the system to create its location in 3D space so that the extraction of forest structure variables will have a precise 3D position as the ground needs real-time observation with the main ground control point using (GNSS) receivers.

Apart from that, during the ground measurement of forest structure variables, the DGPS was set up with a duration of two hours to achieve mm accuracy for each plot. At the same time, the other DGPS was set up at the Danum Centre weather station. The location for each plot was also recorded by the DGPS for the calibration of the total plot readings. However, due to the complexity of the canopies, some of the DGPS points in the plots received minor satellite interference from the adjacent trees.

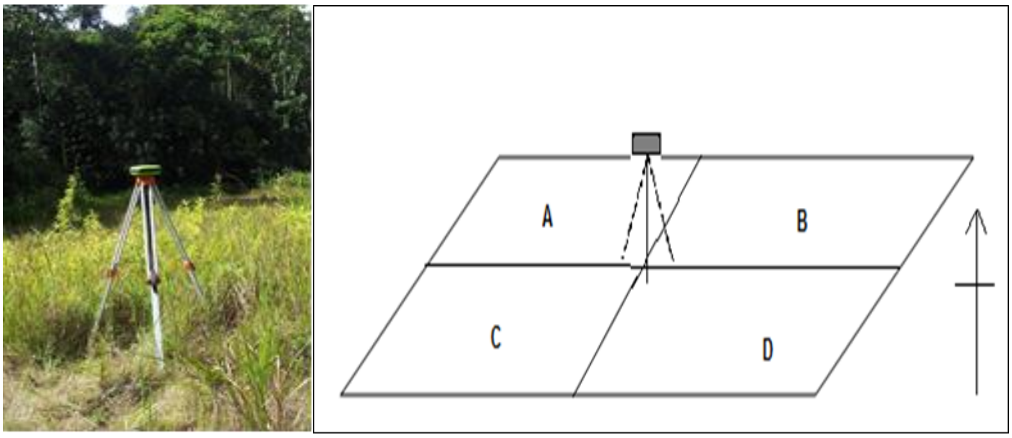


Figure 2. DGPS setup at the JUPEM GPS control points and at the centre of the plots.

2.4. Raw LiDAR data
This study collects raw LiDAR data from laser scanning until it produces point clouds in a readable format as product data. Figure 3 shows the process flow of raw LiDAR data from the laser pulse to registered point clouds data. LiDAR data is processed using Justin Javad, Pospac MMS, LMS, TerraScan and TerraMatch. The adjusted GPS data and the IMU data were combined in order to obtain a fixed-wing position, altitude, and to keep track of the aircraft rotations in the x, y, z axis and the GPS to keep track of the actual location of the aircraft in space. The output point clouds in binary format are obtained using this software. After pre-processing, the product data are exported to the micro station for the cleaning process to remove the noise.

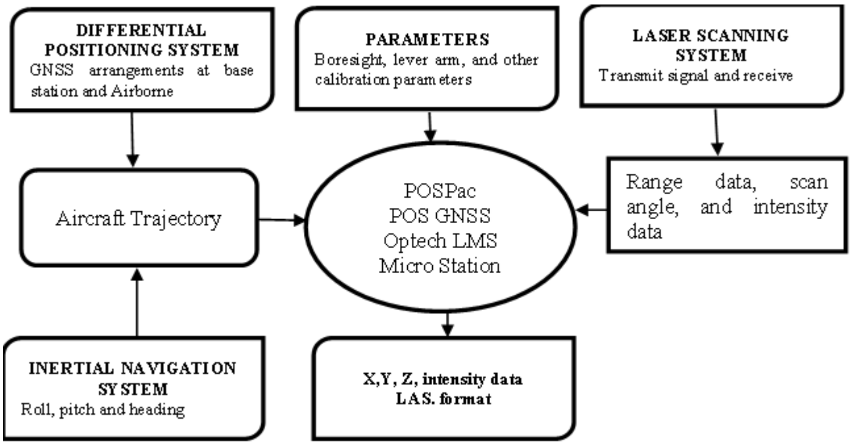


Figure 3. The workflow processing of the LiDAR system.

