



## LiDAR patch metrics for object-based clustering of forest types in a tropical rainforest



Cici Alexander<sup>a,d,\*</sup>, Amanda H. Korstjens<sup>a</sup>, Graham Usher<sup>b</sup>, Matthew G. Nowak<sup>b,c</sup>, Gabriella Fredriksson<sup>b</sup>, Ross A. Hill<sup>a</sup>

<sup>a</sup> Bournemouth University, Department of Life and Environmental Sciences, Talbot Campus, Poole, Dorset, BH12 5BB, United Kingdom

<sup>b</sup> The PanEco Foundation - Sumatran Orangutan Conservation Programme, Chileweg 5, Berg am Irchel 8415, Switzerland

<sup>c</sup> Southern Illinois University, Department of Anthropology, 1000 Faner Drive, Carbondale, IL 62901, USA

<sup>d</sup> Aarhus University, Aarhus Institute of Advanced Studies (AIAS), Høegh-Guldbergs Gade 6B, DK-8000 Aarhus C, Denmark

### ARTICLE INFO

#### Keywords:

Sumatra  
Batang Toru  
Canopy Height Model  
Classification  
ALS  
Habitat

### ABSTRACT

Tropical rainforests support a large proportion of the Earth's plant and animal species within a restricted global distribution, and play an important role in regulating the Earth's climate. However, the existing knowledge of forest types or habitats is relatively poor and there are large uncertainties in the quantification of carbon stock in these forests. Airborne Laser Scanning, using LiDAR, has advantages over other remote sensing techniques for describing the three-dimensional structure of forests. With respect to the habitat requirements of different species, forest structure can be defined by canopy height, canopy cover and vertical arrangement of biomass. In this study, forest patches were identified based on classification and hierarchical merging of a LiDAR-derived Canopy Height Model in a tropical rainforest in Sumatra, Indonesia. Attributes of the identified patches were used as inputs for k-medoids clustering. The clusters were then analysed by comparing them with identified forest types in the field. There was a significant association between the clusters and the forest types identified in the field, to which arang forests and mixed agro-forests contributed the most. The topographic attributes of the clusters were analysed to determine whether the structural classes, and potentially forest types, were related to topography. The tallest clusters occurred at significantly higher elevations (> 850 m) and steeper slopes (> 26°) than the other clusters. These are likely to be remnants of undisturbed primary forests and are important for conservation and habitat studies and for carbon stock estimation. This study showed that LiDAR data can be used to map tropical forest types based on structure, but that structural similarities between patches of different floristic composition or human use histories can limit habitat separability as determined in the field.

### 1. Introduction

Tropical rainforests are relatively poorly understood, despite supporting a large proportion of the Earth's plant and animal species within a restricted global distribution, and influencing the global carbon and hydrological cycles (Corlett, 2016; Lawrence and Vandecar, 2014; Thomas and Baltzer, 2001). Human activities such as logging and clearing of forests for agriculture and agro-forestry continue to alter their extent and composition, leading to fragmentation of habitats. This makes it difficult for species, including those on the brink of extinction, to move to safer or more suitable locations in response to the destruction of their habitats or changes in climate (Corlett, 2016; Whitmore, 1990). Mapping of tropical rainforests is a pre-requisite for a better understanding of these resources so that fragmentation can be monitored, and strategies for conservation can be devised.

Tropical rainforests are also considered to play an important role in regulating the Earth's climate by being a large sink for carbon dioxide. An accurate estimation of carbon components within a forest is a first step in the recent United Nations initiative for Reducing carbon Emissions from Deforestation and forest Degradation (REDD). However, limited knowledge about the quantity and spatial distribution of biomass at the landscape level has led to considerable uncertainties in the estimation of carbon stocks (Rödig et al., 2018). Even within the same region, there can be large variations in carbon stock based on factors such as topography, geology and natural or anthropogenic structural changes (Laumonier et al., 2010). Accurate maps at the landscape level could provide information about the different forest types so that the variability of biomass among the different forest types is taken into consideration for estimations of carbon stock.

Remote sensing, based on satellite imagery, has been extensively

\* Corresponding author at: Aarhus Institute of Advanced Studies (AIAS), Aarhus University, Høegh-Guldbergs Gade 6B, DK-8000 Aarhus C, Denmark.  
E-mail address: [cici@aias.au.dk](mailto:cici@aias.au.dk) (C. Alexander).

used for large-scale mapping and monitoring of global forest cover (Hansen et al., 2013; Kim et al., 2014; Sexton et al., 2015). Field-based surveys of habitats or forest types could be more difficult for tropical forests than temperate, in terms of access and a potential imbalance between high species diversity and limited knowledge of taxonomy. Remote sensing can be an efficient source of information for mapping tropical forests, and to identify forest types for more detailed field surveys (Moran et al., 1994; Salovaara et al., 2005). Broad tropical forest types have been successfully classified using medium-resolution optical imagery (Hill and Foody, 1994), although the availability of cloud-free data and topographic effects may pose problems for passive remote sensing (Asner, 2001; Miettinen et al., 2014).

Airborne Laser Scanning (ALS), based on the active remote sensing technique of Light Detection and Ranging (LiDAR), has advantages over other remote sensing techniques especially in the case of forests. ALS data are less influenced by cloud cover than satellite data since the flying altitude can be below the cloud layer. ALS provides information about the three-dimensional structure of forests, which provides an additional perspective of the habitat needs and requirements of species compared to two-dimensional satellite imagery (Bergen et al., 2009; Coops et al., 2016). Information about heights can be derived from stereo-imagery, but their use may be limited in dense tropical forests, where the ground is obscured by vegetation or shadows. ALS is considered to be the best suited for generating terrain models in forests due to its ability to collect data from the forest floor, which is not possible with passive optical remote sensing. Terrain models are required to estimate canopy height and structure, which could be used for classifying forest types (Wulder et al., 2008), and subsequently assessing their carbon content (Asner et al., 2018; Mohd Zaki et al., 2018).

Tropical forests have been classified based on elevation (lowlands, hills, sub-montane, montane), soil type (clayey, sandy), location with respect to streams or rivers (water-logged, floodplain, terra firma), and corresponding differences in structural characteristics. These structural characteristics, including canopy height, canopy cover and density of under-storey vegetation (Foody and Hill, 1996; Laumonier, 1997; Thomas and Baltzer, 2001), can be derived from ALS data, and in turn could be used for the classification of tropical forest types, but has been less explored. Kennel et al. (2013) undertook a technical analysis of ALS derived height distribution statistics and texture measures (Haralick, Fourier transform-based, and wavelet-based features) with different classification approaches (linear discriminant analysis, random forest and support vector machine), to map small sample areas representing five structurally different but low-growth forest types in French Guiana. For boreal forests in Norway, Sverdrup-Thygeson et al. (2016) used 30 vertical ALS metrics, based on the laser height distribution, and 25 horizontal metrics, based on the crown characteristics and the distribution of canopy gaps, to classify near-natural and old-growth forests. However, with the exception of Ioki et al. (2016), who assessed tree community composition using 16 ALS variables for a disturbed forest in Borneo, nothing similar has been attempted with tropical rainforests.

Classification algorithms are used to group grid cells/pixels into labelled classes when accurate field data are available for all the major land cover types in the landscape, while clustering algorithms can be used to group unlabelled data. Heterogeneity within forest types can reduce the accuracies in pixel-based classifications. Large pixel sizes, low-pass filtering and segmentation have been used to reduce noise in landscape-level mapping of vegetation. However, large pixel sizes may result in multiple forest types within a pixel, while low-pass filtering may blur the boundaries reducing classification accuracies (Hill, 1999; Hou et al., 2013; Salovaara et al., 2005). Patches of different canopy structure and composition are often visible in Canopy Height Models derived from ALS data, which may be difficult to classify using a pixel-based approach. Horizontal spatial patterns resulting from changes in land cover, forest type or habitat can be characterised by patch or landscape metrics (Turner et al., 2001).

