Mapping fine-scale variation in diverse tropical forests with distinct ecological dynamics requires few leaf traits and structural attributes

Elsa Ordway[1], Gregory Asner[2], David Burslem[3], Simon Lewis[4], Reuben Nilus[5], Roberta Martin[2], Michael O'Brien[6], Oliver Phillips[4], Lan Qie[7], Nick Vaughn[2], and Paul Moorcroft[1]

[1]Harvard University
[2]Arizona State University
[3]University of Aberdeen School of Biological Sciences
[4]University of Leeds School of Geography
[5]Sabah Forestry Department
[6]Universidad Rey Juan Carlos
[7]University of Lincoln

March 07, 2024

Abstract

Remote sensing is a powerful tool for characterizing ecosystems at large scales. However, the relative importance of leaf traits and canopy structure in characterizing the spatial distribution of functionally distinct tropical forests – the most diverse, structurally complex, and heterogeneous ecosystems on Earth – remains under-explored. Using satellite-resolution LiDAR and imaging spectroscopy metrics, we map spatial turnover in tropical forest function, examine the relative importance of leaf traits and canopy structure, and analyze differences in aboveground carbon and demography. We find that leaf phosphorus, LMA, and canopy height are key distinguishing properties of forest types, achieving accuracies of 85-96% and correspond to differences in community growth and mortality rates. Our remotely sensed forest types align with ground-based forest definitions but enable mapping of their entire extent. At 30 m resolution, our method can be used at large scales with spaceborne data to reveal important differences in structure and function across tropical forests.

Article Type: Method

Mapping fine-scale variation in diverse tropical forests with distinct ecological dynamics requires few leaf traits and structural attributes

Short Title: Mapping fine-scale tropical forest variation

Elsa M. Ordway[1,2]*, Gregory P. Asner[3], David F.R.P. Burslem[4], Simon L. Lewis[5,6], Reuben Nilus[7], Roberta Martin[3], Michael J. O'Brien[8], Oliver L. Phillips[5], Lan Qie[9], Nicolas R. Vaughn[3], Paul R. Moorcroft[1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

[2]Department of Ecology and Evolutionary Biology, UCLA, 612 Charles E. Young Drive South, Los Angeles, CA 90095, USA

[3]Center for Global Discovery and Conservation Science, Arizona State University, 1001 McAllister Ave., Tempe, AZ 85281, USA

[4]School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, U.K.

[5]School of Geography, University of Leeds, Leeds LS2 9JT, U.K.

[6]Department of Geography, University College London, London. WC1E 6BT.

[7]Sabah Forestry Department, Forest Research Centre, Sandakan, Sabah, MY

[8]Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, c/ Tulipán s/n., E-28933 Móstoles, Spain

[9]School of Life Sciences, University of Lincoln, Lincoln LN6 7DL, U.K.

Elsa M. Ordway, elsa.ordway@gmail.com

Gregory P. Asner, gregasner@asu.edu

David F.R.P. Burslem, d.burslem@abdn.ac.uk

Simon Lewis, S.L.Lewis@leeds.ac.uk
Reuben Nilus, Reuben.Nilus@sabah.gov.my
Roberta Martin, Roberta.Martin@asu.edu
Michael J. O'Brien, mikey.j.obrien@gmail.com
Oliver L. Phillips, O.Phillips@leeds.ac.uk
Lan Qie, lqie@lincoln.ac.uk
Nick Vaughn, nickvaughn@asu.edu
Paul R. Moorcroft, paul_moorcroft@harvard.edu

Statement of Authorship: EO and PM designed the study. DB, SL, RN, MO, OP, and LQ collected and provided the inventory plot data. GA and RM led the collection of remote sensing data and foliar trait data. RM and NV processed the remote sensing data and foliar chemical data. EO analyzed the output data, performed the statistical modeling work, and wrote the first draft of the manuscript.

Data Accessibility Statement: Should this manuscript be accepted, the data supporting results will be archived in Dryad or Figshare, with a DOI that will be included at the end of the article.

Abstract Word Count: 150
Main Text Word Count: 4999
Text Box Word Count: N/A
Number of References: 68
Number of Figures: 6
Number of Tables: 0
Number of Text Boxes: 0

***Corresponding Author**: Elsa M. Ordway, 26 Oxford Street, Suite 43, Cambridge, MA 02138; 616-443-0141; e-mail: elsa.ordway@gmail.com

Abstract: Remote sensing is a powerful tool for characterizing ecosystems at large scales. However, the relative importance of leaf traits and canopy structure in characterizing the spatial distribution of functionally distinct tropical forests – the most diverse, structurally complex, and heterogeneous ecosystems on Earth – remains under-explored. Using satellite-resolution LiDAR and imaging spectroscopy metrics, we map spatial turnover in tropical forest function, examine the relative importance of leaf traits and canopy structure, and analyze differences in aboveground carbon and demography. We find that leaf phosphorus, LMA, and canopy height are key distinguishing properties of forest types, achieving accuracies of 85-96% and correspond to differences in community growth and mortality rates. Our remotely sensed forest types align with ground-based forest definitions but enable mapping of their entire extent. At 30 m resolution, our method can be used at large scales with spaceborne data to reveal important differences in structure and function across tropical forests.

Keywords: Airborne remote sensing, imaging spectroscopy, LiDAR, Borneo, Malaysia

Introduction

Tropical forests are the most biologically diverse biome on Earth (Myers 1988), encompassing an estimated 96% of all tree species (Corlett 2016). Through their differences in structure and functional traits, variation in species composition can directly influence ecosystem processes in tropical forests (e.g., Osborne *et al.* 2020). Tropical forest canopy structure and function vary geographically by climate (Givnish 1999), topography (Jucker et al. 2018), and edaphic conditions (Townsend *et al.* 2008; Hulshof & Spasojevic 2020), as well as different natural and anthropogenic disturbance histories and regimes (Chazdon 2003; Brando *et al.* 2019). However, comprehensive knowledge of tropical forest diversity remains largely limited to field studies that cover a small fraction of the biome. While networks of tropical forest inventory plots offer invaluable ground observations and insights into fine-scale mechanisms and processes, remote sensing data, increasingly available at spatial resolutions relevant to organisms, can be used to scale these insights to entire landscapes and regions, serving as powerful tools to measure and map forest function (Schimel *et al.* 2013; Jetz *et al.* 2016).

Imaging spectroscopy (i.e., hyperspectral remote sensing) and light detection and ranging (LiDAR) offer capabilities for measuring, mapping, monitoring, and understanding tropical forest functional diversity, structure, vertical light environments, leaf traits, and aboveground carbon stocks beyond plot boundaries. These data can inform ecological understanding (Bongalov *et al.* 2019; Draper *et al.* 2019), support conservation efforts (Asner *et al.* 2017), and constrain terrestrial biosphere models (Antonarakis *et al.* 2014). In the tropics, airborne imaging spectroscopy has recently been used to map patterns of diversity across forest communities in Amazonia (Féret & Asner 2014; Draper *et al.* 2019) and Borneo (Bongalov *et al.* 2019), and spectral measures of

tropical forest α and β diversity have been shown to correlate with traditional taxonomically-based estimates of these quantities.

Airborne imaging spectroscopy measurements have also been used to characterize the leaf traits of tropical forest canopies and identify relationships between these traits and underlying environmental drivers including soil biogeochemistry, topography, hydrology, and climate. For example, Asner *et al.* (2016, 2017) identified relationships between imaging spectroscopy derived estimates of foliar traits and variation in geology, topography, hydrology, and climate across the Peruvian Amazon, and sorted the region into 36 distinct forest types using hierarchical clustering. In Malaysia, airborne imaging spectroscopy and LiDAR data have been used to demonstrate a strong influence of fine-scale topography on forest structure, composition and diversity (Jucker *et al.* 2018b), and role of geomorphology on topographic controls on canopy foliar traits across larger elevation gradients (Chadwick & Asner 2020). In a similar manner, LiDAR measurements have been used to evaluate variation in tropical forest height and carbon stocks with forest succession (Dubayah *et al.* 2010), fine-scale topography (Muscarella *et al.* 2020), and spatial variation in vertical leaf area density profiles (Detto *et al.* 2015).

The recent surge in ecologically orientated satellite remote sensing missions, including the operational PRISMA (ESA 2021b) and DESIS (GAC & TBE 2021) spectrometers, NASA's GEDI spaceborne LiDAR (Dubayah *et al.* 2020b, a, c), and the planned NASA SBG (NASA JPL 2021) and European Space Agency CHIME (ESA 2021a) satellite-based spectrometers, make this a critical moment to assess the relative importance of forest structure and canopy leaf traits for characterizing tropical forest function. These instruments will overcome airborne campaign

limitations, which are expensive and restricted in spatial extent, by providing extensive coverage over tropical forest regions. However, the data from these sensors will be at spatial resolutions of ~30 m, far coarser than the 1-5m resolution data used in the studies described above. In addition, the above-mentioned studies have demonstrated the capacity to map spatial variation in tropical forest species composition and functional and structural diversity using remote sensing data. However, the relative importance of different leaf traits and forest structural attributes in determining differences between distinct tropical tree communities remains largely unknown. In this study, we combine imaging spectroscopy-derived leaf trait measurements with lidar-derived measurements of canopy structure to 1) identify, characterize, and map structurally and functionally distinct tropical forests across two landscapes in Malaysian Borneo; 2) examine the feasibility of conducting these analyses at resolutions corresponding to new satellite missions; 3) determine the key leaf traits and canopy structural attributes that distinguish different forest types; and 4) integrate inventory plot data to explore differences in forest dynamics across mapped forest types.

Materials and Methods

Study Landscapes

The study landscapes are in Sabah, Malaysian Borneo, encompassing forests in Danum Valley with the tallest trees in the tropics (Shenkin *et al.* 2019), and nutrient-poor *kerangas* forests with stunted canopies and unique floristic composition (Newbery 1991). The first landscape is Sepilok, a 4,500 ha reserve of lowland mixed dipterocarp forests spanning varying topography and soil nutrients (Fox 1973, Nilus 2004, Dent *et al.* 2006; Jucker *et al.* 2018b). The second landscape is

Danum, a 44,000 ha conservation area with predominantly lowland, intact tropical rainforest. In this study, we focus on the 50-ha ForestGEO inventory plot located in the eastern part of Danum.

Both landscapes exhibit differences in structure, function, and composition that correspond to underlying soil and geologic substrate (Fox 1973, Nilus 2004, Dent & Burslem 2016; Coomes *et al.* 2017; Jucker *et al.* 2018b). Sepilok is characterized by three forest types: alluvial forests on fertile ultisols along alluvial flats and gentle slopes; sandstone forests on well-drained, nutrient-poor ultisols along steep ridges; and *kerangas* forests that dominate acidic, extremely nutrient-poor podosols along lower dip slopes of cuesta landforms (DeWalt *et al.* 2006; Dent & Burslem 2016). Total P, nitrate, and base cations are significantly higher in alluvial soils than in the sandstone and more acidic *kerangas* forest soils, influencing community differences in species composition, leaf traits, and stand structure (Dent *et al.* 2006; Dent & Burslem 2009). An earlier field study also identified mudstone hills within the alluvial forests as being further distinguishable in terms of soil chemistry and plant growth (Nilus 2004), although mudstone and alluvial areas in Sepilok are typically characterized as a single forest type (e.g., Coomes *et al.* 2017; Jucker *et al.* 2018b).

Airborne remote sensing data

To measure forest structure and foliar traits, we used co-aligned LiDAR and imaging spectroscopy data collected by the Global Airborne Observatory (GAO) in April 2016 (Asner *et al.* 2012). We examined ten forest structure variables and canopy foliar characteristics that are strongly linked to ecosystem function and have demonstrated measurability with high accuracy using airborne remote-sensing techniques (Table S1, Supplementary Figure S1). Variation in canopy structure

169 was characterized using five metrics: 99th percentile of total canopy height (Max H, m), leaf area
170 index (LAI, m^2 m^{-2}), the peak height of LAI ($H_{peak\ LAI}$, m), a measure of canopy architecture
171 indicating the vertical distribution of plant foliage (P) relative to the total canopy height (P:H ratio),
172 and the fraction of canopy cover taller than 20 m height above the ground ($Cover_{20}$, %). Variation
173 in canopy leaf traits were analyzed based on differences in leaf mass per area (LMA, g DM m^{-2}),
174 foliar nitrogen (N, %) and phosphorus (P, %) concentrations, and foliar N:P ratios. To assess
175 differences in maximum photosynthetic capacity, V_{cmax} was estimated from foliar N and P
176 concentrations using the equation in Table 3, model 1 from (Walker *et al.* 2014). To examine the
177 feasibility of conducting these analyses at coarser resolutions, we resampled data and ran analyses
178 at resolutions ranging from 16 m^2 - 40,000 m^2. LiDAR and imaging spectroscopy data and
179 processing are described in Supplementary Methods.
180
181 ***Characterizing functionally distinct forests***
182 We mapped forest types across Sepilok and Danum. At Danum, we restricted our analysis to the
183 50-ha ForestGEO plot location and a 1-km buffer around the plot. To characterize functional and
184 structural diversity across all pixels, we 1) conducted a principal component analysis (PCA) to
185 reduce dimensionality of all ten canopy leaf traits and structural attributes (hereafter canopy
186 properties), and 2) ran a *k*-means cluster analysis (Hartigan & Wong 1979) on the first two
187 principal components to categorize pixels into distinct functional communities. PCA and *k*-means
188 cluster analysis data processing is described in Supplementary Methods.
189
190 The primary metric for identifying the appropriate number of clusters (*k*) was the gap statistic
191 (Gap_k), which defines the number of clusters based on the first local and global maxima (Tibshirani

192 *et al.* 2001). We also evaluated output for $k = k_s + 1$ and for $k = k_s - 1$, where k_s represents the
193 number of clusters selected using Gap_k. Two secondary cluster metrics were also considered: 1)
194 the elbow approach using the within group sum of squares (W_k), and 2) the between cluster sum
195 of squares (*BSS*) divided by the total sum of squares (*TSS*). A higher value of *BSS/TSS* indicates
196 improved fit of the cluster analysis to the data (Milligan & Cooper 1985). Because *BSS/TSS*
197 increases monotonically as k increases, we evaluated the k at which *BSS/TSS* increases flattened,
198 in addition to Gap_k and the W_k elbow approach (Tibshirani *et al.* 2001).
199
200 We visually evaluated cluster results against inventory plot data from forest ecosystems that have
201 been studied extensively in the field and exhibit clear differences in structure and function.
202 Significant differences in canopy properties between clusters were calculated based on one-way
203 ANOVAs using the *aov* and *TukeyHSD* functions in R. To explore the minimum number of canopy
204 properties required to capture differences in forest types, we evaluated cluster results using only
205 LiDAR variables (structural attributes), only imaging spectroscopy variables (leaf traits), and
206 reduced combinations of canopy properties. To evaluate these reduced models, we calculated
207 overall accuracy as the proportion of pixels mapped the same as the full 10-variable model.
208
209 ***Inventory plot data***
210 To evaluate cluster analysis performance, we compared our forest functional composition maps to
211 inventory plot data at Danum and Sepilok. Our plot dataset consisted of nine existing 4-ha forest
212 inventory plots distributed across alluvial ($n = 3$), sandstone ($n = 3$), and *kerangas* ($n = 3$) forests
213 at Sepilok, and one 50-ha plot at Danum. Data from the nine 4-ha Sepilok plots and the Danum
214 50-ha plot were from the ForestPlots.net online repository (Lopez-Gonzalez *et al.* 2009, 2011) and

the ForestGEO online repository (ForestGEO 2021), respectively. The datasets include stem diameter measurements and taxonomic identification to species level for every tree ≥ 1 cm and ≥ 5 cm in diameter in the ForestGEO and ForestPlots.net plots, respectively. Census years from each plot were as follows: alluvial – 2001, 2009, 2014; sandstone – 2001/03, 2008/09, 2013/14; *kerangas* – 2001, 2008/10, 2014/15; Danum – 2011/15, 2019. The GAO airborne campaign in Sabah was conducted in 2016.

Observed differences in ecosystem dynamics

In lieu of direct measurements of ecosystem function at the study locations (e.g., net primary productivity), we quantified differences in three related ecosystem dynamics: aboveground carbon, growth, and mortality. We compared stand-level growth and mortality rates calculated from forest inventory data and remotely sensed estimates of aboveground carbon density (ACD, Mg C ha^{-1}) at plot locations within the inventory plots, and across all mapped pixels within each forest type to examine differences in aboveground carbon beyond the plots. ACD at 30 m resolution was estimated from the GAO top-of-canopy height (TCH) and $Cover_{20}$ data following (Jucker *et al.* 2018a), described in (Asner *et al.* 2018). The method involves estimating ACD from a network of 0.25 to 1-ha field plots using the BIOMASS workflow described in (Réjou-Méchain *et al.* 2017) in conjunction with the Chave *et al.* (2014) pantropical biomass allometry. Equations from (Asner & Mascaro 2014) were used to estimate ACD from the TCH data, modified based on (Jucker *et al.* 2018a) to incorporate $Cover_{20}$ as a proxy for stand-level basal area. Annual relative DBH growth rates and annual mortality rates were calculated from plot data (stems ≥ 10 cm) following (Condit *et al.* 2006). When calculating growth rates, we excluded trees with broken or resprouted

stems and stems that grew > 7.5 cm yr^{-1} or shrunk > 25% of their initial DBH following (Condit *et al.* 2006). Negative growth rates < 25% of initial DBH were converted to zero.

