

	GlobBiomass		Page 1
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

DUE GlobBiomass

D6 – Global Biomass Map

Algorithm Theoretical Basis Document

(incorporating the Design Justification File D7)

Prepared for European Space Agency (ESA-ESRIN)

In response to ESRIN/Contract No. 4000113100/14/I_NB



	GlobBiomass		Page 2
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

Revision History

Deliverable	D6, D7 – Global Biomass Map		
Work Package	3000		
Due date	KO+25		
Authors	Shaun Quegan, Yrjö Rauste, Alexandre Bouvet, Joao Carreiras, Oliver Cartus, Nuno Carvalhais, Thuy LeToan, Stephane Mermoz, Maurizio Santoro		
Distribution	FSU: Christiane Schmallius, Evelin Matejka, Carsten Pathe, Christian Thiel	ESA: Frank Martin Seifert; Nathalie Boisard	
Reason for change			
Issue			
Revision			
Date			
Release	1		
Version	03		

	GlobBiomass		Page 3
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

Contents

1	Global Biomass Map.....	6
1.1	Issues associated with global biomass mapping using Earth Observation data	6
1.2	Possible solutions towards a novel global biomass retrieval scheme.....	8
1.3	Datasets.....	10
1.3.1	Input datasets.....	10
1.3.1.1	Envisat ASAR (C-band, wavelength 5.6 cm).....	10
1.3.1.2	ALOS PALSAR (L-band, wavelength 23 cm)	12
1.3.1.3	Landsat reflectance mosaics	13
1.3.2	Datasets used for training	14
1.3.2.1	In situ measurements	14
1.3.2.2	MODIS Vegetation Continuous Fields	14
1.3.2.3	Landsat canopy density and density change	14
1.3.2.4	ICESAT GLAS.....	15
1.3.2.5	CCI Land Cover map.....	16
1.3.2.6	ERA Interim air temperature	17
1.3.2.7	Worldclim Bioclimatic Variables.....	17
1.3.2.8	FAO Global Ecological Zones dataset	18
1.3.2.9	Global forest canopy height	18
1.4	Methods	18
1.4.1	Data pre-processing.....	18
1.4.1.1	Envisat ASAR.....	18
1.4.1.2	ALOS PALSAR mosaics	22
1.4.1.3	Landsat data products.....	25
1.4.1.4	ICESAT GLAS.....	25
1.4.1.5	MODIS Vegetation Continuous Fields	26
1.4.1.6	CCI Land Cover.....	26
1.4.1.7	ERA Interim.....	26
1.4.1.8	Global Ecological Zones	26
1.4.1.9	WorldClim Data Base.....	26
1.4.1.10	GSV of dense forest	27
1.4.1.11	Global forest canopy height	29
1.4.2	Biomass estimation algorithms	30
1.4.2.1	BIOMASAR-C.....	33
1.4.2.1.1	Estimation of σ_{gr}^0	35

	GlobBiomass		Page 4
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.4.2.1.2	Estimation of σ_{df}^0	37
1.4.2.1.3	Estimation of β	37
1.4.2.1.4	Estimation of σ_{veg}^0	41
1.4.2.1.5	Inverting the forest backscatter model	41
1.4.2.1.6	Multi-temporal combination of GSV estimates	42
1.4.2.2	BIOMASAR-L	42
1.4.2.2.1	Estimating σ_{gr}^0 and σ_{df}^0	45
1.4.2.2.1.1	Data preparation	45
1.4.2.2.1.2	Identification of areas of low and high canopy density	46
1.4.2.2.1.3	Model parameter estimation	46
1.4.2.2.2	Estimating σ_{veg}^0	47
1.4.2.2.3	Estimation of β	50
1.4.2.2.4	Retrieval of GSV	53
1.4.2.2.4.1	Maximum retrievable GSV	53
1.4.2.2.4.2	Comparison of BIOMASAR-L retrieval results with existing maps	54
1.4.2.2.4.3	Multi-polarization/multi-temporal retrieval strategy	58
1.4.2.2.5	Limitations of BIOMASAR-L when using PALSAR mosaics	60
1.4.2.3	Upscaling BIOMASAR-C GSV estimates	64
1.4.2.4	CESBIO	69
1.4.3	Methods to derive a merged GSV data product	69
1.4.3.1	Inter-comparison of individual biomass datasets	69
1.4.3.2	Merging biomass estimates	79
1.4.4	Conversion of GSV to AGB	85
1.4.5	Methods to assign accuracies to the retrieved AGB	100
1.4.5.1	BIOMASAR-C	100
1.4.5.2	BIOMASAR-L	102
1.4.5.3	BIOMASAR-C+	103
1.4.5.4	CESBIO	105
1.4.5.5	Integrated GSV data product	105
1.4.5.6	Conversion of GSV to AGB	105
1.5	Products	105
2	Annex A	107
3	Annex B	111
4	References	112

	GlobBiomass		Page 5
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

List of acronyms

AGB	Above ground biomass
ALOS	Advanced Land Observing Satellite
APM	Alternating Polarization Medium
ASAR	Advanced synthetic aperture radar
ATBD	Algorithm Theoretical Basis Document
CART	Classification and Regression Trees
CCI	Climate Change Initiative
DEM	Digital Elevation Model
ECMWF	European Centre for Medium-range Weather Forecast
Envisat	Environmental satellite
EO	Earth Observation
ESA	European Space Agency
FAO	Food and Agriculture Organization (
FIA	Forest Inventory and Analysis
GLAS	Geoscience Laser Altimeter System
GM1	Global Monitoring
GSV	Growing stock volume
ICESAT	Ice, Cloud and Land Elevation Satellite
IIASA	International Institute of Applied System Analysis
IMM	Image Mode Medium
IPCC	Intergovernmental Panel on Climate Change
JAXA	Japan Aerospace Exploration Agency
<i>k</i> NN	<i>k</i> -Nearest Neighbour
LC	Land Cover
MIPERS	Multi-static Interferometric and Polarimetric Electromagnetic model for Remote Sensing
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NBCD	National Biomass Carbon Dataset
NDVI	Normalized Difference Vegetation Index
NPP	Net Primary Production
PALSAR	Phased Array type L-band Synthetic Aperture Radar
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
SAR	Synthetic aperture radar
SRTM	Shuttle Radar Topography Mission
USGS	United States Geological Survey
VCF	Vegetation Continuous Fields
WCM	Water Cloud Model
WSM	Wide Swath Mode

	GlobBiomass		Page 6
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1 Global Biomass Map

One of the key objectives of the GlobBiomass project is to define methods leading to a global map of AGB with improved accuracy and reduced uncertainty with respect to currently available estimates. The global biomass mapping approach has been developed within the context provided by the Product Specification Document derived from user requirements (WP1000) and the available ground and space data database (WP2000). The output will consist of a global dataset of spatially explicit estimates of forest AGB, together with estimates of its accuracy.

During the first year, the focus was on assessing the strengths and weaknesses of existing retrieval approaches from a global perspective, and the form of a global mapping algorithm was derived. During the second project year, the algorithms selected were further developed to increase their robustness with respect to correct modelling of the input remote sensing observables as a function of biomass. Rather than focusing on a single approach to be applied globally, multiple frameworks have been tested for reliability in a global biomass retrieval scheme and a number of these have been further analysed, developed and assessed. As none of them is yet capable of providing a true global mapping without systematic errors, a rule-based combination of EO-based estimates of biomass from the different algorithms has been developed to generate a final synergistic set of biomass estimates. Accordingly, estimates of accuracy have been generated. While it is believed that the algorithm to retrieve biomass globally has been cast into its best possible form, it is already known that the estimates have limitations, expressed as biases and large uncertainties, primarily in regions where the remote sensing data available to this project is of limited usefulness to resolve forest structures, i.e., to sense biomass more precisely.

1.1 Issues associated with global biomass mapping using Earth

Observation data

Investigations on the derivation of forest biomass from Earth Observation data in the early 1990s agreed on the benefit of SAR, long wavelengths and cross-polarization, and used regression or semi-empirical approaches to estimate biomass (Le Toan et al., 1992; Dobson et al., 1992; Ranson et al., 1994; Imhoff et al., 1995; Pulliainen et al., 1996). What prevented SAR becoming part of an operational retrieval scheme at national/continental scale was a combination of patchy data coverage, difficulties in accessing data and the fact that the backscatter in a single image from any of the existing spaceborne SAR systems typically saturated at low biomass. The combination of optical and airborne LIDAR data is also widely used (Boudreau et al., 2008). However, radar was little used in biomass mapping until the potential of hyper-temporal ASAR C-band, SRTM, ALOS PALSAR and TanDEM-X data was realised (Santoro et al., 2011; Kellndorfer et al., 2012; Askne et al., 2013; Rosenqvist et al., 2014). In contrast, despite the lack of a direct physical relationship between the remote sensing observable and biomass, optical remote sensing has been used by several countries in support of forest inventory to map biomass and/or related forest variables at national scale (Reese et al, 2003; Tomppo et al., 2013; Beaudoin et al., 2014).

Globally available EO datasets have specific strengths and weaknesses with respect to mapping biomass:

- C-band SAR backscatter has been used to map the biomass in northern hemisphere boreal and temperate forests but these methods are unlikely to extend to dense tropical forest because of saturation effects;
- L-band SAR backscatter has greater sensitivity to biomass than C-band but still saturates at higher biomass levels;

	GlobBiomass		Page 7
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- ICESAT GLAS has been used to measure vertical canopy structure and forms the crucial dataset underlying the Saatchi et al. (2011) and Baccini et al. (2012) pan-tropical biomass maps;
- Optical data (e.g., MODIS, Landsat, SPOT etc.) are sensitive to some forest properties, allowing for biomass to be inferred by regression under some conditions;
- SRTM interferometric height, available for latitudes between 60°S and 60°N, is related to canopy height and density but needs conversion to obtain AGB
- TanDEM-X interferometric height is closely related to canopy height but it requires a DEM of the underlying surface and a value for the extinction of the X-band microwave in the specific forest structure to allow an estimate of forest height. In addition, conversion to AGB relies on allometry, a conversion that at global scale contains large uncertainties.

As a result of this analysis, the strategy followed to generate a global dataset of forest biomass in the GlobBiomass project was to identify the potential of each dataset and/or their combination to estimate biomass.

It is preferred to estimate biomass from EO datasets rather than combining existing datasets reporting biomass in order to avoid propagating errors and uncertainties of each individual biomass estimate in a new data product. The latter approach was recently followed under the GEOCARBON project to generate a global forest biomass dataset from three existing wall-to-wall datasets (Saatchi et al., 2011; Baccini et al., 2012; Santoro et al., 2015a), weighted with local reference datasets to minimize the impact of errors on the final biomass estimates in the tropics (Avitabile et al., 2015).

The strength of the GEOCARBON biomass dataset is that it for the first time provides global estimates at moderate resolution (0.01°). This improved the knowledge about the spatial distribution of biomass with respect to the only global dataset of biomass estimates available until then, namely the IIASA FAO global dataset of forest variables (Kindermann, 2008). Biomass attributes (aboveground and growing stock volume) were derived from NPP and FAO FRA estimates with a spatial resolution of 0.5° (Kindermann, 2008). Besides the coarse resolution, the major caveat was that the product embodies the errors and uncertainties in the values reported by individual nations for the FRA.

Nonetheless, the GEOCARBON dataset has a number of weaknesses:

- Errors and uncertainties in the tropics, where two datasets were available, were minimized only in areas where reference datasets were available, which covered a minor fraction of the tropical zone. A comparison of the pan-tropical maps by Saatchi et al., 2011, and Baccini et al., 2012, revealed large discrepancies at local scale (Mitchard et al., 2013). A list of possible explanations is provided in Mitchard et al. (2013). In our opinion, the major weaknesses of the estimation approaches in the two papers are the assumptions that: (i) continental scale models for converting GLAS height metrics to biomass can be considered valid across all tropical landscapes; and (ii) optical data can be used to reliably extrapolate estimates of biomass starting from allegedly well mapped individual pixels where height metrics from ICESAT were converted to biomass. Only at the level of nationwide averages did the two datasets agree, suggesting that the pan-tropical datasets are unbiased, but it is unclear which dataset, if either, portrays the spatial distribution of biomass correctly. Interestingly, the spatial distribution of biomass in the GEOCARBON dataset also differs from the two input datasets. From the point of view of a project aiming at generating a global dataset of biomass, these discrepancies indicate that an approach different to those used in Saatchi et al., 2011, Baccini et al., 2012, and, ultimately, Avitabile et al., 2015, needs to be considered.
- For latitudes north of 30°N, the GEOCARBON dataset corresponds to the biomass estimated from the estimates of forest growing stock volume (GSV) described in Santoro et al., (2015a) and converted to AGB using the biomass conversion and expansion factors (BCEF) published by the IPCC (IPCC, 2006). Unlike the pan-tropical maps, the estimation of biomass for the northern hemisphere forests relied on Envisat ASAR C-band backscatter measurements,

	GlobBiomass		Page 8
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

which are supposedly more related to forest structure, and thus to biomass, than optical data. Nonetheless, the GSV estimates were reported to be underestimated in dense mature forest due to the very weak sensitivity of C-band backscatter to forest structural variables at high biomass levels. Underestimation also occurred in patchy forest landscapes. Large uncertainties were reported in temperate and sub-tropical forest where the forest/non-forest backscatter contrast was rather small. Finally, the conversion from GSV to AGB was based on simple, biome-specific BCEFs that do not take into account the complexity of the forest landscape in terms of genus and wood density, as is done by Thurner et al. (2014). Here, the authors derived carbon stocks of the northern hemisphere from the same GSV dataset but using multiple available datasets on structural and biophysical forest variables. While such experimental datasets exist for boreal and temperate forest, they are scarce for other biomes, which is why ultimately biomass was estimated from GSV with the IPCC BCEFs to obtain the dataset input to the GEOCARBON AGB map.

- The three wall-to-wall datasets (Saatchi et al., 2011; Baccini et al., 2012; Santoro et al., 2015a) were based on EO images acquired in different epochs (year 2000, 2007-2008 and 2010) and thus do not represent the true distribution of forest biomass at one point in time globally.

Recently, a global dataset of forest aboveground biomass with a 0.01° spatial resolution has been released (Hu et al., 2016). Biomass was estimated with a non-parametric approach applied to spaceborne LiDAR, optical imagery (NDVI) and forest inventory data. The authors reported improved estimates with respect to existing continental and global datasets. Despite the tremendous effort, our first analysis of the dataset indicated some deficiencies, which in our understanding are due to incomplete characterization of the training dataset and the well-known limitations of NDVI to resolve biomass levels.

1.2 Possible solutions towards a novel global biomass retrieval scheme

Regression-based techniques based on parametric semi-empirical or empirical regressions or non-parametric regression trees have been widely used to model various EO observables as a function of biomass or growing stock volume (Reese et al., 2003; Houghton et al., 2007; Blackard et al., 2008; Nelson et al., 2009; Tomppo et al., 2013; Gallaun et al., 2010; Lucas et al., 2010; Saatchi et al., 2011; Baccini et al., 2012; Kellndorfer et al., 2012; Avtar et al., 2013; Carreiras et al., 2013; Beaudoin et al., 2014; Cartus et al., 2014; Mermoz et al., 2014). This requires the model to be trained with the aid of a training dataset that should be representative of the entire range of biomass in the area covered by the image and the number of samples should be large enough to ensure that the uncertainty in the model parameter estimates is small. A major problem is the lack of *in situ* AGB data allowing for a global model to be adapted to local conditions, although there are large databases from inventory (Reese et al., 2003; Blackard et al., 2008; Tomppo et al., 2013; Beaudoin et al., 2014). Local estimates have also been extrapolated over continental-scale regions (Saatchi et al., 2011; Baccini et al., 2012).

Methods that do not require a training dataset can be based on physically-based models that relate one or more SAR observables to a forest variable, usually GSV or AGB (Askne et al., 2009; Cartus et al., 2011; Santoro et al., 2011; Askne & Santoro, 2012; Cartus et al., 2012; Santoro et al., 2015b). Here the model parameters correspond to properties of the scattering from a forest, e.g., in the Water Cloud Model (Attema & Ulaby, 1978; Pulliainen et al., 1996; Askne et al., 1997) there are parameters corresponding to the backscatter from unvegetated surfaces and the vegetation canopy. Auxiliary information can be used to estimate the parameters locally based on statistics of the SAR data. Such methods have been shown to perform well in boreal and temperate forests at regional (Cartus et al., 2011; Cartus et al., 2012; Santoro et al., 2013) and continental (Thurner et al., 2014; Santoro et al., 2015b) scales, but they have not yet been demonstrated in tropical environments

(though there are initial studies using ALOS PALSAR (Santoro et al., 2014) and Envisat ScanSAR data (Santoro et al., 2015b)).

These considerations are brought together in Figure 1-1, which illustrates our understanding of how a global biomass approach should be structured and which features are necessary. The input consists of a suite of EO data acquired during 2010 and/or around the 2010 epoch, as well as a suite of auxiliary datasets used in support of model training. The list of sensors in Figure 1-1 represents the current availability of EO data at moderate resolution (< 100 m) and with global coverage for the selected epoch. Classification and segmentation was originally conceived to generate clusters for which one of the inversion models would be applied. Allowing a retrieval algorithm to be driven by ancillary data can cause estimation errors because of the wrong allocation of the inversion algorithm. For this reason, we prefer to operate on a pixel-by-pixel basis, adapting the retrieval and inversion models to perform at pixel level. Accordingly, the final AGB map is seen as a result of merging the individual estimates from the inversion methods with rules defined at pixel level. Multiple inversion models are here considered since biomass retrieval approaches developed so far have always been applied at regional and continental level, or at biome level at most. It is unrealistic to believe that a single approach, i.e., a single model, can meet the requirements of this project (retrieval error below 30% with a spatial resolution better than 500 m) because of the heterogeneity of forests worldwide and the sub-optimal EO datasets available to map biomass.

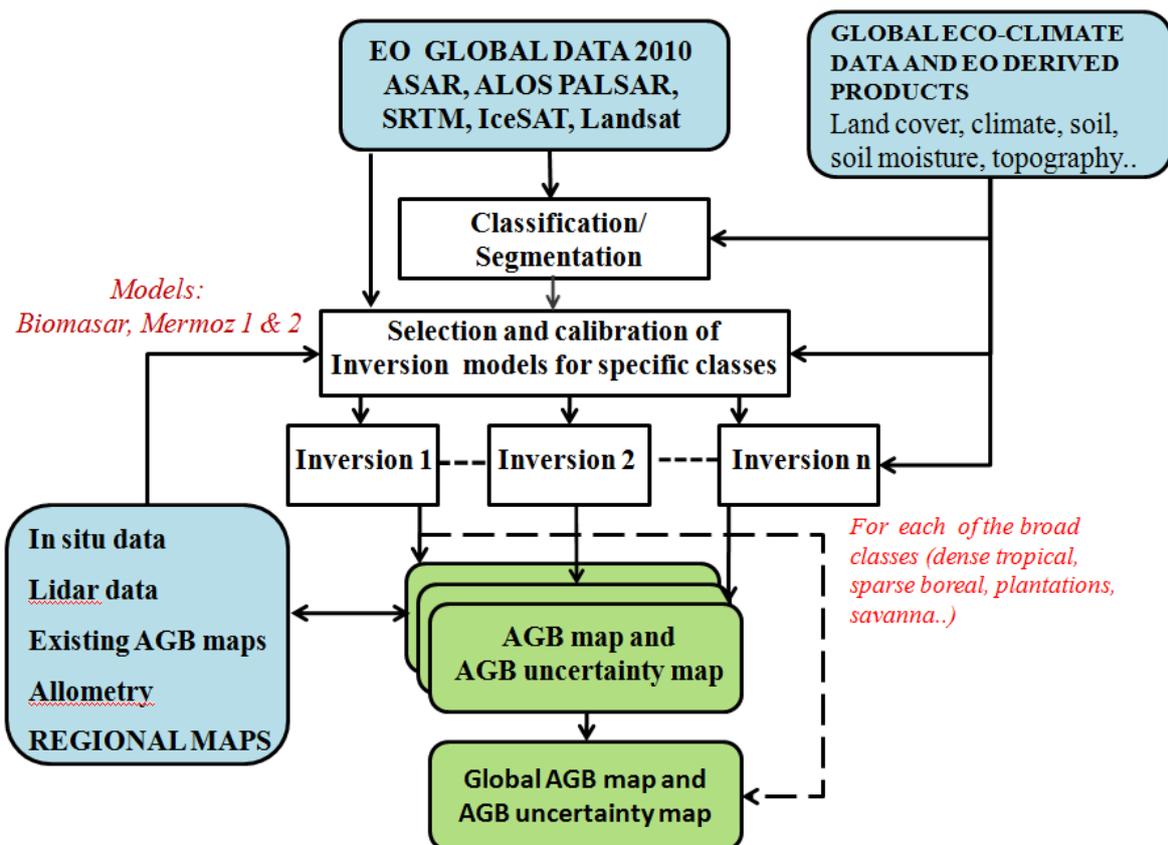


Figure 1-1. Originally proposed concept of the global mapping approach; Figure 1-20 gives the current structure considered for global biomass mapping.

	GlobBiomass		Page 10
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.3 Datasets

This section provides details on the datasets used in the global biomass retrieval algorithm developed and implemented in GlobBiomass. We differentiate between datasets used by the retrieval algorithms (observations) and datasets used in support of the training of the algorithms (auxiliaries).

1.3.1 Input datasets

1.3.1.1 Envisat ASAR (C-band, wavelength 5.6 cm)

During the Envisat mission (2002-2012), the ASAR instrument operated over land in four modes. Image Mode and Alternating Polarization Mode (spatial resolution < 30 m, swath width approximately 100 km) provided frequent monitoring, but with spatial coverage too sparse for global biomass estimation. The 150 m resolution Wide Swath Mode (WSM) had a 405 km swath and provided multiple observations of a target during the repeat-pass cycle (35 days until October 2010, 30 days thereafter). For example, at 60°N, observations every three days were possible, albeit at incidence angles between 18° and 43°. ESA also generated Image Mode Medium (IMM) and Alternating Polarization Medium (APM) datasets at the same resolution as the WSM (Desnos et al., 2000). The Global Monitoring (GM1) mode, which was operated as a background mission when the instrument was not meeting other requests, also had a 405 km swath but with 1 km resolution. From the 10 years of acquisitions, the ASAR archive contains acquisitions in one specific mode for virtually anywhere on Earth. These are publicly available on a rolling archive in ESA's Grid Processing On Demand (G-POD) facility, which is designed for processing large amounts of data using code provided by the user. During the Envisat mission, the ASAR archive was populated with data starting July 2005 for IMM, APM and WSM and November 2004 for GM1. Recently, ESA completed the archive with remaining data.

Figures 1-2 to 1-4 show the number of ASAR observations over land for IMM, WSM and GM1, respectively, in 1°×1° grid-cells for the time period July 2005-April 2012 (<https://wiki.services.eoportal.org/rss-storage-data.php>). Near-global, dense datasets of backscatter were obtained in GM1, and when complemented by WSM and IMM, there are no significant gaps (Figure 1-5), except over islands in Oceania and isolated islands in the southern hemisphere. The spatial distribution of observations applies for each full year between 2006 and 2011 as well.

IMM, WSM and GM1 data were acquired as either HH or VV polarization. Both polarizations give a similar relationship between backscatter and forest variables over the 20° to 45° range of incidence angles of ScanSAR (Castel et al., 2001), so they are not differentiated in the retrieval approach.

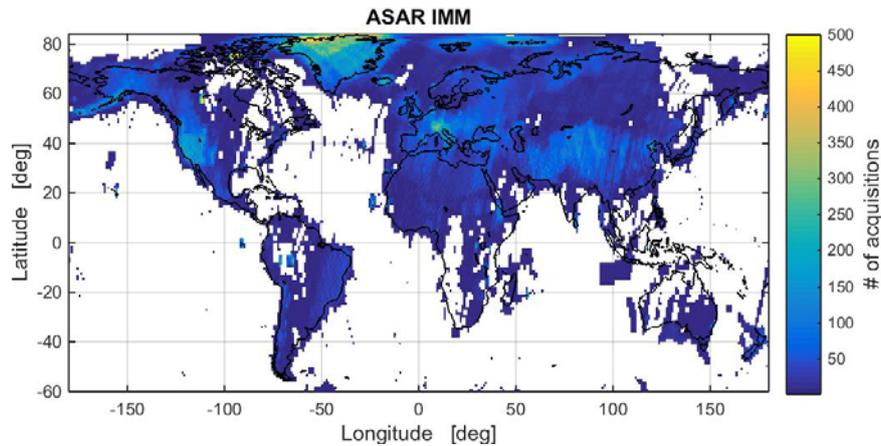


Figure 1-2. Total number of IMM observations per $1^{\circ} \times 1^{\circ}$ grid-cell (July 2005 - April 2012).

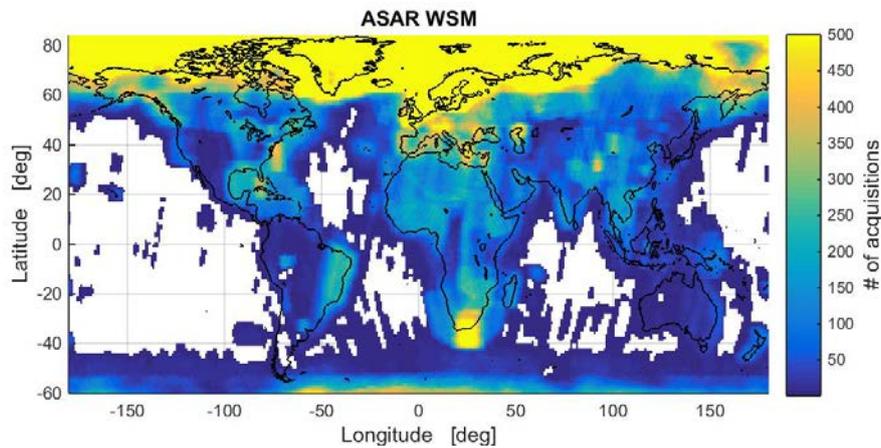


Figure 1-3. Total number of WSM observations per $1^{\circ} \times 1^{\circ}$ grid-cell (July 2005 - April 2012). The colour bar has been constrained to 500 observations to enhance the contrast in regions with a small number of observations. The largest number of observations was obtained for north Greenland (9,680).

