communications earth & environment

Article

A Nature Portfolio journal

https://doi.org/10.1038/s43247-025-02976-w

Characterizing the vertical structure of forests in the Brazilian Amazon

Denis Valle[1] ✉, Leo Haneda[1], Ismael Verrastro Brack[1], Jean Ometto[2], Ovidiu Csillik[3], Marcos Longo[4,5], Michael Keller[6,7] & Danilo Almeida[8,9]

Little is known about the structure of tropical forests despite its critical role in the provisioning of ecosystem services. Here we assess the vertical structure of forests in the Brazilian Amazon with a large-scale airborne LiDAR dataset. We show that fire has greater impact in the lowest forest strata, differently from selective logging and windthrow. We also find that secondary forests quickly recover or even exceed reference areas at the 1-10 m height stratum but that full recovery for the 20-30 m height stratum has not been achieved even after 35 years. Our modeling results suggest that proximity to roads, elevation, precipitation, soil pH, and proportion of sand in the soil are the most important predictors of forest structure. Finally, we identify 5 forest structural types (FSTs) and use them to visualize the spatial distribution of forest structure. This study provides important information for forest monitoring, management, and conservation.

Forests play an important role worldwide, holding a large fraction of the biodiversity of terrestrial ecosystems and providing numerous ecosystem services[1–3]. However, forests are experiencing substantial losses due to a variety of human activities[4–6] and remaining forest fragments are often degraded due to edge effects, selective logging, fire, and cattle grazing within fragments[7–9]. Unfortunately, it is challenging to detect signs of forest degradation, such as changes in forest structure, with traditional remote sensing techniques based solely on canopy surface reflectance[10]. The monitoring of forest structure is also important because it has a strong influence on several ecosystem services such as timber production, carbon storage, and water quality[11,12]. For example, forests have a fundamental role in regulating climate at local, regional, and continental scale[13–15], with forest structure strongly mediating these effects[16–18]. Furthermore, forest structure can play an important role in preventing soil erosion by holding the soil, promoting infiltration, and slowing down water flow[19,20]. Finally, forest structure complexity is strongly associated with wildlife occurrence and diversity (e.g., bats and birds[21–23]).

Active laser sensors, particularly light detection and ranging (LiDAR), have been shown to substantially improve upon passive optical remote sensing systems in characterizing forest structure[24,25]. LiDAR sensors operate by emitting high-density laser pulses towards the ground and measuring the height at which these pulses are reflected[24,26], resulting in a detailed three-dimensional point cloud containing information on the spatial and vertical arrangement of vegetation[27]. Despite LiDAR's ability to improve the understanding of the vertical structure of forests, most of the research conducted with LiDAR has focused solely on top-of-the-canopy height[28]. Importantly, although categorizing forests into forest types/vegetation classes based on their structure would be helpful for the development of more tailored conservation, management, and restoration plans and actions, existing vegetation classes might not be as informative about forest structure as they could be. For example, to date, the most used vegetation classification in the Brazilian Amazon is the one provided by the Brazilian Institute of Geography and Statistics (IBGE), developed based on airborne radar data and field plots measured in the 1970s and 1980's[29–31]. Because multiple criteria were used to create the IBGE classes (e.g., latitude, altitude, and climate), some of these classes might be floristically distinct but not necessarily structurally distinct.

In this study, we characterize the 3D (spatial and vertical) forest structure in the Brazilian Amazon based on an unprecedented large-scale LiDAR dataset covering approximately 188,000 ha. This characterization of forest structure is based on the modeling of return probability (RP), estimated as the number of returns within a given height class divided by the number of incoming returns (i.e., returns with height below the upper limit of the target height class)[32,33]. We chose this metric because it is

[1]School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL, USA. [2]Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Sao Paulo, Brazil. [3]Environment and Sustainability Studies Program, Wake Forest University, Winston Salem, NC, USA. [4]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Divisão de Modelagem Numérica do Sistema Terrestre, Instituto Nacional de Pesquisas Espaciais (INPE), Cachoeira Paulista, São Paulo, Brazil. [6]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [7]United Stated Department of Agriculture Forest Service, International Institute of Tropical Forestry, Rio Piedras, Puerto Rico. [8]BRCarbon, Piracicaba, Sao Paulo, Brazil. [9]Bioflore, Belo Horizonte, Minas Gerais, Brazil. ✉e-mail: drvalle@ufl.edu

straightforward to calculate and interpret, and it can be used as a proxy for the amount of vegetation in a given height stratum. Importantly, this approach allows us to account for the occlusion effect of vegetation above this stratum[32–36]. On the other hand, more traditional LiDAR metrics focused on characterizing the vertical structure of forests (e.g., kurtosis of the distribution of all return heights, foliage height diversity, plant area index, and plant area volume density[37–39]) often do not share these characteristics.

We start this article by showing how our characterization of the vertical structure of the forest using RP can yield more insights than that obtained by only using top of canopy height. We illustrate this by comparing results for secondary forests of different ages and for forests impacted by different types of disturbances. Using the full LiDAR data, we assess the predictors of the vertical structure of the forest. Finally, we identify forest classes with similar vertical biomass profiles (i.e., forest structural types (FST)), mapping their spatial distribution across the region.

Results

Comparison of top-of-canopy height (TCH) and vertical distribution of return probability (RP)

We rely on RP as a measure of biomass in each height strata. In this section, we compare top-of-canopy height (TCH) and RP for areas with secondary forests of different ages in 2019 with areas that were never clear cut between 1984 and 2019 (the time period encompassed by the Mapbiomas product on secondary forest age). We refer to these areas that were never clear cut as reference areas. To this end, we calculate the difference in TCH between secondary forests and the median TCH for reference areas within the same transect, denoting this difference as ΔTCH. Similarly, we calculate the difference in RP between secondary forests and the median RP for reference areas within the same transect, denoting this difference as ΔRP. Based on these calculations, we find that, as expected, ΔTCH and ΔRP tend to increase as secondary forests grow older (Fig. 1). However, even after 35 years (the maximum age of secondary forests in our dataset), we find that the ΔRP and ΔTCH of the 20–30 m height class have not fully recovered when compared to reference areas in the same transect (Fig. 1c, d). On the other hand, our results reveal that ΔRP in the 1–10 m height class recovers quickly, taking 5–10 years to reach levels comparable to those found in surrounding reference areas (Fig. 1a). Interestingly, our results suggest that secondary forests that are more than 15 years old often have greater RP in the 1–10 m height class when compared to surrounding areas that were never clear cut (Fig. 1a).

We find that ΔRP provides unique information regarding the lower forest strata. More specifically, the correlation between ΔTCH and ΔRP for the 1–10 m height class was only 0.18, indicating that the 1–10 m RP contains information that is distinct from that provided by top-of-canopy measures. On the other hand, ΔTCH has an overall pattern that is very similar to that of ΔRP in the 10–20 m and 20–30 m height classes (Fig. 1b–d). Indeed, we find a strong positive correlation (Pearson correlation ≈ 0.7) between ΔTCH and ΔRP for these higher strata, suggesting that these metrics provide similar insights regarding biomass in the top strata of the forest.

Our second example relies on pixels impacted by a single forest disturbance (i.e., selective logging, fire, or windthrow) and that were flown twice (once before and once after the disturbance). We calculated RP for each height class and each flight, enabling the assessment of how much RP changed (i.e., ΔRP) and how much TCH changed (ΔTCH) for the impacted areas. The ΔTCH results reveal that forests impacted by logging tended to have a slightly greater decrease in forest height when compared to forests impacted by fire and windthrows (Fig. 2b). On the other hand, the Δ*RP* results reveal more detailed information. For example, fire had the greatest impact in the 1-10 class, with diminishing impacts for higher height classes (Fig. 2a), an expected pattern given that forest fires are typically disturbances that arise from the forest floor. On the other hand, the vertical distribution of Δ*RP* reveals that pixels affected by logging or windthrows tended to have more even impacts in the different height classes (Fig. 2a).

Recall that secondary forest age and forest disturbance data were independent of the LiDAR information. As a result, similar to the findings in our previous work[32], the comparative results described in this section provide strong support for the ability of RP to capture vertical patterns in forest biomass that are not evident from top-of-canopy information.

