scientific data

OPEN

DATA DESCRIPTOR

Comprehensive uncrewed aerial system data for Amazon rainforest at Tiputini Biodiversity Station, Ecuador

Minyoung Jung[1], Anjin Chang[2] ✉, Charles H. Cannon[3], Gonzalo Rivas-Torres[4,5] & Jinha Jung[1]

The Amazon rainforest, the most biodiverse biome on Earth, represents a critical natural laboratory for studying species interactions, ecosystem functioning, and vulnerability to climate change. High-quality datasets are crucial for advancing a detailed understanding of this complex biome; however, open-access uncrewed aerial system (UAS) datasets for this region remain notably scarce. Here, we present UAS-derived high-resolution products that seamlessly cover over 700 hectares of the upper and hyperdiverse Amazon. To the best of our knowledge, this is the first large-scale, comprehensive, open, and seamless UAS dataset of the Amazon, integrating multispectral imagery and lidar (light detection and ranging) data collected during a single mission, ensuring temporal consistency. The dataset provides complementary spectral and structural information from the Tiputini Biodiversity Station, one of the regions with the highest tree species richness per unit area on Earth. This resource supports diverse applications such as biodiversity and forest-structure analyses and establishes a methodological reference for future large-scale UAS research in the Amazon.

Background & Summary

The Amazon rainforest is the most biodiverse ecosystem on Earth, supporting unparalleled species richness and playing a central role in global ecological and climatic processes. Despite covering only about 1% of Earth's surface, Amazonia is home to an estimated 10% of all known species[1], earning it the title of the "*world's richest repository of life*." Its biodiversity (including countless trees, plants, insects, birds, mammals, and amphibians that form complex food webs and ecological interactions) underpins essential ecosystem services such as carbon storage, hydrological regulation, and climate stabilization, which are vital at both regional (for indigenous cultures and local communities) and global scales. Given its ecological significance, research in the Amazon has predominantly centered on understanding biodiversity patterns[2], quantifying carbon stocks[3], and monitoring forest structure and dynamics[4] in the context of climate change and increasing anthropogenic pressures.

The Amazon forest spans over 5.5 million square kilometers, covering a significant part of the South American continent. Monitoring its biodiversity is a daunting task due to its vast extent, dense canopy, and limited accessibility, while large-scale field-based surveys are logistically infeasible. As a result, remotely sensed data have been widely employed, particularly satellite imagery that covers large areas at moderate spatial resolution. A wide range of sensors, including multi-spectral[5–7], light detection and ranging (lidar)[8,9], and synthetic aperture radar (SAR)[10,11], have been used to characterize this data-scarce region. Typically, the Amazon forest has been monitored by public organizations or government agencies. The Brazilian National Institute for Space Research (INPE) launched the Program for Calculating Deforestation in the Amazon (PRODES) and the Real-Time Deforestation Detection System (DETER)[6,7]. The Global Ecosystem Dynamics Investigation (GEDI)

[1]Lyles School of Civil and Construction Engineering, Purdue University, West Lafayette, IN, 47907, USA. [2]Department of Biosystems & Agricultural Engineering, College of Agriculture & Natural Resources, Michigan State University, East Lansing, MI, 48824, USA. [3]Forest and Tree Research, Singapore Botanic Gardens, National Parks Board, Singapore. [4]Universidad San Francisco de Quito, Estación de Biodiversidad Tiputini, Av. Diego de Robles e Interoceánica, Quito, Ecuador. [5]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig, 04103, Germany. ✉e-mail: changanj@msu.edu

mission by NASA has provided over 7 billion full-waveform lidar measurements, creating unprecedented datasets for forest structure analysis that strongly correlate with patterns of species diversity[12,13]. Additionally, SAR has been commonly used to monitor land-use change, forest degradation, and regrowth in the Amazon[14–17]. Because SAR signals can penetrate dense canopies and are sensitive to woody components, key indicators of forest biomass (i.e., trunks and branches), SAR represents a valuable data source for biomass estimation in closed-canopy forests such as the Amazon, as further highlighted by the BIOMASS mission led by the European Space Agency (ESA)[14]. These satellite-based studies have provided critical insights into large-scale changes in forest cover, biomass dynamics, and climatic feedbacks. However, their relatively coarse resolution constrains the ability to capture the Amazon's structurally complex and ecologically diverse landscapes, prompting continued calls for higher-resolution data. A promising resource in this regard was Planet's Norway International Climate and Forests Initiative (NICFI) program, which provided free access to tropical multispectral satellite mosaics (red, green, blue, and near-infrared) at a spatial resolution of 4.7 m from September 2020 to January 2025[18], demonstrating the value of publicly available high-resolution remote sensing data for complex rainforest monitoring[19–24].

In the meantime, airborne platforms equipped with various sensors have been extensively adopted to provide finer spatial detail[25–27]. More recently, advances in uncrewed aerial systems (UAS) have demonstrated their value in efficiently surveying remote and inaccessible parts of the Amazon, offering spatial resolutions far beyond those of satellites and traditional airborne systems[28–32]. However, UAS data are inherently constrained in coverage, making spatially continuous monitoring across the Amazon basin infeasible. Consequently, most UAS-based studies in Amazonian forests have focused on relatively small, localized areas (typically a few to several tens of hectares), with large-scale applications remaining rare. Although such research has demonstrated the utility of UAS, study areas spanning only tens of hectares are often insufficient to represent the multiple habitat types that define Amazonian ecosystems[33–35]. As a result, biodiversity studies based on such spatially limited data can fail to capture the heterogeneity and sub-environmental gradients characteristic of the Amazon. Addressing this limitation requires efforts to acquire seamless UAS datasets covering larger areas, thereby supporting a broader range of studies in the Amazon, including fine-scale assessments of ecological transitions at habitat boundaries.

Most high-quality datasets collected in the Amazon have not been made publicly available and are often accessible only through derived products presented within publications, while raw data (e.g., orthomosaics, point clouds) remain extremely limited. To the best of our knowledge, only a few datasets are openly available (specifically, two[36,37]). This restricted accessibility is largely due to the substantial data volumes involved, the sensitivity of surveyed locations, and restrictions on access to protected areas. Yet, as demonstrated by the uptake of other openly available satellite data, such as NICFI imagery, open access to high-quality UAS raw datasets could facilitate a broad range of downstream applications.

Therefore, in this study, we collected high-quality UAS data over the Amazon rainforest, specifically at the Tiputini Biodiversity Station (TBS) in the Ecuadorian Amazon, which is recognized as one of the most biodiverse regions on Earth. We openly released both the raw data and derivative data products (i.e., orthomosaics; point clouds; digital elevation models, DEMs; and a canopy height model, CHM) in an efficient and accessible format. Our dataset represents an exceptional resource for Amazon in the following aspects:

1. Comprehensive information: Two different sensors (multispectral and lidar) were deployed during the same campaign, enabling both spectral and structural information to be obtained within a short temporal window (up to three days).
2. Spatial detail: The dataset comprises seven-band multispectral orthomosaics with a spatial resolution of 5 cm and point clouds with more than 300 points per square meter, along with their derivative products.
3. Spatial coverage: The dataset seamlessly covers an area of more than 700 ha, encompassing multiple habitat types.

This study describes the details of this exceptional UAS dataset, including data quality, and documents the methodology used to generate seamless, large-area data products. The dataset provides a valuable resource for biodiversity monitoring, carbon and biomass estimation, and structural and compositional analyses, while also serving as a reference for future large-scale UAS campaigns in tropical forest ecosystems.

Methods

Site description. The TBS is a remote research facility located within the Yasuní Biosphere Reserve, in a minimally disturbed primary rainforest of the Ecuadorian Amazon (0°38′ S, 76°09′ W; at a mean elevation of approximately 230 m above sea level), accessible only by boat along the Tiputini River (Fig. 1). The TBS's climate is perhumid tropical with a mean annual precipitation of 3,100 mm/year[38]. The yearly mean temperature is approximately 27 °C, with a stable annual variation. The TBS is located in an area with the highest recorded tree species richness per hectare on Earth[39], encompassing terra firme and white-water-inundated forests, palm swamps and other wetlands, as well as natural succession areas[40]. Its isolation, biological richness, and well-established infrastructure (e.g., two observation towers, about 40 km of well-marked trails, laboratories, and geo-referenced research plots, including two 100 ha units with a 100 m grid system) make it a premier site for scientific investigation and environmental data collection in tropical ecosystems, such as long-term ecological research, biodiversity monitoring, and conservation studies.