2.5. *Mission planning*

Before the aircraft scans the study area, mission planning was carried out to ensure the accuracy of data collected and to set all the parameters needed for data collection. ALTM-NAV Planner software was used in the mission planning for data logging, flight management, and survey planning (refer to Figure 4). In ALTM-NAV Planner software, the study area was identified, and several parameters were set to ensure that all necessary data was collected. The essential parameters were recorded, including the number of flight lines, scan angle, flight speed, overlap percentage, flight altitude, flight duration, and point density.

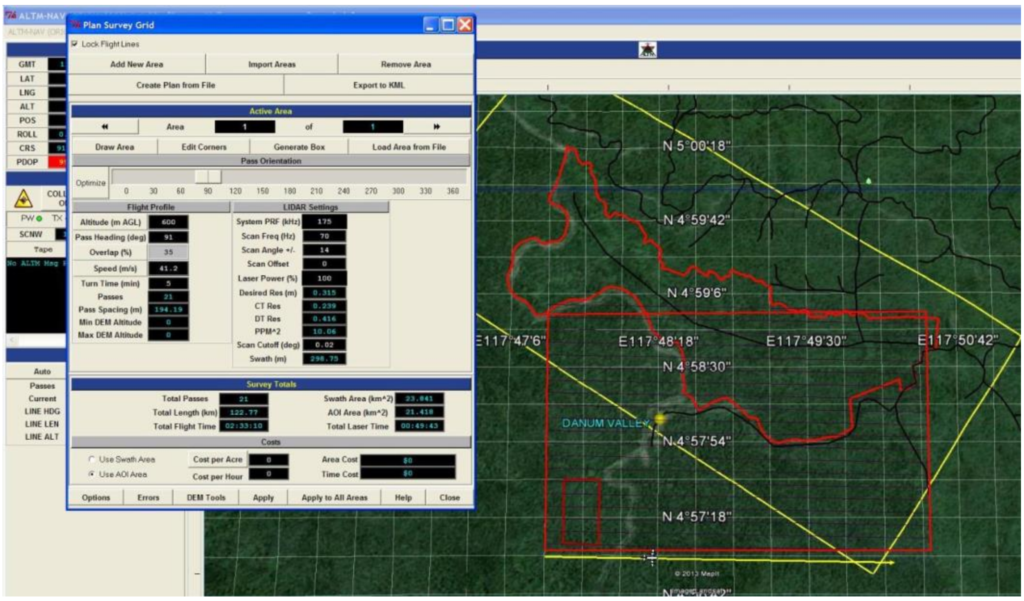


Figure 4. Mission planning in ALTM-NAV Planner for the study area.

IOP Conf. Series: Earth and Environmental Science **1412** (2024) 012005 doi:10.1088/1755-1315/1412/1/012005

2.6 LiDAR data scanning

After the mission planning was successfully set up, on 10 October 2013, aircraft LiDAR sensor scanned two areas, including Danum Valley, which covers the primary and secondary forests, and Tawau city, which were used for calibration purposes. There are 21 flight lines scanned for the study area and 7 flight lines for the calibration area. Table 1 shows the LiDAR metadata in the study area.

Table1. Metadata for airborne LiDAR in the study area.

	Tawau City	**Danum Valley**
Acquisition Date	9 October 2013	11 October 2013
Laser Scanner	Optech Orion-C200	
Aircraft	Nomad N22C	
Flight Altitude	900 m	500 m
Flight Speed	120 kts	90 kts
Scan Angle	20°	14.2°
Scan Frequency	35 Hz	70 Hz
Point per square metre (PPM^2)	1.75	8.00
Point repetitive frequency (PRF)	100 KHz	175 z

2.7 LiDAR data preparation

After the data collection, the pre-processing task followed with trajectory processing and data calibrations of the point cloud. The final output after pre-processing stages will produce a registered coordinate system of the point cloud as product data in .las file format. Figure 5 below shows the flow chart process of raw LiDAR data.

