



OPEN

DATA DESCRIPTOR

# LiDAR-based reference aboveground biomass maps for tropical forests of South Asia and Central Africa

Suraj Reddy Rodda *et al.*<sup>#</sup>

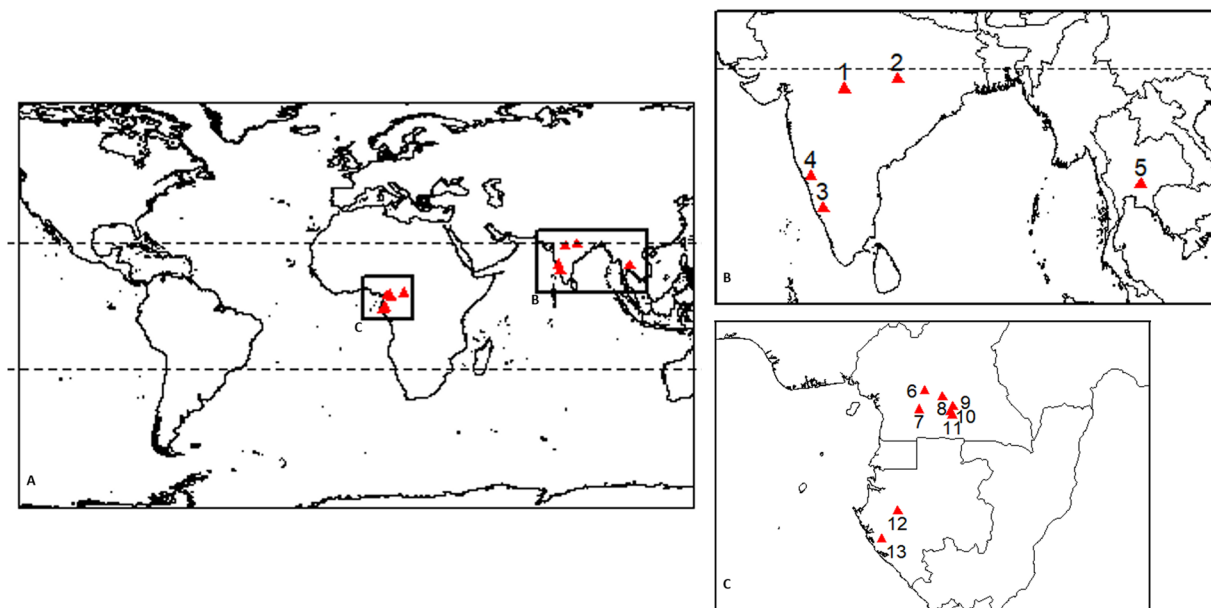
Accurate mapping and monitoring of tropical forests aboveground biomass (AGB) is crucial to design effective carbon emission reduction strategies and improving our understanding of Earth's carbon cycle. However, existing large-scale maps of tropical forest AGB generated through combinations of Earth Observation (EO) and forest inventory data show markedly divergent estimates, even after accounting for reported uncertainties. To address this, a network of high-quality reference data is needed to calibrate and validate mapping algorithms. This study aims to generate reference AGB datasets using field inventory plots and airborne LiDAR data for eight sites in Central Africa and five sites in South Asia, two regions largely underrepresented in global reference AGB datasets. The study provides access to these reference AGB maps, including uncertainty maps, at 100 m and 40 m spatial resolutions covering a total LiDAR footprint of 1,11,650 ha [ranging from 150 to 40,000 ha at site level]. These maps serve as calibration/validation datasets to improve the accuracy and reliability of AGB mapping for current and upcoming EO missions (*viz.*, GEDI, BIOMASS, and NISAR).

## Background & Summary

Tropical forests play a vital role in the Earth's carbon cycle and contribute largely to uncertainties in the global carbon budget<sup>1</sup>. Methods to accurately map and monitor tropical forest carbon – or aboveground biomass (AGB) – are thus urgently needed to improve Earth system models and to help design carbon emission mitigation strategies in the context of Reducing Emissions from Deforestation and forest Degradation (REDD+)<sup>2,3</sup>. In the last decade, spaceborne Earth Observation (EO) data in combination with forest inventory measurements have been extensively used to generate spatially continuous AGB maps at pan-tropical scale using different modelling strategies<sup>3–6</sup>. However, existing broad-scale maps show divergent estimates among themselves and differ from field-derived forest AGB stocks at different spatial scales<sup>4,5,7</sup>, indicating the presence of high uncertainties in prediction maps. To improve the accuracy and reliability of AGB maps over the tropics, several ongoing and upcoming EO missions (NASA's GEDI, ESA's BIOMASS, NASA-ISRO's NISAR and JAXA's ALOS-4 missions, notably) have been specifically designed to collect satellite data sensitive to forest structure, hence to forest AGB<sup>6,8–10</sup>. While these new spaceborne datasets will undoubtedly revolutionise broad-scale forest AGB mapping, a network of high-quality reference data is needed to calibrate and validate the mapping algorithms<sup>11,12</sup>. Besides, using the same sets of reference data across different EO missions would vastly improve the comparability and confidence in the derived AGB maps, enabling their use in a wide range of science, policy, and management applications<sup>13</sup>.

Generating reference AGB observations over a given area is challenging since forest AGB is not directly measured through destructive sampling, but instead estimated from tree inventories and a series of statistical models propagating substantial uncertainties<sup>5</sup>. It is therefore required to reduce as far as possible and quantify the uncertainty on reference AGB predictions. In this context, the Committee on Earth Observation Satellite (CEOS) has established a good practice protocol for generating reference AGB dataset, to facilitate the

<sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.



**Fig. 1** (A) Overview map showing the locations of sampling sites ( $n = 13$ ) used in the current study. Outlined regions are expanded in (B): South Asian region and in (C): Central African region. Sampling site names and descriptions associated with site numbers are provided in Table 1.

production and warrant the consistency of next-generation biomass products<sup>14</sup>. The protocol suggests developing reference AGB maps over sizable areas using local forest sample plots and LiDAR data acquired using aerial platforms (hereafter airborne LiDAR). Airborne LiDAR data is currently the most informative data type for characterizing forest structure and deriving AGB maps at the landscape scale, provided they are adequately calibrated with respect to local environmental gradients and forest structural and species variability<sup>15</sup>. Besides, the high spatial resolution of airborne LiDAR data (or of derived AGB predictions) can easily be aggregated to coarser resolutions, thus bridging the scale gap between field data and the resolution of upcoming EO sensors (e.g. 100 m for NISAR, 200 m for BIOMASS).

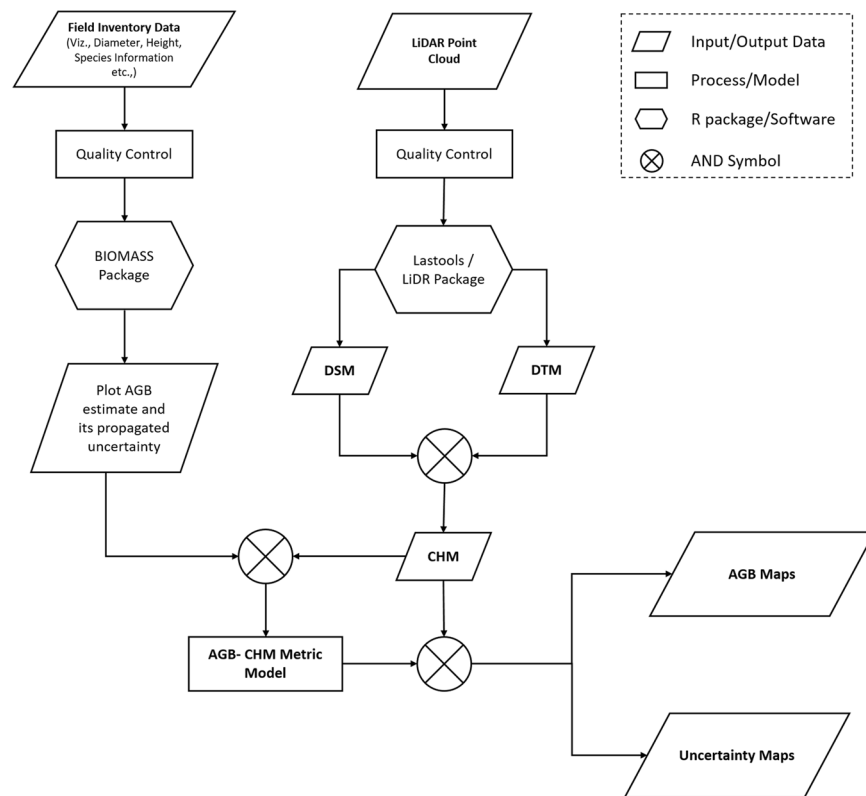
The establishment and long-term maintenance of a network of reference forest AGB observatories across the tropics entails a myriad of challenges, particularly concerning the representativeness of the network<sup>12</sup>. Ideally, the network should be relatively evenly distributed in space and cover the main environmental gradients. While scientific discussions on site selection are on-going<sup>12</sup>, the Global Ecosystem Dynamics Investigation (GEDI) sensor on-board the International Space Station has already acquired data for a longer period than its initially projected lifetime. Data users would benefit from open-access reference AGB data, particularly in Asia where large geographic regions are not represented in the calibration/validation dataset of GEDI biomass mapping algorithm<sup>16,17</sup>. Besides the notion of spatial representativeness, hurdles related to the temporal mismatch between reference AGB and EO data should not be neglected. Rapid growth in regenerating forests or forest clearing/degradation – which notably characterise rural landscapes around central African cities, where slash-and-burn agriculture induces relatively fast dynamics – could rapidly make tens of thousands of GEDI data shots unusable. We argue that airborne LiDAR data acquired during GEDI lifetime over rapidly changing landscapes are invaluable and should be utilized to improve GEDI biomass mapping algorithms, notably on the lower-end of the forest biomass gradient to best capture forest degradation and regeneration gradients.