Predictors of return probability (RP)

We used a generalized additive model (GAM) to explore predictors of RP for the entire LiDAR dataset. Our GAMs revealed that all spline terms were statistically significant, regardless of height class. This is expected given the large number of observations (~750,000) available for each model. We find that proximity to roads tended to decrease RP across all height classes

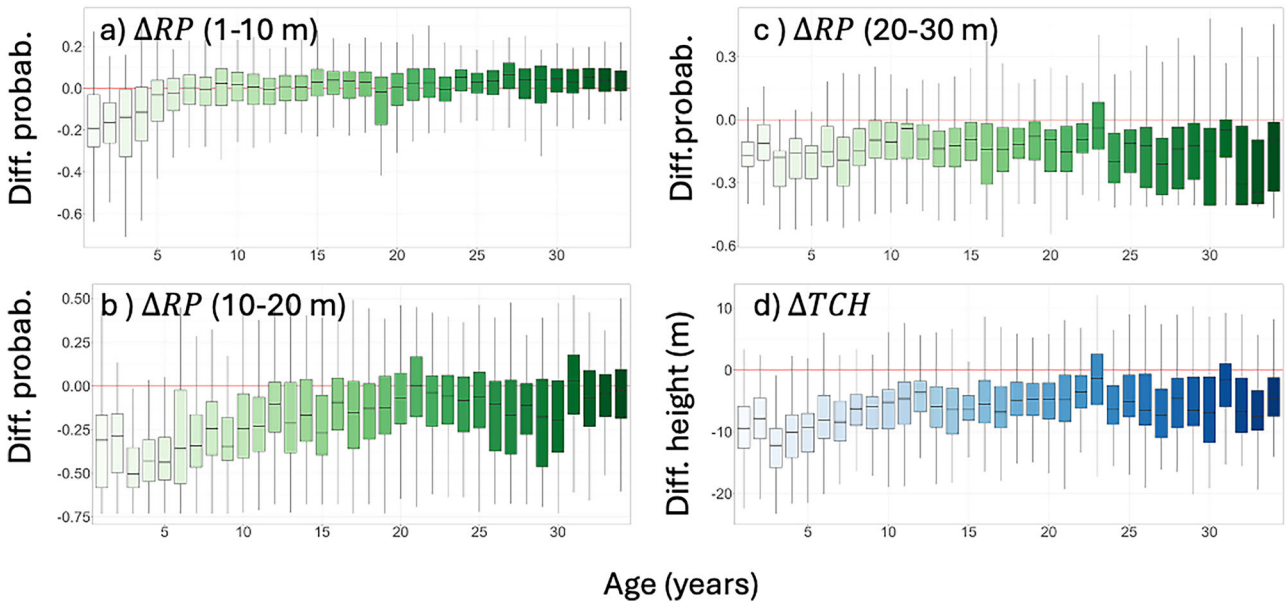

Fig. 1 | Comparison of the difference in return probability (ΔRP) and difference in top of canopy height (ΔTCH) as a function of secondary forest age (*x*-axis). Differences were calculated as the observed metric minus the median value for the corresponding reference areas. Values close to zero (horizontal red line) indicate that the target metric is close to the value observed for reference areas. ΔRP results are shown separately for different height classes: 1–10 m (**a**), 10–20 m (**b**), and 20–30 m (**c**). Results for 30–40 m and 40–50 m are not shown because the difference in return probabilities was very small and uninformative. **d** The corresponding results for the difference in top of canopy height (Δ*TCH*).

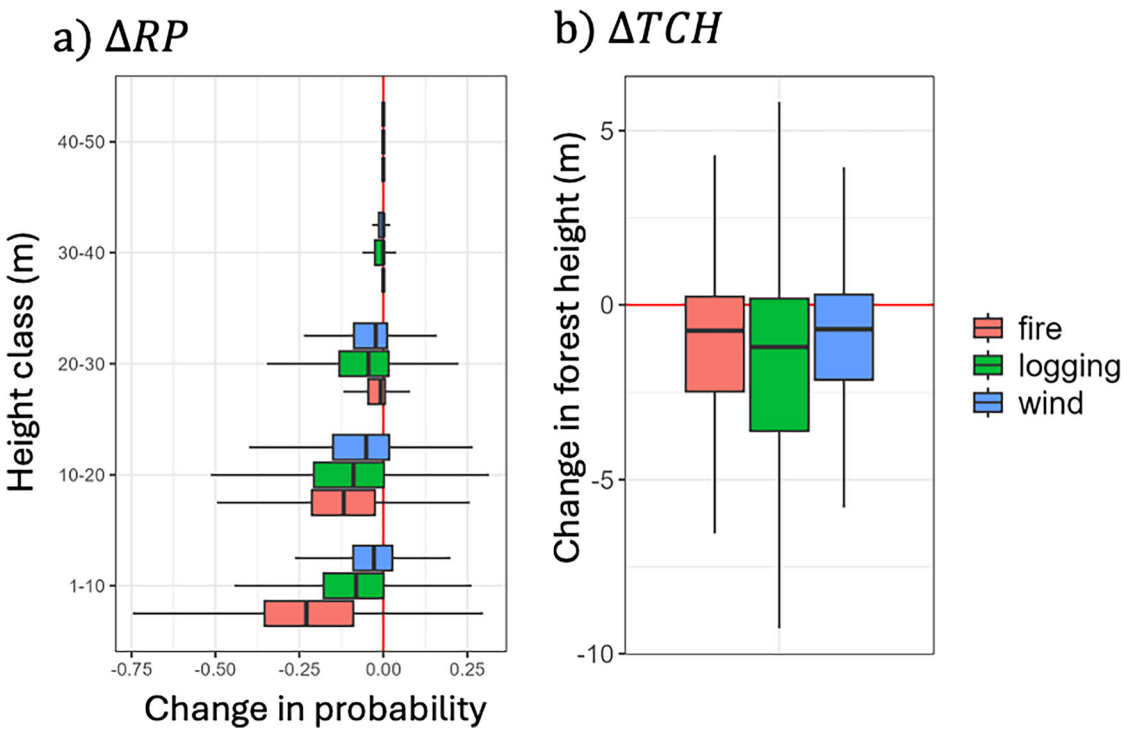

Fig. 2 | Impact of fire, logging, and windthrow on the vertical distribution of return probability (RP) and top-of-canopy height (TCH). Changes in RP (**a**) and TCH (**b**) were calculated as $\Delta RP = RP_{ith,after} - RP_{ith,before}$ and $\Delta\ TCH = TCH_{it,after} - TCH_{it,before}$, respectively, for each 50 × 50 m pixel.

(Fig. 3), a pattern that we attribute to the direct effects (e.g., logging and fire) and indirect effects (e.g., edge effects) of human activities. The effect of proximity to roads tended to be stronger and have longer spatial range (i.e., greater distances) for the 1–10 m and 10–20 m height strata, suggesting that anthropogenic disturbances that tend to predominantly come from below (e.g., fire and understory vegetation clearing by humans or cattle) have greater spatial reach. Critically, these lower forest strata impacts are more likely to go undetected by optical remote sensing, suggesting that the extent of anthropogenic disturbances is likely to be much larger than what has currently been estimated using optical sensors.

Aside from proximity to roads, the variables that in general had the greatest effect (determined as the variables resulting in the greatest predicted RP range) across most of the height classes were elevation, soil pH, proportion of sand, and precipitation (Fig. 3). Interestingly, we observe that elevation and soil pH tend to have relatively consistent effects across the different height classes. For example, low elevation areas (e.g., areas close to the main rivers) tend to have lower biomass across all height strata when compared to high elevation areas. Similarly, areas with more acidic soils tended to have greater biomass across all height strata. On the other hand, the relationship between RP and proportion of sand and precipitation depended on the height class being examined. More specifically, areas with low proportion of sand tended to have higher RP for taller classes (i.e., 30–40 m and 40–50 m height classes) but somewhat lower probabilities for shorter classes (i.e., 10–20 m and 20–30 m height classes). This pattern suggests a shift in forest structure, where areas with low proportion of sand generally have taller forests with less understory whereas areas with a high proportion of sand have shorter forests with greater density of understory. Similarly, our results for precipitation also suggest a shift in forest structure, where areas with lower average precipitation (i.e., ~150 mm per month) generally have shorter forests with greater density of understory whereas areas with higher precipitation (i.e., ~240 mm per month) have taller forests with less understory. To help the interpretation of the results in Fig. 3, we have qualitatively categorized each relationship (positive, negative, constant, or unimodal) in Supplementary Note 4.

The amount of explained deviance ranged between approximately 20% for the 1–10 and 10–20 m height strata to approximately 35–38% for the taller strata. These results suggest that the adopted explanatory variables, together with other omitted spatial variables captured by the spatial spline in GAM, are more strongly related to biomass in the higher strata when compared to the lower strata. The substantial unexplained variance, particularly at the lower strata, might be due to local anthropogenic activities (e.g., logging and fire) that are not fully captured by our distance to road variable and/or by local high soil fertility patches (e.g., due to the presence of terra preta potentially created by ancient human populations[40]).

Forest structural types (FSTs) and their spatial distribution

We used a modified LidarLDA model to identify forests that have similar vertical structure (Forest structural types; FSTs) and to estimate the proportion of these forest types in each LiDAR transect. Importantly, although this model assumed a maximum of five FSTs, it allowed for the possibility of there being fewer FSTs. Our results suggest that all five FSTs are important as they were common in different regions of the landscape. These FSTs have relatively distinct forest structure as shown with their height profiles in Fig. 4a, with FSTs 1 through 5 representing a general gradient from tall to short forest types. There are also important differences between FSTs regarding RP in the intermediate height classes. For example, FST 3 represents forests that are stockier (i.e., slightly shorter but with much greater biomass between 10 and 30 m) when compared to the slender FST 2 (i.e., slightly taller but with substantially less biomass between 10 and 30 m). Finally, FSTs 4 and 5 are much shorter forest types when compared to the other FSTs, with FST 5 being the shortest FST.