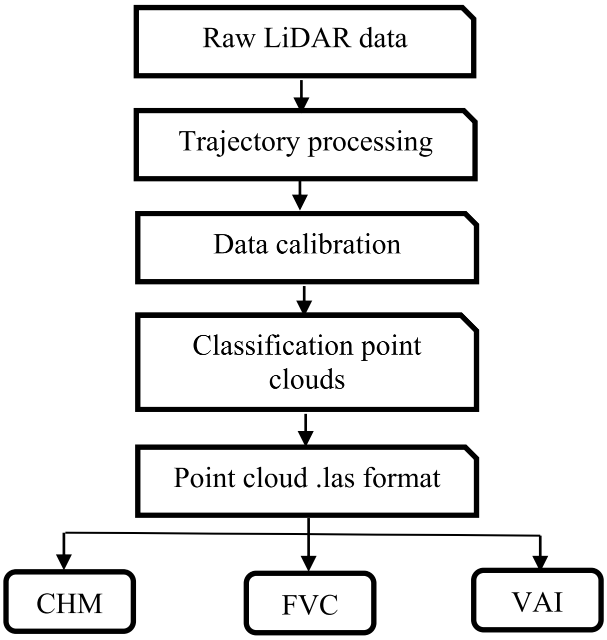


Figure 5. The workflow processing of raw LiDAR data.

2.8 Trajectory processing

LiDAR trajectory processing is the process to combine the POS data from GPS and IMU systems that determine the positional accuracy of LiDAR data. The IMU tracks the tilt in the skies as the aircraft is used for LiDAR scanning flies and calculates the accuracy of elevation. The GPS (TRIUMPH-VS DGPS) was also set up on the ground during the LiDAR scan. In this study, Taliwas JUPEM benchmark (M415) is the nearest GPS control point from the study area and is used as reference. Both the GPS and IMU systems were processed using POSPac MMS 6.1 to link the LiDAR point clouds to the real-world coordinate system. The processing of GPS and IMU results in the final orientation parameters consisting of X, Y, Z coordinates and their orientation angles for each LiDAR points to the location on the ground.

2.9 Data calibration

After the point cloud has been registered, the data is calibrated. Optech LiDAR Mapping Suite Manager (Optech LMS Manager) software is used to calibrate LiDAR data and process the data in a readable format (.las format). In order to correct the angular misalignment between the IMU and the laser, and also to check the consistency of the collected data, the calibration building was carried out by having the aircraft fly several passes over the building before and after data acquisition. The correction process is called bore sighting and usually involves the correction of four scanning errors known as heading, pitch, roll, and scale.

Tawau City was used to calculate all misalignment errors, and the final results were used in the Danum Valley study area to correct any misalignment that occurred within the overlapping LiDAR flight strips. Further calibration procedures were carried out using Terrascan and Terramatch extension tools in Microstation to increase the accuracy of LiDAR data. The dz RMS value was used to evaluate the accuracy of the calibration. The lower final value of the dz RMS compared to the initial value indicates that the misalignment had decreased (error reduced). The general acceptance level of error is less than 0.15m.

2.10 LiDAR data processing

2.10.1 Classification point clouds

The processed LiDAR points in.las format was then classified using the Microstation V8i software. Data in point-cloud form have been classified into the ground and non-ground points. Figure 6 shows the classification of the point cloud to ground and non-ground points.