Results

We identified between two and four distinct forest types in Sepilok (Figure 1). The Gap_k metric identified three clusters (*BSS/TSS* = 68.5%). However, the W_k elbow and *BSS/TSS* metrics suggest that Sepilok can also be characterized as two (*BSS/TSS* = 51.9%) or four (*BSS/TSS* = 76.7%) distinct forest types based on the magnitude of the decline in W_k, and gains in *BSS/TSS* before the values of both metrics level-off with increasing k (Figure S4-S5). Correspondence between mapped forest type boundaries and inventory plots show that the series of clustered forest types align closely with existing forest community definitions (Figure 1). Cluster analysis results for differing values of k indicate a nested hierarchy of forest types at Sepilok: the highest level ($k = 2$) distinguished the alluvial from the sandstone and *kerangas* forest communities; $k = 3$ distinguished sandstone forests from *kerangas* forests; and $k = 4$ partitioned the alluvial forest into two forest types, revealing the less-well known mudstone community as distinct from the interspersed alluvial forest.

At Danum, the Gap_k metric identified a single cluster (*BSS/TSS* = 0.0%); however, the W_k elbow *BSS/TSS* methods both indicate that Danum can be characterized as three distinct forest types (*BSS/TSS* = 61.3%; Figure 1; Figure S4-S5). Two of these forest types were found within the 50-ha plot (white rectangle in Figure 1). The plot is dominated by one forest type (Danum 2), although the northeast corner was identified as distinct (Danum 1) when $k = 2$ and $k = 3$ (Figure 1; Figure S8).

260

261 ***Distinguishing characteristics of forest types***

262 The first principal component (PC1) corresponded to leaf economic spectrum traits (LMA, N, P).
263 The second principal component (PC2) reflected variation in canopy stature (Max H, $Cover_{20}$) and
264 architecture (P:H), as well as photosynthetic capacity (V_{cmax}). These patterns were consistent at
265 Danum and Sepilok (Figure 2; Figure S6). LAI explained little variation across the forest types,
266 with weak loading values (PC3 at Sepilok, PC4 at Danum; Figure S6). Figure 3 shows variation
267 in canopy properties across forest types, shown for the largest number of forest types identified at
268 each landscape (i.e., k = 3 and k = 4; see Figures S7-S9 for results from other values of k). The
269 sandstone and *kerangas* forests had the lowest mean foliar nutrient concentrations and
270 photosynthetic capacities (Figures 3 – Foliar N, Foliar P, V_{cmax}). Despite having lower canopy
271 height than other forest types, the sandstone and *kerangas* forests had the highest fraction of
272 canopy cover above 20 m, high P:H values, and the highest peak height of LAI (Figure 3 – $Cover_{20}$,
273 P:H, $H_{peak\ LAI}$).

274

275 Strong gradients in LMA, N, and P leaf traits were observed across all forest types. The highest
276 foliar nutrient concentrations and the lowest average LMA were observed in the three Danum
277 forest types, and the Sepilok mudstone and alluvial forests (Figure 3 – LMA). These patterns were
278 consistent across different values of k (Figures S7-S8). Average leaf N and P in the mudstone
279 forest were equivalent to or higher than the alluvial forest, yet the mudstone forest had significantly
280 lower V_{cmax}. Significantly lower maximum canopy heights (max H) and greater foliage density
281 near the ground (lower P:H) also distinguished the mudstone and Danum 1 forests from the alluvial
282 and Danum 2-3 forests. The Danum 1 forest (when k = 2 or 3) was structurally similar to the

mudstone forest; however, the two communities differed in leaf economic spectrum traits (Figures 3 – LMA, Foliar N, Foliar P).

While average canopy LAI was similar across communities (Figure 3 – LAI), ranging from 5.5 to 6.3, (coefficient of variation (CV) = 0.05), the average height of maximum LAI ($H_{peak\ LAI}$), canopy architecture (P:H), and canopy cover at 20 m ($Cover_{20}$) all exhibited much greater variation across communities (CV = 0.48; 0.12; 0.25 respectively). Vertical LAI patterns further illustrated differences in structure across forest types despite similar total LAI (Figure 4, Figure S10), with strong clumping in the understory and the upper canopy at the alluvial and Danum forests. Vertical LAI profiles indicated less height heterogeneity in the sandstone and *kerangas* forests (Figure 4). Maximum canopy height, which varied significantly across clusters, was correlated with V_{cmax} between the different forest types (R^2 = 0.72, p = 0.017) and at the pixel scale (R^2 = 0.24, p < 0.0001) (Figure S9).

Aboveground carbon, an emergent property of ecosystem function, differed significantly across clustered forest types, with high values on average in sandstone forests and widely varying values in the alluvial and Danum 2-3 forest types (Figure 5a). Aboveground carbon density within the inventory plots generally corresponded to aboveground carbon distributions derived from the entire forest type (Figure 5a). The one exception was the alluvial forest. When three forest types were distinguished at Sepilok (k = 3), the alluvial forest inventory plot had significantly higher aboveground carbon than the cluster-derived alluvial forest extent (Figure 5a, p < 0.001). However, when the mudstone and alluvial forests were differentiated (k = 4), the inventory plot aboveground

carbon distribution was comparable to aboveground carbon in the clustered alluvial forest extent, while the mudstone forest encompassed significantly lower aboveground carbon densities.

Differences in annual relative growth and mortality rates were also observed across forest types within the inventory plots (Figure 5b). Growth rates differed significantly across all forest types, corresponding inversely to mean aboveground carbon at the sandstone (232 MgC ha^{-1}), alluvial (223 MgC ha^{-1}), and Danum 50-ha (194 MgC ha^{-1}) inventory plots (Figure 5a-b). The *kerangas* forest did not follow this trend, exhibiting an intermediate plot-level growth rate despite lower average aboveground carbon (180 MgC ha^{-1}). Mortality rates were similar in the alluvial and Danum 50-ha plots, and significantly higher than the mortality rates in the sandstone or *kerangas* plots.

The relative importance of leaf traits and structural attributes

Cluster analyses conducted with only structural attributes, only leaf traits, or reduced combinations of leaf traits and structural attributes, indicated that leaf P, LMA, maximum canopy height and $Cover_{20}$ are critical for capturing the observed forest types (Figure 6). Clustering with LMA, P, $Cover_{20}$, and maximum height resulted in similar forest types to those identified when ten canopy properties were used (overall accuracies (OA) of 96.0% and 86.0% for $k = 2$ and $k = 4$ respectively) at Sepilok (Figure 6a; Figure S11a), as well as higher *BSS/TSS* values at both Sepilok (Figure S12a) and Danum (Figure S12b). At Danum, LMA, P, and $Cover_{20}$ alone yielded the strongest similarity to the cluster results with all ten variables (OA = 88.0%; Figure 6b, Figure S11b). The highest overall accuracy for $k = 3$ at Sepilok was achieved with the three leaf economic spectrum traits, equal to 85.9%, although the combination of maximum height, LMA and P (OA = 84.8%),

and just LMA and P (OA = 84.7%) yielded similar results (Figure 6a). We were unable to obtain the observed patterns using structural attributes alone. The inclusion of leaf P improved output in all cases in terms of correspondence with plot locations and noise (speckling) reduction.

Discussion

Our analysis of LiDAR and imaging spectroscopy data at satellite-scale resolution reveals that a few key remotely sensed canopy properties – foliar P, LMA, Max H, $Cover_{20}$ – can be used to successfully identify ecologically-distinct forest types at two tropical forest sites in Malaysian Borneo. The forest types identified using these remotely sensed traits closely align with forest communities defined from field-based floristic surveys and plot-based measurements of their growth and mortality rates. However, our approach enables mapping of their entire extent and reveals important structural and functional variation within areas characterized as a single forest community in previous studies. The ability to do so using remote sensing measurements at 30 m resolution means that our method can be applied to emerging spaceborne LiDAR and imaging spectroscopy data to reveal important differences in structure and function across the world's tropical forests.

Nested functional communities revealed

The cluster analyses at Sepilok and Danum revealed nested distinctions between forest types. The Sepilok mudstone forest was nested, both spatially and statistically, within the alluvial forest type. For k = 2 and 3, the two forests were aggregated as a single forest type, although k = 4 revealed forests with significant differences in leaf economic spectrum traits and canopy structure (Figure 3). This finding is consistent with independent field-research at Sepilok. Mudstone hills were first

identified as distinct from surrounding alluvial forests by (Nilus 2004; Nilus *et al.* 2011), who found differences in soil cation exchange capacity, pH, and nutrient concentrations that translated into intermediate plant growth rates in mudstone forests, between higher and slower growth rates in alluvial and sandstone forests respectively. More recently, (Bartholomew *et al.* in press) found higher clay fractions and higher exchangeable Mg, Ca, and K at varying soil depths in Sepilok mudstone forest compared to alluvial forests. In addition to differences in foliar N and P concentrations, consistent with our results, (Bartholomew *et al.* in press) found that leaf Ca concentrations were higher in mudstone forests than alluvial, sandstone, and *kerangas* forests.

Our findings also reveal that mudstone forests have much lower aboveground carbon than the intermingled low-lying alluvial forests. The lower aboveground carbon may be due to lower soil nutrients and higher acidity, as well as differences in hydrology. (Born *et al.* 2014, 2015) found that differences in growth and mortality responses to flooding at seedling and sapling stages are relevant to the community assembly of species in Sepilok mudstone and alluvial forests. High mortality was observed for some species in alluvial areas immediately after ephemeral flooding events, suggesting that soil water relations might play a significant role in differential survival of forest specialist seedlings and saplings. However, for saplings that survive to later growth stages, the higher water availability in alluvial forests may be an important contributing factor to the tall tree heights that we observed from the LiDAR data, which contributes directly to higher aboveground carbon densities in alluvial forests. Because the mudstone forests in Sepilok are also generally closer to anthropogenic forest edges than alluvial forests, edge effects, which have been shown to significantly influence large tree mortality and lower aboveground carbon, cannot be ruled out (Laurance *et al.* 2000; Qie *et al.* 2017; Ordway & Asner 2020).

At Danum, our results indicate that the region is comprised of one to three forest types that differ in canopy height, vertical structure, LMA, and foliar N and P. Two of these forest types (Danum 1 and 2) are found within the Danum 50-ha plot (Figure 1). Interestingly, this finding of two distinct forest types within the 50-ha plot aligns with recently identified differences in species composition and soil characteristics between the northeast corner and the remainder of the 50-ha plot (Cardon Pocovi 2019). The northeast corner (Danum 1) has lower species richness, diversity, stem density, and basal area compared to the rest of the plot (Danum 2), linked to less acidic soils with a higher cation exchange capacity and higher Ca, Mg, and Ni content (Cardon Pocovi 2019).

The implications of k selection

Rather than making an *a priori* decision about the number of clusters (k), we deliberately explored the capacity of remotely sensed data to reveal variation in ecological communities. Because the choice of k directly influences analysis outcomes, the method used for selecting k is important. The Gap_k and W_k elbow methods yielded different optimal numbers of clusters for Danum (1 versus 3 respectively). Similarly, a comparison of results based on k = 2, 3, and 4 in Sepilok revealed ecologically meaningful and interesting structural and functional differences in forest communities, consistent with a general hierarchical organization of forest community types at this site. In both cases, Gap_k pointed to an optimal number of clusters, and the reality of graduated transitions between forest communities on the ground at both sites emerged from our results when evaluating possible alternative values of k. Applying this methodology at broader scales will require similar decisions about k, which will either require user input, or the development of robust

396 automated algorithms for selecting the value of *k*. Our results indicate that the exploration of traits
397 that aggregate or separate communities as *k* changes is a valuable exercise.
398
399 ***Linking remote sensing and ground-based studies***
400 Our finding that aboveground carbon estimates derived from within plot boundaries corresponded
401 to estimates derived from larger mapped forest areas suggests that the inventory plots in this study
402 and the corresponding mapped forest types capture similar landscape-scale patterns. We found
403 significant differences in aboveground carbon and growth and mortality rates between the mapped
404 forests. The Sepilok alluvial and Danum 2 forest plots had similar aboveground carbon on average
405 (Sepilok alluvial: 231 Mg C ha^{-1}, Danum 2: 203 Mg C ha^{-1}). Both forest types are dominated by
406 large and fast-growing dipterocarp species, although the plots exhibited different stand-level
407 relative growth rates. Lower LMA and significantly higher leaf P and N, as well as a lower N:P in
408 Danum 2 compared to the Sepilok alluvial forest are consistent with the higher observed growth
409 rates. Similar mortality rates, despite varying growth, suggests high turnover rates in both forests,
410 perhaps with a greater influence of exogenous disturbance processes on mortality in the alluvial
411 forest (Margrove *et al.* 2015). The lack of structural differences between Danum 2 and alluvial
412 forests, despite significant differences in all leaf traits, suggests a strong control of trait driven
413 differences on growth even under similar vertical light environment conditions.
414
415 Higher aboveground carbon corresponded to lower mortality rates, except at the *kerangas* forest.
416 These *kerangas* forests, which had the highest LMA, lowest foliar P and N, and the lowest plot-
417 level aboveground carbon density (186 Mg C ha^{-1}), are known to have higher stem densities, lower
418 canopy heights, and long-lived leaves (Fox 1973, Dent *et al.* 2006; Jucker *et al.* 2018b), suggesting

419 well-developed strategies for nutrient retention (Turner *et al.* 1993; Turner 1994). In contrast, the
420 Sepilok sandstone forests, comprised of slow-growing dipterocarp species (Dent & Burslem 2009,
421 2016), had the highest median aboveground carbon density (236 Mg C ha^{-1}), with higher foliar P
422 and N, and lower LMA. Despite significant differences in aboveground carbon and demography,
423 the *kerangas* and sandstone forests did not differ in their LAI or canopy architecture (P:H);
424 although, maximum height, $Cover_{20}$, and $H_{peak\ LAI}$ were significantly higher in the sandstone forest.
425 The taller canopy and lower leaf nutrient concentrations are consistent with the low growth rate in
426 the sandstone forest, indicating a slow-growth strategy yielding larger trees and higher
427 aboveground carbon stocks. Similar LAI between the sandstone and *kerangas* forests, despite
428 differences in ecosystem dynamics, highlights a need to account for differences beyond LAI when
429 scaling processes from leaves to ecosystems.
430
431 ***Remotely sensed metrics beyond LAI***
432 LAI is considered one of the most important ecophysiological attributes of vegetation, and is
433 widely used in terrestrial ecosystem and biosphere models to upscale estimates of leaf-level
434 processes to ecosystem scales and model land atmosphere interactions (Jarvis & McNaughton
435 1986; Bonan *et al.* 1993). While there is significant variation in LAI between the world's major
436 biomes (Fang *et al.* 2019), we found that community scale differences in LAI across lowland
437 tropical forests in this study failed to capture important variation in canopy architecture, and thus
438 likely important differences in vertical light environments, between forest types. Instead, our
439 findings emphasize the importance of using additional LiDAR-derived metrics – maximum height,
440 $Cover_{20}$, P:H, and $H_{peak\ LAI}$ – and leaf traits to identify differences in forest canopy structure and
441 function.

Previous studies have emphasized the importance of the distribution of leaf area vertically for many canopy processes since the total amount of leaf surface area and its vertical organization can vary independently (Wu *et al.* 2000; Frolking *et al.* 2009; Shugart *et al.* 2010). Parker (2020) suggested that total LAI may not be directly relevant for many processes in ecosystems beyond LAI of three. Our findings provide additional evidence that vertical foliar distributions may be more important than the absolute amount of leaf area for characterizing differences across ecosystems.

Forest communities revealed by the cluster analyses were distributed along the leaf economic spectrum. On one end of the spectrum, the Danum 1 and 2 forests exhibited high nutrient concentrations and low LMA, while the sandstone and *kerangas* forests exhibited low nutrient concentrations and high LMA (Figure 5 and S6). Differences in forest structure varied across forest communities in ways that were orthogonal to the variation in leaf economic spectrum traits. Our PCA findings are consistent with the growth-survival and stature-recruitment (longevity-reproduction) tradeoff hypotheses (Díaz *et al.* 2016; Rüger *et al.* 2020). Interestingly, variation in V_{cmax} across communities exhibited significant correlation with stature (max H), which somewhat complicates the distinction between growth and stature tradeoffs (Figure S9). Bartholomew *et al.* (in press) found that, in Sepilok, variation in V_{cmax} and LMA was more related to nutrient availability than tree height, suggesting limited plasticity with changes in light availability and that responses to light availability in these ecosystems are likely constrained by nutrient availability. Importantly, we were able to detect and map these patterns at 30 m resolution, which will be available with spaceborne data. Since the main axes of variation in canopy properties correspond

to quantities that are measurable from spaceborne LiDAR and imaging spectroscopy, our approach offers a framework for large-scale mapping of functionally distinct tree communities that can be employed across highly diverse tropical forest ecosystems at regional and global scales.