Using a random forest model to create spatial predictions of the proportion of each FST, we find that the five FSTs have very distinct spatial distributions (Fig. 4c–g). For example, the tallest forest type (i.e., FST 1) is very common in western Amapá and northwestern Pará. On the other hand, the slightly shorter FST 2 is more common in Acre whereas the stockier FST 3 is much more common in Amazonas state. The shorter FSTs (i.e., FSTs 4 and 5) are more common along the main rivers in the region and in the eastern/southeastern edges of the study area, representing the

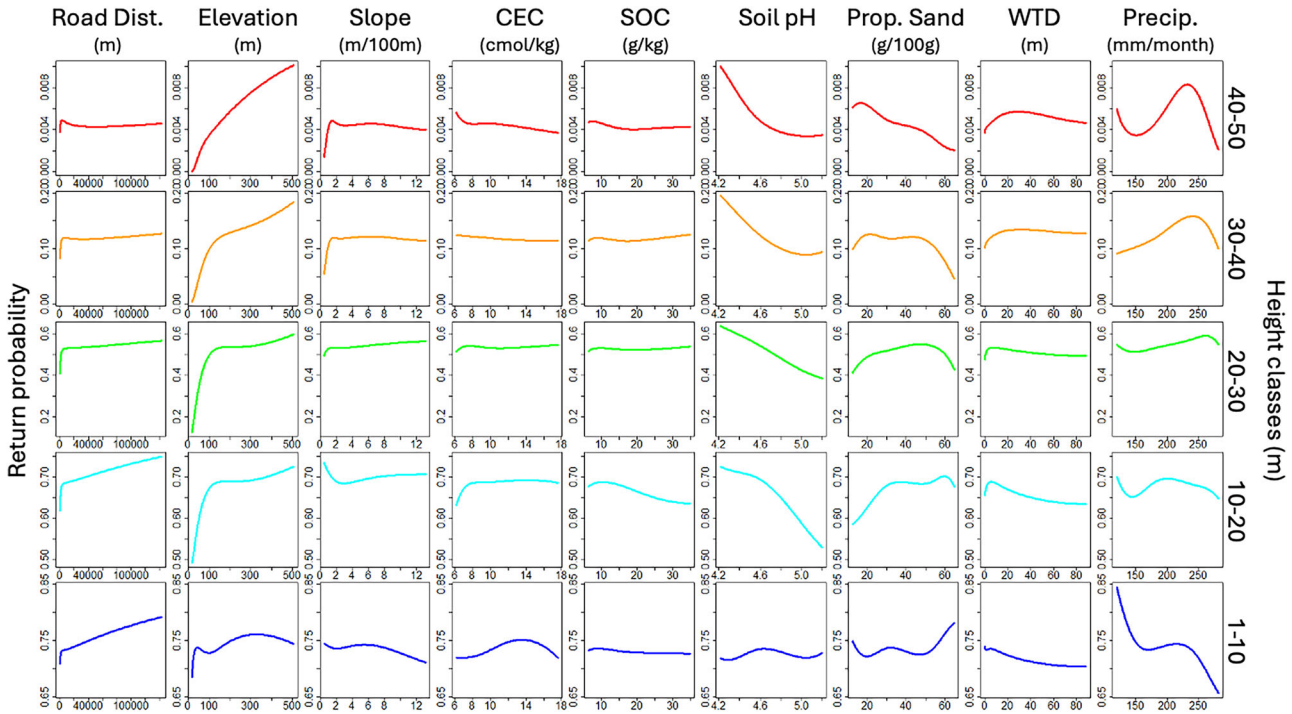

Fig. 3 | Relationship between return probability (RP) for different height classes and potential environmental predictors. Top to bottom panels show results for different height classes, whereas the effect of environmental predictors are shown from left to right panels. The range of the *x*-axis of each panel was restricted to the 1st and 99th percentiles of the corresponding covariate because this is the range most informed by the data. We show the corresponding pointwise 95% confidence intervals but because these intervals are very narrow, they cannot be visually distinguished from the mean relationships. CEC, SOC, and WTD stand for cation exchange capacity, soil organic carbon, and water table depth, respectively.

transition zone between the Amazon and the Brazilian savanna (i.e., Cerrado) biomes.

To summarize this information, we also created a single map containing the dominant FST for each pixel (Fig. 4b). Interestingly, there seems to be a strong correspondence of the spatial distribution of some of the FSTs with geological formations. For example, FST 1 seems to correspond well with the Guiana shield formation, FST 2 with the Brazilian shield, and FST 3 with the Amazon basin lowlands and Pebas formation. This map also allows the comparison of the spatial distribution of FSTs and the IBGE vegetation classes, revealing that, despite some similarities, there is no one-to-one relationship between IBGE vegetation classes and our FSTs. These results suggest that the identified FSTs provide distinct information than that provided by the IBGE vegetation classes (Supplementary Note 3). We also compare these classification schemes regarding their ability to capture large-scale patterns of TCH and biomass based on independent global maps. We find that the IBGE vegetation classes had an R^2 of 52% and 23% for biomass and TCH, respectively. On the other hand, the estimated FSTs had an R^2 of 48% and 45% for biomass and TCH, respectively. In short, the estimated FSTs perform substantially better than the IBGE vegetation classes in explaining forest height but has similar performance regarding forest biomass.

Discussion

We use an unprecedented airborne LiDAR dataset from the Brazilian Amazon to assess the 3D structure of Brazilian Amazon forests and identify the environmental variables that are strongly associated with it. We focus on RP as a measure of biomass in a particular height stratum because this is a straightforward metric that avoids the occlusion effect of vegetation above the height stratum of interest. We show how the vertical distribution of RP can provide more information about forests than TCH by characterizing the structural recovery of secondary forests through time and the distinct vertical impacts of fire, selective logging, and windthrows. Using the full LiDAR dataset, we model RP as a function of a series of covariates, finding that proximity to roads, elevation, precipitation, soil pH, and proportion of sand were the variables that had the greatest effects on RP across most of the height classes. Finally, using a modified LidarLDA model, we identified 5 FSTs, enabling the visualization of the spatial distribution of forest structure across the Brazilian Amazon.

By analyzing the vertical profile of secondary forests in the region, we were able to characterize the regrowth dynamics of these forests over different height strata. More specifically, our results indicate that, even after 35 years, most secondary forest areas have not fully recovered their TCH, an outcome that is corroborated by the finding that RP in the 20–30 m height class also did not fully recover (Fig. 1). Similar to these results[41], find that canopy height of secondary forests plateaus at 60% of intact forest after 10–15 years and that aboveground biomass is only 40% when compared to intact forest after 20 years. However, our RP results for the 1–10 m height class reveal that the biomass of secondary forests in this height class has already fully recovered after 5–10 years and that secondary forests >15 years old often have more biomass in this height class than the surrounding forest that was never clear cut (Fig. 1). The patterns regarding vegetation regrowth in different height strata can serve as baseline information to determine how successful restoration projects in the region have been in increasing the recovery rate of these forests. More specifically, region-specific baselines could be created given that climate and other environmental factors strongly influence forest recovery rate[42]. These region-specific baselines are likely to be useful for the large-scale assessment of restoration projects, complementing the plot-based ecological indicators that have recently been put forward for the Brazilian Amazon[43].

There are multiple types of forest disturbance in the Amazon region[44]. For example, despite the fact that humid tropical forests are not naturally fire prone, fire has become increasingly more prevalent in the Amazon region and it has been estimated that fire has impacted an area of approximately 69,000 km^2 in 2019[45]. Selective logging is another important disturbance as it has been estimated to affect 12,000–20,000 km^2 per year between 1999 and 2002[10]. Finally, convective storms can create windthrows, resulting in