Since its establishment, valuable ecological studies have been conducted through repeated field surveys in the region[40,41–46]. However, no high-resolution dataset has been available to provide wall-to-wall spatial information for this large-scale research facility. Although time-series analyses using medium-resolution satellite data such as Landsat and Sentinel are feasible, their coarse spatial resolution limits the ability to investigate this region's rich

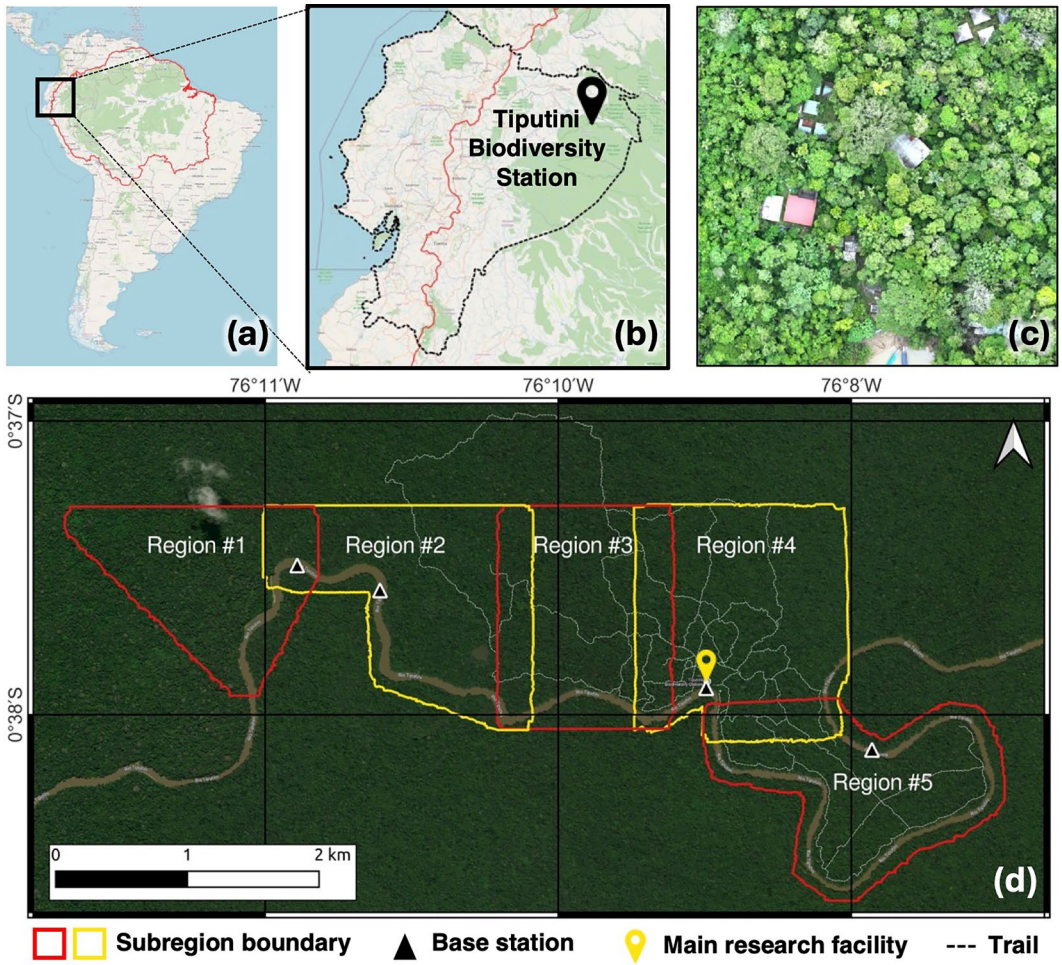

Fig. 1 (**a**) Amazonia region[80] (red line); (**b**) Ecuador (dashed line) with the Tiputini Biodiversity Station (TBS); (**c**) TBS main research facilities; and (**d**) TBS region divided into five subregions for uncrewed aerial system (UAS) missions.

biodiversity beyond broad regional patterns. Structural information, such as CHMs, is even more constrained. Despite the increasing use of GEDI Level 2A relative height products (with a spatial resolution of approximately 25 m), high-quality data remain scarce—only 110 footprints have a '*quality_flag*' of 1, indicating higher scientific reliability[47]—thereby constraining their broader applicability across research fields, particularly for wall-to-wall analyses. Accordingly, this study collected a high-quality, comprehensive UAS dataset covering approximately 7 km^2 of this intact tropical forest.

As illustrated in Fig. 1(d), the study area encompasses the primary research infrastructure at TBS. The site spans 712 ha, with a maximum length of 7.5 km (west to east) and a width of 2.8 km (north to south). Due to the extensive size of the study area, which could not be fully surveyed in a single day or from a single base location, the region was divided into five subregions for UAS surveys. Areas spanning approximately 200 m along both banks of the Tiputini River in the southern portion of each region are classified as seasonally inundated forest. The northern regions of Subregions 1, 3, and 4 are characterized as terra firme forest, while the northern region of Subregion 2 includes areas of swamp forest. These three ecosystems represent the dominant forest types across the Amazon biome.

UAS equipment and sensors. Two UAS with different sensors (multispectral camera and lidar) were used to collect comprehensive, high-quality UAS data for the study site. Multispectral images were collected using a quadcopter DJI Mavic 3 Multispectral edition. It consists of two different cameras, RGB and multispectral. The RGB camera is a 20 MP sensor with 4/3" CMOS, a 24 mm equivalent, adjustable aperture lens (*f*/2.8 to *f*/11), and a mechanical shutter allowing a shutter speed of up to 1/2000 seconds. The multispectral camera includes four 5 MP bands with 1/2.8" CMOS: Green (560 ± 16 nm), Red (650 ± 16 nm), Red Edge (730 ± 16 nm), and Near-Infrared (NIR; 860 ± 26 nm). Different vegetation indices, such as NDVI (normalized difference vegetation index), NDRE (normalized difference red edge), GNDVI (green normalized difference vegetation index), and SAVI (soil adjusted vegetation index), can be derived from the four multispectral bands, thereby effectively supporting different vegetation and/or tree-based applications, such as forest canopy health analysis and crop mapping and prescription.

Sub-region	Area (km²)	Multispectral			Lidar	
		Date	Flight Time	# Images	Date	Flight Time
1	1.19	2024/06/09	45 m 04 s	1,348	2024/06/09	30 m 30 s
2	1.78	2024/06/09	1 h 06 m 31 s	1,990	2024/06/09	41 m 58 s
3	1.73	2024/06/10	1 h 04 m 52 s	1,940	2024/06/08	40 m 01 s
4	2.14	2024/06/11	1 h 18 m 42 s	2,354	2024/06/08	48 m 43 s
5	1.49	2024/06/10	58 m 08 s	1,739	2024/06/10	38 m 23 s

Table 1. Information on UAS missions for each subregion.

	Multispectral	Lidar
Altitude (m)	150	150
Flight Speed (m/s)	13.4	15
Forward Overlap (%)	80	70 (RGB)
Side Overlap (%)	80	61 (RGB) / 50 (lidar)
Altitude Mode	AGL	AGL
Return Mode	—	Penta Return
Scanning Mode	—	Non-repetitive
Sampling Rate (kHz)	—	240
Expected GSD (cm/pixel)	6.92 (multispectral camera)	4.03 (RGB camera)
Expected Purse Density (purse/m²)	—	122

Table 2. Flight configuration of UAS missions. The forward overlap for the lidar is only related to the additional RGB camera of the DJI Zenmuse L2 sensor. AGL and GSD indicate above ground level and ground sampling distance, respectively. For the AGL mode, the ASTER GDEM version 3[79] was used in the mission planning procedure.

The 3D point cloud was collected using a DJI Zenmuse L2 sensor mounted on a quadcopter DJI Matrice 350RTK (real-time kinematic) platform. The DJI Zenmuse L2 sensor integrates a high-precision IMU and an additional RGB mapping camera. The lidar supports recording up to 5 returns, which allows detailed canopy penetration capability, and delivers centimeter-level geolocation accuracy (vertical 4 cm, horizontal 5 cm at 150 m). The additional RGB sensor is a 20 MP sensor with 4/3" CMOS, which enables the capture of RGB color information and thereby generates RGB-colored point clouds. Both UASs can be associated with the DJI D-RTK 2 Mobile Station, which supports multiple GNSS (Global Navigation Satellite System) constellations, including GPS, GLONASS, BeiDou, and Galileo, ensuring fast initialization and robust correction signals even in challenging environments.

UAS data acquisition. Due to the extensive size of the study area, the region was divided into five subregions for UAS surveys conducted over four days, as shown in Fig. 1(d). Each subregion overlaps with its neighboring ones and has a single base location, where both the DJI D-RTK 2 Mobile Station was set up and both UAS took off (Subregions 3 and 4 shared the same base station). The details of UAS missions for each subregion are provided in Table 1. The flight configuration, as summarized in Table 2, was consistent across the flight missions. To maximize the capture of under-canopy structures, like understory and tree trunks, the lidar was operated in non-repetitive scanning mode, recording up to five returns per pulse[48]. To minimize environmental variability, such as variations in water levels and atmospheric conditions, and ensure high consistency among the collected data, UAS missions were conducted under sunny or partly cloudy conditions, which meant there was no significant precipitation throughout the data collection.

As the study site is located deep within the Amazon rainforest, no '*known point*' for RTK GPS (global positioning system) mode was available, and there was also no access to an RTK correction source, such as NTRIP (Networked Transport of RTCM via Internet Protocol). Alternatively, a single averaged coordinate determined by the DJI D-RTK 2 Mobile Station was used as the known point (base station) to operate the UAS in RTK mode under these dense forest conditions. The base station was installed in relatively open areas, such as sand dunes or a wooden deck along the riverside. To reduce the duration of signal blackout between the base station and the UAS (rover) during flight missions, the flight paths were planned so that each strip included as many portions as possible close to the base station, maximizing the time spent within a stronger signal range. Consequently, the flight paths were oriented north–south in Subregion 4 and east–west in the other subregions for both UAS platforms.

The expected absolute geolocation accuracy of the single average mode is approximately ±31.5 m horizontal and ±3 m vertical[49], which was considerably coarser than the expected GSD (ground sampling distance); in contrast, the relative positioning accuracy between the base station and the UAS is still expected to be fairly high. GCPs (ground control points) could be useful in UAS-based surveying with high-precision georeferencing. However, since the study site is too broad and inaccessible to evenly distribute GCPs across the entire region, and it was expected to be nearly impossible to clearly observe the GCPs from the collected UAS raw image due

Subregion	# Total images	# Fixed images	# Float image
1	1,354	1,351	3
2	1,993	1,992	1
3	2,763	872	1,891
4	2,390	149	*2,241
5	1,744	329	1,415
Total	10,244	4,693	5,551

Table 3. Number of total raw images, successfully corrected (fix), and not corrected (float) images. * includes one single point.

to the dense vegetation, this study could not use GCPs. Consequently, the collected UAS data inevitably contained absolute positioning errors, thereby resulting in geospatial misalignments or shifts between neighboring subregions. To improve the geo-locational accuracy and spatial consistency of the collected UAS data, this study employed two additional processing steps, PPK (post-processed kinematics) for raw images and image registrations between regions, as presented in the following section.