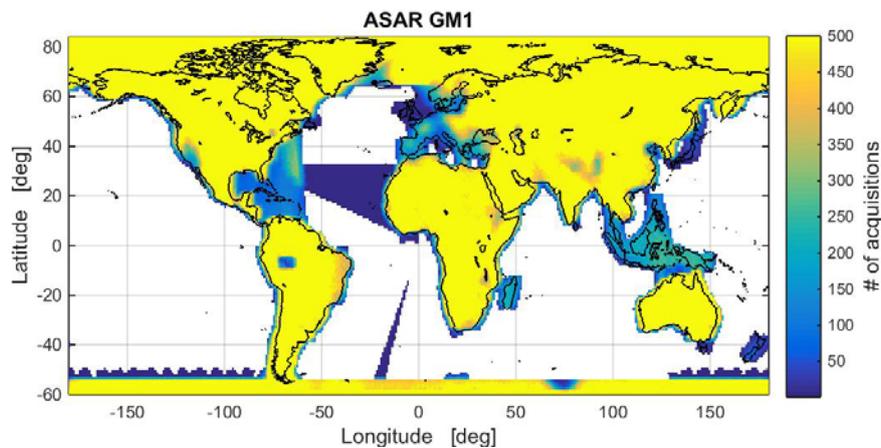


Figure 1-4. Total number of GM1 observations available per $1^{\circ} \times 1^{\circ}$ grid-cell (July 2005 - April 2012). The colour bar has been constrained to 500 observations to enhance the contrast in regions with a small number of observations. The largest number of observations was achieved over the arctic regions of Russia and Alaska (12,990).

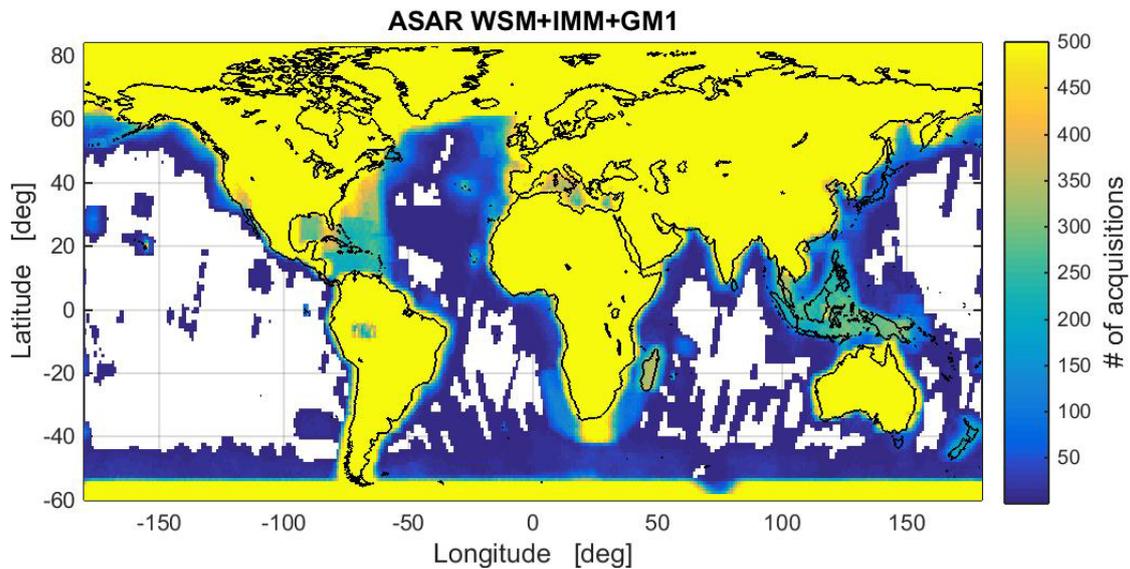


Figure 1-5. Total number of combined IMM, WSM and GM1 observations per 1°x1° grid-cell (July 2005 - April 2012). The colour bar is constrained to 500 observations to emphasize the spatial distribution. The largest number of observations was obtained for north Greenland (15,000).

As a joint endeavour of the ESA Climate Change Initiative Land Cover (CCI-LC), BIOMASAR and GlobBiomass projects, GAMMA has processed all IMM, WSM and GM1 images acquired over land between July 2005 and April 2012 to form stacks of co-registered images of SAR backscatter. The global biomass map for 2010 uses ASAR data acquired during 2010 with additions from 2009 and 2011 to achieve a hyper-temporal dataset sufficiently reliable to retrieve biomass. For the northern hemisphere, a 17-months time interval including two winter seasons (2009-2010 and 2010-2011) is used to maximize the amount of images acquired under the most favourable conditions to retrieve biomass (Santoro et al., 2015a). Accordingly, for the tropics, the hyper-temporal dataset is tailored to include two dry seasons, which increases the sensitivity of the SAR backscatter to biomass compared to wet environmental conditions. In regions with scarce data coverage around 2010, the ASAR dataset is complemented with observations from other years to avoid gaps (e.g., St. Pierre and Miquelon, Canada, with best coverage during 2007-2008). Adding IMM data fills some tectonic areas with sparse ScanSAR coverage (e.g., Italy, California) where ASAR operated almost exclusively in Image Mode. WSM and IMM data were processed on G-POD and shipped to GAMMA. GM1 were already available at GAMMA and processed on local servers.

1.3.1.2 ALOS PALSAR (L-band, wavelength 23 cm)

The ALOS dataset consists of yearly mosaics of HH and HV polarized backscatter acquired in Fine Beam Dual (FBD) mode between 2007 and 2010 and produced by JAXA (Shimada & Ohtaki, 2010; Shimada et al., 2014). Each includes data acquired between May and October of the given year. To achieve global land coverage, gaps were filled with data from other years. JAXA reported that the gaps were minor so that each mosaic truly represents the backscatter for a given year. For each pixel, the mosaic dataset provides:

- the HH and HV backscatter
- the local incidence angle with respect to the orientation of the pixel, derived from a Digital Elevation Model, (3-arcsec SRTM or 1-arcsec ASTER DEM), as well as layover/shadow masks
- the date of acquisition of the image

	GlobBiomass		Page 13
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- indication of whether the pixel is land or water

Figure 1-6 shows the HV backscatter for 2010; bright tones correspond to forested areas, while dark tones correspond to bare areas and water surfaces.

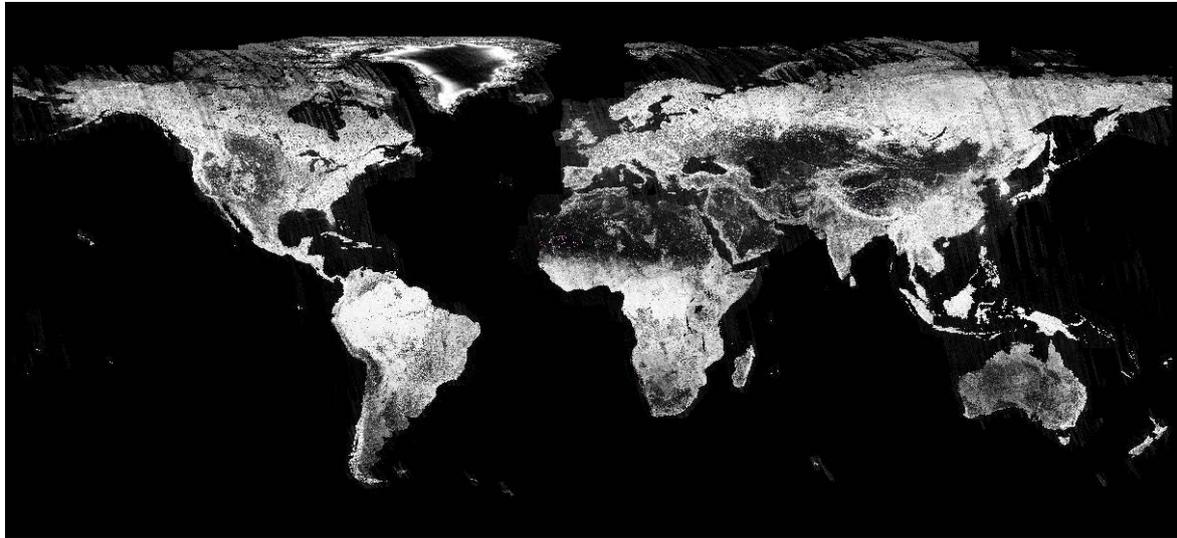


Figure 1-6. ALOS PALSAR mosaic of HV images acquired during 2010.

The data were processed to γ^0 , i.e., σ^0 divided by the cosine of the local incidence angle (Shimada, 2010), and were provided as $1^\circ \times 1^\circ$ tiles for all land masses with a pixel size of $1/4000^{\text{th}}$ of a degree in both latitude and longitude, corresponding to roughly 25 m at the Equator.

The weighted feathering approach applied by JAXA to smooth the backscatter differences between adjacent orbital tracks distorts the original backscatter and could cause errors in retrieval of biomass. In addition, although JAXA corrected the backscatter for slope-induced effects, residual striping is clearly visible in areas of strong topography, particularly in northern hemisphere forests.

As for C-band ASAR data, all available mosaic data are considered to avoid large biomass errors from using only single backscatter observations.

1.3.1.3 Landsat reflectance mosaics

In 2008, USGS released the multi-decadal archive of Landsat imagery free of cost. The Global Forest Change project of the University of Maryland (Hansen et al., 2013) has provided global (excluding Antarctica and Greenland) mosaics of Landsat reflectances in bands 3, 4, 5 and 7 for 2000 and 2013. Bands 1 and 2 were not considered because they are most affected by atmospheric effects (e.g., haze) and are least relevant for forest applications. The pre-processing of the archive of Landsat-7 Enhanced Thematic Mapper Plus (ETM+) observations in Google Earth Engine included image resampling; calibration to top of atmosphere (TOA) reflectances; and cloud/shadow/water detection. The mosaicking comprised: identification of cloud-free Landsat observations acquired throughout the growing season; image normalization with the aid of MODIS data (Potapov et al., 2012); and per-pixel calculation of the median reflectance of cloud-free growing season reflectances in each of the four bands. To fill gaps in the mosaics due to persistent cloud cover in the target years 2000 and 2013, observations from previous years were used. The global reflectance mosaic for band 3 is illustrated in Figure 1-7. The mosaics were released with a pixel spacing of 1 arc-second, i.e., roughly 30 m at the

	GlobBiomass		Page 14
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

equator. For GlobBiomass, the global reflectance mosaic for 2013 was downloaded from Google Earth Engine at:

http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.1.html.

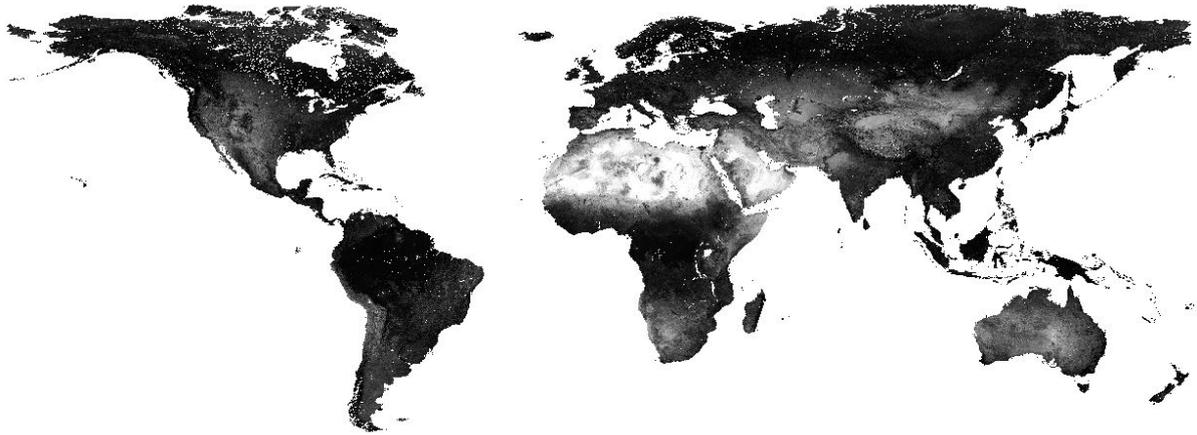


Figure 1-7. Landsat Band 3 growing season TOA reflectances for the year ca. 2013.

1.3.2 Datasets used for training

1.3.2.1 *In situ* measurements

One major aspect of the retrieval approaches here presented for global biomass retrieval is the independence from *in situ* data or reference maps.

1.3.2.2 MODIS Vegetation Continuous Fields

The MODIS VCF product (DiMiceli et al., 2011) estimates the percentage cover of woody vegetation, herbaceous vegetation and bare ground. It is derived from all seven bands of MODIS (Hansen et al., 2003) and is available annually for 2000 to 2010 with a spatial resolution of 250 m. Collection 5 is the current version. Data are provided in an equiangular projection in tiles of 12°×12° with a pixel size of 0.00208° from the Global Land Cover Facility (GLCF) and can be accessed via an ftp server (ftp://ftp.glcg.umd.edu/glcg/Global_VCF/).

1.3.2.3 Landsat canopy density and density change

Following initial tests for the US and Mexico (Hansen *et al.*, 2011), a first global Landsat-based canopy density map representing forest state in 2000 (Figure 1-8) was released in the framework of the Global Forest Change project (Hansen et al., 2013). To map canopy density, a suite of multi-temporal reflectance metrics (maximum, minimum, various percentiles) was calculated for the global Landsat dataset (see Section 1.1). These were used in regression tree models, trained with the aid of very high resolution imagery (e.g., Quickbird) classified to forest/non-forest classes, to map canopy density globally at 1 arc-second pixel posting. More recently, a global canopy density map for the year 2010 has been released.

The same multi-temporal metrics were also used to produce global 30 m maps of forest cover change, including information about annual forest cover loss between years 2000 and 2013, as well as gain throughout the entire thirteen year period. The forest cover change database includes a 30 m water body map, but no information was available about how this map was generated. Visual

inspection and comparison of the product with high resolution imagery in Google Earth showed good agreement with and better accuracy than, e.g., the water body map provided by JAXA with the PALSAR mosaics.

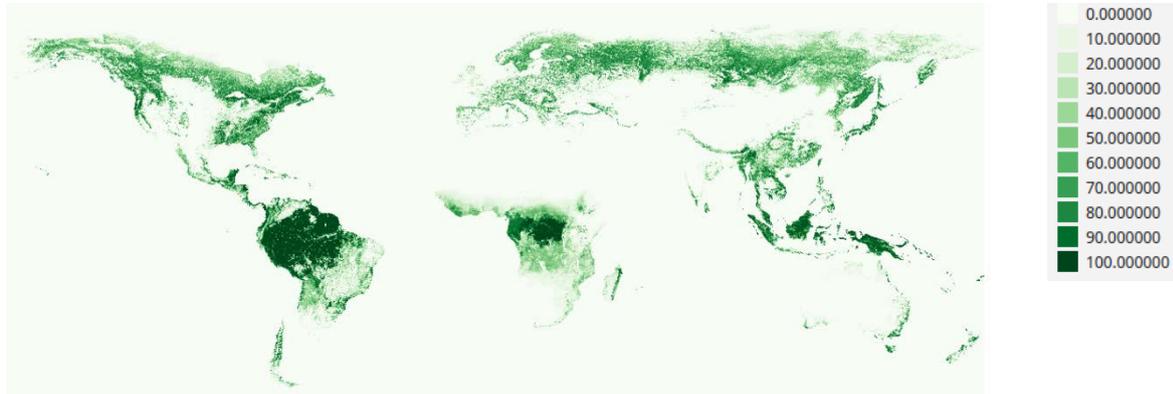


Figure 1-8. Canopy density map for the year 2000 produced from Landsat (Hansen et al., 2013) (

1.3.2.4 ICESAT GLAS

Although primarily designed for altimetry, between 2003 and 2009 the Geoscience Laser Altimeter System (GLAS) on board ICESAT collected information about the vertical structure of forests in ca. 65 m large footprints collected every 170 m along track. The distance between tracks was of the order of tens of km and increased towards the equator. When forest cover lay within a footprint, the returned signal reflected the vertical distribution of matter, with the density, shape and reflectivity of leaves, needles and branches in each layer of the forest canopy determining the strength of the reflected signal from the respective layer. An example of the vertical distribution of energy returned from a forest (the “waveform”) is shown in Figure 1-9. Depending on the height and structure of the forest, the waveform will present a different extent and shape as well as a different number of peaks. The beginning and end of the waveform are determined based on a threshold defined relative to the noise floor. The height of the first (from the top of the canopy) and last (from the forest floor) returns was defined as the heights where the signal energy exceeded 4.5 times the mean noise level (Los et al., 2012). Their difference is referred to as the waveform length.

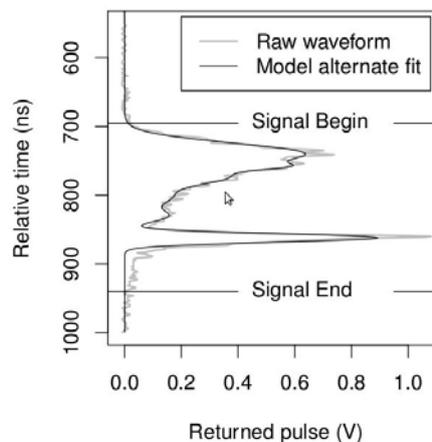


Figure 1-9. An ICESAT GLAS waveform showing the vertical distribution of returned energy from a forest (from Los et al., 2012).

	GlobBiomass		Page 16
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

ICESAT GLAS measurements can be downloaded at different processing levels. The GLA01 and associated GLA05 and GLA06 products contain Level-1 altimetry data for the transmitted and received raw waveforms (grey curve in Figure 1-9) as well as information required for geolocation, terrain and atmospheric corrections. In GlobBiomass we use the GLA14 product (version 34), which provides altimetry data for land surfaces only and for which geodetic, instrument, and atmospheric corrections have already been applied. The entire global archive of GLA14 data was downloaded from the National Snow & Ice Data Center (NSIDC). This provides information about the waveforms in the form of the parameters of a multi-Gaussian model fitted to the raw waveforms (black curve in Figure 1-9). For each footprint, up to six Gaussians were used to model the raw waveform, as described in Hofton et al. (2000). Table 1-1 provides an overview of the major parameters provided for each footprint in the GLA14 product. In addition, GLA14 provides a number of parameters relevant for identifying footprints affected by the atmosphere, signal saturation, low signal-to-noise ratio, etc. (not included in Table 1-1).

Table 1-1. Description of the major GLA 14 parameters.

GLA14 parameter	Description
i_lat	Latitude at footprint center
i_lon	Longitude at footprint center
i_DEM_elev	Elevation according to external DEM
i_elev	Waveform reference elevation (waveform centroid)
i_gpCntRngOff	Centroid Range
i_SigBegOff	Signal Begin Offset
i_SigEndOff	Signal End Offset
i_IdRngOff	Land Range Offset
i_numPk	Number of Gaussians fitted to the waveform
i_gpCntRngOff	Range Offset for each Gaussian
i_Gsigma	Sigma for each Gaussian
i_Gamp	Amplitude for each Gaussian
i_Garea	Area under each Gaussian

1.3.2.5 CCI Land Cover map

Land cover information is used during training to mask out specific classes, e.g., the BIOMASAR approach requires the identification of areas with low percentage canopy cover. At C-band, the SAR

backscatter of water bodies presents some of the highest and lowest values observed. The backscatter of urban areas is mostly higher than in forests. The backscatter of permanent snow/ice cover presents strong variability in time and space depending on the wet/dry conditions of the snow layer and the structure of the ice. If such cases are not accounted for, the estimate of the backscatter of unvegetated areas would be distorted, thus causing incorrect modelling of the forest backscatter and biased GSV estimates.

The Climate Change Initiative Land Cover (CCI-LC) project has recently released three global land cover maps for epochs 2000, 2005 and 2010 (), based on 300 m MERIS and 1 km SPOT-VEGETATION data. Envisat ASAR data were used to improve detection of water bodies. The overall accuracy of the 2010 land cover dataset is 76%, which is higher than other global land cover datasets. Nonetheless, commission and omission errors occur, particularly in mixed classes or areas of strongly heterogeneous land cover. A data layer giving the classification certainty and a set of quality flags are provided.

The land cover maps are provided in equiangular projection with a pixel size of 1/360th of a degree in latitude and longitude.

1.3.2.6 ERA Interim air temperature

The SAR backscatter is affected by environmental conditions at the time of image acquisition, e.g., C-band backscatter is more sensitive to GSV under frozen conditions (Santoro et al., 2011) because larger gaps in and increased transparency of the forest canopy allow greater microwave penetration. To account for this, the parameters that describe forest transmissivity in the model relating biomass to backscatter are forced to depend frozen/unfrozen conditions.. The environmental conditions are quantified by the "air temperature at 2 meters" surface parameter of the ERA-Interim reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF). This is provided globally with a time-step of 6 hours and a pixel spacing of 0.75°×0.75° in latitude and longitude.

1.3.2.7 Worldclim Bioclimatic Variables

The Worldclim data base (www.worldclim.org) developed by Hijmans, et al. (2005) includes a set of global maps on major bioclimatic variables that are derived from weather station data measured in the time frame 1950 to 2000. Data from various weather station networks, such as the Global Historical Climate Network or WMO Climatological Normals, were considered to produce a total of nineteen interpolated maps at 1 km pixel scale that provide information about annual climatic trends in terms of temperature and precipitation, such as the annual mean or the seasonal variability (Table 1-2). In GlobBiomass, the maps are used to support the estimation of the maximum GSV across forests in different ecosystems (see below).

Table 1-2. Overview of BioClim variables

BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter

	GlobBiomass		Page 18
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

1.3.2.8 FAO Global Ecological Zones dataset

The relationship between backscatter and biomass depends on forest structure. It is therefore necessary to stratify the model calibration according to forest type. The Global Ecological Zones (GEZ) dataset produced by the FAO (FAO, 2001) divides the land surface into 20 zones with “broad yet relatively homogeneous natural vegetation formations, similar (but not necessarily identical) in physiognomy. Boundaries of the EZs approximately coincide with the map of Köppen-Trewartha climatic types, which was based on temperature and rainfall. An exception to this definition is “Mountain systems”, classified as a separate EZ in each Domain and characterized by high variation in both vegetation formations and climatic conditions caused by large altitude and topographic variation” (Simons, 2001). The GEZ dataset is publicly available as a vector dataset, in equiangular map projection.

1.3.2.9 Global forest canopy height

Global estimates of forest height have been derived by Simard et al., 2011, using IceSAT GLAS waveform and MODIS products. The estimates are provided with a spatial resolution of 1,000 m and are representative for the year 2005. The estimates were validated with field measurements from 66 FLUXNET sites showing a RMSE of 6.1m with underestimation in dense, tall, broadleaf forest (e.g., Amazon). Spatially explicit estimates of forest canopy height are here used in the support of merging GSV estimates obtained with different retrieval approaches.

1.4 Methods

In this Section, the methods used for data handling (import, transformation, error characterization) are first described. We refer to these as pre-processing methods. Then, the methodology of the global biomass mapping algorithm is presented.

1.4.1 Data pre-processing

Pre-processing is the phase during which data, including both primary satellite data and auxiliary datasets, are converted to a form suitable for ingestion in the retrieval algorithms.

1.4.1.1 Envisat ASAR

The ASAR data are processed to obtain calibrated and speckle filtered images with sub-pixel co-registration accuracy, arranged in a structure that allows easy access and management (Figure 1-10) (Santoro et al., 2011; Santoro et al., 2015c). To this end, a global grid tied to the geographic reference system and having its origin at 90° N, 180° W, with tiles of relatively small size is used. The IMM and

WSM images (spatial resolution approximately 150m×150m and pixel size 75m×75m) are geocoded to a pixel size of 1/720th of a degree in latitude and longitude, and tiled into 1°×1° tiles (i.e., 720×720 pixels). The GM1 images (spatial resolution approximately 1 km×1 km and pixel size 500m×500m) are geocoded to a pixel size of 1/100th of a degree in latitude and longitude, and tiled into 2°×2° tiles (i.e., 200×200 pixels).