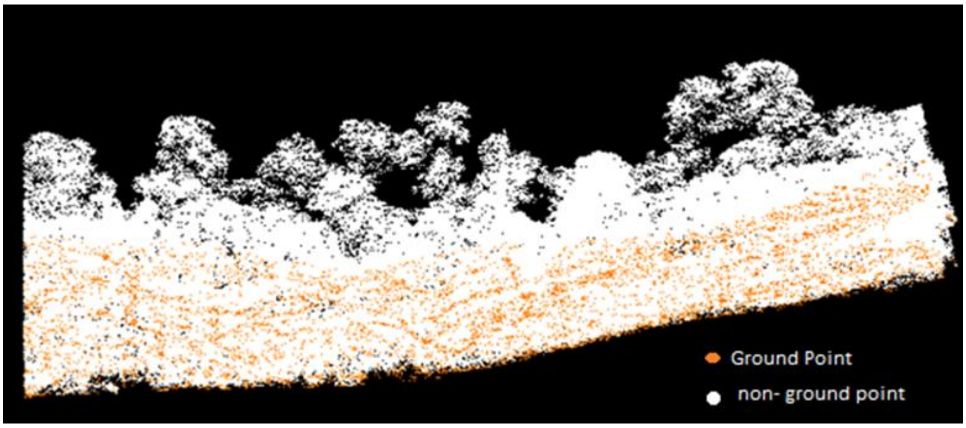
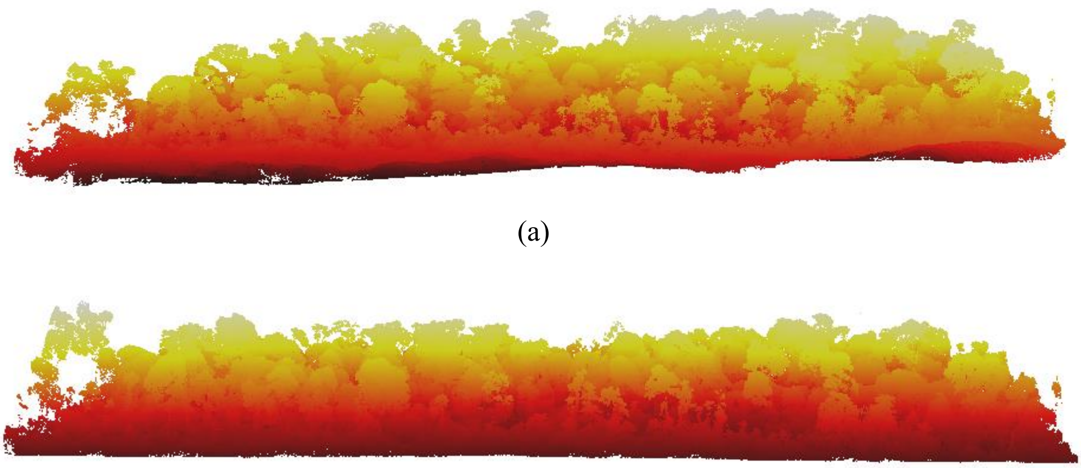


Figure 6. Classification of the point cloud to ground and non-ground.

IOP Conf. Series: Earth and Environmental Science **1412** (2024) 012005 doi:10.1088/1755-1315/1412/1/012005

After the point cloud has been classified, the next step is to normalise the LiDAR points. Lasheight in the lastools is used to normalise the point cloud. During this process, the undesired point cloud derived from LiDAR raw data will be eliminated. Figure 7 shows the difference between the normalised point cloud data starting from zero is relative to the similar ground altitude datum. This differs with the raw data that starts with a negative value due to the influence of the terrain. Normalised vegetation point cloud data were then used for further analysis in which various LiDAR metric information were extracted as an input rainfall interception estimation model.



(a)

(b)

Figure 7. (a) Raw point cloud, (b) Normalized point cloud.

3. Results and Discussions

3.1 Accuracy assessment raw LiDAR data

To achieve the objective, this study processed the raw LiDAR data from laser penetration until it produced data point clouds. This section presents the results of GPS network adjustment, data calibration, strip adjustment and classification of point clouds.

3.2 GPS network adjustment

The aircraft GPS trajectories were differentially corrected to the ground GPS control point. The corrected GPS control point at Danum Weather Station (DWS) was used for differential correction to the other 30 plots to ensure reliable differential processing of aircraft GPS data. To determine the coordinate for the individual points, the Inertial Measurement Unit (IMU) and Digital Global Positioning System (DGPS) in the aircraft were combined with the Position and Orientation System (POS) data. The final geographic coordinates with the corresponding mean sea level and ellipsoid vertical elevations are shown below in Table 2. GPS stations were observed in WGS 84 format and have an ellipsoidal height for each plot that are relative to Malaysian primary GPS network. An ellipsoidal height recorded a range between 1.27m to 1.77m recorded by plot DV144 and DV19, respectively. The ultimate height value shall be within 0.15m accurate to actual value RMSE or any other value as specified. From the results obtained, it is shown that the final adjusted baselines have low standard deviations and RMS errors (Table 3). All of the baselines meet and exceed the first order network standards as specified by the Federal Geodetic Control Committee (FGCC) Standards and Specifications for Geodetic Control Networks.