The composition and structure of trees occurring in forest patches, generated due to natural or anthropogenic causes, can have an influence on the associated animal species. For example, bald-faced saki monkeys (*Pithecia irrorata*) in the Peruvian Amazon have a preference for sites with the tallest, most homogeneous canopies (Palminteri et al., 2012). Bornean agile gibbons (*Hylobates albibarbis*) are considered to select continuous canopies over discontinuous, and higher canopies over lower for locomotion, often following established routes through trees referred to as ‘arboreal highways’ (Cheyne et al., 2013; Chivers, 1974). For three sympatric primate species, white-faced capuchin monkeys (*Cebus capucinus*), mantled howler monkeys (*Alouatta palliata*), and black-handed spider monkey (*Ateles geoffroyi*), in a tropical forest at Barro Colorado Island, movement patterns were shown to correlate with canopy height and distance to gaps, which themselves were related to forest maturity and lateral connectivity (McLean et al., 2016). Using ALS data in Central Kalimantan, Bornean agile gibbons and red langurs (*Presbytis rubicunda*) were shown to have almost non-overlapping habitats for sleeping, ranging and feeding, with both forest structure variables (e.g. canopy height) and landscape-scale measures relating to anthropogenic disturbance (e.g. distance to forest edge or burnt forest) explaining habitat suitability patterns (Singh et al., 2018).

The three-dimensional structure of forests, with respect to the habitat requirements of different species, can be defined by just three attributes—canopy height, canopy cover and vertical arrangement of biomass (Coops et al., 2016; Guo et al., 2017). In this study, we limit the attributes derived from ALS data to these three broad categories, even though it is possible to derive a large number of metrics from an ALS point cloud, using the heights of points within a grid cell (McGaughey, 2015). Thresholds for estimating canopy cover in temperate, boreal or managed forests may not be suitable for tropical forests, the trees being dense and multi-layered, and optimum thresholds have to be determined.

The aim of this study was to identify and characterise forest patches based on structural attributes derived from ALS data in a tropical rainforest in Sumatra, Indonesia. The first objective was to develop a method to identify homogeneous forest patches based on a Canopy Height Model; generation of canopy attributes using a Canopy Height Model is less complex than from the ALS point cloud. The second objective was to determine whether forest types identified in the field corresponded to structural classes based on ALS data. Land cover changes in tropical forests have been observed to have a correlation with elevation, with increased deforestation at lower elevations for settlement, agriculture and plantations (Kanade and John, 2018; Putz et al., 2018). Global tree cover is also correlated with slope in areas with a history of anthropogenic land cover changes, where steep ground acts as refuge for trees (Sandel and Svenning, 2013). The third objective was therefore to analyse whether the natural clusters based on the canopy attributes were related to topography.

## 2. Study Area and Datasets

The study area is in Batang Toru (1° 52'N, 99° 3'E), in the Indonesian province of North Sumatra (Fig. 1). The Batang Toru forests are home to a number of unique plant and animal species including the newly identified (Nater et al., 2017) and critically endangered Tapanuli orangutan (*Pongo tapanuliensis*), Malayan tapirs (*Tapirus indicus*) and Sumatran tigers (*Panthera tigris sumatrae*); it is the southernmost known habitat of orangutans in Sumatra, with the largest known population south of Lake Toba (Wich et al., 2016). The study area covers approximately 162 km<sup>2</sup>, where a history of logging and clearing of land for agro-forestry, selective logging to establish “forest gardens” and natural dynamics have created a mosaic of forest patches.

ALS data were collected between 23<sup>rd</sup> March and 4<sup>th</sup> April 2015, using a Leica ALS-70 HP LiDAR system from a fixed wing aircraft. The flying height was between 900 m and 1350 m above ground level, and the scan half angle was 22.5°. This generated an ALS point cloud with

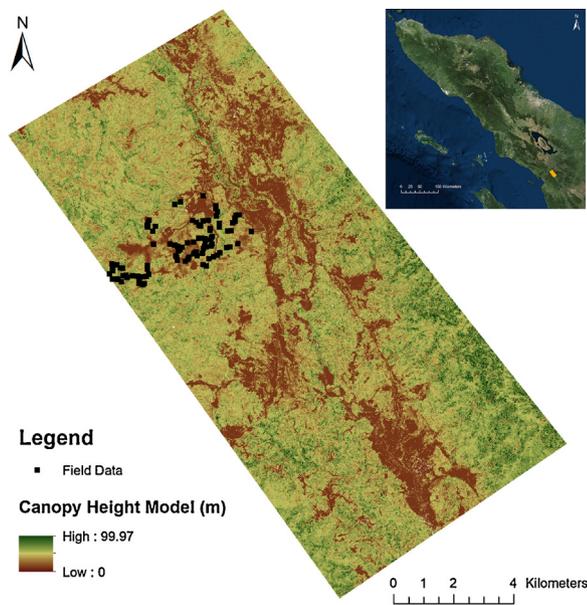


Fig. 1. Location of field data within the study area and location of the study site in North Sumatra (inset); Base map from ArcMap™ 10.1.

an average density of 12 returns  $m^{-2}$ . The returns were classified into ground and non-ground using an algorithm based on adaptive TIN filtering implemented in Terrasolid software, and divided into 240 ( $1\text{ km} \times 1\text{ km}$ ) tiles for both the ground returns and non-ground returns (Axelsson, 2000; PTMcElhanney, 2015). The software also identified error or noise points within the data (labelled as Class 7 in LAS format), which were subsequently removed from further analysis.

Field data were collected between 16<sup>th</sup> and 21<sup>st</sup> November 2013, and between 25<sup>th</sup> July and 16<sup>th</sup> August 2016 by the surveying team of the Sumatran Orangutan Conservation Programme (SOCP). The field data were collected for a rapid assessment of habitat types for an environmental management project, and covered less than 10% of the study area. There were 505 data points classified into primary, ‘Kayu Arang’, secondary/disturbed forests and mixed agro-forests.

- Primary forests are relatively undisturbed with no history of logging. The dominant family of trees present is the Dipterocarpaceae, with individual trees potentially growing to over 60 m in height.
- ‘Kayu Arang’ (hereafter referred to as arang) forests form a natural, edaphically-determined, low-growth forest of high conservation value. Within this area of Batang Toru, arang forest is characterised by a relatively low level of diversity, with woody species such as *Vaccinium heterophylla*, *Rhododendron* spp. and *Rhodoleia championii* being particularly prevalent.
- Secondary/disturbed forests have been logged in the past, and have regenerated largely through natural processes.
- Agro-forests or “forest gardens” in the study area resulted from selective logging and planting of trees such as rubber (*Hevea brasiliensis*), benzoin (*Styrax benzoin*) and durian (*Durio zibethinus*), the latter being a fruit tree.

### 3. Methods

#### 3.1. Generation of grid-level metrics from ALS point cloud data

Although descriptive statistics (metrics) at a resolution of 1 m would provide a detailed description of the study area, the scale may not be suitable for classifying habitat types. A minimum cell size of 15 m is recommended for computing ALS metrics for forests (McGaughey, 2015), and thus a cell size of 20 m was used in this study to better match the tile size of  $1\text{ km} \times 1\text{ km}$ . Note that the metrics were

generated with a cell size of 1 m, and later aggregated to 20 m (Section 3.3).

The classified ground returns had an average density of 1–2 returns  $m^{-2}$ . A Digital Terrain Model (DTM) with a grid cell size of 1 m was generated from the ground returns in FUSION software (McGaughey, 2015). The average elevation of all ground returns within a cell was used to compute the elevation of each cell, and the cells that did not contain any returns were filled by interpolation. The ground and non-ground returns were merged, and the 95<sup>th</sup> percentile height of returns within a cell size of 1 m was generated, with the elevation of returns and the DTM as the inputs. The 95<sup>th</sup> percentile height was used instead of the maximum to exclude outliers. The generated output will be referred to as the Canopy Height Model (CHM) in the following sections.