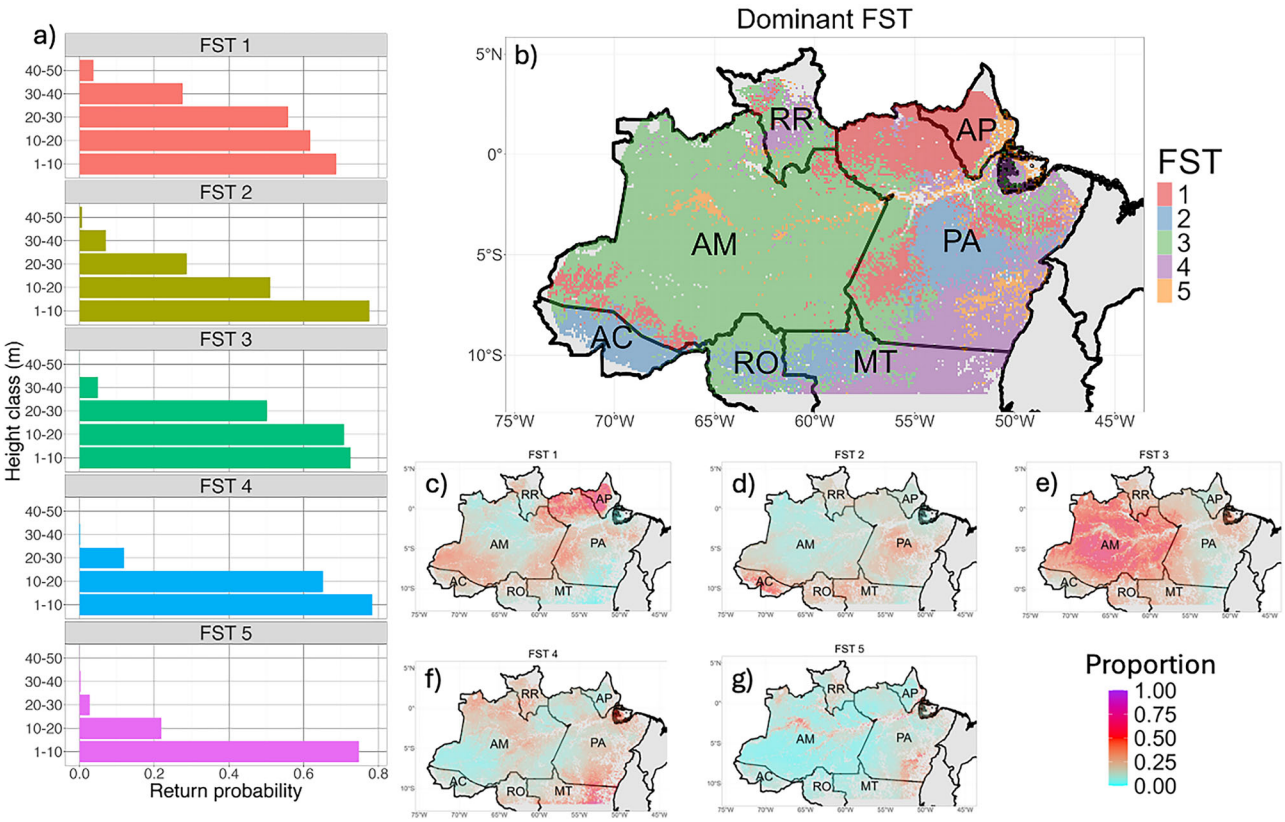

Fig. 4 | Height profiles of the forest structural types (FSTs) identified by the modified LidarLDA model, and the predicted spatial distribution of the proportion of each FST according to our random forest model. The height profile of each FST is displayed in (**a**), the spatial distribution of the predicted proportion of each FST is shown in (**c**–**g**) (proportion legend), and a summary of the information from **c** to **g** created by displaying the dominant FST in each pixel is provided in (**b**) (FST legend). Extrapolations were avoided by restricting predictions to pixels that have predictor values within the range of values in the training data and that were within the convex hull of the LiDAR data. We only plot predictions for areas classified as forests in 2019. Abbreviations refer to the following Brazilian states: AM Amazonas, AC Acre, RR Roraima, RO Rondonia, MT Mato Grosso, PA Pará, and AP Amapá. The size of each pixel in these maps is 0.1 × 0.1°.

substantial tree mortality[46] and recent modeling results suggest that windthrows are likely to increase in frequency due to an increase in the area with favorable conditions for extreme storms[47]. Three-dimensional forest structure information is important to shed light on the vertical distribution of the impact of these forest disturbances. For example, using field data from a fire experiment[48], showed that fires resulted in higher mortality of smaller trees, particularly at forest edges. These findings were corroborated by our results showing greater fire impact on the lower height strata of the forest (Fig. 3). More generally, our GAM analysis reveals the strong negative influence of proximity to roads on RP, with a particularly pronounced effect and a greater distance reach for the lower height strata (i.e., 1–10 m and 10–20 m; Fig. 3). We attribute this distance effect to anthropogenic disturbances such as understory fire, logging, and edge effects. However, we note that our road network dataset is likely to represent only a fraction of the existing roads at the time of LiDAR data collection. Unfortunately, there is no up-to-date, complete dataset of the road network in the region which also includes so-called unofficial roads[49] and, as a consequence, our distance-to-road results likely overestimate the reach of anthropogenic disturbances. Nevertheless, because we found a greater effect of proximity to road for the lower height strata, it is possible that there might be substantial underestimation of forest disturbance area if one relies solely on top-of-canopy information (e.g., information that arise from passive optical remote sensing satellite imagery[10,50]).

This differential impact of disturbances across height strata is also likely to have strong implications for determining the effect of these disturbances on forest carbon storage. For example[51], determined carbon loss associated with a range of disturbances by comparing pre- and post-disturbance TCH. However, given the differential impact of fire, logging, and windthrows in the different height strata, it is possible that their quantification of carbon losses could be improved. For example, the impact of fire is particularly acute for the 1–10 and 10–20 m height strata, suggesting that carbon storage losses associated with fire are much more pronounced than what would be predicted solely based on decreases of TCH. More broadly, it is possible that large discrepancies between field biomass and biomass map products derived from TCH are associated with areas that have substantially different vertical distribution of biomass, either due to natural forest characteristics or anthropogenic disturbances. If this hypothesis is correct, this would suggest that prediction models used to create biomass/carbon maps based on LiDAR data would benefit from including additional metrics that better capture the vertical distribution of biomass information (e.g., RP for different height strata).

Several studies have assessed large-scale spatial patterns of forest characteristics in tropical forests, such as species richness and composition[52,53], biomass and carbon density[54–56], and TCH[57,58]. However, to our knowledge, this is the first study to explore in detail the vertical structure of these forests in the region using airborne LiDAR data. Importantly, despite the fact that the modified LidarLDA model adopted here did not explicitly include spatial information, the maps depicting the predicted distribution of the identified FSTs reveal striking spatial patterns, some of which are corroborated by the literature. For example, a widely acknowledged pattern is that flooded forests in the Amazon tend to be shorter and have lower biomass when compared to upland (Terra Firme) forests[59,60], agreeing with our finding that the shortest FST is very common in areas close to the major rivers.

We also compare the identified FSTs with the IBGE vegetation classes. The IBGE vegetation classes have been the most used classification in the

Brazilian Amazon, being adopted for a wide range of applications. Furthermore, these IBGE vegetation classes are a key component of prominent worldwide vegetation classification schemes[61,62]. Nevertheless, there have been relatively few comparisons of these vegetation classes in the literature, particularly for the Brazilian Amazon. Our comparison of the identified FSTs with the IBGE vegetation classes reveals that, despite some similarities, there is no one-to-one relationship between these classes. Critically, using independent data, we demonstrate that the identified FSTs better represent TCH when compared to the IBGE vegetation classes whereas there was little difference regarding biomass. Importantly, there might be a decoupling between forest structure and biomass where forests with relatively distinct structure might nevertheless have similar overall biomass. Indeed, we find that field plots in areas dominated by FSTs 1 and 3 tend to have similar biomass values, whereas FST 2 areas tend to have lower biomass, despite FST 2 generally having taller trees than FST 3. Future studies using forest structure metrics measured on the field will be critical to assess if the identified FSTs indeed better describe forest structural characteristics when compared to the IBGE vegetation classes.

Amazon forest types are broadly determined by soil and hydrological conditions (e.g., presence or absence of seasonal or permanent flooding)[52,53,63]. Our GAM analysis corroborates these results by identifying the strong role of elevation (a proxy for hydrological condition), as well as soil pH and proportion of sand, in determining RP. These results agree with those from other local studies in the region that have found that biomass tends to be negatively associated with proportion of sand ([64,65] but see ref. 66) and a regional analysis that found that maximum tree height is positively associated with soil clay content and elevation[67]. Our GAM analysis also identified the strong influence of precipitation, with areas with low precipitation having a much higher RP in the 1–10 m class but lower RP for the 30–40 and 40–50 m classes than areas with high precipitation. Indeed, experimental[68] and observational studies using remote sensing[69] have reported on decreases of Amazonian forest biomass and vegetation indices due to decreased precipitation. However, similar to the quadratic biomass-rainfall and maximum tree height-rainfall relationships in ref. 70 and ref. 67, respectively, we find that the relationship between RP for the 30–40 and 30–40 m height class peaks around 230 mm per month, declining thereafter. In other words, areas with even higher monthly precipitation apparently do not support higher RP, an intriguing finding that requires more exploration.

Most studies examining large-scale patterns in Amazonia have relied on relatively large but sparse field plot data[52,71], limiting their ability to characterize forests in this vast region. We believe that the LiDAR data from the EBA project data can complement these existing datasets by providing large-scale standardized information on forest structure. Nevertheless, it is important to note that this project did not control for leaf phenology and, as a consequence, some of the unexplained geographical and before-and-after differences in our results might be attributed to some trees having fewer leaves during the time of the year that their data were collected. We do not expect this to be a major problem for most of the study region, but we acknowledge that the proportion of deciduous trees might be non-negligible for some areas (e.g., the southeast area of the Brazilian Amazon). Finally, another important limitation is that, while we explicitly avoided data from secondary forests when identifying forest types, our data might still contain degraded forests, potentially impacting our ability to accurately characterize some of the forest types (e.g., forest types 4 and 5).

