Past decade above-ground biomass change comparisons from four multi-temporal global maps

Arnán Araza a,⁎, Martin Herold b,c, Sytze de Bruin a, Philippe Ciais d, David A. Gibbs e, Nancy Harris f, Maurizio Santoro g, Jean-Pierre Wigneron h, Hui Yang i, Natalia Málaga a, Karimón Nesha a, Pedro Rodríguez-Veiga j,k,l, Olga Brovkina j, Hugh C.A. Brown k,l, Milén Chanov m, Zlatomir Dimitrov m, Lachezar Filchev m, Jonas Fridman n, Mariano García o, Alexander Gikov m, Leen Govaere p, Petar Dimitrov m, Fardin Moradi q, Adriane Esquivel Muellert f, Jan Novotný j, Thomas A.M. Pugh s,t, Mart-Jan Schelhaas u, Dmitry Schepaschenko v, Krzysztof Stereńczak w, Lars Hein b

⁎ Corresponding author at: Laboratory of Geo-Information and Remote Sensing, Wageningen University and Research, The Netherlands.

A R T I C L E   I N F O

Dataset link: http://dx.doi.org/10.6084/m9.figshare.22349218

Keywords:
Above-ground biomass
Above-ground biomass change
Carbon flux
Map assessment
Global carbon cycle
Earth observation

A B S T R A C T

Above-ground biomass (AGB) is considered an essential climate variable that underpins our knowledge and information about the role of forests in mitigating climate change. The availability of satellite-based AGB and AGB change (ΔAGB) products has increased in recent years. Here we assessed the past decade net ΔAGB derived from four recent global multi-date AGB maps: ESA-CCI maps, WRI-Flux model, JPL time series, and SMOS-LVOD time series. Our assessments explore and use different reference data sources with biomass re-measurements within the past decade. The reference data comprise National Forest Inventory (NFI) plot data, local AGB maps from airborne LiDAR, and selected Forest Resource Assessment country data from countries with well-developed monitoring capacities. Map to reference data comparisons were performed at levels ranging from 100 m to 25 km spatial scale. The comparisons revealed that LiDAR data compared most reasonably with the maps, while the comparisons using NFI only showed some agreements at aggregation levels <10 km. Regardless of the aggregation level, AGB losses and gains according to the map comparisons were consistently smaller than the reference data. Map-map comparisons at 25 km highlighted that the maps...
1. Introduction

The above-ground biomass (AGB) content in forests represents the amount of carbon they store and hence changes in AGB correspond to the amount of CO₂ emitted to or removed from the atmosphere. This function of AGB and their changes (ΔAGB) defines them as an essential climate variable, being an important input for global climate models and a necessity for countries in their mandated carbon accounting (Herold et al., 2019). Monitoring the spatial and temporal dynamics of AGB benefits from Earth Observation (EO) and a wider range of observations from space that shall enable a more accurate estimation of AGB is anticipated (Rodríguez-Vega et al., 2017).

Two approaches are commonly used to obtain an estimate of ΔAGB. The stock change approach consists of estimating ΔAGB by differencing AGB maps from different epochs. Sources of such an approach include the Climate Change Initiative (CCI) Biomass AGB maps for the epochs 2010, 2017 and 2018 with a 100 m pixel size produced using radar remote sensing (Santoro and Cartus, 2021). Another example is the recent L-band Vegetation Optical Depth (LVOD) global AGB dataset with a pixel size of 25 km (Wigneron et al., 2021; Yang et al., 2022). In contrast, with the "gain-loss" approach, land use-specific carbon emission and removal factors are used to derive an estimate of ΔAGB starting from an initial estimate of AGB for a given epoch (McRoberts et al., 2020). One example is Harris et al. (2021), wherein a baseline AGB map, annual forest loss and gain maps, and activity data specific to deforestation, fire, agriculture and forestry were used to assess net carbon fluxes in the past two decades at 30 m spatial resolution. Xu et al. (2021) used a similar approach that paired annual AGB maps with activity data for carbon fluxes analysis from 2000 to 2019 at the spatial scale of 10 km.