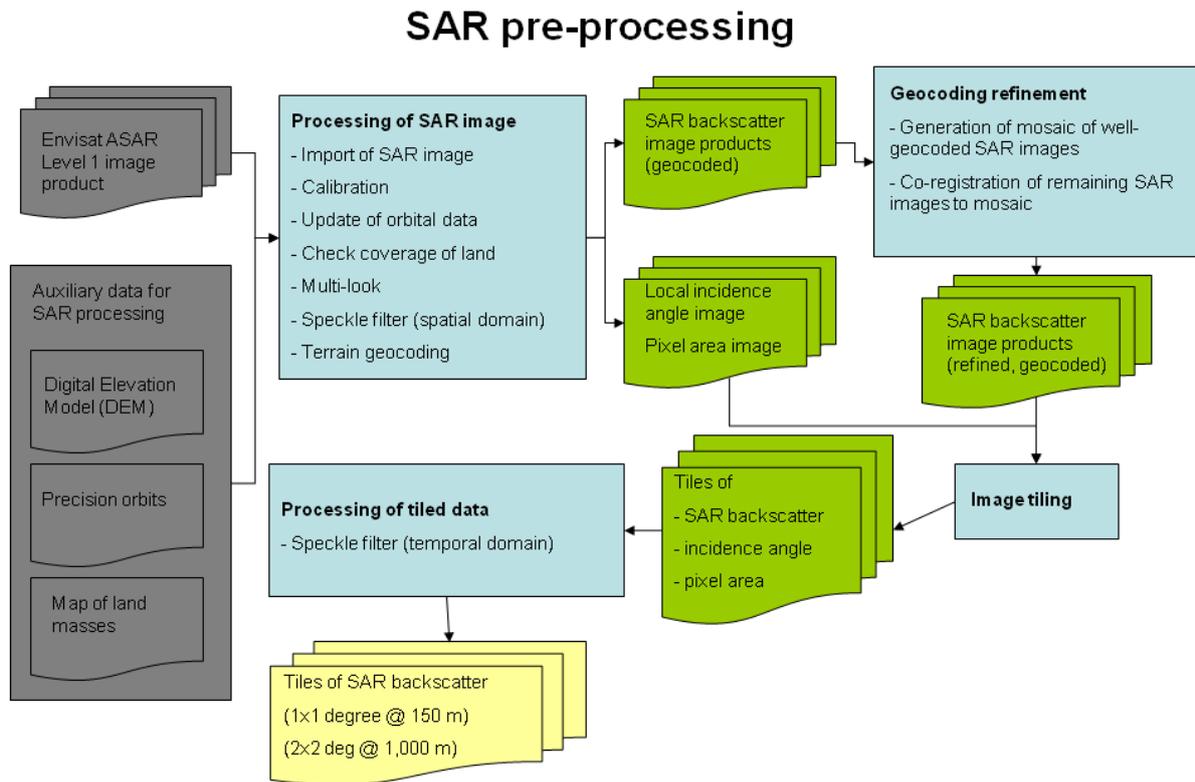


Figure 1-10. Flowchart of ASAR pre-processing.

The ASAR dataset consists of images of radar backscattered intensity (β^0) in ground range geometry. Each image is first calibrated to σ^0 using factors provided by ESA in the image metadata. Orbital state vectors are improved or extended using DORIS Precise Orbit State Vectors (<https://earth.esa.int/web/guest/data-access/browse-data-products/-/article/doris-precise-orbit-state-vectors-1502>). At this stage, the coverage of the ASAR image is checked against a coarse resolution map of land masses; images acquired over the ocean are discarded. The remaining images are spatially averaged using a 2×2 window to obtain a pixel size closer to the original spatial resolution. A gamma MAP filter (Lopes et al., 1990) is then applied to obtain an estimate of the local average backscatter, which is used as a weight in multi-channel speckle filtering.

Each ASAR image is geocoded to the geographic projection using a look-up table. The look-up table has the same geometry as the data in the output map geometry and provides at each pixel the coordinates of the pixel in the radar image. The look-up table is generated from orbital state vectors and viewing geometry parameters for the specific image (contained in the image metadata); the DEM is used to correctly locate each pixel on the Earth's surface. Cross-correlation between the SAR image and an image of the backscattered intensity simulated from the DEM is used to correct for errors in the look-up table arising from errors in the image metadata (Wegmüller, 1999). This procedure typically gives a geocoding accuracy below 0.3 times the size of the pixel.

	GlobBiomass		Page 20
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

We used the following DEMs to assemble a global DEM with a pixel size of 1/720th of a degree (i.e., nearly 150 m at the Equator):

- 3-arcsecond SRTM DEM (Rabus et al., 2003) for latitudes south of 60° N,
- digitized topographic maps of Eurasia (de Ferranti, 2009),
- Canadian Digital Elevation Database (CDED) (CCG, 2007)
- 7.5 minutes native DEM of Alaska (USGS, 2001) north of 60°N

The tiles were first mosaicked, preserving the original spatial resolution, and then resampled to 3-arcsecond. To obtain a corresponding version at 0.01°, each of the DEMs was averaged with a factor 11 and resampled to 0.01° using bilinear interpolation. The four datasets of elevation were finally mosaicked to obtain a global dataset of elevations with a pixel size of 0.01°×0.01°.

The geocoding procedure also generated images of local incidence angle and pixel area (relative to flat terrain). For images covering mostly flat terrain and characterized by poor co-registration with the DEM, the geocoding was refined by co-registering the geocoded SAR image with a mosaic of well geocoded ASAR images (Santoro et al., 2011). This ensures that all images are geocoded with sub-pixel accuracy.

Each ASAR image and the corresponding images of local incidence angle and pixel area are tiled to the predefined grid to obtain a multi-year data stack of observations of the ASAR backscatter with corresponding local incidence angles and pixel area. For each tile, a slightly modified version of the multi-channel filtering approach in Quegan & Yu (2001) is applied to the stack of backscatter images. It should be noted that this filter could not be applied before, e.g., in the radar geometry, because the stack of backscatter observations originate from images acquired along multiple orbital tracks both along ascending and descending orbits. In Quegan & Yu (2001), the local estimates of the backscattering coefficient are obtained by means of a moving average applied to the intensity images. Here, as noted above, we use the gamma MAP filtered images as local estimates of the backscattering coefficient. It is here remarked that that gamma MAP filtered images were obtained in the radar geometry (5 × 5 window) so that the number of looks of the filtered images were not affected by the different size of the pixel for different latitudes.

As shown in Figure 1-5, global coverage with hyper-temporal ASAR backscatter observations is only obtained when considering images from the IMM, WSM and GM1 modes. The correct procedure to obtain a data stack at 0.01° would require multi-looking each IMM, WSM or GM1 image to 1 km and geocoding to 0.01°. Substantial pre-processing activities are, however, beyond the scope of the project. Hence, integration of data from the three modes of ASAR can be achieved only by starting from the multi-resolution data stacks of images processed in the context of the CCI Land Cover project. The multi-channel filtered geocoded WSM and IMM images and the corresponding images of incidence angle and pixel area are multi-looked with factor 7×7 and resampled to the 0.01° pixel size using bi-cubic interpolation. A consequence of this procedure is that the normalization of the backscatter for images acquired in WSM and IMM modes might not be entirely correct. All scattering area corrections should be made as close as possible to the original resolution and then averages of the corrected power values should be taken. Averaging the power and area data separately and then making the corrections is further away from the observation physics of the SAR sensor.

This merged dataset is used to estimate biomass from the ASAR data, rather than estimating biomass at the original scale of the SAR data and then averaging, because of the difficulty of avoiding gaps in the estimates of the model parameters with the BIOMASAR algorithm at moderate resolution

	GlobBiomass		Page 21
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

(Santoro et al., 2011). In addition, multi-looking the ASAR data and estimating GSV from the 0.01° dataset is significantly faster than estimating GSV at the original scale of the ASAR data and averaging estimates to 0.01°.

To account for slope-induced distortions, the backscatter is normalized by compensating for the effective pixel scattering area and local incidence angle (Wiesmann et al.; 2004; Ulander, 1996; Castel et al., 2001).

$$\gamma^0 = \sigma^0 \frac{A_{flat}}{A_{slope}} \left(\frac{\cos\theta_{ref}}{\cos\theta_{loc}} \right)^n \quad (1-1)$$

In Equation (1-1), θ_{loc} and θ_{ref} represent the local incidence angle and a reference incidence angle (e.g., the incidence angle at mid-swath) respectively. A_{slope} and A_{flat} represent the true pixel area and the local pixel area for theoretically flat terrain respectively. The images of the area normalization factor (A_{flat}/A_{slope}) and the local incidence angle can be obtained from the DEM and orbital information (Wegmüller, 1999). For bare surfaces, the exponent n is equal to 1. For vegetated surfaces, n expresses the variation of the scattering mechanism due to the presence of a volume on sloping terrain, so is related to the optical depth of the vegetation. For C-band co-polarized data, it can be assumed to be equal to 1 (Ulander, 1996; Castel et al., 2001). Hence, the compensation corrects for the effect of terrain slopes on the backscatter but not for object-specific modulations of the backscatter due to slope and orientation (e.g., the effect of slope and orientation of trees on the backscatter).

The topographic normalization performed well for slopes tilted away from the sensor, but for slopes facing the sensor, the high sensitivity of the pixel area estimates to errors in the estimates of the local incidence angle caused residual distortions of the normalized backscatter. Therefore, steep slopes facing the sensor are masked out in GSV retrieval. Pixels with area ratio $A_{flat}/A_{slope} < 0.85$ usually distort the histogram of "ground" and "dense forest" pixels and are therefore excluded from the model training procedure.

Other methods to compensate for slope and orientation effects on backscatter (Small, 2011; Frey et al., 2013) give improvements at high resolution, but have little impact at kilometric scale. The two-step approach in Hoekman & Reiche (2015) tunes the correction of slope-induced effects with land-cover based empirical functions, but requires the biomass to be known a priori to select the appropriate function.

One known issue with the ASAR ScanSAR data is the sub-optimal inter-calibration of the swaths forming an image. The result is an offset of the backscatter across the seam between two adjacent swaths typically of the order of a few decimals of a dB but it can be several dB. This issue is critical in forests where the range of the C-band backscatter between unvegetated areas and dense forest is often 1-2 dB (see Table A1). Figure 1-11 shows an example for an ASAR GM1 image covering the rain forest of the Brazilian Amazon. The area covered by the image corresponds to very dense tropical forest (percent tree cover from the MODIS VCF data = 80%). The panel on the left shows a clear offset of the backscatter along a diagonal line, which corresponds to the seam between adjacent swaths of the ASAR image. A profile of the backscatter values along the dashed line superimposed on the ASAR image shows the clear offset at the swath intersection. The profile also shows additional effects which, however, do not seem to depend upon incorrect calibration of the data.

It is not possible to reverse the calibration of the data but we still want to avoid the offset in the backscatter image translating to offsets in the biomass map, so we applied a 1-D moving median

filter (length: 11 pixels) on each line of a geocoded image. The filtering is limited to forests that are most prone to retrieval errors, i.e., tropical rainforest and moist forest. The filter was therefore applied only to pixels that belonged to these classes according to the GEZ dataset. Figure 1-12 shows the result of filtering. The strong median filter reduces seams but also small-scale features, thus limiting the possibility to resolve small variations of biomass. This was considered to be of minor importance compared to producing a map with artefacts, given that C-band should not be able to estimate biomass with high accuracy in high biomass tropical forest.

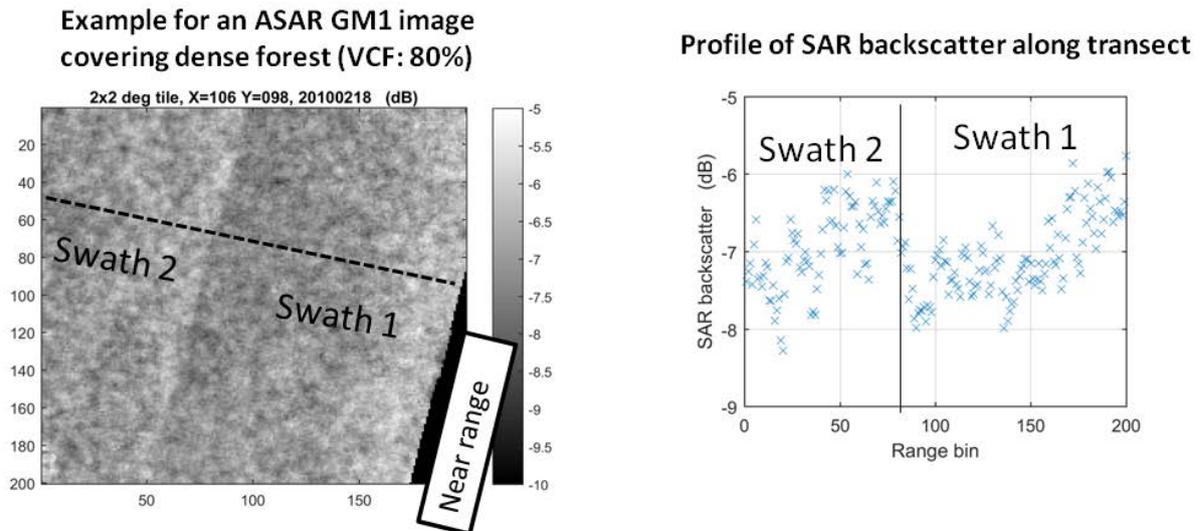


Figure 1-11. Illustration of backscatter offset along the seam between two adjacent swaths of an ASAR GM1 image covering dense tropical forest in the Amazon (left panel). The profile of the backscatter along the dashed line superimposed to the ASAR image is shown in the panel on the right hand-side.

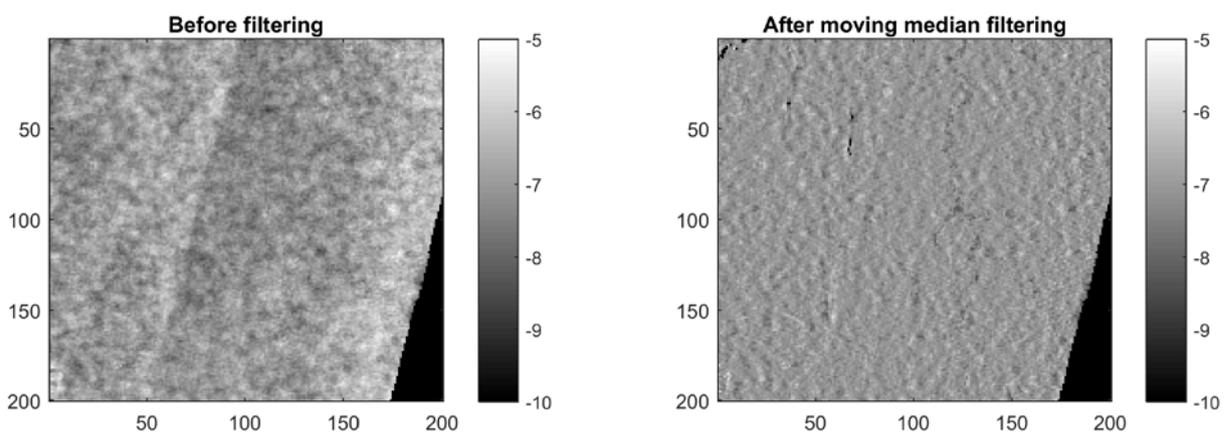


Figure 1-12. Example of the ASAR GM1 used in Figure 1-11 before and after filtering with a moving median filter.

1.4.1.2 ALOS PALSAR mosaics

The ALOS PALSAR datasets were already geocoded, orthorectified and calibrated. They had also been compensated for variations in the pixel scattering area due to topography and for the dependence of

backscatter on the local incidence angle (Shimada & Ohtaki, 2010). Pre-processing was therefore limited to speckle filtering the PALSAR mosaics using the multi-channel filter described in Quegan & Yu (2001). The filter requires an estimate of the local backscattering coefficient for each pixel and image, for which we used the local average in 9x9 pixel moving windows. After filtering, each of the M output images will have the same equivalent number of looks (ENL). The ENL after filtering using all four annual dual-polarization PALSAR mosaics (so M = 8) was estimated in selected polygons with homogeneous forest cover in the Amazon basin, and at full 25 m resolution found to be ~60. When aggregating the PALSAR images to coarser pixel sizes, the ENL increased linearly with the pixel posting (i.e., with the square root of the number of pixels that were aggregated, see Figure 1-13). As for the Landsat datasets, global PALSAR mosaics of the filtered backscatter intensity in virtual raster datasets (VRT) format were generated from the 1°x1° tiles provided by JAXA.

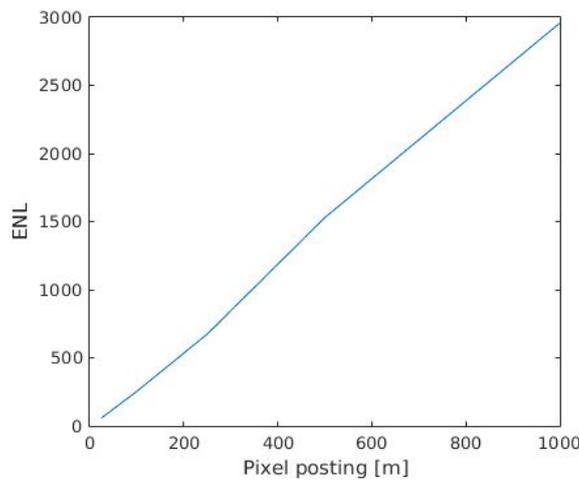


Figure 1-13. ENL of the PALSAR intensity images after multi-channel filtering at different pixel aggregation scales. The ENL was estimated for a number of homogeneous forest patches across the Amazon basin.

Since the PALSAR and Landsat mosaics are used jointly for estimating biomass, accurate co-registration of the two datasets is required. Cross-correlating small image chips distributed evenly across the images indicated sub-pixel co-registration accuracy, so no further co-registration was performed.

The yearly ALOS PALSAR mosaics were created using dual-polarization imagery acquired primarily throughout the northern hemisphere summer. Since they were acquired under similar environmental conditions, the differences in the 2009 and 2010 PALSAR mosaics were in most cases small. However, the sensitivity of the L-band backscatter to biomass can present substantial differences depending on environmental conditions, mostly for frozen/unfrozen and wet/dry conditions at the time of image acquisition (Harrel et al., 1997; Cartus et al., 2012; Santoro et al., 2006, 2015). Due to the limited availability of multi-temporal observations per year (Rosenqvist et al., 2007), JAXA had to use imagery that had been acquired under markedly different conditions than summer/fall for creating wall-to-wall mosaics locally. At high latitudes, for instance, some orbits acquired under frozen or freeze/thaw conditions were used. As a result, the mosaics present distinct striping effects because images acquired under frozen or freeze/thaw conditions exhibit significantly lower backscatter. In order to achieve the best possible retrieval accuracy, imagery acquired in 2010 was replaced with imagery from 2009 when the 2010 imagery presented clear environmental effects due to freezing

	GlobBiomass		Page 24
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

conditions or increased wetness. Images were replaced only if the 2009 imagery was not affected by similar distortions (Figure 1-14).

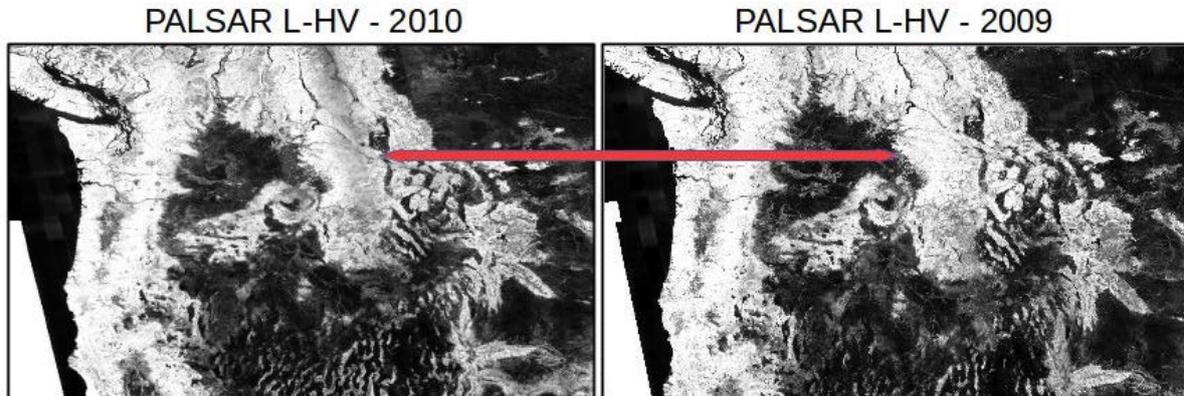


Figure 1-14. ALOS PALSAR L-HV mosaic for 2010 (left) and 2009 (right). The red arrow denotes an area in the mosaics for which differences in the imaging conditions between 2009 and 2010 lead to significant backscatter differences.

In the tropics, however, simply replacing 2010 PALSAR imagery with imagery acquired in previous years, i.e., acquisitions contained in the 2009 mosaic, would not allow for producing radiometrically balanced mosaics since all four available PALSAR mosaics presented similar levels of distortions.

In this case, we opted to minimize the pronounced striping effects observed over dense tropical rainforest using an empirical normalization approach similar to the procedure implemented for the ASAR data. For each image line along the range direction this:

- 1) identifies trends in L-HV backscatter over closed tropical forest canopies (i.e., only where Landsat reports 100% canopy density) using a line-by-line moving median filter of 500 pixels window length
- 2) normalizes the L-HV backscatter observed over closed tropical forest to the average backscatter level by subtracting line-by-line the median-filtered values from the actual observations and adding the residuals back to the average backscatter

The effect of the normalization is illustrated in Figure 1-15, which clearly shows that the normalization results in more consistent backscatter mosaics over the tropics. A drawback of the normalization is, however, that smaller scale variability in backscatter is smoothed out, i.e., some spatial detail over closed tropical forest canopies, which may or may not be associated with different biomass levels, is lost. While certainly not ideal, it is preferred to use the normalized backscatter mosaic because:

- 1) the modelling results presented above suggest that there is no significant sensitivity of L-band backscatter to GSV or AGB over closed tropical forest canopies
- 2) in a global map product, the visual effect of striping distracts users from the potential strengths of the product outside dense tropical forests

	GlobBiomass		Page 25
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

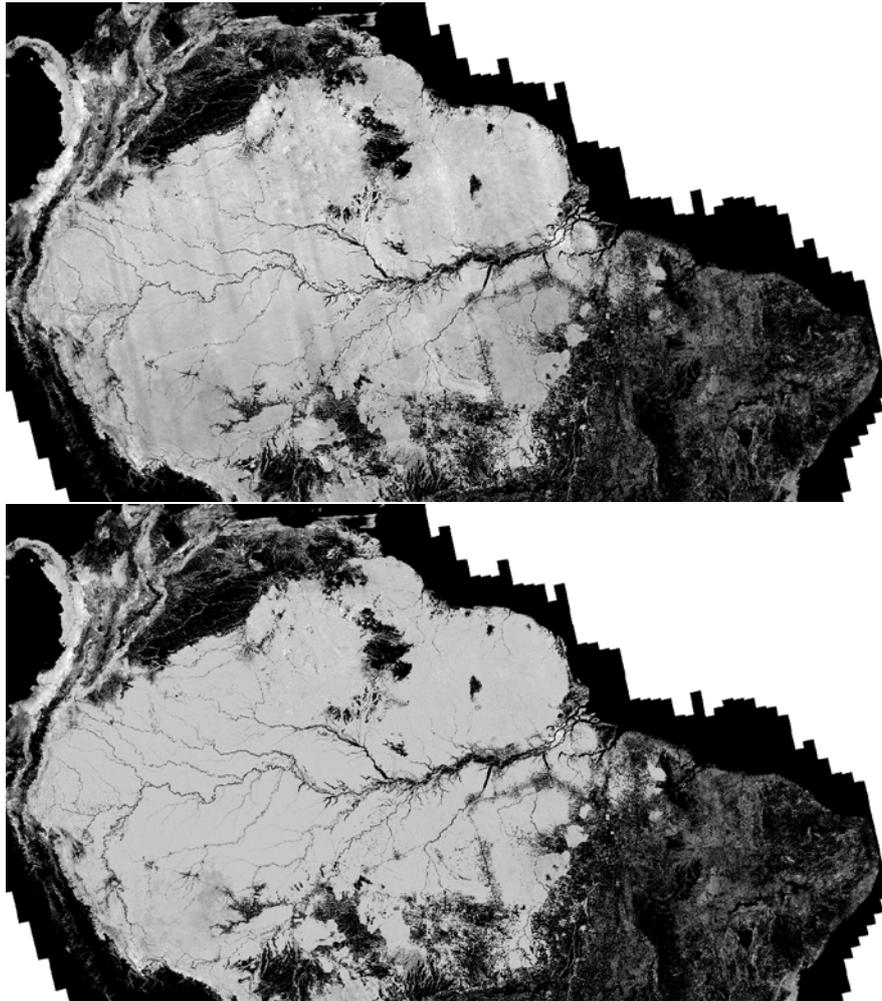


Figure 1-15. ALOS PALSAR L-HV mosaic for the year 2010 before (top) and after (bottom) radiometric balancing.

1.4.1.3 Landsat data products

The Landsat datasets were provided as 10°×10° tiles. Global mosaics for each of the four reflectance bands as well as the forest cover and forest cover change maps were generated in the form of VRT datasets using GDAL tools (www.gdal.org).

1.4.1.4 ICESAT GLAS

Within GlobBiomass, we are using GLAS data to estimate forest height following the approaches in Simard et al. (2011) and Los et al. (2012), which calculated RH100 globally and defined a set of filters to discard footprints affected by topography and various noise sources in the waveforms. We followed the filtering procedure in Los et al. (2012) to discard footprints with: i) erroneous terrain elevation estimates (based on GLAS parameters *i_elev* and *i_DEM_elev* in Table 1-1); ii) uncertain estimates of the ground peak location due to low amplitude and area under the ground peak (based on GLAS parameters *i_Garea* and *i_Gamp* in Table 1-1); and iii) unrealistic Gaussian fits in terms of the standard deviation (based on GLAS parameter *i_Gsigma* in Table 1-1). Following Simard et al. (2011), we also discarded footprints with: i) low signal-to-noise ratio; and ii) slopes exceeding 5°. The GLA14 product provides slope estimates based on the 3 arc-second SRTM DEM (for latitudes < 60° N). Using the recently released 1 arc-second DEM (SRTM-1), we re-computed the slope for each GLAS

footprint. Since only footprints covering forest are of interest, we also discarded all footprints with only one peak (GLAS parameter *i_numPk*). The remaining GLAS database contained estimates of RH100 for ca. 26.5 million footprints; their distribution is illustrated in Figure 1-16. In addition, we computed an estimate of the canopy density for each footprint as the ratio of energy received from the canopy (i.e., returns from above the ground peak) to the total energy received.