Our methodology has relied on RP, a measure of biomass that takes into account the occlusion effect of the vegetation above the height stratum of interest[32–36]. Importantly, we show that RP is particularly effective in characterizing the lower height strata of the forest, something that TCH cannot capture. While several LiDAR metrics have been proposed and used in the past to capture the vertical structure of forests (e.g., see refs. 37,38,72,73), we believe that RP is a useful addition to these metrics because it is straight-forward to calculate and interpret and, as we have shown, it can provide rich information regarding the vertical distribution of biomass in the forest. We believe there are two important areas for future research. First, it would be fruitful to use terrestrial laser scanning (TLS) data to validate RP-based results regarding the vertical structure of forests. TLS has been extensively used to characterize the vertical structure of tropical forests and the impact of disturbances (e.g., logging and edge effects) on forest structure[74–76]. Second, understanding how metrics from spaceborne LiDAR sensors (e.g., GEDI[77] and ICESat-2[78]) relate to RP-based results has the potential to allow for the mapping of forest structure even in areas that lack ALS data.

We model RP as a function of several environmental variables using GAMs and we identify FTSs using a modified LidarLDA model. This latter approach is particularly useful because it enables the straightforward visualization of the spatial distribution of the vertical profile of forests. We believe that this methodology is generalizable and can be used to characterize forest structure in other areas that have large-scale airborne LiDAR data. Aside from methodological advances, we believe that the information derived from our analyses will be of wide use to managers, decision makers, and other researchers interested in forest structure and its relation to ecosystem services, quality of restoration projects, and impacts of forest disturbances.

Methods

LiDAR data

We rely on airborne LiDAR data from the "Improving Biomass Estimation Methods for the Amazon" (EBA) project conducted by the Brazilian National Institute for Space Research (INPE)[54,67]. These data were collected between 2016 and 2018 with full-waveform laser pulses being emitted from a Trimble Harrier 68i sensor (Timble; Sunnyvale, CA) onboard a Cessna aircraft (model 206). Each transect corresponded to a single flight line and covered an area of 375 ha (12.5 km × 300 m). The average flight height was 600 m above ground, the field of view was 30°, and the average point density was set to 4 points/m^2 (additional technical information regarding the LiDAR system and these flights can be found in published studies[54,79]). In total, LiDAR data were collected for 906 transects across the entire Brazilian Amazon. Most of these transects were located at random within areas classified as forests in 2015 by INPE, but some of these transects were also strategically located in areas with ongoing research projects from partner organizations.

These data have been used to create a regional biomass map[54], assess dynamics of forest canopy gaps[80] and how these gaps relate to tree mortality[79], determine carbon loss associated with anthropogenic and natural disturbances[51], and investigate the factors that influence forest height[67] within the Brazilian Amazon. However, it is important to note that all of these past studies have relied on top-of-the-canopy information whereas here we focus on exploring forest structure information for multiple height strata.

Pre-processing of LiDAR data

We subset the LiDAR data to retain only those returns associated with absolute scan angles lower than 5 degrees. We adopt this criterion to improve our calculation of the number of incoming returns for each height stratum (see "Return probability" description below). All vegetation returns were normalized to account for ground height based on a digital terrain model with a 1 m resolution. Furthermore, all normalized returns below 0 m were truncated to zero. Finally, we aggregated these data by counting the number of returns within 50 × 50 m horizontal pixels and within the following height bins: 1–9.9, 10–19.9, 20–29.9, 30–39.9, 40–49.9 m, and >50 m.

We excluded some of the LiDAR transects due to one of the following reasons. First, transects with problematic projection information (e.g., transects with no projection information or transects in which the projection information implied that the coordinates were outside the Amazon region) were excluded. Second, we only retained transects that were randomly located to avoid over-representing some research areas with multiple transects. Finally, because one of our predictor variables (i.e., distance to the road network) was not available for the eastern most part of the Brazilian Amazon (i.e., Maranhao and Tocantins states), we also removed transects in these states. Out of the original 906 transects, we retained 725 transects. Within these transects, we only kept pixels classified as forests by the MapBiomas land-use/land-cover (LULC) map for 2019[81]. Altogether, the

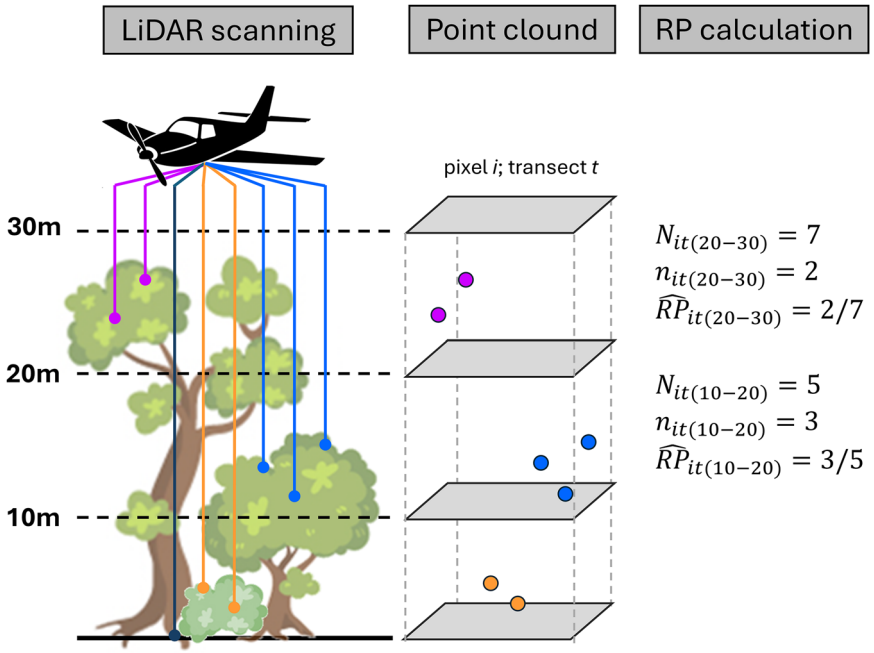

Fig. 5 | Schematic representation of the calculation of return probability (RP) between 10–20 m and 20−30 m for a set of 3 stacked voxels. The number of incoming returns N_{ith} is given by the sum of all returns below the upper height threshold.

final dataset represented a total area of approximately 188,000 ha (see spatial distribution of LiDAR transects in Supplementary Note 1).

Return probability (RP)

We focus on modeling RP_{ith} for each pixel i, transect t, and height class h. This probability can be estimated as the number of returns within a given height class divided by the number of incoming returns. More specifically, let n_{ith} denote the number of returns in pixel i in transect t and height strata h, and N_{ith} be the number of incoming returns for that height stratum, calculated as $N_{ith} = \sum_{j=0}^{h} n_{itj}$. RP can therefore be estimated as $\widehat{RP}_{ith} = \frac{n_{ith}}{N_{ith}}$. We illustrate the calculation of RP in Fig. 5. For example, if there are 3 returns between 10 and 20 m and a total of 5 returns below 20 m, then RP between 10 and 20 m (i.e., $\widehat{RP}_{it(10-20)}$) is estimated as $\frac{3}{5}$. Similarly, if there are 2 returns between 20 and 30 m and a total of 7 returns below 30 m, then RP between 20 and 30 m (i.e., $\widehat{RP}_{it(20-30)}$) is estimated as $\frac{2}{7}$.

The benefit of focusing on RP is that it inherently takes into account the occlusion effect of the vegetation above each height stratum (an important problem for ALS data[82]) by conditioning on the number of returns that reach that stratum. Furthermore, this probability can be used as a proxy of the amount/density of vegetation in different height strata[32–36]. For example[33], called this metric the "relative density model" (RDM) and showed how RDM between 1 and 5 m can be used to successfully map understory disturbances associated with selective logging (e.g., roads, skid-trails, landings, and harvested tree gaps) in the Brazilian Amazon region.

Comparison of top-of-canopy height (TCH) and the vertical distribution of return probability (RP)

RP is a measure that can be used to assess the vertical distribution of forest biomass, providing more information than what can be obtained from just top-of- canopy height (TCH) information. To obtain the TCH information, we first created a canopy height model (CHM) by computing the maximum of the normalized cloud (digital surface model minus digital terrain model) within 1 × 1 m pixels. Then, TCH was calculated by averaging the CHM results within 25 × 25 m pixels. We illustrate the advantage of RP relative to TCH by giving two examples: (1) comparing forests that were never clear cut to secondary forests of different ages; and (2) characterizing the impact of different forest disturbances.

For our first example, we relied on Mapbiomas product containing the 2019 age of secondary forests. This age is estimated based on the pixel-level time-series of land-use and land-cover (LULC) classes[83] and ranges from 0 (areas that were never clear cut) to 35 years of age. For each transect, we calculate the median RP and median TCH for areas that were never clear cut during this 35-year period (hereafter referred to simply as reference areas) and use this information as the baseline against which we compare the results for secondary forests of different ages. We expect to see increasing RP and TCH as secondary forests grow older, with more rapid recovery of RP to baseline levels in the lower height classes. We also hypothesize that the TCH recovery patterns will not correlate well with RP recovery patterns at the lower height classes, revealing the limitations associated with using only top-of-canopy information.