Importantly, leaf P and LMA were critical for mapping functionally distinct tropical forests. Maximum canopy height and the fraction of canopy cover taller than 20 m were important for distinguishing forest types, although variation in structure alone was insufficient to capture observed differences in forest types. The accurate mapping of leaf P and LMA using imaging spectroscopy data from spaceborne sensors will thus be essential for ecological applications. Our remote sensing-based results re-affirm findings from field studies and yield new insights into the spatial turnover of canopy structure and functional traits, and the potential to reveal unstudied ecological communities across the tropics. In doing so, our results underscore potential synergies between ground-based and remote-sensing ecological analyses, whereby landscape-scale remote surveys can efficiently pinpoint locations that can be targeted as high priority for discovery-oriented fieldwork and plot measurements.

Acknowledgments

EO was supported by the Harvard University Center for the Environment Postdoctoral Fellowship program. The airborne science was completed with funding from the United Nations Development Programme and donations to GA from the Avatar Alliance Foundation, Margaret A. Cargill Foundation, David and Lucile Packard Foundation, Gordon and Betty Moore Foundation, Grantham Foundation for the Protection of the Environment, W. M. Keck Foundation, John D. and Catherine T. MacArthur Foundation, Andrew Mellon Foundation, Mary Anne Nyburg Baker

and G. Leonard Baker Jr., and William R. Hearst III, and Arizona State University. Plots at Sepilok were funded by the British Ecological Society, with the re-census in 2013 carried out by RN and LQ funded by an ERC Advanced Grant to OLP (291585, "T-FORCES"). The Danum plot is a core project of the Southeast Asia Rain Forest Research Partnership (SEARRP). We thank SEARRP partners, especially Yayasan Sabah for their support, and HSBC Malaysia and the University of Zurich for funding. We are grateful to the research assistants who are conducting the census, in particular the team leader Alex Karolus, and to Mike Bernados and Bill McDonald for species identifications. We thank Stuart Davies and Shameema Esufali for advice and training. This contribution is an output of ForestPlots.net approved research project no. 46. "Unraveling the role of prior disturbance histories in heterogeneous tropical forest responses to climate change."

References Cited

Antonarakis, A.S., Munger, J.W. & Moorcroft, P.R. (2014). Imaging spectroscopy- and lidar-derived estimates of canopy composition and structure to improve predictions of forest carbon fluxes and ecosystem dynamics. *Geophys. Res. Lett.*, 41, 2535–2542.

Asner, G.P., Brodrick, P.G., Philipson, C., Vaughn, N.R., Martin, R.E., Knapp, D.E., *et al.* (2018). Mapped aboveground carbon stocks to advance forest conservation and recovery in Malaysian Borneo. *Biol. Conserv.*, 217, 289–310.

Asner, G.P., Knapp, D.E., Boardman, J., Green, R.O., Kennedy-Bowdoin, T., Eastwood, M., *et al.* (2012). Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sens. Environ.*, 124, 454–465.

Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C.B., Sinca, F., *et al.* (2017). Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide

conservation. *Science*, 355, 385–389.

Asner, G.P. & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sens. Environ.*, 140, 614–624.

Bonan, G.B., Pollard, D. & Thompson, S.L. (1993). Influence of subgrid-scale heterogeneity in leaf area index, stomatal resistance, and soil moisture on grid-scale land-atmosphere interactions. *J. Clim.*, 6, 1882–1897.

Bongalov, B., Burslem, D.F.R.P., Jucker, T., Thompson, S.E.D., Rosindell, J., Swinfield, T., *et al.* (2019). Reconciling the contribution of environmental and stochastic structuring of tropical forest diversity through the lens of imaging spectroscopy. *Ecol. Lett.*, 22, 1608–1619.

Born, J., Bagchi, R., Burslem, D., Nilus, R., Tellenbach, C., Pluess, A.R., *et al.* (2015). Differential Responses of Dipterocarp Seedlings to Soil Moisture and Microtopography. *Biotropica*, 47, 49–58.

Born, J., Pluess, A.R., Burslem, D.F.R.P., Nilus, R., Maycock, C.R. & Ghazoul, J. (2014). Differing Life History Characteristics Support Coexistence of Tree Soil Generalist and Specialist Species in Tropical Rain Forests. *Biotropica*, 46, 58–68.

Brando, P.M., Paolucci, L., Ummenhofer, C.C., Ordway, E.M., Hartmann, H., Cattau, M.E., *et al.* (2019). Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis. *Annu. Rev. Earth Planet. Sci.*, 47, 555–581.

Cardon Pocovi, J.M. (2019). Drivers of species richness , tree community composition and diversity in a lowland dipterocarp forest in Sabah, Malaysia. *MSc thesis, Univ. Aberdeen, UK*, 1–46.

Chadwick, K.D. & Asner, G.P. (2020). Geomorphic transience moderates topographic controls

on tropical canopy foliar traits. *Ecol. Lett.*, 23, 1276–1286.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., *et al.* (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.*, 20, 3177–3190.

Chazdon, R.L. (2003). Tropical forest recovery: legacies of human impact and natural disturbances. *Perspect. Plant Ecol. Evol. Syst.*, 6, 51–71.

Condit, R., Ashton, P., Bunyavejchewin, S., Dattaraja, H.S., Davies, S., Esufali, S., *et al.* (2006). The importance of demographic niches to tree diversity. *Science*, 313, 98–101.

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., *et al.* (2017). Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sens. Environ.*, 194, 77–88.

Corlett, R.T. (2016). Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Divers.*, 38, 10–16.

Dent, D.H., Bagchi, R., Robinson, D., Majalap-Lee, N. & Burslem, D.F.R.P. (2006). Nutrient fluxes via litterfall and leaf litter decomposition vary across a gradient of soil nutrient supply in a lowland tropical rain forest. *Plant Soil*, 288, 197–215.

Dent, D.H. & Burslem, D.F.R.P. (2009). Performance trade-offs driven by morphological plasticity contribute to habitat specialization of bornean tree species. *Biotropica*, 41, 424–434.

Dent, D.H. & Burslem, D.F.R.P. (2016). Leaf traits of dipterocarp species with contrasting distributions across a gradient of nutrient and light availability. *Plant Ecol. Divers.*, 9, 521–533.

Detto, M., Asner, G.P., Muller-Landau, H.C. & Sonnentag, O. (2015). Spatial variability in

tropical forest leaf area density from multireturn lidar and modeling. *J. Geophys. Res. Biogeosciences*, 120, 294–309.

DeWalt, S.J., Ickes, K., Nilus, R., Harms, K.E. & Burslem, D.F.R.P. (2006). Liana habitat associations and community structure in a Bornean lowland tropical forest. *Plant Ecol.*, 186, 203–216.

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., *et al.* (2016). The global spectrum of plant form and function. *Nature*, 529, 167–171.

Draper, F.C., Baraloto, C., Brodrick, P.G., Phillips, O.L., Martinez, R.V., Honorio Coronado, E.N., *et al.* (2019). Imaging spectroscopy predicts variable distance decay across contrasting Amazonian tree communities. *J. Ecol.*, 107, 696–710.

Dubayah, R., Hofton, M., Blair, J.B., Armston, J., Tang, H. & Luthcke, S. (2020a). *GEDI L2A Elevation and Height Metrics Data Global Footprint Level V001 [Data set]*. *NASA EOSDIS L. Process. DAAC*. Available at: https://doi.org/10.5067/GEDI/GEDI02_A.001. Last accessed 19 June 2021.

Dubayah, R., Luthcke, S., Blair, J.B., Hofton, M., Armston, J. & Tang, H. (2020b). *GEDI L1B Geolocated Waveform Data Global Footprint Level V001 [Data set]*. *NASA EOSDIS L. Process. DAAC*. Available at: https://doi.org/10.5067/GEDI/GEDI01_B.001. Last accessed 19 June 2021.

Dubayah, R., Tang, H., Armston, J., Luthcke, S., Hofton, M. & Blair, J.B. (2020c). *GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V001 [Data set]*. *NASA EOSDIS L. Process. DAAC*. Available at: https://doi.org/10.5067/GEDI/GEDI02_B.001. Last accessed 19 June 2021.

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., *et al.* (2010).

Estimation of tropical forest height and biomass dynamics using lidar remote sensing at la Selva, Costa Rica. *J. Geophys. Res. Biogeosciences*, 115, 1–17.

ESA, E.S.A. (2021a). *CHIME (Copernicus Hyperspectral Imaging Mission for the Environment)*. *eoPortal Dir.* Available at: https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/chime-copernicus. Last accessed 19 June 2021.

ESA, E.S.A. (2021b). *PRISMA (Hyperspectral Precursor and Application Mission)*. *eoPortal Dir.* Available at: https://directory.eoportal.org/web/eoportal/satellite-missions/p/prisma-hyperspectral. Last accessed 19 June 2021.

Fang, H., Baret, F., Plummer, S. & Schaepman-Strub, G. (2019). An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Rev. Geophys.*, 57, 739–799.

Féret, J.B. & Asner, G.P. (2014). Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecol. Appl.*, 24, 1289–1296.

ForestGEO. (2021). *ForestGEO Data Portal*. *Glob. Earth Obs. Netw.* Available at: http://ctfs.si.edu/datarequest/. Last accessed 13 July 2020.

Frolking, S., Palace, M.W., Clark, D.B., Chambers, J.Q., Shugart, H.H. & Hurtt, G.C. (2009). Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure. *J. Geophys. Res. Biogeosciences*, 114.

GAC, G.A.C. & TBE, T.B.E. (2021). *DESIS: DLR Earth Sensing Imaging Spectrometer*. *Teledyne*. Available at: https://tbe.com/geospatial/desis. Last accessed 19 June 2021.

Givnish, T.J. (1999). On the causes of gradients in tropical tree diversity. *J. Ecol.*, 87, 193–210.

Hartigan, J.A. & Wong, M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C (Applied Stat.*, 28, 100–108.

Hulshof, C.M. & Spasojevic, M.J. (2020). The edaphic control of plant diversity. *Glob. Ecol. Biogeogr.*, 29, 1634–1650.

Jarvis, P.G. & McNaughton, K.G. (1986). Stomatal Control of Transpiration: Scaling Up from Leaf to Region. *Adv. Ecol. Res.*, 15, 1–49.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., *et al.* (2016). Monitoring plant functional diversity from space. *Nat. Plants*, 2, 1–5.

Jucker, T., Asner, G.P., Dalponte, M., Brodrick, P.G., Philipson, C.D., Vaughn, N.R., *et al.* (2018a). Estimating aboveground carbon density and its uncertainty in Borneo's structurally complex tropical forests using airborne laser scanning. *Biogeosciences*, 15, 3811–3830.

Jucker, T., Bongalov, B., Burslem, D.F.R.P., Nilus, R., Dalponte, M., Lewis, S.L., *et al.* (2018b). Topography shapes the structure, composition and function of tropical forest landscapes. *Ecol. Lett.*, 21, 989–1000.

Laurance, W.F., Delamonica, P., Laurance, S.G., Vasconcelos, H.L. & Lovejoy, T.E. (2000). Rainforest fragmentation kills big trees. *Nature*, 404, 836.

Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M. & Phillips, O.L. (2011). ForestPlots.net: A web application and research tool to manage and analyse tropical forest plot data. *J. Veg. Sci.*, 22, 610–613.

Loubota Panzou, G.J., Fayolle, A., Jucker, T., Phillips, O.L., Bohlman, S., Banin, L.F., *et al.* (2020). Pantropical variability in tree crown allometry. *Glob. Ecol. Biogeogr.*, geb.13231.

Margrove, J.A., Burslem, D.F.R.P., Ghazoul, J., Khoo, E., Kettle, C.J. & Maycock, C.R. (2015). Impacts of an Extreme Precipitation Event on Dipterocarp Mortality and Habitat Filtering in

a Bornean Tropical Rain Forest. *Biotropica*, 47, 66–76.

Milligan, G.W. & Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50, 159–179.

Muscarella, R., Kolyaie, S., Morton, D.C., Zimmerman, J.K. & Uriarte, M. (2020). Effects of topography on tropical forest structure depend on climate context. *J. Ecol.*, 108, 145–159.

Myers, N. (1988). Threatened biotas: "Hot Spots" in tropical forests. *Environmentalist*, 8, 187–208.

NASA JPL, N.J.P.L. (2021). *Surface Biology Geology*. Available at: https://sbg.jpl.nasa.gov/. Last accessed 19 June 2021.

Newbery, D.M. (1991). Floristic variation within kerangas (heath) forest: re-evaluation of data from Sarawak and Brunei. *Vegetatio*, 96, 43–86.

Nilus, R. (2004). Effect o f edaphic variation on forest structure, dynamics, diversity and regeneration in a lowland tropical rainforest in Borneo. *PhD thesis, Univ. Aberdeen, UK*.

Nilus, R., Maycock, C.R., Majalap-Lee, N. & Burslem, D.F.R.P. (2011). Nutrient limitation of tree seedling growth in three soil types found in Sabah. *J. Trop. For. Sci.*, 23, 133–142.

Ordway, E.M. & Asner, G.P. (2020). Carbon declines along tropical forest edges correspond to heterogeneous effects on canopy structure and function. *Proc. Natl. Acad. Sci. U. S. A.*, 117, 7863–7870.

Osborne, B.B., Nasto, M.K., Soper, F.M., Asner, G.P., Balzotti, C.S., Cleveland, C.C., *et al.* (2020). Leaf litter inputs reinforce islands of nitrogen fertility in a lowland tropical forest. *Biogeochemistry*, 147, 293–306.

Parker, G.G. (2020). Tamm review: Leaf Area Index (LAI) is both a determinant and a consequence of important processes in vegetation canopies. *For. Ecol. Manage.*, 477.

Qie, L., Lewis, S.L., Sullivan, M.J.P., Lopez-Gonzalez, G., Pickavance, G.C., Sunderland, T., *et al.* (2017). Long-term carbon sink in Borneo's forests halted by drought and vulnerable to edge effects. *Nat. Commun.*, 8.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J. & Hérault, B. (2017). BIOMASS: an R package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods Ecol. Evol.*, 8, 1163–1167.

Rüger, N., Condit, R., Dent, D.H., DeWalt, S.J., Hubbell, S.P., Lichstein, J.W., *et al.* (2020). Demographic tradeoffs predict tropical forest dynamics. *Science (80-.).*, 368, 165–168.

Schimel, D.S., Asner, G.P. & Moorcroft, P. (2013). Observing changing ecological diversity in the Anthropocene. *Front. Ecol. Environ.*, 11, 129–137.

Shenkin, A., Chandler, C.J., Boyd, D.S., Jackson, T., Disney, M., Majalap, N., *et al.* (2019). The World's Tallest Tropical Tree in Three Dimensions. *Front. For. Glob. Chang.*, 2, 1–5.

Shugart, H.H., Saatchi, S. & Hall, F.G. (2010). Importance of structure and its measurement in quantifying function of forest ecosystems. *J. Geophys. Res. Biogeosciences*, 115, 1–16.

Tibshirani, R., Walther, G. & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Stat. Methodol.*, 63, 411–423.

Townsend, A.R., Asner, G.P. & Cleveland, C.C. (2008). The biogeochemical heterogeneity of tropical forests. *Trends Ecol. Evol.*, 23, 424–431.

Turner, I.M. (1994). A Quantitative Analysis of Leaf Form in Woody Plants from the World's Major Broadleaved Forest Types. *J. Biogeogr.*, 21, 413.

Turner, I.M., Choong, M.F., Tan, H.T.W. & Lucas, P.W. (1993). How tough are sclerophylls? *Ann. Bot.*

Walker, A.P., Beckerman, A.P., Gu, L., Kattge, J., Cernusak, L.A., Domingues, T.F., *et al.*

(2014). The relationship of leaf photosynthetic traits - Vcmax and Jmax - to leaf nitrogen, leaf phosphorus, and specific leaf area: A meta-analysis and modeling study. *Ecol. Evol.*, 4, 3218–3235.

Wu, J., Liu, Y. & Jelinski, D.E. (2000). Effects of leaf area profiles and canopy stratification on simulated energy fluxes: The problem of vertical spatial scale. *Ecol. Modell.*, 134, 283–297.