In this context, we aim to generate reference biomass datasets over the tropics for eight sites in Central Africa and five sites in South Asia (Fig. 1) by calibrating airborne LiDAR data with locally established field plots. This paper briefly describes (1) the details of the study sites and the datasets used, (2) the methodology used to generate the reference AGB maps (Fig. 2), and (3) the Monte Carlo simulation workflow used to generate uncertainty maps along with each reference AGB map. Finally, this paper provides access to these reference AGB datasets generated at 100 m and 40 m spatial resolutions over airborne LiDAR footprints ranging from 100 to 40,000 ha.

## Methods

**Sampling sites and associated inventory and LiDAR datasets.** We compiled co-located forest inventory and LiDAR datasets from 13 sampling sites in Central Africa and South Asia encompassing an array of abiotic conditions, forest types and structures (Fig. 1; Tables 1 and 2). Forest inventories were carried out at each site, and LiDAR datasets were obtained with an absolute temporal difference of  $2.2 \pm 1.9$  years (range: 0–6.2 years) from the field measurements.

Forest inventories were conducted by different teams but followed similar protocols. In each plot, the diameter at breast height (DBH or referred to as  $D$  in this study, with  $D \geq 10$  cm) and the taxonomic identification of each tree were recorded. Tree relative coordinates within the plots were measured either at the individual or at the  $20 \times 20$  m quadrat level. For a subsample of trees within the plots, tree height ( $H$ ) was measured using a laser



**Fig. 2** Flow chart depicting workflow of the data analysis procedure to generate reference AGB datasets.

rangefinder device. Finally, plot geographic coordinates were determined using points measured every 20 m along the plot borders using a combination of differential GPS measurement system and electronic total station (in Asia) or a regular GPS system (in Africa), to warrant an accurate link between ground and remote-sensing data. The complete inventory dataset includes information on D and H measurements for respectively 97,251 and 13,303 trees, and identification rates of 89% at the species level and 92% at the genus level (8% of the trees were left unidentified). The number, size and layout of the inventory plots are uneven across sampling sites with, e.g., a single large 25-ha plot in the Forest-Geo “Rabi” site, a large 30-ha plot and smaller plots of 1-ha and 0.48-ha in the “Khao Yai” site, or a varying number of scattered 1-ha plots (ranging from 2 to 16 in the “Atout” and “Achanakmar” site, respectively). In general, the inventoried extent per site is smaller in Africa ( $9 \pm 8$  hectares) than in Asia ( $27 \pm 13$  hectares). For a breakdown of plot number, size, tree measurements and identification rates per sampling site, please refer to Table 3.

The LiDAR data at each sampling site were acquired between 2012 and 2022 using either aircraft or unmanned aerial vehicles (UAV, Table 1).

**Inventory data processing: computation of reference AGB predictions.** Forest inventories were first split into 1-ha (i.e.  $100 \times 100$  m) and 0.16-ha (i.e.  $40 \times 40$  m) plots, using information on tree location recorded in the field (i.e. either individual tree location or quadrat number). The two plot sizes correspond to the two mapping resolutions considered in this study. The 40-m resolution was chosen to account for plots where individual tree locations were only recorded at  $20 \times 20$  m quadrat-level. In cases where the original plot size was not a multiple of the desired output size (typically when splitting  $100 \times 100$  m plots into  $40 \times 40$  m plots), subplots of the desired outputs size were selected at the edges of the original plot, thus leaving-out parts of the original inventory dataset (20 m wide bands in the center as per the previous example). The resulting number of 1-ha and 0.16-ha plots compiled at each sampling site is provided in Table 3.

Subsequently, the BIOMASS R package<sup>18</sup> (version 2.1.8) within the R statistical platform (version 4.1.3) was used to compute reference AGB predictions for forest inventory plots at the two spatial resolutions (1-ha and 0.16-ha). To that end, we differentiated sites with a cumulated forest inventory area of 10 ha or more (i.e., 8 out of 13 sites, Tables 1 and 3) from those with less than 10 ha of cumulated forest inventory area (i.e., 5 sites). In the former case, we developed site-specific tree height-diameter (H-D) allometric models using second-order polynomials on log-transformed data (modelHD function in the BIOMASS package) and these models were used to predict the height of trees without H measurements in each respective site. In the latter case, which pertained to sites located in moist dense forests of Cameroon (SiteIDs 7 to 11 in Table 1), all inventory data from that country and biome were pooled into a single training dataset and the same H-D modelling procedure was applied. The resulting country- and biome-specific model was then used for predicting tree height at those sites. The H-D model coefficients for these site-level and Cameroon level model are presented in Table 4. Next, a wood density (WD) estimate was attributed to each tree based on its taxonomic identification using the getWoodDensity function.

Sno	Site name	Region	PFT	INV <sub>Date</sub>	Area <sub>INV</sub>	N <sub>range</sub>	BA <sub>range</sub>	LiDAR <sub>Date</sub>	LiDAR <sub>Area</sub>	Platform
1	Betul	S-As	DBT, GSW	02-2016 to 03-2016	46	150–437	7.2–20.2	04-2014	10000	Aircraft
2	Achanakmar	S-As	EBT, DBT	12-2017 to 01-2018	16	186–509	10.0–31.3	10-2015	11000	Aircraft
3	Yellapur	S-As	EBT, DBT, GSW	01-2015 & 01-2016	15	140–749	20.3–43.6	11-2017	40000	Aircraft
4	Uppangala	S-As	EBT	03-2013 & 03-2014	23	370–838	10.3–62.4	11-2013	900	Aircraft
5	Khaoyai	S-As	EBT	11-2015 to 04-2018	34.84	394–1015	20.8–38.1	04-2017	4500	Aircraft
6	Nachtigal	C-Af	DBT, GSW	02-2018 to 05-2018	13.16	216–538	23.4–35.0	01/2012	25000	Aircraft
7	Mbalmayo	C-Af	EBT	04-2021 to 07-2021	9	442–596	20.1–32.5	02-2023	400	UAV
8	Atout	C-Af	EBT	07-2017	2	475–574	30.8–32.0	11-2021	150	UAV
9	Kompia	C-Af	EBT	12-2018	2	423–519	29.0–32.3	12-2019	400	UAV
10	Somalomo	C-Af	EBT	09-2022	8	325–508	25.4–33.6	04-2022	800	UAV
11	Bouamir	C-Af	EBT	12-2018	4	362–552	24.9–35.8	12/2018	1000	UAV
12	Mabounié	C-Af	EBT	04-2012 to 10-2012	11	222–492	17.7–31.4	11/2007	16000	Aircraft
13	Rabi	C-Af	EBT	2015	25	392–532	20.5–36.6	10/2012	1000	Aircraft

**Table 1.** Sampling site details on forest types, inventory statistics and characteristics of the LiDAR acquisitions. Area<sub>INV</sub> indicates the total area of field inventories, LiDAR<sub>Date</sub> indicates the month and year of acquisition of LiDAR data and LiDAR<sub>Area</sub> indicates the total area covered by LiDAR data over the site. N<sub>range</sub> and BA<sub>range</sub> indicate the range in number of trees and basal area per hectare across the inventoried area, respectively. The associated plant functional types (PFT's) for each site are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type product (MCD12Q1) which follows Land Cover Type 5 Classification Scheme; a similar strategy is adopted by GEDI Mission.

Site No.	Site name	Region	Elevation	Slope	MAT	MAP
1	Betul	S-As	487 ± 48	7.8 ± 5.7	25.6 ± 0.3	1266 ± 27
2	Achanakmar	S-As	813 ± 165	8.8 ± 6.8	23.3 ± 0.7	1328 ± 42
3	Yellapur	S-As	459 ± 110	8.7 ± 7.8	24.4 ± 0.5	2383 ± 421
4	Uppangala	S-As	377 ± 122	18.6 ± 8.5	25.1 ± 0.5	3789 ± 63
5	Khaoyai	S-As	757 ± 34	6.8 ± 4.2	23.3 ± 0.1	1127 ± 7
6	Nachtigal	C-Af	527 ± 50	4.5 ± 3.4	24.4 ± 0.5	1588 ± 17
7	Mbalmayo	C-Af	662 ± 11	5.2 ± 3.3	23.6 ± 0.1	1706 ± 1
8	Atout	C-Af	715 ± 11	5.2 ± 3.0	23.7 ± 0.1	1572 ± 1
9	Kompia	C-Af	705 ± 20	5.7 ± 3.3	23.5 ± 0.1	1606 ± 2
10	Somalomo	C-Af	656 ± 13	4.7 ± 3.0	23.7 ± 0.1	1602 ± 3
11	Bouamir	C-Af	696 ± 14	5.3 ± 3.1	23.5 ± 0.1	1616 ± 2
12	Mabounié	C-Af	88 ± 31	7.3 ± 4.6	26.1 ± 0.1	2034 ± 16
13	Rabi	C-Af	68 ± 15	6.3 ± 3.8	25.6 ± 0.1	1826 ± 4

**Table 2.** Environmental conditions across sampling sites. Over the LiDAR acquisition area, the statistics (mean ± standard deviation) of elevation and slope are computed using SRTM at 30-m spatial resolution (V3 product), Mean Annual Temperature (MAT) and Mean Annual Precipitation (MAP) are computed using WorldClim Version 2.1 data.