In our second example, we compare information from RP in each height strata with TCH information for transects impacted by different forest disturbances. More specifically, we assess the effect of fire, selective logging, and windthrows (three disturbances likely to have very different vertical profiles) for transects that were flown twice, before and after the disturbance. To this end, we reanalyzed the data used by ref. 51, in which disturbances were identified and categorized by a detailed examination of high-resolution Sentinel-2 imagery and PlanetScope NICFI mosaics for a subset of 99 randomly selected LiDAR transects from the EBA project. For our analysis, we restrict these data to only pixels that were flown twice and that were only impacted by a single type of disturbance, resulting in a total of approximately 5300, 5200, and 1200 pixels for fire, windthrows, and selective logging, respectively. For each of these pixels, we calculated the difference in pre- and post-disturbance RP and TCH. We expect that fire impact will be higher for lower height classes due to the bottom-up nature of this disturbance, whereas we expect that windthrows will have a greater impact in the upper height classes. Finally, we do not have a clear expectation regarding the effect of selective logging; while tree felling is likely to result in major changes at the top of the forest canopy, the infrastructure needed to remove the felled trees (i.e., roads and skid trails) is likely to have strong impact on the lower height classes.

Predictors of return probability (RP)

We relied on a generalized additive model (GAM[84]) to understand why some areas have higher RP than others. More specifically, we assume that:

$$n_{ith} \sim Binomial\left(N_{ith}, RP_{ith}\right) \quad (1)$$

where RP_{ith} is the RP. We assume that RP is given by:

$$logit\left(RP_{ith}\right) = \beta_{0h} + \sum_{p=1}^{P} s_{1hp}\left(x_{itp}\right) + s_{2h}\left(std.lat_{it}, std.long_{it}\right) \quad (2)$$

where $s_{1hp}(x_{itp})$ denotes a 1-D B-spline with 5 basis functions for predictor x_{itp} and height class h. We also include a 2-D thin-plate spline for the standardized latitude and longitude coordinates (i.e., $s_{2h}\left(\text{std.lat}_{it}, \text{std.long}_{it}\right)$) to account for spatial correlation arising from additional factors not included in the model. We fit an independent GAM for each height strata using the function "bam" within the R package "mgcv"[84].

The predictor variables for the GAM model consisted of:

Topographic variables: we extracted elevation (in meters) and slope information (m/100 m) for each pixel from the 1 km spatial resolution products provided by ref. 85. These products are based on the near global 90 m digital elevation model product of the Shuttle Radar Topography Mission (SRTM).

Precipitation: we first extracted monthly precipitation information (in mm per month) for each year and month between 2010 and 2019 from WorldClim[86] to then summarize this information by calculating its average across all 10 years. Using these data, we also calculated the maximum cumulative water deficit (MCWD[87]) as a proxy for rainfall seasonality.

Soil characteristics: we extracted soil properties information from SoilGrids[88]. SoilGrid maps have a 250 m spatial resolution and are based on prediction models fitted to over 230,000 soil profile observations from the WoSIS database (see WoSIS details in ref. 89). We relied on the following soil properties: cation exchange capacity (cmol/kg), soil pH, proportion of sand particles (g/100 g), and soil organic carbon content (g/kg). Information on these soil characteristics was available for different soil depths (i.e., 0–5, 5–15, 15–30, 30–60, and 60–100 cm). To summarize each metric across all these depths, we calculated a weighted average where the weights were given by the height of each soil layer (e.g., the weights for the 0–5 cm and 30–60 cm layers were 5/100 and 30/100, respectively).

Water table depth (WTD): WTD estimates (in meters) are provided by ref. 90 and are based on more than a million well measurements from across the world.

Road distance: To account for human influence on forest structure (e.g., through selective logging, understory fire, or other edge effects), we relied on the road network mapped by IMAZON[91] to quantify the distance (in meters) of each pixel to the nearest road.

We checked for multicollinearity problems with these predictor variables by calculating their variance inflation factor (VIF). All the variables, except MCWD, had a VIF less than 3.5 and thus were kept in the model. The spatial distribution of these variables is depicted in Supplementary Note 1. For the purposes of our GAMs, variables with very skewed distributions were transformed (i.e., $w = \log(q+1)$, where q and w are the original and transformed variables, respectively) to ensure a more even distribution of values. The only exception was WTD. Because this is a variable with negative values, we relied on the transformation $w = \log(-q+1)$ to make its distribution less skewed. All predictor variables were subsequently standardized (i.e., $x = \frac{w-\bar{w}}{sd_w}$, where $\bar{w}$ and sd_w are the mean and standard deviation of w) prior to model fitting.

Forest structural types (FST) and their spatial distribution

To simplify the characterization of the 3D forest structure, we identified clusters of pixels with similar vegetation type using a modified version of a model called LidarLDA[32]. The LidarLDA model is based on the Latent Dirichlet Allocation (LDA) model originally developed for text-mining[92] and later applied to a wide range of fields. In LidarLDA, a type of mixed membership model, a given pixel can be assigned to multiple clusters. While the original LidarLDA model assigned each return to a cluster, we modified this approach such that all returns within a 50 × 50 m pixel are assigned to a single cluster instead. This modification substantially reduces the number of latent variables in the model, making it more computationally efficient, while helping to ensure that results are more interpretable. We call the identified clusters FSTs because these forests have similar vertical vegetation profiles (i.e., similar RPs in the different height classes). We developed the Monte Carlo Markov Chain (MCMC) algorithm needed to fit the modified LidarLDA model using R[93] and C++[94]. A more detailed description of this modified LidarLDA model can be found in Supplementary Note 2.

To avoid taller height classes from overly influencing the identified FSTs due to their higher number of returns, we set the maximum number of incoming returns to 100, adjusting the number of returns within each height class accordingly. More specifically, we set the number of incoming returns to $\widetilde{N}_{ith} = 100$ and the number of returns per height strata to $\widetilde{n}_{ith} = \text{round}(\frac{n_{ith}}{N_{ith}} \times 100)$ whenever $N_{ith} > 100$. In this expression, $round(x)$ is a function that rounds x to the nearest integer. This approach ensures that different height classes generally have the same number of incoming returns, having similar influence when determining FSTs. We set the maximum number of FSTs to 5 because there are 5 main vegetation types in the region based on the IBGE classification (Supplementary Note 3). However, the modified LidarLDA model can identify fewer FSTs if that is best supported by the data because it relies on the truncated stick-breaking prior, a sparsity-inducing prior that is widely used to estimate the number of clusters[95]. Finally, because we are interested in the biogeography of the natural FSTs in the region, we fit this model only using data from forested areas that were not deemed to be secondary forests in 2019 by Mapbiomas[83].

The modified LidarLDA model characterizes each FST in terms of their height profiles (i.e., distribution of RP for each height class) and estimates the proportions of each FST in each transect. These FST proportion estimates were subsequently interpolated by creating spatial predictions using a random forest regression model[96]. More specifically, the random forest model was used to relate the estimated proportions of each FST to the predictor variables listed above and the spatial coordinates of the center of each transect. The fitted random forest model was then used to make predictions for systematically distributed points throughout the study region, enabling the creation of maps depicting the spatial distribution of each FST. To avoid extrapolations in covariate space, we only make predictions for areas in which the covariate values are within the range of the training dataset. Furthermore, to avoid spatial extrapolation, we restrict predictions to the convex hull derived from all the observations used to train the model. Finally, we restrict predictions to areas that were known to be forested in 2019 according to the LULC classification by MapBiomas.

We relied on two independent datasets to compare the ability of the estimated FSTs and the IBGE vegetation classes in summarizing forest structure in the region. The first dataset is a global map of TCH derived from the Copernicus Sentinel-2 mission, operated by the European Space Agency (ESA), and NASA's Global Ecosystem Dynamics Investigation (GEDI)[57]. The second dataset consisted of forest inventory data from ref. 38. This dataset contained information on the location (i.e., spatial coordinates), aboveground carbon density, plot size, and disturbance history (burning and selective logging) of each plot. We restricted our analysis to only plots located in areas with no forest disturbance (i.e., classified as intact forest by Longo and colleagues[38]) and to plots with 2300–2700 m^2 in size. Ultimately, this filtering process resulted in a dataset with 121 plots spread throughout the Brazilian Amazon region.