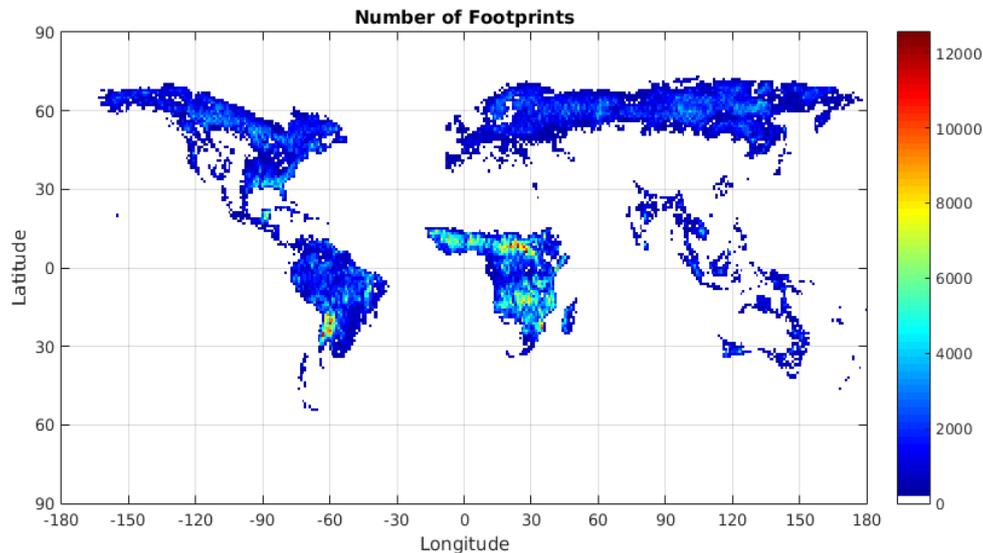


Figure 1-16. Distribution of GLAS footprints for which forest height was estimated. Darker colours indicate a large number of available estimates.

1.4.1.5 MODIS Vegetation Continuous Fields

The MODIS VCF estimates of percent cover of woody vegetation for 2010 are used to help train the ASAR forest backscatter model. This is undertaken at 1 km pixel size, so the original VCF estimates of percent canopy cover are aggregated to the pixel size closest to 0.01° and resampled to 0.01°.

1.4.1.6 CCI Land Cover

For calibrating the BIOMASAR algorithm using C-band data, the CCI land cover dataset for 2010 was resampled from its original pixel size of 0.002777° in latitude and longitude to 0.01° posting with majority voting. For this, the program *gdal_warp* of the Geospatial Data Abstraction Library was used. The original dataset is used to support modelling with the fine-resolution EO datasets.

1.4.1.7 ERA Interim

For a given pixel in an ASAR image, the air temperature nearest in location and closest in time is extracted from the ERA Interim dataset.

1.4.1.8 Global Ecological Zones

The GEZ dataset has been rasterized with a posting of 0.01° in latitude and longitude with the program *gdal_rasterize* of GDAL.

1.4.1.9 WorldClim Data Base

The nineteen maps depicting annual trends in temperature and precipitation have been aggregated to a 2x2 degree grid.

	GlobBiomass		Page 27
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.4.1.10 GSV of dense forest

Three of the four retrieval algorithms proposed in this document aim at minimizing the dependence on in situ information about biomass. Specifically, they all retrieve GSV from a set of input EO observations. Information about the local distribution of GSV in terms of the maximum value is required in order to constrain the retrieval based on spaceborne SAR imagery to a realistic range of biomass values. To this end, a data base on the maximum retrievable GSV has been compiled. A single value of the GSV assumed to be representative for densest possible forest, V_{df} , was assigned to the centre of each tile in a regular $2^{\circ} \times 2^{\circ}$ grid across the northern hemisphere. Where available, *in situ* measurements from field plots or spatially explicit datasets of GSV were used to estimate V_{df} , which was defined to be the 90th percentile of the GSV distribution within a tile (Santoro et al., 2011). In areas where datasets of AGB were available, the BCEF factors reported by the IPCC were applied. Elsewhere, V_{df} was estimated with an empirical piece-wise linear function (Santoro et al., 2015b) starting from values of the average GSV reported at provincial or national levels. For tiles including several provinces or nations, the average GSV representative for the tile was obtained by weighting the individual averages by the area of each within the tile. For areas lacking whatsoever form of biomass estimates, the IIASA FAO 0.5° dataset of global GSV was used. The dense forest GSV was set equal to the maximum of the 16 values within the $2^{\circ} \times 2^{\circ}$ large tile. A list of information sources used to compile the global $2^{\circ} \times 2^{\circ}$ data base on V_{df} is provided in Table 1-3. Figure 1-17 illustrates the data base in form of a raster map.

The assumptions behind the parameterization of V_{df} were very crude and should be considered a first-order approximation. Attaching a single value of V_{df} to areas covering thousands of km^2 was assumed to be sufficient to characterize V_{df} in regions including a single type of forest ecosystem, given that the variability of GSV for the densest forests is likely to be small at kilometric resolution. This approximation instead was incorrect in regions with transitions of ecosystems (e.g., tropical forest and savannah). Although, it was tried to ensure the spatial consistency of the estimates, this dataset of V_{df} has intrinsic errors and uncertainties propagating from the input datasets, the rules implemented to estimate V_{df} from the data available in a given tile and the BCEF values.

To reduce the caveats of the dataset, additional processing steps were then carried out. More specifically, the aim was to:

- 1) Improve the reliability of the data base, which in some areas, such as large parts of Africa or Southeast Asia, had to rely on the assumption that linear relationships between reported average stocks and local maxima exist
- 2) Fill gaps in the 2 degree data base
- 3) Increase the resolution to better depict smaller scale variations in the maximum GSV such as those expected for transition areas between tropical rainforest and savannah

The maximum GSV that is reached by forests across different ecoregions is expected to depend on natural factors such as temperature, precipitation or disturbance regime as well as on anthropogenic factors such as varying types forest management. In order to verify/improve the initial estimates for V_{df} and to fill gaps in the global database, a data base of predictor layers that are expected to have predictive power for V_{df} was compiled at $2^{\circ} \times 2^{\circ}$ resolution, including the nineteen WorldClim bioclimatic variables as well as ICESAT GLAS. For each $2^{\circ} \times 2^{\circ}$ degree grid cell, metrics were calculated from the local ICESAT GLAS footprints that characterize the distribution of forest height and density, i.e., quartiles of the distribution of GLAS height metrics RH100 and HOME (i.e., the Height Of Median Energy).

	GlobBiomass		Page 28
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

RandomForest (Breiman, 2001) models were then developed for each FAO ecoregion using the initial estimates for V_{df} (Figure 1-17) as response and the WorldClim and ICESAT GLAS layers as predictors. Once calibrated, the models were then used to predict V_{df} globally at a resolution of $0.2^\circ \times 0.2^\circ$ (Figure 1-18); note that for mapping at 0.2° scale with ICESAT GLAS interpolation was required to bridge gaps in the 0.2° maps of GLAS metrics associated with low counts of GLAS footprints. Figure 1-19 illustrates the predictive performance of the models for each FAO ecoregion at 2×2 degree resolution with the comparison of Out-Of-Bag model predictions (i.e., bootstrap aggregation of predictions from 500 regression tree models) versus the initial estimates for V_{df} . The high correlation of the initial estimates for V_{df} and the randomForest predictions based on GLAS and WorldClim variables suggests, with some regional limitations in particular in the subtropics:

- 1) high plausibility of estimates for V_{df} in the initial $2^\circ \times 2^\circ$ data base (Figure 1-17), and
- 2) high predictive power of WorldClim and GLAS forest structural variables for V_{df}

The reference data base linked with GLAS and BioClim variables would potentially allow for new insights into the role of environmental factors (precipitation, temperature regime) and anthropogenic disturbance on forest growth; such an investigation is however beyond the scope of this study.

The comparison of the reference database and randomForest predictions based on environmental and structural variables suggests that the standard error associated with randomForest predictions for V_{df} is of the order of 15 to 60 % with the largest error of 40 to 60 % for sub-tropical and tropical dry forests.

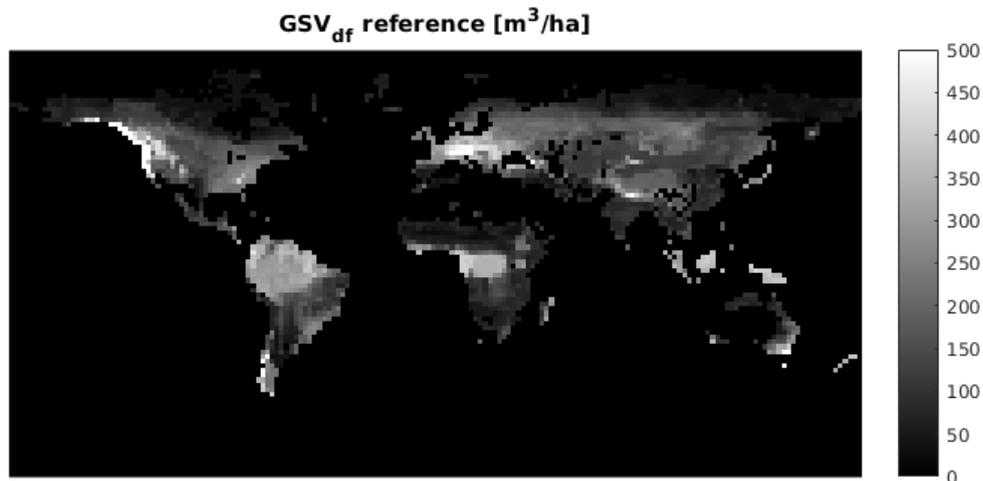


Figure 1-17. Estimates of V_{df} derived from inventory datasets, existing maps, or official forest statistics.

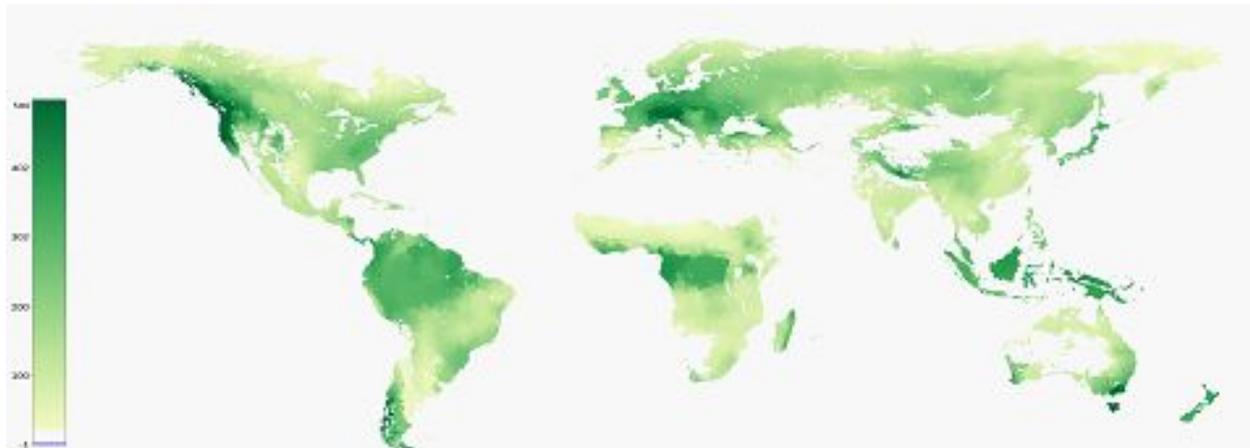


Figure 1-18. Map of the GSV of dense forests at 0.2°x0.2° resolution.

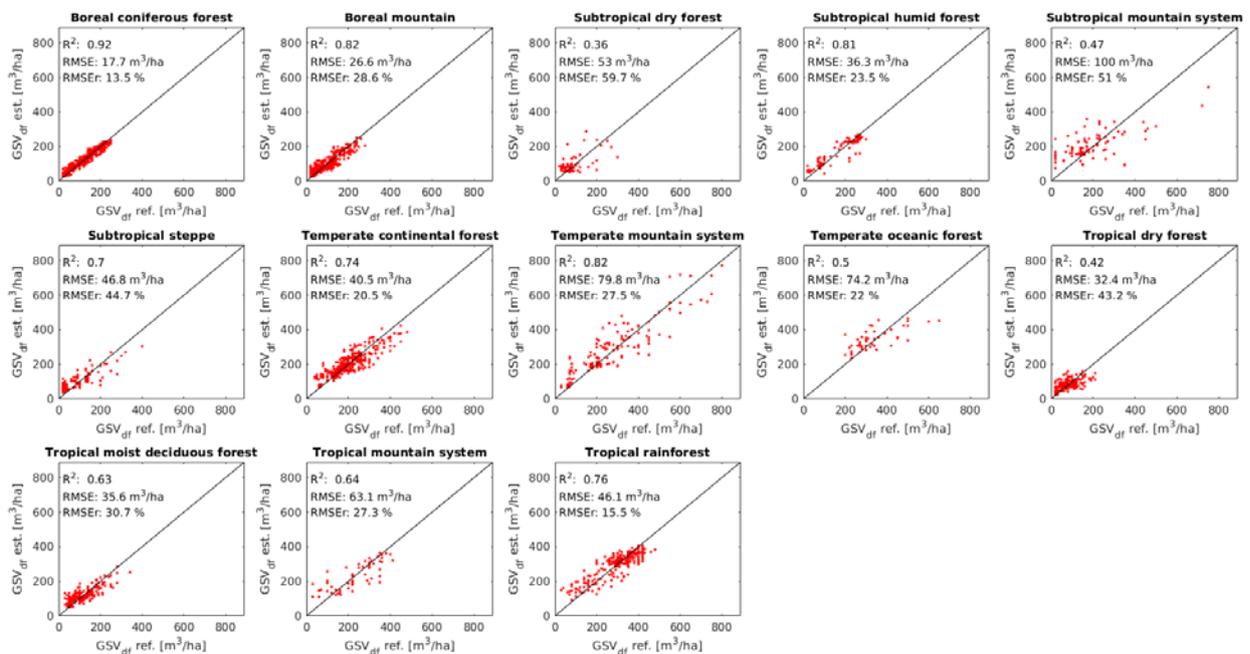


Figure 1-19. RandomForest predictions for the GSV of dense forests per 2°x2° grid cell derived from BioClim and ICESAT GLAS versus estimates derived from inventory, provincial/state reports or existing maps.

1.4.1.11 Global forest canopy height

Global estimates of forest height have been derived by Simard et al. (2011) using IceSAT GLAS waveform and MODIS reflectances. The estimates are provided with a spatial resolution of 1,000 m and are representative for the year 2005. The estimates were validated with field measurements from 66 FLUXNET sites showing a RMSE of 6.1m with underestimation in dense, tall, broadleaf forest (e.g., Amazon). Spatially explicit estimates of forest canopy height are here used in the support of merging GSV estimates obtained with different retrieval approaches.

	GlobBiomass		Page 30
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.4.2 Biomass estimation algorithms

The GlobBiomass estimation algorithm is the result of the detailed review of existing algorithms outlined at the beginning of this Section and their projection to a global context. In particular, a biomass retrieval algorithm that is expected to perform globally will need to adapt to forest types and environmental conditions worldwide. In this project, the idea was not to come up with a completely new approach, given the large range of existing, tested, calibrated, validated and inter-compared approaches and the lack of remote sensing observations that are markedly different from those supporting the investigations that led to the range of algorithms summarized at the beginning of Section 1. It is instead believed that algorithms suitable for global forest biomass retrieval exist; our aim is to identify and improve those with greatest potential and make them truly perform globally.

Parametric approaches are preferred to non-parametric approaches because the operator is in full control of the development of the algorithm and of its calibration. Physically-based approaches relating the SAR observables to the forest variable of interest are furthermore preferred to empirical relationships because empirical model and regression approaches are often developed at local sites and have not been generalized to a level that can be considered of sufficient performance in a global perspective. A final note regards the procedure to train the model relating the SAR observable to the forest variable of interest. Model training requires a dataset of *in situ* observations of the forest variable of interest, with high accuracy, with the same spatial scale as the SAR observations and with a representative distribution within the range of biomass values in the area of interest. While such requirements may be met locally, inventoried forests are a small fraction of forests worldwide. It is believed that accurate estimates of model parameters can be obtained for forests with *in situ* data, but estimates for forests that are under-represented or not represented at all in the training dataset may be erroneous because the model parameters are based on reference material insufficiently descriptive of the behaviour of biomass as a function of the SAR observables.

As a result of this analysis of the advantages and disadvantage of existing retrieval algorithms, we have selected four approaches. Each was already demonstrated to perform at regional to continental scale (Santoro et al., 2011; Cartus et al., 2012; Santoro et al., 2013; Mermoz et al., 2014; Santoro et al., 2015a); however, none can be considered the optimal solution to the challenge of generating a high-resolution, high-accuracy, global dataset of forest biomass. This is mostly because the EO datasets currently available at global scale are sub-optimal for retrieving forest biomass, especially in dense tropical forest. In addition, they all to a large extent generalize the relationship between biomass and SAR observables, so are unable to reproduce subtle differences in forest structure, which ultimately correspond to differences in forest biomass.

To overcome the limitations of each algorithm, we propose linkages to exploit the salient aspect of each in order to improve the biomass estimates with respect to those obtained with each of the algorithms. Figure 1-20 shows the structure of the proposed biomass retrieval algorithm, which contains three successive stages:

1. A global dataset of GSV is derived from the hyper-temporal dataset of ASAR backscatter images with a pixel spacing of 0.01° using the BIOMASAR algorithm (Santoro et al., 2011; Santoro et al., 2015a). To avoid confusion with other implementations of the same approach using data from other sensors, we refer to this as BIOMASAR-C. This approach was shown to perform well in the northern hemisphere, reproducing correctly the spatial distribution of the forest biomass and providing accurate estimates of GSV (Santoro et al., 2015a) and AGB

	GlobBiomass		Page 31
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

(Turner et al., 2014) at regional scale by averaging individual retrieved values. Despite the strong performance of the retrieval based on the BIOMASAR-C algorithm and hyper-temporal observations of C-band backscatter, it does not suffice to fulfil the requirement of a global map of forest AGB with an accuracy of 30% or better and a spatial resolution below 500 m. In addition, there are evident limitations of C-band SAR for estimating GSV in dense forests and tropical environments (Santoro et al., 2015b) or in patchy landscapes with a mix of forest and other land cover types. A global GSV map obtained with BIOMASAR-C is therefore treated in the first instance as an indicator of biomass, supporting the retrieval with EO data at higher spatial resolution and stronger sensitivity to forest biomass. The retrieval approach is outlined in Section 1.4.2.1. This indicator is as input in an upscaling approach together with high resolution L-band and optical data (Section 1.4.2.3).

2. The bulk of the GlobBiomass retrieval corresponds to the second stage which includes three retrieval approaches applied to high-resolution SAR data. Global estimates of biomass are obtained with two of the three approaches. Multiple approaches are considered to reduce potential flaws in one of the approaches due either to the input dataset or the simplified modelling. ALOS PALSAR backscatter is used as a predictor in a model-based approach (referred to as BIOMASAR-L) and in a re-scaling approach of the BIOMASAR-C estimates together with Landsat reflectances. Both approaches are designed to be calibrated without *in situ* measurements and retrieve GSV at the spatial resolution of the input EO data, i.e., approximately 25 m. BIOMASAR-L mimics the BIOMASAR algorithm developed for C-band and is therefore referred to as BIOMASAR-L (see Section 1.4.2.2). The re-scaling of BIOMASAR-C estimates of biomass is tested with two algorithms. Each upscales the BIOMASAR-C estimates from kilometric to decametric resolution (see Section 1.4.2.3). Strengths and weaknesses of each of the two re-scaling solutions are discussed in Section 1.4.2.3. The re-scaling approach is referred to as BIOMASAR-C+ and has been developed to complement the retrieval with BIOMASAR-L in areas of poor performance of the retrieval based on single L-band observations (e.g., very low biomass) or systematic effects, e.g., due to topography or events altering the ALOS PALSAR backscatter on the specific acquisition date. The third approach referred to as the CESBIO method (Section 1.4.2.4) inverts a parametric model with a Bayesian approach to retrieve AGB. Here, the model training phase is based on *in situ* measurements of biomass and an electromagnetic scattering model. The retrieval approach is parameterized only for a limited range of biomass (< 100 Mg/ha) appropriate to the dry tropics. Hence the AGB estimated with the CESBIO approach is used to benchmark the performance of the other two high-resolution approaches in tropical forest and the merging of estimates. To allow for a comparison of estimates, the GSV obtained with the BIOMASAR-L and the BIOMASAR-C+ approaches are transformed to AGB with BCEF published by IPCC (IPCC, 2006).
3. For each pixel the final estimate of biomass is obtained by weighting the BIOMASAR-L and the BIOMASAR-C+ estimates with weights defined by their similarity to theoretical behaviour when related to forest canopy height, percent tree cover and forest transmissivity. The integration of the biomass estimates is parameterized at the ecozone level. The integration is currently implemented to combine biomass expressed as GSV. In this way, the impact of errors and uncertainties of the GSV to AGB conversion are decoupled from the retrieved values. Conversion of GSV to AGB is implemented as the last stage.

	GlobBiomass		Page 32
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

Each map of GSV and AGB has a corresponding map of the estimates' accuracy (Figure 1-20).

The biomass (GSV and AGB) estimates are obtained at 0.00022° corresponding roughly to 25 m. Given that the retrieval is based on a single L-band observation and C-band estimates rescaled from 1,000 m to 25 m, the information content of the estimates at this resolution is believed to be poor. To fulfil the requirement of a maximum 30% retrieval error set in the Statement of Work, the final estimates are aggregated to lower resolutions generating a set of estimates at degraded resolution starting with 100 m (i.e., 0.001°) up to 1,000 m (i.e., 0.01°).

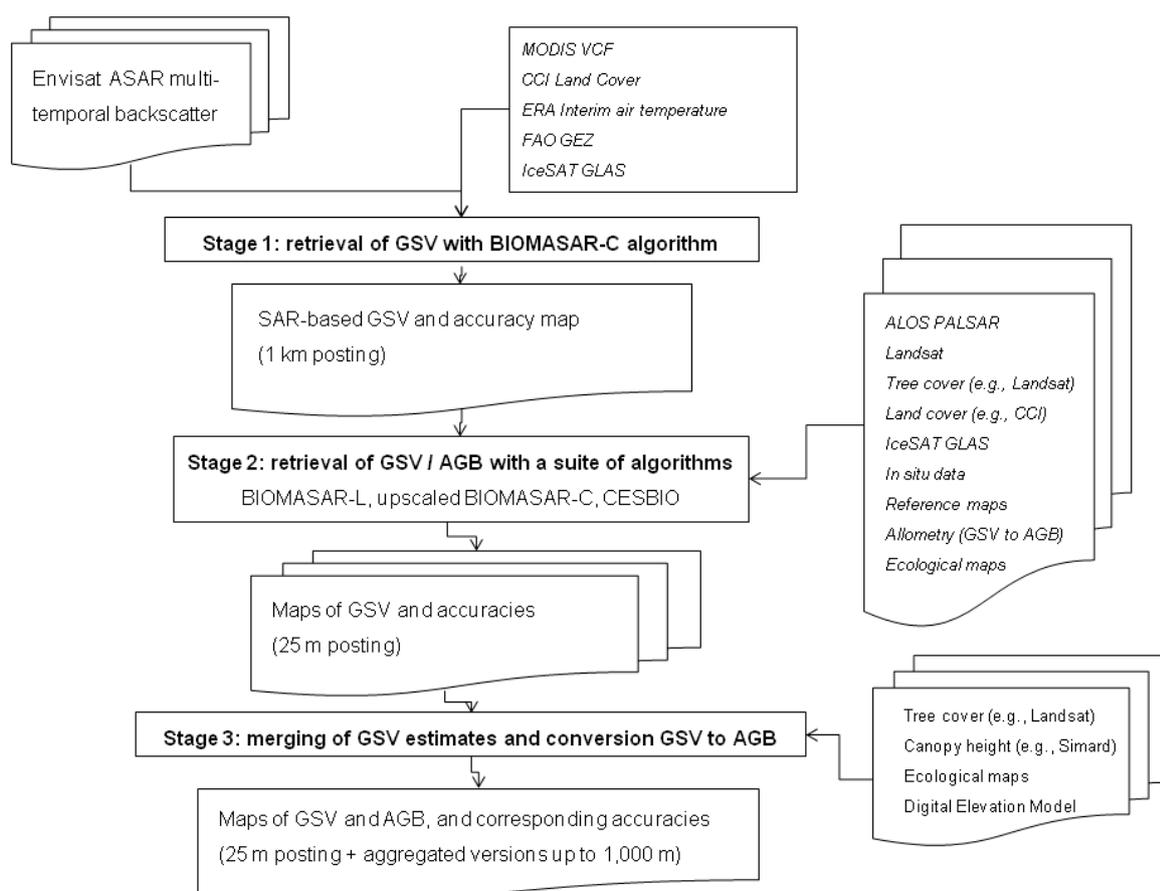


Figure 1-20. Flowchart of GlobBiomass global biomass retrieval algorithm (version of August 2017).

Another representation of the global biomass retrieval algorithm is provided by the more functional flowchart in Figure 1-21, showing the interdependencies of datasets and algorithms. Herewith we summarize the main aspects of the global biomass retrieval algorithm on the basis of this flowchart.

- The BIOMASAR-C GSV data product at 1,000 m spatial resolution serves as an indicator of biomass for the re-scaling approach.
- BIOMASAR-L and BIOMASAR-C+ produce estimates of GSV at 25 m spatial resolution using two different approaches on nearly the same input EO datasets.
- The CESBIO algorithm is included to provide an additional set of biomass estimates in tropical dry forest because it implements a more rigorous and advanced model training and inversion approach (tested so far only in tropical forest) compared to BIOMASAR-L and BIOMASAR-C+.

