Article

Multi-scale lidar measurements suggest miombo woodlands contain substantially more carbon than thought

Check for updates

Miro Demol [1] ✉, Naikoa Aguilar-Amuchastegui [2], Gabija Bernotaite [1], Mathias Disney [3,4], Laura Duncanson [5], Elise Elmendorp [1], Andres Espejo [2], Allister Furey[1], Steven Hancock [6], Johannes Hansen [1], Harold Horsley [1], Sara Langa[7], Mengyu Liang [5], Annabel Locke [1], Virgílio Manjate[8], Francisco Mapanga[9], Hamidreza Omidvar [1], Ashleigh Parsons[1], Elitsa Peneva-Reed [2], Thomas Perry [1], Beisit L. Puma Vilca [1], Pedro Rodríguez-Veiga [1,10], Chloe Sutcliffe[1], Robin Upham [1], Benoît de Walque [1] & Andrew Burt [1]

Miombo woodlands are integral to livelihoods across southern Africa, biodiversity in the region, and the global carbon cycle, making accurate and precise monitoring of their state and change essential. Here, we assembled a terrestrial and airborne lidar dataset covering 50 kha of intact and degraded miombo woodlands, and generated aboveground biomass estimates with low uncertainty via direct 3D measurements of forest structure. We found 1.71 ± 0.09 TgC was stored in aboveground biomass across this landscape, between 1.5 and 2.2 times more than the 0.79–1.14 TgC estimated by conventional methods. This difference is in part owing to the systematic underestimation of large trees by allometry. If these results were extrapolated across Africa's miombo woodlands, their carbon stock would potentially require an upward revision of approximately 3.7 PgC, implying we currently underestimate their carbon sequestration and emissions potential, and disincentivise their protection and restoration.

Miombo woodlands, the dry tropical forests spanning large areas of southern Africa, directly support many millions of livelihoods in various ways including supply of plant-based materials, fertile soils for agriculture, and grazing lands[1]. These ecosystems also hold cultural and spiritual significance, provide habitat for substantial plant and animal biodiversity, and regulate both the climate and water resources[2]. These landscapes, however, are changing because of human activities, with cover reducing from approximately 2.7 to 1.9 million km[2] between 1980 and 2020[3]. Owing to both their importance and dynamic nature, it is therefore crucial to monitor how the world's miombo woodlands are changing.

One essential climate variable that requires accurate and precise monitoring is the aboveground biomass (AGB) and carbon stored in these woodlands[4]. Any uncertainty that exists in the quantification of these stocks has consequences, particularly regarding misinformed policy and decision making towards them, as well as the misallocation of funding and resources[5,6]. Carbon markets for example, through

programmes such as Reducing Emissions from Deforestation and Degradation (REDD+)[7], require low uncertainty in estimates of carbon stocks if they are to properly incentivise direct climate benefits, and co-benefits including biodiversity and ecosystem services, by safeguarding these woodlands. Further, intended outcomes from international climate agreements towards greenhouse gas emissions reductions, such as the Paris Agreement including individual countries' Nationally Determined Contributions, are premised on forest carbon accounting with low uncertainty[8]. That is, both high accuracy and precision, quantitatively expressed as a bias and variance, respectively, are usually important for any estimate of forest AGB stocks in these contexts. Whilst accuracy is the principal concern in accounting (systematic over- or under-estimation commensurately misleads understanding of forest carbon sequestration and emissions potential)[9], precise estimates are also important, including from the requirement to detect change over time (it can be problematic to interpret differences between observations with low precision)[10]. This is

[1]Sylvera Ltd, London, UK. [2]The World Bank Group, Washington, DC, USA. [3]Department of Geography, University College London, London, UK. [4]NERC National Centre for Earth Observation (NCEO), Leicester, UK. [5]Department of Geographical Sciences, University of Maryland, College Park, USA. [6]School of GeoSciences, University of Edinburgh, Edinburgh, UK. [7]Independent researcher, Maputo, Mozambique. [8]Independent researcher, Boane, Mozambique. [9]Independent researcher, Inhaca, Mozambique. [10]School of Geography, Geology and the Environment, University of Leicester, Leicester, UK. ✉e-mail: miro@sylvera.io

particularly the case for miombo woodlands given the aforementioned pace of their anthropogenic change.

The conventional approach to quantifying region-scale forest AGB stocks across miombo woodlands, and forests generally, within the context of UNFCCC- and IPCC-compliant greenhouse gas inventories, sees the combination of activity data and emissions factors (EF): remotely sensed estimates of forest area are multiplied by values of expected AGB per unit area of forest[11]. These expected values, based on in-situ measurements, might be generated from National Forest Inventories (NFI), or alternatively, where such data are unavailable, taken from the literature, such as IPCC defaults[12]. While this overall approach can be readily implemented it does have limitations, including: (i) restricted ability to describe AGB variations within forest types; (ii) EFs not being representative of the forest in question; and (iii) failing to detect change beyond binary transition between forest and non-forest (e.g. degradation).

For example, when focusing solely on the EF, and ignoring immediate questions surrounding the representativeness of applying a single value to any particular region of miombo woodland, uncertainties arise from the methods used to gather the in-situ data from the forest plots underlying the EFs themselves[13]. A ubiquitous feature of such measurements is the application of allometric models to estimate individual tree AGB. These models characterise the correlations that exist between tree shape and mass, enabling AGB estimation from more readily-measurable predictor variables such as stem diameter and tree height[14]. Such allometrics are themselves calibrated using hard-won destructive weighing measurements collected from a limited number of harvested trees that then must represent the entire variability of the specific taxa or region where that model is subsequently applied.

Uncertainties in allometric-derived AGB predictions therefore arise from the selection, measurement and modelling of these calibration trees, and the measurement of the predictor variables of any out-of-sample tree[15]. Several studies have explored the precision of allometric predictions of tropical and subtropical forests, where the expectation is that uncertainties range from 10 to 40% of the estimate itself at the hectare-scale[16,17]. Research has also explored their accuracy, with a particular focus on the selection and modelling of allometric calibration data[18], which are routinely heavily skewed towards small trees owing to their relative ease of harvesting. It has been hypothesised that this, combined with inadequate statistical methods, might cause biased AGB predictions for underrepresented larger trees[19]. Concurrently, independent lidar-based methods for AGB estimation have shown large differences versus allometry, estimating up to 1.77 times greater stocks at the plot-scale[20]. These potential uncertainties in allometric predictions are problematic as they would propagate directly into derived EFs and their aforementioned applications.

Here, we present the first (to our knowledge) mapping of region-scale AGB stocks generated entirely independent of the above conventional methods, including activity data, EFs and allometrics, using 3D multi-scale lidar (MSL) data acquired across 50 kha of forests in and around Gilé National Park, Mozambique (Fig. 1). The continuous region of interest (ROI) where these data were collected was selected such that it ranged from intact forests, secondary forests in various states of degradation, through to clearland, resulting in data that could reasonably be considered representative of miombo woodland landscapes more widely. Across the ROI, the MSL dataset (approximately 450 billion measurements) comprised helicopter-based airborne laser scanning (ALS) across its entirety, unoccupied aerial vehicle laser scanning (UAV-LS) from six 300 ha sections, and terrestrial laser scanning (TLS) and conventional forest inventory measurements from six coincident 1 ha plots.

An inverted pyramid approach was used to estimate AGB stocks and uncertainties across the ROI from these MSL data (Fig. 2), whereby each layer calibrated the next, commencing with the TLS point clouds that estimated individual tree AGB via both quantitative structural models (QSMs) explicitly reconstructing whole-tree woody architecture and volume[21] (Fig. 3), and estimates of basic woody tissue density[22]. These estimates were themselves calibrated and validated using a literature sample of destructive measurements[23]. TLS-derived AGB was gridded and used to train extreme gradient boosting machine learning models[24] using predictor variables retrieved from the UAV-LS point clouds that describe forest structure[25,26] (e.g., canopy height, tree fractional cover, and voxel occupancy rates describing the 3D distribution of woody plant material), with the step repeated to upscale to the ALS. The optimisation and performance of these models was evaluated using spatial cross-validation methods[27], with meaningful confidence intervals capturing uncertainties arising from both the QSM-derived training data and the upscaling itself, providing a robust understanding of the uncertainty in AGB predictions.

These new MSL-derived AGB estimates provide insights into the accuracy and precision of current best practices, and show that between 51 and 118% more AGB is stored across these miombo woodlands than conventional methods suggest. This is first demonstrated at the tree-scale by directly comparing TLS and allometric estimates for 1000+ trees (Fig. 3). We then show how these tree-level discrepancies, in part, translate into large differences at the region-scale, by comparing MSL-derived AGB stocks across the 50 kha ROI with counterparts estimated from both activity data and EFs (Fig. 4), and more direct mapping methods, using AGB products from the NASA GEDI spaceborne lidar mission[28] (Fig. 5). In the discussion, we explore the likely drivers of these differences, and examine these results in the context of miombo woodlands and global change more widely, particularly the consequences for their protection and restoration.