We compare the ability of the estimated FSTs and the IBGE vegetation classes in characterizing biomass and TCH by calculating R^2 for each classification approach, given by:

$$R^2_{IBGE} = 1 - \frac{\sum_i \sum_t \left(y_{it} - \bar{y}^{IBGE}_{k[it]}\right)^2}{\sum_i \sum_t \left(y_{it} - \bar{y}\right)^2} \quad (3)$$

$$R^2_{LidarLDA} = 1 - \frac{\sum_i \sum_t \left(y_{it} - \bar{y}^{LidarLDA}_{k[it]}\right)^2}{\sum_i \sum_t \left(y_{it} - \bar{y}\right)^2} \quad (4)$$

where y_{it} is the TCH (or biomass) for pixel i in transect t and $\bar{y}$ is the overall average. In these expressions, $\bar{y}_{k[it]}^{\text{IBGE}}$ is the median TCH (or biomass) for the IBGE vegetation class assigned to this pixel and transect (i.e., class k). Similarly, $\bar{y}_{k[it]}^{\text{LidarLDA}}$ is the median TCH (or biomass) for the FST assigned to this pixel and transect (i.e., FST k). Values of R^2 close to 1 indicate that the corresponding classification scheme explains almost all the variation in forest structure whereas values close to zero indicate that the classification scheme poorly represents forest structure.

All results and figures in this article were created using R and the following packages: ggplot2[97], sf[98,99], terra[100], mgcv[84], and randomforest[96].

Data availability

The LiDAR data from the EBA project can be found in the links below: a. Overview here: https://zenodo.org/records/4557390. b. Acre, Rondonia: https://zenodo.org/records/7689909. c. Mato Grosso, Amazonas, Para: https://zenodo.org/records/7636454. d. Roraima, Amapa: https://zenodo.org/records/7689693. The data used to create Figs. 2 and 3 can be found in the link https://zenodo.org/records/17487287.

Code availability

The modified LidarLDA model is fitted in a Bayesian framework with a Gibbs sampler. All the necessary code to run this model can be found in the public repository https://github.com/drvalle1/mLidarLDA. This repository also includes a brief tutorial describing how to fit the model and interpret its results when applied to simulated data.

Received: 21 February 2025; Accepted: 31 October 2025;
Published online: 11 December 2025

References

1. Food and Agriculture Organization (FAO). *The State of the World's Forests 2020. Forests, Biodiversity and People* (FAO, 2020).
2. Brickerhoff, E. et al. Forest biodiversity, ecosystem functioning and the provision of ecosystem services. *Biodivers. Conserv.* **26**, 3005–3035 (2017).
3. Houghton, R. A. & Nassikas, A. A. Negative emissions from stopping deforestation and forest degradation, globally. *Glob. Change Biol.* **24**, 350–359 (2018).
4. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
5. Hansen, M. C., Stehman, S. V. & Potapov, P. V. Quantification of global gross forest cover loss. *Proc. Natl Acad. Sci. USA* **107**, 8650–8655 (2010).
6. Keenan, R. J. et al. Dynamics of global forest area: results from the FAO Global Forest Resources Assessment 2015. *For. Ecol. Manag.* **352**, 9–20 (2015).
7. Ondei, S., Prior, L. D., Vigilante, T. & Bowman, D. M. J. S. Fire and cattle disturbance affects vegetation structure and rain forest expansion into savanna in the Australian monsoon tropics. *J. Biogeogr.* **44**, 2331–2342 (2017).
8. Benitez-Malvido, J. & Martinez-Ramos, M. Impact of forest fragmentation on understory plant species richness in Amazonia. *Conserv. Biol.* **17**, 389–400 (2003).
9. Broadbent, E. N. et al. Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. *Biol. Conserv.* **141**, 1745–1757 (2008).
10. Asner, G. P. et al. Selective logging in the Brazilian Amazon. *Science* **310**, 480–482 (2005).
11. Felipe-Lucia, M. R. et al. Multiple forest attributes underpin the supply of multiple ecosystem services. *Nat. Commun.* **9**, 4839 (2018).
12. Shah, N. W. et al. The effects of forest management on water quality. *For. Ecol. Manag.* **522**, 120397 (2022).
13. De Frenne, P. et al. Global buffering of temperatures under forest canopies. *Nat. Ecol. Evol.* **3**, 744–749 (2019).
14. Spracklen, D. V., Arnold, S. R. & Taylor, C. M. Observations of increased tropical rainfall preceded by air passage over forests. *Nature* **489**, 282–285 (2012).
15. Ellison, D. et al. Trees, forests and water: cool insights for a hot world. *Glob. Environ. Change* **43**, 51–61 (2017).
16. Jucker, T. et al. Canopy structure and topography jointly constrain the microclimate of human-modified tropical landscapes. *Glob. Change Biol.* **24**, 5243–5258 (2018).
17. Longo, M. et al. Impacts of degradation on water, energy, and carbon cycling of the Amazon tropical forests. *J. Geophys. Res.* **125**, e2020JG005677 (2020).
18. Pinage, E. R. et al. Surface energy dynamics and canopy structural properties in intact and disturbed forests in the Southern Amazon. *JGR Biogeosci.* https://doi.org/10.1029/2023JG007465 (2023).
19. Celentano, D. et al. Degradation of riparian forest affects soil properties and ecosystem services provision in eastern Amazon of Brazil. *Land Degrad. Dev.* **28**, 482–493 (2017).
20. Holz, D. J., Williard, K. W. J., Edwards, P. J. & Schoonover, J. E. Soil erosion in humid regions: a review. *J. Contemp. Water Res. Educ.* **154**, 48–59 (2015).
21. Jung, K., Kaiser, S., Bohm, S., Nieschulze, J. & Kalko, E. K. V. Moving in three dimensions: effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. *J. Appl. Ecol.* **49**, 523–531 (2012).
22. Muller, J., Moning, C., Bassler, C., Heurich, M. & Brandl, R. Using airborne laser scanning to model potential abundance and assemblages of forest passerines. *Basic Appl. Ecol.* **10**, 671–681 (2009).
23. Martinuzzi, S. et al. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sens. Environ.* **113**, 2533–2546 (2009).
24. Lim, K., Treitz, P., Wulder, M. A., St-Onge, B. & Flood, M. LiDAR remote sensing of forest structure. *Prog. Phys. Geogr.* **27**, 88–106 (2003).
25. Lefsky, M. A., Cohen, W. B. & Spies, T. A. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Can. J. For. Res.* **31**, 78–87 (2001).
26. Dubayah, R. & Drake, J. B. Lidar remote sensing for forestry. *J. For.* **98**, 44–46 (2000).
27. Bergen, K. M. et al. Remote sensing of vegetation 3-D structure for biodiversity andhabitat: Review and implications for lidar and radar spacebornemissions. *J. Geophys. Res.* **114**, G00E06 (2009).
28. Coops, N. C. et al. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sens. Environ.* **260**, 112477 (2021).
29. Brasil. *Projeto RADAMBRASIL. Vol. 1-34*. (Depto Nacional de Producao Mineral, 1978).
30. IBGE. *Manual técnico da vegetação brasileira: sistema fitogeográfico: inventário das formações florestais e campestres: técnicas e manejo de coleções botânicas: procedimentos para mapeamentos*. IBGE. Coordenação de Recursos Naturais e Estudos Ambientais. (2012).
31. Momsen, R. P. Jr. Projeto RADAM: a better look at the Brazilian tropics. *GeoJournal* **3**, 3–14 (1979).
32. Valle, D., Silva, C., Longo, M. & Brando, P. The latent dirichlet allocation model applied to airborne LiDAR data: a case study on mapping forest degradation associated with fragmentation and fire in the Amazon region. *Methods Ecol. Evol.* **13**, 1329–1342 (2022).
33. d'Oliveira, M. V. N., Reutebuch, S. E., McGaughey, R. J. & Andersen, H.-E. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sens. Environ.* **124**, 479–491 (2012).

