Analyzing canopy height variations in secondary tropical forests of Malaysia using NASA GEDI

E Adrah ¹, W S Wan Mohd Jaafar ², S Bajaj ¹, H Omar ³, R V Leite ⁴, C A Silva ⁵, ⁶, A Cardil ⁷, ⁸, ⁹ and M Mohan ¹, ¹⁰*

¹ United Nations Volunteering Program, Morobe Development Foundation, Lae 00411, Papua New Guinea
² Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia
³ Geoinformation Programme, Forest Research Institute Malaysia, 52109 Kepong, Selangor Darul Ehsan, Malaysia
⁴ Department of Forest Engineering, Federal University of Viçosa, Viçosa 36570-900, MG, Brazil
⁵ School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA
⁶ Department of Geographical Sciences, University of Maryland, College Park, MD, USA
⁷ School of Agrifood and Forestry Science and Engineering, University of Lleida, Av. de l’Alcalde Rovira Roure, 191, 25198 Solsona (Lleida), Spain
⁸ Tecnosylva, Parque Tecnológico de León, 24009 León, Spain
⁹ Forest Science and Technology Centre of Catalonia (CTFC), Ctra. Sant Llorenç de Morunys, Km 2, 25280 Solsona Lleida), Spain
¹⁰ Department of Geography, University of California—Berkeley, Berkeley, CA 94709, USA

*Corresponding E-mail: mid_mohan@berkeley.edu

Abstract. Tropical forests play a significant role in regulating the average global atmospheric temperature encompassing 25% of the carbon present in the terrestrial biosphere. However, the rapid change in climate, arising from unsustainable human practices, can significantly affect their carbon uptake capability in the future. For understanding these deviations, it is important to identify and quantify the large-scale canopy height variations arising from previous anthropogenic disturbances. With the advent of NASA GEDI spaceborne LiDAR (light detection and ranging), it is now possible to acquire three-dimensional vertical structural data of forests globally. In this study, we evaluate the applicability of GEDI for analyzing relative canopy height variations of secondary tropical forests of different age groups located across multiple geographical regions of peninsular Malaysia. The results for RH98 GEDI metric trends for the lowland and hill forests category across 4 different disturbance groups show a positive correlation between mean relative height and secondary forest ages. The consistency of these findings with previous studies in the region indicate the usefulness of GEDI to provide valuable insights into the patterns and drivers of forest height variation. Thus, this study contributes toward the operationalization of spaceborne LiDAR technology for monitoring forest disturbances and measuring biomass recovery rates and should help support large-scale sustainable forest management initiatives with respect to the tropical forests of Malaysia.
Keywords: Global Ecosystem Dynamics Investigation (GEDI), disturbed forests, Southeast Asia, forest canopy structures, logging

Track Name: Land, Water, Forests and Food Security

1. Introduction

Tropical forest ecosystems especially in Southeast Asian countries such as Malaysia face challenges from selective logging which has been a widespread modus operandi for commercial timber production [1-2]. Logging influences forest ecosystem dynamics - including soil greenhouse gas flux, spatial and temporal variations in light environment - as well as forest structural attributes such as canopy gap, canopy height, crown surface area, stem density to name a few [1-4]. A recent study by Okuda et al. (2019) investigated the potential impact of logging on canopy height recovery over selectively logged and primary forest regions in Pasoh Forest Reserve Peninsular Malaysia between 2003-2011. Herein, the authors employed airborne LiDAR (light detection and ranging) and found that average canopy height along with coefficient of variation and height diversity in logged forest was significantly lower than the primary forest plot suggesting that even after 53 years of logging, previously disturbed regions were not able to fully recover. Through simulation, the authors predicted that it would take another 16 years for full recovery from the previous canopy height mapping in 2011 [5]. Similar findings regarding mean canopy height were observed in one of Okuda’s previous studies as well which utilized aerial triangulation and photographs to measure impact of logging in a lowland dipterocarp forest in peninsular Malaysia [2].