#### 3.2. Segmentation of the Canopy Height Model into forest patches

The CHM was used to generate homogeneous forest patches. The CHM, with a cell size of 1 m was reclassified into a binary raster in ArcGIS™ 10.1 with height cut-off at 5 m, to separate the forest and non-forest areas. This threshold was based on the definition of forests by the Food and Agriculture Organisation of the UN (FAO). All the grid cells in the CHM with height less than 5 m would now have a value of 0, and those equal to or above 5 m would have a value of 1. This binary raster was then aggregated to 20 m cells, by adding the values of the 1 m cells (equivalent to counting the number of 1 m cells with height equal to or above 5 m within a 20 m cell), and the values ranged from 0 to 400. This 20 m raster was again reclassified into a binary raster with a value of 1 if at least 75% (or 300) of the 1 m grid cells had a height above the 5 m threshold. Eight more similar reclassified raster layers were generated with height thresholds from 10 m to 45 m, at 5 m intervals.

These nine binary raster layers were then added together, generating a single output with values in the range 0–9; where a value of 0 represents 20 m cells that do not have at least 75% of 1 m cells with a 95<sup>th</sup> percentile height > 5 m and a value of 9 represents cells that have at least 75% of 1 m cells with a 95<sup>th</sup> percentile height > 45 m. This raster layer was converted to polygons for further processing, such that neighbouring cells of the same class were grouped into a single spatial feature. Their areas were computed, and the polygons with an area less than ca. 0.25 ha were reallocated to the surrounding class with a lower value. Since the polygons were generated from  $20\text{ m} \times 20\text{ m}$  cells, this translated to reallocating polygons with an area less than or equal to  $2400\text{ m}^2$  (i.e. six  $20\text{ m} \times 20\text{ m}$  cells).

The above process was done sequentially starting from cells with a value of 9. A few iterations had to be done at each step to remove all the small polygons, since the initial merging could still generate polygons with an area less than 0.24 ha. Those small polygons that were not surrounded by cells with the next lowest height threshold were not merged in this process, but were merged with the polygons with the largest shared border as the final step, to generate homogeneous forest patches.

#### 3.3. Generation of patch-level metrics

The nine binary 1 m raster layers, with height thresholds from 5 m to 45 m, produced in the previous step were used to generate patch-level CHM-based metrics. The percentage of 1 m cells in each binary layer was computed for each patch. This provided an estimate of canopy cover for each height threshold (CanCov5, ..., CanCov45), based only on the CHM, which itself was based on the 95<sup>th</sup> percentile height per 1 m grid cell. Within each patch, the mean and standard deviation of the 95<sup>th</sup> percentile values (HtMean and HtStd) were calculated, with the 1 m CHM and patch polygons as inputs. Thus there were 11 variables in total based on the CHM, out of which nine were canopy cover at different height thresholds. The canopy cover variables for clustering were selected based on their inter-correlations.

**Table 1**  
Attributes generated from the Canopy Height Model.

No:	Variable	Description
1	HtMean	Mean height based on the Canopy height Model
2	HtStd	Standard deviation of height based on the Canopy height Model
3–11	CanCov5, CanCov10, ..., CanCov40, CanCov45	Canopy cover at height thresholds of 5–45 m at 5 m intervals based on the Canopy Height Model (9 attributes)

### 3.4. Clustering of the patches using k-medoids

The forest patches were treated as separate objects, without any pre-assigned class, for clustering based on their structural attributes (Table 1). K-medoids was selected over the more popular k-means for clustering the patches since it uses actual data points as centroids, instead of virtual points as in the case of k-means. K-medoids is also considered to be more robust to noise and outliers than k-means. The optimal number of clusters was selected using the Calinski-Harabasz clustering evaluation criterion, an index using a ratio of a between-cluster-means and a within-cluster-sum-of-squares statistic (Caliński and Harabasz, 1974; de Amorim and Hennig, 2015).

### 3.5. Analysis of the clusters

ANOVA (one-way analysis of variance) was used to test for differences in the LiDAR structure variables between the derived k-medoids clusters, using Scheffe's procedure for post hoc pair-wise comparisons. Box-and-whisker plots of the attributes grouped into the derived k-medoids clusters were generated for visual analysis. The importance of the individual attributes for the classification was estimated using a Random Forest classifier. Height profile transects were extracted across the study site covering different clusters to understand their structural differences.

Cluster labels from the k-medoids output were extracted for all points corresponding to the field data collection, and a correspondence between cluster names and forest types identified in the field was assessed using Chi-square tests. Tree species composition influences the structural composition of forest patches, which in turn is influenced by factors including topography. ANOVA was used to test whether there were significant differences between the elevations and slopes occupied by the different clusters. All statistical analyses were performed in MATLAB, with  $\alpha$  set to 0.001.

## 4. Results

### 4.1. Identification of forest patches

The terrain elevation at the field site ranged from 453 m to 1417 m. The output from combining the nine different canopy cover maps, based on the CHM, had 101,816 patches, which were merged to generate 7994 forest patches (Fig. 2). There were 413 non-forest patches which covered a total area of approximately 36 km<sup>2</sup> of the 162 km<sup>2</sup>.

### 4.2. Analysis of the clusters

The number of canopy cover variables was reduced from nine to five, at 10 m intervals, since adjacent canopy cover variables based on 5 m height intervals were highly correlated (> 90%). Grouping the 7581 forest patches using the seven variables (HtMean, HtStd, CanCov5, CanCov15, CanCov25, CanCov35 and CanCov45) into six clusters produced the highest Calinski-Harabasz index (Fig. 3), and six was therefore selected as the number of groups for k-medoids clustering.

Forests covered a major part (77.76%) of the study area. The number of patches in each of the six forest clusters were 1414 (25.11 km<sup>2</sup>), 1802 (40.93 km<sup>2</sup>), 1072 (10.55 km<sup>2</sup>), 267 (1.76 km<sup>2</sup>), 800 (8.54 km<sup>2</sup>) and 2226 (39.30 km<sup>2</sup>) respectively. The mean canopy heights of the six clusters were 25.27 ± 2.19 m, 20.54 ± 1.84 m, 12.17 ± 1.85 m, 40.03 ± 5.84 m, 30.03 ± 2.31 m and 16.09 ± 1.94 m respectively. High elevations occurred in the north-west and south-east of the study area, with the high elevations in the south-east also associated with steep slopes. The tallest clusters—Cluster 4 and Cluster 5—predominantly occurred in the south-east, while the shortest cluster—Cluster 3—was almost absent there (Fig. 4).

The mean and standard deviation of heights and canopy cover at 5, 15, 25, 35 and 45 m were significantly different between the six clusters (all  $p < 0.001$ ;  $F_{5, 7575} = 12,797.56$ ;  $F_{5, 7575} = 150.03$ ;  $F_{5, 7575} = 1433.9$ ;  $F_{5, 7575} = 13,265.36$ ;  $F_{5, 7575} = 17,986.5$ ;  $F_{5, 7575} = 8521.41$ ;  $F_{5, 7575} = 2060.57$ ). When the clusters were compared pairwise, all differences were significant for mean height and canopy cover at 25 m, but the other variables (standard deviation and canopy cover at 5 m, 15 m, 35 m and 45 m) varied between clusters in their separability (Table 2).

Cluster 3 could be separated from all the other clusters based on canopy cover at 15 m, although there was some overlap with Cluster 6. Cluster 4 and Cluster 5 could be separated from all other clusters, except Cluster 1, using canopy cover at 25 m (Fig. 5A–G). Canopy cover values at 15 m and 25 m were also the most important predictors when a Random Forest classifier was used to estimate the importance of the variables for the classification (Fig. 5H).