34. Hosoi, F. & Omasa, K. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning Lidar. *IEEE Trans. Geosci. Remote Sens.* **44**, 3610–3618 (2006).
35. Moran, C. J., Rowell, E. M. & Seielstad, C. A. A data-driven framework to identify and compare forest structure classes using LiDAR. *Remote Sens. Environ.* **211**, 154–166 (2018).
36. Pimont, F., Allard, D., Soma, M. & Dupuy, J.-L. Estimators and confidence intervals for plant area density at voxel scale with T-LiDAR. *Remote Sens. Environ.* **215**, 343–370 (2018).
37. Atkins, J. W. et al. Scale dependency of lidar-derived forest structural diversity. *Methods Ecol. Evol.* **14**, 708–723 (2023).
38. Longo, M. et al. Aboveground biomass variability across intact and degraded forests in the Brazilian Amazon. *Glob. Biogeochem. Cycles* **30**, 1639–1660 (2016).
39. Fedrigo, M. et al. Predicting temperate forest stand types using only structural profiles from discrete return airborne lidar. *ISPRS J. Photogramm. Remote Sens.* **136**, 106–119 (2018).
40. Schmidt, M. J. et al. Intentional creation of carbon-rich dark earth soils in the Amazon. *Sci. Adv.* **9**, eadh8499 (2023).
41. Bourgoin, C. et al. Human degradation of tropical moist forests is greater than previously estimated. *Nature* **631**, 570–576 (2024).
42. Jacon, A. D. et al. Characterizing canopy structure variability in Amazonian secondary successions with full-waveform airborne LiDAR. *Remote Sens.* **16**, 2085 (2024).
43. Giles, A. L. et al. Simple ecological indicators benchmark regeneration success of Amazonian forests. *Commun. Earth Environ.* **5**, 780 (2024).
44. Lapola, D. M. et al. The drivers and impacts of Amazon forest degradation. *Science* **379**, eabp8622 (2023).
45. Cardil, A. et al. Recent deforestation drove the spike in Amazonian fires. *Environ. Res. Lett.* **15**, 121003 (2020).
46. Urquiza-Munoz, J. D. et al. Increased occurrence of large-scale windthrows across the Amazon Basin. *AGU Adv.* https://doi.org/10.1029/2023AV001030 (2024).
47. Feng, Y., Negron-Juarez, R. I., Romps, D. M. & Chambers, J. Q. Amazon windthrow disturbances are likely to increase with storm frequency under global warming. *Nat. Commun.* **14**, 101 (2023).
48. Brando, P. M. et al. Abrupt increases in Amazonian tree mortality due to drought-fire interactions. *Proc. Natl Acad. Sci. USA* **111**, 6347–6352 (2014).
49. Brandao, A. O. Jr. & Souza, C. M. Jr. Mapping unofficial roads with Landsat images: a new tool to improve the monitoring of the Brazilian Amazon rainforest. *Int. J. Remote Sens.* **27**, 177–189 (2006).
50. Matricardi, E. A. T. et al. Long-term forest degradation surpasses deforestation in the Brazilian Amazon. *Science* **369**, 1378–1382 (2020).
51. Csillik, O. et al. A large net carbon loss attributed to antrhopogenic and natural disturbances in the Amazon arc of deforestation. *Proc. Natl Acad. Sci. USA* **121**, e2310157121 (2024).
52. ter Steege, H. et al. Mapping density, diversity and species-richness of the Amazon tree flora. *Commun. Biol.* **6**, 1130 (2023).
53. Marca-Zevallos, M. J. et al. Local hydrological conditions influence tree diversity and composition across the Amazon basin. *Ecography* **2022**, e06125 (2022).
54. Ometto, J. P. et al. A biomass map of the Brazilian Amazon from multisource remote sensing. *Sci. Data* **10**, 668 (2023).
55. Baccini, A. et al. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nat. Clim. Change* **2**, 182–185 (2012).
56. Xu, L. et al. Changes in global terrestrial live biomass over the 21st century. *Sci. Adv.* **7**, eabe9829 (2021).
57. Lang, N., Jetz, W., Schindler, K. & Wegner, J. D. A high-resolution canopy height model of the Earth. *Nat. Ecol. Evol.* **7**, 1778–1789 (2023).
58. Tolan, J. et al. Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sens. Environ.* **300**, 113888 (2024).
59. Myster, R. W. The physical structure of forests in the Amazon Basin: a review. *Bot. Rev.* **82**, 407–427 (2016).
60. Targhetta, N., Kesselmeier, J. & Wittmann, F. Effects of the hydroedaphic gradient on tree species composition and aboveground wood biomass of oligotrophic forest ecosystems in the central Amazon basin. *Folia Geobot.* **50**, 185–205 (2015).
61. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51**, 933–938 (2001).
62. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534–545 (2017).
63. Ter Steege, H., Jetten, V. G., Polak, A. M. & Werger, M. J. A. Tropical rain forest types and soil factors in a watershed area in Guyana. *J. Veg. Sci.* **4**, 705–716 (1993).
64. Laurance, W. F. et al. Relationship between soils and Amazon forest biomass: a landscape-scale study. *For. Ecol. Manag.* **118**, 127–138 (1999).
65. Toledo, J. J., Castilho, C. V., Magnusson, W. E. & Nascimento, H. E. M. Soil controls biomass and dynamics of an Amazonian forest through the shifting of species and traits. *Braz. J. Bot.* **40**, 451–461 (2017).
66. Williams, M. et al. Heterogeneity of soils and vegetation in an Eastern Amazonian rain forest: implications for scaling up biomass and production. *Ecosystems* **5**, 692–704 (2002).
67. Gorgens, E. B. et al. Resource availability and disturbance shape maximum tree height across the Amazon. *Glob. Change Biol.* **27**, 177–189 (2021).
68. Brando, P. M. et al. Drought effects on litterfall, wood produciton and below ground carbon cycling in an Amazon forest: results of a throughfall reduction experiment. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 1839–1848 (2008).
69. Hilker, T. et al. Vegetation dynamics and rainfall sensitivity of the Amazon. *Proc. Natl Acad. Sci. USA* **111**, 16041–16046 (2014).
70. Poorter, L. et al. Diversity enhances carbon storage in tropical forests. *Glob. Ecol. Biogeogr.* **24**, 1314–1328 (2015).
71. Malhi, Y. et al. An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). *J. Veg. Sci.* **13**, 439–450 (2002).
72. Almeida, D. R. A. et al. Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and lidar fusion. *Remote Sens. Environ.* **264**, 112582 (2021).
73. Burns, P., Hakkenberg, C. & Goetz, S. J. Gridded GEDI vegetation structure metrics and biomass density at multiple resolutions (2024).
74. Santos, E. G., Nunes, M. H., Jackson, T. & Maeda, E. E. Quantifying tropical forest disturbances using canopy structural traits derived from terrestrial laser scanning. *For. Ecol. Manag.* **524**, 120546 (2022).
75. Maeda, E. E. et al. Shifts in structural diversity of Amazonian forest edges detected using terrestrial laser scanning. *Remote Sens. Environ.* **271**, 112895 (2022).
76. Ehbrecht, M. et al. Global patterns and climatic controls of forest structural complexity. *Nat. Commun.* **12**, 519 (2021).
77. Dubayah, R. et al. The global ecosystem dynamics investigation: high-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* **1**, 100002 (2020).
78. Palm, S. P. et al. ICESat-2 atmospheric channel description, data processing and first results. *Earth Space Sci.* **8**, e2020EA001470 (2021).
79. Dalagnol, R. et al. Large-scale variations in the dynamics of Amazon forest canopy gaps from airborne lidar data and opportunities for tree mortality estimates. *Sci. Rep.* **11**, 1388 (2021).

80. Dalagnol, R. et al. Mapping tropical forest degradation with deep learning and Planet NICFI data. *Remote Sens. Environ.* **298**, 113798 (2023).
81. Souza Jr., C. M. et al. Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat archive and Earth Engine. *Remote Sens.* **12**, 2735 (2020).
82. Crespo-Peremarch, P., Fournier, R. A., Nguyen, V.-T., van Lier, O. R. & Ruiz, L. A. A comparative assessment of the vertical distribution of forest components using full-waveform airborne, discrete airborne and discrete terrestrial laser scanning data. *For. Ecol. Manag.* **473**, 118268 (2020).
83. Silva, Jr. C. H. L. et al. Benchmark maps of 33 years of secondary forest age for Brazil. *Sci. Data* **7**, 269 (2020).
84. Wood, S. N. *Generalized Additive Models: An Introduction With R* (CRC Press, 2017).
85. Amatulli, G. et al. Data descriptor: a suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci. Data* **5**, 180040 (2018).
86. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
87. Aragao, L. E. O. C. et al. Spatial patterns and fire response of recent Amazonian droughts. *Geophys. Res. Lett.* **34**, L07701 (2007).
88. Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240 (2021).
89. Batjes, N. H. et al. WoSIS: providing standardised soil profile data for the world. *Earth Syst. Sci. Data* **9**, 1–14 (2017).
90. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table Depth. *Science* **339**, 940–943 (2013).
91. ImazonGeo. Instituto do Homem e Meio Ambiente da Amazonia (IMAZON) (Geoportal ImazonGeo, 2022).
92. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
93. R Core Team. R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2020).
94. Eddelbuettel, D. & Francois, R. Rcpp: seamless R and C++ integration. *J. Stat Softw.* **40**, 1–18 (2011).
95. Valle, D. et al. Automatic selection of the number of clusters usingBayesian clustering and sparsity-inducing priors. *Ecol. Appl.* **32**, e2524 (2022).
96. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).
97. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
98. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J* **10**, 439–446 (2018).
99. Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R* 1st edN (Chapman and Hall/CRC, 2023).
100. Hijmans, R. J. terra: spatial data analysis. (2024).

Acknowledgements

This work was supported by the US Department of Agriculture National Institute of Food and Agriculture McIntire–Stennis project 1005163, US National Science Foundation award 2040819, and NASA's Carbon Monitoring System grant 80NSSC23K1257 to DV. This manuscript has authors at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. O.C., M.L. and M.K. were supported by the Next Generation Ecosystem Experiments-Tropics, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges, that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

Author contributions

D.V. conceptualized the article, developed the methodology, performed the analysis, created all figures (except for Fig. 5), and wrote the first draft. L.H. and I.V.B. created Fig. 5. J.O. and M.K. acquired the LiDAR data. L.H., I.V.B., J.O., O.C., M.L., M.K., and D.A. reviewed the article and provided feedback that substantially improved it.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-025-02976-w.

Correspondence and requests for materials should be addressed to Denis Valle.

Peer review information *Communications Earth & Environment* thanks Jose Urquiza-Muñoz and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary handling editors: Tobias Landmann and Mengjie Wang. A peer review file is available.