Data processing. *Post-processed kinematic (PPK).* PPK is a GNSS correction method that significantly improves geolocation accuracy, often to within a few centimeters. To apply PPK correction, a base station (i.e., a stationary GNSS receiver) is first established at a known position. During the survey, both the base station and the rover (i.e., a UAS) simultaneously record raw GNSS satellite signals. After the survey, the rover's data is post-processed against the base station's data to compute precise positions for each measurement (i.e., image capture), correcting for potential satellite and atmospheric errors. Because PPK performs corrections after the mission, it is more robust in environments with limited connectivity and does not require a real-time radio or internet link, unlike RTK correction. It also reduces reliance on GCPs, making it a practical and reliable solution for achieving accurate geolocation in remote, densely forested regions such as the Amazon.

As described in earlier sections, the study site was divided into five subregions for the UAS data acquisition. A base station was installed at four different unknown locations during UAS missions with RTK mode and logged satellite data for 2 to 3 hours at each site (Fig. 1(d)). We used the IBGE-PPP service to determine the base station's coordinates after UAS missions. IBGE-PPP is a free online service for post-processing GNSS data provided by the Instituto Brasileiro de Geografia e Estatística (Brazilian Institute of Geography and Statistics, https://www.ibge.gov.br/). It uses the CSRS-PPP (GPS Precise Point Positioning) program developed by NRCan (Geodetic Survey Division of Natural Resources of Canada). The IBGE-PPP processes GNSS data (GPS and GLONASS) that were collected by single- or dual-frequency receivers in static or kinematic mode. Since our study site is outside of Brazil, the IBGE-PPP service calculated the solution using the 'FINAL' orbit type, which is available 12–20 days after the end of the survey with an accuracy of up to 2.5 cm.

Subsequently, the exterior orientations (including 3D coordinates) of each raw multispectral image were computed using the REDtoolbox software (REDcatch GmbH, Innsbruck, Austria, v.3.3.0) with PPK correction. The base station coordinates obtained from IBGE-PPP were used as the known positions for PPK processing. Raw satellite data from both the base station and the rover were post-processed, and the corrected image's exterior orientations were extracted for further multispectral data processing. As shown in Table 3, 45.8% of the total 10,244 image capture locations were successfully corrected (fixed), while the remainder could not be solved with comparable precision (float), likely due to subregional variation and environmental interference near the base stations. The corrected coordinates of the base stations were used for PPK processing to update the locations of UAS images. Most of the image locations were fixed conditions in the western Subregions 1 and 2, but there are more 'float' or 'single' conditions in the eastern Subregions 4 and 5. Most images on the west side had fixed coordinates with small errors (<3 cm), but fewer images in Subregions 3, 4, and 5 were successfully corrected. In Subregions 1 and 2, the base station was set up on a sand dune right next to the river with more open sky. The surroundings of the base station had a significant impact on PPK processing. The error amount for each image location was also input into Agisoft Metashape, allowing the software to update the external orientations used for the simultaneous processing of the five subregion datasets to generate seamless multispectral orthomosaic images.

Multispectral data. Since the raw multispectral UAS data consisted of a series of continuous individual images, a photogrammetry process was required to generate meaningful geospatial data products, including 3D point clouds, digital surface models (DSMs), and orthomosaics (i.e., orthorectified mosaic images). Photogrammetric reconstruction was performed by the Structure from Motion (SfM) algorithm using Agisoft Metashape Professional (Agisoft LLC, St. Petersburg, Russia, v.2.1.0). Although the RGB camera of the multispectral UAS captures higher-resolution images than its multispectral counterpart, the multispectral data provided sufficient spatial resolution for the purposes of this study (the expected GSD of orthomosaics for the multispectral camera is about 7 cm/pixel at the altitude of 150 m, as presented in Table 2); therefore, an RGB orthomosaic was not separately generated. The SfM processing workflow followed the typical photogrammetric procedures outlined in the official user manual[50] with two additional steps: camera alignment optimization and radiometric calibration. Parameters used for the initial multispectral data processing are summarized in Table 4. Notably, to improve georeferencing accuracy, the adjusted image locations (X, Y, and Z; longitude, latitude, and height), which were derived from the PPK processing, for the individual raw images were used instead of the initial estimates

1. Align photos	Accuracy	High
	Generic preselection	Enabled
	Reference preselection	Enabled (Source)
	Key point limit	40,000
	Tie point limit	4,000
	Apply masks to	None
	Exclude stationary tie points	Enabled
	Guided image matching	Disabled
	Adaptive camera model fitting	Enabled
2. Optimize Camera Alignment	Selected components to optimize	Focal length (*f*)
		Principal point (*cx, cy*)
		Lens radial distortion (*k1, k2, k3*)
		Tangential distortion (*p1, p2*)
3. Build Point Cloud	Quality	High
	Depth filtering	Mild
	Calculate point colors	Enabled
	Calculate point confidence	Enabled
4. Build Digital Elevation Model (DEM)	Source data	Point cloud
	Interpolation	Enabled
5. Build Orthomosaic	Surface	DEM
	Blending mode	Mosaic
	Refine seamlines	Disabled
	Hole filling	Enabled
	Ghosting filter	Disabled

Table 4. Photogrammetry parameters in Agisoft Metashape Professional.

embedded in the images' metadata. Additionally, since the multispectral UAS is equipped with a sunlight sensor that captures solar irradiance during data acquisition, radiometric calibration was performed using the recorded irradiance data in each image file. As a result, the final multispectral orthomosaics represent reflectance values suitable for spectral analysis-based research.

As this study focused on providing geometrically and radiometrically continuous maps for the TBS region, all the raw multispectral data of the five subregions were processed simultaneously via a single photogrammetry pipeline. Given the computational demands of processing large-scale multispectral datasets (over ten thousand images), we implemented a private high-performance computing (HPC) cluster composed of five local GPU (GeForce RTX 4090) workstations connected with 10 Gigabit Ethernet. The parallel processing pipeline automatically distributed processing tasks across nodes by running the Agisoft Metashape in network mode. This distributed setup allowed the raw data collected during separate UAS missions across the five subregions to be seamlessly mosaiced, resulting in a continuous orthoimage covering the entire 712-hectare study area.

Lidar data. The collected lidar data were reconstructed into 3D point cloud data using DJI Terra software (DJI, Shenzhen, China, v. 4.4.0). The corrected coordinates of each base station by IBGE-PPP were used instead of the initially recorded ones to improve the 3D point cloud's absolute geolocation accuracy. The reconstruction parameters were set as in Table 5 to increase the point density and the overall consistency and accuracy of the output point cloud[51]. Ground points were classified during reconstruction, and a 50-centimeter digital terrain model (DTM) was also generated. The heights of the point clouds were normalized by subtracting the DTMs to be heights above ground using pdal *filters.hag_dem*[52], and, finally, the 25-centimeter CHM for each subregion was produced. Considering the presence of random noise, canopy height was defined as the 98th percentile of normalized lidar point heights within each pixel, consistent with previous studies[53,54].

Unlike the multispectral data, the lidar data was processed by each UAS mission due to the extremely large data volume and the lack of supportable computational resources, resulting in separate geospatial products for each subregion. To generate seamless lidar-derived data products, including 3D point cloud, DTM, and CHM, across the entire study region (comparable to the multispectral data products), it is necessary to merge the lidar data products from all five subregions.

Although the positions of base stations corrected by the IBGE-PPP service were used during 3D point cloud reconstruction in the Terra software, misalignments between adjacent subregions were observed, likely due to intermittent signal degradation caused by canopy interference, weak GNSS reception in the dense rainforest environment, and uncertainty of geolocation positioning in GNSS system. Therefore, point cloud registration was conducted to minimize these misalignments and ensure geometric consistency prior to merging. Since the structural complexity and visual homogeneity of the rainforest landscape made manual selection of corresponding features between subregions nearly impossible, we implemented an automated registration approach. A standard solution involves registering the original 3D point clouds. However, achieving precise registration requires an iterative approach, and conventional point cloud registration methods (e.g., iterative closest point; ICP[55])

Point Cloud Density		By percentage (100%)
Point Cloud Effective Distance Range		3–300 m
Optimized Point Cloud Accuracy		Enabled
Ground Point Classi-fication	Ground Type	Gentle slope
	Building Max Diagonal	20 m
	Iteration Angle	6 degrees
	Iteration Distance	0.5 m
	DEM GSD	0.5 m

Table 5. Reconstruction parameters of DJI Terra.

Subregion		**1**	**3**	**4**	**5**
East	Mode	0.75	1.00	1.25	−0.75
	Standard Deviation	0.73	0.66	0.76	1.29
North	Mode	−0.25	0.25	0.00	0.88
	Standard Deviation	0.69	0.65	0.64	0.87
Vertical		0.65	1.75	−0.70	3.05

Table 6. Estimated shifts according to the reference Subregion 2 (unit: meter).

demand significant computational time and high-performance computing resources, especially for large areas with high point density, such as the dataset used in this study. Therefore, instead of registering the original point clouds directly, in this study, the amounts of spatial misalignments were estimated by using raster data (DTMs and CHMs) derived from the point clouds.