Results

Divergence between small and large trees

A total of 1071 individual trees were explicitly matched in both the TLS and inventory data, with their TLS-derived AGB summing to 462.0 Mg, compared with 450.8, 421.9, and 414.0 Mg (2.5%, 9.5%, and 11.6% smaller, respectively) predicted from two miombo woodland-specific and one widely-used pan-tropical allometric models, respectively[29,30] (Fig. 3b, c). Approximately 50% of AGB was stored in the largest 115 trees by stem diameter (i.e., 11% of trees). Here, the differences in AGB predictions between both methods were more marked, summing to 232.0 Mg vs. 215.0, 198.6 Mg, and 197.8 Mg (7.9%, 16.9% and 17.3% smaller, respectively). That is, a systematic trend was observed whereby allometric predictions produced similar estimates for small trees, but smaller estimates for large trees.

Region-scale differences

MSL-derived AGB (Fig. 2c) across the ROI summed to 3.85 Tg ± 11.0% (uncertainty expressed as 90% confidence intervals), with uncertainty at individual pixel-level (10 m resolution) averaging 60.4%. This AGB estimate reduced to 3.65 Tg, with average AGB density of 98.4 Mg/ha when considering forested area only (37 kha, derived via a MSL-based forest/non-forest mask defined as tree fractional cover greater than or equal to 30% using a canopy threshold of 5 m at 10 m resolution[31]). Estimated AGB densities were 99.2, 100.3, and 86.5 Mg/ha for core park, buffer zone and beyond, respectively. Conventional AGB estimates via activity data and EFs ranged from 1.67 to 2.42 Tg (mean: 2.10 Tg) across the ROI (Fig. 4c), generated using the same mask, and four representative EF values of 65.2, 62.2, 45.1, and 54.0 Mg/ha (IPCC default for African subtropical dry forests, Mozambique's Forest Reference Emission Levels, and literature on miombo woodlands, respectively[12,31–33]).

Comparison with GEDI

Overall, there was some agreement between MSL- and GEDI-derived AGB, with mean densities of 78.1 and 68.5 Mg/ha, respectively, for the 18,611 GEDI footprints available across the ROI, although it was observed that MSL estimates were generally larger for densities greater than 50 Mg/ha (Fig. 5b). These differences were explored by considering the African deciduous broadleaf forests model underlying GEDI-derived AGB (predictor variables: aboveground relative heights)[34], and simulating GEDI-perceived waveforms and metrics from the ALS data[35]. This provided insight into the performance of this model, whose predictions were typically larger for lower densities (< 50 Mg/ha) and vice versa for higher densities (Fig. 5c)
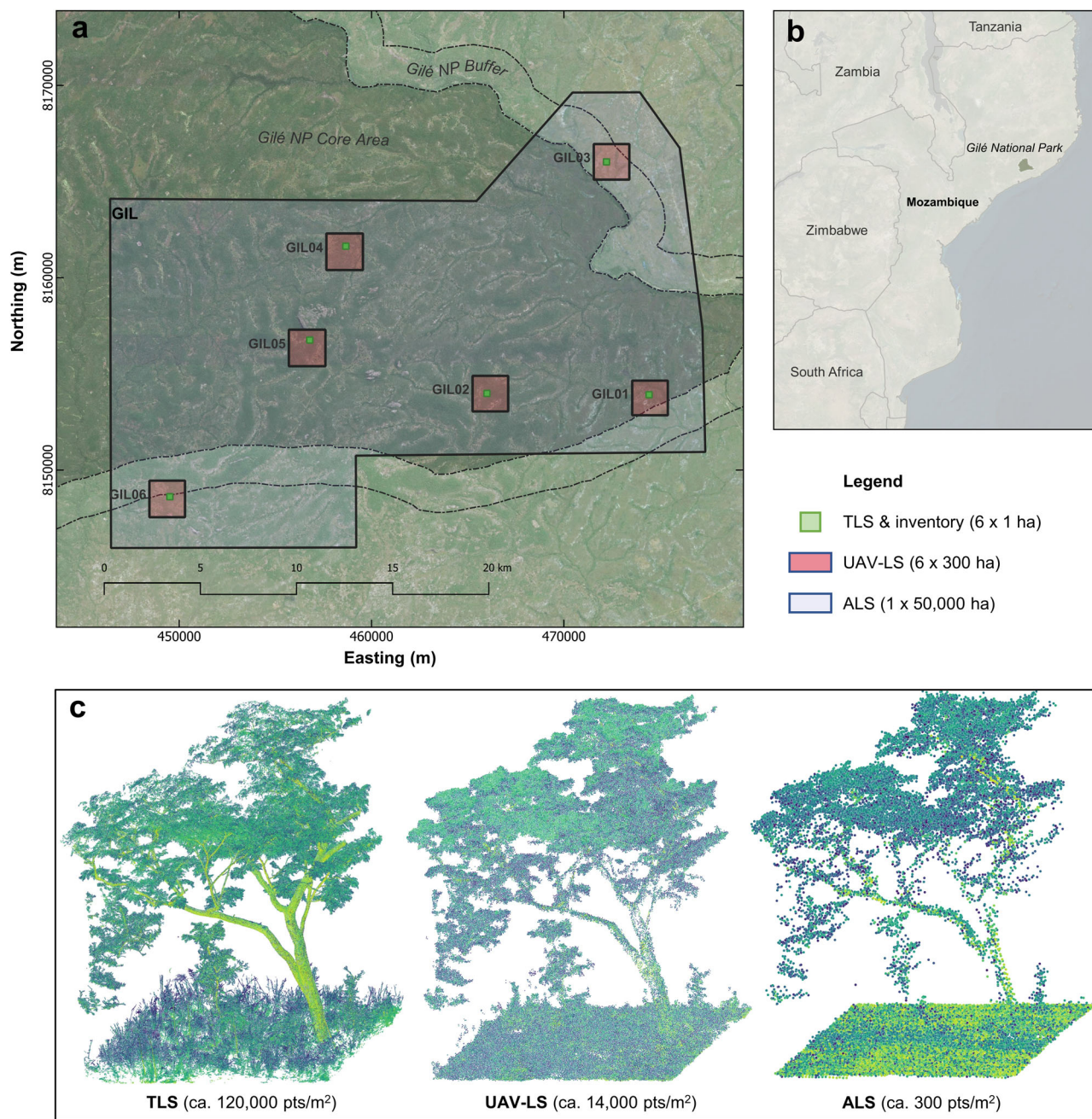
Fig. 1 | Multi-scale lidar across Gilé National Park, Mozambique.
a Approximately 450 billion laser scanning measurements were acquired in a 50 kha region of interest (ROI) located across the southeast corner of the park, capturing core area, buffer zone and beyond, such that the ROI encompassed intact miombo woodlands through to clearland (CRS is EPSG:32737). Helicopter-based airborne laser scanning (ALS) data were collected across its entirety, whilst slow-flying unoccupied aerial vehicle laser scanning (UAV-LS) data were acquired across six 300 ha sections. Terrestrial laser scanning (TLS) and conventional inventory data were collected in six 1 ha plots coinciding with these sections. b Location of Gilé National Park in wider Mozambique. c Example of coincident TLS, UAV-LS and ALS point clouds from a 10 m² section of forest (coloured by reflectance). Map data © 2024 Microsoft.

compared to MSL counterparts. That is, the overall agreement was in part owing to differences at high and low densities offsetting one another.