The canopy of Cluster 3, the shortest cluster, seems to be open, based on the few sample plots, with many points close to the ground (Fig. 6A & B). Although the mean canopy height of Cluster 4 is 40.03 m, there are patches with heights well above 60 m with different canopy layers visible in their height profiles (Fig. 6C). Gridding into 20 m cells and merging small polygons to generate forest patches inevitably caused overlaps between clusters, especially at the boundaries (Fig. 6).

### 4.3. Comparison with identified forest types

The number of locations identified in the field for primary, arang, secondary/disturbed forests and mixed agro-forests were 91, 84, 171 and 159 respectively. None of the field data points occurred in Cluster 4. There were only three locations in Cluster 5, and they were excluded from the significance test, since the expected values were less than 5. There was a significant association between the four remaining clusters—Cluster 1, Cluster 2, Cluster 3 and Cluster 6—and the identified forest types:  $\chi^2(9, N = 502) = 81.33, p < 0.001$ . This was mainly due to the associations between arang forests and Cluster 3, and between mixed agro-forests and Cluster 6 (Fig. 7A). Arang forests seemed to have an agreement with Cluster 3 (52.63%), which also had the lowest mean canopy height among the six clusters. Cluster 6 is mostly mixed agro-forest (41.82%). Secondary/disturbed forests occur in the highest proportions in Cluster 1 (47.37%) and Cluster 2 (38.03%). All the points in Cluster 5 belonged to primary forests, although it was only 3.3% of this field type (Fig. 7B).

### 4.4. Relation between clusters and topography

The tallest among the six clusters—Cluster 4 and Cluster 5—occupied the highest elevations and the steepest slopes. Cluster 3, the shortest cluster, occupied the lowest slopes and the lowest elevations, the latter along with Cluster 6 (Fig. 8). The elevations occupied by Cluster 1 and Cluster 2, Cluster 3 and Cluster 6, and Cluster 4 and Cluster 5 were not significantly different from each other (at  $p = 0.001$ ). The slopes occupied by all the clusters were significantly different from one another (at  $p = 0.001$ ).

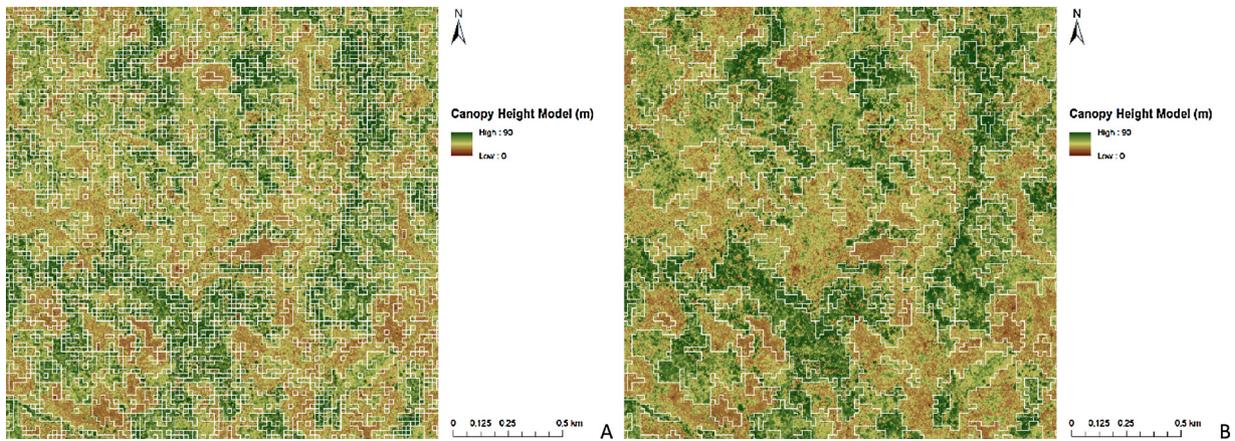


Fig. 2. Forest patches generated from the combined canopy cover maps at different height thresholds (A), and after the merging process (B).

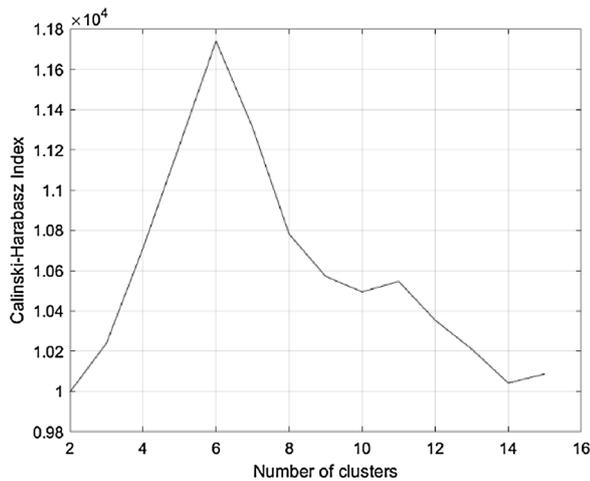


Fig. 3. Calinski-Harabasz indices for two to fifteen clusters.

5. Discussion

Segmentation based on the CHM provided a simple, yet reasonably efficient, method for identifying forest patches using basic GIS software. The developed segmentation and hierarchical merging method created forest polygons which were as homogeneous in structure as possible, with minimised heterogeneity caused by small (i.e. < 0.24 ha) patches of taller trees. The forest patches corresponded reasonably well with patches that could be visually identified from the CHM (Fig. 2). The segmentation method could work in other forests in the region due to

the prevalence of logging, shifting cultivation and subsequent regrowth, leading to a mosaic landscape.

Using the CHM is less time- and resource-intensive than working directly with the ALS point cloud or individual trees (Alexander et al., 2017) for delineating forest patches, especially for large areas. It should however be noted that a CHM with a grid cell size of 1 m was used for generating the patch-level attributes. This could be equivalent to using an ALS point cloud with a point density of 1 point m<sup>-2</sup>. Canopy cover based on the CHM is in some ways a combination of the horizontal and vertical components at a certain threshold. It is possible that this makes it useful for identifying natural clusters in the dataset.

Canopy cover can be calculated at any height threshold, and 1 m is often used as the threshold in managed forests to separate ground from canopy returns (Solberg et al., 2009). A height threshold of 5 m was used in this study to separate forests and non-forests, based on the definition of forests by the FAO. Almost 84% of the forest patches had canopy cover above 90% based on the CHM at a threshold of 5 m, and this would have made it less important for the classification. Canopy cover at higher thresholds was therefore used as additional attributes, and canopy cover at thresholds of 15 m and 25 m were found to be important for the classification (Fig. 5H).

The significant association between the forest structure, represented by the natural clusters, and the forest types in the study area was mainly due to the association between arang forests and Cluster 3. Arang forests are considered to have lower biodiversity than rainforests, but are characterised by high uniqueness of species, with a few (such as Wegner’s glass lizard (*Dopasia [Ophisaurus] wegneri*) and the endemic pitcher plant *Nepenthes tobaica*) of High Conservation Value. However, there have been very few studies on the species composition and extent of these habitats (Sarulla Operations Ltd., 2015). Forest structure

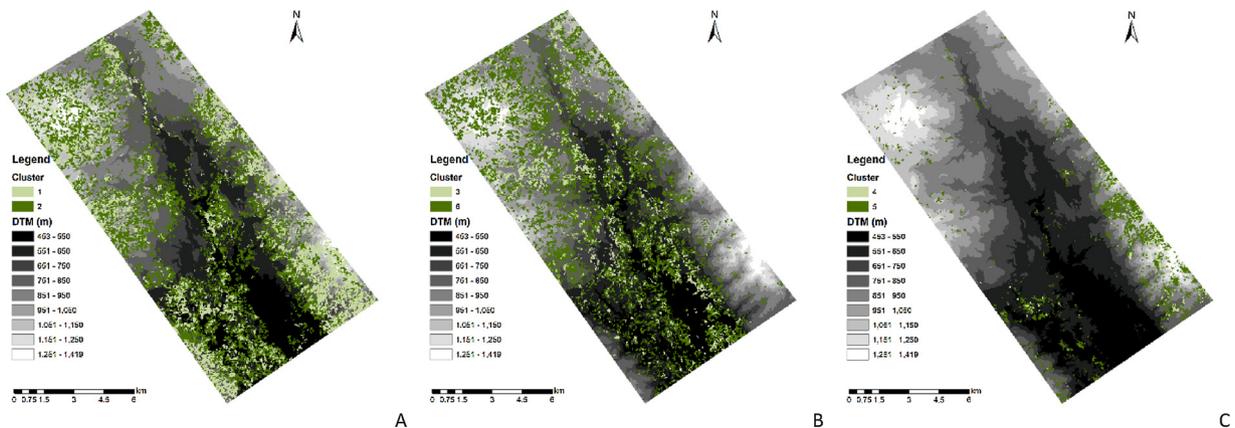


Fig. 4. Forest patches belonging to clusters 1 & 2 (A), 3 & 6 (B) and 4 & 5 (C) overlaid on the Digital Terrain Model.