Tables & Figures

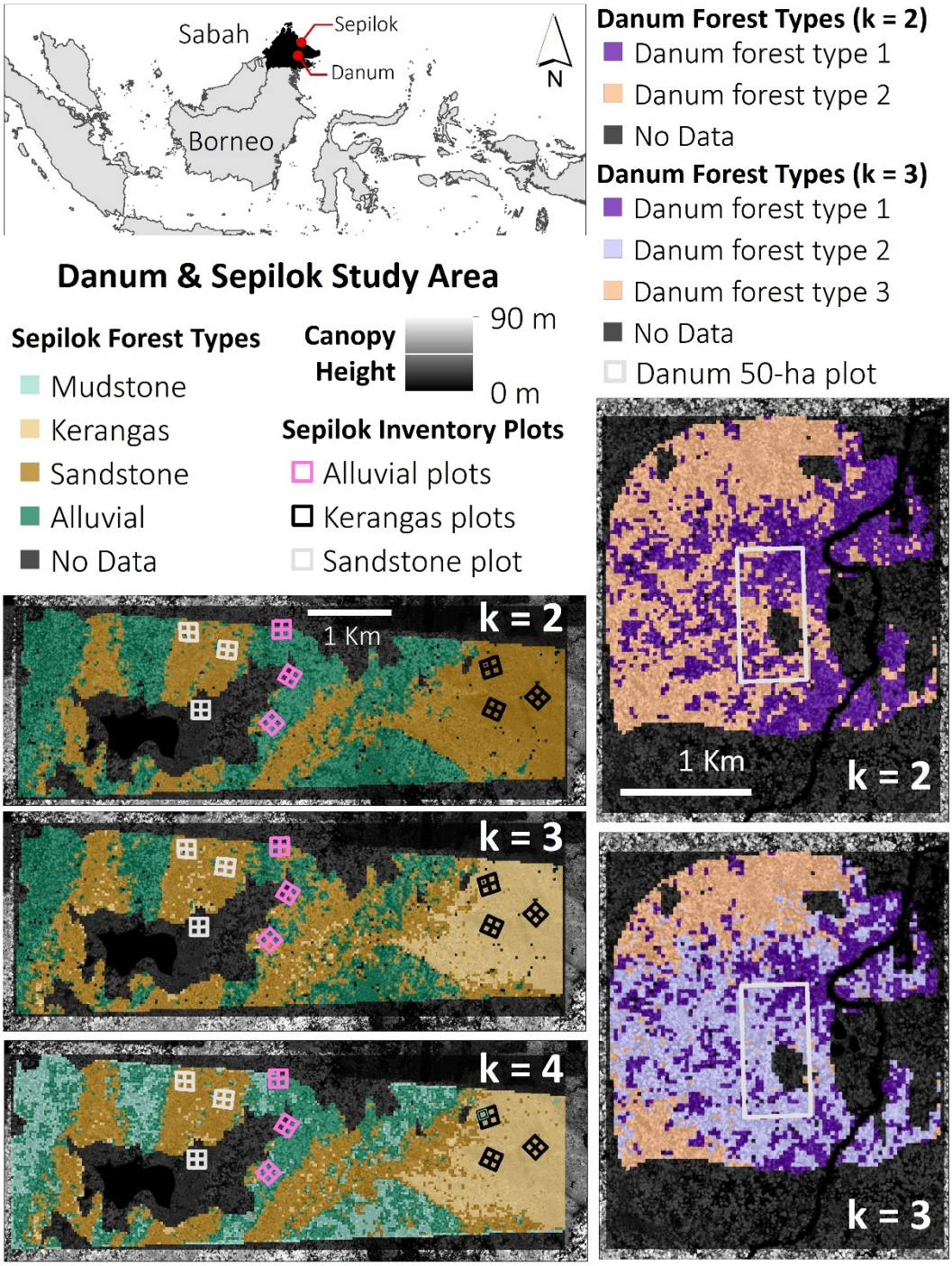

Figure 1. Results from PCA and *k*-means clustering of 10 variables across forest ecosystems in Sepilok Forest Reserve for $k = 2$, 3, and 4, and in Danum Valley Conservation Area around the 50-ha ForestGEO plot for $k = 2$ and 3. The partitioning of the alluvial forest into alluvial and mudstone forest types is revealed with $k = 4$. No Data indicates omitted pixels and pixels that were cloud, cloud shadow, and water masked.

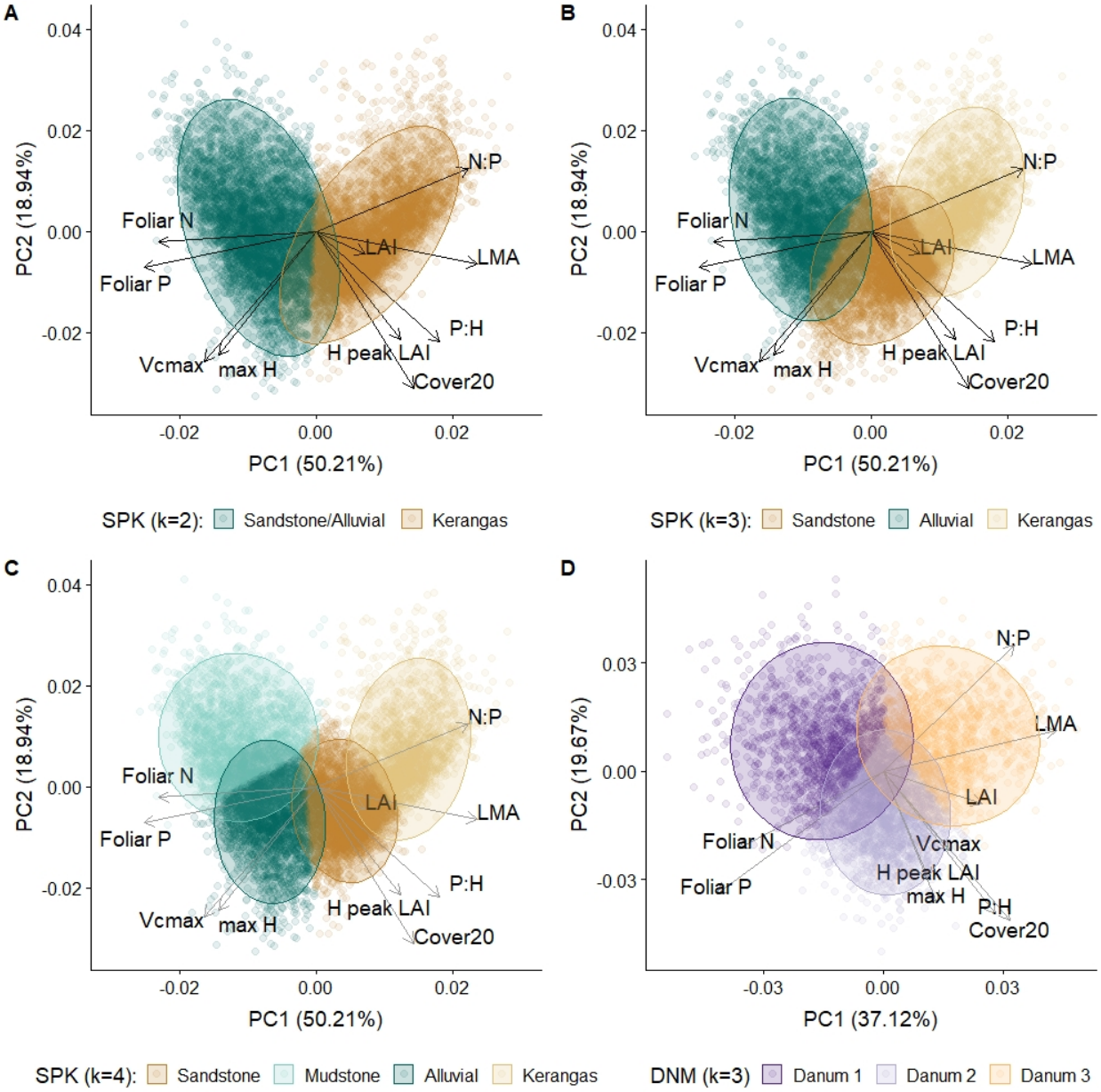

Figure 2. The first two loadings from the principal component analysis at Sepilok (a-c) and Danum (d). (a-c) illustrate the partitioning of pixels into k = 2, 3, and 4 clusters at Sepilok. (d) illustrates k = 3 clusters at Danum.

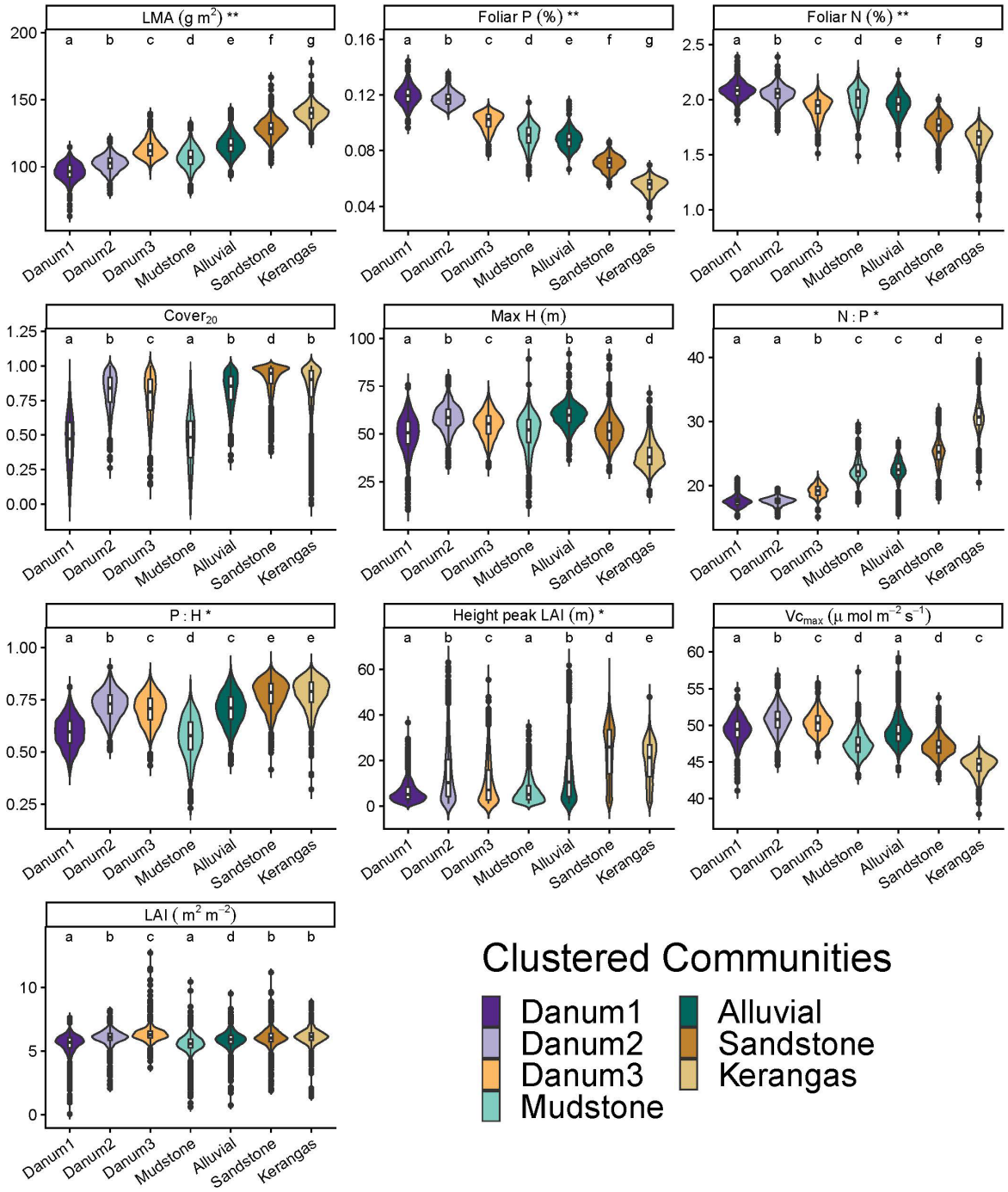

Figure 3. Trait distributions by cluster for Sepilok $k = 4$ and Danum $k = 3$. Forest communities are ordered based on their median LMA to illustrate differences in traits for communities that vary along the leaf economics spectrum. Identical letters represent clusters where there is no significant difference between forests based on one-way ANOVA tests ($p < 0.01$). ** = traits that varied significantly between all seven forest types. * = traits that varied significantly between at least five forest types.

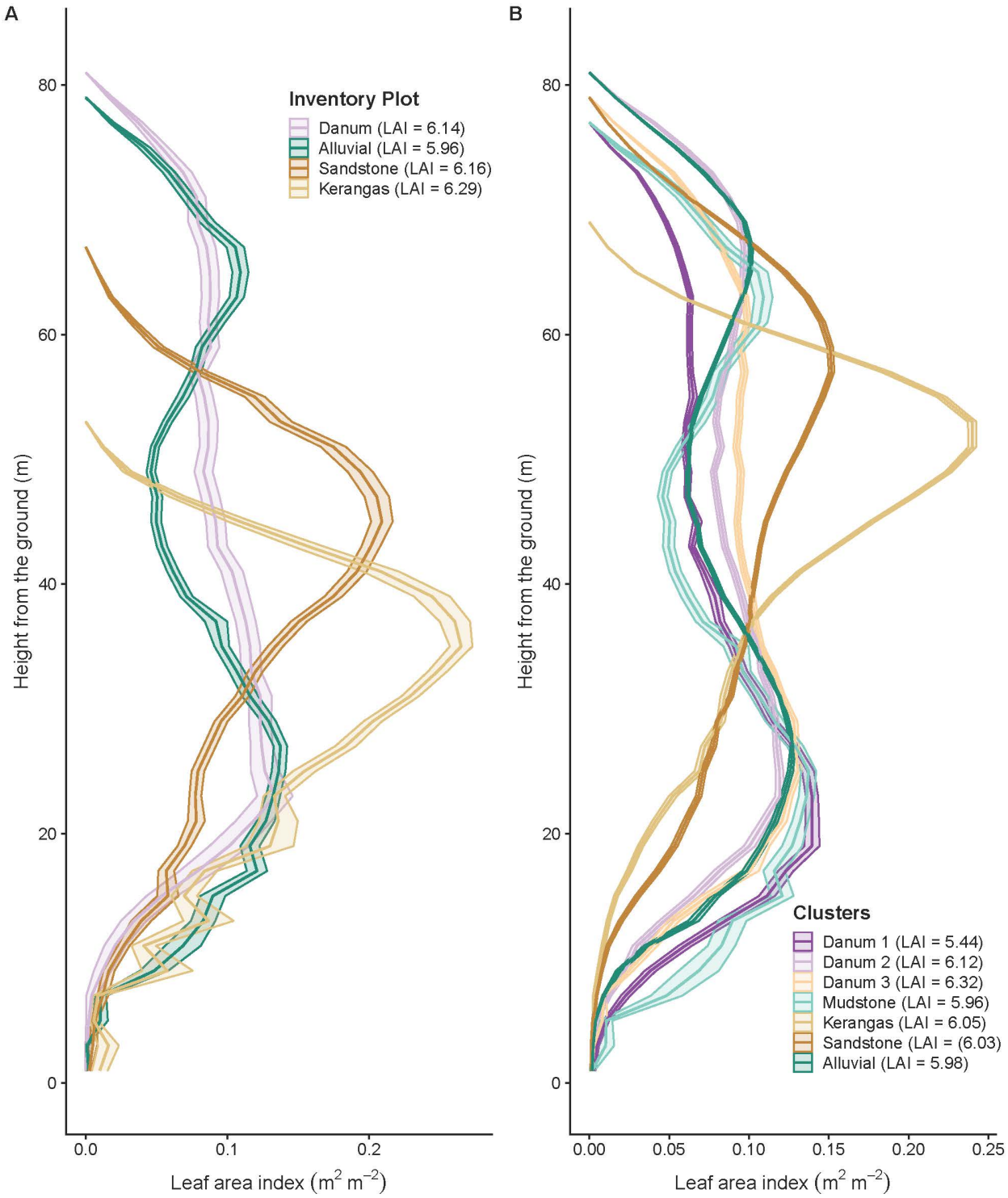

Figure 4. Vertical LAI profiles for all pixels within each inventory plot (a) and forest community identified based on $k = 3$ clusters at Danum and $k = 4$ clusters at Sepilok (b).