Considering that tree AGB prediction is associated with various sources of uncertainty (including measurement errors of the independent variables such as tree diameter, height, and wood density, as well as prediction errors of the H-D models and the AGB allometric model)<sup>5,14</sup>, we used a Monte Carlo approach for uncertainty propagation. Specifically, we employed the AGBmonteCarlo function of the BIOMASS package<sup>18</sup>, which allows propagating the above-mentioned sources of uncertainty and outputs 1000 tree-level and subsequently plot-level AGB predictions. Tree AGB predictions were made using the pantropical AGB allometric model (i.e., Equation-4 in Chave *et al.*<sup>19</sup>). For each plot, the 1000 AGB predictions were (i) averaged to obtain a reference plot-level AGB density (hereafter AGB<sub>REF</sub>) for the development of LiDAR-AGB models and (ii) used for the propagation of uncertainties to the final AGB maps (see section “mapping forest AGB and prediction uncertainty”).

**LiDAR data processing: computation of canopy height metrics.** LiDAR data from African and Asian sites were processed using LAStools (version 201124) and the lidR R package (version 4.0.1), respectively. The same processing chain was applied to generate the canopy metrics in both cases. First, a digital surface model (DSM) free of pits and spikes was generated at a 1-m resolution by interpolating the highest points on a 1-m grid. Second, a ground point classification was performed on the point cloud and a digital terrain model (DTM) was interpolated from ground-points. The canopy height model (CHM) was then derived by subtracting the DTM from the DSM. Finally, the 1-m CHM was used to compute 15 canopy metrics for each plot (Table 5) as candidate predictors of forest AGB.

Sno	Site	Field Plots description	N <sub>1ha</sub>	N <sub>0.16ha</sub>	N <sub>Trees</sub>	N <sub>Tree_hts</sub>	Species [%]	Genus [%]	Family [%]
1	Betul	Single plot of 34-ha and 12 distributed plots of 1-ha	34	227	15672	677	95%	96%	96%
2	Achanakmar	16 distributed plots of 1-ha	16	64	5750	1546	88%	88%	88%
3	Yellapur	15 distributed plots of 1-ha	15	60	8519	4932	97%	97%	97%
4	Uppangala	Single plot of 10-ha and 13 distributed plots of 1-ha	23	112	14967	1729	61%	61%	61%
5	Khaoyai	Single plot of 30-ha, 1 plot of 1-ha and 8 distributed plots of 0.48 ha	31	192	19461	517	100%	100%	100%
6	Nachtigal	11 distributed plots of 1-ha & 18 distributed plots of 0.16 ha	11	62	4315	703	86%	96%	97%
7	Mbalmayo	9 distributed plots of 1-ha	9	36	4378	532	74%	97%	99%
8	Atout	2 distributed plots of 1-ha	2	8	1049		96%	99%	100%
9	Kompia	2 distributed plots of 1-ha	2	8	942	49	95%	98%	98%
10	Somalomo	8 distributed plots of 1-ha	8	32	4564	215	93%	99%	99%
11	Bouamir	4 distributed plots of 1-ha	4	16	1784	171	92%	98%	98%
12	Mabounié	11 distributed plots of 1-ha	11	44	4425	570	93%	100%	100%
13	Rabi	Single plot of 25-ha	25	144	11425	1662	94%	100%	100%

**Table 3.** Site-level details on field plot layout description, the number of compiled plots at 1-ha (N<sub>1ha</sub>) and 0.16-ha (N<sub>0.16ha</sub>), number of total trees across all plots (N<sub>Trees</sub>), number of trees measured for height (N<sub>Tree\_hts</sub>). Species, Genus and Family (%) stands for the identification rate (in %) at the given taxonomic level.

Sno	Site	a	b	c	sigma [ε]	R <sup>2</sup>	RMSE [m]
1	Betul	−1.163	1.857	−0.194	0.148	0.71	2.37
2	Achanakmar	−1.314	1.770	−0.159	0.250	0.63	4.06
3	Yellapur	0.563	0.851	−0.050	0.231	0.60	3.75
4	Uppangala	0.164	1.123	−0.077	0.249	0.67	4.33
5	Khaoyai	1.094	0.479	0.025	0.329	0.66	4.90
6	Nachtigal	−0.446	1.457	−0.127	0.223	0.72	5.06
7	Mbalmayo	−0.081	1.286	−0.104	0.243	0.71	4.61
8	Atout						
9	Kompia						
10	Somalomo						
11	Bouamir						
12	Mabounié	0.283	1.102	−0.083	0.236	0.66	5.47
13	Rabi	1.234	0.614	−0.022	0.224	0.58	4.46

**Table 4.** H-D Model coefficients (a, b, c) of the 2<sup>nd</sup> order log-log polynomial model form ( $\ln(H) = a + b \times \ln(D) + c \times \ln(D^2) + \varepsilon$ ), where H is the height of the tree and D is the tree diameter. ε is the normally distributed error to be used during back-transformation for Baskerville correction.

**Specification of a general AGB model form.** While LiDAR-based AGB mapping models were trained at the site or regional level (for some Cameroon sites), to minimise local bias in model predictions<sup>14,20</sup>, we privileged the use of (i) a single AGB model form across all sites to facilitate sites inter-comparison and the subsequent use of AGB predictions for spaceborne products calibration/validation and (ii) a simple, parametric modelling approach, keeping the number of predictors to a minimum to avoid overfitting and multicollinearity issues. To specify the AGB model form, we used linear mixed-effects models to identify the most predictive LiDAR-derived canopy height metrics (LCMs) on AGB<sub>REF</sub> variation while accounting for the hierarchical spatial structure of the data. In practice, we built 15 linear mixed-effects models (one for each LCM) on the log-transformed variables of AGB<sub>REF</sub> and LCM (Eq. 1):

$$\log(AGB_{REF}) = a + b \times \log(LCM) + RE_{site} + \varepsilon \quad (1)$$

where *a* and *b* are the model's coefficients, LCM represents the Lidar-derived Canopy Metric, AGB<sub>REF</sub> corresponds to the field-derived AGB prediction at a given spatial resolution (i.e. 0.16- or 1-ha), RE<sub>site</sub> denotes the random site effect used in linear mixed-effects modelling and ε is the error term, assumed to follow a normally distribution with a mean of zero and a standard error σ. Based on the AIC criterion, the meanTCH metric (i.e. the mean of all CHM values in the plot area) emerged as the best predictor of AGB<sub>REF</sub> variation at both 1-ha and 0.16-ha spatial resolutions (Table 6).

A similar procedure was run on AGB<sub>REF</sub> prediction models combining each pair of LCMs rather than a single predictor. At both spatial resolutions, the best two-predictor model resulted in a modest improvement in relative RMSE (i.e., <0.2%, Table 7) compared to the model based on meanTCH only. The latter model form was thus selected for biomass mapping. In line with the H:D modelling procedure, LiDAR-based AGB mapping models were either trained at the site-level (for sites with a cumulated forest inventory area of 10 ha more) or on a pooled training dataset containing all inventory data from Cameroonian moist dense forests (for sites with a cumulated

LiDAR Canopy Metric (LCM)	Description
H40	Percentile of CHM values (ex. H98 for the 98th percentile, in m)
H50	
H60	
H70	
H80	
H90	
H98	
meanTCH or meanH	
sdH	Standard deviation of CHM values (in m)
CV	Coefficient of variation of CHM values (meanTCH divided by sdH)
QMCH	Quadratic mean of CHM values
CCF2	Percentage of CHM values above 2, 5 and 10 m (in %)
CCF5	
CCF10	
rumple	Roughness of CHM surface (rumple_index function in lidR R package)

**Table 5.** List of canopy metrics derived from LiDAR-derived CHMs over forest plots extent.

Plot Size (1-ha)					Plot Size (0.16-ha)				
LCM	AIC	R <sup>2</sup>	RMSE [Mg ha <sup>-1</sup> ]	RMSE [%]	LCM	AIC	R <sup>2</sup>	RMSE [Mg ha <sup>-1</sup> ]	RMSE [%]
meanH	-180.42	0.89	43.62	15.2	meanH	149.94	0.71	88.55	32.4
RH60	-169.22	0.88	44.75	15.6	QMCH_chm	194.11	0.70	89.31	32.7
RH50	-167.22	0.88	45.27	15.8	RH90	379.89	0.65	96.81	35.4
RH70	-151.53	0.87	45.83	16.0	RH98	642.43	0.59	105.29	38.5
QMCH_chm	-148.49	0.88	45.30	15.8	RH40	869.25	0.60	106.49	39.0
RH80	-133.52	0.87	47.14	16.5	RH50	889.42	0.59	106.84	39.1
RH40	-114.58	0.86	48.46	16.9	CCF10	928.45	0.50	116.11	42.5
RH90	-102.74	0.85	50.54	17.7	RH60	975.37	0.59	106.68	39.0
RH98	-23.33	0.79	59.26	20.7	RH80	1033.34	0.62	103.08	37.7
CCF10	15.74	0.72	70.11	24.5	RH70	1036.05	0.59	106.41	38.9
CCF5	66.87	0.63	79.56	27.8	CCF5	1101.04	0.46	120.43	44.1
CV	85.22	0.61	81.18	28.4	CCF2	1105.63	0.45	121.80	44.6
CCF2	90.21	0.60	82.20	28.7	CV	1265.00	0.41	132.29	48.4
sdH	114.75	0.62	79.52	27.8	sdH	1365.98	0.41	128.62	47.1
rumple	131.76	0.58	83.77	29.3	rumple	1473.62	0.39	130.34	47.7

**Table 6.** LiDAR-AGB Linear Mixed Effects Model performance statistics at 1-ha and 0.16-ha plot sizes. The table is sorted in ascending order based on the column “AIC” (Akaike information criterion) when the respective LiDAR Canopy Metric (LCM) is used for Eq. 1. R<sup>2</sup> and RMSEs (in Mg ha<sup>-1</sup> and in %) are computed on back-transformed predictions.

forest inventory area smaller than 10 ha), henceforth referred to as the “regional” AGB model. It is noteworthy that including sites as an additional fixed-effect covariate in the regional model did not yield significant effects for this variable at a 5% risk (neither in terms of site-level intercepts nor in terms of interactions between sites and the meanTCH predictor), suggesting a minimal site effect on the regional model’s predictions, if any.