With increasing global deforestation and degradation along with climate change, the trend of tropical forests making an approximate net neutral contribution towards the global carbon cycle may change and tropical forests may act as net carbon sources [6]. Remote sensing technologies are widely used for monitoring tropical forest biomass over time which help predict the change in carbon sequestration capability of forests experiencing climate anomalies and/or anthropogenic disturbances [7-9]. In this regard, LiDAR sensors represent an efficient option due to its capability to directly retrieve vegetation vertical structure and have been applied extensively using aerial platforms [10-11]. Nonetheless, airborne LiDAR may be limited when it is necessary to assess large areas (e.g. regional/global scale), and hence, spaceborne data may be required. The first spaceborne LiDAR sensors onboard IceSat-1 and IceSat-2 satellites were used in several forest related studies to estimate plant traits [12-13]. However, their characteristics were not developed to optimally retrieve forest characteristics, especially over dense canopies. The recently launched Global Ecosystem Dynamics Investigation (GEDI) was designed to penetrate forest with about 98% canopy cover [14] and has been used to estimate forest biomass in different ecosystems [15-17]. Recent studies also highlighted its potential for mapping forest height and obtaining disturbance history [18-19]. For instance, Potapov et. al (2021) used GEDI data with time-series optical imagery (Landsat) for monitoring forest height and its dynamics [18]. However, its potential to characterize tropical forests’ height variation of secondary forests has not been explored.

To the best of our knowledge, GEDI has yet not been employed to study forest structural attributes over the diverse tropical forests of Southeast Asia and through this pilot study we intend to explore the same. The aim of this study is to observe the trends for the chosen GEDI metric (RH98) and evaluate how post-logging forest recovery rates vary with respect to time in the various chosen geographical regions. In a prior study by Potapov et al. (2021), relative height (RH) metrics were found to be correlated with the 90th percentile of ALS-based canopy height per Landsat data pixel [18]. In this study, we compare the height distribution and growth rates of secondary forests of age groups 10-20, 20-30 and 30+ years for three different geographical regions of peninsular Malaysia and to demonstrate the applicability of GEDI for optimizing future forest management operations and supporting next generation earth system predictive models.
2. Study Area, Data and Methods

This exploratory study takes into consideration ‘logged land forests, lowland forests and hill forests’ forest category as per the classifications proposed by the Forest research institute of Malaysia (FRIM). Four pre-identified disturbance groups based on logging time were sampled across the aforementioned forest category in three different geographical zones (North, South, and West) as shown in figure 1. The logging time considered was derived from the FRIM classification and the disturbance groups taken into account are (A) logged eleven to twenty (11-20) years ago, (B) logged twenty-one to thirty (21-30) years ago, (C) logged more than thirty (>30) years ago, and (D) Protection Forests. The protected forest refers to virgin forest that has never been logged (at least in the past 100 years).

The GEDI data were downloaded and subsetted from NASA EARTH DATA through the GEDI finder web tool. Level 2A data are processed waveforms providing directly measured metrics from which canopy height metrics were extracted for each 25 m diameter GEDI footprint. Level 2B products contain calculated biophysical attributes, such as plant area index (PAI) and canopy cover, while Level 1B contains geolocated corrected waveforms and geophysical correction for all shots in the GEDI beams. The GEDI instrument produces a total of eight ground tracks by three lasers, two full-power lasers dithered into four and one coverage laser diverged and dithered to produce another four beams. The produced GEDI slots are spaced 600 m apart in the cross track direction and approximately 60 m apart along the track direction.

Ground transects used in this study were filtered and considered only the full power beams to disregard and mitigate any uncertainty in the measurement due to cloud cover, day noise and scattering. Shots with low quality or incomplete information were filtered and disregarded as well. The data were then clipped, the height metrics were extracted and grouped based on the disturbance groups. Herein, RH98, the height at which 98 percentiles of energy is reached relative to the center of the ground return, was employed as a proxy for canopy height. Concerning the comparison of canopy heights between different disturbance groups, the study area was divided into a polygon grid to have a relatively similar climate and ecological zones; this helps minimize the influence of external factors such as precipitation, soil characteristics, slope, etc. that otherwise might affect the growth and recovery rates of secondary forests. The GEDI data were then overlapped over these zones individually, and subsetted for its coverage over each disturbance group after performing standard quality assessment. Three zones (North, South and West) were considered for this study, wherein a random sample of 100 GEDI footprints from each disturbance group was available after quality assessment (totalling 400 GEDI shots per zone) and the respective descriptive statistics were then calculated. A similar analysis is done considering all the zones together with 1200 GEDI footprints in total.
Figure 1. Study areas; a) zone A (west); b) zone B (north); c) zone C (south); the blue dots represent the GEDI footprints.
3. Results and Discussion