As the relative absolute positioning accuracy is substantially precise, the misalignments among adjacent lidar data products can be assumed as simple constant shifts in all three directions–horizontal (east-to-west and north-to-south) and vertical–without changes in rotation or scale. The horizontal shifts were estimated first by measuring the similarity in the overlapped areas between neighboring CHMs, given that CHMs represent normalized heights from the ground and are thus independent of vertical positioning errors.

Mutual Information[56] (MI) was selected as the similarity metric (as defined in Eq. (1)), as it is known to be robust to noise[57]. MI is an area-based approach and thus requires a defined spatial extent (i.e., window). To reduce computational cost, we did not evaluate the similarity of all possible windows across the entire overlapping area. Instead, the MI was computed for a set of randomly selected windows of size 20 × 20 m. To ensure even spatial distribution of the windows, the overlap area was divided into 50 × 50 m grids, and a single window was randomly placed within each grid. The similarity between two adjacent CHMs was then iteratively calculated by gradually shifting one CHM (the reference image) in the horizontal direction. The shift that yielded the maximum similarity was assumed to represent the horizontal misalignment between the two adjacent subregions. The final horizontal shifts were determined as the mode (most frequent value) of all estimates from the randomly selected windows.

$$\mathrm{MI}(A,\ B) = \sum_{a\in A}\sum_{b\in B} p_{AB}(a,\ b) \cdot \log\left(\frac{p_{AB}(a,\ b)}{p_A(a)p_B(b)}\right) \tag{1}$$

where A and B denote two adjacent CHMs. $p_{AB}(a,\ b)$ is the joint probability of height values between A and B, and $p_A(a)$ and $p_b(b)$ are their marginal probabilities.

The vertical shifts were estimated using the DTMs, considering that the canopy heights are less stable than the ground heights. Once the horizontal shifts were determined, the DTMs were horizontally aligned accordingly. Vertical shifts were then calculated using a histogram-based method[58] applied to the horizontally aligned DTMs. In contrast to horizontal shift estimation, which was conducted only for the randomly selected areas, vertical shifts were estimated using the entire overlapping area, as ground heights are more stable and less affected by noise.

Since the PPK correction results for Subregion 2 were stable compared to the other subregions (Table 3), we designed Subregion 2 as the reference and sequentially aligned the point cloud data of the remaining four subregions to this reference. Table 6 summarizes the estimated positional shifts relative to Subregion 2. Despite the applied PPK corrections for GNSS coordinates, around 1 meter positioning errors remained in this remote and inaccessible rainforest. These findings highlight the necessity of additional registration steps, as described in this section, to produce seamless maps across the broader region.

Finally, using the estimated shifts, the point clouds of individual subregions were spatially aligned and merged into a seamless point cloud covering the TBS region. From this merged LiDAR dataset, we regenerated the DTM and CHM for the entire region using *PDAL writers.raster* and *filters.hag_dem*, rather than simply mosaicking the existing subregional products. The DSM was likewise generated from the merged point cloud, based on the 98th percentile of point heights within each pixel, consistent with the CHM definition. Figure 2

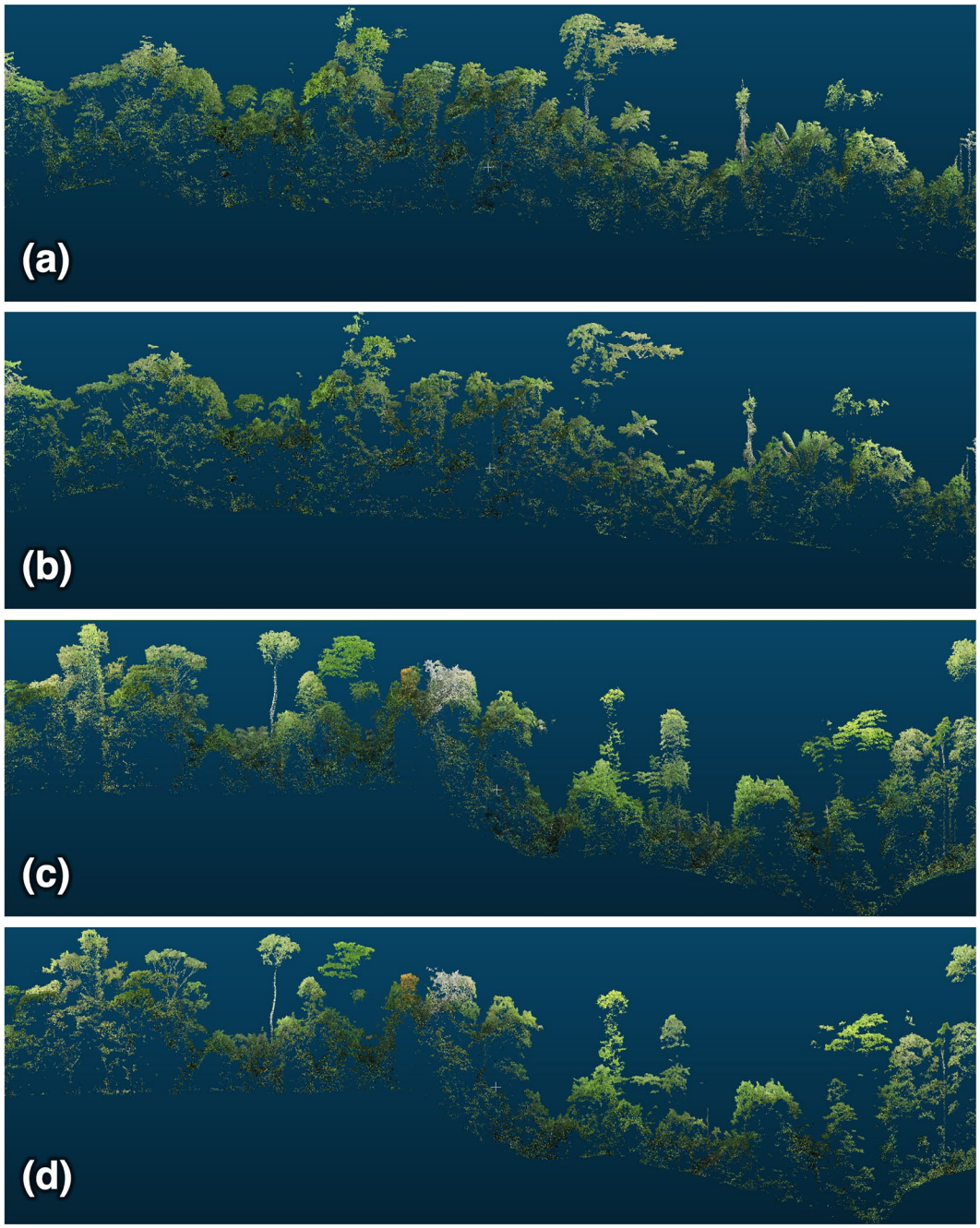

Fig. 2 Lidar registration results. (**a,b**) Cross-sections along the West–East axis before and after registration, respectively; (**c,d**) Cross-sections along the North–South axis before and after registration, respectively. Each panel shows a 200 m transect.

illustrates the registration results, comparing LiDAR products before and after alignment and highlighting the effectiveness and necessity of the registration step.

Data Records

Table 7 summarizes the UAS data products processed and published in this study. All data except for the raw UAS data (i.e., initially collected data) are provided in the WGS 84 / UTM Zone 18S coordinate system (EPSG:32718), which is consistent with the coordinate reference system commonly used in publicly available satellite datasets for monitoring the Amazon forest. Given the substantial size of the data products, this study adopts cloud-optimized formats — Cloud Optimized GeoTIFF (COG,*.tif*) and Cloud Optimized Point Cloud (COPC,*.laz*, LAS header v1.4) — to facilitate efficient data sharing. These formats enable partial access to spatial or attribute data subsets and support rapid visualization without requiring full dataset downloads. The data produced in this study are hosted in the publicly accessible Data-to-Science (D2S) STAC (SpatioTemporal Asset Catalog) repository (https://doi.org/10.4231/FV2H-VR18)[59]. STAC is a standardized metadata format designed

Spatial extent	Sensor	Item name	Spatial detail	Format	Size
Entire region	Multi-spectral	Orthomosaic	0.05	COG	65.9 GB
		Digital_Surface_Model_MS	0.10	COG	3.8 GB
	Lidar	Point_Cloud	395.51	COPC	30.9 GB
		Digital_Terrain_Model	0.50	COG	132 MB
		Digital_Surface_Model	0.25	COG	591 MB
		Canopy_Height_Model	0.25	COG	655 MB
Subregion	Multi-spectral	*Region_N*_Orthomosaic*	0.05	COG	—
		*Region_N*_DSM_MS	0.10	COG	—
	Lidar	*Region_N*_Point_Cloud*	—	COPC	—
		*Region_N*_DTM	0.50	COG	—
		*Region_N*_CHM	0.25	COG	—

Table 7. Summary of published data products. *Region_N* denotes the subregion number. COG and COPC denote Cloud Optimized GeoTIFF and Cloud Optimized Point Cloud, respectively. DTM, DSM, and CHM represent digital terrain model, digital surface model, and canopy height model, respectively. The sizes of subregion data products vary across subregions and sensors, and the point densities of the point clouds also vary. Raw data can be downloaded via the link in "*Additional Resources*" under the orthomosaics or point clouds of each subregion (denoted with *). Units of spatial details are m/pixel for raster data (COG) and points/m^2 for point clouds (COPC).

to support FAIR (Findable, Accessible, Interoperable, and Reusable) geospatial data. A single STAC asset can contain multiple data products (i.e., items); accordingly, this study published a single STAC asset titled '*Tiputini Biodiversity Station*', which aggregates all UAS data products generated from the two sensors.