- A set of auxiliary datasets is necessary to calibrate the models taking into account that, except for the CESBIO approach, the retrieval algorithms do not rely on in situ measurements for training.
- The models implemented in the BIOMASAR-C and BIOMASAR-L algorithms retrieve GSV rather than AGB because SAR senses forest structure and GSV is the prime forest variable related to structure. In this way, it was ensured that the scattering physics are represented in the model functional dependencies and we have better control of the model parameter values.
- The GSV estimates of BIOMASAR-L and BIOMASAR-C+ are merged to allow for reduction of systematic errors in one or the other dataset.
- A conversion from GSV to AGB is implemented at the end of the retrieval as a separate step. The characterization of the AGB errors can therefore rely on separate estimates related to the retrieval algorithm and the conversion factors.
- Accuracy is characterized at each step shown in Figure 1-21. An estimate of the AGB accuracy is attached to each pixel at 25 m.
- Spatial aggregation (averaging) is applied to reduce pixel-wise retrieval errors and increase the accuracy so as to meet the requirement of a global AGB product with at most 30% error with a spatial resolution better than 500 m.

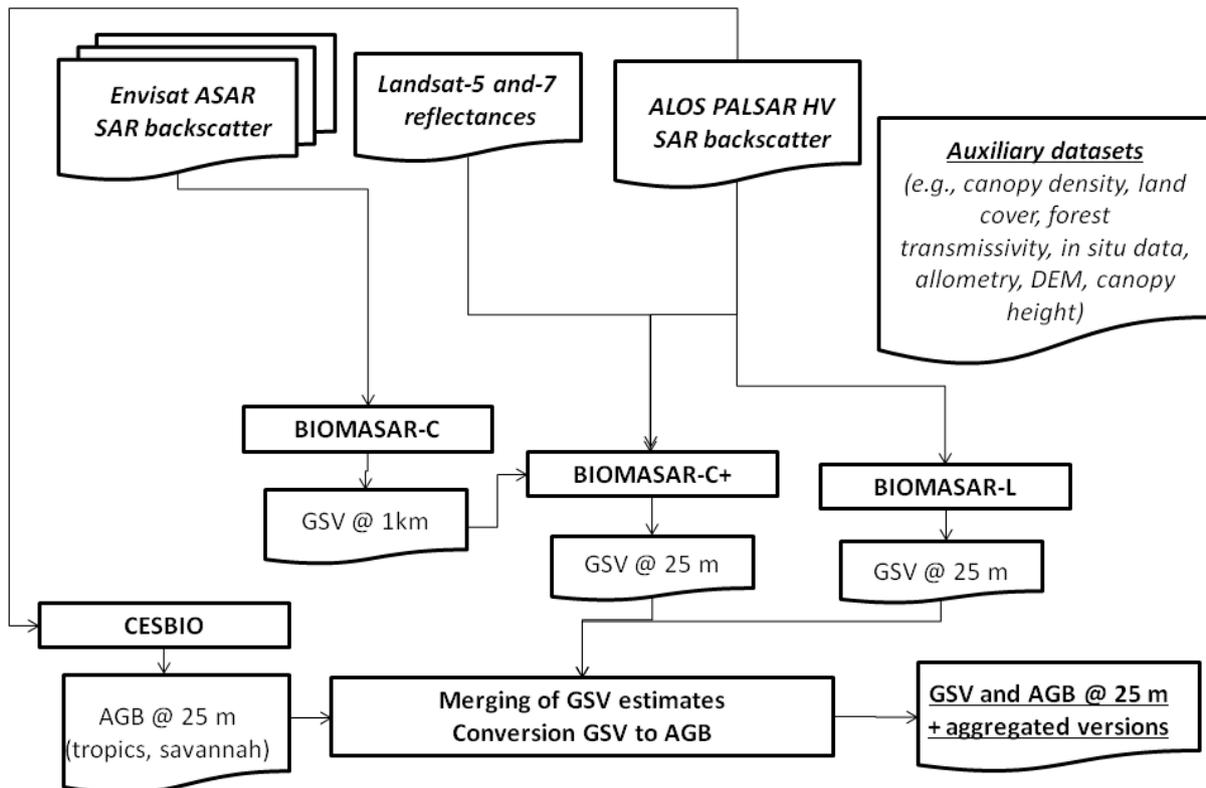


Figure 1-21. Functional dependencies of datasets and approaches forming the GlobBiomass global biomass retrieval algorithm.

1.4.2.1 BIOMASAR-C

At C-band, spatial and temporal variability of the backscatter make empirical modelling of GSV derived using *in situ* measurements almost useless if the aim is to produce large-scale estimates

	GlobBiomass		Page 34
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

based on a small set of reference measurements, which is very often the reality. Hence robust retrieval of GSV from backscatter should be based on a physically-based model that expresses the backscatter in terms of the main scattering mechanisms in as general a manner as possible.

The Water-Cloud Model with gaps (Attema & Ulaby, 1978; Askne et al., 1997) given as Equation (1-2) expresses the total forest backscatter as the sum of direct scattering from the ground through gaps in the canopy, ground scattering attenuated by the canopy and direct scattering from the vegetation:

$$\sigma_{for}^o = (1-\eta)\sigma_{gr}^o + \sigma_{gr}^o T_{tree} + \sigma_{veg}^o (1-T_{tree}) \quad (1-2)$$

Here θ is the area-fill or tree cover factor, representing the fraction of the area covered by vegetation, σ_{gr}^o and σ_{veg}^o , are the backscattering coefficients of the ground and vegetation layer, respectively, and T_{tree} is the two-way tree transmissivity, which can be expressed as $e^{-2\alpha h}$, where α is the two-way attenuation per meter through the tree canopy and h is the depth of the attenuating layer.

In practice, Equation (1-2) is not useful since the area-fill factor is not a parameter of interest to foresters. For retrieval purposes, it is more convenient to describe the backscatter as a function of GSV, V , as in Pulliainen et al. (1994):

$$\sigma_{for}^o = \sigma_{gr}^o e^{-\beta V} + \sigma_{veg}^o (1 - e^{-\beta V}) \quad (1-3)$$

where β is an empirically defined coefficient expressed in ha/m^3 . The link between β , η and α is given by (Santoro et al., 2002):

$$\eta = \frac{1 - e^{-\beta V}}{1 - e^{-\alpha h}} \quad (1-4)$$

Equation (1-3) contains three unknowns that need to be estimated: β , σ_{veg}^o and σ_{gr}^o . Although these can be estimated by least squares regression using a dataset of forest GSV measurements (Pulliainen et al., 1994; Fransson & Israelsson, 1999; Santoro et al., 2002), this approach is unfeasible for large areas because it requires a dense network of training sites. This can be overcome by a model training approach that does not rely on *in situ* measurements. The procedure outlined below is applied at pixel level.

To estimate σ_{gr}^o and σ_{veg}^o , we use the MODIS VCF percentage tree cover to identify pixels with negligible and dense canopy cover; these are referred to as “ground” and “dense forest” pixels respectively. The estimate of σ_{veg}^o is obtained after correcting the values of the backscatter for dense forest pixels, σ_{df}^o , for a residual contribution from the ground. The compensation requires knowledge of two additional parameters: the forest transmissivity and a GSV representative of dense forest. The procedure to train the backscatter model and retrieve GSV is outlined in (Figure 1-22). The estimation of the two parameters σ_{gr}^o and σ_{df}^o as well as the procedure to derive estimates of the forest transmissivity and the GSV of dense forests are described below. Once the model is trained, it is inverted to estimate GSV from corresponding observations of the SAR backscatter.

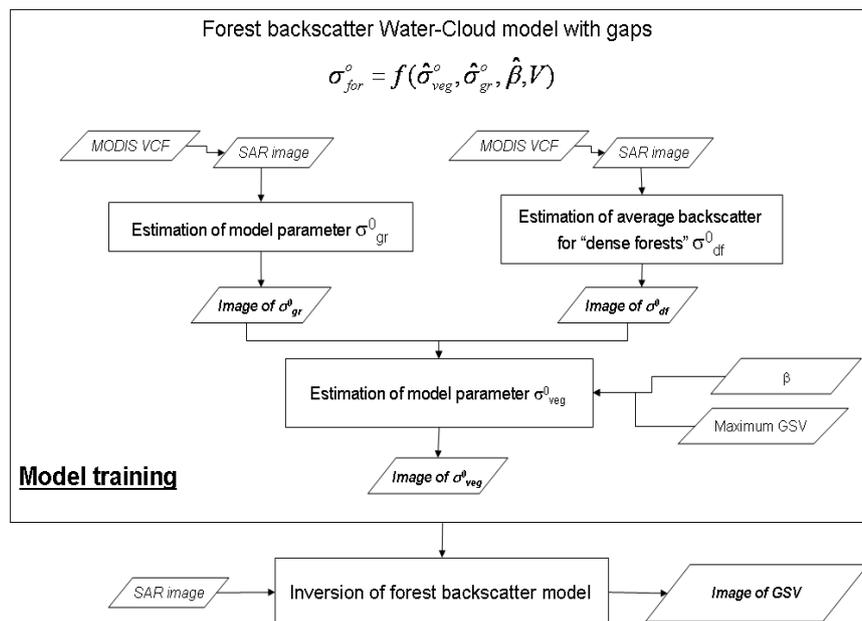


Figure 1-22. Model training and retrieval approach implemented in BIOMASAR-C (Santoro et al., 2011).

1.4.2.1.1 Estimation of σ_{gr}^0

σ_{gr}^0 is estimated as the median backscatter of pixels labelled as ground within a window centred on the pixel of interest (Santoro et al., 2011). To prevent distortion of the histogram of the ground pixels, land-cover classes not related to vegetation areas (e.g., water bodies, urban areas, exposed rocks, permanent snow or ice), are masked out using a land cover product (e.g., the CCI land cover product). As illustrated in Figure 1-23, starting from a minimum VCF threshold and radius of the estimation window, the percentage of ground pixels is computed, and the window is enlarged until at least 2% of the pixels are labelled as “ground”. The window size is limited by an upper threshold varying between 50 and 150 pixels which depends on the proportion of unvegetated areas within the study region. If the percentage of ground pixels is still below 2% at this threshold, the VCF threshold (which is also limited by a threshold) is increased and the window size is reset to the minimum radius. The 2% level is commonly reached with a VCF threshold of 30-35%. Higher thresholds are avoided to prevent the set of backscatter observations representative of unvegetated conditions being affected by observation of pixels with partial vegetation cover. This would lead to distortion of the histogram of the “ground” pixels. Pixels are labelled as “ground” when the percentage tree cover in the MODIS VCF product is below a spatially adaptive threshold (15% up to 30%).

If the 2% level cannot be reached for the highest tree cover, the value of the estimate is set initially to not-a-number (NaNs). NaNs are then replaced by extrapolating from surrounding areas ($10^\circ \times 10^\circ$ region) with valid estimates (<https://ch.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans>). This solution allows obtaining an estimate of σ_{gr}^0 in regions with unbroken dense forest cover (e.g., Amazon, Congo Basin). While recovering values everywhere, the extrapolation misses local variability of the backscatter from the forest floor. This however is assumed to have a negligible effect on the retrieved GSV because of the small proportion of the backscatter contribution from the ground in very dense forest, i.e., high biomass forest.

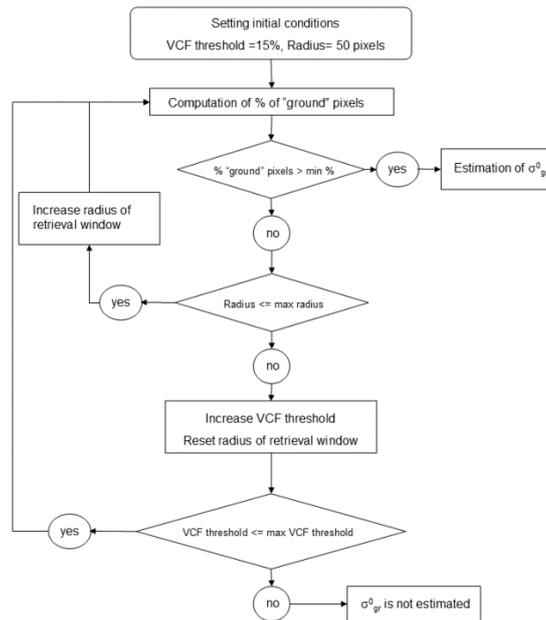


Figure 1-23. Flowchart illustrating the procedure to estimate σ_{gr}^0 in BIOMASAR-C (Santoro et al., 2011).

Figure 1-24 shows the result of extrapolation of values for a $2^\circ \times 2^\circ$ tile in the Amazon basin. The tile is covered by an ASAR image except for its western edge. The panel at the top shows that estimates of σ_{gr}^0 were not obtained for the southern part of the tile. By using the values estimated with the procedure described in Figure 1-23 from this tile as well as the neighbouring tiles in a $10^\circ \times 10^\circ$ area, the gap could be filled (bottom panel of Figure 1-24).

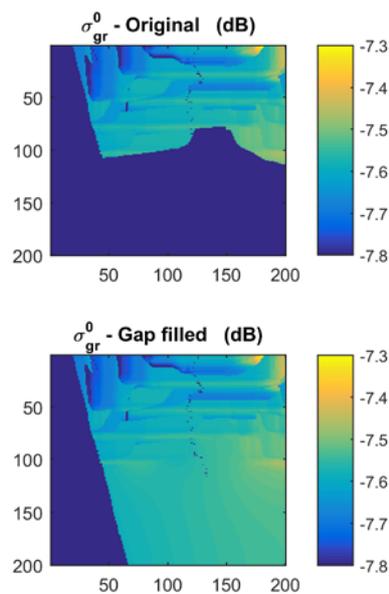


Figure 1-24. Example of gap filling of σ_{gr}^0 . Top panel: estimates of σ_{gr}^0 using the procedure described in Figure 1-23. Bottom panel: result of extrapolation of values using existing σ_{gr}^0 estimates.

	GlobBiomass		Page 37
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.4.2.1.2 Estimation of σ_{df}^0

σ_{df}^0 is defined as the average SAR backscatter for dense forests. Similarly to the estimation of σ_{gr}^0 , an estimation window is defined and pixels within it with a VCF value above a pre-defined threshold are selected to form the “dense forest” class. The average of the backscatter values for the pixels forming the “dense forest” class is associated with the parameter σ_{df}^0 . Window size and VCF threshold are predefined. A dense forest is defined as having a canopy cover above 75% of the maximum value within a $1^\circ \times 1^\circ$ large window around the pixel of interest. All pixels satisfying this criterion within a 200×200 pixel window (i.e., $2^\circ \times 2^\circ$) around the same pixel of interest are then selected to be included in the “dense forest” class (Santoro et al., 2011). Given that the C-band SAR backscatter of dense forests is less affected by spatial variability than unvegetated terrain, the estimation rules for the backscatter of dense forests can be kept simple. As in the case of σ_{gr}^0 , NaN values in areas of predominantly bare soil are replaced by extrapolating from valid values within a $10^\circ \times 10^\circ$ large window.

1.4.2.1.3 Estimation of β

β is related to the vegetation dielectric and forest structure, including seasonal effects such as frozen/unfrozen and leaf-on/leaf-off conditions (Pulliainen et al., 1996; Santoro et al., 2002), so must adapt to local environmental and forest conditions. Using Equation (1-4), canopy cover can be simulated for a given GSV by setting values for the coefficient of the tree transmissivity, α , and the forest transmissivity, β . Taking into account that the sensitivity of the model in Equation (1-4) to α is weak for $\alpha > 0.5$ dB (i.e., at C-band), information on plausible values for β in a given forest environment can be obtained by relating canopy cover observations to GSV observations. This approach was attempted in the development phase of the BIOMASAR algorithm. Analysis of moderate resolution SAR data from boreal, temperate sub-tropical and, in part, tropical forests was used to establish the range of plausible values for β in the northern hemisphere. Except for tropical wet and moist forest, a constant value of 0.006 ha/m^3 appeared to be plausible (Santoro et al., 2011; Santoro et al., 2015b). To assess the impact of assuming a constant value of β , the sensitivity of the GSV retrieval to β was tested by training the model in Equation (1-4) with different values of β and inverting it. The resulting GSV estimates differed by less than 10%, regardless of the study area. In humid tropical forests higher β values appeared to be more plausible ($> 0.01 \text{ ha/m}^3$) (Santoro et al., 2015a). Tropical forests, however, have not been the topic of GSV retrieval in past studies.

This approach has now been extended to all forests globally and stratification in terms of ecological zone has been applied in order to provide a more realistic representation of the spatial distribution of β . Simulations of area-fill from Equation (1-4) were generated using β values between 0.004 and 0.014 ha/m^3 and α values between 0.2 and 2 dB/m . To assess which value of β would be most appropriate to describe the relationship between GSV and area-fill in a given eco-region, the simulations were compared to estimates of canopy cover in the MODIS VCF dataset and spatially explicit estimates of GSV or AGB mostly derived from remote sensing. When AGB was available, it was converted to GSV using the BCEF published by the IPCC (IPCC, 2006). This can be considered sub-optimal because of intrinsic errors and uncertainties of the AGB/GSV estimates and the conversion factors. Nonetheless, it was found to be the only practicable way to assess the spatial variability of the coefficient of the forest transmissivity given the lack of large-scale datasets of *in situ* observations of GSV. Both the GSV and the MODIS VCF values were aggregated to 1 km pixel size and resampled to 0.01° . Aggregation helped to decrease uncertainties in the canopy cover and GSV estimates. Aggregation to coarse resolution, however, also implied that the estimates of β would be representative for usage in the retrieval at coarse scale. The height in Equation (1-4) is represented

	GlobBiomass		Page 38
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

by 1-km averages of ICESAT vegetation height. Hence, the GSV and VCF datasets include only observations from pixels corresponding to the area covered by the 1-km average of ICESAT vegetation heights.

Figure 1-25 shows an example of model fits between GSV and VCF using Equation (1-4) after stratification in terms of ecological zone and tree type (broadleaf or conifer) for Sweden. The modelled canopy cover is represented by curves of different colours for different values of β . Curves with the same colour have the same β value but different α values; the steepest curve corresponds to the lowest value of α . Figure 1-25 indicates that percentage tree cover depends on ecological zone and, to a lesser extent, on tree type. Notwithstanding the values of β for the intermediate range of VCF values, which are known to be positively biased, the trends for forests in Sweden indicate values for β between 0.006 ha/m^3 and 0.008 ha/m^3 . Under unfrozen conditions in boreal forest (when $\alpha > 0.5 \text{ dB/m}$, i.e. the flattest curve for each given colour), the model fit with $\beta = 0.008$ was closer to the observations than when $\beta = 0.006$, while $\beta = 0.006 \text{ ha/m}^3$ was plausible for temperate forests. For frozen conditions, $\alpha < 0.5 \text{ dB/m}$ is more realistic (Santoro et al., 2002), so the observations need to be compared with the uppermost curves for a given colour. Then $\beta \sim 0.006 \text{ ha/m}^3$ appears to give the best fit. These values are confirmed when computing the Root Mean Square Difference (RMSD) between the modelled and the VCF values of canopy cover.

An example for sub-tropical and tropical forest types is shown in Figure 1-26. Here, the dependency upon α is disregarded given that the sensitivity of the modelled canopy cover upon GSV is not impacted by values of α which are assumed to be greater than 1 dB/m . Tropical rainforest and tropical moist forest presented the largest β values (~ 0.014) in low GSV forest. For higher GSV (above $100 \text{ m}^3/\text{ha}$), the decrease of VCF cannot be explained by the model and it is unclear what causes this trend. A comparison with the Landsat VCF did not reveal the same trend but rather a slow increase of percent tree cover with GSV although with smaller β values. Not having available a third source of information did not allow us to conclude which is the correct relationship between GSV and percent tree cover. It should be noted that this kind of result was consistent at different sites including the ecological zones of tropical rainforest and moist forest. Taking into account that there can be multiple reasons for the ambiguous relationship identified between percent tree cover and GSV (errors and uncertainty of input dataset, BCEF, segmentation into ecological zones), it was decided to set β to 0.012 in tropical rainforest and moist forest, this representing a compromise between the results obtained from our analysis. For drier forest types, the increase of VCF with GSV is steady and is well characterized by the model relating percent tree cover and GSV with β values in the range $0.004 - 0.006$.

As a result of the comparison of MODIS and Landsat VCF percent tree cover against GSV estimate, most plausible β values for each ecozone of the FAO GEZ dataset have been identified. Accordingly, a look-up table relating β to the corresponding GEZ class has been established. Plausible means that the estimate of β is not strictly the one for which the RMSD between percent tree cover and GSV is the minimum; it is rather the value for which the model in Equation (1-4) describes the relationship between percent tree cover and GSV in a reasonable manner for all sites given a GEZ class. In addition, some adjustments were undertaken to avoid discrepancies with β estimates obtained with the alternative approach proposed for the BIOMASAR-L algorithm.

The look-up table in Table 1-3 shows the most plausible values for β at C-band (i.e., $\alpha > 1 \text{ dB}$) for each ecological zone in the FAO GEZ dataset.

Table 1-3. Estimates of β by ecological zone.

Ecological zone according to FAO GEZ dataset	β [ha/m ³]
Tropical rain forest	0.012
Tropical moist deciduous forest	0.012
Tropical dry forest	0.008 (North America) 0.006 (Eurasia)
Tropical shrubland	0.004
Tropical desert	0.003
Tropical mountain systems	0.010
Subtropical humid forest	0.008
Subtropical dry forest	0.004
Subtropical steppe	0.004
Subtropical desert	0.004
Subtropical mountain systems	0.005
Temperate oceanic forest	0.007
Temperate continental forest	0.0085 (North America), 0.007 (Eurasia)
Temperate steppe	0.006
Temperate desert	0.005
Temperate mountain systems	0.006
Boreal coniferous forest	0.006
Boreal tundra woodland	0.005
Boreal mountain systems	0.006
Polar	0.004

β varied between 0.012 ha/m³ in tropical rain forest and 0.004 ha/m³ in arid environments. In general, higher β values were estimated in forest with a more closed canopy, which agrees with theory. Nonetheless, this procedure of estimating β is strongly influenced by uncertainties in the spatially explicit datasets of GSV and AGB, the conversion factors between AGB and GSV, and the VCF dataset. The procedure is furthermore influenced by the zoning and resolution of the GEZ dataset, and by classification errors in the land cover dataset. The values reported above should be treated as indicative, and further studies are desirable; these would require *in situ* observations of gap fraction, GSV and attenuation.

The VCF product was derived for leaf-on conditions. The canopy cover in leaf-off conditions, however, can differ substantially from leaf-on conditions. For forests in Southwest U.S., Twedt et al. (2015) reported how much canopy cover differs between leaf-on and leaf-off conditions. To the best

of our knowledge, this is the only study reporting measurements of canopy cover under leaf-on and leaf-off conditions. The sparse experimental evidence concerning the link between leaf-on and leaf-off conditions did not allow us to infer a “leaf-off VCF product” from which estimates of β could be derived. As a consequence, we do not separate between leaf-on and leaf-off conditions when estimating the model parameters. Similarly, we do not account for differences of β depending on environmental conditions, since the coefficient of the tree transmissivity α should not be less than 0.5 dB at C-band and the modelled VCF as a function of GSV did not differ for $\alpha \geq 0.5$ dB (Figure 1-25).

The spatially explicit dataset of β values associates a single value of β with the corresponding ecological zone in the FAO GEZ dataset. To avoid discontinuities (steps) in β due to the categorical classification of the ecological zones, an average filter is applied to the original estimates of β using a 5x5 window.

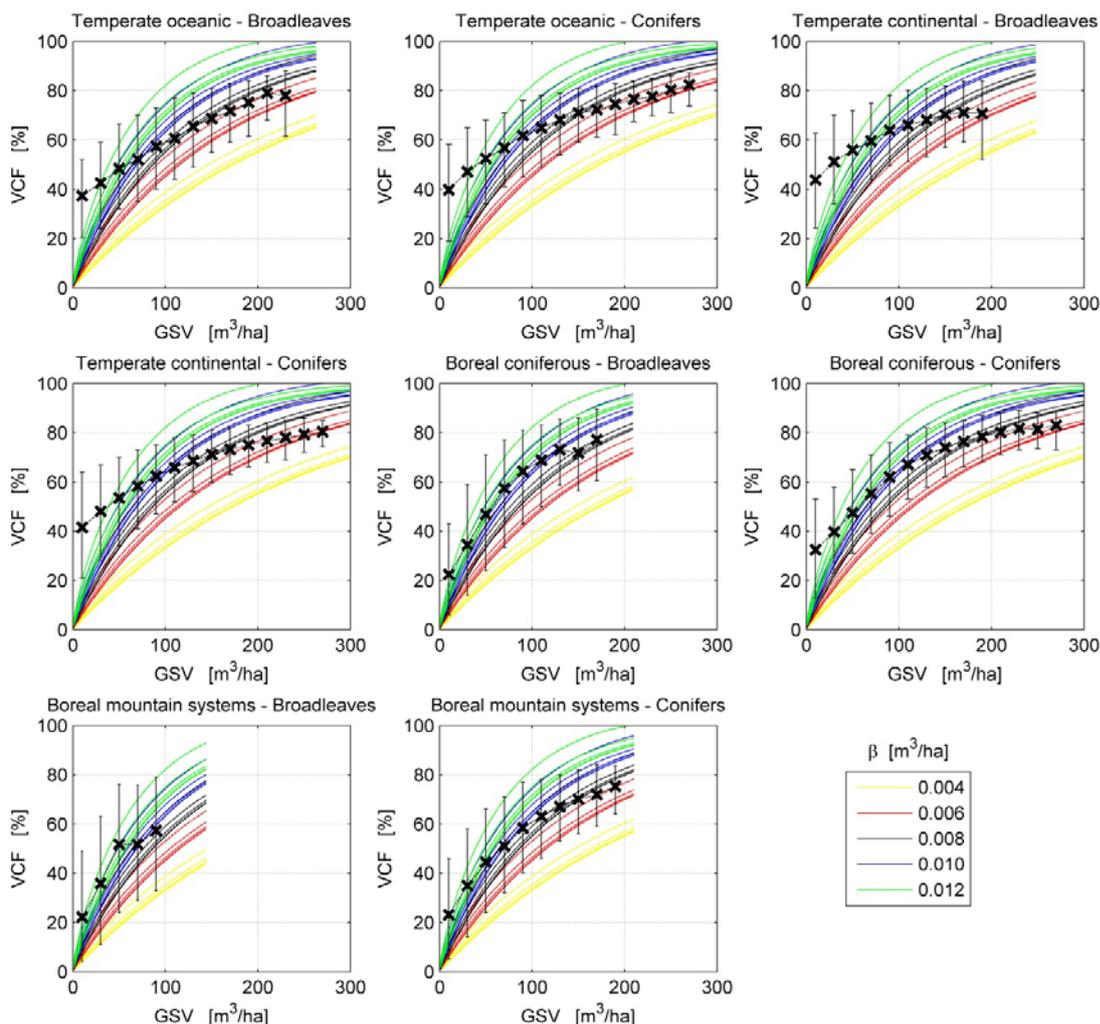


Figure 1-25. Curves of percent tree cover modelled with Equation (1-4) as a function of GSV in Sweden for different values of α and β . Crosses and vertical bars represent the median and the 10-90 percentile interval of MODIS VCF percent tree cover in 20 m³/ha intervals of GSV (aggregated kNN product for Sweden). Each panel corresponds to an ecological zone (FAO, 2001) and tree genus (conifers or broadleaves).