Discussion

Here, we presented the first region-scale mapping of forest AGB stocks driven by direct 3D measurements of forest structure, independent of conventional methods, including allometrics. Importantly, these estimates have a credible estimate of uncertainty being 11.0% of the region-scale AGB estimate itself. We note that even with these first-of-their-kind MSL measurements capturing samples of the structure of each individual tree across

the 50 kha region, pixel-level (10 m resolution) uncertainty frequently exceeded 60% (averaging out to approximately 36% and 27% when aggregating to 30 and 100 m resolution, respectively). That is, these miombo woodlands exhibit pronounced structural and woody tissue density variation, part of which remained uncaptured by either the MSL data themselves, or more likely, the developed processing methods. These maps then, are likely inappropriate for small-scale applications such as individual tree AGB estimation[36], but suitable for enabling accurate local and regional carbon accounting through calibration and validation of Earth observation instrumentation such as GEDI, and the upcoming ESA BIOMASS mission[37].
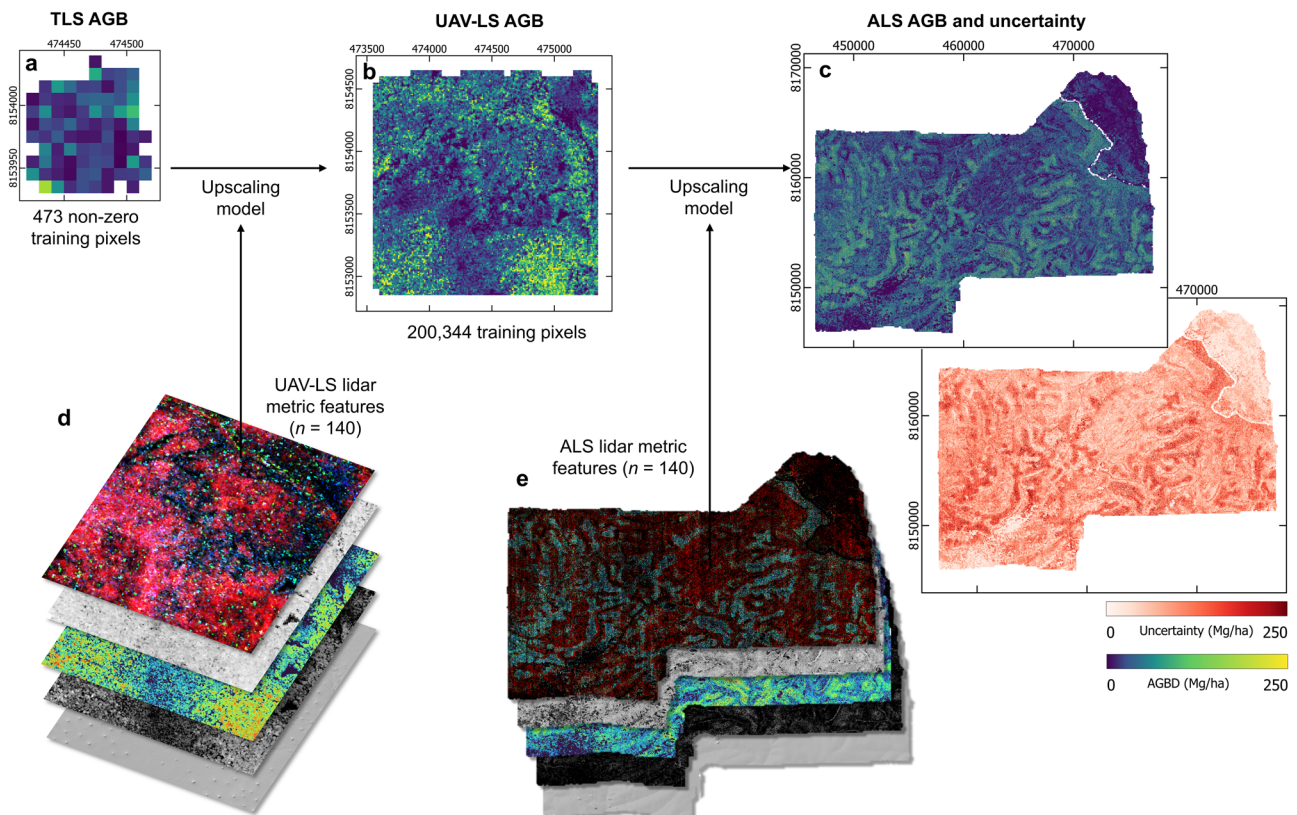
Fig. 2 | Estimating forest aboveground biomass from multi-scale lidar. a TLS-derived estimates of gridded AGB (10 m resolution) were generated for the six 1 ha plots (Fig. 1a) via quantitative structural models describing the woody architecture and volume of individual trees (Fig. 3a). **b** Upscaled AGB across the six 300 ha sections was estimated through gradient boosting machine learning, using TLS estimates as training data, and metrics of forest structure retrieved from the UAV-LS data as predictor variables. **c** This upscaling step was repeated to produce AGB estimates across the ROI, and also shown here are the uncertainties associated with each pixel prediction. Models were evaluated using spatial cross-validation methods, and uncertainty quantification captured components arising from both the upscaling and the underlying TLS training data. **d, e** Examples of the predictor variables generated from the UAV-LS and ALS point clouds including but not limited to canopy height, tree fractional cover and voxel occupancy rates (a proxy for the 3D distribution of woody volume).

The principal insight from these MSL methods is that 51–118% more AGB is stored across the ROI than predicted by conventional methods. A key driver of these differences was observed at the tree-level, where TLS-derived AGB estimates were systematically greater than allometric counterparts for large trees. This was also reflected in the GEDI analysis, where the AGB model for African deciduous broadleaf forests[34], itself underpinned by allometry, generally predicted lower values than MSL-derived methods for higher densities (> 50 Mg/ha). The importance of this point is magnified when considering the disproportionate contribution of large trees to upscaled AGB, as observed here (i.e., 11% of trees contributed greater than 50% of AGB across the six 1 ha stands) and described in literature[38]. We cannot definitively state whether this is due to over- or under-estimation of either method, or some mixture thereof, as accompanying destructive harvest data were not acquired. However, our TLS-derived estimates were calibrated using a representative sample of destructive measurements from the literature[23]. The trend for allometric and TLS methods to produce biased and unbiased estimates for large trees, respectively, is consistent with studies where estimates from both methods were coincident with destructive measurements across various forests[39,40].

The cause of this potential systematic underestimation remains an open question. One possibility however, is that all widely-used allometrics are modelled via log-transformed linear regression[41], and then applied to trees of all sizes (oftentimes caveated that any given model should not be used to predict an out-of-sample tree if its size falls outside the range of the calibration data[42]). It is implicitly assumed then, that model parameters are as equally suitable for large trees, as they are for small trees, within the context of the accuracy of predictions. However, the underlying calibration data are, as a rule, skewed towards smaller trees, often necessarily because of

the increasing difficulty of harvesting larger trees[43]. For example, for the miombo woodland specific and pan-tropical models considered here[29,30], the median stem diameter of the 167 and 4004 trees comprising the calibration data was 30 and 15 cm (mean: 35 and 24 cm), respectively. This compares with a median and mean stem diameter of 48 and 50 cm for the largest 10% of trees across the six 1 ha stands. The nature of linear regression whereby each observation is usually assigned equal weight[44], therefore suggests these aggregate models are unlikely to be representative of large trees, thus leading to biased predictions if this parameter invariance assumption is invalid[19]. Owing to the aforementioned context of large trees driving AGB distributions, and limited study on this subject[15,18,20], an argument from parsimony would be that large tree allometric predictions are less certain than small tree estimates, unless proven otherwise.