The coefficients and calibration statistics of LiDAR-based AGB mapping models are provided in Table 8, while Fig. 3 shows scatterplots of ‘reference’ against predicted AGB values.

**Mapping forest AGB and prediction uncertainty.** We mapped forest AGB and prediction uncertainty over the extent of airborne LiDAR data at each site using a Monte Carlo approach similar to that used to compute plot-level AGB<sub>REF</sub>. More specifically, we used the 1000 plot-level AGB predictions generated at the first modelling level (i.e., from tree to plot) to build 1000 LiDAR-based models per site (or at “regional” level for Cameroonian sites with less than 10 ha of cumulated forest inventory area). At the second modelling level (i.e., from plot to landscape), pixel AGB predictions derived from LiDAR-based models suffer from additional uncertainty associated to the LiDAR-based models themselves. To propagate this additional uncertainty, we mimicked the procedure used in BIOMASS to propagate the uncertainty associated to the tree-level AGB allometric model (see Appendix S1 of Réjou-Méchain *et al.*<sup>18</sup> for codes and details), which entailed using a Markov chain Monte Carlo algorithm to infer the uncertainty on Lidar-based models’ parameters (i.e., models’ coefficients and associated RSE). The Markov chain outputted 1000 sets of model parameters per model. For each of the 1000 LiDAR-based model at each site, we then (1) randomly selected a set of parameters among the 1000 available sets, (2) used the



Plot Size (1-ha)					Plot Size (0.16-ha)				
LCMs	AIC	R <sup>2</sup>	RMSE [Mg ha <sup>-1</sup> ]	RMSE [%]	LCMs	AIC	R <sup>2</sup>	RMSE [Mg ha <sup>-1</sup> ]	RMSE [%]
RH50, RH90	-192.69	0.89	43.75	15.3	meanH, CCF2	114.03	0.71	88.00	32.2
RH50, RH80	-191.69	0.88	43.84	15.3	RH40, meanH	131.70	0.71	88.09	32.2
RH50, QMCH_chm	-188.18	0.88	43.87	15.3	RH98, QMCH_chm	134.13	0.71	87.66	32.1
RH50, RH70	-188.15	0.88	44.18	15.4	meanH, CCF10	136.58	0.71	88.07	32.2
RH50, meanH	-186.09	0.89	43.67	15.3	RH50, meanH	137.09	0.71	88.22	32.3
RH40, RH80	-185.00	0.88	43.96	15.4	meanH, QMCH_chm	139.23	0.71	88.29	32.3
RH40, RH50	-183.00	0.88	44.67	15.6	RH70, meanH	139.36	0.71	88.44	32.4
RH50, RH60	-182.72	0.88	44.37	15.5	RH60, meanH	139.42	0.71	88.36	32.3
RH40, RH90	-182.32	0.88	44.15	15.4	meanH, CCF5	141.40	0.71	88.14	32.3
RH40, RH70	-181.96	0.88	44.26	15.5	RH90, meanH	147.02	0.71	88.54	32.4

**Table 7.** LiDAR-AGB Linear Mixed Effects Model performance statistics at 1-ha and 0.16-ha plot sizes using two LCMs as predictive variables. The table is sorted in ascending order based on the column “AIC” (Akaike information criterion) when the respective LiDAR Canopy Metrics (LCM) are used in Eq. 1. R<sup>2</sup> and RMSEs (in Mg ha<sup>-1</sup> and in %) are computed on back-transformed predictions.

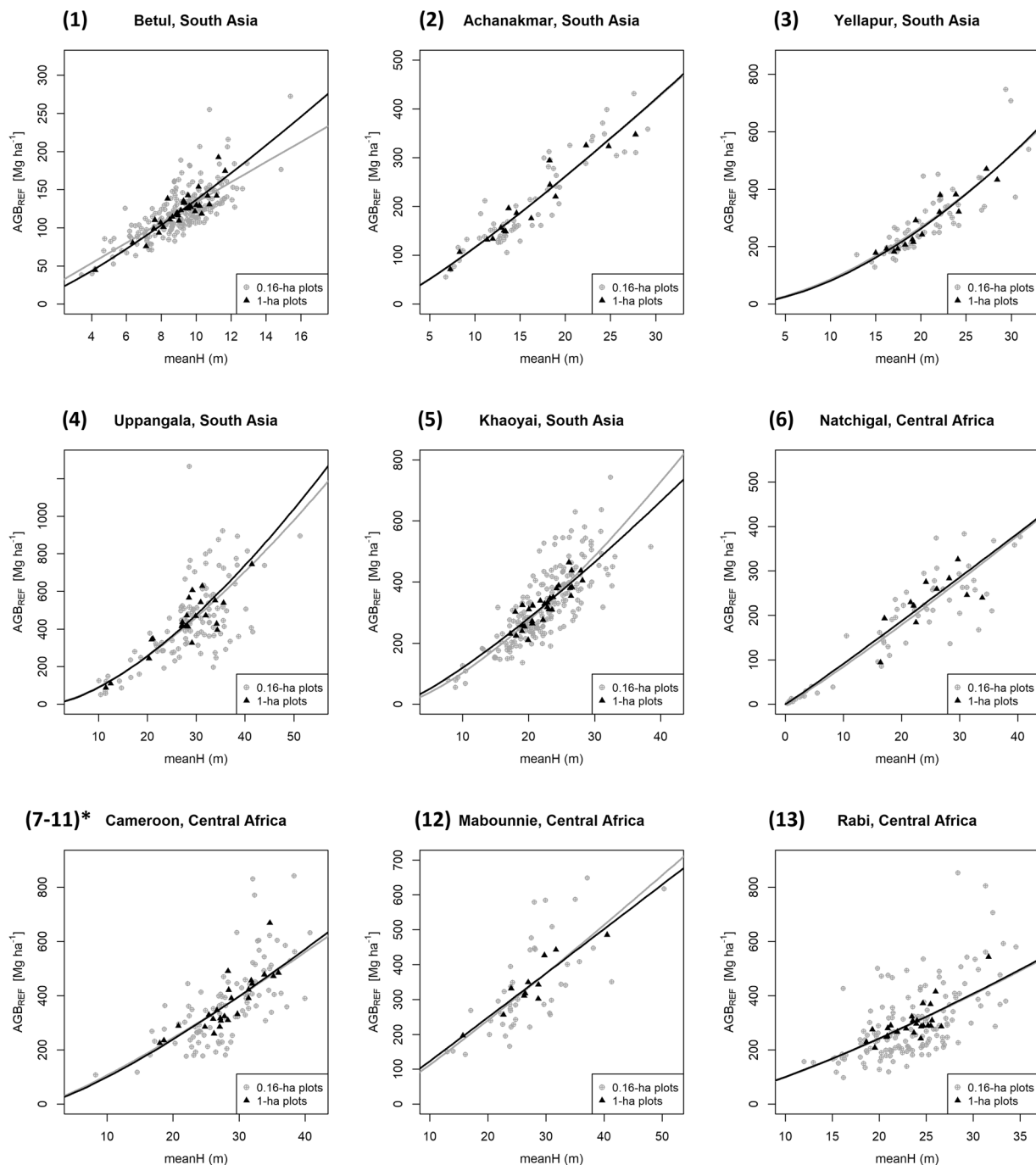
Sno	Site	Plot Size (1-ha)						Plot Size (0.16-ha)					
		a (se)	b (se)	sigma	R <sup>2</sup>	RMSE [Mg ha <sup>-1</sup> ]	RMSE [%]	a (se)	b (se)	sigma	R <sup>2</sup>	RMSE [Mg ha <sup>-1</sup> ]	RMSE [%]
1	Betul	2.043 (0.191)	1.247 (0.087)	0.095	0.87	12.04	9.9	2.605 (0.107)	0.988 (0.049)	0.162	0.65	21.08	17.3
2	Achanakmar	2.046 (0.219)	1.173 (0.080)	0.113	0.94	23.51	11.7	2.058 (0.175)	1.165 (0.064)	0.185	0.84	36.58	18.3
3	Yellapur	0.500 (0.478)	1.691 (0.158)	0.111	0.90	31.07	11.0	0.684 (0.380)	1.632 (0.126)	0.188	0.74	66.65	23.1
4	Uppangala	0.969 (0.442)	1.523 (0.134)	0.192	0.86	82.16	18.9	1.087 (0.335)	1.464 (0.100)	0.328	0.66	145.85	32.6
5	Khaoyai	1.934 (0.445)	1.236 (0.144)	0.109	0.72	32.27	9.9	1.360 (0.189)	1.411 (0.061)	0.219	0.74	74.08	23.1
6	Nachtigal	2.009 (1.004)	1.037 (0.314)	0.233	0.55	43.21	18.6	1.902 (0.081)	1.083 (0.029)	0.284	0.82	48.28	27.9
7	Mbalmayo												
8	Atout												
9	Kompia	1.721 (0.517)	1.253 (0.155)	0.135	0.74	52.62	14.1	1.230 (0.396)	1.39 (0.119)	0.282	0.59	108.54	29.1
10	Somalomo												
11	Bouamir												
12	Mabounié	2.471 (0.447)	1.015 (0.136)	0.100	0.86	32.26	9.4	2.159 (0.498)	1.098 (0.152)	0.265	0.55	94.23	28.0
13	Rabi	1.267 (0.671)	1.397 (0.213)	0.123	0.65	38.78	13.1	1.386 (0.407)	1.344 (0.130)	0.312	0.43	99.63	33.8

**Table 8.** Model coefficients along with standard errors (in brackets) for site-wise level models at 1-ha and 0.16-ha resolution. For Cameroon sites listed from 7-11 in column “Sno”, a single regional model is employed. Sigma is the model residual standard error in log-transformed units. R<sup>2</sup> and RMSEs (in Mg ha<sup>-1</sup> and in %) are computed on back-transformed predictions.

model coefficient selected in (1) to predict pixels AGB and (3) added to all pixels an error term randomly drawn from a normal distribution  $N(0, RSE_i)$  where  $RSE_i$  is the model RSE selected in (1). This procedure led to 1000 predictions of pixels AGB embedding the prediction uncertainty from both the first and second modelling levels. Finally, reference AGB maps and associated spatial uncertainty maps were generated as the mean and standard deviation of the 1000 pixel AGB predictions, respectively. Hereafter, we refer to pixels mean AGB prediction as  $AGB_{PRED}$ .