Through the workflows described in the *Data processing* section, the seamless geospatial products for the entire study region were published. Two geospatial products derived from the multispectral UAS were published: a 5-cm resolution orthomosaic and a 10-cm resolution DSM, as shown in Fig. 3(a) and (b). The multispectral orthomosaic, which minimizes geometric distortions and displacements, contains eight bands in the following order: red, blue, green, multispectral green, multispectral red, multispectral red edge, multispectral near infrared, and alpha. The last alpha band defines the spatial extent of the TBS site, where the pixel value is 65,535. The four multispectral bands were radiometrically calibrated using the onboard sunlight sensor, resulting in approximate surface reflectance values, whereas the three RGB bands were color-balanced relative to one another. The DSM from the multispectral UAS represents elevation information over the TBS area, derived from a dense point cloud generated through the photogrammetric workflow. In contrast to the lidar-based DTM described below, this DSM includes the ellipsoidal heights of above-ground features (e.g., trees). It was also used as an intermediate data product in the generation of the final photogrammetric output, the orthomosaic.

From the UAS lidar data, four geospatial data products were published: 3D point cloud, 50-cm resolution DTM, and 25-cm resolution DSM and CHM (Fig. 4). The point cloud includes RGB color information collected by the RGB camera mounted on the L2 sensor as well as two extra-dimensional attributes: '*Height Above Ground*' and '*Point Source ID*', to provide as much information as possible. *Height above ground* values the elevation relative to the terrain, based on the DTM. The *point source ID* presents the subregion to which the corresponding point is assigned. In addition, ground points are denoted as class number 2 (the other points are denoted as class number 0, following the ASPRS LAS Specification[60]). The DTM represents ellipsoidal terrain heights based on the WGS84 reference system, providing detailed topographic information of the TBS region, such as hills and plains. The DTM, DSM, and CHM derived from the point cloud use a no-data value of −9999, indicating areas where no returns were recorded during the UAS surveys (e.g., over waterbodies) or outside the study region. Finally, to maximize the usability of the acquired data, this study also publishes the raw data for each subregion (initially collected by the UAS platforms, provided as *.zip* files) and the derived geospatial products for each subregion processed without PPK, as listed in Table 7.

Technical Validation

Quality of georeferencing. Due to the lack of available references for georeferencing accuracy near the TBS, it is not possible to precisely assess the absolute georeferencing accuracy of the dataset. Alternatively, the IBGE-PPP service provides expected accuracy depending on the location and time[61]. The locations of the base stations were calculated using the 'FINAL' orbit type for the outside of Brazil in the static mode. The IBGE-PPP service provides an accuracy level of up to 2.5 cm for all GNSS data collected in this study. Table 8 shows the sigma value (i.e., 95% standard deviation) of each determined coordinate during the PPK process. Although they do not indicate the absolute accuracy of the coordinates, we can estimate the internal reliability of the static processing to be less than one meter. Due to surroundings (e.g., occlusion by dense canopies, atmospheric conditions, and logging periods), the base stations located in the eastern subregions, especially Subregions 4 and 5, were less stable than those in the western subregions. For example, during the UAS mission for Subregion 5, the base station was set up right next to denser canopies and rocks due to environmental challenges. In contrast, the base stations for Subregions 1 and 2 were set up on sand hills at the riverside.

Quality of registered lidar data. Due to the highly complex canopy structure in the TBS region, manually identifying corresponding points between point clouds from different subregions is nearly

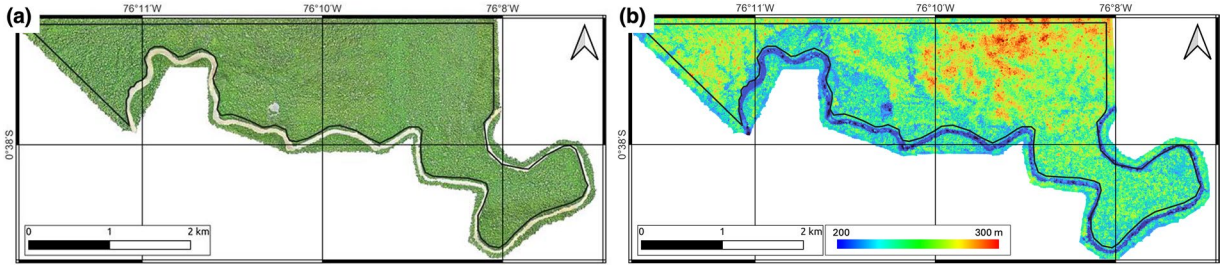

Fig. 3 Overview of multispectral data products: (**a**) RGB orthoimage and (**b**) DSM.

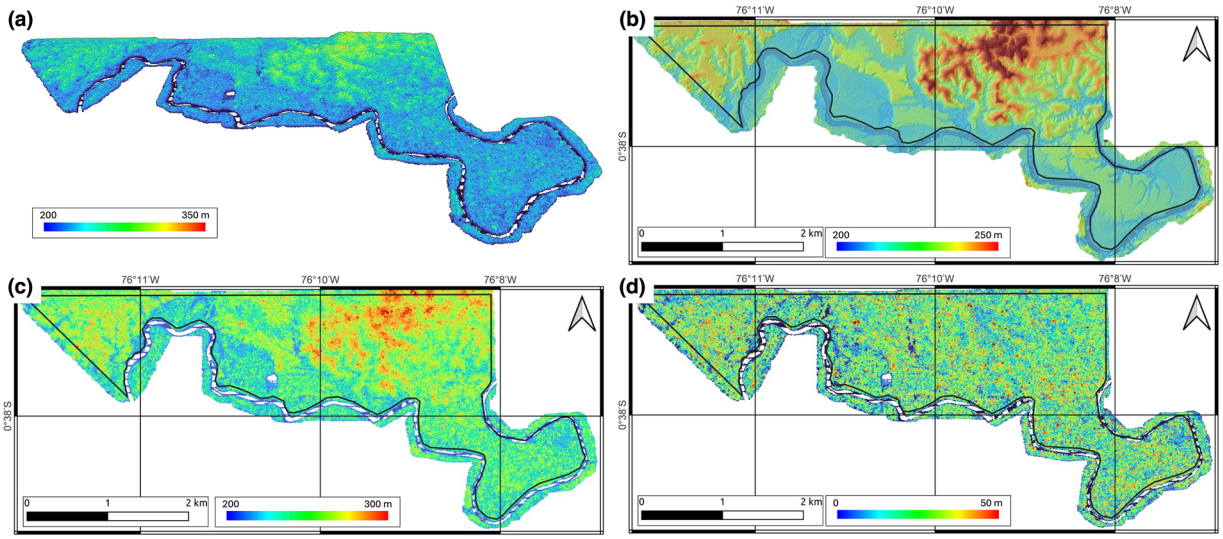

Fig. 4 Overview of lidar data products: (**a**) point cloud; (**b**) DTM with hillshade; (**c**) DSM; and (**d**) CHM. The color of (**a**–**c**) indicates ellipsoidal heights, while (**d**) indicates normalized height from the ground.

Base stations by subregion	Easting	Northing	Altitude
1	0.040	0.020	0.088
2	0.228	0.114	0.397
3 & 4	0.208	0.107	0.209
5	0.533	0.318	0.720

Table 8. The 95% standard deviation of the coordinates of each base station (unit: meter).

infeasible, making direct quantitative assessment of registration quality impractical. As an alternative, registration results were visually evaluated prior to merging the lidar products from each subregion. While direct quantitative assessment was not feasible, we additionally examined the distribution of estimated spatial shifts across randomly selected windows as an indirect measure of registration quality. In the *Data processing* section, Table 6 summarizes the standard deviations of these shifts, which correspond to several pixels in the 25-cm resolution CHMs. Figure 5 presents the histograms of the estimated shifts within each overlapping region. Overall, the shift distributions are consistent across overlaps. Some outliers appear near subregion boundaries or on the water bodies, where point cloud density is inherently low, while others likely result from transient environmental disturbances such as wind. The consistency of the estimated shifts indicates that registration was successfully achieved at a global scale across the dataset. For applications requiring sub-pixel alignment (e.g., within a 25-cm threshold), we recommend applying more precise, point cloud–level registration to each individual subregion. These individual point clouds are provided as part of the published dataset.

Spatial alignment between multispectral and lidar datasets. As described in the *Data processing* section, the multispectral raw images were processed simultaneously through the single photogrammetry pipeline using the corresponding exterior orientations refined through the PPK processing. In contrast, the lidar data were processed separately by subregion, using only refined base station coordinates, and then merged. Consequently, the outputs from these two contrasting data processing workflows were differently