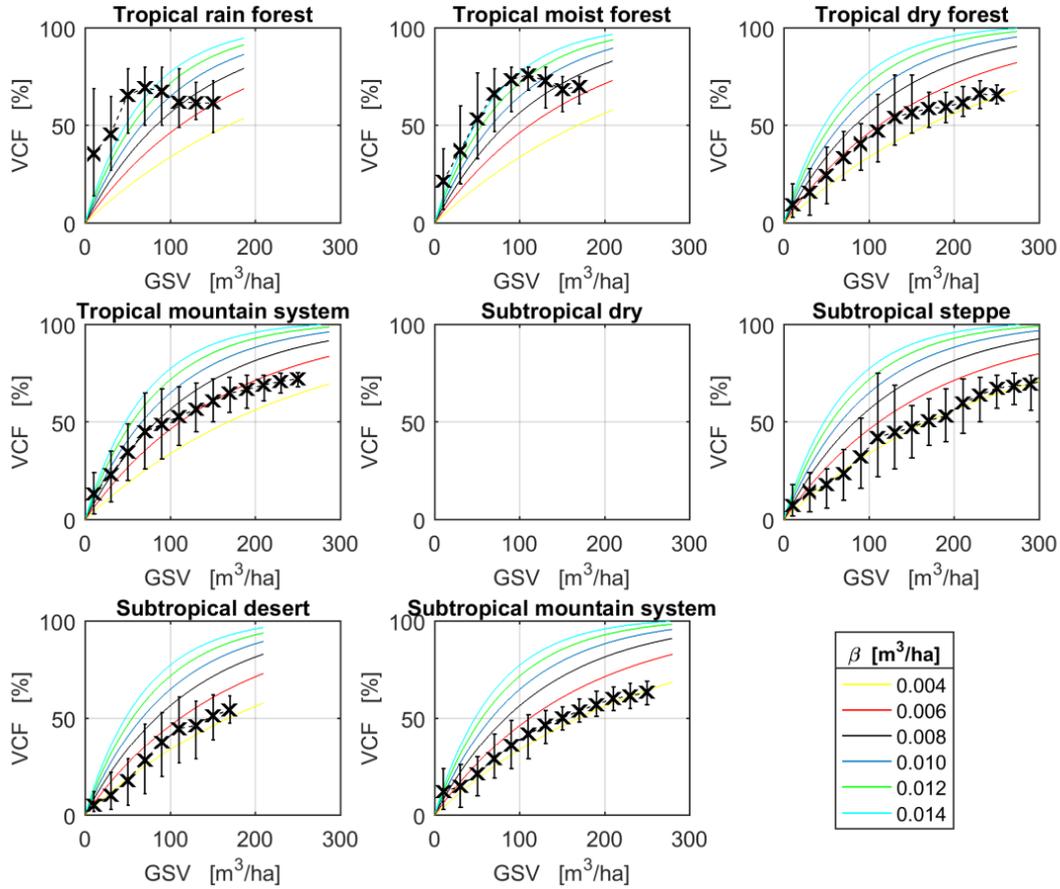


Figure 1-26. Curves of percent tree cover modelled with Equation (1-4) as a function of GSV in Mexico for different values of β . Crosses and vertical bars represent the median and the 10-90 percentile interval of MODIS VCF percent tree cover in $20 \text{ m}^3/\text{ha}$ intervals of GSV. The GSV estimates were obtained from the AGB dataset by Cartus et al., 2014, aggregated to 1 km and converted to GSV using IPCC BCF. Each panel corresponds to an ecological zone (FAO, 2001).

1.4.2.1.4 Estimation of σ_{veg}^0

To estimate σ_{veg}^0 , the measured backscatter from dense forests must be compensated for backscatter from the ground seen through gaps in the canopy. Since σ_{gr}^0 is known at this stage, Equation (1-3) can be inverted to obtain σ_{veg}^0 from the backscatter of the pixels forming a dense forest” class:

$$\sigma_{veg}^0 = \frac{\sigma_{df}^0 - \sigma_{gr}^0 e^{-\beta V_{df}}}{1 - e^{-\beta V_{df}}} \quad (1-5)$$

where σ_{df}^0 represents a measurement of the backscatter of the dense forest class and V_{df} is the GSV value representative of dense forest.

1.4.2.1.5 Inverting the forest backscatter model

Once the model parameters have been estimated, the model in Equation (1-3) can be inverted to derive an estimate of GSV,, from a measurement of the backscatter, σ_{meas}^0 :

	GlobBiomass		Page 42
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

$$\hat{V} = -\frac{1}{\beta} \ln \left(\frac{\sigma_{meas}^o - \sigma_{veg}^o}{\sigma_{gr}^o - \sigma_{veg}^o} \right) \quad (1-6)$$

At C-band, it is likely that the measured backscatter is not within the range of modelled backscatter values, especially in areas with high GSV where the backscatter typically saturates. This requires the inversion to be constrained to certain ranges of backscatter values and GSV (Santoro et al., 2011). Assuming that backscatter increases with increasing GSV, the retrieved GSV is set to 0 m³/ha or the maximum retrievable GSV when the measurement falls within a buffer zone 0.6 dB either below or above the modelled backscatter. No GSV is retrieved for backscatter outside this range. The maximum retrievable GSV is defined as $V_{df} + 50 \text{ m}^3/\text{ha}$ (Santoro et al., 2011).

1.4.2.1.6 Multi-temporal combination of GSV estimates

Given N individual estimates of GSV from Equation (1-3), a weighted linear combination of the estimates is used to obtain the final estimate of GSV, V_{mt} (Equation (1-7)). This reduces the retrieval error with respect to each of the individual GSV estimates (Kurvonen et al., 1999; Santoro et al., 2002; Santoro et al., 2011).

$$V_{mt} = \frac{\sum_{i=1}^N w_i v_i}{\sum_{i=1}^N w_i} \quad (1-7)$$

The weights, w_i , in Equation (1-7) are defined as the vegetation-to-ground backscatter difference in dB, $\sigma_{veg}^o - \sigma_{gr}^o$, normalized by the maximum backscatter difference:

$$w_i = \frac{\sigma_{veg,i}^o - \sigma_{gr,i}^o}{\max(\sigma_{veg,i}^o - \sigma_{gr,i}^o)} \quad (1-8)$$

In boreal forests, Santoro et al. (2011) found that the RMS difference between retrieved and *in situ* GSV decreased with increasing backscatter difference. GSV estimates obtained from SAR images with absolute backscatter difference less than 0.5 dB were discarded because they caused the performance of the GSV retrieval to deteriorate.

1.4.2.2 BIOMASAR-L

Many studies have documented the sensitivity, as well as the limitations, of L-band backscatter to forest biophysical parameters such as GSV or AGB across a wide range of forest ecosystems. The existing studies generally report a higher sensitivity of L-band to GSV or AGB than shorter wavelength radars because of its increased ability to penetrate forest canopies (Dobson et al., 1992; Harrel et al., 1997; Fransson & Israelsson, 1999; Saatchi & Moghaddam, 2000; Hyypä et al., 2000; Hoekman & Quinones, 2000; Castel et al., 2001; Sandberg et al., 2011). The highest sensitivity was usually reported for the L-band cross-polarized intensity (Le Toan et al., 1992; Dobson et al., 1992; Ranson & Sun, 1994, 2000; Harrel et al., 1997; Hoekman & Quinones, 2000; Castel et al., 2001; Balzter et al., 2002; Kellndorfer et al., 2003).

While an increase of L-band backscatter with increasing GSV or AGB was consistently observed, as well as with increasing canopy density and height, the backscatter contribution from the forest floor declines and the volume scattering contribution from the canopy increases, and forest structural

	GlobBiomass		Page 43
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

differences have been shown to affect the functional relationship between backscatter and GSV or AGB. Saturation, i.e., the GSV or AGB level beyond which the measurements no longer show any sensitivity to biomass, depends on the forest structure and has been reported to occur in biomass ranges of 80 to 150 t/ha for savannah (Lucas et al., 2006, 2010; Mitchard et al., 2009), 40 to 150 t/ha for boreal and temperate forests (LeToan et al., 1992; Dobson et al., 1992; Ranson & Sun, 1994; Imhoff et al., 1995; Ranson et al., 1995; Sandberg et al., 2011) and 40 to 150 t/ha in the tropics (Imhoff, 1995; Luckman et al., 1997; Saatchi et al., 2011). In addition, L-band backscatter from forested terrain was observed to depend on the imaging conditions, in particular the wet/dry or frozen/unfrozen state of the soils and vegetation or, if present, the properties of the snow cover (Rignot et al., 1994a, 1994b; Kwok et al., 1994; Way et al., 1994; Harrel et al., 1995; Ranson et al., 1995, 2000; Pierce et al., 1998; Pulliainen et al., 1999; Saatchi & Moghaddam, 2000; Salas et al., 2002; Askne et al., 2003; Santoro et al., 2006; Lucas et al., 2010; Kasischke et al., 2011; Sandberg et al., 2011). Although overall more consistent over time than C-band observations over forest (Askne et al., 2003; Santoro et al., 2006, 2009), the imaging conditions were found to significantly affect: 1) the overall backscatter strength from the forest floor and the canopy; and 2) the sensitivity to forest biophysical parameters (Rignot et al., 1994b; Harrel et al., 1995, 1997; Pulliainen et al., 1999; Ranson & Sun, 2000; Rauste, 2005; Santoro et al., 2006).

Algorithms aiming at exploiting the sensitivity of L-band backscatter to GSV or AGB for large-scale retrieval therefore need to be calibrated adaptively to local forest structure as well as the prevalent imaging conditions. The limited availability of *in situ* information (e.g., inventory plots) prevents adaptive calibration of retrieval algorithms using conventional approaches. In many areas, in particular the tropics, the number of available plots is very limited so that models may only be calibrated using reference information collected over large areas. Conventional approaches generally necessitate working with mosaics of L-band backscatter imagery that are compiled in a way that potential differences in the imaging conditions between the orbits/acquisition dates used for generating the backscatter mosaic are minimized. When multi-temporal observations are available, mosaicing entails careful selection of imagery to minimize between-orbit radiometric differences. Alternatively, empirical inter-orbit normalization techniques may be applied (e.g., De Grandi et al., 2011; Shimada et al., 2010).

Approaches that potentially allow more spatial adaptivity are:

- 1) two-stage upscaling approaches exploiting the sensitivity of air- or spaceborne Lidar measurements to GSV or AGB by using GSV/AGB estimates derived for Lidar footprints as a surrogate reference for calibrating models relating spaceborne imagery to GSV/AGB (e.g. Baccini et al., 2012; Saatchi et al., 2011).
- 2) the BIOMASAR approach presented in the previous sections for retrieving GSV globally from hyper-temporal observations of ENVISAT ASAR ScansAR C-band observations.

Two-stage upscaling approaches exploiting the synergy of spaceborne Lidar (so far only ICESAT GLAS) and imagery (optical or radar) require a separate set of models for estimating GSV/AGB from Lidar metrics, such as the canopy height or various percentiles characterizing the vertical shape of the waveforms, and thus a set of *in situ* measurements co-located with the Lidar measurements.

The application of the BIOMASAR algorithm, instead, allows for spatially adaptive calibration of models relating backscatter observations to GSV while minimizing the need for reference information, as calibration is achieved through inter-comparison of radar and optical remote sensing products. The BIOMASAR approach is thus, unlike calibration using the available *in situ*

	GlobBiomass		Page 44
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

measurements, a candidate for mapping GSV globally from the available ALOS PALSAR mosaics of L-HH and L-HV backscatter. Although developed for C-band, the algorithm has already been tested successfully in boreal and temperate forest areas for ALOS PALSAR L-band data (Cartus et al., 2012; Santoro et al., 2014). Only limited experience on the performance of the approach with respect to its ability to describe/predict the relationship of L-band backscatter and GSV in other forest ecosystems, in particular the (sub-)tropics, has yet been gained.

In the following sections, we describe the modifications to the BIOMASAR-C algorithm that are needed to retrieve GSV globally from the available ALOS PALSAR L-band backscatter mosaics provided by JAXA. This modified version of the algorithm will be referred to as BIOMASAR-L. The modifications concern the model calibration and inversion approach as well as the general implementation of the algorithm, the latter being required because the PALSAR data were provided by JAXA in the form of 1°x1° tiles, each containing a mosaic of images acquired from two to four different orbits.

One of the major components of the BIOMASAR algorithm is the possibility to improve the retrieval performance by integrating multi-temporal stacks of SAR backscatter observations. In general, the retrieval is performed separately for each image in a multi-temporal stack and then the retrieval results for each image are combined in a weighted manner, giving more weight to those acquisitions with the best sensitivity to GSV. Major improvements in performance have been reported at C-band (Santoro et al., 2011), but improvements could also be achieved at L-band (Santoro et al., 2006, 2014; Cartus et al., 2012), albeit less pronounced than at C-band. For GlobBiomass, we exploit the annual L-band mosaics generated by JAXA. If we strictly limit the retrieval to 2010, improvements through multi-temporal/multi-seasonal combination are not possible, since the retrieval performance then completely depends on the single backscatter observation in the 2010 mosaic. However, the advantage of combining observations from two consecutive years, at least regionally where the imagery in the 2010 mosaic was acquired under adverse (i.e., wet or frozen) conditions, is explored.

To model the relationship between L-band backscatter observations and GSV, we use the same Water Cloud type of model as for C-band (Equation (1-3)). One of the underlying assumptions of this model is that higher order scattering can be neglected. Although under typical conditions (rough forest floor, substantial attenuation in the canopy) stem-ground interactions can be neglected at L-band (Dobson et al., 1992; Pulliainen et al., 1999), a significant contribution from higher order scattering may arise from the canopy. Models indicate that higher order scattering effects are negligible for co-polarization, but not necessarily for cross-polarization. In Wang et al. (1998), higher order scattering increased the modelled L-HV backscatter from pine forest by 1.5 to 2 dB (at ~ 35° incidence angle). Karam et al. (1992) noted that higher order scattering from walnut orchards had a significant effect at HV polarization only for X-band, not L-band, regardless of the incidence angle. For a forested site in France, Picard et al. (2004) observed underestimation of L-HV backscatter when only first order scattering in a model was considered. However, when multiple scattering effects were included, the backscatter was overestimated compared to SIR-C L-HV measurements at 26° and 54° incidence angle. Picard et al. concluded that the improvements in the modelling by including multiple scattering were minor compared to the overall uncertainty in the modelling and that, in contrast to C-HV, the modelled relationship between L-HV backscatter and GSV containing only first order scattering depicted well the observed relationship of SIR-C L-HV intensity and GSV.

	GlobBiomass		Page 45
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.4.2.2.1 Estimating σ_{gr}^0 and σ_{df}^0

One of the key elements of the BIOMASAR algorithm is the adaptive estimation of the backscattered power arising from the forest floor and closed forest canopies, i.e., the model parameters σ_{gr}^0 and σ_{df}^0 in Equation (1-3), by analysing histograms of L-band backscatter in areas of low and high canopy density according to globally available optical remote sensing products such as canopy density maps produced from MODIS or Landsat data.

While the concept of the BIOMASAR-L approach is essentially the same as for BIOMASAR-C, some modifications were necessary due to the higher resolution of the data and the fact that it was being delivered as a global mosaic of ~ 27000 $1^\circ \times 1^\circ$ tiles, each containing imagery acquired from two to four different orbits. Figure 1-27 shows a PALSAR image mosaic with ~ 25 m pixel posting in a $1^\circ \times 1^\circ$ degree tile and a map of the acquisition dates of the images used to create the mosaic (i.e., 3 different acquisition dates in the example in Figure 1-27). The acquisition date map allows for model parameters to be estimated for each orbit in a tile. The weighted feathering approach that was applied by JAXA locally to smooth out some of the backscatter differences observed between adjacent orbital tracks may however lead to local artefacts because along the edges of adjacent orbits the backscatter statistics will be distorted. Additional datasets required for model training are the local incidence angle and layover/shadow maps provided by JAXA, the Landsat canopy density and water body maps, and the CCI land cover map.

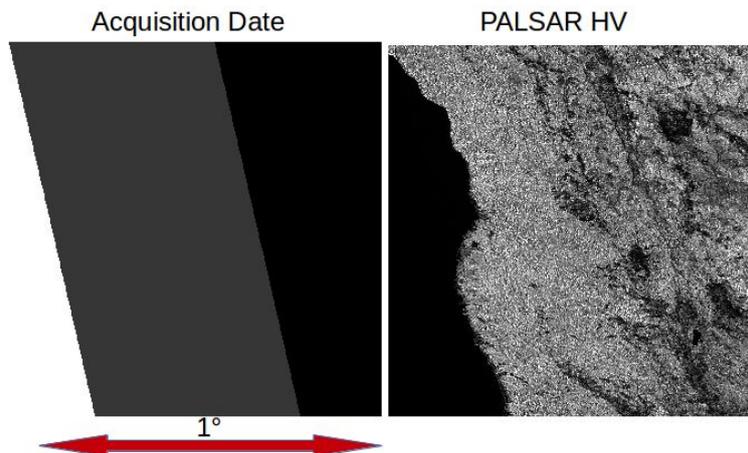


Figure 1-27. $1^\circ \times 1^\circ$ tile of ALOS PALSAR HV backscatter (right) and the acquisition dates of the images used to create the mosaic (left). The acquisition date map allows the acquisition date for each pixel in the tile to be identified. In this example, images from three different acquisition dates (illustrated in white, gray and black) have been used.

The workflow to estimate σ_{gr}^0 and σ_{df}^0 for each orbit in a $1^\circ \times 1^\circ$ PALSAR tile comprises the following three steps.

1.4.2.2.1.1 Data preparation

The $1^\circ \times 1^\circ$ PALSAR tile for which the model parameters are to be estimated, the associated ancillary files (acquisition date map, local incidence angle map, layover/shadow mask) and the corresponding subset of the Landsat canopy density map from 2010 are aggregated to the $\sim 300 \times 300$ m² pixel size ($0.00277^\circ \times 0.00277^\circ$) of the CCI land cover map. After aggregating the datasets, CCI land cover is used to mask out cropland, urban areas, impervious surfaces, and permanently and temporarily flooded land cover types. Layover/shadow areas were masked based on maps provided by JAXA. Although

	GlobBiomass		Page 46
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

topographic corrections had been applied by JAXA, the PALSAR mosaics still exhibited significant topographic effects, in particular for steep slopes facing the sensor. Areas imaged with a local incidence angle of less than 28°, as well as areas of layover and shadow, are therefore masked.

1.4.2.2.1.2 Identification of areas of low and high canopy density

Using the Landsat canopy density maps, areas of sparse and dense forest cover in each orbit and tile are identified. Tests indicated that a certain number of pixels are required to estimate σ_{gr}^0 and σ_{df}^0 from backscatter observations in areas of low and high canopy density, respectively. The canopy density threshold used to delineate areas of low canopy density is increased from 0% in steps of 1% until the required number of pixels with low canopy density (2000 pixels) is found. The maximum allowed canopy density threshold is set to 20%. The canopy density threshold used to delineate dense forest is reduced in steps of 1% from 100% until the required number of pixels (1000 pixels) with a canopy density above the selected threshold is found. The minimum threshold is set to 70%. If this requirement cannot be fulfilled, the area within which the model parameters are estimated for each orbit in the 1°×1° tile is increased in steps of 1° (i.e., image tiles within a buffer of 1° with backscatter observations from the same orbit are considered) as shown in Figure 1-28. This is repeated for each orbit crossing the tile considering the neighbouring tiles in a 3°×3° window or, in another iteration, a 5°×5° window. The largest window sizes for estimating the model parameters are required only in dense tropical forest regions, such as the Amazon Basin or Central Africa, because continuous forest cover may not allow estimation of σ_{gr}^0 at more local scale.

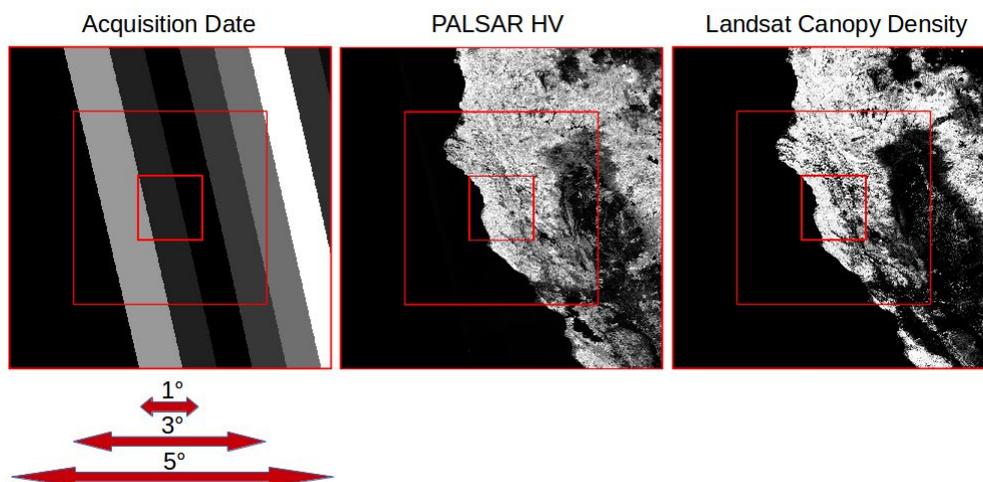


Figure 1-28. 5°×5° subsets of the global PALSAR HV backscatter mosaic (centre) and Landsat canopy density maps (right). The red boxes denote the variable areas around the 1° x 1° centre tile that, depending on the local canopy density distribution, are considered for estimating the model parameters for each orbit crossing the centre 1°×1° tile. The PALSAR orbits that were used to generate the mosaic in this area are illustrated on the left with different shades of gray.

1.4.2.2.1.3 Model parameter estimation

In BIOMASAR, σ_{gr}^0 and σ_{df}^0 are estimated from the histograms of backscatter observations in areas of low and high canopy density, respectively, with metrics denoting the central tendency of the backscatter distribution (Figure 1-29). For the estimation of σ_{df}^0 in BIOMASAR-L, we use the mode of the backscatter distribution in areas of high canopy density once at least 1000 pixels corresponding to the dense forest class have been identified in the imagery with the aid of the Landsat canopy density map and after applying a kernel smoothing algorithm to the histogram to avoid spurious effects on

	GlobBiomass		Page 47
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

the estimation of the mode. The parameter σ_{gr}^0 is, instead, estimated with the median of the backscatter distribution in areas of low canopy density.

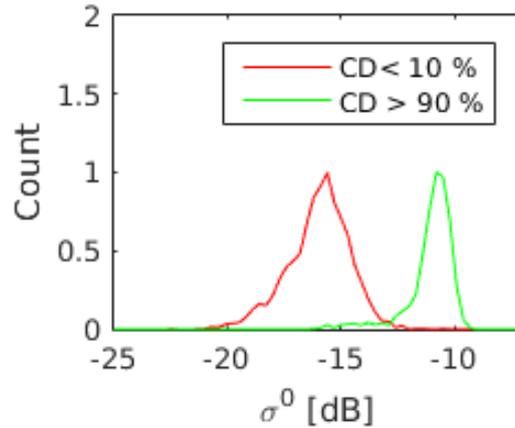


Figure 1-29. Histograms of L-HV backscatter in areas of low (red) and high (green) canopy density according to Landsat.

1.4.2.2.2 Estimating σ_{veg}^0

The model parameter σ_{veg}^0 represents the volume scattering power from an ideal opaque forest canopy. The backscatter observed over dense forest, σ_{df}^0 , which was estimated with the aid of an optical canopy density map, therefore needs to be compensated for residual ground contributions in the measured backscatter, i.e., for a non-zero forest transmissivity due to residual gaps in the canopy as well as the finite depth of the canopy layer, to obtain σ_{veg}^0 . σ_{veg}^0 can be estimated based on the model in Equation (1-2) describing backscatter as a function of the forest transmissivity of the densest forests in the area of interest, which itself is considered a function of residual canopy gap fraction, η_{df} , and canopy height, h_{df} , when inverting the model for σ_{veg}^0 :

$$\sigma_{veg}^0 = \frac{\sigma_{df}^0 \sigma_{gr}^0 T_{df}}{1 - T_{df}} \quad (1-9)$$

with

$$T_{df} = (1 - \eta_{df}) + \eta e^{-\alpha h_{df}} \quad (1-10)$$

Figure 1-30 illustrates the concept for estimation of σ_{veg}^0 from σ_{df}^0 . σ_{gr}^0 and σ_{df}^0 are estimated from the backscatter distribution in areas of low (red histogram) and high canopy density (green histogram) as explained above. The residual transmissivity, T_{df} , is then estimated from the ICESAT GLAS based estimates of canopy height and density for footprints covering the densest forests in the particular area. In the example shown in Figure 1-30, the compensation for the residual transmissivity leads to an estimate for σ_{veg}^0 which is ~ 1.5 dB higher than σ_{df}^0 .