The MSL methods used here provide capabilities to resolve this issue and further enhance conventional methods. Whilst the airborne components presented here are at a more experimental stage, TLS methods are closer to operational readiness and less cost prohibitive. These data can be collected from 1 ha plots, within days, using sampling protocols complying with adopted good practices[45]. Data processing methods including segmentation and structural modelling are complex, but substantial progress has been made in recent years, particularly on both automation[46], and the avoidance of overestimating the volume of smaller trees and high-order branches[47]. TLS methods can be deployed in two ways: first, and most direct, estimating plot-scale AGB by summing contributions from individually modelled trees. This, applied across NFI networks for example, would enable updated EFs to be generated. Second, improving existing allometric models by augmenting their calibration datasets[48]. The key here would be generating uniform datasets (i.e., across tree size) through the addition of
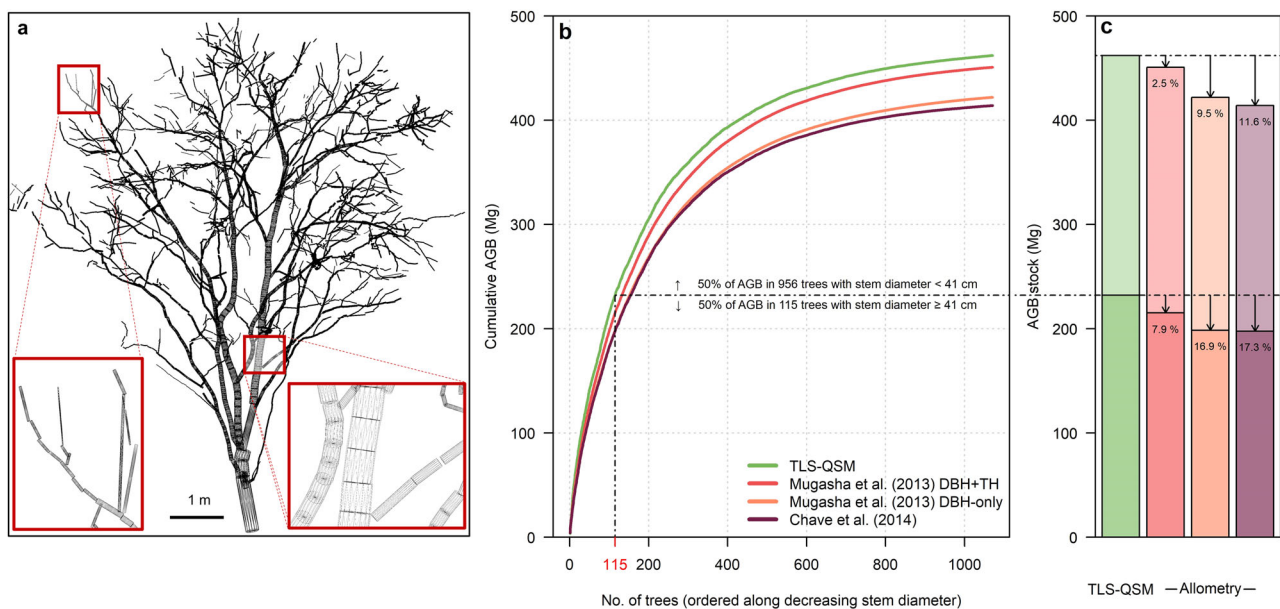
Fig. 3 | Deviations between large tree aboveground biomass estimates from terrestrial laser scanning and allometry. a Illustration of one quantitative structural model derived from the TLS point clouds (here of a *Pterocarpus angolensis*) that, coupled with species-specific basic woody tissue density, enables estimation of tree-scale AGB. **b** The cumulative distribution of TLS- and allometric-derived AGB across the 1,071 trees matched in both the TLS and inventory data across the six 1 ha plots (Fig. 1), ordered by decreasing stem diameter. Allometric estimates were generated from three appropriate models: two miombo woodland specific models[29] (predictor variables: stem diameter only [red], and stem diameter and tree height [orange]) and a pan-tropical allometry[30] (predictor variables: stem diameter, tree height and basic woody tissue density [purple]). **c** Summed AGB estimates across these trees, whereby the percentage decrease between TLS- and allometric estimates is shown. The dotted lines show the contribution to AGB from the 115 largest trees (by stem diameter), where it can be seen that: (i) these ~11% of trees contributed ~50% of summed AGB, and (ii) allometric estimates were systematically smaller than TLS counterparts.

larger trees. This is appealing as it would both reduce the uncertainty in highly-practicable allometric methods, and leverage the value of historic datasets. However, appending calibration datasets with TLS-derived AGB observations is non-trivial. Assumptions in linear regression include that the mean of the distribution of error in the dependent variable (i.e., AGB) is zero, and ideally errors are not autocorrelated or heteroscedastic[44]. Therefore, such efforts require thoughtful undertaking.

Returning then to the region-scale predictions, while the divergence between TLS- and allometric-derived AGB explains part of the overall difference, they are also driven by the selected EFs. That is, these values are not only underpinned by allometrics, but also the sampling pattern of their underlying field plots, and how that differs from the composition of the ROI considered here. This was partially unpicked by stratifying the ROI into core park, buffer zone and beyond, where it was observed that AGB densities were still an increase of 75%, 77%, and 53% more than the mean of the selected EFs. An important contributor here is the long-tail of large MSL-derived AGB observations driven by aforementioned large trees (e.g., predictions greater than 150 Mg/ha contributed 36% to the total 3.65 Tg), and the statistical likelihood these would be undersampled by randomly distributed field plots.

The question remains then: what are the implications of these observed differences in AGB stocks for our understanding of miombo woodlands? That largely depends on the transferability of our results to the world's 1.9 million km[2] of these forests[3]. The ROI where data were collected was deliberately positioned to capture as much of the range of states and successions, and therefore variance in AGB, across the wider region as possible. This capture of structural and taxonomic variation is illustrated by the inventory measurements across the six 1 ha plots (Table S1), where stem count, stem diameter and basal area ranged from 56 to 349, 10 to 75 cm and 1.9 to 20.9 m^2/ha, respectively, and that 81 of the estimated 334 unique species across miombo woodlands were observed, including from the dominant *Brachystegia* and *Julbernardia* genera[1]. Further, tree fractional cover across the 50 kha ranged from 0 to 1 with a mean of 0.59. These traits coincide with the ranges observed more widely across the continent[49], so we

therefore suggest it is not unreasonable to consider these sampled forests as being at least somewhat representative of miombo woodlands more broadly.

Speculatively then, if we were to extrapolate our results across the world's miombo woodlands, they potentially store in the region of 3.7 PgC more carbon in their AGB than currently estimated, assuming the mean of the considered EFs (56.6 Mg/ha) is uplifted by 74% (assuming 47% carbon content). It is also noteworthy that MSL methods detected an additional 0.20 Tg AGB stored across the ROI in land classified as non-forest, potentially increasing this delta still further, and emphasising that fragments of miombo woodlands have the potential to store significant quantities of carbon[50].

Whilst such extrapolation requires additional data for confirmation, the magnitude of this difference suggests our understanding of the role these forests play in global change requires a rethink, considering this overall increase is approaching the current annual global atmospheric increase (5.1 PgC/yr)[51]. That is, these forests could have a more potent ability to sequester carbon from afforestation and reforestation efforts, albeit equally the reverse, that their loss leads to increased emissions. Finally, an uplift in the carbon density of these forests per unit area could correspond to a proportional factor increase of 1.5 to 2.2 in their value on carbon markets, thus better incentivising their protection and restoration, and disincentivizing the value extracted from their deforestation[52].

Materials and methods
Site description
The 50 kha region of interest (ROI) where data were collected (Fig. 1a) was located on the southeastern border of Gilé National Park, Zambezia Province, Mozambique. The forests here feature woodlands, riverine forests, and wooded savannas, dominated by trees from the *Brachystegia* and *Julbernardia* genera, characteristic of the more broad classification of miombo woodland[1]. Mean annual precipitation is between 800 and 1000 mm, with a dry season May–October[53]. Mean monthly temperature varies from low teens to high thirties, the terrain is largely flat, and soils comprise sandy loam and sandy clay[54].

Fig. 4 | Multi-scale lidar predicts consistently larger aboveground biomass stocks across these miombo woodlands than conventional methods using activity data and emissions factors. a MSL-derived forest/non-forest map across the 50 kha region of interest, overlaid on the boundaries of Gilé National Park, generated by thresholding tree fractional cover greater than or equal to 30% using a canopy threshold of 5 m at 10 m resolution. **b** Distribution and mean (green) of MSL-derived aboveground biomass density predictions (Fig. 2) for pixels considered forested, versus a representative selection of four EFs (red, orange, purple and blue) taken from IPCC defaults, Mozambique's Forest Reference Emission Levels, and literature on miombo woodlands[12,31–33]. **c** Summed AGB stocks across the ROI (including uncertainty in the MSL-derived estimate), whereby these EFs were combined with the FNF map. It is observed that the MSL approach estimated AGB stocks between 51 and 118% larger than these conventional methods, with a mean increase of 74%. Map data © 2024 Microsoft.
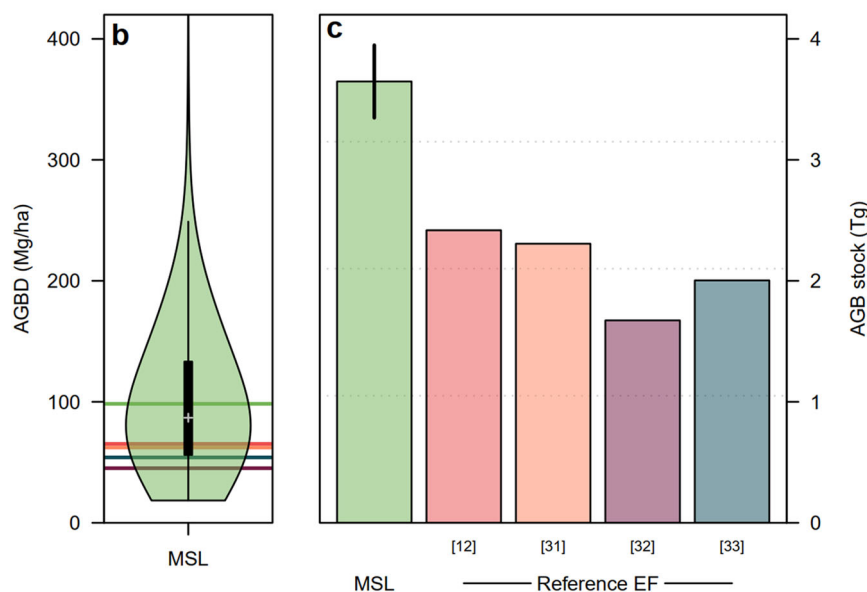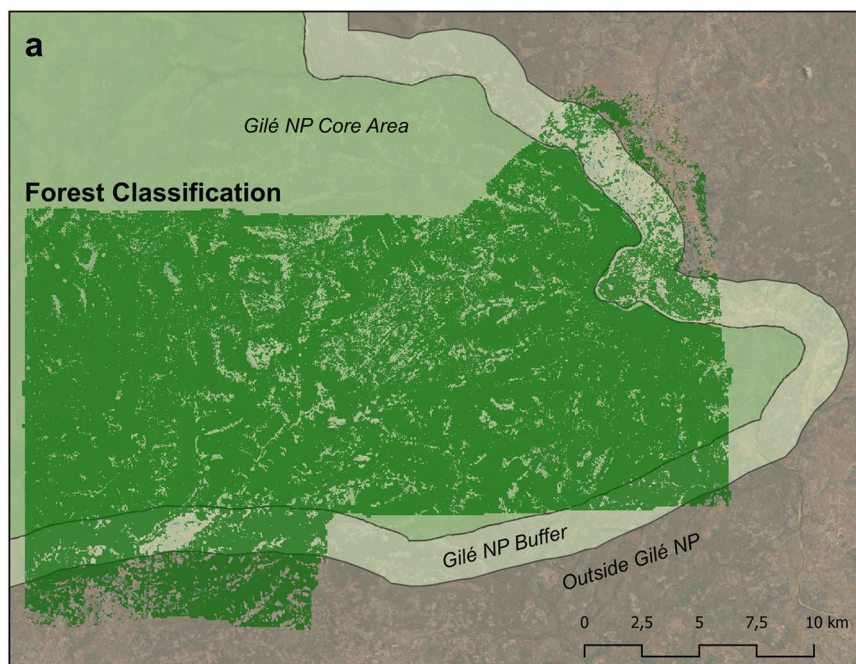
Study design