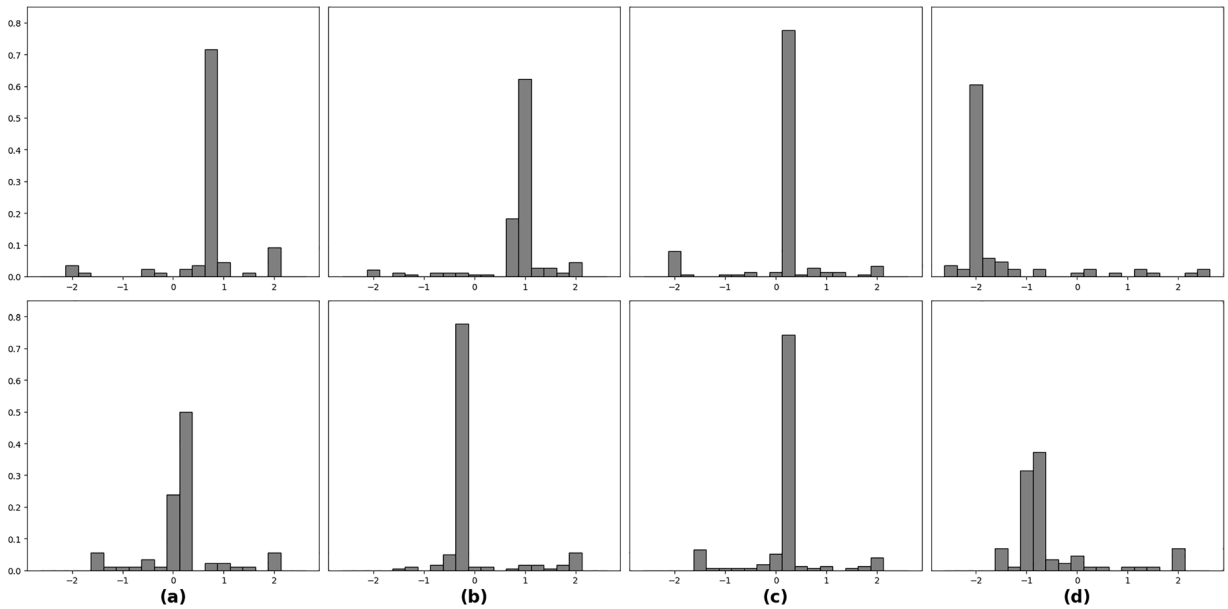

Fig. 5 Relative frequency histograms of the estimated spatial shifts derived from the mutual information-based registration approach. Results are shown for evenly distributed and randomly sampled windows in each subregion: (**a**) Subregion 1 and (**b**) Subregion 3, both aligned to reference Subregion 2; (**c**) Subregion 4 aligned to Subregion 3; and (**d**) Subregion 5 aligned to Subregion 4. The top row illustrates Easting shift distributions, and the bottom row illustrates Northing shift distributions. The x- and y-axis indicate the amount of shift (meters) and relative frequency (%), respectively.

affected by the PPK results, and their spatial alignment varied across subregions. Figure 6 shows the degree of spatial consistency between the multispectral orthomosaic and the lidar-derived CHM near the center of each subregion. The first row displays the tree crowns delineated from the orthomosaic, while the second row shows the same segments overlaid on the CHM. Misalignments between the red segments and the CHM in the second row indicate spatial discrepancies between the two products. In all subregions, subtle spatial misalignments were observed, which are attributable to inevitable errors introduced in each processing workflow. To quantify the spatial inconsistency in each subregion, the misalignments were estimated using the MI-based approach as applied to register lidar products between the multispectral-derived and the lidar-derived DSMs, as summarized in Table 9. Overall, the multispectral and lidar products exhibit subtle spatial misalignment, with mean easting and northing shifts of 0.27 and 0.31 meters, respectively. It is noteworthy that certain areas within Subregions 3 and 5 exhibit relatively larger spatial misalignments (exceeding 1 meter), highlighting the importance of recognizing misalignments when comprehensively utilizing data products from both sensors. Therefore, particular caution is advised when integrating multispectral and lidar data in these subregions. For applications requiring highly precise spatial alignment, we recommend using the datasets processed separately for each subregion and sensor without PPK correction, which are also made publicly available through the STAC asset in Table 7.

Usage Notes

Data visualization and analysis. It is recommended to use any GIS (geographic information system) software, such as QGIS (https://qgis.org/), to visualize and analyze 2D raster data (orthoimages, DSMs, DTMs, and CHMs); while CloudCompare (https://www.cloudcompare.org/) is recommended for 3D point clouds. The COG URL (Uniform Resource Locator) for each published data product is also available on the same STAC repository, so the data published in this study can be streamed rather than downloaded. In addition, researchers who analyze our data with emerging deep learning can easily stream our data instead of downloading the entire dataset to Python environments via a Python package, *d2spy* (https://pypi.org/project/d2spy/), which can interact with the D2S STAC repository.

Downstream research applications. The published UAS dataset represents one of the most detailed datasets ever collected for the Amazon rainforest in terms of spatial resolution and coverage. Unlike publicly available spaceborne data (e.g., Landsat, Sentinel, and GEDI), our comprehensive dataset provides spectral and structural information at the individual tree level, along with topographic information derived from the DTM, all captured with minimal temporal discrepancies (up to three days).

Given the vast number of trees living within the TBS site, this dataset is expected to serve as an ideal training resource for a range of deep learning applications, thereby facilitating the development of AI models specifically optimized for tropical rainforests. Potential use cases include tree canopy segmentation (e.g., DeepForest[62]), tree species identification (particularly palm trees[32] and other distinctive species), and canopy height estimation

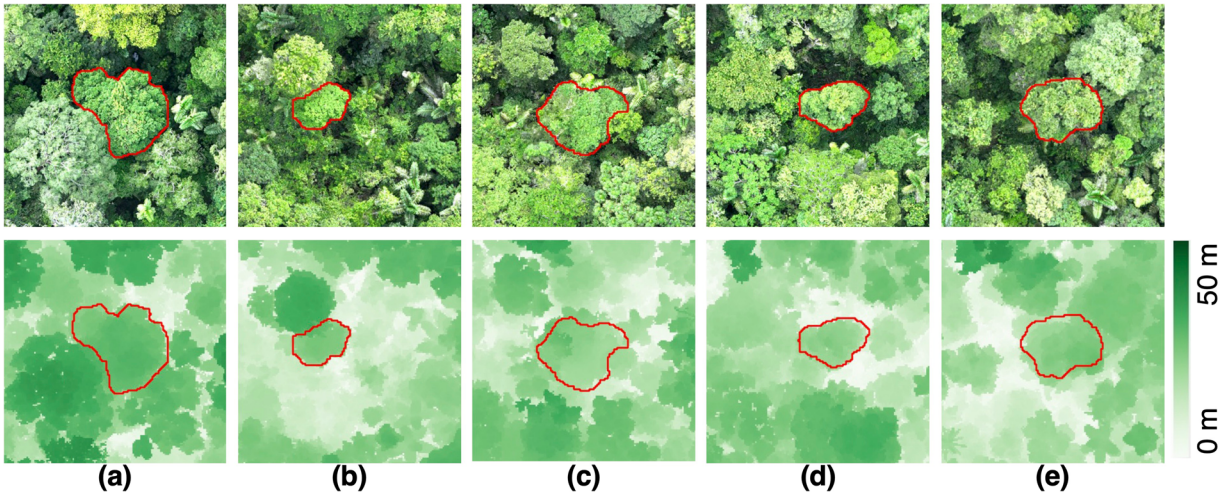

Fig. 6 Comparison of the multispectral orthomosaic and the lidar-derived CHM for five subregions: (**a**) Subregion 1, (**b**) Subregion 2, (**c**) Subregion 3, (**d**) Subregion 4, and (**e**) Subregion 5. The red polygons represent tree crowns delineated from the orthomosaic. Each image spans approximately 40 m in width.

Shift	1	2	3	4	5	Overall
Mean easting	0.52	0.00	**1.09**	0.36	−0.53	**0.27**
Mean northing	−0.13	0.06	0.13	0.33	**1.03**	**0.31**
RMS easting	0.61	0.36	**1.21**	0.94	0.64	**0.80**
RMS northing	0.35	0.34	0.49	0.51	**1.23**	**0.67**

Table 9. Spatial misalignment between multispectral and lidar DSMs based on mean and root mean squared (RMS) easting and northing shifts (unit: meter).

from RGB imagery (e.g., DepthAnything[63]). Such AI-driven approaches will enable detailed extraction of tree-level spectral and 3D structural information, facilitating biodiversity monitoring based on species[64] and enabling carbon estimation[65–68] using structural metrics such as canopy height and cover.

The dataset's fine resolution and seamless coverage of tree species composition and structural characteristics also provide an exceptionally rich environmental context for ecological research. This includes assessments of contemporary forest dynamics[69,70] (e.g., gap formation and turnover), tree mortality[71,72], and physiological variation[73] across distinct ecosystems. Moreover, results derived from our dataset can be readily incorporated into ongoing research efforts at TBS[40,41–46,74–77], including studies of phenology, tree and fauna dynamics, and forest ecosystem processes, as well as long-term monitoring programs of understory bird and canopy insect populations. These ecological insights can strategically guide targeted studies of specific ecological functional groups, population-level distributions, and endangered species.

Furthermore, by enabling detailed analyses of vegetation–topography interactions within a consistent climatic zone[78], this resource provides a valuable baseline for future multitemporal assessments of climate change impacts. Beyond advancing our understanding of ecosystem processes in the Ecuadorian Amazon rainforest, this dataset will support long-term monitoring of the major Amazonian biomes. Its value will continue to grow as natural disturbances such as storms and droughts reveal the resilience and vulnerabilities of different structural and compositional elements within the forest.

Data availability

All UAS data generated and published in this study (as listed in Table 7) are publicly available at https://doi.org/10.4231/FV2H-VR18. No additional data were used in this study.

Code availability

All custom code used to produce the published datasets is available at https://github.com/gdslab/tbs-data-code. Note that the point clouds are voluminous, so processing them requires substantial computational resources. Example scripts for *d2spy* to systematically import the published data are available in the public documentation on the official *d2spy* website (https://py.d2s.org/guides/).