The results for the variation of the RH98 metric are divided into 2 sections - first, across the 4 disturbance groups (A, B, C and D) considering all the 3 chosen zones (Zone A, Zone B, Zone C) together and second considering each zone separately. The results for the former are presented in figure 2 whereas for the latter are presented in figure 3. Various parameters (min, max, median, mean, interquartile range) relevant to the box plots are enumerated in table 1 and table 2 respectively.

From Table 1, we observe a positive correlation between relative mean RH98 and secondary forest age groups, when all the zones are considered together. The upper and lower whisker extends from lowest to the highest value respectively within the interquartile range between 16.4 m to 38.83 m. As expected, protected forests had the highest average and group A (age group 11-20 years) had the lowest average for RH98. Regarding heights variations (interquartile span) within the group, it was the highest for group B (21-30 years) and lowest for group A (11-20 years).

| Disturbance group | All zones |
|-------------------|-----------|
|                   | Min | 1st Quartile | Median | Mean | 3rd Quartile | Max | Interquartile Range |
| A (11 - 20 Years) | 1.6 | 16.4 | 23.04 | 23.14 | 29.92 | 52.07 | 13.52 |
| B (21 - 30 Years) | 2.99 | 13.52 | 22.94 | 23.73 | 31.88 | 65.19 | 18.36 |
| C (> 30 Years)    | 2.54 | 17.95 | 25.77 | 25.66 | 33.05 | 65.95 | 15.1  |
| D Protected Forests | 3.17 | 22.2 | 30.22 | 30.59 | 38.83 | 72.2  | 16.63 |

Table 1. Summary statistics of height variations (for all the zones combined).
Figure 2. Boxplots representing the height variations across various disturbance groups while considering all the three zones together.

The results for the variation of RH98 metric across the different disturbance groups considering the 3 zones individually are presented in table 2 and figure 3. Positive correlation patterns are also observed between mean RH98 and secondary forest age groups and these trends persist across all zones except for the western zone (zone A) where the group A (11-20 years) have a slightly higher mean value than the group B (21-30 years) (figure 3). This can be due to differences in logging intensities experienced and/or the type of forest species predominant in these forest regions and will require further investigation before making any conclusions.

Table 2. Summary statistics of height variations (for individual zones).

| Zone     | Disturbance group | Min | 1st Qu | Median | Mean  | 3rd Qu | Max  | Interquartile Range |
|----------|-------------------|-----|--------|--------|-------|--------|------|---------------------|
|          | (11 - 20 Years)   | 2.77| 12.43  | 21.64  | 21.26 | 27.57  | 47.76| 15.14               |
| Zone A   | (21 - 30 Years)   | 2.99| 11.67  | 20.25  | 20.46 | 26.79  | 54.57| 15.12               |
|          | (> 30 Years)      | 4.08| 16.24  | 20.95  | 23.37 | 29.72  | 52.95| 13.48               |
|          | Protected Forests | 6.06| 22.06  | 30.46  | 30.01 | 37.02  | 57.99| 14.96               |
| Zone B   | (11 - 20 Years)   | 4.94| 17.93  | 23.69  | 24.36 | 30.52  | 52.07| 6.83                |
|          | (21 - 30 Years)   | 4   | 14.38  | 25.18  | 25.44 | 35.25  | 58.75| 10.07               |
Another explicit observation which can be inferred from the results show that the relative heights metric for the protected forest category have the highest values across all zones when compared to disturbance groups A, B and C. In other words, the RH98 metric values for all disturbance categories are well below when compared to that of protected forest when all the zones are considered, individually and together. This implies that the forests have yet not fully recovered to their original potential. In sum, the results based on the RH98 GEDI metric are consistent with previous studies in the region which took into account the actual canopy height and show a similar trend with respect to the recovery status [2,5].