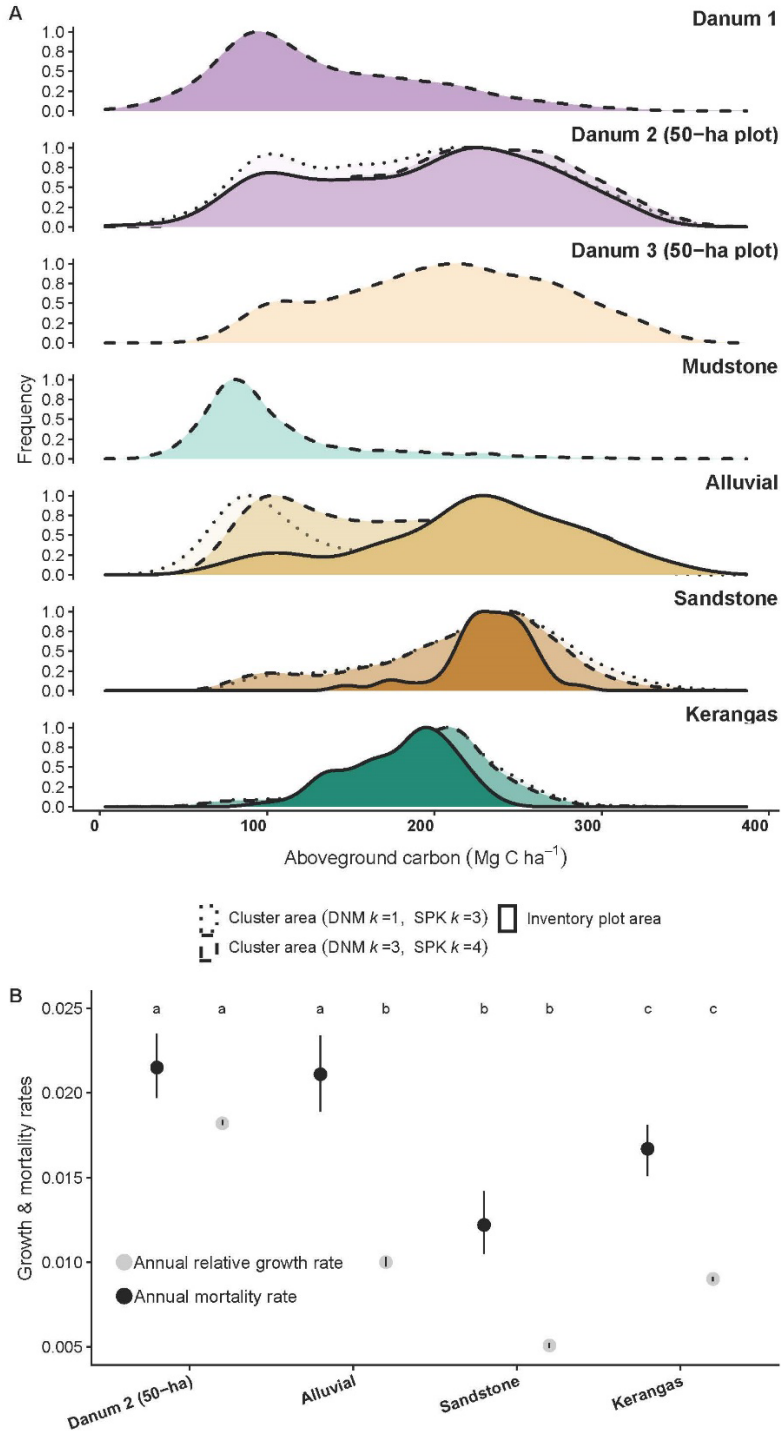

Figure 5. (a) Aboveground carbon density for each field inventory plot (solid line) compared to aboveground carbon for the entire forest type based on cluster results where $k = 1$ for Danum and $k = 3$ for Sepilok (dotted line) and $k = 3$ for Danum and $k = 4$ for Sepilok (dashed line). (b) Annual relative growth (grey) and mortality (black) rates for each forest type calculated from forest inventory plot data. Identical letters represent inventory plots with no significant difference in terms of carbon, mortality rates, and growth rates respectively, based on one-way ANOVA tests ($p < 0.01$).

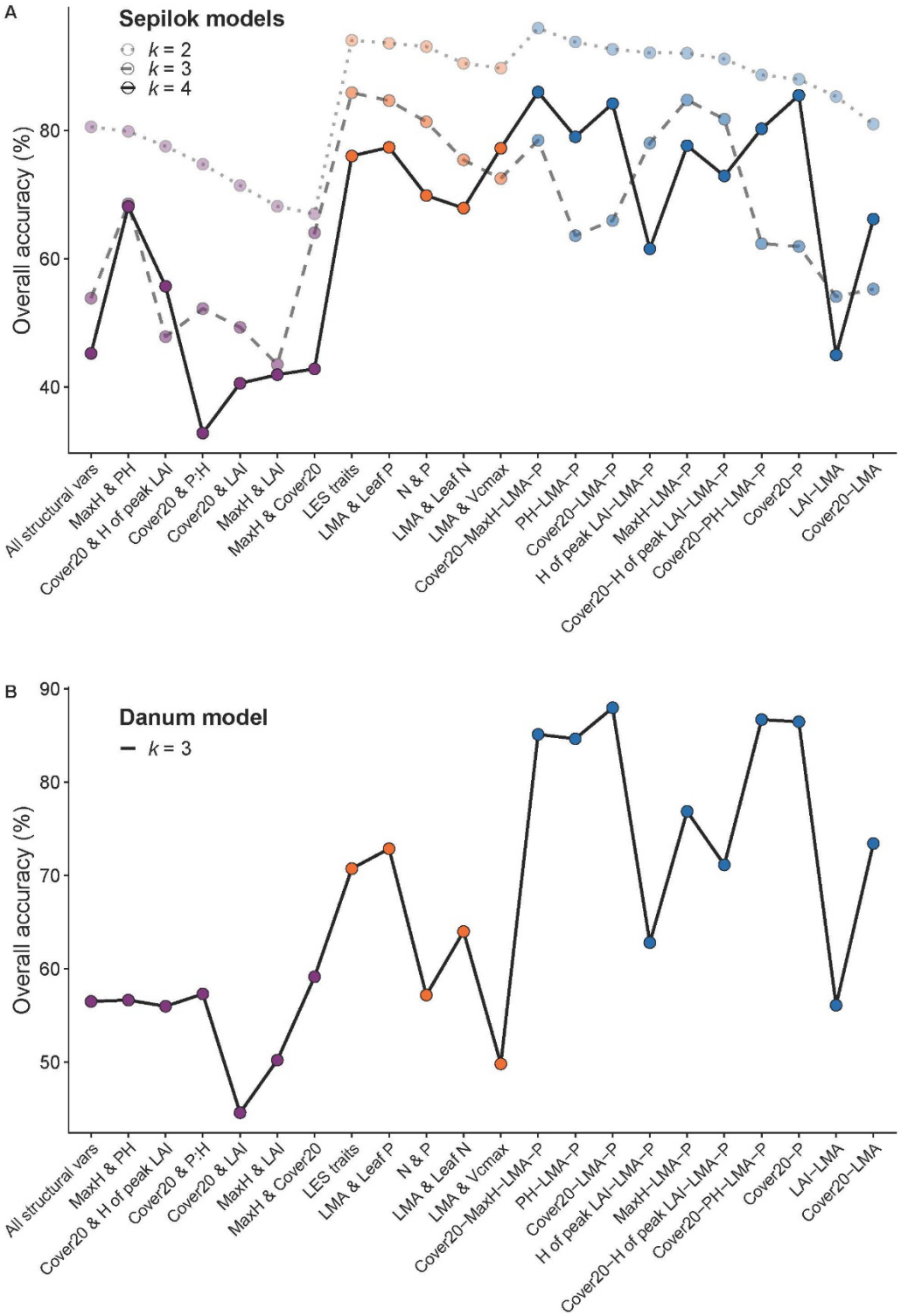

Figure 6. Change in overall accuracy for reduced *k*-means clustering models using structural variables (purple), leaf trait variables (orange), and combinations of structural and leaf trait variables (blue) for k = 2, 3, and 4 for Sepilok and k = 3 for Danum. All are compared to the full 10-variable *k*-means clustering analysis for Sepilok (A) and Danum (B). LES: leaf economic spectrum.

SUPPORTING INFORMATION

Mapping fine-scale variation in diverse tropical forests with distinct ecological dynamics requires few leaf traits and structural attributes

Elsa M. Ordway[1,2]*, Gregory P. Asner[3], David F.R.P. Burslem[4], Simon L. Lewis[5,6], Reuben Nilus[7], Roberta Martin[3], Michael J. O'Brien[8], Oliver L. Phillips[5], Lan Qie[9], Nicolas R. Vaughn[3], Paul R. Moorcroft[1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

[2]Department of Ecology and Evolutionary Biology, UCLA, 612 Charles E. Young Drive South, Los Angeles, CA 90095, USA

[3]Center for Global Discovery and Conservation Science, Arizona State University, 1001 McAllister Ave., Tempe, AZ 85281, USA

[4]School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, U.K.

[5]School of Geography, University of Leeds, Leeds LS2 9JT, U.K.

[6]Department of Geography, University College London, London. WC1E 6BT.

[7]Sabah Forestry Department, Forest Research Centre, Sandakan, Sabah, MY

[8]Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, c/ Tulipán s/n., E-28933 Móstoles, Spain

[9]School of Life Sciences, University of Lincoln, Lincoln LN6 7DL, U.K.

This file includes:

Supplementary Methods

Tables S1-S2

Figures S1-S12

Supplementary Methods

Airborne remote sensing data and processing

LiDAR data were collected at a minimum pulse density of 1.14 pulses m^{-2} (4.5 returns m^{-2} in forested regions) and processed to top-of-canopy height (TCH, m) at 2 m resolution using the LAStools software suite (Rapidlasso, GmbH, Gilching, Germany). Using the LiDAR top-of-canopy height (TCH) data, maximum height was calculated as the 99th percentile of TCH for every resampled 30 m pixel. The 2 m TCH data was also used to calculate the fraction of each 30 m resolution pixel that exceeded 20 m, known as $Cover_{20}$ (Coomes *et al.*, 2017; Jucker *et al.*, 2018). Canopy cover at or above 20 m aboveground correlates with plot level basal area in the region (Coomes *et al.*, 2017). Vertical LAI profiles, estimated from the LiDAR data using the spherical theoretical leaf angle distribution method described in (Detto *et al.*, 2015), and binned vertically every 2 meters, were used to identify the height aboveground where maximum (i.e., peak) LAI occurred. We calculated the P:H ratio at 5 m resolution using the method described in (G. P. Asner *et al.*, 2014), where P refers to the height aboveground at maximum canopy volume within the 5 m pixel resolution, and H is the 99th percentile of total canopy height. Areas with high P:H values correspond to forests with foliage vertically partitioned high in the canopy, while low P:H values indicate foliage vertically partitioned nearer to the ground.

Imaging spectroscopy data were collected at 4 m ground-level resolution using a visible to shortwave (VSWIR) imaging spectrometer that measures spectral radiance in 427 channels at 5 nm bandwidths from 350-2485 nm. Radiance data were averaged to 10 nm bands, atmospherically corrected using the ACORN-6LX software, and transformed to apparent surface reflectance. After averaging the radiance data to 10 nm bands, the ACORN-6LX atmospheric correction software was used to transform the imaging spectroscopy radiance data to apparent surface reflectance

(Imspec LLC, Glendale, CA USA). Each study site was processed through ACORN using mean flight conditions (elevation, collection altitude, sensor and solar view angles, and time) specific to that site.

Crown-level foliar chemical traits and LMA were estimated by linking spectral observations with field-based measurements of foliar characteristics (Martin *et al.*, 2018), summarized here. Individual trees identified as visible within the imaging spectroscopy reflectance data were sampled across 13 field locations in Sabah, including 13 crowns in the *kerangas* forest, 35 crowns in the alluvial forest, 14 crowns in the sandstone forest, and 76 crowns in Danum Valley (Table S1). Mature top-of-canopy leaf samples were collected from at least two fully sunlit branches of each tree. Leaf samples were scanned, weighed, and dried for at least 72 hours before dry mass (DM) was measured. Leaf mass per area (LMA) was calculated as g DM m^{-2}. Detailed descriptions of chemical analysis protocols, standards, and instruments used to extract total element concentrations of N and P are described in (Gregory P. Asner *et al.*, 2014; Gregory P. Asner & Martin, 2011, 2016).

To ensure accurate comparison between laboratory measurements of N, P, and LMA and the corresponding airborne spectroscopy data, the spectral data were restricted to fully sunlit portions of tree crowns. After applying a hand-generated cloud and cloud-shadow mask, spectral data were filtered based on a 2 m height requirement to exclude bare ground and non-forest vegetation, and a Normalized Difference Vegetation Index (NDVI) threshold of ≥ 0.75 to ensure sufficient foliar cover for pixels included. Spectral bands in the 440-1320, 1500-1760, and 2040-2440 nm

wavelengths were omitted due to high atmospheric water absorption. Filtered spectral data were brightness normalized to eliminate anomalously low or high reflectance values.

A partial least squares regression model was generated to relate the brightness normalized surface reflectance spectra to lab-assayed foliar traits across the state of Sabah, Malaysia, and this model was subsequently applied across surface reflectance imagery to generate foliar trait maps. Crown-level mass-based foliar N (%) and P (%) concentrations and LMA values were predicted with R^2 = 0.54, 0.65, 0.81 and normalized root mean squared error (RMSE) = 0.43, 0.03, and 23.90 (Martin *et al.*, 2018). The mapped foliar traits were used to calculate foliar N:P ratios across the Sepilok study area. N:P ratios are broadly used to infer the potential limitation of N or P with respect to primary productivity (Koerselman & Meuleman, 1996; Tessier & Raynal, 2003). Low N:P values, less than circa 14, are considered to indicate N limitation, while values > 16 indicate P limitation (Townsend *et al.*, 2007).

PCA and k-means cluster data processing

We conducted the principal component and clustering analyses at the following spatial resolutions (m): 4, 8, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 170, 200. Prior to analysis at each resolution, all variables were resampled to the same resolution and stacked. The collection of variables at each pixel location was treated as a sample for subsequent analysis. The height of peak LAI and N:P were log transformed, and $Cover_{20}$ and P:H were cube root transformed to normalize their distributions. Each variable was then centered and scaled across all samples. Pixels without a value for every single trait were omitted, and a 5 m height requirement was applied to remove bare ground and non-forest vegetation. Improvement in the degree of explained variance saturated around 20-40 m resolution (Figure S3), which corresponds to the maximum crown diameter for

canopy trees in the region (Loubota Panzou *et al.*, 2020; Shenkin *et al.*, 2019). We report results from analyses at 30 m resolution to assess the feasibility of this framework at the planned 30 m resolution of SBG imaging spectroscopy data.

References Cited

Asner, G. P., Anderson, C. B., Martin, R. E., Knapp, D. E., Tupayachi, R., Sinca, F., & Malhi, Y. (2014). Landscape-scale changes in forest structure and functional traits along an Andes-to-Amazon elevation gradient. *Biogeosciences*, *11*(3), 843–856. https://doi.org/10.5194/bg-11-843-2014

Asner, Gregory P., & Martin, R. E. (2011). Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytologist*, *189*(4), 999–1012. https://doi.org/10.1111/j.1469-8137.2010.03549.x

Asner, Gregory P., & Martin, R. E. (2016). Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. *Global Ecology and Conservation*, *8*, 212–219. https://doi.org/10.1016/j.gecco.2016.09.010

Asner, Gregory P., Martin, R. E., Tupayachi, R., Anderson, C. B., Sinca, F., Carranza-Jiménez, L., & Martinez, P. (2014). Amazonian functional diversity from forest canopy chemical assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15), 5604–5609. https://doi.org/10.1073/pnas.1401181111

Coomes, D. A., Dalponte, M., Jucker, T., Asner, G. P., Banin, L. F., Burslem, D. F. R. P., Lewis, S. L., Nilus, R., Phillips, O. L., Phua, M. H., & Qie, L. (2017). Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sensing of Environment*, *194*, 77–88. https://doi.org/10.1016/j.rse.2017.03.017

Detto, M., Asner, G. P., Muller-Landau, H. C., & Sonnentag, O. (2015). Spatial variability in tropical forest leaf area density from multireturn lidar and modeling. *Journal of Geophysical Research: Biogeosciences*, *120*, 294–309. https://doi.org/10.1002/2014JG002774.Received

Jucker, T., Asner, G. P., Dalponte, M., Brodrick, P. G., Philipson, C. D., Vaughn, N. R., Arn Teh, Y., Brelsford, C., Burslem, D. F. R. P., Deere, N. J., Ewers, R. M., Kvasnica, J., Lewis, S. L., Malhi, Y., Milne, S., Nilus, R., Pfeifer, M., Phillips, O. L., Qie, L., … Coomes, D. A. (2018). Estimating aboveground carbon density and its uncertainty in Borneo's structurally complex tropical forests using airborne laser scanning. *Biogeosciences*, *15*(12), 3811–3830. https://doi.org/10.5194/bg-15-3811-2018

Koerselman, W., & Meuleman, A. F. M. (1996). The Vegetation N:P Ratio: a New Tool to Detect the Nature of Nutrient Limitation. *British Ecological Society*, *33*(6), 1441–1450.