Received: 15 September 2025; Accepted: 11 February 2026;
Published online: 26 February 2026

References

1. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448, https://doi.org/10.1038/nature11631 (2012).
2. ter Steege, H. *et al*. Mapping density, diversity and species-richness of the Amazon tree flora. *Commun Biol* **6**, 1130, https://doi.org/10.1038/s42003-023-05514-6 (2023).
3. Heinrich, V. H. A. *et al*. Large carbon sink potential of secondary forests in the Brazilian Amazon to mitigate climate change. *Nat Commun* **12**, 1785, https://doi.org/10.1038/s41467-021-22050-1 (2021).
4. Flores, B. M. *et al*. Critical transitions in the Amazon forest system. *Nature* **626**, 555–564, https://doi.org/10.1038/s41586-023-06970-0 (2024).
5. Matt, F. *et al*. Combating deforestation: From satellite to intervention. *Science* **360**, 1303–1305, https://doi.org/10.1126/science.aat1203 (2018).
6. Parente, L. *et al*. Quality assessment of the PRODES Cerrado deforestation data. *Remote Sens. Appl. Soc. Environ.* **21**, 100444, https://doi.org/10.1016/j.rsase.2020.100444 (2021).
7. Assunção, J., Clarissa, G. & Romero, R. DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement. *Am. Econ. J. Appl. Econ.* **15**, 125–56, https://doi.org/10.1257/app.20200196 (2023).
8. Bauer, L., Knapp, N. & Fischer, R. Mapping Amazon forest productivity by fusing GEDI lidar waveforms with an individual-based forest model. *Remote Sens.* **13**, 4540, https://doi.org/10.3390/rs13224540 (2021).
9. Milenković, M. *et al*. Assessing Amazon rainforest regrowth with GEDI and ICESat-2 data. *Sci. Remote Sens.* **5**, 100051, https://doi.org/10.1016/j.srs.2022.100051 (2022).
10. Saatchi, S. S., Soares, J. V. & Alves, D. S. Mapping deforestation and land use in Amazon rainforest by using SIR-C imagery. *Remote Sens. Environ.* **59**, 191–202, https://doi.org/10.1016/S0034-4257(96)00121-9 (1997).
11. Dell'Amore, L., Bueso-Bello, J. L., Klenk, P., Reimann, J. & Rizzoli, P. Characterization of the Amazon rainforest backscatter at X-band using TanDEM-X data. IEEE J. Sel. Top. Appl. Earth Obs. *Remote Sens.* **17**, 1673–1690, https://doi.org/10.1109/JSTARS.2023.3341234 (2023).
12. Burns, P., Hakkenberg, C. R. & Goetz, S. J. Multi-resolution gridded maps of vegetation structure from GEDI. *Sci. Data* **11**, 881, https://doi.org/10.1038/s41597-024-03668-4 (2024).
13. de Conto, T., Armston, J. & Dubayah, R. Characterizing the structural complexity of the Earth's forests with spaceborne lidar. *Nat. Commun.* **15**, 8116, https://doi.org/10.1038/s41467-024-52468-2 (2024).
14. Quegan, S. *et al*. The European Space Agency BIOMASS mission: measuring forest above-ground biomass from space. *Remote Sens. Environ.* **227**, 44–60, https://doi.org/10.1016/j.rse.2019.03.032 (2019).
15. Karaman, K., Sainte Fare Garnot, V. & Wegner, J. D. Deforestation detection in the Amazon with Sentinel-1 SAR image time series. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-1/W1-2023, 835–842, https://doi.org/10.5194/isprs-annals-X-1-W1-2023-835-2023 (2023).
16. Pfefer, A. *et al*. How Sentinel-1 time series can improve the implementation of conservation programs in Brazil. *Remote Sens. Appl. Soc. Environ.* **35**, 101241, https://doi.org/10.1016/j.rsase.2024.101241 (2024).
17. European Space Agency, *Algorithm theoretical bases documents (ATBD): Multi-temporal change detection product (MCD Amazonas)* https://drive.google.com/file/d/1WEEpEbnQBfYomVdiyu4mR31psVM07bMR/view (2022).
18. Planet. Tropical forest monitoring with NICFI satellite data. Planet University (2020–2025). https://university.planet.com/page/tfo.
19. Wagner, F. H. *et al*. Mapping tropical forest cover and deforestation with Planet NICFI satellite images and deep learning in Mato Grosso State (Brazil) from 2015 to 2021. *Remote Sens.* **15**, 521, https://doi.org/10.3390/rs15020521 (2023).
20. Dalagnol, R. *et al*. Mapping tropical forest degradation with deep learning and Planet NICFI data. *Remote Sens. Environ.* **298**, 113798, https://doi.org/10.1016/j.rse.2023.113798 (2023).
21. Yang, F. *et al*. Improved fine-scale tropical forest cover mapping for Southeast Asia using Planet NICFI and Sentinel-1 imagery. *J. Remote Sens.* **3**, 0064, https://doi.org/10.34133/remotesensing.0064 (2023).
22. Zhang, Y., Wang, X., Li, X., Du, Y. & Atkinson, P. M. Monitoring monthly tropical humid forest disturbances with Planet NICFI images in Cameroon. *Agric. For. Meteorol.* **341**, 109676, https://doi.org/10.1016/j.agrformet.2023.109676 (2023).
23. Braga, D.*et al*. Monitoring sustainable logging in an Amazon forest under management using deep learning andPlanet NICFI imagery. *Int. J. Remote Sens.* **47**, 1287–1312, https://doi.org/10.1080/01431161.2025.2603696 (2026).
24. CTrees. Amazon canopy height dataset. AWS Open Data Registry, https://registry.opendata.aws/ctrees-amazon-canopy-height/ (2023).
25. Rangel Pinagé, E. *et al*. Long-term impacts of selective logging on Amazon forest dynamics from multi-temporal airborne LiDAR. *Remote Sens.* **11**, 709, https://doi.org/10.3390/rs11060709 (2019).
26. Broadbent, E. N. *et al*. Linking rainforest ecophysiology and microclimate through fusion of airborne LiDAR and hyperspectral imagery. *Ecosphere* **5**, 57, https://doi.org/10.1890/ES13-00255.1 (2014).
27. 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, https://doi.org/10.1038/s41598-020-80809-w (2021).
28. Simonetti, A. *et al*. Canopy gaps and associated losses of biomass—combining UAV imagery and field data in a central Amazon forest. *Biogeosciences* **20**, 3651–3666, https://doi.org/10.5194/bg-20-3651-2023 (2023).
29. Messinger, M., Asner, G. P. & Silman, M. Rapid assessments of Amazon forest structure and biomass using small unmanned aerial systems. *Remote Sens.* **8**, 615, https://doi.org/10.3390/rs8080615 (2016).
30. Albuquerque, R. W. *et al*. Mapping key indicators of forest restoration in the Amazon using a low-cost drone and artificial intelligence. *Remote Sens.* **14**, 830, https://doi.org/10.3390/rs14040830 (2022).
31. Morales, G. *et al*. Automatic segmentation of Mauritia flexuosa in unmanned aerial vehicle (UAV) imagery using deep learning. *Forests* **9**, 736, https://doi.org/10.3390/f9120736 (2018).
32. Tagle Casapia, X. *et al*. Effective integration of drone technology for mapping and managing palm species in the Peruvian Amazon. *Nat. Commun.* **16**, 3764, https://doi.org/10.1038/s41467-025-58358-5 (2025).
33. Laurance, W. F. *et al*. Ecosystem decay of Amazonian forest fragments: a 22-year investigation. *Conserv. Biol.* **16**, 605–618, https://doi.org/10.1046/j.1523-1739.2002.01025.x (2002).
34. Barlow, J. *et al*. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* **535**, 144–147, https://doi.org/10.1038/nature18326 (2016).
35. Rödig, E., Cuntz, M., Heinke, J., Rammig, A. & Huth, A. Spatial heterogeneity of biomass and forest structure of the Amazon rainforest: linking remote sensing, forest modelling and field inventory. *Glob. Ecol. Biogeogr.* **26**, 1292–1302, https://doi.org/10.1111/geb.12633 (2017).
36. Tagle Casapia, M. X. *et al*. Data underlying the publication: effective integration of drone technology for mapping and managing palm species in the Peruvian Amazon. 4TU.ResearchData, Version 1 https://doi.org/10.4121/70a8cec0-dfa7-4963-ba8a-612e738ec0cb.v1 (2025).
37. Ometto, J. *et al*. L1A – discrete airborne LiDAR transects collected by EBA in the Brazilian Amazon (Mato Grosso, Amazonas and Pará) (Version v20230303). *Zenodo* https://doi.org/10.5281/zenodo.7636454 (2023).
38. Blake, J. G. *et al*. Temporal activity patterns of terrestrial mammals in lowland rainforest of eastern Ecuador. *Ecotropica* **18**, 137–146 (2012).
39. Bass, M. S. *et al*. Global conservation significance of Ecuador's Yasuní National Park. *PLoS ONE* **5**, e8767, https://doi.org/10.1371/journal.pone.0008767 (2010).