**Additional metadata for the AGB maps.** LiDAR-based AGB maps produced in the present study are intended to support calibration and validation efforts of spaceborne data. To maximise their usefulness, we provide additional information that users may require – depending on their study’s objective and methodological choices – to facilitate their integration with spaceborne data and/or develop comprehensive uncertainty propagation schemes up to the final, spaceborne-derived AGB map.

A first challenge users may face relates to the computation of the uncertainty associated with the mean AGB of arbitrary subregions of LiDAR AGB maps. Such subregions could for instance correspond to the footprints of spaceborne data unit pixels. Estimating the total mean squared error associated with a map (sub) population mean requires access to the matrix of pairwise population unit covariances, which is rarely communicated by map makers to users because of its large size. Yet, McRoberts *et al.*<sup>21</sup> recently showed that pairwise population unit covariances could largely contribute to total mean squared error, and proposed an averaging and binning approach to drastically reduce the matrix size, thus facilitating its publication along with AGB maps. While we refer interested readers to McRoberts *et al.*<sup>21</sup> for methodological details, we provide in Supplementary data all information recommended by the authors to allow map users to comply with IPCC good practice guidelines for



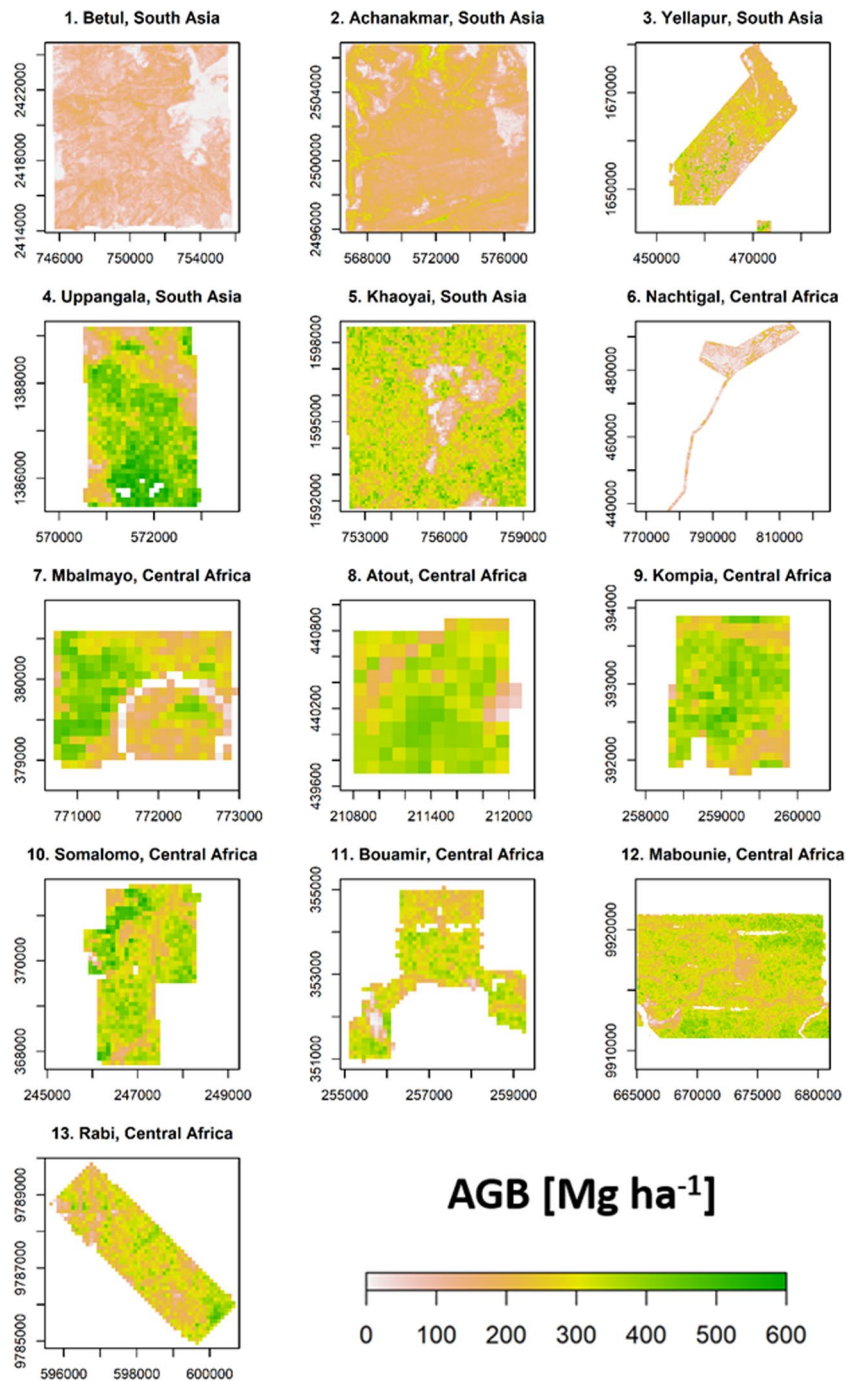
**Fig. 3** LiDAR-AGB models of Asian and African sites at 1-ha and 0.16-ha resolutions. The numbers for each site refer to Table 1. (7–11)\* refers to the regional model established over moist dense forests of Cameroon.

greenhouse gas inventories. We note that for each pixel of the LiDAR-based AGB maps provided in this study, a bin number is available in the third map layer.

Another challenge lies in the propagation of uncertainties in multi-level hierarchical modelling, which is a likely use-case of the LiDAR-based maps we produced. These maps were generated by applying two hierarchically nested models: a tree allometric model linking field measurements to tree AGB, and a mapping model linking plot AGB to LiDAR data. LiDAR-based AGB maps users may employ a three-steps hierarchical modelling approach and add as a third step a model linking high resolution AGB predictions from the LiDAR-based maps to the coarser resolution of spaceborne data. An example of such an approach is presented in detail in Saarela *et al.*<sup>22</sup> and referred to as “three-phase hierarchical model-based inference”. The uncertainty assessment in such a nested modelling approach requires information at the two first modelling steps that goes beyond the results of the Monte Carlo simulation we used to produce pixel-level uncertainty estimates. While we refer interested readers to Saarela *et al.*<sup>22</sup> for methodological details, we provide in Supplementary data all information allowing users



# 1-ha Maps

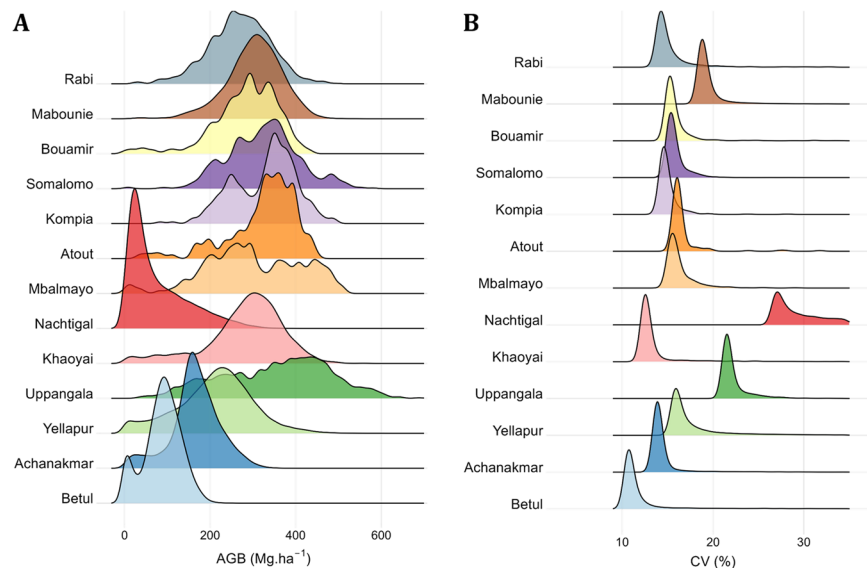


**Fig. 4** Reference AGB maps of Asian and African sites at 1-ha spatial resolution.

to assess uncertainty as described in Saarela *et al.*<sup>22</sup>. This information notably includes the variance-covariance matrix of model parameters for each sampling site as well as statistics on parameters (DBH, AGB, pixels' height from CHM, etc.,) used at various levels in the chain of hierarchical models.

## Data Records

For each site, AGB and uncertainty maps are distributed as a single GeoTiff file at the two spatial resolutions (1- and 0.16-ha) through Dataverse<sup>23</sup>. Each file comprises three individual layers. The two first layers named meanAGB and sdAGB correspond to the mean and standard deviation of AGB predictions over the 1000 Monte Carlo simulations, respectively. The file projection system is Universal Transverse Mercator. The third layer



**Fig. 5** Density distributions of (A) mean pixel AGB and (B) AGB uncertainty, expressed as a coefficient of variation (CV, in %), at 1-ha mapping resolution across sites.

named *Nbin* corresponds to the bin number each map pixel is associated with in the binning approach proposed by McRoberts *et al.*<sup>21</sup> to allow users reconstituting a matrix of pairwise population unit covariances estimates.