Table 2. Final Geographic Coordinates.

Plot	Y	X	Z	Height (m)
DV05	549381.1	590924.5	273.67	1.76
DV06	549187.5	590864.3	237.61	1.74
DV07	549350.8	591590.5	300.81	1.66
DV105	550336.2	592664.2	304.63	1.56
DV107	550117.1	592332.2	340.33	1.52
DV11	549556.2	590004.3	234.23	1.34
DV117	549947.1	592081.0	323.97	1.65
DV12	549584.2	590487.5	238.02	1.58
DV144	549596.3	591350.3	256.84	1.27
DV145	549560.5	591498.4	267.58	1.28
DV17	549721.9	589916.7	244.95	1.59
DV19	550094.1	590403.4	280.54	1.77
DV20	549726.0	591559.8	276.80	1.61
DV203	550110.5	592855.3	365.43	1.68
DV204	550364.4	591965.7	269.50	1.52
DV205	550159.2	592017.9	284.95	1.46
DV207	549983.3	591159.5	352.51	1.34
DV208	549988.5	591289.1	353.56	1.66
DV23	549396.3	590604.1	256.00	1.66
DV301	549331.0	591368.7	268.64	1.64
DV302	549179.2	591473.3	298.21	1.72
DV303	549777.2	590870.5	322.37	1.59
DV304	549598.5	590973.7	300.10	1.56
DV306	549963.5	590299.6	260.96	1.59
DV306	549743.7	590358.5	266.91	1.70
DV309	549748.9	592061.4	303.52	1.60
DV31	549903.9	590449.9	275.24	1.68
DV40	549583.2	592168.1	289.68	1.61
DV42	549937.4	592733.4	351.45	1.41
DV43	550086.3	592544.2	337.38	1.77

3.3 Data calibration and strip adjustment

The coordinates of the point of the overlap strip are not the same due to differences in trajectory measurement for each strip. By comparing the points in the overlapping region, the misalignment error is corrected. The alternative is to use LiDAR Mapping Suite (LMS) tools for automated LiDAR point cloud rectification, which can remove surface features per flight line and generate geo-metrically correct point clouds. The automated system calibration includes heading, roll, pitch, and scanner mirror. The values of calibration are examined on the flight-by-flight basis. Table 3 shows the final RMS value of 0.10m.

Table 3. Report of HRPM and Z shift.

Starting RMS	0.0950
Final RMS	0.1021
Standard error	0.0453
Average magnitude	0.11765

3.4 Classification point clouds

The final data product point clouds are produced in .las format file. There are 21 flight lines scanned for the study area and produced 246,132,031 point clouds. As presented in Table 4, there are 97.6% of point clouds classified as non-ground and 2.4% as ground point clouds. Every point cloud penetrates forest structure according to the return pulse. Table 5 shows the percentage of point clouds classified into first, second, third, and fourth returns are 49.71%, 31.63%, 14.13 %, and 4.53%, respectively. In forested areas, the first return typically would come from the tree canopy, the second from the lower branches, and the third or fourth return from the ground. This study utilized all point clouds to maximize the accuracy extraction of variables of forest structure. The number of ground points in the forested areas increases greatly when first and last returns are combined (Wehr and Lohr, 1999). Moreover, the density of point clouds in this study is sufficient to derive the variables of forest structures.

Table 4. Report of classification point clouds.

Classification	**Point Count**	**Z Min**	**Z Max**	**Min Intensity**	**Max Intensity**
1 Non-Ground	240,229,435	194.06	461.94	1	2464
2 Ground	5,902,234	193.89	408.48	1	4085

Table 5. Report of point clouds returns.