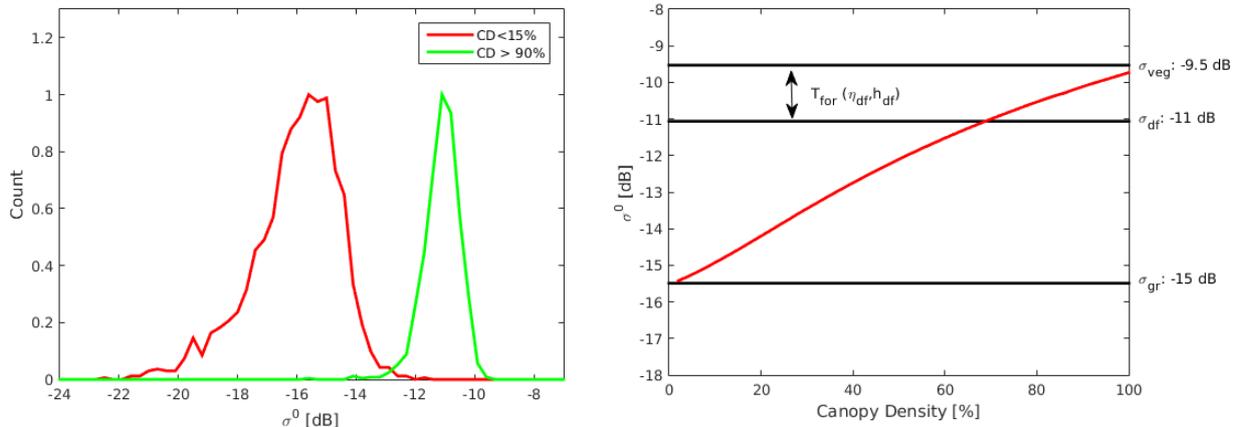


Figure 1-30. Left: Histogram of L-HV backscatter in areas of low (red) and high (green) canopy density. Right: Modelled relationship of L-HV backscatter as function of canopy density (red line) with σ_{gr}^0 and σ_{df}^0 derived from the histograms and σ_{veg}^0 derived from σ_{df}^0 with the aid of ICESAT GLAS based estimates of the residual transmissivity.

To simulate T_{for} at L-band from ICESAT GLAS, information about the two-way tree transmissivity, α , is required. Only a few measurements of the attenuation at L-band have been reported in the literature. Existing measurements and modelling results suggest that at L-band the two-way signal attenuation may be of the order of ~ 0.5 dB/m (Ulaby et al., 1990; Chauhan et al., 1991; Shinohara et al., 1992; Sheen et al., 1994; Kurum et al., 2009; Praks et al., 2012). Unfortunately, the information on attenuation for the major forest types is too sparse to permit adjusting α specifically for different forest types and imaging conditions. A sensitivity analysis performed in temperate, boreal and (sub-)tropical biomes showed however that α hardly affected the σ_{veg}^0 estimates. For instance, when using $\alpha = 1$ dB/m instead of 0.5 dB/m, σ_{veg}^0 changed by only up to 0.3 dB. Significant effects on the estimates of σ_{veg}^0 were only observed for values of α well below 0.5 dB/m which, based on our current understanding, may be associated with images acquired under frozen conditions. The PALSAR data used to generate the mosaics were primarily acquired under unfrozen conditions so that the use of a fixed value for α of 0.5 dB/m appears to be justified.

When estimating the transmissivity of the densest forests globally (i.e., with $> 90\%$ density according to Landsat) with the aid of ICESAT GLAS (Figure 1-31), we find that the transmissivity can be expected to be low across most of the tropical and temperate biomes, regardless of the value for the two-way tree attenuation α . The compensation of σ_{df}^0 for the residual transmissivity to derive σ_{veg}^0 will thus have a minor effect. However, in the boreal and sub-tropical zones, the effect of the correction is significant. Differences in the transmissivity due to varying two-way tree attenuation ($\alpha = 0.5$ or 1 dB/m) are expected to be minor.

	GlobBiomass		Page 49
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

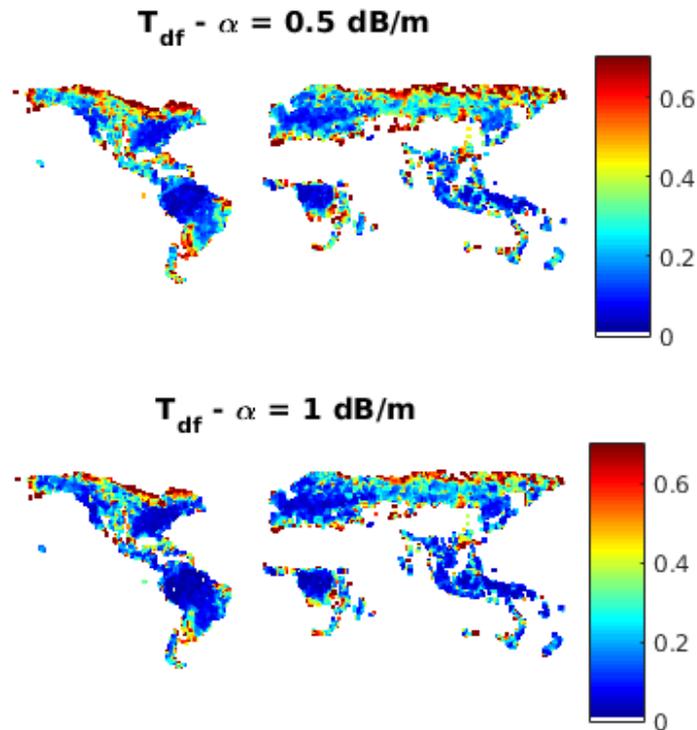


Figure 1-31. L-band transmissivity of dense forests modelled based on ICESAT GLAS based estimates of canopy density and height when assuming a two-way tree attenuation of 0.5 (top) and 1 dB/m (bottom), respectively.

The importance of accounting for the transmissivity of dense forests when estimating the model parameter σ_{veg}^0 with respect to the retrieval of GSV is illustrated in Figure 1-32 for $1^\circ \times 1^\circ$ PALSAR tiles covering forest areas in Montana (USA), Lapland (Sweden), and the Eastern Cape Province (South Africa). The retrieval of GSV was done with the BIOMASAR-L algorithm, in one case considering the transmissivity of the densest forests in the estimation of σ_{veg}^0 (Model 1) and in the other case assuming that σ_{df}^0 and σ_{veg}^0 are identical (Model 2); for the details about the model inversion approach, the reader is referred to the following sections. The examples show that Model 2 underestimates the backscatter as function of GSV and as a result leads to a systematic overestimation of GSV.

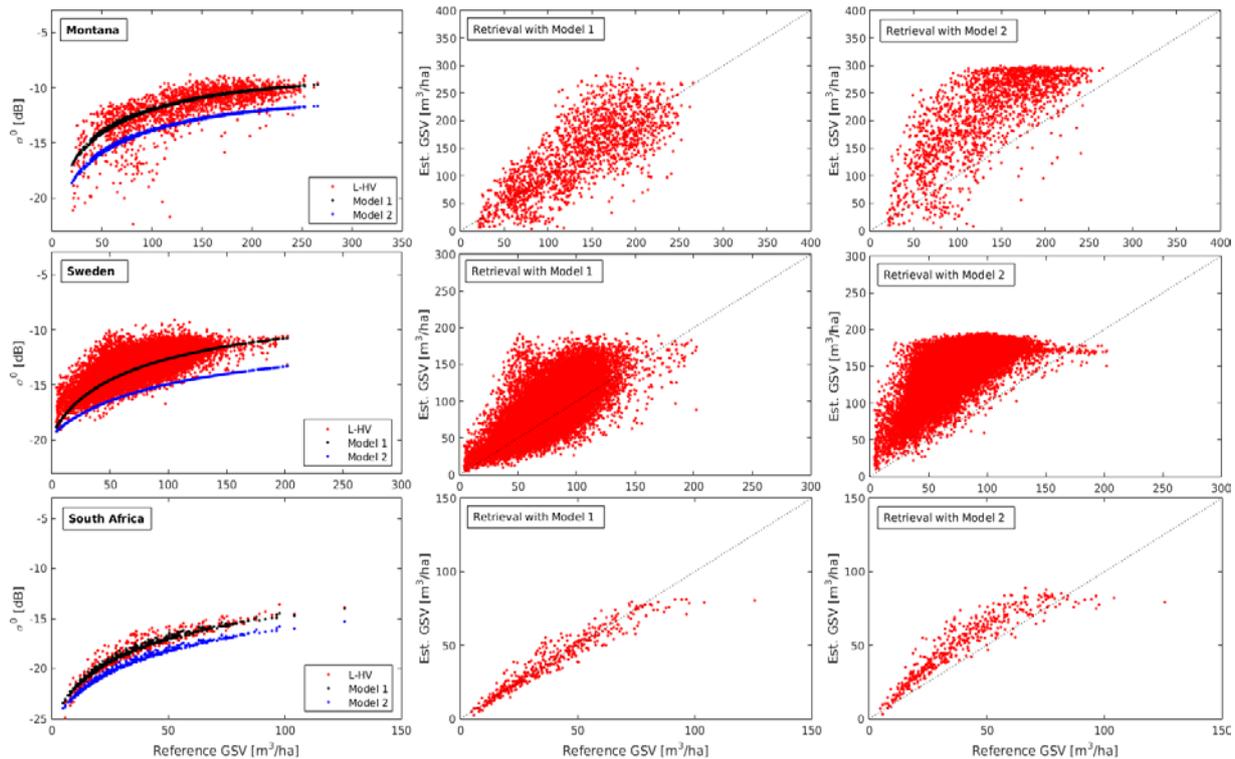


Figure 1-32. Observed and modelled relationship of L-HV backscatter and GSV for a forest area in Montana (top row) Sweden (centre row), and South Africa (bottom row) using BIOMASAR-L for model calibration (left). Model 1 accounts for the transmissivity of dense forests when estimating σ_{veg}^0 whereas Model 2 does not. The effect on the GSV retrieval is demonstrated in the centre and right plots.

1.4.2.2.3 Estimation of β

Studies concerned with modelling L-band backscatter as a function of GSV in boreal forest areas suggested that β was generally of the order of 0.004 ha/m^3 (Askne et al., 2003; Santoro et al., 2006) with no apparent dependence on the forest type, e.g., deciduous vs. evergreen, but with potential variations associated with changing imaging conditions, in particular in the case of frozen conditions for which an increased transmissivity of trees alters the relationship between transmissivity and GSV. Limited experience has been gathered so far for temperate, tropical and sub-tropical forests.

In order to determine which value for β is required across the major forest biomes to maximize the agreement of BIOMASAR-L GSV estimates with existing regional maps, the GSV retrieval with BIOMASAR-L was carried out repeatedly for the United States, Central Mexico and Yucatan, Colombian Amazon, Peru, Eastern South Africa, Sweden, Poland, and Kalimantan using values for β between 0.003 and 0.015 ha/m^3 . For a detailed description of the retrieval procedure, the reader is referred to the next section. The resulting GSV maps were compared against the existing regional maps (Kellndorfer et al., 2012; Asner et al., 2012, 2014, and the GlobBiomass regional maps produced by SLU, UoL, CESBIO, IGIK, RSS) after aggregating all maps to a common pixel size of $0.01^\circ \times 0.01^\circ$. The values for β which maximized the agreement of the BIOMASAR-L GSV estimates with the regional maps for each orbit and $1^\circ \times 1^\circ$ tile were then investigated on the basis of the FAO eco-region map.

Consistent with reports in the literature, values for β of $\sim 0.004 \text{ ha/m}^3$ were required in the boreal zone to maximize the agreement in terms of the root mean square difference (RMSD) with the

regional map (i.e., for Sweden). In the case of temperate and sub-tropical forests, the optimal values for β varied in a range between 0.004 and 0.008 ha/m³. In the tropics, instead, significantly higher values for β of on average 0.011 ha/m³ were generally required (Figure 1-33). The range of the optimal values for β was mostly consistent between the different forest regions and reference maps considered. However, discrepancies for some eco-regions were noticed. For the FAO eco-region “Tropical Mountain System”, for instance, the optimal value for β in Colombia was 0.012 ha/m³, whereas in the case of Mexico the optimal β for the same eco-region was 0.006 ha/m³. At this point it is unclear if such differences are a consequence of actual forest structural differences (in other words the FAO eco-region database is too generalised to capture forest structural differences affecting the relationship of L-band backscatter and GSV) or if the differences are due to the fact that specific eco-region differences had not been accounted for in the production of the regional maps, for instance due to the lack of *in situ* data in the respective eco-region.

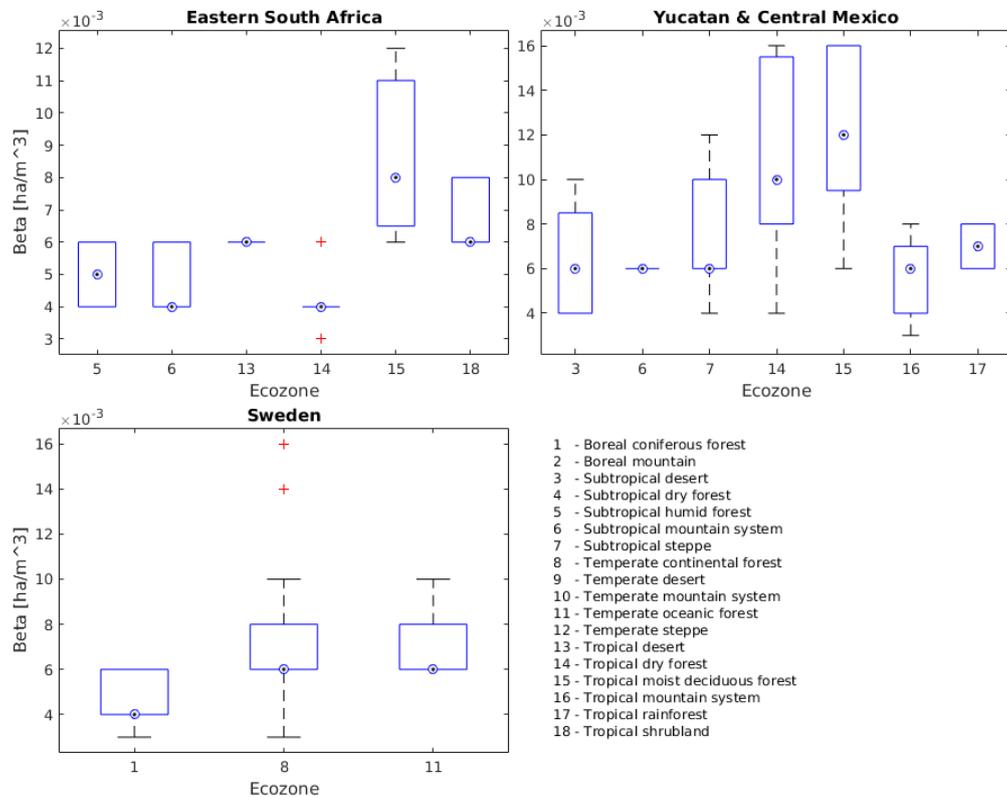


Figure 1-33. Values for the forest transmissivity parameter β required in different forest eco-regions to maximize the agreement between reference GSV maps for South Africa, Mexico, and Sweden and GSV maps produced from the 2010 ALOS PALSAR mosaics with BIOMASAR-L.

The results presented above suggest that there are systematic forest type-dependent differences in the relationship of the transmissivity at L-band and GSV. In order to verify this assumption, we analysed the relationship between T_{for} and GSV for each eco-region in the FAO eco-region map. T_{for} was simulated for each GLAS footprint with:

$$T_{for} = (1 - \eta) + \eta e^{ah} \quad (1-11)$$

where the GLAS height estimates were used to characterize the canopy height, h , and the corresponding average canopy density in each 65 m footprint according to the global Landsat product was used as estimate for η . The two-way tree transmissivity α was again assumed to be 0.5 dB/m. For each GLAS footprint covering one of the available GSV reference maps, the corresponding estimate of GSV was extracted to analyze the trend of T_{for} as function of GSV and to estimate the transmissivity parameter β by fitting Equation (1-12):

$$(1-12)$$

to the observed trend by means of non-linear regression.

Figure 1-34 illustrates the relationship of T_{for} and GSV as well as the fit of the model in Equation (1-12) for seven FAO eco-regions. The trend of the simulated T_{for} as function of GSV and the derived estimates for β confirmed the previous observation that significant differences between the eco-regions existed. The comparison indicated as well that low values for β of ~ 0.004 - 0.005 ha/m³ characterize best the relationship between T_{for} and GSV for boreal and sub-tropical dry forests, whereas for temperate, sub-tropical humid, and tropical forests, higher values for β in the range of ~ 0.006 to 0.011 ha/m³ are needed. In the analysis, it was assumed that a two-way tree transmissivity α of 0.5 dB/m would be adequate globally. When analysing the trend of T_{for} as a function of GSV simulated from ICESAT GLAS with a value for α of 1 dB/m, the derived estimates for β changed only for about 0.001 to 0.002 ha/m³. Larger differences of up to 0.004 ha/m³ were observed only in the tropics, i.e., the estimates for β reached values of up to 0.015 ha/m³. The high values obtained when assuming an α of 1 dB/m were however significantly higher than the values needed to maximize the agreement between BIOMASAR-L GSV estimates and the reference maps (Figure 1-34).

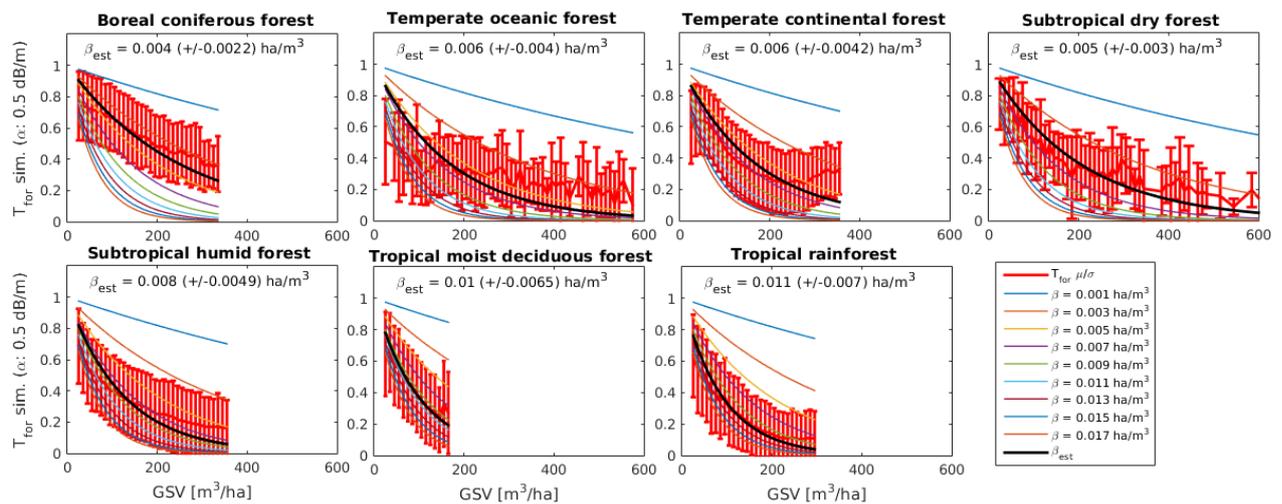


Figure 1-34. Estimates for β (incl. 95% confidence bounds) obtained for different FAO eco-regions by fitting the model in Equation (1-12) to the observed relationship of the forest transmissivity, simulated with the aid of ICESAT GLAS and Landsat canopy densities (Equation (1-11)), and the corresponding GSV in the regional map products.

Based on the results presented above, a global look-up table for the forest transmissivity coefficient β was compiled (Figure 1-35). For boreal and dry sub-tropical forests, β in the range of 0.004 to 0.005 ha/m³ will be used in BIOMASAR-L. In temperate and sub-tropical humid forest regions, higher values in the range of 0.006-0.008 ha/m³ and in the tropics values of 0.011 ha/m³ will be used. In some eco-

	GlobBiomass		Page 53
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

regions for which the optimal value of β for the GSV retrieval depended strongly on the reference map (e.g., the maps for Colombia or Mexico), we took into account if either of the maps had been produced using in situ data collected in that particular eco-region when deciding for the most plausible value of β .

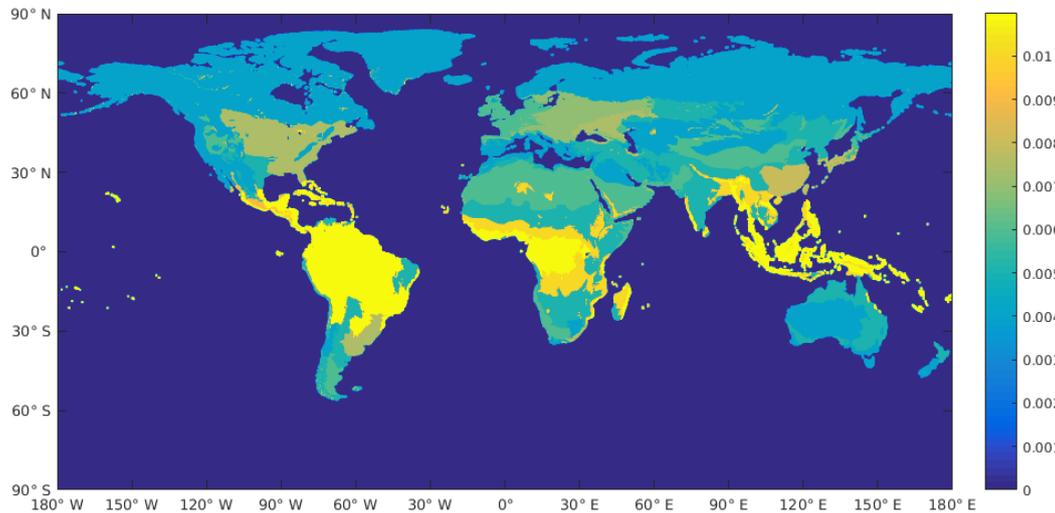


Figure 1-35. Values for the forest transmissivity coefficient β [ha/m^3] used for retrieving GSV from ALOS PALSAR data globally.

1.4.2.2.4 Retrieval of GSV

The model inversion approach for estimating GSV from individual L-band observations is generally the same as for BIOMASAR-C, i.e., the Water-Cloud model calibrated locally is inverted to estimate GSV at pixel level (Equation (1-6)). In this section, we focus on aspects of the retrieval not yet addressed in the context of BIOMASAR-C or specific to the implementation of BIOMASAR-L:

- 1) the maximum retrievable GSV
- 2) comparison of BIOMASAR-L retrieval results with reference maps
- 3) the multi-temporal/-polarization retrieval strategy

1.4.2.2.4.1 Maximum retrievable GSV

In the retrieval, it has to be considered that for intensity values exceeding σ_{veg}^0 , inversion of the model is not possible and for intensities slightly lower than σ_{veg}^0 , the inversion could result in GSV estimates far exceeding the range of realistic GSV values, which is why a maximum GSV level, GSV_{max} , has to be defined, up to which the inversion is carried out (cf., Santoro et al., 2011).

In BIOMASAR-C, GSV_{max} is determined with a constant positive offset of $50 \text{ m}^3/\text{ha}$ to the GSV corresponding to dense forests in a particular region, GSV_{df} . Since with BIOMASAR-L we aim to retrieve GSV at $\sim 25 \text{ m}$ pixel scale, it has to be considered that the maximum GSV that can be observed is scale-dependent. Figure 1-36 illustrates this for the National Biomass and Carbon Dataset for the United States (Kelldorfer et al., 2012). When comparing in a $1^\circ \times 1^\circ$ grid the maximum GSV that can be observed when analysing the map at full ($\sim 30 \text{ m}$) resolution or at aggregated 1 km scale, a ca. 30 % difference was found. The scale-dependent differences in the maximum GSV were generally similar for other forest areas, which is why for the GSV retrieval with the 25 m PALSAR mosaics we define GSV_{max} with:

	GlobBiomass		Page 54
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

$$GSV_{max} (@25m) = 1.3 * (GSV_{df} + 50m^3 / ha) \quad (1-13)$$

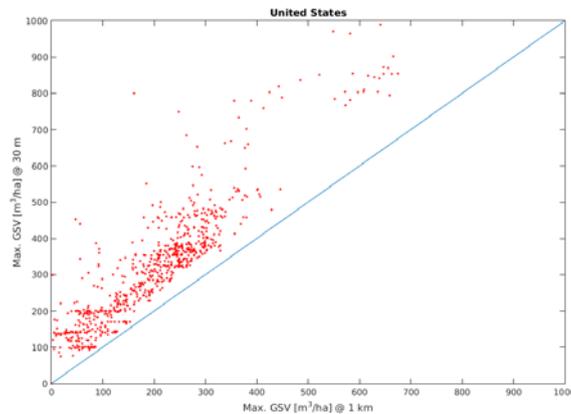


Figure 1-36. Scale dependence of the maximum GSV observed in the NBCD data set for the United States at 30 m and 1 km pixel scale, respectively.

1.4.2.2.4.2 Comparison of BIOMASAR-L retrieval results with existing maps

For an initial assessment of the performance of BIOMASAR-L, GSV maps were produced from the 2010 PALSAR L-HV mosaics for the conterminous United States, Central Mexico and Yucatan, the Colombian Amazon, Peru, South Africa, Poland, Sweden, and Kalimantan (Figure 1-37).