Loubota Panzou, G. J., Fayolle, A., Jucker, T., Phillips, O. L., Bohlman, S., Banin, L. F., Lewis, S. L., Affum-Baffoe, K., Alves, L. F., Antin, C., Arets, E., Arroyo, L., Baker, T. R., Barbier,

N., Beeckman, H., Berger, U., Bocko, Y. E., Bongers, F., Bowers, S., … Feldpausch, T. R. (2020). Pantropical variability in tree crown allometry. *Global Ecology and Biogeography*, *January*, geb.13231. https://doi.org/10.1111/geb.13231

Martin, R. E., Dana Chadwick, K., Brodrick, P. G., Carranza-Jimenez, L., Vaughn, N. R., & Asner, G. P. (2018). An approach for foliar trait retrieval from airborne imaging spectroscopy of tropical forests. *Remote Sensing*, *10*(2), 199. https://doi.org/10.3390/rs10020199

Shenkin, A., Chandler, C. J., Boyd, D. S., Jackson, T., Disney, M., Majalap, N., Nilus, R., Foody, G., bin Jami, J., Reynolds, G., Wilkes, P., Cutler, M. E. J., van der Heijden, G. M. F., Burslem, D. F. R. P., Coomes, D. A., Bentley, L. P., & Malhi, Y. (2019). The World's Tallest Tropical Tree in Three Dimensions. *Frontiers in Forests and Global Change*, *2*(June), 1–5. https://doi.org/10.3389/ffgc.2019.00032

Tessier, J. T., & Raynal, D. J. (2003). Use of nitrogen to phosphorus ratios in plant tissue as an indicator of nutrient limitation and nitrogen saturation. *Journal of Applied Ecology*, *40*(3), 523–534. https://doi.org/10.1046/j.1365-2664.2003.00820.x

Townsend, A. R., Cleveland, C. C., Asner, G. P., & Bustamante, M. M. C. (2007). Controls over foliar N:P ratios in tropical rain forests. *Ecology*, *88*(1), 107–118. https://doi.org/10.1890/0012-9658(2007)88[107:COFNRI]2.0.CO;2

Supporting Tables

Table S1. Ten variables used in the PCA and k-means cluster analysis.

Cluster Analysis Traits	Abbreviation	Unit	Resolution*	Description
Maximum height	Max H	m	2	Maximum height of the forest canopy.
Canopy cover at 20 m	$Cover_{20}$	%	30	Fraction of canopy cover ≥ 20 m height aboveground.
Leaf area index	LAI	$m^2\ m^{-2}$	50	Leaf area index estimated using the spherical theoretical leaf angle distribution method.
Height of peak LAI	$H_{peak\ LAI}$	m	50	Height above ground of the peak LAI from the vertical LAI profile.
Canopy shape ratio	P:H	unitless	5	P: Height above ground at max. canopy volume. H: 99th percentile of total canopy height.
Leaf mass per area	LMA	$g\ m^2$	4	The ratio of leaf dry mass to leaf area.
Foliar nitrogen	N	%	4	Mass-based foliar nitrogen concentration.
Foliar phosphorus	P	%	4	Mass-based foliar phosphorus concentration.
Nitrogen to phosphorus ratio	N:P	unitless	4	Foliar nitrogen to phosphorus ratio.
Max. photosynthetic capacity	V_{cmax}	$\mu mol\ m^{-2}\ s^{-1}$	4	Maximum rate of Rubisco carboxylase activity, a metric of photosynthetic capacity.

*Original resolution of data prior to resampling for analysis.

Table S2. Number of tree crowns and species sampled per site for remotely sensed canopy foliar trait estimation as described in (Martin *et al.*, 2018).

Site	**N crowns**	**N species**	**Example species**
Sepilok alluvial/ mudstone	30	24	*Dipterocarpus applanatus, Dryobalanops lanceolata, Eusideroxylon zwageri, Parashorea tomentella, Shorea johorensis, Shorea leprosula*
Sepilok sandstone	14	11	*Dipterocarpus acutangulus, Hopea baccarina, Shorea beccariana, Shorea multiflora, Shorea smithiana,*
Sepilok *kerangas*	13	11	*Cotylelobium melanoxycan, Ixonanthus reticulata, Shorea multiflora, Koompassia malaccensis*
Danum Valley	58	40	*Eusideroxylon zwageri, Koompassia excelsa, Samanea saman, Shorea faguetiana, Shorea johorensis, Shorea leprosula, Shorea parvifolia*

Supporting Figures

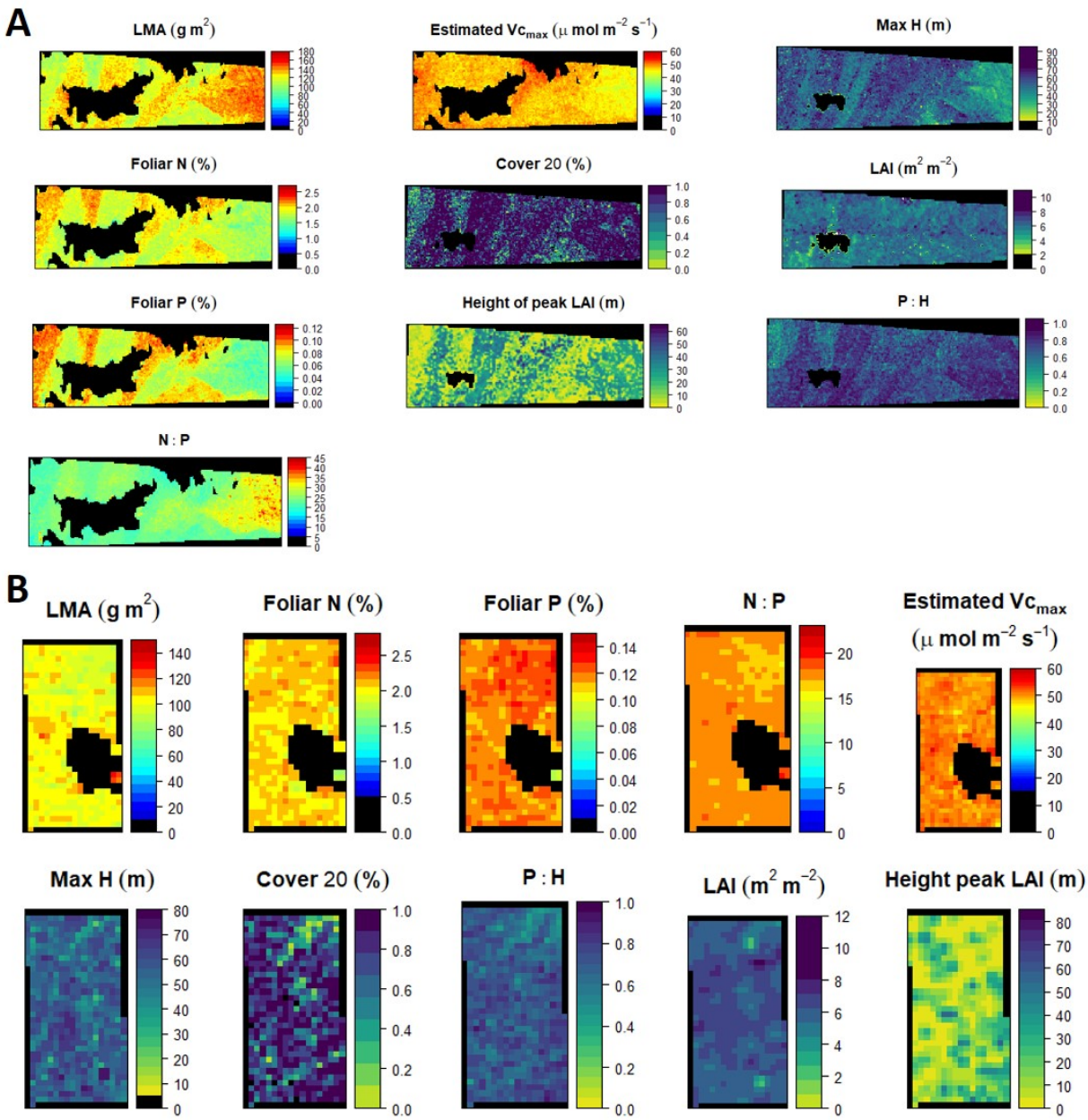

Figure S1. Ten community, plant, and leaf traits used in the forest functional mapping at Sepilok (a) and Danum (b). LiDAR and imaging spectroscopy data were collected in 2016 by the Global Airborne Observatory. LMA: leaf mass per area, N: nitrogen, P: phosphorus, N:P: nitrogen to phosphorus ratio, Max H: maximum height, Cover 20: canopy gap fraction at 20 m, LAI: leaf area index, P:H: ratio describing the vertical partitioning of foliage in the canopy. Black areas indicate No Data.

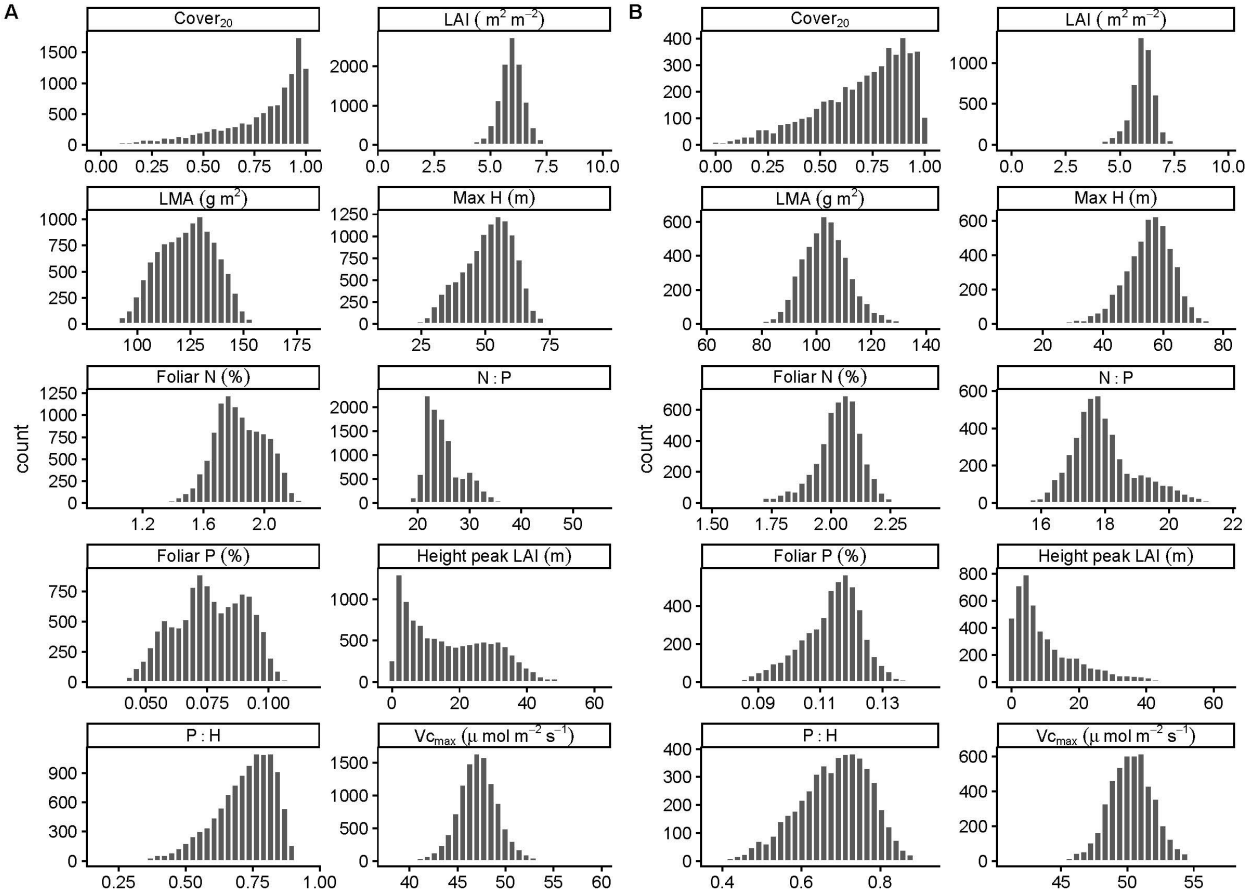

Figure S2. Histograms of untransformed distributions for all ten traits used in the PCA and *k*-means cluster analysis for Sepilok (a) and Danum (b).

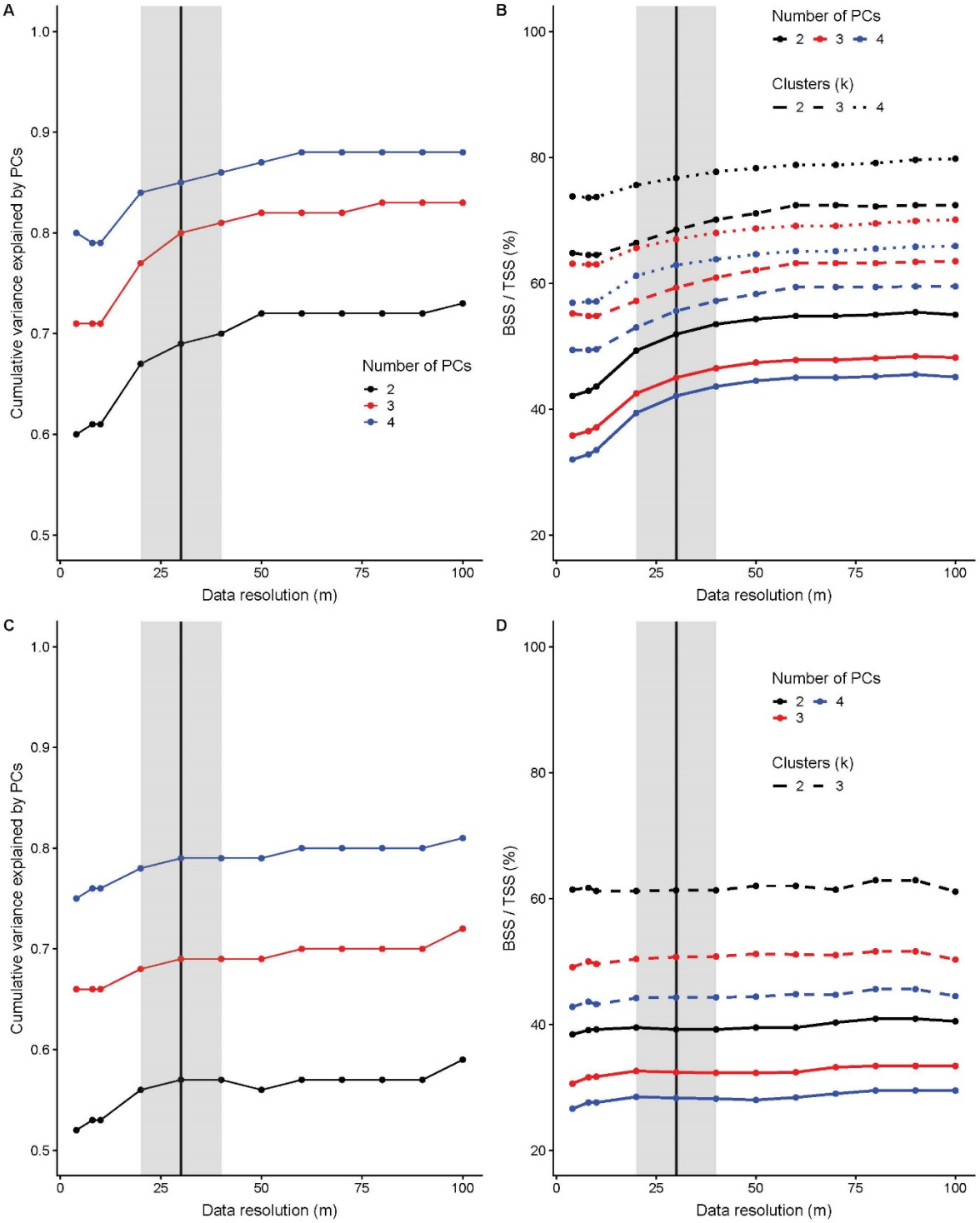

Figure S3. The influence of data spatial resolution on the degree of variance explained for 2-3 principal components (PCs) (a, c) and the *k*-means *BBS/TSS* (b, d) for both Sepilok (a-b) and Danum (c-d). The grey shaded area highlights saturation for both metrics around 20-40 m resolution. The 30 m analysis resolution is indicated by the black vertical line.