Return	**Point Count**	**Z Min**	**Z Max**
First	122,362,879,	193.89	461.94
Second	77,851,278	193.9	460.84
Third	34,770,827	193.91	459.33
Fourth	11,147,047	193.95	455.66
Last	122,447,636	193.89	461.92
Single	44,570,150	193.89	461.92
First of Many	77,792,729	194.89	461.94
Last of Many	77,877,486	193.9	460.31
All	246,132,031	193.89	461.94

Figure 8 displays a histogram analysis illustrating intensity metrics reflecting the strength of the laser pulse returned from forest structures. The mean intensity values extracted from LiDAR data range from 0.68 to 76.60, spanning from the 1st to the 99th intensity percentiles. This indicates a moderate density of canopy and vegetation cover in the study area, attributed to its status as a secondary forest in the process of recovery following previous logging activities. The standard deviation of intensity reveals considerable variability in vegetation densities, canopy heights, and structural complexities, ranging from 8.68 to 31.45. The standard error recorded for intensity metric ranged from 0.14 to 0.68 This variability underscores the presence of diverse ecosystems characterized by heterogeneous vegetation composition.

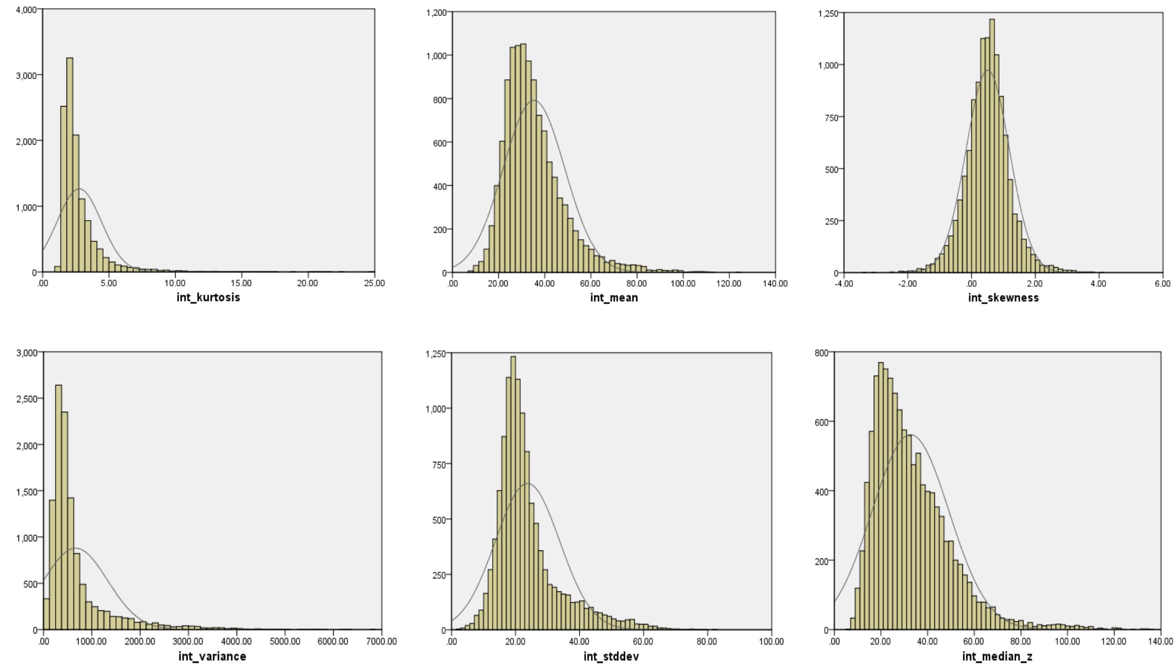


Figure 8. Histogram analysis for intensity metrics.