Independent map assessments should be a standard operating procedure for EO-based products (Nightingale et al., 2010; Duncanson et al., 2021). Map assessment in this context involves the comparison of ΔAGB from maps and a reference dataset, i.e., using in-situ data from repeated AGB measurements. A common source of reference data used is represented by measurements collected at sample plots by a National Forest Inventory (NFI) data. Data from NFIs in extra-tropical countries are commonly being used for ΔAGB assessments and subsequent analysis mainly because these NFIs are well-established. Most NFIs, however, are not open-access data, which constrains their use for assessments beyond country scales. Yet, countries have used NFI data to report AGB statistics (means and totals) every five years as part of United Nation’s Forest Resource Assessment (FRA). The FRA data was used in global-scale studies related to ΔAGB analysis, e.g., as a basis for comparison with map-based estimates aggregated to countries in multiple periods (Xu et al., 2021; Araza et al., 2022b). At local scales, ΔAGB from airborne LiDAR-based maps provide precise estimates. While forest height derived from LiDAR is highly correlated with AGB, the local estimates they provide are considered more reliable than the ΔAGB from plot data as – unlike plots – LiDAR covers whole landscapes (Næsset et al., 2015). These features have allowed LiDAR as a reliable AGB reference data (Duncanson et al., 2021). Given that most LiDAR campaigns are being repeated for monitoring forest ecosystems, a long list of local LiDAR-based ΔAGB assessments is available. A good example is the study of Nguyen et al. (2020) that used periodic LiDAR-based AGB to validate an AGB time series.

The ΔAGB map assessments concern the spatial resolution i.e., aggregation level over which ΔAGB is assessed. The choice of spatial resolution often depends on the map use (Quegan and Ciais, 2018; Herold et al., 2019). Global map users such as carbon and climate modellers commonly use coarse resolution maps (>1 km) as inputs to global models (Quegan and Ciais, 2018). In contrast, most country-level applications prefer fine-resolution maps that are more informative about highly localized forest area changes such as small-scale mining and slash-and-burn farming. For instance, high-resolution ΔAGB maps serve as input for carbon monitoring, reporting and verification (MRV) (Csillik et al., 2022), and ecosystem accounting (Hein et al., 2020). Because of this, it makes sense to iterate map assessments over a range of aggregation levels to determine at what resolution the plot and pixel averages best match (Moreno et al., 2016; Santoro et al., 2022a). Such an exercise would also inform the map users how the maps compare with certain reference data.

Global ΔAGB assessment is challenged by several factors. First, there is a lack of readily available and globally suitable reference data because global mapping of multi-temporal AGB has only recently started. Second, the uncertainty about ΔAGB assessed from a global product is large. For instance, changes driven by slow regrowth and degradation are challenging to be detected from satellites (Ryan et al., 2011; Santoro and Cartus, 2021). Third, the current good practices for AGB map validation (Duncanson et al., 2021) and map-reference data comparisons (Araza et al., 2022b) concern AGB maps of single epochs. Multi-date AGB mapping studies have either assessed their products for single epochs or skipped map assessments all together. Lastly, not all ΔAGB data sources directly provide ΔAGB. Converting CO₂ and carbon into AGB is straightforward, but sometimes the map products combine the carbon of above-ground and below-ground components (Xu et al., 2021) and occasionally even soil components are included (Harris et al., 2021).

Here we present an exploratory assessment of four ΔAGB maps that represent the past decade. The maps were specifically derived from the CCI maps; carbon flux produced using the Harris et al. (2021) method; the Xu et al. (2021) maps and the LVOD maps herein called as CCI, Flux, JPL, LVOD, respectively. Given the lack of a consistent global ΔAGB reference dataset, map-map comparisons are used to assess consistency among the maps in reporting AGB gains and losses. We (1) compile several ΔAGB reference data sources and assess their suitability for map assessment; (2) compare the ΔAGB between reference data and maps at different spatial aggregation levels; and (3) spatially assess the mutual consistency of ΔAGB maps.

2. Materials and methods

2.1. Overview of the methods

The assessment of four global ΔAGB Mg ha⁻¹ is outlined in Fig. 1. Preliminary steps that include forest masking, modification of the original carbon flux model and other pre-processing steps all made sure that the maps have comparable ΔAGB. To assess the ΔAGB maps, we used three types of AGB dataset with repeated measurements as reference data: airborne LiDAR-based AGB maps (LiDAR), NFI plots and FRA.
To make the map-reference data comparisons meaningful, we further selected subsets of the reference data based on data quality criteria and we applied the same forest mask (used for the maps) based on a forest definition set as >30% tree cover at 30 m pixel size (Hansen et al., 2013). Then, the ΔAGB of the reference data and maps were aggregated and compared at five aggregation levels between 100 m and 25 km. The reference data uncertainty were estimated and used as weights for the aggregation step. At 25 km, we assessed the level of agreement among them through spatial analysis and cross-correlation. Note that the units of the aggregation level and map pixel size are both referred similarly e.g., 25 × 25 km² and 25 km.