Besides data access through Rodda *et al.*<sup>23</sup>, data from Asian sites can be accessed and visualized through the Bhuvan Portal (<https://bhuvan-app3.nrsc.gov.in/data/download/index.php>). To access the visualization/download through Bhuvan Portal, select the ISRO Geosphere-Biosphere Programme under the Program option and then choose the group Above Ground Biomass (AGB) Data.

In addition, we provide two supplementary data files (in excel format) that provide additional metadata details on site-level binned covariance matrices and variance-covariance matrices and summary of all the parameters used in the present study.

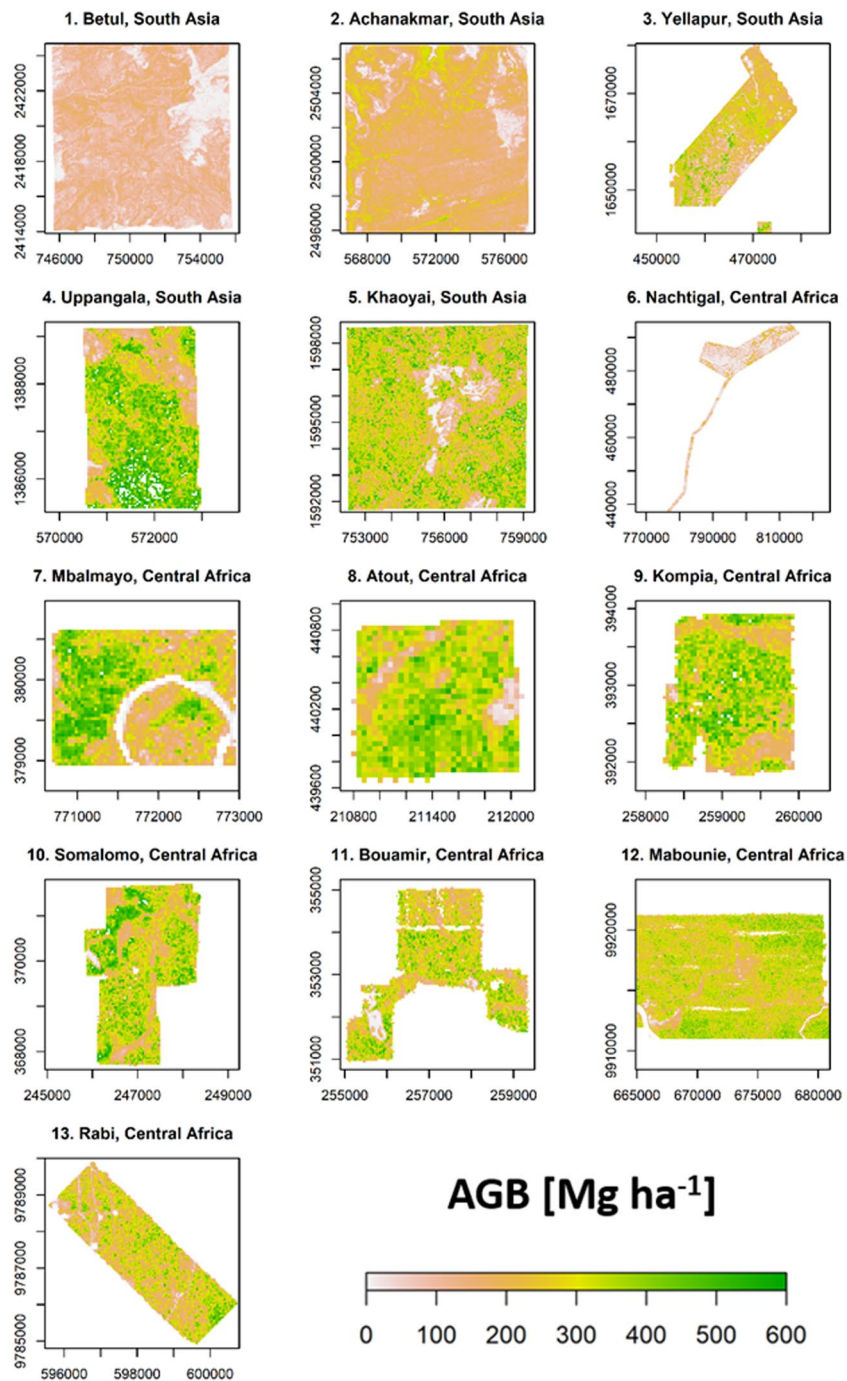
### Technical Validation

Reference AGB maps at 1-ha resolution are shown in Fig. 4 and the density distributions of 1-ha AGB maps are represented in Fig. 5-A along with uncertainty levels in Fig. 5B expressed as a coefficient of variation (CV, in % of mean AGB) (see Fig. 6 and Fig. 7 for AGB maps and respective density distributions at 0.16-ha resolution). Figure 5-B shows that the mean uncertainty across sites is 15.4%, with site-level mean uncertainty ranging from 10.8 to 31%. It can be observed that Nachtigal and Uppangala sites have larger mean uncertainties than other sites, with 31% and 20.1%, respectively. This can be explained by larger LiDAR-AGB model uncertainties at these sites and mapping resolution (see models' "sigma" in Table 8).

In addition to the per pixel estimates of uncertainty accompanying AGB maps, we hereafter provide (i) an assessment of mapping model predictive performances using a spatial model cross-validation technique<sup>24</sup>, to provide additional insights into the reliability of AGB predictions on each map and (ii) an assessment of mapping models extrapolation at each sampling site, which may be useful to help users for filtering-out pixels where extrapolation occurred and only retaining the highest quality AGB predictions for spaceborne products calibration/validation.

**Model spatial cross-validation.** Model calibration statistics in Table 8 likely overestimate model predictive performance on pixels that are not used for model training, that is, on most maps' pixels. We performed a cross-validation (CV) of each model to provide more reliable insights into model predictive performance. Field plots at each site are iteratively split into training and test data and model CV statistics are built on the set of test data predictions. Regarding CV design, we selected a buffered leave-one-out cross-validation (LOO-CV<sup>25</sup>) where a spatial buffer around test data is used to exclude from model training dataset observations located at the neighbourhood of test data, thus avoiding inflation in CV statistics due to spatial autocorrelation in forest AGB<sup>24</sup>. As a compromise between the diversity in terms of number and spatial arrangement of field data across sites (e.g. multiple individual 1-ha plots vs. single large plot), the consistency of the CV approach across sites, as well as our expectation for a relatively weak spatial autocorrelation in forest AGB at the high resolution of the maps (<100 m<sup>2</sup>)<sup>26</sup>, we selected a LOO-CV with a 100 m buffer radius for all sites and mapping resolutions (i.e. 100 × 100 m and 40 × 40 m). This CV design notably implies that (i) when a test observation came from a large field plot (i.e. >1-ha, e.g. the 25-ha plot at Rabi), subplots at its direct neighbourhood were not used for model training (i.e., all subplots intersecting a 100 m circular buffer around the center of a test subplot were excluded from the training set, regardless of the mapping resolution), and (ii) at the 40 × 40 m mapping resolution, when a test observation came from a 1-ha field plot, the remaining three subplots of that 1-ha plot were not used for model training. The results of the buffered LOO-CV are presented in Table 9. They show that the predictive

# 0.16-ha Maps



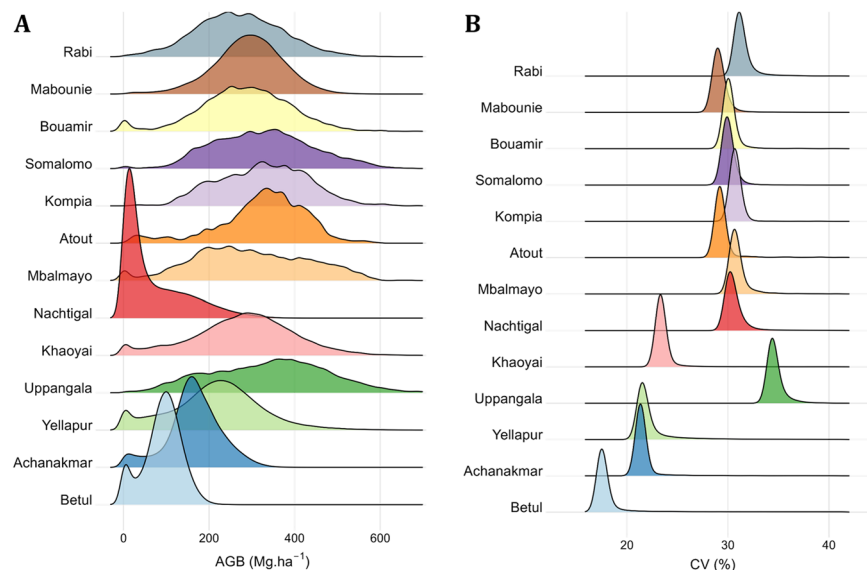
**Fig. 6** Reference AGB maps of Asian and African sites at 0.16-ha spatial resolution.

performances of mapping models developed in this study are comparable to those found in the literature (i.e. 15–20% on average for the tropical forest biome<sup>15</sup>) with relative RMSEs ranging from 10.6 to 20.1% (mean across sites: 14.1%) at 1-ha and 17.7 to 33.7% (mean across sites: 25.7%) at 0.16-ha.