The ROI was positioned such that it covered core park, buffer zone and beyond (Fig. 4a). The dataset (Fig. 1a) comprised airborne laser scanning (ALS) data across the entirety of the ROI (designation: GIL), unoccupied aerial vehicle laser scanning (UAV-LS) data from six 300 ha sections (designation: GIL01 to GIL06), and terrestrial laser scanning (TLS) measurements and inventory data from six 1 ha plots coincident with these sections (designation: GIL01-01 to GIL06-01). These sections and plots were strategically located to capture variations in forest state, succession, structure, and taxonomy across the ROI.

Data collection

Data were acquired between June-November 2022. The six 100×100 m planimetric plots were established and inventoried using RAINFOR protocol[55]. Measurements of each tree inside these plots with stem diameter ≥ 10 cm included: (i) stem diameter via a circumference/diameter tape, (ii) point of measurement of the stem diameter (either 1.3 m above

ground or 0.5 m above-buttress), (iii) taxonomic identity determined by a single trained botanist, and (iv) x-y coordinates using eye-estimation.

TLS data were collected in GIL01-01 through GIL06-01 using a RIEGL VZ-400i laser scanner. Sampling followed established protocol[56]. That is, in accordance with the CEOS Aboveground Woody Biomass Product Validation Good Practices Protocol[45]. In particular, scans were acquired from 121 locations at 10 m intervals across each plot, with upright and tilt scans acquired at each location to capture a complete sample of the scene. The instrument pulse repetition rate was 300 kHz and the angular step between sequentially fired pulses was 0.04 degrees. The laser pulse had a wavelength, pulse width, beam divergence and exit footprint diameter of 1550 nm, 3.0 ns, 0.30 mrad, and 7.0 mm, respectively. Coarse georeferencing of scans was generated from an onboard GNSS receiver obtaining real-time differential corrections from a nearby static Emlid Reach RS2 GNSS receiver.

UAV-LS data were acquired across GIL01 through GIL06 using a RIEGL VUX-120 and Trimble Applanix APX-20 GNSS/INS kinematic laser scanning system mounted on a hybrid-electric drone, in a 50 m

Fig. 5 | Comparison between aboveground biomass estimates from multi-scale lidar and the GEDI spaceborne lidar mission. a Illustration of the 18,611 GEDI footprints available across the ROI overlaid on the MSL-derived predictions (Fig. 2). **b** Comparison between AGB density estimated via MSL and GEDI methods for these footprints. GEDI estimates are derived from a model for deciduous broadleaved trees in Africa (DBT.Af) using aboveground relative heights as predictor variables[34]. It is observed that MSL predictions are generally larger for densities greater than 50 Mg/ha. **c** This difference was examined by simulating GEDI waveforms and retrieving GEDI-perceived relative height metrics from the ALS data[35], that were subsequently used to generate estimates of AGB density from the DBT.Af model. This subplot compares these estimates with MSL-derived AGB, whereby they are typically larger for lower densities and smaller for higher densities. That is, the agreement between MSL- and GEDI-derived AGB is part owing to these differences offsetting one another. Both scatter plots comprise an identity line (green), linear regression with free intercept (black) and statistics including concordance correlation coefficient (CCC), and root mean square difference (RMSD).
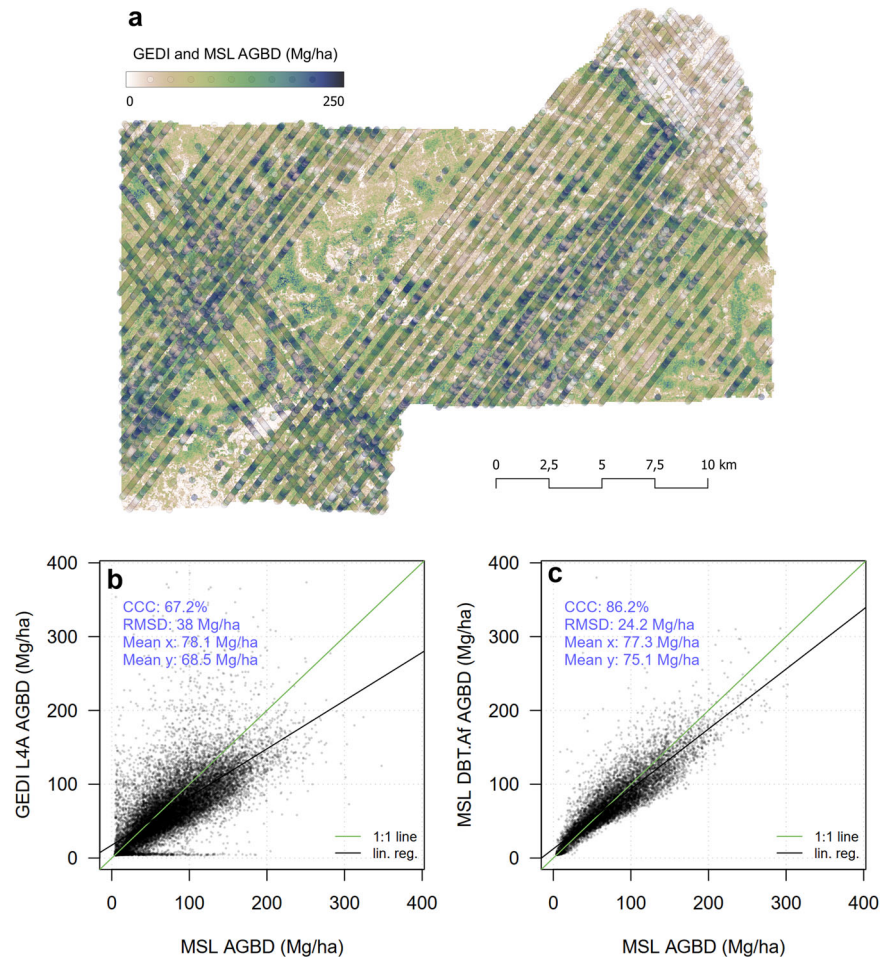
double-gridded configuration at 5.0 m/s velocity with an aboveground level of 100 m. The instrument pulse repetition rate, field of view and scan rate were 1800 kHz, 90 degrees and 315 lines per second, respectively. The laser pulse had a wavelength, pulse width, beam divergence and exit footprint diameter of 1550 nm, 3.0 ns, 0.38 mrad and 5.7 mm, respectively.

ALS data were acquired across GIL using the same kinematic laser scanning system mounted on a helicopter, in a 127 m spaced parallel line configuration at 41.2 m/s velocity with an aboveground level of 160 m. In this configuration, the pulse repetition rate, field of view and scan rate were 1200 kHz, 100 degrees and 396 lines per second, respectively. A nearby static Stonex S900A GNSS receiver collected observables throughout UAV-LS and ALS data acquisition for georeferencing purposes.