2.2. ΔAGB maps

CCI Biomass maps

The European Space Agency Climate Change Initiative (CCI) Biomass dataset consist of annual global AGB maps with a pixel size of 100 m. The most recent version includes maps for the epochs 2010, 2017 and 2018, and we selected the 2010 and 2018 epochs for our analysis. Each map was derived from synthetic aperture radar (SAR) data where the signal is in the form of backscatter intensity. The SAR input data are then integrated into a physically-based model, accounting for the individual contributions to the backscattered signal from the forest canopy and the ground below the canopy (Santoro and Cartus, 2021; Santoro et al., 2021). The CCCI modelling approach is reinforced by allometric equations that relate the canopy density to forest height and reduces the inter-annual variability in the original time series. The JPL product (Xu et al., 2021) retrieves AGB from remote sensing data following a two-step approach. First, in every 10 km grid cell covering field plots, AGB mapping is performed based on the relationship between field plot AGB and vegetation height derived from spaceborne LiDAR and backscatter intensity from spaceborne radar data. The resulting grid cells are then used to train and test machine learning models using covariates from MODIS optical data, topographic and climatic variables. At the pixel level, Xu et al. (2021) estimated and summed the uncertainties related to model residual, model parameter estimation error and plot measurement error. The 10 km time series of above-ground carbon maps were first divided by 0.49 to obtain AGB. We used the average of 2009–2011 and 2017–2019 to derive ΔAGB. This step allows comparison with the other ΔAGB maps and reduces the inter-annual variability in the original time series.

SMOS-LVOD time series AGB

Lastly, we used the 25 km LVOD time series AGB products (Wigneron et al., 2021; Yang et al., 2022) derived from observations by the Soil Moisture and Ocean Salinity (SMOS) satellite. Spatially explicit estimates of LVOD were derived every year between 2010 and 2019. Filtering individual images was done to mitigate the signal noises but resulted in partial coverage of east Asia. Temporal aggregates of the LVOD product are results of a temporal decorrelation method to reduce the seasonal effects related to water content. The uncertainty of the LVOD dataset primarily comes from the AGB reference biomass map used. The LVOD dataset was originally projected to an Equal-Area Scalable Earth (EASE) system. We re-projected the map to WGS 84 using bilinear interpolation. Note that owing to data availability, 2011 was used as the starting point for ΔAGB assessment by LVOD.
More information about the ΔAGB maps are shown in Table S2 including the different remote sensing data applied, spatial resolution, temporal resolution, validation method and forest masking.

2.3. ΔAGB reference data

NFIs

The first set of reference data consists of re-measured National Forest Inventories (NFI) plot data from Belgium, the Netherlands, Philippines and Sweden where the first and second measurements were surveyed at least five years apart. The re-measurements allowed estimation of ΔAGB. The NFIs have plot-level AGB estimated by the data sources, but without uncertainty estimates. We then estimated plot uncertainty from measurement and allometric model error as a function of AGB, eco-zone and plot size. We used the prediction model described in Araza et al. (2022b) developed from an extensive plot database spread over all major eco-zones. The model followed an error propagation method of parameters from tree measurements such as diameter, height and wood density and the parameters of the AGB allometric model such as the model coefficients and residual standard errors. The AGB of the reference data was also adjusted to reduce the effects of temporal mismatches between the reference and maps. The AGB of datasets surveyed ± 2 years apart from the map epoch were subjected to this step using growth data from the 2019 Intergovernmental Panel on Climate Change (IPCC) (Buendia et al. 2019). Biomass is either added or reduced depending on the number of years between the inventory date and map epoch. The associated annual uncertainty owing to these adjustments was also estimated based on the reported IPCC growth data uncertainties.

LiDAR

The second reference data comprises local AGB maps derived in forests with re-measured plot inventories and two airborne LiDAR campaigns between 2010 and 2019. Maps of ΔAGB were derived by differencing AGB maps from the two survey periods. Data sources include maps from Brazil (Longo et al., 2016) and the USA (Johnson et al., 2010) where AGB mapping involved calibration of LiDAR height and plot AGB using power-law models. We also used LiDAR-based maps from research projects in Bulgaria, Czech Republic, Costa Rica, Poland and Spain derived using regression models that relate height and AGB. The LiDAR maps resampled to 100 m were used. Some of these maps have associated SD layers all estimated using Monte Carlo error propagation involving errors from the calibration dataset and the associated height-AGB model parameters.