**Table 2**

Results of ANOVA using Scheffe’s procedure for post hoc pair-wise comparisons. The six symbols in each cell show the results of the comparison of each cluster with the other five; ✓ - significantly different; × - not significantly different with  $\alpha$  set to 0.001; □ - not relevant.

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
HtMean	□✓✓✓✓✓	✓□✓✓✓✓	✓✓□✓✓✓	✓✓✓□✓✓	✓✓✓✓□✓	✓✓✓✓✓□
HtStd	□✓✓✓×✓	✓□✓✓✓✓	✓✓□✓✓✓	✓✓✓□✓✓	×✓✓✓□×	✓✓✓✓×□
CanCov5	□×✓×✓✓	×□✓✓✓✓	✓✓□✓✓✓	×✓✓□×✓	✓✓✓×□✓	✓✓✓✓□□
CanCov15	□✓✓✓✓✓	✓□✓✓✓✓	✓✓□✓✓✓	✓✓✓□×✓	✓✓✓×□✓	✓✓✓✓✓□
CanCov25	□✓✓✓✓✓	✓□✓✓✓✓	✓✓□✓✓✓	✓✓✓□✓✓	✓✓✓✓□✓	✓✓✓✓✓□
CanCov35	□✓✓✓✓✓	✓□✓✓✓✓	✓✓□✓✓×	✓✓✓□✓✓	✓✓✓✓□✓	✓✓×✓✓□
CanCov45	□✓✓✓✓✓	✓□×✓✓×	✓×□✓✓×	✓✓✓□✓✓	✓✓✓✓□✓	✓××✓✓□

characterisation using ALS data could therefore be useful for the delineation and conservation of these unique habitats.

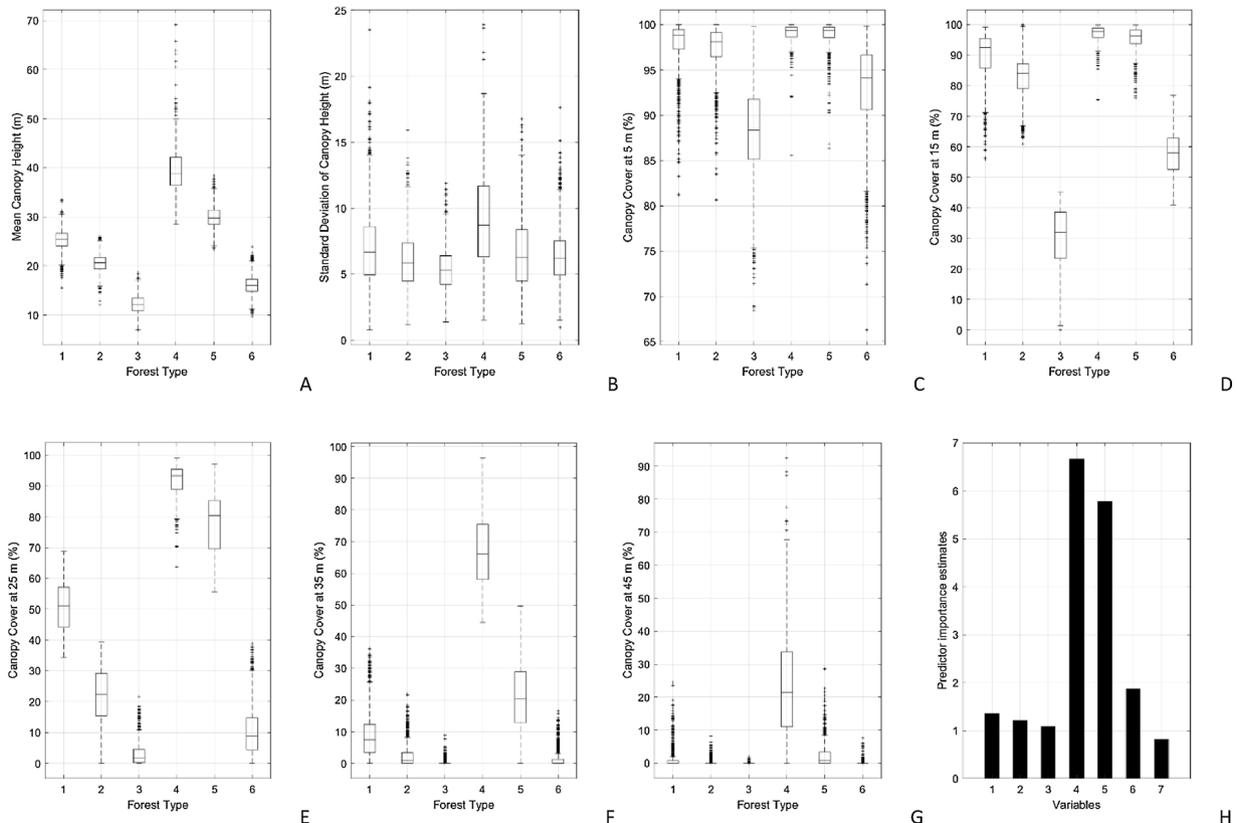
The forest types were determined in the field based on a mixture of floristic and structural criteria, where very mature agroforestry systems may appear structurally similar to tall primary forests. Benzoin and durian are native species and agro-forests would therefore be structurally not very different from secondary/disturbed forests. Cluster 6 was one of the shortest clusters, and had the highest association with agro-forests. It is possible that Cluster 6 represents young agro-forests, with mature agro-forests being located in the taller clusters (especially Clusters 1 and 2).

Field data were collected in the more accessible parts of the study area, and had only very few points in the tallest clusters—three in Cluster 5, and none in Cluster 4. Clustering was therefore very useful in identifying the natural clusters in the whole study area, which would not have been possible using a supervised classification. A number of factors would have influenced the accuracy of assigning cluster labels to the field points. The use of handheld GPS, with low accuracies in rugged terrain under thick canopy, would influence the location accuracy of field points. Many of the field points were close to cluster

boundaries, and an error in the location could get them assigned to adjacent clusters. Points located within the small polygons that were merged to generate the forest patches would also reduce the accuracy of labelling. In addition to this, there could be high subjectivity in the determination of different forest types.

The tallest clusters—Cluster 4 and Cluster 5—had mean elevations above 875 m (Fig. 8); forests in Sumatra that occur between 800/900 m and 1300/1400 m are characterised as sub-montane (Laumonier, 1997). Canopy cover is complementary to gap fraction, and gap fraction is often higher in selectively logged than old-growth forests (Kent et al., 2015). The mean canopy cover for Cluster 4, the tallest cluster, was above 90% even at a height threshold of 25 m. It is highly likely that these tall clusters with the highest canopy cover represent the few remaining patches of primary forest left in the region protected due to the limited accessibility, although this can be ascertained only with field surveys.