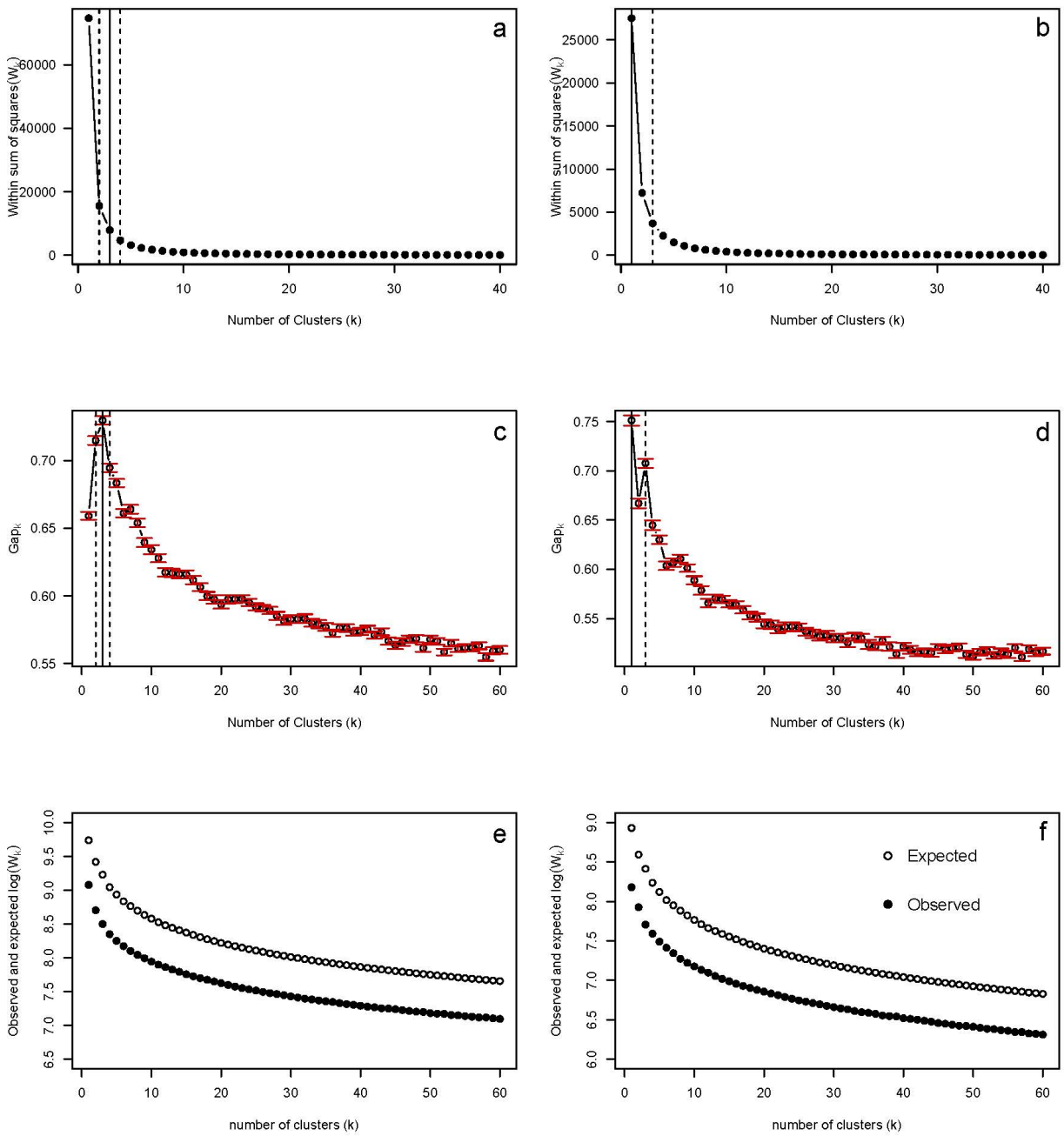

Figure S4. Comparison of the and within group sum of squares (W_k) (a-b) and gap statistic (c-d) used to determine the number of *k*-means clusters for Sepilok (a,c,e) and Danum (b,d,f). The solid vertical lines in panels b and e indicate the number of clusters selected using the first local and global maxima, while the dashed lines indicate additional *k* values explored. Panels e-f show the observed and expected log(W_k), where the optimal number of clusters is the value of *k* for which observed log(W_k) falls the farthest below expected log(W_k).

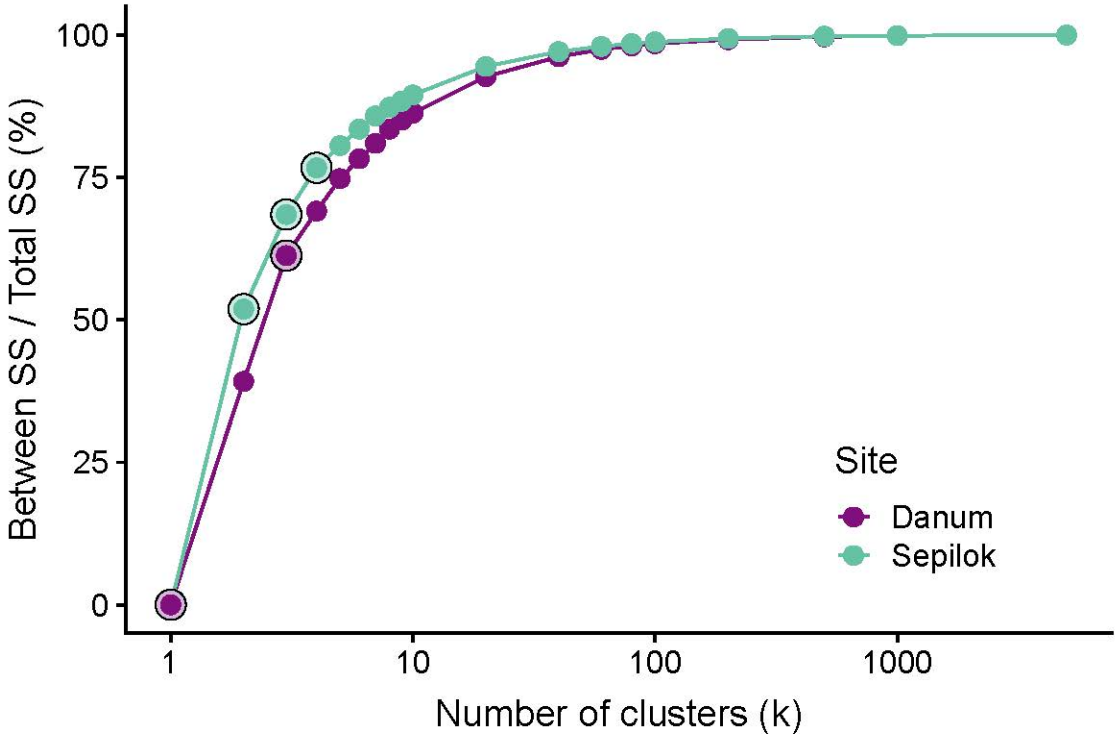

Figure S5. Between cluster sum of squares (SS) divided by the total SS for k values between 1 and 5000. The points circled in black indicate the number of clusters analyzed for each site (Danum = 1, 3; Sepilok = 2, 3, 4).

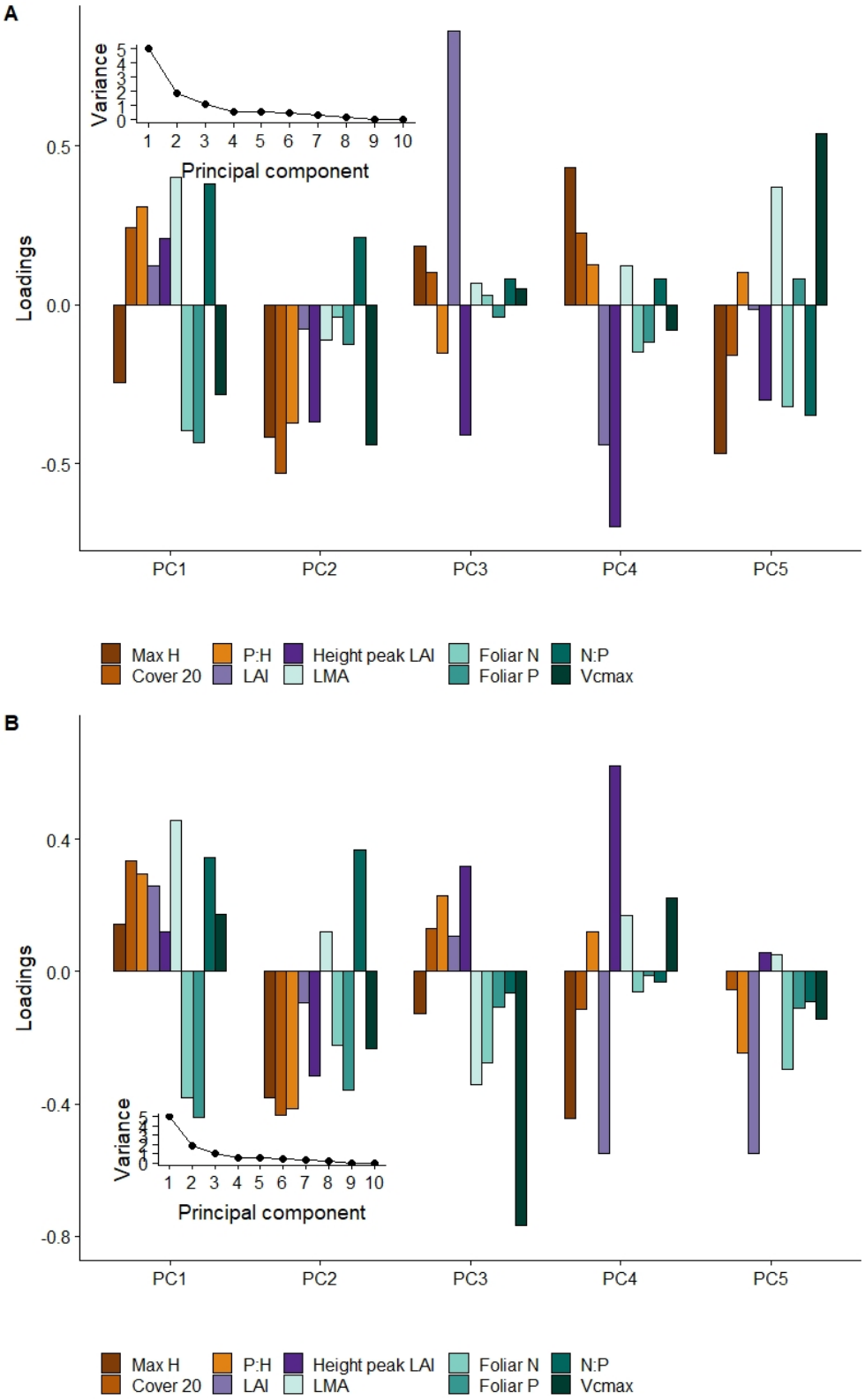

Figure S6. Principal components from PCA of the 10 foliar, plant, and community traits at Sepilok Forest Reserve (a) and Danum Valley (b).

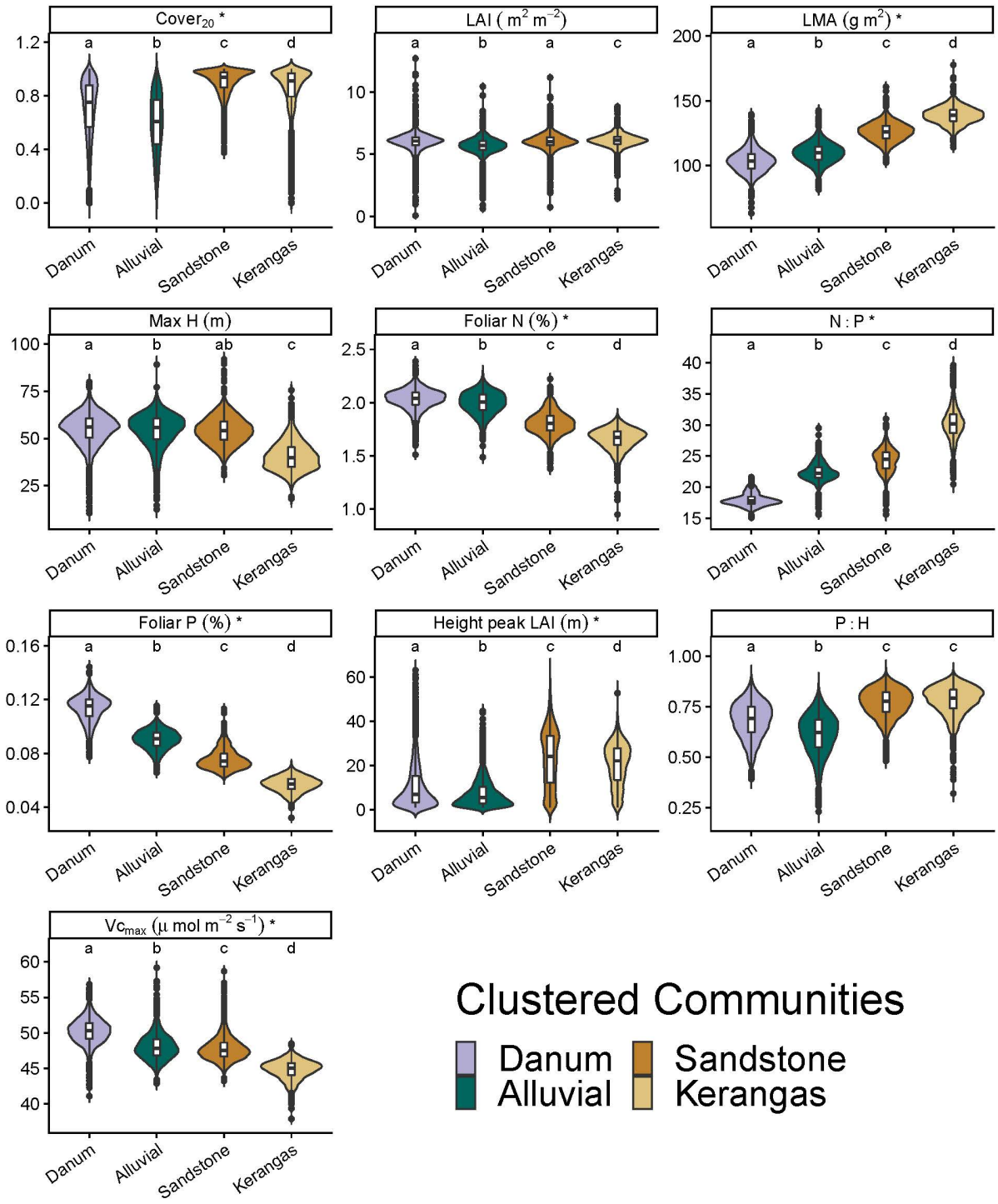

Figure S7. Trait distributions by cluster for Sepilok $k = 3$ and Danum $k = 1$. Forest communities are ordered based on their median LMA to illustrate differences in traits for communities that vary along the leaf economics spectrum. Identical letters represent clusters where there is no significant difference between forests based on one-way ANOVA tests ($p < 0.01$).

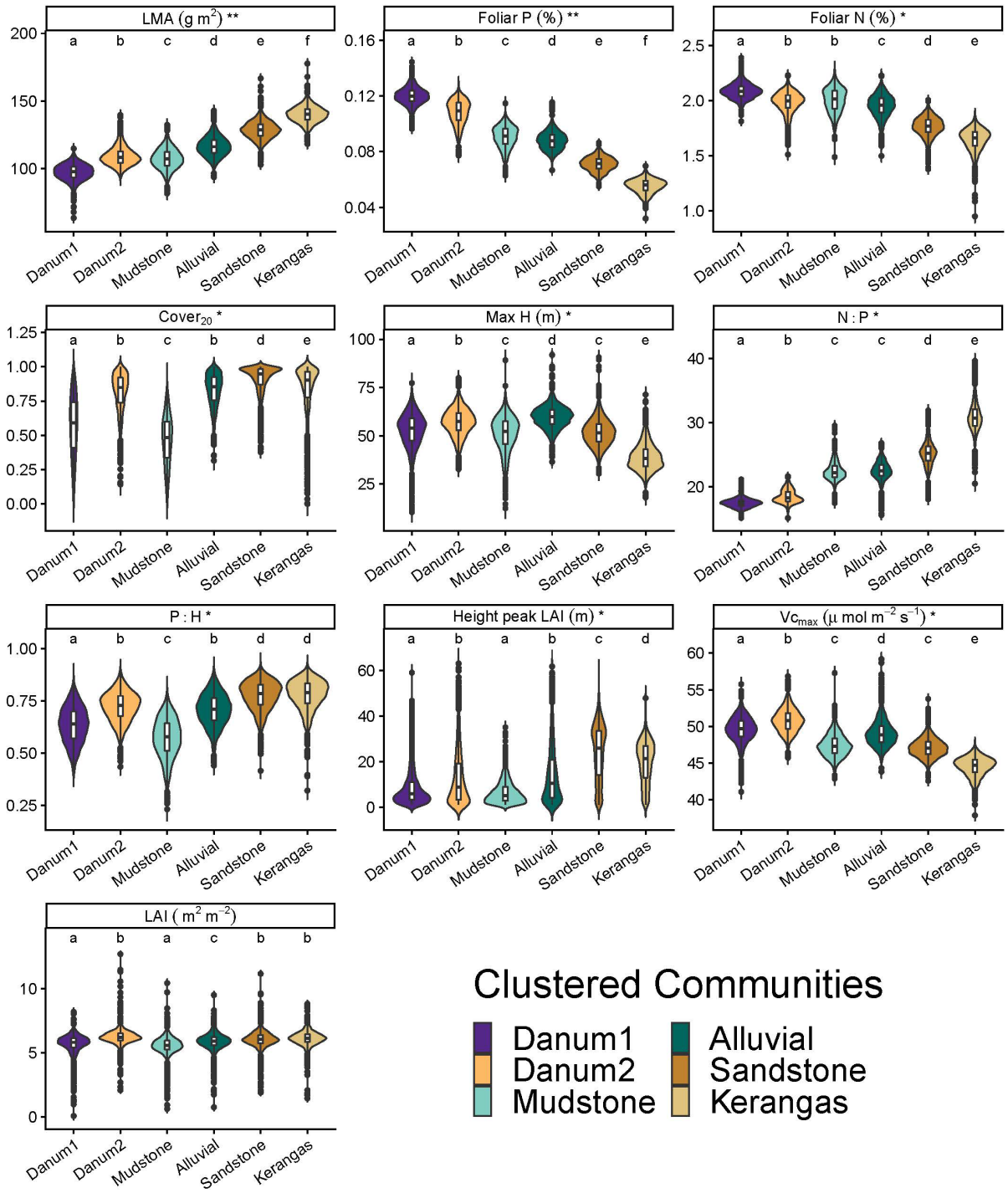

Figure S8. Trait distributions by cluster for Sepilok $k = 4$ and Danum $k = 2$. Forest communities are ordered based on their median LMA to illustrate differences in traits for communities that vary along the leaf economics spectrum. Identical letters represent clusters where there is no significant difference between forests based on one-way ANOVA tests ($p < 0.01$). ** traits that varied significantly between all six forest types. * traits that varied significantly between at least four forest types.