Data preprocessing
Inventory data (1406 trees) were manually digitised from field sheets, with accuracy assessed via a second operator digitising a randomly selected 5% subset. Errors in taxonomic identity were resolved using the Taxonomic Name Resolution Service[57]. Estimates of basic woody tissue density were derived from the mean of entries available in the Global Wood Density Database[58], whereby attribution was made if possible at the species-, or else genus-level (84.4 and 14.2% of the trees, respectively). If no taxonomic attribution could be made, basal area-weighted plot average wood densities were used (1.4%).

TLS data were co-registered into georeferenced (EPSG: 32737) tiled point clouds (Fig. 1c) using RIEGL RiSCAN Pro (v2.15) via its Automatic Registration 2 and Multi Station Adjustment 2 modules. Airborne mission trajectories were refined to survey-grade accuracy and precision via Applanix POSPac UAV (v8.8) using GNSS observables from the base station, whose absolute positioning was refined using the AUSPOS service[59].

Lidar data were united with these trajectories and merged into georeferenced tiled point clouds (Fig. 1c) using RIEGL RiPROCESS (v1.9.2.2) including its RiUNITE (v1.0.3.3) and RiPRECISION (v1.4.2) modules. TLS point cloud georeferencing was refined by aligning with the UAV-LS data using an iterative closest point algorithm implemented in CloudCompare (v2.12.4). Noise in TLS, UAV-LS and ALS point clouds was labelled using reflectance and deviation thresholding[60], and statistical outlier filtering. Incompletely sampled tiles were discarded based on point density and morphological erosion.

Data processing
TLS-derived tree-scale aboveground biomass (AGB) estimates were generated in six steps. First, point clouds representing individual trees either inside, or part of whose AGB fell inside the plot, were segmented from the tiled point clouds (1,339 trees). This was undertaken both manually in CloudCompare (v2.12.4) and using the Forest Structural Complexity Tool[46]. Second, point clouds were manually linked with inventory data via stem maps, whereby due to edge effects around the plot (i.e. trees with stem base inside but crown partly growing outside the plot, and vice versa) and multi-stemmed trees, there were somewhat fewer point clouds than census entries (1339 vs 1406 respectively). Third, leafy material, which had distinctively lower apparent reflectance than the woody material, was segmented from point clouds via thresholding based on evaluating single-tree reflectance histograms. Fourth, quantitative structural models (QSMs) (Fig. 3a) were constructed for woody point clouds using TreeQSM[21] (v2.4.1). QSMs were inspected by eye, and validated via comparison to the input point clouds (i.e., point-to-cylinder distances). Fifth, potential overestimations of small branch volume arising from wind, co-registration error, properties of the laser pulse itself, or some mixture thereof, were negated using a post-

processing step based on metabolic scaling theory[61]. This limited the maximum diameter of third order and above branches to half the diameter of their parent, and the diameter of cylinders intra-branch to no more than their parent. This was validated by comparing open-access data from previous studies where destructive data were available[23]. Sixth, AGB was estimated from QSM-derived volume and basic woody tissue density, the latter obtained from the established link between point clouds and inventory data.

Metrics describing forest structure (Fig. 2d, e) were generated from the UAV-LS and ALS point clouds at 10 m resolution unless stated otherwise. The metrics, described in[25], and retrieved using methods similar to those implemented in lidR[62], comprised: digital terrain and canopy height models (1 m), relative height, tree fractional cover, canopy height rugosity, fixed and variable gap fraction, canopy closure, canopy ratio, z-entropy, skewness and kurtosis. Additionally, voxel-based metrics describing the 3D distribution of woody plant material were also retrieved[26]. This was undertaken by segmenting leafy material via reflectance thresholding, and partitioning the UAV-LS and ALS point clouds into voxels with 0.1 and 0.5 m edge length, respectively. Volumes of each voxel comprising at least one point were then aggregated in 3D grids of 1 m and 5 m resolution, respectively.

Aboveground biomass modelling

TLS-derived gridded estimates of AGB across GIL01-01 through GIL06-01 (Fig. 2a) were estimated by constructing a georeferenced 10 m grid across each plot, decomposing each QSM into its constituent cylinders, and allocating volume to respective cells. Numerical approximation estimated the quantity of intra-cylinder volume assigned to multiple cells. Cells were considered valid only when all AGB from trees with stem diameter ≥10 cm was captured. This included contributions from trees outside the plot whose stem or crown only partially fell inside the cell in question. For such trees, QSMs were produced with the encroaching cylinder volumes attributed to the relevant cell, assuming a wood density equal to the plot basal area-weighted wood density. In total, 568 cells, of which 473 were non-zero, were created with biomass contributions from 1,259 QSMs.

Gridded estimates of AGB (10 m resolution) across GIL01 through GIL06 (Fig. 2b) were retrieved using extreme gradient boosting machine learning via XGBoost[63] (v1.6.2), which has previously been applied to AGB modelling[24]. The TLS-derived gridded estimates of AGB were used as training data, and the spatially coincident UAV-LS-derived forest structure metrics were used as predictor variables. The choice of metric and the produced biomass map resolution (10 m) was informed by balancing having plentiful training pixels (i.e. higher resolution) versus the information contained within each pixel to predict biomass, using a rule of thumb that there be at least ten times more training pixels than features. Optimised hyperparameters and feature selection were found by minimising the root mean square error of validation folds within a spatial cross-validation framework[27] using GIL01-01 through GIL06-01 as separate folds. For this, a random grid search of the following six hyperparameters was undertaken: (i) learning rate (a step shrinkage weight; used to make the boosting process more conservative and reduce the risk of overfitting); (ii) minimum loss reduction (a decision parameter to make a partition on a leaf node of the tree); (iii) maximum depth of a tree (a proxy for the complexity, and hence overfitting risk, of the model); (iv) minimum child weight (minimum sum of the hessian required in each child, with higher values again resulting in more conservative models); (v) subsample of the training instances (fraction sampled from the training data prior to tree growing); and (vi) column sampling ratio (fraction of the features to be subsampled). Further performance metrics were also generated, including bias, and two based on the log of the accuracy ratio: (i) median symmetric accuracy, and (ii) symmetric signed percentage bias. These two metrics are well-suited to assessing predictions potentially spanning several orders of magnitude[64].

Gridded estimates of AGB (10 m resolution) across GIL (Fig. 2c) were retrieved by repeating this step, where the UAV-LS-derived gridded estimates of AGB were the training data, and the spatially coincident ALS-derived forest structure metrics were the predictor variables. Fig. S1 and

Fig. S2 present cross-validation statistics, the weight and gain assigned to each predictor variable, and scatter plots illustrating predicted versus reference AGB, for the UAV-LS and ALS models, respectively.

A direct TLS-to-ALS upscaling model was also tested (i.e., skipping the intermediate UAV-LS layer). The direct model was slightly more biased (cross-validation bias with TLS labels: −3.63%) and less accurate (cross-validation RMSE with TLS labels: 62.6 Mg/ha), thus not considered further in this study. Further methodological improvements to the MSL workflow could encompass more sophisticated lidar features, perhaps tailored to the extreme high point density of UAV-LS, that are likely to correlate more with AGB than the ones used in our study. Additionally, spatially explicit models, such as convolutional neural networks, can take into account spatial context rather than only the pixel values themselves. Here, spatial information is only used in the cross-validation of the XGBoost models.

Uncertainty quantification

Uncertainty in TLS-derived tree-scale AGB arises from the underlying point cloud itself, quantitative structural modelling and basic woody tissue density estimation[43]. Uncertainty from these sources was implicitly captured by modelling the expected distribution of error using existing data where TLS-derived AGB estimates were available alongside reference measurements derived from destructive harvesting and weighing (391 trees from 111 species)[23]. To increase the representativity of these harvested trees to the ROI, the dataset was subsetted to contain only trees with stem diameter <75 cm, scanned in leaf-on conditions, and to exclude trees from boreal and temperate regions (n = 174). The error distribution was modelled via the mean and variance of residuals, as a function of stem diameter, using linear and non-linear quantile regression, respectively. The appropriate mean residual was subtracted from the raw AGB estimate for each tree to remove bias known to arise in TLS-derived volume estimates, especially in smaller branches[47], complementing the QSM postprocessing based on metabolic scaling. Uncertainty in TLS-derived gridded AGB was derived as a volume-weighted combination of tree-scale AGB uncertainty, by modelling the true AGB of each tree as following a Gaussian distribution dependent on its TLS-estimated AGB, and independent of all other trees. This assumption of independence refers only to effects causing a discrepancy between TLS-estimated AGB and true AGB (i.e., imperfections in lidar scanning, QSM reconstruction and wood density assignment) and does not neglect spatial correlation of true AGB (e.g., arising from the effect of trees on each other), which is inherited by TLS-estimated AGB, but remains an approximation since spatial correlation may exist in some residual effects such as wind noise. Consequently, the only modelled source of covariance is due to a given tree spanning more than one pixel. This enabled calculation of the full pixel covariance matrix, capturing both the marginal uncertainty in each pixel and the correlation between pixels.