Figure 9 showed the scatter plots relationship between intensity metrics and percentiles. scatter plot reveals that there is a linear relationship between different percentiles and intensity metrics, it suggests a consistent pattern across the distribution. For instance, plotting mean intensity against the 50th percentile intensity reveals that there is a linear relationship between the central tendency of intensity values and their median values. it suggests a consistent pattern across the distribution. However, there are outliers that appear as data points that deviate significantly from the general pattern observed in the scatter plot. For instance, the 30th percentile intensity. These outliers could result from measurement errors, environmental factors, or unique characteristics of the scanned area. Identifying outliers is crucial for ensuring data quality for analysis.

IOP Conf. Series: Earth and Environmental Science **1412** (2024) 012005 doi:10.1088/1755-1315/1412/1/012005

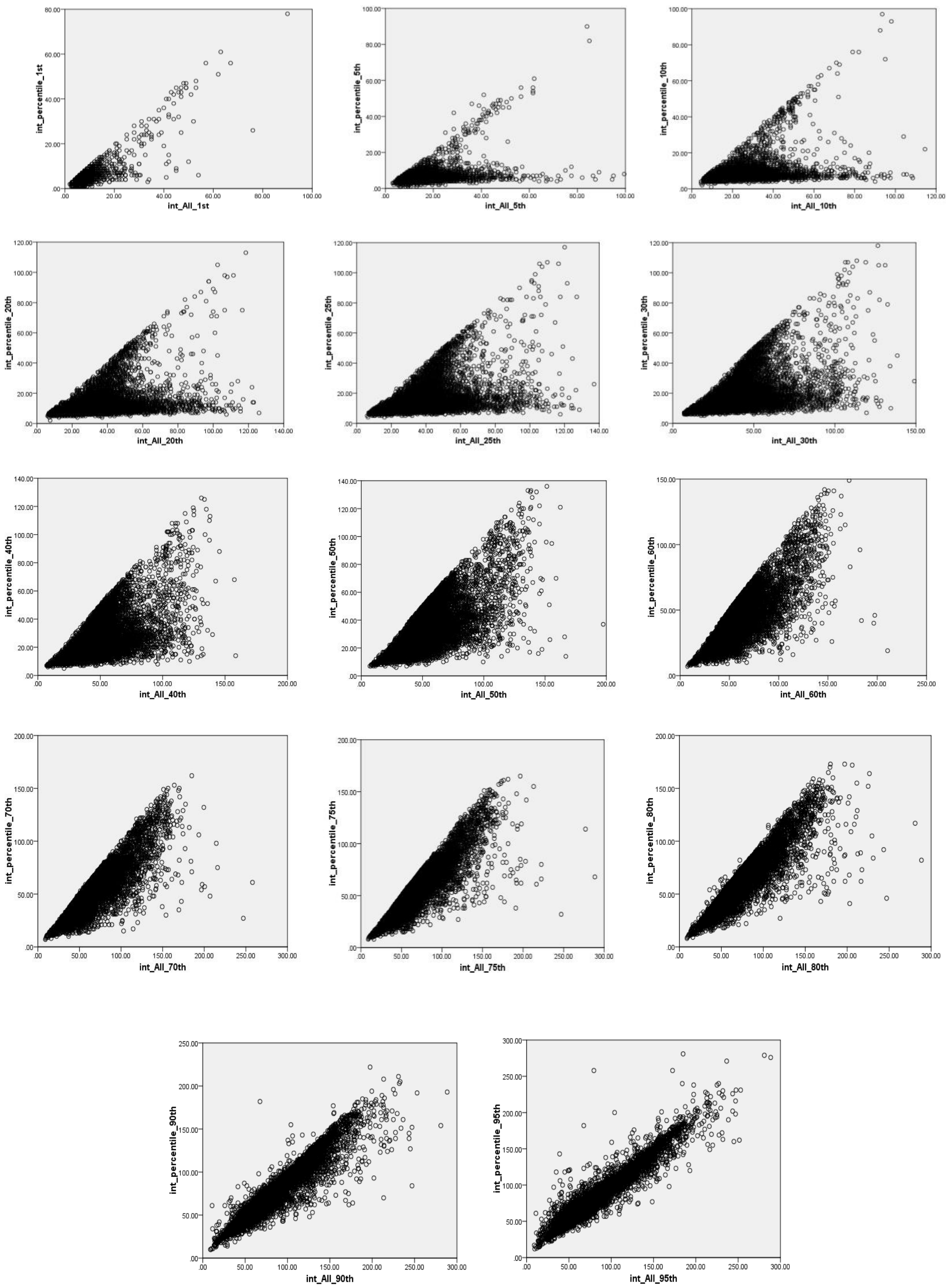


Figure 9. Scatter plots of intensity metrics of percentiles.