FRA

The third set of reference data are country-level estimates of ΔAGB from FRA reports, derived by differencing the reported AGB 2018 and 2010, where 2018 was computed as the average of 2015 and 2020 AGB. We selected the FRA data based on the capability of countries to conduct NFIs and derive FRA variables using remote sensing. We followed the FRA capacity categories based on a scale of 1 to 5 (1=very poor; 5=very good). Selected countries belonging to east Asia were excluded because of incomplete map coverage as explained in Section 2.2. The AGB from the FRA data does not have associated uncertainty estimates.

Reference data quality assessment

We further selected subsets of the reference data:

1. NFI plots were filtered using tree cover loss datasets (Hansen et al., 2013) to retain only plots without forest area changes after the latest measurement and prior to 2018 map epoch. We also discarded plots more than 10 years apart from the map epoch;
2. LiDAR pixels were discarded if there are AGB values in one epoch but without values in the other epoch;
3. FRA data were limited to countries with re-measured NFI or with “very good” FRI reporting capacity since 2010.

We reported the number of reference data retained after quality filtering compared to the original data, and mapped the coverage of the selected reference data over eco-zones based on Whittaker’s biome (Whittaker, 1975). The coverage per eco-zone and country determined the suitability of reference data for global map assessments. For each reference dataset, histograms of the AGB distribution in two epochs are shown in Figure S1. We also derived the ΔAGB density for NFI and LiDAR to assess the ΔAGB distribution at every aggregation level (Section 2.4). We aim to assess how the ΔAGB (losses, gains and no changes) are depicted from fine to coarse levels depending on the reference data. More information about the reference data is shown in Table S1 and their maps are shown in Figure S2.

2.4. Map-reference data ΔAGB comparisons

The past decade net ΔAGB from the map products and the reference data were compared at different spatial resolutions i.e., grid cells. The details of the comparisons are shown in Table 1 showing the grid cell selection.

Grid cells were used if they met the minimum number of reference data inside grid cells (Xu et al., 2021; Araza et al., 2022b), see 2nd column of Table 1. This way, grid cells > 1 km with very few reference data were excluded from the analysis. The AGB averages per epoch from NFI plots and LiDAR pixels at grid cells were estimated as weighted means where reference data with high uncertainty received smaller weights in the averaging. The weights W(x) are inversely proportional to the variance SD^2 of an NFI plot or a LiDAR pixel x as shown in Eq. (1) (Araza et al., 2022b).

\[ W(x) = 1/SD^2(x) \]

The AGB averages of both reference data and maps were assured to correspond to a forest mask defined as 30 m pixels > 30% tree cover (Hansen et al., 2013). This step was separately done per epoch and for each aggregation level. Subjected to this harmonization are the ΔAGB maps from CCI and LVOD (see Table S2) and the LiDAR reference data. Particularly, the maps were resampled to 30 m to match the forest mask pixel size. All non-forest pixels were masked out then the remaining pixels were averaged for each aggregation level. The assessment at 25 km level included FRA as reference data. We simply used the FRA country average AGB for the needed epochs, which are subject to the country forest definition.

The ΔAGB comparisons of grid cell averages between the reference data and maps were assessed using statistical measures including mean difference (MD), Root Mean Square Difference (RMSD), coefficient of determination (R^2) and Nash–Sutcliffe Efficiency (NSE). These evaluations were implemented for each aggregation level and summarized as colour-coded matrices per ΔAGB bin. Herein, ΔAGB comparisons pertain to map-reference data ΔAGB comparisons, which were also displayed as scatterplots and per ΔAGB bin. We also refer to map AGB loss and gain underestimation whenever ΔAGB estimates from reference data are higher than the map estimates.

2.5. Map-map ΔAGB comparisons

The 25 km ΔAGB maps were masked using the LVOD product for geographical comparability. The latitudinal profiles of the masked ΔAGB maps were graphed to depict regional patterns of ΔAGB. Then, the level of agreement among the ΔAGB maps was spatially assessed depending on the sign of the ΔAGB. We first classified ΔAGB into “loss” (net loss), “gain” (net gain) and “no change”. Here we assume ΔAGB 7 to −7 Mg ha⁻¹ as “no change” based on a conservative SD of 9-year growth rate
defined in Table 14 of IPCC 2019 for global analysis (Buendia et al., 2019). The threshold avoids erroneous labelling of small ΔAGB values, which can be very uncertain (Ryan et al., 2011, Santoro et al. (2022a)). The ΔAGB pixels were finally classified as follows: (1) all products agree on “loss”; (2) all products agree on “gain”; (3) all products agree on “no change”; (4) 2 products agree on “loss”, other 2 not “loss” and disagree with each other; (5) 2 products agree on “gain”, other 2 not “gain” and disagree with each other; (6) 2 products agree on “no change”, other 2 not “no change” and disagree with each other; (7) 3 products disagree. We also assessed cross-correlation among the 25 km ΔAGB maps (using all pixels) indicated by Pearson’s correlation coefficient (r).