Uncertainty in UAV-LS-derived gridded estimates of AGB was modelled as the sum of two independent components: measurement variance and model variance. Measurement variance was estimated using a Monte Carlo random sampling approach. 100 samples of each of the six TLS-derived gridded AGB estimates were generated from their underlying multivariate Gaussian distribution, described above. Each set of six gridded AGB estimates was used to train a separate XGBoost model, producing 100 gridded predictions of AGB for each UAV-LS section, the sample variance across which was taken as the measurement variance, per pixel. Model variance was estimated as a zero-intercept linear function of predicted pixel AGB, with slope calibrated from the cross-validation data described in the previous section. Uncertainty in ALS-derived gridded estimates of AGB (Fig. 2c) was estimated by repeating this process, using the 100 UAV-LS-derived gridded AGB estimates as training data to a further ensemble of 100 XGBoost models. Uncertainty is expressed as 90% confidence intervals throughout.

Conventional methods

Allometric-derived tree-scale AGB estimates were produced by three allometric models. These models are widely used for carbon stock estimation in the region. First, the pan-tropical allometry described in[30] that considers the

predictor variables stem diameter, tree height and basic woody tissue density, itself calibrated from the harvest of 4,004 trees across the tropics and subtropics. Second, two miombo woodland specific allometries described in[29], themselves calibrated from the harvest of 167 trees in Tanzania, that consider: (i) stem diameter only, and (ii) stem diameter and tree height. Tree height was derived from TLS data. Individual trees with non-matching stem diameters between the inventory and TLS were excluded from the tree-level AGB comparison (i.e., stem diameters with >5 cm difference; $n = 188$) to ensure potential errors in linking both data sets were omitted.

A selection of four representative emission factors (EFs) were gathered to enable conventional region-scale estimation of AGB. We used values from: (i) the IPCC default for African subtropical dry forest described in[12] (using a combination of L- and C-band radar); (ii) Mozambique's Forest Reference Emission Levels[31] for semi-deciduous forest including Miombo (based on the country's National Forest Inventory and the allometrics described in[29]); (iii) specific to Zambezia province described in[32] (obtained from L-band radar and a network of forest plots using the allometrics of[30]); and (iv) Mozambique-wide described in[33] (destructive harvesting coupled to 27 ha forest inventory).

Comparison with GEDI

GEDI L2A (version 2) and L4A (version 2.1) products (i.e., relative height metrics and AGB density, respectively) were downloaded from NASA's Earthdata[65]. These data were filtered to only include observations between day-of-year 275 and 365 (to match seasonality) from the years 2018–2022, with a sensitivity greater than 90%, spatially overlapping with the ALS data, and with 98% relative heights under 35 m. This resulted in 18,611 GEDI observations (Fig. 5a) from the available 72,101. Coincident MSL-derived AGBD was retrieved by simulating a circle at the GEDI footprint coordinates (12.5 m radius), and performing a weighted average extraction from the gridded 10 m resolution lidar-derived AGB predictions. No geospatial aligning of GEDI footprints with the ALS data was applied.

Additionally, we tested the influence of the AGB model underpinning GEDI's L4 product in the region (i.e. the African dry broadleaf forest model (DBT.Af)[34]). Since relative heights retrieved from the ALS data (calculated from the height distributions within a point cloud) are not comparable to GEDI-derived relative height (derived from a waveform from a single laser pulse), we simulated the ALS data into GEDI-perceived waveforms for 10,000 randomly distributed 25 m diameter point cloud sections within the ROI using gediRat[35]. These resulting simulated waveforms were directly comparable with GEDI's waveforms and thus were used as input to GEDI's DBT.Af model, to predict AGBD.

Data availability

Gridded estimates of AGB and its uncertainty (10 m resolution) from the TLS, UAV-LS and ALS data, and gridded metrics of forest structure (10 m resolution) from the UAV-LS and ALS data, for the section GIL04, and the data to produce the graphs and charts, are available at https://doi.org/10.5281/zenodo.11072918. These data are distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Code availability

Semantic and instance segmentation was undertaken using a modified version of *FSCT*[46]. Quantitative structural models were generated using a modified version of *TreeQSM*[21]. Forest structure metrics were retrieved via a C++ implementation of *lidR*[62]. Extreme gradient boosting models were constructed using *XGBoost*[63]. These underlying packages are available at https://github.com/SKrisanski/FSCT, https://github.com/InverseTampere/TreeQSM, https://github.com/r-lidar/lidR and https://github.com/dmlc/xgboost, respectively.

References

1. Abbot, J. et al. *The Miombo in transition: woodlands and welfare in Africa. The Miombo in Transition: Woodlands and Welfare in Africa* (Center for International Forestry Research CIFOR, 1996). https://doi.org/10.17528/cifor/000465.
2. Ryan, C. M. et al. Ecosystem services from southern African woodlands and their future under global change. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150312 (2016).
3. Ribeiro, N. S., Silva de Miranda, P. L. & Timberlake, J. Biogeography and Ecology of Miombo Woodlands. in *Miombo Woodlands in a Changing Environment: Securing the Resilience and Sustainability of People and Woodlands* (eds. Ribeiro, N. S., Katerere, Y., Chirwa, P. W. & Grundy, I. M.) 9–53 (Springer International Publishing, 2020). https://doi.org/10.1007/978-3-030-50104-4_2.
4. 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).
5. Houghton, R. A. Aboveground forest biomass and the global carbon balance. *Glob. Chang. Biol.* **11**, 945–958 (2005).
6. Pan, Y. et al. A large and persistent carbon sink in the World's forests. *Science* **333**, 988–993 (2011).
7. Yanai, R. D. et al. Improving uncertainty in forest carbon accounting for REDD+ mitigation efforts. *Environ. Res. Lett.* **15**, 124002 (2020).
8. Grassi, G. et al. The key role of forests in meeting climate targets requires science for credible mitigation. *Nat. Clim. Chang.* **7**, 220–226 (2017).
9. Clark, D. B. & Kellner, J. R. Tropical forest biomass estimation and the fallacy of misplaced concreteness. *J. Veg. Sci.* **23**, 1191–1196 (2012).
10. Giménez, M. G. et al. *Assessment of Innovative Technologies and Their Readiness for Remote Sensing-Based Estimation of Forest Carbon Stocks and Dynamics*. Report no. 160649, 1-41. http://documents.worldbank.org/curated/en/305171624007704483/Assessment-of-Innovative-Technologies-and-Their-Readiness-for-Remote-Sensing-Based-Estimation-of-Forest-Carbon-Stocks-and-Dynamics (2021).
11. Espejo, A. et al. *Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and guidance from the Global Forest Observations Initiative*. https://www.reddcompass.org/mgd/resources/GFOI-MGD-3.1-en.pdf (2020).
12. Rozendaal, D. M. A. et al. Aboveground forest biomass varies across continents, ecological zones and successional stages: refined IPCC default values for tropical and subtropical forests. *Environ. Res. Lett.* **17**, 014047 (2022).
13. Maniatis, D. & Mollicone, D. Options for sampling and stratification for national forest inventories to implement REDD+ under the UNFCCC. *Carbon Balance Manag.* **5**, 9 (2010).
14. Brown, S., Gillespie, A. J. R. & Lugo, A. E. Biomass estimation methods for tropical forests with applications to forest inventory data. *For. Sci.* **35**, 881–902 (1989).
15. Burt, A. et al. Assessment of bias in pan-tropical biomass predictions. *Front. For. Glob. Chang.* **3**, 12 (2020).
16. 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).
17. Picard, N., Boyemba Bosela, F. & Rossi, V. Reducing the error in biomass estimates strongly depends on model selection. *Ann. For. Sci.* **72**, 811–823 (2015).
18. Picard, N., Rutishauser, E., Ploton, P., Ngomanda, A. & Henry, M. Should tree biomass allometry be restricted to power models?. *For. Ecol. Manage.* **353**, 156–163 (2015).
19. Zhou, X. et al. Dynamic allometric scaling of tree biomass and size. *Nat. Plants* **7**, 42–49 (2021).