Overall, it would seem reasonable to label Clusters 1, 2 and 6 as ‘secondary/disturbed or mixed agro-forest’, Cluster 3 as ‘arang or young secondary/disturbed forest’, and Clusters 4 and 5 as ‘primary sub-montane forest’. Although identified in the field as separate habitat



**Fig. 5.** Box-and-whisker plots of mean canopy height (A), standard deviation of canopy height (B) and canopy cover, at height thresholds of 5, 15, 25, 35 and 45 m (C–G), for the clustering into six forest types, and estimates of predictor importance based on a Random Forest classifier (H).

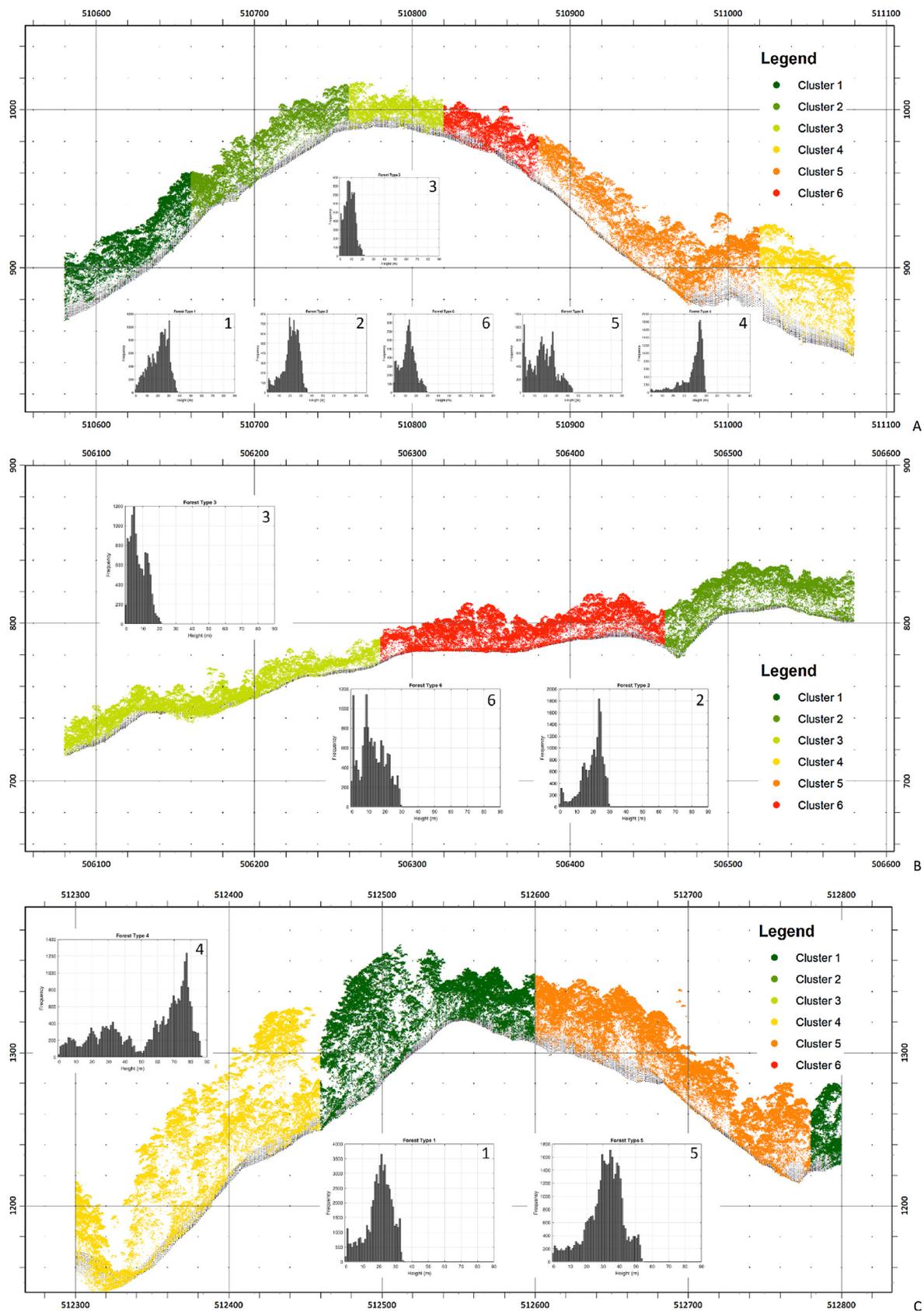


Fig. 6. Vertical profiles through the point cloud from an area of 500 m × 20 m with all the six clusters (A), with the three shortest clusters (B) and the three tallest clusters (C), with points extracted from the Digital Terrain Model (black); Sample height profiles based on ALS point cloud (40 m × 20 m) are also shown.

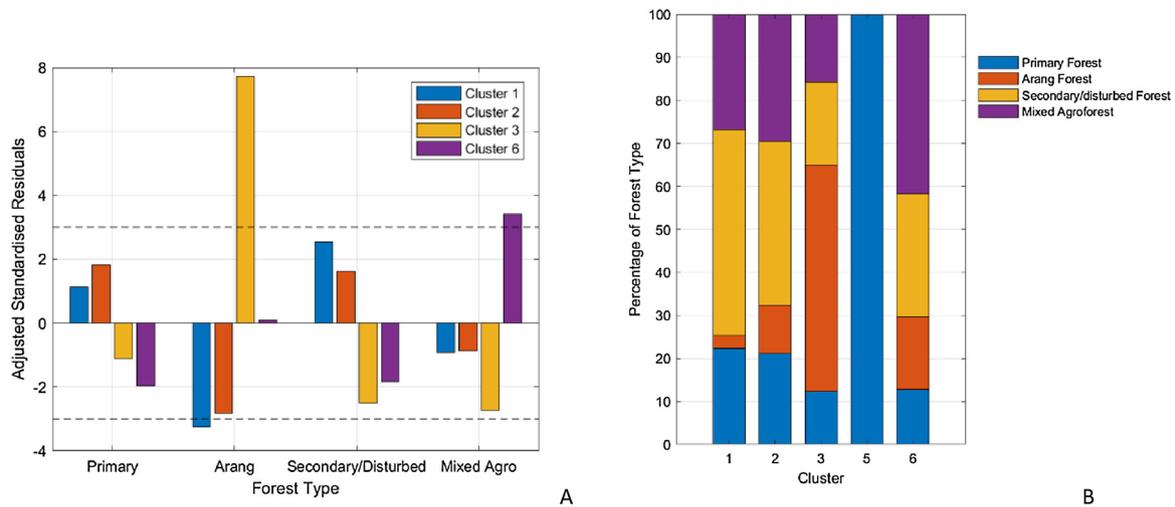


Fig. 7. Adjusted standardised residuals (significant values  $\geq |3.0|$ ) from a chi-squared analysis of clusters by forest type:  $\chi^2(9, N = 502) = 81.33, p < 0.001$  (A); Percentage of field-based forest types occurring in each cluster (B). The number of locations identified as primary, arang and secondary/disturbed forests and mixed agro-forests are 88, 84, 171 and 159 respectively. The number of locations in the different clusters are 67, 213, 57, 0, 3 and 165 respectively.

types, agro-forests in this study area are a form of disturbed forest because they do not have a history of clear felling. Thus, what may be clearly distinguishable in the field based on floristic composition may be less distinguishable based on structure.

The tallest patches in the study area are also largely confined to the south-east of the study area with steep slopes. Mapping and monitoring the extent of these tall patches is important since large trees account for most of the biomass in tropical forests. They also serve as a focal point for biological activity, and create large gaps at death altering the forest structure dynamics and releasing the sequestered carbon (Chambers et al., 2007; Ferraz et al., 2016).