3. Results

3.1. Reference data for ΔAGB global comparisons

The spatial distribution of the reference datasets are shown in Fig. 2. The number of discarded data was largest for FRA (90%) since most countries do not have repeated NFIs (see Table S3). More than half of the NFI plots (56%) were excluded either because they were outdated (see Section 2.3) or the sites were deforested after the 2nd measurement and before 2018. Almost no LiDAR pixels (<1%) were filtered out as reference since the repeated LiDAR surveys took all place in the past decade and almost all pixels had valid data in both epochs. The reference data are mostly found in the temperate and tropics but under-represents them as well as the other eco-zones. The selected FRA data, though small in size, come from all eco-zone. Despite its smaller size, the NFI dataset has broader eco-zone coverage than the LiDAR dataset. That is because NFIs are surveyed over entire countries while LiDAR campaigns are typically confined to forests. We had no access to NFIs and LiDAR data from Africa and Australia.

The ΔAGB distributions from LiDAR and NFI data at different aggregation levels are shown in Fig. 3. The highest density of data is observed for small ΔAGB but there are also several reference data implying large AGB gains and losses. Owing to spatial averaging, the density of data increases towards small ΔAGB from fine to coarse aggregation levels, especially for NFI data. The NFI captured larger AGB gains (until 1 km), while LiDAR captured more AGB losses throughout all aggregation levels. These results suggest LiDAR data showed consistent ΔAGB distributions across the aggregation levels. These observations are influenced by the forest types where the reference data are located, i.e., forest plantations for NFIs and disturbed forests for LiDAR (Table S1).

3.2. ΔAGB comparisons at different aggregation levels

Fig. 4 compares the ΔAGB derived from the CCI and Flux maps with respect to the corresponding NFI and LiDAR values for all aggregation levels (left to right). While the averaging resulted in a decrease in scattering especially in small ΔAGB (i.e., RMSD), the mean difference (MD) is still prominent particularly for AGB loss regardless of the spatial scale. The map-based estimates of ΔAGB agreed most with the LiDAR-based values regardless of the spatial scale (e.g., 0.1–0.44 R²). When compared to ΔAGB values derived from NFI data, the agreement was instead moderate until 1 km (e.g., at most 0.1 R²) and poor for coarser spatial scales (e.g., at most 0.03 R²). Often, the number of NFI plots inside coarse grid cells is limited e.g., 5–6 plots per 10 km cell (Figure S2). Another observation in Fig. 4 particularly for the 100–500 m comparisons is the different distribution of ΔAGB between CCI and Flux ΔAGB bins with no to minimal changes based on reference data. The effect of spatial averaging is further observed in the coloured matrices of MD and RMSD per ΔAGB bin in Fig. 5.

Fig. 6 shows the ΔAGB comparisons for all map products against reference data aggregated to a common spatial resolution of 25 km. At such a coarse level, map estimates show some level of agreement with LiDAR data. For all maps, agreement is substantially reduced when using NFI data as the reference. The comparisons with the FRA show that countries with re-measured NFI (mostly reporting gains) agree differently with the maps (e.g., 0.03–0.28 R²). These variations among maps will be the focus in the spatial analysis results.

3.3. ΔAGB spatial analysis and map-map comparisons

Fig. 7(a) shows the inter-comparison of ΔAGB maps. The magnitude and ranges of ΔAGB vary among the maps. The CCI and Flux maps display the largest AGB changes in time as indicated by higher colour contrast in the map and higher fluctuations in their latitudinal profile graphs (Figure S3). Examples are depicted in the CCI results for regions such as east US, south Amazon basin and Madagascar (~18°), and South American countries along ~42°; and for high gains in the temperate region for the Flux results. These are regions where the CCI and Flux disagree. Except for LVOD AGB loss, the changes are not very evident in the two time series products (JPL and LVOD). The maps in Fig. 7(a) mostly agree in regions where net AGB loss is prominent such as southwest Amazon, Siberian boreal forests, west and central Africa and Indonesia. Consensus about net forest gains is evident in China, western Canada, African savannahs and in a few patches within Europe and Amazon basin; but disagreement is evident in central Amazonia. Separate maps of these hotspot regions are shown in Figure S4. Map disagreements are further emphasized in Figure S5 where correlation coefficients (r) among maps range from 0.11 to 0.29. The majority of the pixels are classified as “no change” as shown in Fig. 7(b). They constitute 66% of total while the other classes constitute: disagree = 21%, gain = 8% and loss = 5%.