4. Conclusions

Processing raw LiDAR data to analyze intensity in tropical forests presents significant challenges, particularly due to variations in targets and echo types. However, the analysis of LiDAR intensity metrics is crucial for discerning forest structures, composition, and overall health, thereby enhancing sustainable forest management practices. Furthermore, monitoring LiDAR intensity metrics facilitates the creation of predictive models and decision support tools, which are invaluable for optimizing forest management operations and conserving biodiversity. By leveraging these metrics, stakeholders can make informed decisions and implement strategies that promote both ecological health and resource sustainability within tropical forest environments.

References

[1] Jamru, L. R., & Rahaman, Z. A. (2018). Combination of spatial logistic regression and geographical information systems in modelling wetland changes in Setiu basin, Terengganu. In *IOP Conference Series: Earth and Environmental Science* (Vol. 169, No. 1, p. 012106). IOP Publishing.

[2] Linda Roziani, J., Zullyadini, A. R., & Wan Ruslan, I. (2013). Permodelan perubahan tanah bencah di lembangan sungai Setiu, Terengganu menggunakan logistik regresi dan aplikasi Sistem Maklumat Geografi (GIS). *International Journal of Environment, Society and Space*, *1*(1), 75-97.

[3] Jamru, L. R., Sharil, M. N., & Yusoh, M. P. (2023). Assessing the evolution of paddy cultivation in kota belud, sabah using gis and remote sensing techniques. *Planning Malaysia*, *21*.

[4] Jamru, L. R., Sharil, M. N., Jafar, A., Eboy, O. V., Atang, C., & Talib, M. A. (2023). The effectiveness of remote sensing techniques for land use classification in Kota Belud, Sabah. *Asian Journal of Research in Education and Social Sciences*, *5*(2), 90-97.

[5] Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote sensing of environment*. 80(1), 88–99.

[6] Wehr, A., & Lohr, U. (1999). Airborne laser scanning—an introduction and overview. *ISPRS Journal of photogrammetry and remote sensing*, *54*(2-3), 68-82.

[7] Persson, Å. J. U. H., Holmgren, J., Söderman, U., & Olsson, H. (2004). Tree species classification of individual trees in Sweden by combining high resolution laser data with high resolution near-infrared digital images. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*(8), 204-207.

[8] Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, *29*(5), 1339-1366.

[9] Holmgren, J., Nilsson, M. and Olsson, H. (2003). Estimation of tree height and stem volume on plots using airborne laser scanning. *Forest Science*. 49(3), 419–428.

[10] Jamru L. R, Hashim. 2019. LiDAR Sensing for biophysical structure of vegetation in Sub-urban Kluang, Johor, International Graduate Conference of Built Environment & Surveying. Vol 1, 507-515. doi: 10.7717/peerj.4397 ISSN: 2167-8359

[11] Jamru L. R, Hashim. 2018. Treetop detection from LiDAR using local maxima algorithm. ACRS 39th (Remote Sensing Enabling Prosperity). Vol 5, pg 2692-2700

[12] Lindah Roziani Jamru. 2018. Correction pit free canopy height model derived from LiDAR data for the broad leaf tropical forest. *IOP Conf. Ser.: Earth Environ. Sci.* 169 012113

[13] Jamru, L.R. and Hashim, M., 2015. Modeling LiDAR cloud point to estimate tree height and delineation of tree crown for selected vegetation in suburban Kluang, Johor. In *36th Asian Conference on Remote Sensing: Fostering Resilient Growth in Asia*.