## 6. Conclusion

Landscape-level mapping of forest types based on structural characteristics is useful in providing a broader view of the landscape components compared to field-based surveys in tropical forests. These maps can be used for more efficient field-based surveys based on stratified sampling rather than random sampling, which may not take all the forest types into consideration. This is also relevant for carbon stock assessment, where it is necessary to understand the heterogeneity in the landscape for an accurate assessment of biomass, streamlined for different forest types. This could be a step forward in addressing the need for incorporating regional variations into pan-tropical biomass maps

(Mitchard et al., 2014). The structural composition of the patches may be an indication of habitat type and quality for the different species in these forests, which are increasingly under threat from anthropogenic and natural disturbances.

## Acknowledgements

This research received funding through EU Marie Skłodowska-Curie Actions (H2020-MSCA-IF-2014) under grant agreement no [657607], and is part of the LEAP (Landscape Ecology and Primatology) project. We are grateful to Mr. Johannes Sagala (Sarulla Operations Ltd., Indonesia) for providing us with the Airborne Laser Scanner data, and to PT McElhanney, Indonesia for airborne data collection and initial processing to classify ground and non-ground points. We are also grateful to Eka Siswiyati, Dewi Kurnia, Nursaniah Nasution, Ronald AP Siagian, Sugesti Mohamad Arif, Mokhammad Faeasli Khakim, Irvan Sipayung, Hermansyah, Alamsyah Nasution and Nardi Simbolon of PanEco-YEL, and Roma Irama, Kamarudin, Sumurudin, Pardi Sitompul, Hanjandri Matondang and Parel Sitompul of Sarulla Operations Ltd. for field data collection. We would also like to thank the anonymous reviewers for their valuable comments and suggestions.

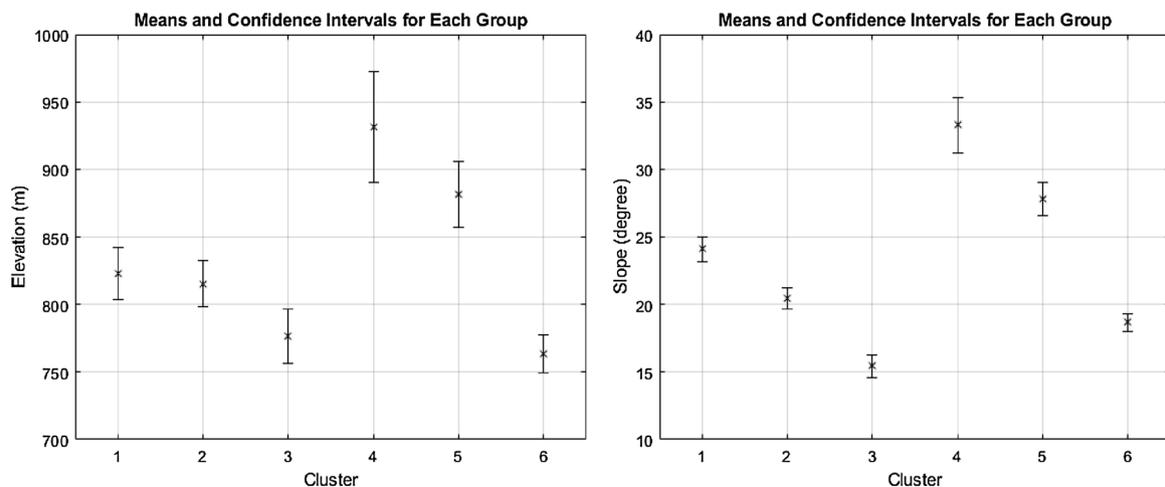


Fig. 8. Mean elevations (A) and mean slopes (B) occupied by the six clusters;  $\alpha = 0.001$ .