**Model extrapolation in the predictor space.** Uncertainty maps, AGB maps, and model CV results provide insights into the reliability of AGB predictions within the calibration domain of mapping models. It is however likely that the entire gradient of forest structure sampled by LiDAR data was not fully sampled in the model's training set, thus leading to situations of predictive extrapolation where prediction uncertainty is unknown. To investigate this issue, we compared the range of vegetation height (i.e., meanTCH) sampled by

Sno	Site	Plot Size (1-ha)			Plot Size (0.16-ha)		
		R <sup>2</sup>	LOOCV-RMSE [Mg ha <sup>-1</sup> ]	LOOCV-RMSE [%]	R <sup>2</sup>	LOOCV-RMSE [Mg ha <sup>-1</sup> ]	LOOCV-RMSE [%]
1	Betul	0.76	12.85	10.5	0.55	21.47	17.6
2	Achanakmar	0.90	27.05	13.5	0.81	39.00	19.5
3	Yellapur	0.85	36.27	12.9	0.62	74.21	25.8
4	Uppangala	0.64	100.44	23.2	0.37	171.68	38.4
5	Khaoyai	0.65	37.40	11.5	0.62	75.05	23.4
6	Nachtigal	0.22	60.49	26.1	0.81	49.85	28.8
7	Mbalmayo	0.66	56.36	15.1	0.46	110.39	29.6
8	Atout						
9	Kompia						
10	Somalomo						
11	Bouamir						
12	Mabounié	0.79	37.47	11.0	0.46	96.32	28.7
13	Rabi	0.54	46.11	15.5	0.39	101.74	34.5

**Table 9.** Error statistics of modified LOO-CV procedure at site-level for 1-ha and 0.16-ha plots.

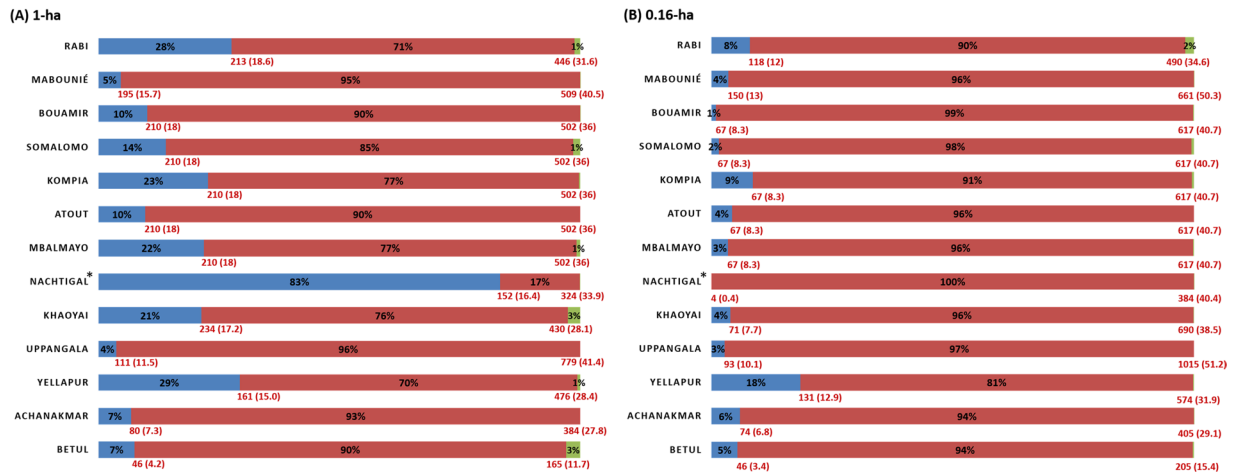


**Fig. 7** Density distributions of (A) mean pixel AGB and (B) AGB uncertainty, expressed as a coefficient of variation (CV, in %), at 0.16-ha mapping resolution across sites.

the training set of each mapping model to the full range found in the LiDAR data, restricting the analysis to pixels considered as vegetated, i.e., with meanTCH  $\geq 2$  m. We found that the proportion of pixels affected by predictive extrapolation strongly varied across sites and at the two mapping resolutions. Generally, the upper range of meanTCH (and thus of AGB<sub>PRED</sub>) found at a landscape scale in the LiDAR data were sampled in the training set (Fig. 8A,B), which probably is a reflection of the “majestic forest bias”<sup>27</sup> – that is, the tendency for researchers to preferentially establish sample plots where forest stands appear the less disturbed (e.g. tallest canopy height, the highest abundance of large trees, etc.). However, a varying and often substantial proportion of maps on the lower end of the meanTCH gradient was outside the model’s calibration domain. For instance, in the Nachtigal site predictive extrapolation occurred on about 83% of the vegetated pixels on the 1-ha AGB map. This can be explained by the nature of this site, a forest-savanna mosaic, where the meanTCH of all herbaceous and shrubby savannas is lower than the height of the smallest 1-ha forest stand (i.e., 16.4 m) found in model training set (Fig. 8A). However, this proportion dropped to 0% at the 0.16-ha mapping resolution thanks to the inclusion into the model training set of 18 additional 0.16-ha plots established in savannas-dominated areas (Fig. 8B, Table 3).

We thus advise potential users of the AGB maps published here to carefully consider the bounds of the mapping model calibration domain at each site (for which we provided corresponding AGB<sub>PRED</sub> values in Fig. 8) when using these maps as reference data for larger-scale product calibration/validation.





**Fig. 8** Proportion (in %) of map pixels outside and inside models calibration domains at 1-ha (panel A) and 0.16-ha (panel B) mapping resolutions. The proportions are computed with respect to the total number of map pixels with CHM > 2 m at the exception of the Natchigal site where a 0.4 m threshold is used so as to account for the nature of the site i.e., a forest-savanna mosaic. The proportion of map pixels within model calibration domains is represented in red. Map pixels below and above the range of model calibration domains are represented in blue and green, respectively.

### Usage Notes

Forest AGB maps released here constitute reference estimations for the community of remote-sensing scientists interested in forest carbon stocks. For instance, we expect these maps to be of utmost usefulness for the calibration and validation of next-generation broad-scale aboveground biomass mapping models based on data from ongoing or upcoming spaceborne missions (viz., NASA’s GEDI, NASA-ISRO’s NISAR and ESA’s BIOMASS missions). These data can also be useful when assessing the accuracy of existing maps or recalibrating them (as in eg. 28,29), especially since study sites presented here are located on renowned data-poor regions<sup>16</sup> and are marked by notable uncertainties in AGB estimates<sup>17</sup>. That said, we encourage users to account for the dates of LiDAR and ground data acquisitions underlying reference AGB estimates (cf. Table 1), as temporal discrepancies with spaceborne signals or products should ideally be accounted for in any calibration or validation exercise.

More broadly, our study highlights sites of potential interest to build a network of “super-sites” (*sensu*<sup>11</sup>) across the tropics, that is sites combining forest inventory data over sizable areas ( $\geq 10$  ha) – ideally featuring multiple forest censuses – with airborne LiDAR data. Such data have been collected thanks to the long-term vision of few organisations, to dedicated experts and to the efforts of trained labour forces in the past decades. Ongoing global changes makes the sustained monitoring of permanent forest plots in long-term study sites critical, so as to allow measuring their impacts on forest ecosystems. In spite of this crucial stake, access to funding in the tropical world for replacement of expertise, training programs and to support field data acquisition campaigns is critically limited. We thus urge National and International research and space agencies to ensure long-term funding for on-ground forest research in the tropics.

### Code availability

All statistical analyses were performed in R (v.4.1.3). The BIOMASS R-package is an open source library available from the CRAN R repository. The BIOMASS vignettes and individual function helps contain detailed notes on usage to derive plot level AGB estimates with uncertainty estimates through error propagation using Monte Carlo method. The codes associated with the error propagation from plot-scale AGB to LiDAR AGB maps, generating additional metadata and binned covariance matrices as described in Methods section are available on GitHub ([https://github.com/surajreddyr/LIDAR\\_AGB/tree/main](https://github.com/surajreddyr/LIDAR_AGB/tree/main)).

Received: 10 July 2023; Accepted: 19 March 2024;

Published online: 04 April 2024

### References

- Mitchard, E. T. A. *et al.* Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. *Carb Bal Manag.*, (2013).
- Schimel, D. *et al.* Observing terrestrial ecosystems and the carbon cycle from space. *Glob. Chang. Biol.* **21**, 1762–1776 (2015).
- Herold, M. *et al.* The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surv. Geophys.* **40**, 757–778 (2019).
- Mitchard, E. T. A. *et al.* Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Glob. Ecol. Biogeogr.* **23**, 935–946 (2014).
- Réjou-Méchain, M. *et al.* Upscaling Forest biomass from field to satellite measurements: sources of errors and ways to reduce them. *Surv. Geophys.* **40**, 881–911 (2019).
- 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).