4. Discussion

4.1. Reference data quality assessment

Current NFI and LiDAR reference data sources under-represent most eco-zones especially in the tropics. We used three European NFIs and one NFI in the tropics, as NFIs are commonly government data and requesting access is often a long process. Aside from crowd-sourcing platforms, online tools such as the Plot2Map tool by Araza et al. (2022a) can facilitate access to NFI data. Through the Multi-Mission Algorithm Platform (Albinet et al., 2019), the use of Plot2Map has been demonstrated in three countries already. Like for NFIs, the LiDAR reference data were currently limited to specific regions, but there has been an increasing interest and coverage of permanent forest plots with LiDAR campaigns worldwide (Chave et al., 2019). We used NFIs mostly located in European forest plantations while LiDAR data mostly overlapped both disturbed and stable forests. The forest type where
the reference data are situated also drives the distribution of $\Delta$AGB observed in Fig. 3. Future reference data should also represent different forest land uses.

As the availability of reference datasets increases, the criteria for selecting them for global map assessments can become stricter. While the size of plots may need to be large enough (i.e., $>1$ ha) to fully cover the pixels of map products (Réjou-Méchain et al., 2019), this will limit the reference data to permanent research plots which are always preferentially sampled. Here the effects of plot size are considered when we weighted our reference data according to plot size-driven uncertainty, and when we excluded grid cells with plots fewer than the minimum number of plots criterion. Stricter grid cell selection can be implied at coarse aggregation levels. The minimum number of plots per grid cell can be increased (Moreno et al., 2016) or grid cells with locally representative plots can only be selected (Araza et al.,
A. Araza et al.

Fig. 4. $\Delta AGB$ comparisons between the originally 100 m map products (CCI and Flux), and the LiDAR and NFI reference data. The binning symbol size (circles) depends on the number of data, with whiskers indicating the 25th and 75th quartile of the $\Delta AGB$ map. The bins for the 25 km results are lesser than in Fig. 6. Be aware also of the different axes for NFI and LiDAR assessment i.e., narrower axis $>1$ km onwards to increase the visibility of comparisons. Coefficients $R^2$ and NSE are also shown.

Fig. 5. Matrices of MD and RMSD derived from the comparisons at different aggregation levels and $\Delta AGB$ bins (Mg ha$^{-1}$) using NFI and LiDAR data as reference. The colour gradient from light to dark depicts low to high values. For the MD matrix, red gradient refers to AGB loss underestimation while yellow gradient depicts AGB gain underestimation.

2022b). Moreover, additional analysis on global data representativeness and sampling intensity could be initiated. Findings from these can support decisions on whether potential data are still necessary for certain regions (Fig. 2) and identify regions that need more attention for data requests and measurement campaigns. In cases where NFIs and $\Delta AGB$ maps need integration for national $\Delta AGB$ estimates and rigorous map accuracy assessments, the NFI sampling design should be considered (Nesha et al., 2022). Further caution is needed because
NFIs are not primarily designed for map comparisons and plots do not properly sample mapped ΔAGB especially at coarse aggregation levels. Moreover, NFIs can have different sub-plot configurations that cannot be tessellated over the entire mapped area. If only LiDAR is available for country ΔAGB estimation, one should also consider the effect of LiDAR data being preferentially sampled in particular forest types. Lastly, only reference data with estimated uncertainties are preferred for map assessments for uncertainty-weighted AGB averaging. Table S4 summarizes the above discussions geared towards current and future reference data suitability.

ΔAGB comparisons at different aggregation levels

Comparisons of ΔAGB between the reference data and maps from 100 m to 25 km levels of aggregation revealed a decrease of scattering in the ΔAGB comparisons (i.e., smaller RMSD), and the persistent underestimation of AGB losses, owing mostly to map systematic differences (MD), see Fig. 5. Hence, the MD problem exists regardless of the map product and spatial resolution. It is expected for AGB maps to overestimate smaller AGB and underestimate greater AGB (Réjou-Méchain et al., 2019; Araza et al., 2022b) and this MD is carried over into ΔAGB, especially for deforestation and plantation clear-cutting events, e.g., a map-based ΔAGB from 350 to 5 Mg ha\(^{-1}\) where reference data ΔAGB shows 450 to 0 Mg ha\(^{-1}\). Similarly, gains spanning 9 years (this study) were also expected to be underestimated because of this systematic error.