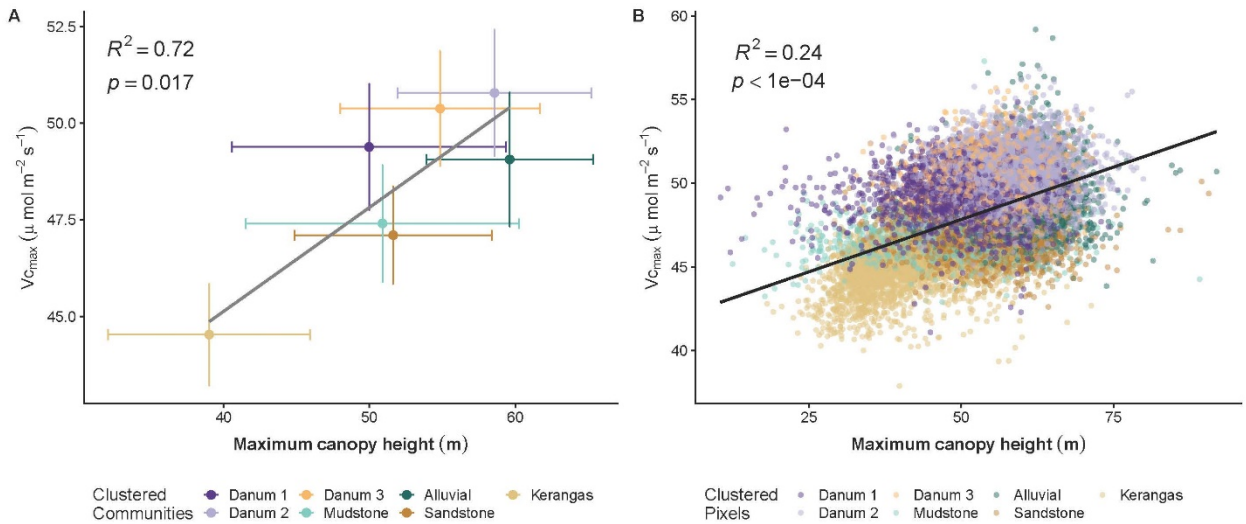

Figure S9. Relationship between V_{cmax} and maximum canopy height at the cluster (a) and pixel (b) level. Colors indicate communities identified for $k = 3$ clusters in Danum and $k = 4$ clusters in Sepilok.

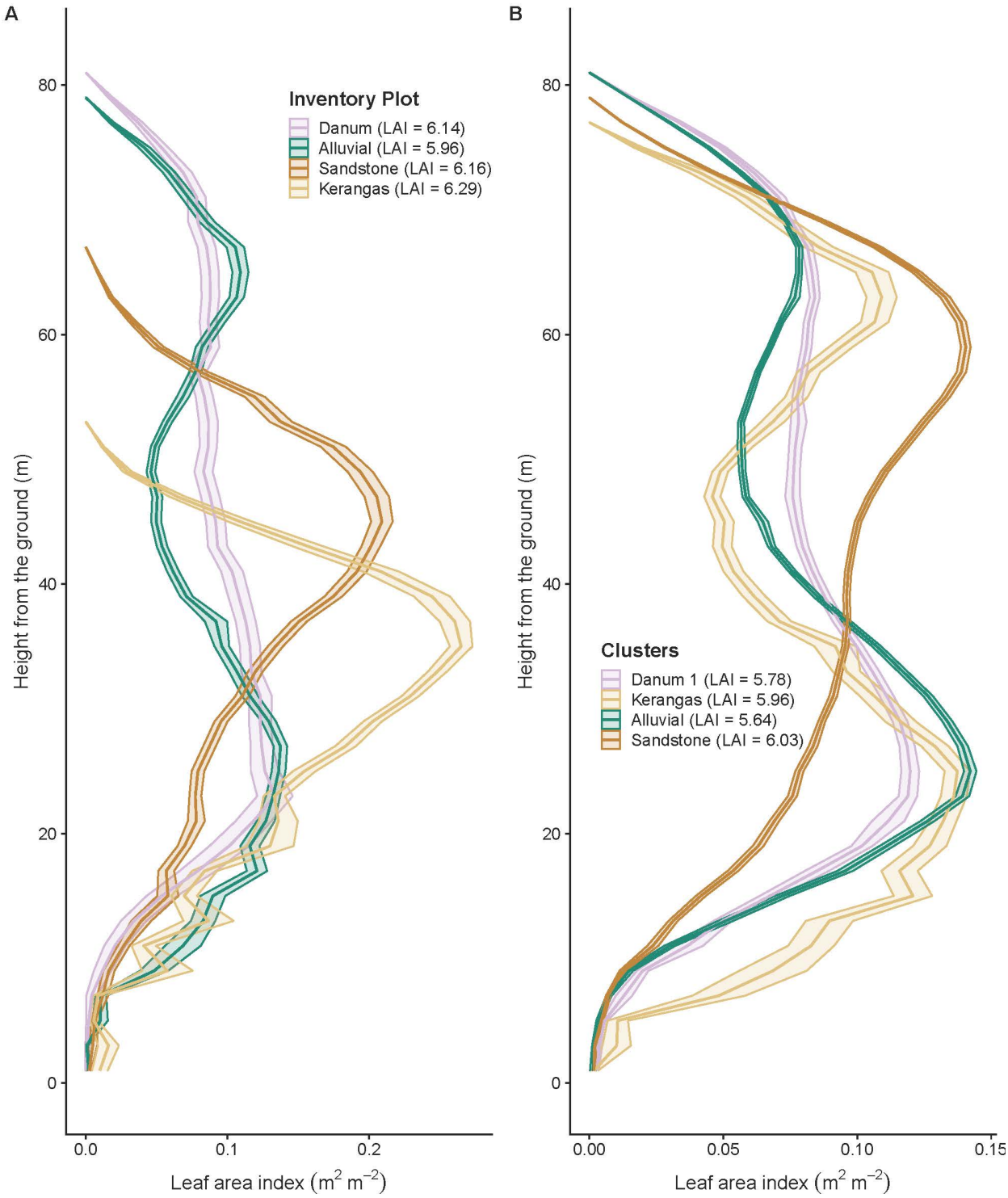

Figure S10. Vertical LAI profiles for all pixels within each inventory plot (a-b) and forest community identified based on $k = 1$ cluster at Danum and $k = 3$ clusters at Sepilok (c-d).

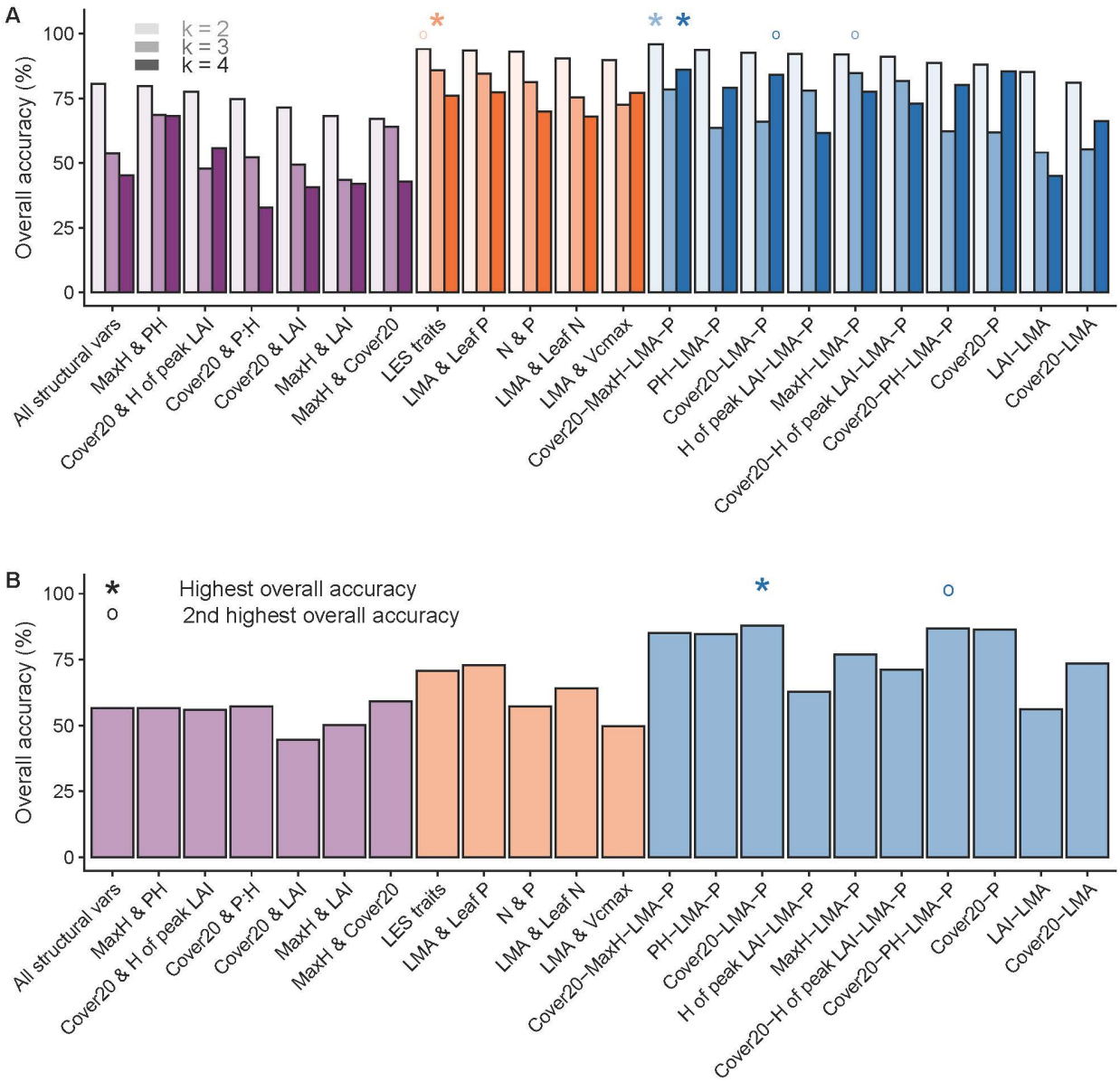

Figure S11. Change in overall accuracy for reduced *k*-means clustering models using structural variables (purple), leaf trait variables (orange), and combinations of structural and leaf trait variables (blue). All are compared to the full 10-variable *k*-means clustering analysis for Sepilok (A) and Danum (B). Asterisks indicate the reduced model with the highest overall accuracy for $k = 2$, 3, and 4 for Sepilok and $k = 3$ for Danum. Dots indicate reduced models with the second highest overall accuracy. Each bar illustrates $k = 2$, 3, 4 from left to right for Sepilok (A) and $k = 3$ for Danum (B). LES: leaf economic spectrum.

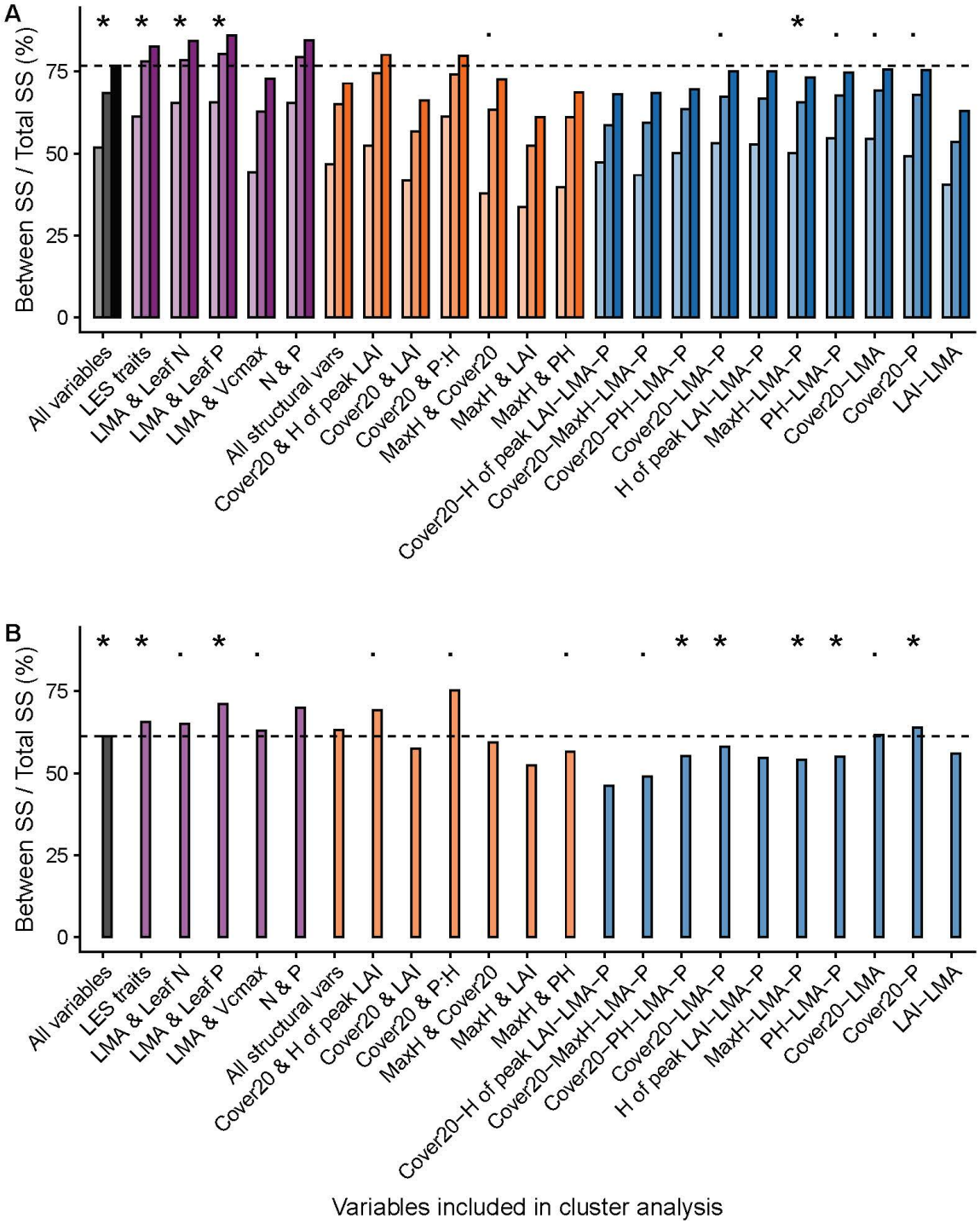

Figure S12. Change in between sum of squares (BSS) divided by total sum of squares (TSS) with variables included in the *k*-means clustering analysis for Sepilok (A) and Danum (B). The horizontal dotted line is the BSS/TSS value with all variables. Asterisks indicate output that captured the different forest types identified with all variables. Dots indicate when a similar pattern was captured but at least one forest type was not distinguished. Each bar illustrates $k = 2$, 3, 4 from left to right for Sepilok (A) and $k = 3$ for Danum (B). LES: leaf economics spectrum.