## References

- Alexander, C., Korstjens, A.H., Hill, R.A., 2017. Structural attributes of individual trees for identifying homogeneous patches in a tropical rainforest. *Int. J. Appl. Earth Observ. Geoinf.* 55, 68–72.
- Asner, G.P., 2001. Cloud cover in Landsat observations of the Brazilian Amazon. *Int. J. Remote Sens.* 22, 3855–3862.
- Asner, G.P., Brodrick, P.G., Philipson, C., Vaughn, N.R., Martin, R.E., Knapp, D.E., Heckler, J., Evans, L.J., Jucker, T., Goossens, B., Stark, D.J., Reynolds, G., Ong, R., Renneboog, N., Kugan, F., Coomes, D.A., 2018. Mapped aboveground carbon stocks to advance forest conservation and recovery in Malaysian Borneo. *Biol. Conserv.* 217, 289–310.
- Axelsson, P., 2000. DEM generation from laser scanner data using adaptive tin models. *Int. Arch. Photogramm. Remote Sens.* XXXIII (Part B4), 110–117.
- Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., Radeloff, V.C., 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *J. Geophys. Res.* Biogeosci. 114, 13.
- Calinski, T., Harabasz, J., 1974. A Dendrite Method for Cluster Analysis.
- Chambers, J.Q., Asner, G.P., Morton, D.C., Anderson, L.O., Saatchi, S.S., Espírito-Santo, F.D.B., Palace, M., Souza Jr, C., 2007. Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests. *Trends Ecol. Evol.* 22, 414–423.
- Cheyne, S., Thompson, C.J.H., Chivers, D.J., 2013. Travel adaptations of Bornean Agile Gibbons *Hylobates albobarbis* (Primates: Hylobatidae) in a degraded secondary forest, Indonesia. *J. Threatened Taxa* 5, 3963–3968.
- Chivers, D.J., 1974. The siamang in Malaya. A field study of a primate in tropical rain forest. *Contrib. Primatol.* 4, 1–335.
- Coops, N.C., Tompaski, P., Nijland, W., Rickbeil, G.J.M., Nielsen, S.E., Bater, C.W., Stadt, J.J., 2016. A forest structure habitat index based on airborne laser scanning data. *Ecol. Indic.* 67, 346–357.
- Corlett, R.T., 2016. Tropical Rainforests and Climate Change. Reference Module in Earth Systems and Environmental Sciences. Elsevier.
- de Amorim, R.C., Hennig, C., 2015. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Inf. Sci.* 324, 126–145.
- Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016. Lidar detection of individual tree size in tropical forests. *Remote Sens. Environ.* 183, 318–333.
- Foody, G.M., Hill, R.A., 1996. Classification of tropical forest classes from Landsat TM data. *Int. J. Remote Sens.* 17, 2353–2367.
- Guo, X., Coops, N.C., Tompaski, P., Nielsen, S.E., Bater, C.W., John Stadt, J., 2017. Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. *Ecol. Inf.* 38, 50–61.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342, 850–853.
- Hill, R.A., 1999. Image segmentation for humid tropical forest classification in Landsat TM data. *Int. J. Remote Sens.* 20, 1039–1044.
- Hill, R.A., Foody, G.M., 1994. Separability of tropical rain-forest types in the Tambopata-Candamo Reserved Zone, Peru. *Int. J. Remote Sens.* 15, 2687–2693.
- Hou, Z., Xu, Q., Nuutinen, T., Tokola, T., 2013. Extraction of remote sensing-based forest management units in tropical forests. *Remote Sens. Environ.* 130, 1–10.
- Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M.-H., Wong, W.V.C., Ling, Z.-Y., Johari, S.A., Korom, A., James, D., Saito, H., Takao, G., 2016. Evaluation of the similarity in tree community composition in a tropical rainforest using airborne LiDAR data. *Remote Sens. Environ.* 173, 304–313.
- Kanade, R., John, R., 2018. Topographical influence on recent deforestation and degradation in the Sikkim Himalaya in India; Implications for conservation of East Himalayan broadleaf forest. *Appl. Geogr.* 92, 85–93.
- Kennel, P., Tramon, M., Barbier, N., Vincent, G., 2013. Canopy height model characteristics derived from airborne laser scanning and its effectiveness in discriminating various tropical moist forest types. *Int. J. Remote Sens.* 34, 8917–8935.
- Kent, R., Lindsell, J.A., Laurin, G.V., Valentini, R., Coomes, D.A., 2015. Airborne LiDAR detects selectively logged tropical forest even in an advanced stage of recovery. *Remote Sens.* 7, 8348–8367.
- Kim, D.-H., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S., Feng, M., Townshend, J.R., 2014. Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sens. Environ.* 155, 178–193.
- Laumonier, Y., 1997. *The Vegetation and Physiography of Sumatra*. Springer Netherlands.
- Laumonier, Y., Edin, A., Kanninen, M., Munandar, A.W., 2010. Landscape-scale variation in the structure and biomass of the hill dipterocarp forest of Sumatra: Implications for carbon stock assessments. *For. Ecol. Manage.* 259, 505–513.
- Lawrence, D., Vandecar, K., 2014. Effects of tropical deforestation on climate and agriculture. *Nat. Clim. Change* 5, 27.
- McGaughey, R.J., 2015. FUSION/LDV: Software for LIDAR Data Analysis and Visualization. Forest Service, United States Department of Agriculture, Pacific Northwest Research Station.
- McLean, K.A., Trainor, A.M., Asner, G.P., Crofoot, M.C., Hopkins, M.E., Campbell, C.J., Martin, R.E., Knapp, D.E., Jansen, P.A., 2016. Movement patterns of three arboreal primates in a Neotropical moist forest explained by LiDAR-estimated canopy structure. *Landsc. Ecol.* 31, 1849–1862.
- Miettinen, J., Stibig, H.-J., Achard, F., 2014. Remote sensing of forest degradation in Southeast Asia—Aiming for a regional view through 5–30 m satellite data. *Glob. Ecol. Conserv.* 2, 24–36.
- Mitchard, E.T.A., Feldpausch, T.R., Brienen, R.J.W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T.R., Lewis, S.L., Lloyd, J., Quesada, C.A., Gloor, M., ter Steege, H., Meir, P., Alvarez, E., Araujo-Murakami, A., Aragão, L.E.O.C., Arroyo, L., Aymard, G., Banki, O., Bonal, D., Brown, S., Brown, F.I., Cerón, C.E., Chama Moscoso, V., Chave, J., Comiskey, J.A., Cornejo, F., Corrales Medina, M., Da Costa, L., Costa, F.R.C., Di Fiore, A., Domingues, T.F., Erwin, T.L., Fredericksen, T., Higuchi, N., Honorio Coronado, E.N., Killeen, T.J., Laurance, W.F., Levis, C., Magnusson, W.E., Marimon, B.S., Marimon Junior, B.H., Mendoza Polo, I., Mishra, P., Nascimento, M.T., Neill, D., Núñez Vargas, M.P., Palacios, W.A., Parada, A., Pardo Molina, G., Peña-Claros, M., Pitman, N., Peres, C.A., Poorter, L., Prieto, A., Ramirez-Angulo, H., Restrepo Correa, Z., Roopsind, A., Roucoux, K.H., Ruidas, A., Salomão, R.P., Schiatti, J., Silveira, M., de Souza, P.F., Steininger, M.K., Stropp, J., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., van Andel, T.R., van der Heijden, G.M.F., Vieira, I.C.G., Vieira, S., Vilanova-Torre, E., Vos, V.A., Wang, O., Zartman, C.E., Malhi, Y., Phillips, O.L., 2014. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Glob. Ecol. Biogeogr.* 23, 935–946.
- Mohd Zaki, N.A., Latif, Z.A., Suratman, M.N., 2018. Modelling above-ground live trees biomass and carbon stock estimation of tropical lowland Dipterocarp forest: integration of field-based and remotely sensed estimates. *Int. J. Remote Sens.* 39, 2312–2340.
- Moran, E.F., Brondizio, E., Mausel, P., Wu, Y., 1994. Integrating Amazonian Vegetation, Land-use, and Satellite Data: Attention to differential patterns and rates of secondary succession can inform future policies. *BioScience* 44, 329–338.
- Nater, A., Mattle-Greminger, M.P., Nurcahyo, A., Nowak, M.G., de Manuel, M., Desai, T., Groves, C., Pybus, M., Sonay, T.B., Roos, C., Lameira, A.R., Wich, S.A., Askew, J., Davila-Ross, M., Fredriksson, G., de Valles, G., Casals, F., Prado-Martinez, J., Goossens, B., Verschoor, E.J., Warren, K.S., Singleton, I., Marques, D.A., Pamungkas, J., Perwitasari-Parajallah, D., Rianti, P., Tuuga, A., Gut, I.G., Gut, M., Orozco-Wengel, P., van Schaik, C.P., Bertranpetit, J., Anisimova, M., Scally, A., Marques-Bonet, T., Meijaard, E., Krützen, M., 2017. Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Curr. Biol.* 27, 3487–3498 e3410.
- Palminteri, S., Powell, G.V.N., Asner, G.P., Peres, C.A., 2012. LiDAR measurements of canopy structure predict spatial distribution of a tropical mature forest primate. *Remote Sens. Environ.* 127, 98–105.
- PTMcElhanney, 2015. LiDAR & DAP Survey Project Report for BUT Sarulla Operations Ltd. In Indonesia.
- Putz, F.E., Ruslandi, Ellis, P.W., Griscorn, B.W., 2018. Topographic restrictions on land-use practices: Consequences of different pixel sizes and data sources for natural forest management policies in the tropics. *For. Ecol. Manage.* 422, 108–113.
- Rödig, E., Cuntz, M., Rammig, A., Fischer, R., Taubert, F., Huth, A., 2018. The importance of forest structure for carbon fluxes of the Amazon rainforest. *Environ. Res. Lett.* 13, 054013.
- Salovaara, K.J., Thessler, S., Malik, R.N., Tuomisto, H., 2005. Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sens. Environ.* 97, 39–51.
- Sandel, B., Svenning, J.-C., 2013. Human impacts drive a global topographic signature in tree cover. *Nat. Commun.* 4, 2474.
- Sarulla Operations Ltd, 2015. Sarulla Geothermal Power Project: Biodiversity Action Plan & Biodiversity Off-set Management Plan. In: Sarulla Operations Ltd (Ed.), Environmental Management Plans.
- Sexton, J.O., Noojipady, P., Anand, A., Song, X.-P., McMahon, S., Huang, C., Feng, M., Channan, S., Townshend, J.R., 2015. A model for the propagation of uncertainty from continuous estimates of tree cover to categorical forest cover and change. *Remote Sens. Environ.* 156, 418–425.
- Singh, M., Cheyne, S.M., Ehlers Smith, D.A., 2018. How conspecific primates use their habitats: Surviving in an anthropogenically-disturbed forest in Central Kalimantan, Indonesia. *Ecol. Indic.* 87, 167–177.
- Solberg, S., Brunner, A., Hanssen, K.H., Lange, H., Næsset, E., Rautiainen, M., Stenberg, P., 2009. Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sens. Environ.* 113, 2317–2327.
- Sverdrup-Thygeson, A., Ørka, H.O., Gobakken, T., Næsset, E., 2016. Can airborne laser scanning assist in mapping and monitoring natural forests? *For. Ecol. Manage.* 369, 116–125.
- Thomas, S.C., Baltzer, J.L., 2001. *Tropical Forests*. eLS: John Wiley & Sons, Ltd.
- Turner, M.G., Gardner, R.H., O'Neill, R.V., 2001. *Landscape Ecology in Theory and Practice: Pattern and Process*. Springer-Verlag, New York.
- Whitmore, T.C., 1990. *An Introduction to Tropical Rain Forests/T.C. Whitmore*. Clarendon Press, Oxford.
- Wich, S.A., Singleton, I., Nowak, M.G., Utami Atmoko, S.S., Nisam, G., Arif, S.M., Putra, R.H., Ardi, R., Fredriksson, G., Usher, G., Gaveau, D.L.A., Kühl, H.S., 2016. Land-cover changes predict steep declines for the Sumatran orangutan (*Pongo abelii*). *Sci. Adv.* 2, e1500789.
- Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., White, J.C., 2008. The role of LiDAR in sustainable forest management. *Forest. Chron.* 84, 807–826.