40. Blake, J. G. *et al*. Yasuní – a hotspot for jaguars Panthera onca (Carnivora: Felidae)? Camera-traps and jaguar activity at Tiputini Biodiversity Station, Ecuador. *Rev. Biol. Trop.* **62**, 689–698 (2014).
41. Bonaccorso, F. J. *et al*. Evidence for exploitative competition: comparative foraging behavior and roosting ecology of short-tailed fruit bats (Phyllostomidae). *Biotropica* **39**, 249–256, https://doi.org/10.1111/j.1744-7429.2006.00251.x (2007).
42. Erwin, T. L., Zamorano, L. S. & Geraci, C. J. Amazonian rainforests and their richness and abundance of terrestrial arthropods on the edge of extinction: abiotic–biotic players in the critical zone. In *Insect Biodiversity: Science and Society* (eds Foottit, R. G. & Adler, P. H.) 65–91, https://doi.org/10.1002/9781118945568.ch4 (Wiley, 2017).
43. Link, A., Milich, K. & Di Fiore, A. Demography and life history of a group of white-bellied spider monkeys (Ateles belzebuth) in western Amazonia. *Am. J. Primatol.* **80**, e22899, https://doi.org/10.1002/ajp.22899 (2018).
44. Blake, J. G. & Loiselle, B. A. Annual and spatial variation in composition and activity of terrestrial mammals on two replicate plots in lowland forest of eastern Ecuador. *PeerJ* **6**, e4241, https://doi.org/10.7717/peerj.4241 (2018).
45. Blake, J. G. Acoustic monitors and direct observations provide similar but distinct perspectives on bird assemblages in a lowland forest of eastern Ecuador. *PeerJ* **9**, e10565, https://doi.org/10.7717/peerj.10565 (2021).
46. Blake, J. G. & Loiselle, B. A. Sharp declines in observation and capture rates of Amazon birds in absence of human disturbance. *Glob. Ecol. Conserv.* **51**, e02902, https://doi.org/10.1016/j.gecco.2024.e02902 (2024).
47. Dubayah, R. *et al*. *GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002. NASA Land Processes Distributed Active Archive Center*, https://doi.org/10.5067/GEDI/GEDI02_A.002 (2021).
48. DJI. Zenmuse L2 User Manual, version 1.2 https://dl.djicdn.com/downloads/Zenmuse_L2/20240718/Zenmuse_L2_User_Manual_v1.2_ENI.pdf (2024).
49. DJI. D-RTK 2 Mobile Station – Product Information. https://www.dji.com/d-rtk-2/info (2025).
50. Agisoft LLC. *Agisoft Metashape user manual: Professional edition, version 2.2* https://www.agisoft.com/pdf/metashape-pro_2_2_en.pdf (2025).
51. DJI. *DJI Terra User Manual v4.0* https://dl.djicdn.com/downloads/dji-terra/20240118/DJI_Terra_User_Manual_v4.0__EN.pdf (2024).
52. PDAL Contributors. filters.hag_dem — PDAL documentation. https://pdal.io/en/stable/stages/filters.hag_dem.html (2025).
53. Lang, N., Jetz, W., Schindler, K. & Wegner, J. D. A high-resolution canopy height model of the Earth. *Nat Ecol Evol* **7**, 1778–1789, https://doi.org/10.1038/s41559-023-02206-6 (2023).
54. Wang, C. *et al*. Interpretable multi-sensor fusion of optical and SAR data for GEDI-based canopy height mapping in southeastern North Carolina. *Remote Sens.* **17**, 1536, https://doi.org/10.3390/rs17091536 (2025).
55. Besl, P. J. & McKay, N. D. Method for registration of 3-D shapes. *Proc. SPIE* **1611**, 586–606, https://doi.org/10.1117/12.57955 (1992).
56. Viola, P. J. & Wells, W. M. III Alignment by Maximization of Mutual Information. *Int. J. Comput. Vis.* **24**, 137–154, https://doi.org/10.1023/A:1007958904918 (1997).
57. Pluim, J. P. W., Maintz, J. B. A. & Viergever, M. A. Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imaging* **22**, 986–1004, https://doi.org/10.1109/TMI.2003.815867 (2003).
58. Jung, M. & Jung, J. A scalable method to improve large-scale lidar topographic differencing results. *Remote Sens.* **15**, 4289, https://doi.org/10.3390/rs15174289 (2023).
59. Jung, M., Chang, A., Jung, J., Cannon, C. & Rivas-Torres, G. Comprehensive high-quality UAS data for Amazon rainforest: Tiputini Biodiversity Station. *Purdue University Research Repository*, https://doi.org/10.4231/FV2H-VR18 (2025).
60. American Society for Photogrammetry and Remote Sensing. *LAS specification v1.4 — R15*. https://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf.
61. Klatt, C. & Johnson, P. A survey of surveys: the Canadian Spatial Reference System Precise Point Positioning Service. *Geomatica* **71**, 27–36, https://doi.org/10.5623/cig2017-103 (2017).
62. Weinstein, B. G. *et al*. DeepForest: A Python package for RGB deep learning tree crown delineation. Methods. *Ecol. Evol.* **11**, 1743–1751, https://doi.org/10.1111/2041-210X.13472 (2020).
63. Yang, L. *et al*. Depth Anything: Unleashing the power of large-scale unlabeled data. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*. 10371–10381, https://doi.org/10.48550/arXiv.2401.10891 (2024).
64. Kacic, P. & Kuenzer, C. Forest biodiversity monitoring based on remotely sensed spectral diversity—A review. *Remote Sens.* **14**, 5363, https://doi.org/10.3390/rs14215363 (2022).
65. Kachamba, D. J., Ørka, H. O., Gobakken, T., Eid, T. & Mwase, W. Biomass estimation using 3D data from unmanned aerial vehicle imagery in a tropical woodland. *Remote Sens.* **8**, 968, https://doi.org/10.3390/rs8110968 (2016).
66. Puliti, S., Breidenbach, J. & Astrup, R. Estimation of forest growing stock volume with UAV laser scanning data: can it be done without field data. *Remote Sens.* **12**, 1245, https://doi.org/10.3390/rs12081245 (2020).
67. Navarro, A. *et al*. The application of unmanned aerial vehicles (UAVs) to estimate above-ground biomass of mangrove ecosystems. *Remote Sens. Environ.* **242**, 111747, https://doi.org/10.1016/j.rse.2020.111747 (2020).
68. Maesano, M. *et al*. Above-ground biomass estimation from UAV high-resolution RGB images and LiDAR data in a pine forest in Southern Italy. *iForest* **15**, 451–457, https://doi.org/10.3832/ifor3781-015 (2022).
69. Nyamgeroh, B. B., Groen, T. A., Weir, M. J. C., Dimov, P. & Zlatanov, T. Detection of forest canopy gaps from very high-resolution aerial images. *Ecol. Indic.* **95**, 629–636, https://doi.org/10.1016/j.ecolind.2018.08.011 (2018).
70. Dobrowolska, D., Piasecka, Ż. & Stereńczak, K. Canopy gap characteristics and regeneration patterns in the Białowieża Forest based on remote sensing data and field measurements. *For. Ecol. Manage.* **511**, 120123, https://doi.org/10.1016/j.foreco.2022.120123 (2022).
71. Brodrick, P. G. & Asner, G. P. Remotely sensed predictors of conifer tree mortality during severe drought. *Environ. Res. Lett.* **12**, 115013, https://doi.org/10.1088/1748-9326/aa8f65 (2017).
72. Bergmüller, K. O. & Vanderwel, M. C. Predicting tree mortality using spectral indices derived from multispectral UAV imagery. *Remote Sens.* **14**, 2195, https://doi.org/10.3390/rs14092195 (2022).
73. Fraser, B. T. & Congalton, R. G. Monitoring fine-scale forest health using unmanned aerial systems (UAS) multispectral models. *Remote Sens.* **13**, 4873, https://doi.org/10.3390/rs13234873 (2021).
74. Blake, J. G. & Loiselle, B. A. Enigmatic declines in bird numbers in lowland forest of eastern Ecuador may be a consequence of climate change. *PeerJ* **3**, e1177, https://doi.org/10.7717/peerj.1177 (2015).
75. Pitman, N. C. *et al*. Distribution and abundance of tree species in swamp forests of Amazonian Ecuador. *Ecography* **37**, 902–915, https://doi.org/10.1111/ecog.00774 (2014).
76. Holbrook, K. M. & Loiselle, B.A. Using toucan-generated dispersal models to estimate seed dispersal in Amazonian Ecuador. In: *Seed Dispersal: Theory and its Application in a Changing World*, 300–321, https://doi.org/10.1079/9781845931650.030 (CABI, 2007).
77. Voyt, R., Di Fiore, S. & Di Fiore, A. Observations of golden-mantled tamarins (*Leontocebus tripartitus*) in Amazonian Ecuador. *Neotrop. Primates* **28**, 73–77, https://doi.org/10.62015/np.2022.v28.229 (2022).
78. Balzotti, C. S. *et al*. Topographic distributions of emergent trees in tropical forests of the Osa Peninsula, Costa Rica. *Ecography* **40**, 829–839, https://doi.org/10.1111/ecog.02062 (2017).
79. NASA/METI/AIST/Japan Spacesystems & U.S./Japan ASTER Science Team. *ASTER Global Digital Elevation Model V003. NASA EOSDIS Land Processes DAAC*, https://doi.org/10.5067/ASTER/ASTGTM.003 (2019).
80. Eva, H. & Huber, O. *A Proposal for defining the geographical boundaries of Amazonia. EUR 21808 EN*. (Office for Official Publications of the European Communities, 2005).

Acknowledgements
This research was partially supported by the Institute for Digital Forestry and the Institute for Plant Sciences at Purdue University, XPrize Rainforest, and the Alana Foundation. We want to thank all the staff at the Tiputini Biodiversity Station who allowed us to perform this work, with special thanks to Ciara Wirth, Tomi Sugahara, Carla Larrea, Carlos Valle, Jose Macanilla, Froilan Macanilla, Ramiro Sanmiguel, Santiago Shiguango, Nicolás Zapata, and RAs who helped with logistics and flights. All the flights and research-related activities were developed under Ecuadorian regulations and according to permit number MAATE-ARSFC-2025-0319 provided by the Ecuadorian Ministry of Environment.

Author contributions
Conceptualization, A.C., C.C., G.R.T., and J.J.; data acquisition, A.C., C.C., G.R.T., and J.J.; data processing, M.J., A.C., and J.J.; technical validation, M.J. and A.C.; formal analysis, M.J., A.C., and J.J.; writing—original draft preparation, M.J. and A.C.; writing—review and editing, all the authors; visualization, M.J. All authors have read and agreed to the published version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.