Concerning heights variation across different geographical zones, as interpreted from table 2, the lowest mean values across all disturbance categories is recorded in the western zone (zone A), while the highest mean across all disturbance categories is recorded in the northern zone (zone B). Northern zone (zone B) is also the one with lowest heights variation (interquartile range) across all disturbance groups. These observations could be explored further to understand the underlying drivers of these variations with more background information about forest disturbances and ecological variants in each zone which we plan to do in our future studies.

**Figure 3.** Boxplots representing the height variations across various disturbance groups considering each zone individually.
Next, the inter-decadal difference was analyzed between the mean relative height (represented by the GEDI metric RH98) of different disturbance groups A, B and C for the 3 zones together and individually. These are presented in table 3. The range for protected forest, as shown in table 3, has little variation across the different geographical zones indicating that the protected forests are near their maturity across the 3 zones. Additionally, we observed that the difference between the inter-decadal relative height mean for disturbance group C & B (i.e. C-B) is greater for Zone A and Zone B when compared to B & A (i.e B-A). Nonetheless, the results observed are reversed for Zone C. The observed results for the zones could be possible due to a number of reasons which we intend to explore in our future studies. For instance, the intensity of logging may vary with time which may ultimately influence the mean relative height. Also, the distribution of age groups within the disturbance groups may not be uniform. For example, the year when logging occurred in a particular zone within different decades could be concentrated at extreme ends of the decade. Thus, in our future studies, we intend to utilize Landsat imagery to gain insights into intensity of logging and exact period when logging occurred. Information about the species specific for the zones could also be gathered to comprehend the results in a better way and this could help us understand how the recovery rate varies across species types.

**Table 3. Height differences (in m) between secondary forests of different age groups and zones.**

| RH98 metric difference across disturbance groups | All zones together | Zone A | Zone B | Zone C |
|------------------------------------------------|-------------------|--------|--------|--------|
| B (21-30 years) – A (11-20 years)              | 0.59              | -0.8   | 1.08   | 1.52   |
| C (>30 years) – B (21-30 years)               | 1.93              | 2.91   | 2.16   | 0.69   |

In subsequent research, calibrating GEDI data using airborne LiDAR data and furthermore, integrating it with Landsat data would be implemented for generating wall-to-wall canopy height maps. In addition, with further availability of new (version 2) and more GEDI data, a similar sampling technique could be conducted to chart the average mean heights against longitude and latitude taking vertical and horizontal tracks covering the extent of peninsular Malaysia. For instance, figure 4 shows the average heights across longitude and latitude for the entire Malaysia peninsular, taking 5000 GEDI shots over the ‘logged land forests, lowland forests and hill forests’ forest category grouped per disturbance groups and sampled equally taking strip polygon grids with 5 km width in each: longitude and latitude directions. This sampling technique could be improved with the availability of more GEDI data and considering further information about the environmental gradients, climate variables, and species, to gain insights into canopy height variation across large landscapes. We also intend to extend our approach to the entire Malaysia and various other forest types and explore the patterns and drivers of canopy height variations by considering the influence of climate variables and environmental gradients.
4. Conclusion

Through this preliminary study we have attempted to explore the relative trends in height variations using the RH98 GEDI metric for the lowland and hill forests category of peninsular Malaysia across 4 different disturbance groups (based on years after logging had happened). Our focus area is narrowed down into 3 different geographical zones after application of various steps such as quality assessment for selecting filtered highest quality GEDI data, and inclusion of a minimum of 100 shots for each type of disturbance group in each zone. The RH98 metric shows a positive correlation between mean relative RH98 and the secondary forest ages and also affirm the trend of protected forests possessing greater relative heights than logged forest categories. The consistency of these patterns with the results of previous studies in the region demonstrate the GEDI capabilities to provide insights on height variation patterns. We also find that the inter-decadal difference between the mean values of the RH98 metric varies spatially. This could be due to several reasons such as varying levels in intensity of logging across zones, non-uniformity in distribution of logging time period within each distribution group to name a few. In our future studies, we intend to explore these variations in recovery rates and height distribution as a function of environmental gradients and climate variables (such as slope and precipitation) and will be extending our analysis to other forest types. In sum, our approach demonstrates the potential of GEDI metrics to quantify the recovery and growth rates of secondary forests which can consequently be useful for estimating aboveground biomass (AGB) and understanding the impact of climate change and anthropogenic activities on tropical forest carbon sequestration capacity.