20. Calders, K. et al. Laser scanning reveals potential underestimation of biomass carbon in temperate forest. *Ecol. Solut. Evid.* **3**, 1–30 (2022).

21. Raumonen, P. et al. Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sens.* **5**, 491–520 (2013).

22. Chave, J. et al. Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12**, 351–366 (2009).

23. Demol, M. et al. Estimating forest above-ground biomass with terrestrial laser scanning: Current status and future directions. *Methods Ecol. Evol.* **13**, 1628–1639 (2022).

24. Li, Y., Li, M., Li, C. & Liu, Z. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Sci. Rep.* **10**, 1–12 (2020).

25. McNicol, I. M. et al. To what extent can UAV photogrammetry replicate UAV LiDAR to determine forest structure? A test in two contrasting tropical forests. *J. Geophys. Res. Biogeosci.* **126**, 1–17 (2021).

26. Whelan, A. W., Cannon, J. B., Bigelow, S. W., Rutledge, B. T. & Sánchez Meador, A. J. Improving generalized models of forest structure in complex forest types using area- and voxel-based approaches from lidar. *Remote Sens. Environ.* **284**, 113362 (2023).

27. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 1–11 (2020).

28. Dubayah, R. et al. GEDI launches a new era of biomass inference from space. *Environ. Res. Lett.* **17**, 095001 (2022).

29. Mugasha, W. A. et al. Allometric models for prediction of above- and belowground biomass of trees in the miombo woodlands of Tanzania. *For. Ecol. Manage.* **310**, 87–101 (2013).

30. Chave, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **20**, 3177–3190 (2014).

31. Governo de Moçambique. *Mozambique's Forest Reference Emission Level for Reducing Emissions from Deforestation in Natural Forests*. https://redd.unfccc.int/media/2018_frel_submission_mozambique.pdf (2018).

32. McNicol, I. M., Ryan, C. M. & Mitchard, E. T. A. Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands. *Nat. Commun.* **9**, 3045 (2018).

33. Ryan, C. M., Williams, M. & Grace, J. Above- and belowground carbon stocks in a miombo woodland landscape of mozambique. *Biotropica* **43**, 423–432 (2011).

34. Kellner, J. R., Armston, J. & Duncanson, L. Algorithm theoretical basis document for GEDI footprint aboveground biomass density. *Earth Sp. Sci.* **10**, 1–20 (2023).

35. Hancock, S. et al. The GEDI simulator: a large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth Sp. Sci.* **6**, 294–310 (2019).

36. Mugabowindekwe, M. et al. Nation-wide mapping of tree-level aboveground carbon stocks in Rwanda. *Nat. Clim. Chang.* **13**, 91–97 (2023).

37. Quegan, S. et al. The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space. *Remote Sens. Environ.* **227**, 44–60 (2019).

38. Slik, J. W. F. et al. Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics. *Glob. Ecol. Biogeogr.* **22**, 1261–1271 (2013).

39. Calders, K. et al. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods Ecol. Evol.* **6**, 198–208 (2015).

40. Gonzalez de Tanago, J. et al. Estimation of above-ground biomass of large tropical trees with terrestrial LiDAR. *Methods Ecol. Evol.* **9**, 223–234 (2018).

41. Kerkhoff, A. J. & Enquist, B. J. Multiplicative by nature: why logarithmic transformation is necessary in allometry. *J. Theor. Biol.* **257**, 519–521 (2009).

42. Chave, J. et al. Error propagation and scaling for tropical forest biomass estimates. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **359**, 409–420 (2004).

43. Burt, A. et al. New insights into large tropical tree mass and structure from direct harvest and terrestrial lidar. *R. Soc. Open Sci.* **8**, 201458 (2021).

44. Hayashi, F. *Econometrics*. (Princeton University Press, 2000).

45. Duncanson, L. et al. Aboveground woody biomass product validation good practices protocol. Version 1.0. *Good Pract. Satell. Deriv. L. Prod. Valid.* 1–236 https://doi.org/10.5067/doc/ceoswgcv/lpv/agb.001 (2021).

46. Krisanski, S., Taskhiri, M. S., Gonzalez Aracil, S., Herries, D. & Turner, P. Sensor agnostic semantic segmentation of structurally diverse and complex forest point clouds using deep learning. *Remote Sens.* **13**, 1413 (2021).

47. Abegg, M., Bösch, R., Kükenbrink, D. & Morsdorf, F. Tree volume estimation with terrestrial laser scanning — Testing for bias in a 3D virtual environment. *Agric. For. Meteorol.* **331**, 109348 (2023).

48. Momo Takoudjou, S. et al. Using terrestrial laser scanning data to estimate large tropical trees biomass and calibrate allometric models: A comparison with traditional destructive approach. *Methods Ecol. Evol.* **9**, 905–916 (2018).

49. The SEOSAW partnership A network to understand the changing socio-ecology of the southern African woodlands (SEOSAW): challenges, benefits, and methods. *Plants People Planet* **3**, 249–267 (2021).

50. Reiner, F. et al. More than one quarter of Africa's tree cover is found outside areas previously classified as forest. *Nat. Commun.* **14**, 2258 (2023).

51. Friedlingstein, P. et al. Global carbon budget 2021. *Earth Syst. Sci. Data* **14**, 1917–2005 (2022).

52. Austin, K. G. et al. The economic costs of planting, preserving, and managing the world's forests to mitigate climate change. *Nat. Commun.* **11**, 5946 (2020).

53. Fusari, A., Lamarque, F., Chardonnet, P. & Boulet, H. *Reserva Nacional do Gilé: Plano de Maneio 2012 − 2021*. 1–143. https://www.biofund.org.mz/wp-content/uploads/2015/03/PM-RNG-2012-2021.pdf (2010).

54. Montfort, F. et al. Regeneration capacities of woody species biodiversity and soil properties in Miombo woodland after slash-and-burn agriculture in Mozambique. *For. Ecol. Manage.* **488**, 119039 (2021).

55. Phillips, O., Baker, T., Feldpausch, T. R. & Brienen, R. *RAINFOR: Field manual for plot establishment and remeasurement. 1–30. RAINFOR, PAN-AMAZONIA Project* https://forestplots.net/upload/ManualsEnglish/RAINFOR_field_manual_EN.pdf (2021).

56. Wilkes, P. et al. Data acquisition considerations for Terrestrial Laser Scanning of forest plots. *Remote Sens. Environ.* **196**, 140–153 (2017).

57. Boyle, B. et al. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinf.* **14**, 16 (2013).

58. Zanne, A. E. et al. Data from: Towards a worldwide wood economics spectrum. *Dryad* https://doi.org/10.5061/dryad.234 (2009).

59. Janssen, V. & McElroy, S. A Practical Guide to AUSPOS. *Proceedings of the 25th Association of Public Authority Surveyors Conference*, 3–28 (2022).

60. Pfennigbauer, M. & Ullrich, A. Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement. in *Laser Radar Technology and Applications XV* (eds. Turner, M. D. & Kamerman, G. W.) vol. 7684 76841F (2010).

61. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664–667 (1999).

62. Roussel, J. R. et al. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* **251**, 112061 (2020).

63. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785–794 (2016).

64. Morley, S. K., Brito, T. V. & Welling, D. T. Measures of model performance based on the log accuracy ratio. *Sp. Weather* **16**, 69–88 (2018).
65. Dubayah, R. et al. GEDI L4A footprint level aboveground biomass density. *Version 2.1.* https://doi.org/10.3334/ORNLDAAC/2056 (2022).

Author contributions
Study design: M. Demol, A.E., A.F., H.O., A.P., E.P.-R., A.B. Data acquisition: E.E., H.H., S.L., M.L., V.M., F.M., B.L.P.V., C.S., B.W. Data analysis: M. Demol, G.B., E.E., H.H., A.L., A.P., T.P., B.L.P.V., C.S., R.U., B.W., A.B. Interpretation of results: M. Demol, N.A.-A., G.B., M. Disney, L.D., S.H., J.H., M.L., P.R.-V., R.U., A.B. Writing—original draft: M. Demol, A.B. Writing—review & editing: all authors.

Competing interests
M. Demol, G.B., E.E., A.F., J.H., H.H., A.L., H.O., A.P., T.P., B.L.P.V., P.R.-V., C.S., R.U., B.W., and A.B. are employees and/or shareowners of Sylvera Ltd. N.A.-A., A.E., and E.P.-R. are employees of the World Bank's Forest Carbon Partnership Facility. The interpretation of the results reflects the position of the authors and has not been endorsed by the World Bank. All other authors have no competing interests to declare.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-024-01448-x.

Correspondence and requests for materials should be addressed to Miro Demol.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Alireza Bahadori and Aliénor Lavergne. A peer review file is available.

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

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