The comparisons from fine to coarse levels were influenced by the reference data (Fig. 4). The comparisons using NFI and LiDAR at aggregation levels up to 1 km showed moderate agreement with maps and were able to assess large changes i.e., deforestation, clear-cuts and regrowth (Fig. 3). These kinds of ΔAGB at fine resolutions are more important when map users require ΔAGB maps such as for carbon MRV reporting and ecosystem accounting. Beyond 1 km comparisons where the ΔAGB between NFIs and maps disagreed, the NFI plots we used likely missed to represent local heterogeneous areas with different forest types and forest management activities in Europe (Figure S2). This result is consistent with (Herold et al., 2022) where the ΔAGB of NFIs and CCI map were similar when using at least 25 European NFI plots in comparisons aggregated to 12 km. Also using all European NFIs, Moreno et al. (2016) reported that increasing the aggregation level from 1 to 5–25 km is optimal because of increased plot representation over grid cells. The country-level ΔAGB comparisons using FRA as reference depicted varying results among maps (Fig. 6). This variation can be attributed to the different forest definitions and AGB estimation methods of countries. Metadata of the FRA dataset is needed as basis to harmonize the map-based estimates and make them more comparable with the FRA. Nevertheless, country-level variations also reflect that there is disagreement of ΔAGB among maps.

Spatial analysis and map-map differences

The purpose of inter-comparing the ΔAGB maps was to assess where maps indicate the same changes. Agreement among maps was gauged by the sign and magnitude of reported ΔAGB, and the ΔAGB cross-correlation results ranged between 0.11 and 0.29 \(r\) (Figure S5). The map products were produced by different remote sensing data types and AGB retrieval methodologies (Section 2.2). The CCI ΔAGB is obtained by subtracting the 2018 and 2010 maps i.e., stock change approach. AGB maps are indirect estimates from remote sensing signals and changes in signals are not always equivalent to ΔAGB. Past and current spaceborne radar datasets used for CCI lack sufficient sensitivity to measure gradual changes associated with regrowth and degradation especially in dense forests (Santoro and Cartus, 2021). For the Flux model, a gain-loss approach, the IPCC 2019 growth rates were one of the bases for years beyond >2012. Such rates can be very uncertain in Europe (Harris et al., 2021). Interestingly, the CCI and Flux yielded contrasting ΔAGB in certain regions (Fig. 7) despite using the same CCI 2010 map as a baseline. This regional variation illustrates the different
results obtained by stock change and gain-loss approaches. The latter can be strongly influenced by the localized activity data i.e., land uses (McRoberts et al., 2020). Compared to the CCI and Flux, the JPL and LVOD time series products reported smaller ΔAGB magnitude especially for AGB gains. Lastly, the ΔAGB differences among the maps may also be affected by the different remote sensing data and mapping methods employed by map producers and the slight variations in the map epochs (Table S2).

The overlap of ΔAGB maps in Fig. 7 showed that current maps mostly capture AGB losses in known deforestation hotspot regions (Feng et al., 2022). It is known that the satellite signal before and after deforestation is less prone to noise than signals concerning gradual changes like regrowth or degradation (Ryan et al., 2011). Moreover, both JPL and Flux used the same tree cover loss product (Hansen et al., 2013), while the LVOD used the CCI precursor, the GlobBiomass (Santoro et al., 2021), for AGB calibration. All maps reported sinks of carbon in parts of Europe, China, Canada and African savannah, which coincides with several global and regional studies that used remote sensing and in-situ data (Bastin et al., 2017; Tubiello et al., 2021). In contrast, map disagreement was evident in the central Amazon basin — being the largest and most complex carbon pool region. Previous work of Phillips et al. (2017) reported contrasting views on whether the basin is a source or sink. Most ΔAGB pixels were within the interval between $-7$ and $7$ Mg ha$^{-1}$, which were classified as “no change”. These pixels occur mostly in intact and non-forest areas that sequester little to no carbon (Lesiv et al., 2022).

**Application of the work and outlook**

This work provides a confluence of evidence from ΔAGB map-reference data and map-map comparisons, which is a source of information for both map producers and users. Map producers can revisit their AGB retrieval process in areas where most maps disagree. We anticipate improvements of the current ΔAGB maps. Future releases of the CCI maps will make use of repeated observations instead of a single observation in the form of an image mosaic. In addition, the AGB retrieval will be supported by a much denser dataset of spaceborne LiDAR observations. The Flux model will implement improved activity data and removal factors. In terms of map users, global carbon and climate modellers mainly use global ΔAGB maps (Herold et al., 2019;
A. Araza et al.