References

[1] Numata S et al 2006 Canopy gap dynamics of two different forest stands in a Malaysian lowland rain forest Journal of Tropical Forest Science 1 109-16.

[2] Okuda T et al 2003 Effect of selective logging on canopy and stand structure and tree species composition in a lowland dipterocarp forest in peninsular Malaysia Forest ecology and management 175 (1-3) 297-320

[3] Yashiro Y et al 2008 The effects of logging on soil greenhouse gas (CO2, CH4, N2O) flux in a tropical rain forest, Peninsular Malaysia. Agricultural and forest meteorology 148 (5) 799-806

[4] Yamada T et al 2014 Spatial and temporal variations in the light environment in a primary and selectively logged forest long after logging in Peninsular Malaysia. Trees 28 (5) 1355-65

[5] Okuda T et al 2019 Canopy height recovery after selective logging in a lowland tropical rain forest. Forest ecology and management 442 117-23.

[6] Mitchard ET 2018 The tropical forest carbon cycle and climate change Nature 559 (7715):527-34.
[7] Sanchez-Azofeifa A et al 2017 Twenty-first century remote sensing technologies are revolutionizing the study of tropical forests Biotropica 49 (5) 604–19
[8] Wan Mohd Jaafar WS et al 2020 Carbon Emissions from Oil Palm Induced Forest and Peatland Conversion in Sabah and Sarawak, Malaysia Forests 11 (12) 1285
[9] Wan Mohd Jaafar WS et al 2020 The influence of deforestation on land surface temperature—a case study of Perak and Kedah, Malaysia Forests 11 (6) 670
[10] Dubayah RO et al 2010 Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica: Forest Dynamics Using Lidar. J Geophys Res [Internet]. 115 (G2)
[11] Silva C et al 2017 Impacts of airborne lidar pulse density on estimating biomass stocks and changes in a selectively logged tropical forest. Remote Sens (Basel) 9 (10) 1068
[12] Duncanson L1 et al 2010 Estimating Forest canopy height and terrain relief from GLAS waveform metrics Remote Sens Environ 114 (1) 138–54
[13] Narine LL et al 2020 Using ICESat-2 to estimate and map forest aboveground biomass: A first example Remote Sens (Basel) 12 (11) 1824.
[14] Dubayah R et al 2020 The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography Science of remote sensing 1 (1) 100002
[15] Silva CA et al 2021 Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. Remote Sens Environ 253 112234
[16] Duncanson L et al 2020 Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California Remote Sens Environ 242 111779
[17] Rishmawi K Huang C and Zhan X 2021 Monitoring key forest structure attributes across the conterminous United States by integrating GEDI LiDAR measurements and VIIRS data Remote Sens (Basel) 13 (3) 442
[18] Potapov P et al 2021 Mapping global forest canopy height through integration of GEDI and Landsat data Remote Sensing of Environment 1 (253) 112165
[19] Sanchez-Lopez N et al 2020 Estimating time since the last stand-replacing disturbance (TSD) from spaceborne simulated GEDI data: A feasibility study Remote Sens (Basel) 12 (21) 3506

Acknowledgments

We would like to thank the Forest Research Institute Malaysia (FRIM) for providing a complete dataset from field data to LiDAR data to other supplementary materials and resources. We also thank members of the Morobe Development Foundation (Papua New Guinea), UN Volunteering Program and NGEE-Tropics (LBNL) for their support and guidance. The authors gratefully acknowledge the funding for this open access publication and related proceedings of the presented work provided under UKM research grants Geran Galakan Penyelidik Muda (GGPM-2020-034) and Ganjaran Penerbitan (GP-2020-K022872).