7. Fararoda, R. *et al.* Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm. *Ecol. Inform.* 101392 (2021).
8. Quegan, S. *et al.* The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space. *Remote Sens. Environ.* **227**, 44–60 (2019).
9. Amelung, F. & others. NASA-ISRO SAR (NISAR) Mission Science Users' Handbook. *Jet Propuls. Lab.* (2019).
10. Motohka, T., Kankaku, Y., Miura, S. & Suzuki, S. Overview of ALOS-2 and ALOS-4 L-band SAR. in *2021 IEEE Radar Conference (RadarConf21)* 1–4 (2021).
11. Chave, J. *et al.* Ground data are essential for biomass remote sensing missions. *Surv. Geophys.* **40**, 863–880 (2019).
12. Labrière, N. *et al.* Toward a forest biomass reference measurement system for remote sensing applications. *Glob. Chang. Biol.* **29**, 827–840 (2023).
13. Duncanson, L. *et al.* The importance of consistent global forest aboveground biomass product validation. *Surv. Geophys.* **40**, 979–999 (2019).
14. Duncanson, L. *et al.* Aboveground Woody biomass product validation good practices protocol 2021.
15. Zolkos, S. G., Goetz, S. J. & Dubayah, R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sens. Environ.* **128**, 289–298 (2013).
16. Duncanson, L. *et al.* Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sens. Environ.* **270**, 112845 (2022).
17. Rodda, S. R., Nidamanuri, R. R., Fararoda, R., Mayamanikandan, T. & Rajashekar, G. Evaluation of Height Metrics and Above-Ground Biomass Density from GEDI and ICESat-2 Over Indian Tropical Dry Forests using Airborne LiDAR Data. *J. Indian Soc. Remote Sens.* 1–16 (2023).
18. Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J. & Hérault, B. biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods Ecol. Evol.* **8**, 1163–1167 (2017).
19. Chave, J. *et al.* Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **20**, 3177–3190 (2014).
20. Asner, G. P. *et al.* A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* **168**, 1147–1160 (2012).
21. McRoberts, R. E., Næsset, E., Saatchi, S. & Quegan, S. Statistically rigorous, model-based inferences from maps. *Remote Sens. Environ.* **279**, 113028 (2022).
22. Saarela, S. *et al.* Three-phase hierarchical model-based and hybrid inference. *MethodsX* **11**, 102321 (2023).
23. Rodda, S. R. *et al.* South Asian and Central African maps from: LiDAR-based reference aboveground biomass maps for tropical forests of South Asia and Central Africa. *Dataverse* <https://doi.org/10.23708/H2MHXF> (2024).
24. Ploton, P. *et al.* Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540 (2020).
25. Parmentier, I. *et al.* Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. *J. Biogeogr.* **38**, 1164–1176 (2011).
26. Réjou-Méchain, M. *et al.* Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences Discuss.* **11**, 5711 (2014).
27. 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).
28. Mitchard, E. T. A. *et al.* Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Glob. Ecol. Biogeogr.* **23**, 935–946 (2014).
29. Avitabile, V. *et al.* An integrated pan-tropical biomass map using multiple reference datasets. *Glob. Chang. Biol.* **22**, 1406–1420 (2016).

## Acknowledgements

S.R.R. received support from French Research Institute for Development (IRD) through an International Travel grant for the execution of this study. We would like to acknowledge the assistance provided by the following entities in various aspects of this research. For the Indian (Asia) sites, we gratefully acknowledge the funding support received from ISRO-Geosphere Biosphere Program (ISRO-GBP) for LiDAR data acquisition and field work at the four Indian sites. We also express our thanks for the consistent support provided by the Indian state forest departments of Karnataka, Madhya Pradesh and Chhattisgarh, which facilitated the forest inventory measurements. We also express our gratitude to Mr. Prashant Kaiwshwar, Scientist (CCOST) for support in the Achanakmar field site in Chhattisgarh, India. Furthermore, we would like to acknowledge the assistance of Aerial Surveys and Digital Mapping Area at the National Remote Sensing Centre (ISRO) for their contributions to flight planning, LiDAR data acquisition and pre-processing. We extend our thanks to the team members who contributed to field measurements at the following locations: (A) Betul, Yellapur and Achanakmar – KV Satish, Wajeed Pasha, M. Suresh, and TR Kiran Chand and (B) Uppangala: S. Aravajy, N. Barathan, Jules Morel, Cedric Vega, A. Dharanidharan, G. Saravanan, K. Ramanujam, Jerome Feret, Ahmed Hamrouni, K. Adimoolam and T. Gopal. For African sites, research was carried out in the framework of the international joint laboratory (LMI) Dycofac, and within the One Forest Vision Initiative. At the exception of Rabi, forest inventories have been conducted thanks to recurrent funding of the IRD. We also acknowledge additional funding from EDF and Nachtigal Hydropower Company (for forest inventories and LiDAR data at Nachtigal) and from ERAMET, Maboumine and the Missouri Botanical Garden (for forest inventories and LiDAR data at Maboumine). We would like to thank Assala Gabon, Compagnie des Bois de Gabon and the Smithsonian National Zoo and Conservation Biology Institute for supporting the collection of the Rabi data in Gabon (contribution No 208 of the Gabon Biodiversity Program). Last, we warmly thank the Ministry of Forestry and Wildlife (MINFOF) of Cameroon for its continuous support to forest scientific research.

## Author contributions

Conceptualization: S.R.R. and P.P.; data curation: R.F., S.R.R., R.G., M.M., G.D., N.B., P.P., N.J., M.R.-M.; formal analysis: S.R.R., P.P., N.J. and R.F.; writing (original draft): S.R.R., R.F., P.P. and M.R.-M.; writing (review and editing): all authors.

## 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/s41597-024-03162-x>.

**Correspondence** and requests for materials should be addressed to S.R.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Suraj Reddy Rodda<sup>1,2</sup>✉, Rakesh Fararoda<sup>1</sup>, Rajashekar Gopalakrishnan<sup>1</sup>, Nidhi Jha<sup>3</sup>, Maxime Réjou-Méchain<sup>4</sup>, Pierre Couteron<sup>4</sup>, Nicolas Barbier<sup>4</sup>, Alonso Alfonso<sup>5</sup>, Ousmane Bako<sup>6</sup>, Patrick Bassama<sup>6</sup>, Debabrata Behera<sup>7</sup>, Pulcherie Bissiengou<sup>8</sup>, Hervé Biyiha<sup>6</sup>, Warren Y. Brockelman<sup>9</sup>, Wirong Chanthorn<sup>10</sup>, Prakash Chauhan<sup>1</sup>, Vinay Kumar Dadhwal<sup>11</sup>, Gilles Dauby<sup>4,12,13</sup>, Vincent Deblauwe<sup>14,15</sup>, Narcis Dongmo<sup>6</sup>, Vincent Droissart<sup>4,12</sup>, Selvaraj Jeyakumar<sup>7</sup>, Chandra Shekar Jha<sup>1</sup>, Narcisse G. Kandem<sup>12</sup>, John Katembo<sup>16</sup>, Ronald Kougue<sup>6</sup>, Hugo Leblanc<sup>4</sup>, Simon Lewis<sup>17,18</sup>, Moses Libalah<sup>12</sup>, Maya Manikandan<sup>1</sup>, Olivier Martin-Ducup<sup>19</sup>, Germain Mbock<sup>6</sup>, Hervé Memiaghe<sup>8</sup>, Gislain Mofack<sup>12</sup>, Praveen Mutyala<sup>1</sup>, Ayyappan Narayanan<sup>7</sup>, Anuttara Nathalang<sup>9</sup>, Gilbert Oum Ndjock<sup>20</sup>, Fernandez Ngoula<sup>12</sup>, Rama Rao Nidamanuri<sup>2</sup>, Raphaël Pélissier<sup>4</sup>, Sassan Saatchi<sup>21</sup>, Le Bienfaiteur Sagang<sup>12,15</sup>, Patrick Salla<sup>12</sup>, Murielle Simo-Droissart<sup>12</sup>, Thomas B. Smith<sup>15</sup>, Bonaventure Sonké<sup>12,13</sup>, Tariq Stevart<sup>22</sup>, Danièle Tjomb<sup>6</sup>, Donatien Zebaze<sup>12</sup>, Lise Zemagho<sup>12</sup> & Pierre Ploton<sup>4,12,13</sup>

<sup>1</sup>Forestry and Ecology Group, National Remote Sensing Centre, ISRO, Hyderabad, 500 037, India. <sup>2</sup>Indian Institute of Space Science and Technology (IIST), Thiruvananthapuram, Kerala, India. <sup>3</sup>College of Forestry, Oregon State University, Corvallis, OR, 97331, USA. <sup>4</sup>AMAP, Univ Montpellier, IRD, CNRS, INRAE, CIRAD, Montpellier, France. <sup>5</sup>Center for Conservation and Sustainability, Smithsonian National Zoo and Conservation Biology Institute, Washington, DC, USA. <sup>6</sup>Ecole Nationale des Eaux et Forêts de Mbalmayo, Ministère Des Forêts Et De La Faune, Mbalmayo, Cameroon. <sup>7</sup>Department of Ecology, French Institute of Pondicherry, Pondicherry, 605 001, India. <sup>8</sup>Institut de pharmacopée et de médecine traditionnelle (Herbier National du Gabon), CENAREST, Libreville, Gabon. <sup>9</sup>National Biobank of Thailand (NBT), National Science and Technology Development Agency, Klong Luang, Pathum Thani, Thailand. <sup>10</sup>Department of Environmental Technology and Management, Faculty of Environment, Kasetsart University, Bangkok, 10900, Thailand. <sup>11</sup>National Institute of Advanced Studies (NIAS), Bangalore, India. <sup>12</sup>Plant Systematics and Ecology Laboratory, Higher Teachers' Training College, University of Yaoundé I, P.O. Box 047, Yaoundé, Cameroon. <sup>13</sup>International Joint Laboratory DYCOFAC, IRD-UYI-IRGM, P.O. Box 1857, Yaoundé, Cameroon. <sup>14</sup>International Institute of Tropical Agriculture (IITA), BP 2008 (Messa), Yaoundé, Cameroon. <sup>15</sup>Center for Tropical Research, Institute of the Environment and Sustainability, University of California Los Angeles, Los Angeles, CA, 90095, USA. <sup>16</sup>Institut Supérieur d'Etudes Agronomiques de Bengamisa, République Démocratique du Congo, Congo, France. <sup>17</sup>Department of Geography, University College London (UCL), London, UK. <sup>18</sup>School of Geography, University of Leeds, Leeds, UK. <sup>19</sup>Ecologie des Forêts Méditerranéennes (URFM), INRAE, F-84914, Avignon, France. <sup>20</sup>Dja Wildlife Reserve, Ministry of Forestry and Wildlife, Yaoundé, Cameroon. <sup>21</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109, USA. <sup>22</sup>Missouri Botanical Garden, Africa & Madagascar Program, 4344 Shaw Blvd., St. Louis, Missouri, 63110, USA. ✉e-mail: [surajreddy.rs@gmail.com](mailto:surajreddy.rs@gmail.com)