The variability in the 25 km map-reference data comparisons and the map-map comparisons reflect potential limitations of the current global AGB assessment. One approach to deal with map disagreements is to produce a new map from an ensemble of existing global AGB maps using model-based approaches (Zhang and Liang, 2020). However, doing so requires a suitable global reference dataset and does not guarantee producing accurate ΔAGB estimates. Worth exploring though is the direct mapping of ΔAGB using the global AGB products and other remote sensing auxiliary data in countries with reference data. The resulting country AGB map would benefit national applications that require AGB map inputs e.g., for carbon MRV (Csislik et al., 2022), and ecosystem accounting (Hein et al., 2020). Such mapping exercise should account for the uncertainty of sampling variability. This caution also applies for sub-national to national carbon accounting using rigorous estimation methods such as an NFI-based model-assisted ΔAGB estimation.

5. Conclusions

We compared ΔAGB derived from four recent multi-date global AGB maps with three reference datasets at spatial scales between 100 m and 25 km. We also compared the 25 km ΔAGB among maps. We conclude the following:

1. The map-based estimates of ΔAGB agreed most with the LiDAR-based values regardless of the spatial scale (e.g., 0.1–0.44 R²). When compared to ΔAGB values derived from NFI data, the agreement was instead moderate until 1 km (e.g., at most 0.1 R²) and poor for coarser spatial scales (e.g., at most 0.03 R²). The comparison with country averages reported in the FRA revealed different levels of agreement depending on the map (e.g., 0.03–0.28 R²). The assessments revealed systematic differences in mapped ΔAGB compared to the reference data, i.e., mapped AGB loss was smaller owing to the underestimation of AGB in baseline maps (2010).

2. The global ΔAGB map assessments should be considered exploratory because the current NFI and LiDAR reference data sources under-represent most eco-zones especially in the tropics. Nonetheless, this limitation will be alleviated by upcoming airborne LiDAR campaigns that cover most eco-zones and world regions. Moreover, the use of NFI data can be enhanced if more countries decide to support online map assessment tools like Plot2Map and data crowd-sourcing platforms. Reference data with uncertainty estimates and within areas where we lack representation in both geographic and feature space (e.g., eco-zones) are preferred for future global ΔAGB map assessments.

3. The maps all identified AGB losses in known deforestation hotspots, while also identifying regions that act as sinks. The remaining disagreement among maps (e.g., in the Amazon basin) reflects the different methodology applied such as stock differencing for CCI and gain-loss for Flux; while the two time series products had smaller ΔAGB magnitudes. To assure comparable ΔAGB among maps and with reference data, preliminary steps were necessary including modification of the original carbon flux model (Flux) and applying a standard forest mask.

The map disagreements and biases need to be understood and addressed to increase the reliability of the maps for applications that require ΔAGB estimation and information.

Models used:
- Modified carbon flux model: https://github.com/arnanaraza/carbon-budget
- Plot2Map: https://github.com/arnanaraza/PlotToMap

CRediT authorship contribution statement


Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Reference data should be requested. Maps are accessible at http://dx.doi.org/10.6084/m9.figshare.22349218

Acknowledgements

This study was partly supported by the (1) IFBN/FOIS (contract no. 4000114425/15/NI/FF/gp), SEN4LDN andCCI Biomass (contract no. 4000123662/18/1-NB) projects funded by the European Space Agency; (2) VERIFY Project: Observation-based system for monitoring and verification of greenhouse gases (GA number 776810, RIA). Acknowledgement is also given to the Open Earth Monitor Project from European Union’s Horizon Europe research an innovation programme (grant agreement 101059548); the World Resources Institute (WRI) land and carbon lab support to WUR and GFZ; and the CGIAT MIT-IGATE+ project. Authors TAMP, AEM and MJS acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 758873, TreeMort). This study contributes to the Strategic Research Areas BECC and MERGE. Czech Republic data was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the CzeCOS program, grant number LIM203048. Polish data was supported by Project LIFE+ ForBioSensing (contract number LIFE13 ENV/PL/000046) and Poland’s National Fund for Environmental Protection and Water Management (contract number 485/2014/WN10/OP-NMLF/D). We sincerely thank Nicolas Labriere for processing and providing the LiDAR reference data. We also thank Maciej Lisiewicz and Lukasz Kuberski, for their assistance in Polish data preparation.