The GSV maps were compared against GSV or AGB maps produced by the GlobBiomass regional projects as well as additional available maps. The AGB maps available for the United States (Kellendorfer et al., 2012) and Poland (IGIK) were converted to GSV using a database of Biomass Expansion and Conversion factors compiled by Thurner et al. (2014) for latitudes > 30°. In the case of the AGB maps for Colombia (Asner et al., 2012), Peru (Asner et al., 2014), Mexico (UoL), South Africa (CESBIO), and Kalimantan (RSS), IPCC default conversion factors were used instead. When comparing the maps, it had to be considered that several of the reference maps reported AGB or GSV only within a predefined forest mask. Where available, the same forest/non-forest mask was applied to the BIOMASAR-L maps. The comparison was carried out at a pixel size of 0.01°x0.01°, i.e., all maps were aggregated to ca. 1km pixel size by simple block averaging.

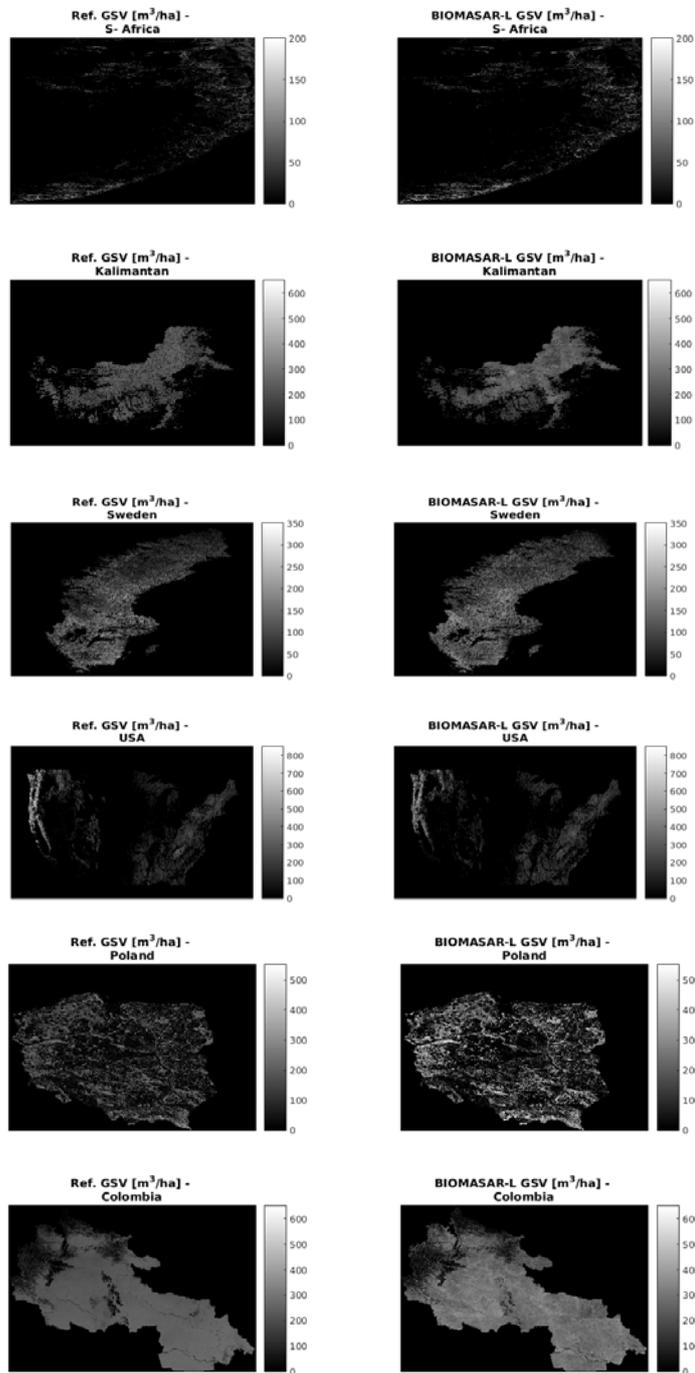
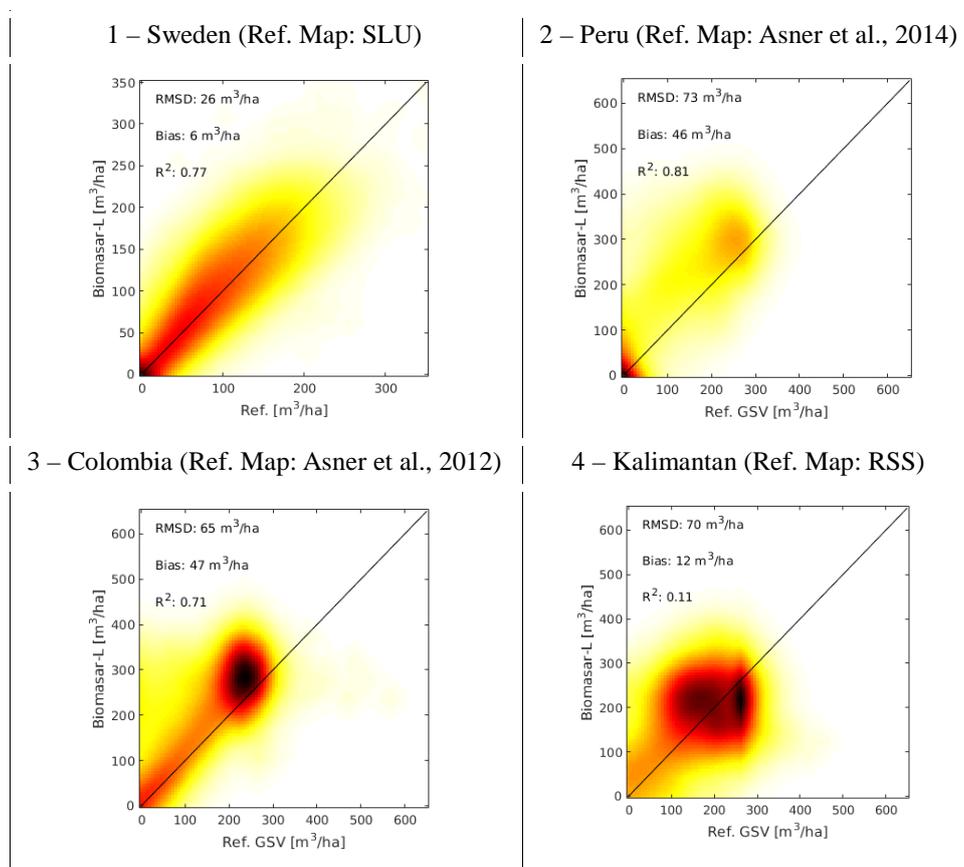


Figure 1-37. BIOMASAR-L GSV maps produced from 2010 ALOS PALSAR HV backscatter imagery (right column) and the available reference maps for the same area (left column).

The comparison of the BIOMASAR-L GSV maps with the set of reference maps (Figure 1-38) in most cases gave reasonable agreement, with values for the coefficient of determination, R^2 , in the range of 0.11 to 0.9. The RMSD varied in a range between 6.6 and 73 m^3/ha and increased with increasing average GSV in the different regions. The bias, i.e., the difference in the average GSV in the BIOMASAR-L and reference maps, was in the range 3 to 47 m^3/ha , with the largest bias observed for Peru and the Colombian Amazon. The agreement along the 1:1 line (Figure 1-38) was generally better for the lower ranges of GSV and regions with an overall lower level of GSV, such as South Africa or

Sweden. Clear limitations of the retrieval of GSV with ALOS PALSAR and BIOMASAR-L were noticed in forest regions characterized by continuous forest cover and high overall levels of GSV. For Peru, the Colombian Amazon and Kalimantan, no correlation between the GSVs in the reference maps and the BIOMASAR-L GSV estimates was found for GSVs in the reference maps beyond $\sim 200 \text{ m}^3/\text{ha}$ and $100 \text{ m}^3/\text{ha}$, respectively. The GSVs in the BIOMASAR-L maps tended to be higher than the GSVs in the reference maps for these areas (Figure 1-39). In the case of Mexico, BIOMASAR-L instead systematically underestimated the GSV of the densest forests with GSVs above $\sim 150 \text{ m}^3/\text{ha}$, which are primarily found at the higher elevations in Central Mexico. For the United States, good agreement along the 1:1 was observed for GSVs up to about $250 \text{ m}^3/\text{ha}$. In regions with higher levels of GSV, found primarily in the Pacific Northwest (Washington, Oregon, and Northern California), we observed general agreement of the BIOMASAR-L and reference map along the 1:1 line up to a GSV of $500\text{-}600 \text{ m}^3/\text{ha}$ (Figure 1-40), albeit with increasing spread for increasing GSVs. Only for GSVs beyond $600 \text{ m}^3/\text{ha}$ did the BIOMASAR-L map clearly underestimate the GSV, which was also obvious when comparing the histograms of the BIOMASAR-L and reference map (Figure 1-40).

It has to be considered that several of the reference maps (Poland, South Africa, Kalimantan, Mexico) had been produced using the same PALSAR data (as one of or even the only spatial predictor dataset) that was used for the retrieval with BIOMASAR-L. This implies that the spatial distribution of GSV is similar if not identical. Nonetheless, the low bias between the reference maps and the BIOMASAR-L maps, at least in the lower ranges of GSV, provides some evidence for the ability of BIOMASAR-L to predict the regional differences in the relationship between L-band backscatter and GSV.



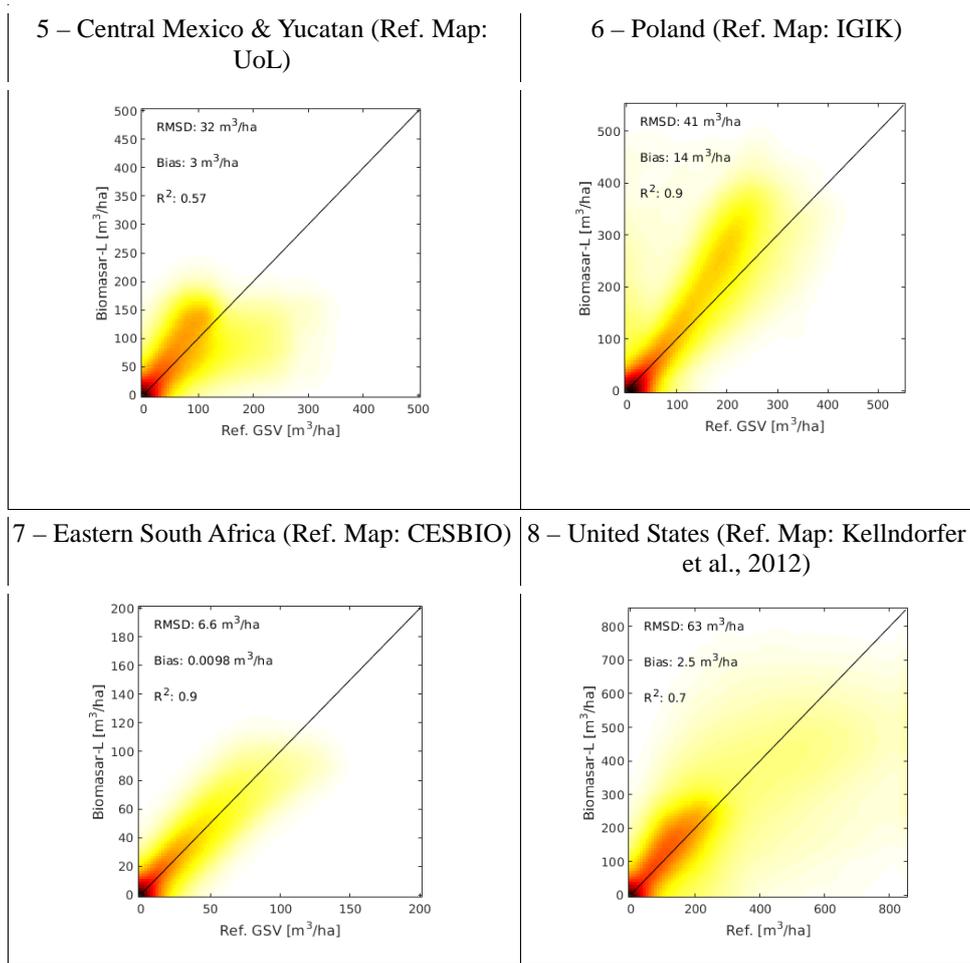


Figure 1-38. Comparison of maps produced with ALOS PALSAR 2010 L-HV backscatter and BIOMASAR-L against reference maps at 1x1 km² pixel size.

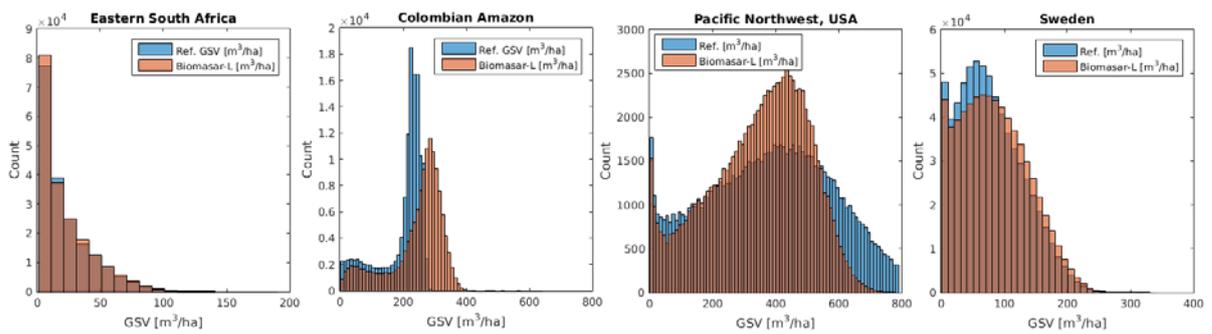


Figure 1-39. Histograms showing the distribution of GSV in the BIOMASAR-L maps and reference maps.

	GlobBiomass		Page 58
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

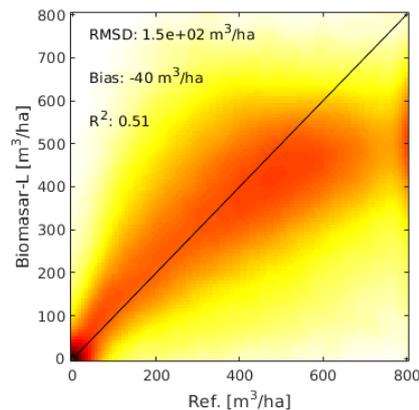


Figure 1-40. Comparison of a GSV map produced with ALOS PALSAR 2010 L-HV backscatter and BIOMASAR-L against a reference map (Kellndorfer et al., 2012) at 1x1 km² pixel size for the Pacific Northwest Region of the United States.

1.4.2.2.4.3 Multi-polarization/multi-temporal retrieval strategy

So far, only the GSV retrieval with the 2010 L-HV PALSAR mosaics has been discussed. Existing studies indicate that GSV retrieval benefits from the use of multi-temporal backscatter observations (Santoro et al., 2006; Cartus et al., 2012). In the BIOMASAR approach, improved estimates of GSV are achieved by retrieving GSV independently for each backscatter observation in a multi-temporal stack followed by weighted averaging of the multi-temporal GSV estimates, Equation (1-7). When averaging, more weight is given to those backscatter observations with the highest sensitivity to GSV, as indicated by the dynamic range (i.e., the difference between σ_{veg}^0 and σ_{gr}^0).

We therefore investigated if improvements in the GSV retrieval performance can be achieved when combining GSV maps obtained from the 2009 and 2010 L-HV mosaics. When testing the multi-temporal retrieval with BIOMASAR-L, we observed however that throughout all study areas in the (sub-)tropical, temperate, and boreal biomes, the GSV estimates that were obtained from the 2009 and 2010 imagery were, with few exceptions, nearly identical. Figure 1-41 illustrates this for the example of a 1°x1° PALSAR L-HV tile covering a forest area in northern Sweden where the 2009 and 2010 GSV estimates were highly correlated (Pearson correlation of 0.9).

The overall high agreement of the GSV estimates obtained from the different L-HV mosaics available for the years 2007 to 2010 was further confirmed when comparing the difference between retrieved GSVs and *in situ* GSV estimates available for ca. 19000 plots from a temporary plot network in Sweden. Figure 1-42 shows the high correlation of the difference between the BIOMASAR-L GSV estimates obtained from the 2007, 2009 and 2010 L-HV backscatter mosaics and the *in situ* GSV. As a consequence, no or only marginal improvements in the retrieval performance could be achieved when combining GSV estimates obtained from the 2009 and 2010 L-HV mosaics by means of weighted averaging. Also for the other study areas in the sub-tropical, tropical, and temperate biomes, no significant differences in the GSV estimates obtained from the 2009 and 2010 L-HV mosaics were observed. Figure 1-43 shows for the Colombian Amazon, Sweden, and South Africa that the agreement of the GSV maps produced from the 2009 PALSAR L-HV mosaics with the respective reference maps was similar to the agreement achieved with the 2010 imagery.

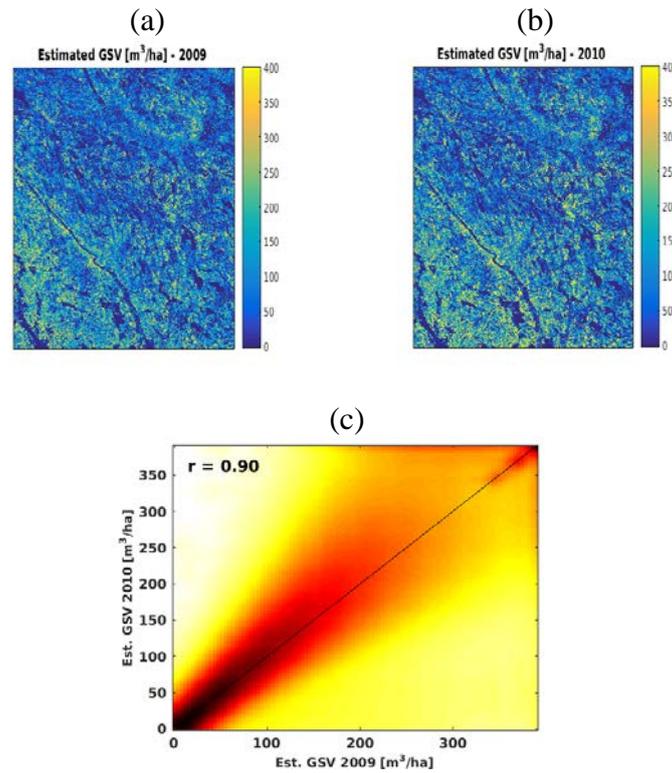


Figure 1-41. GSV estimates derived for an area in Sweden from PALSAR HV images acquired in (a) 2009, (b) 2010, and (c) the correlation of the GSV estimates from 2009 and 2010.

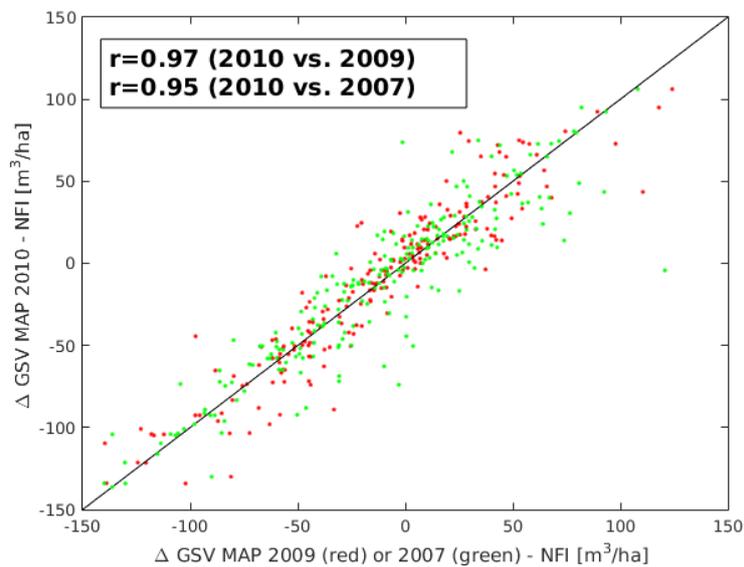


Figure 1-42. Comparison of the difference between GSV estimates obtained from the PALSAR L-HV mosaics for the years 2007, 2009, and 2010 and the BIOMASAR-L algorithm and GSV reference information available for 19.000 plots in Sweden.

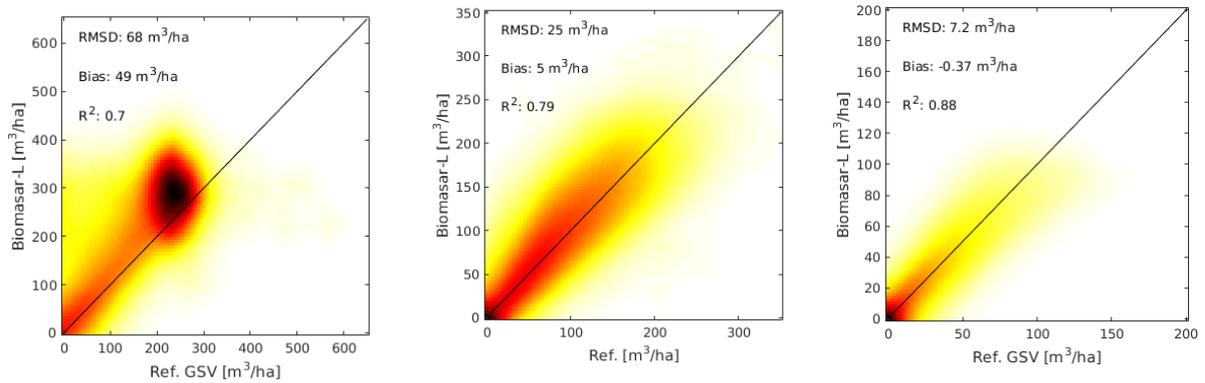


Figure 1-43. GSV estimates derived using the 2009 PALSAR mosaic for Colombia, Sweden, and South Africa.

Accordingly, we tested if the combined use of L-HV and L-HH polarization backscatter data would yield improved estimates of GSV. As for the multi-temporal retrieval, the GSV estimates obtained from the 2010 L-HV and L-HH backscatter mosaics were combined through weighted averaging with the weights determined from the difference of the local estimates for the model parameters σ_{veg}^0 and σ_{gr}^0 . The tests indicated, however, that for none of the study areas gave improvements in the agreement between the multi-polarization GSV maps and the reference maps. It is to be noted however that the retrieval with HH performed only slightly worse than the retrieval with the HV imagery from the same year. Figure 1-44 demonstrates this for the example of the Colombian Amazon and Sweden.

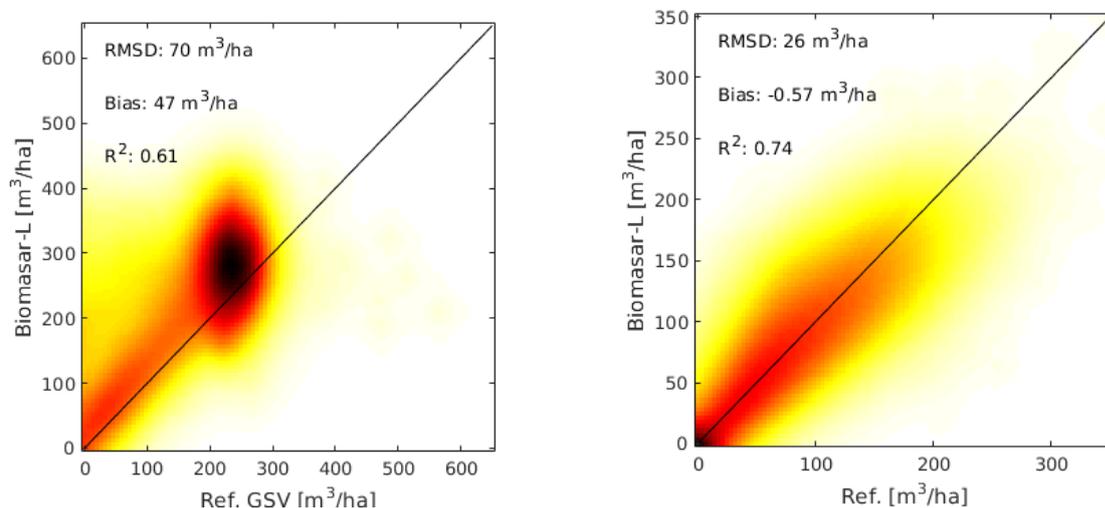


Figure 1-44. Comparison of GSV estimates obtained for Colombia and Sweden using PALSAR L-HH imagery from 2010 against the reference maps.

1.4.2.2.5 Limitations of BIOMASAR-L when using PALSAR mosaics

The development of the BIOMASAR-L algorithm aimed at the optimization of the automatic model calibration approach, which had been initially developed for C-band, to best predict the L-band backscatter to GSV relationship across all major forest ecosystems while minimizing the need for in situ data. One of the assumptions of the BIOMASAR-L algorithm is that the relationship between L-

	GlobBiomass		Page 61
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

band backscatter and GSV varies spatially and temporally with varying environmental imaging conditions, soil types, forest types, etc. and that models relating backscatter to GSV should therefore be as adaptive as possible spatially. Figure 1-45 shows the estimates for the model parameter σ_{gr}^0 derived with the BIOMASAR-L algorithm for the 2010 ALOS PALSAR L-HV backscatter mosaic over Africa. The map shows that the scattering power arising from the forest floor changes significantly across the continent, ranging from values of < -20 dB in Southern Africa to > -16 dB in the African tropics. For σ_{veg}^0 , instead, the derived estimates varied in a comparatively small range between -11.5 (tropical forest) and -10 dB (savannah), which is why the dynamic range (Figure 1-45, right) varied primarily with changes in σ_{gr}^0 . While the comparison of the BIOMASAR-L GSV estimates with regional maps across a wide range of different forest ecosystems suggested an overall reasonable agreement with mostly low bias, and thus that the spatially adaptive BIOMASAR-L parameter estimates are in a reasonable range, systematic validation needs to be done to clarify if the adaptive model calibration outperforms approaches in which field plots collected over large areas are used to calibrate models. The performance of the approach presented in Section 1.4.2.4 therefore serves as a benchmark.

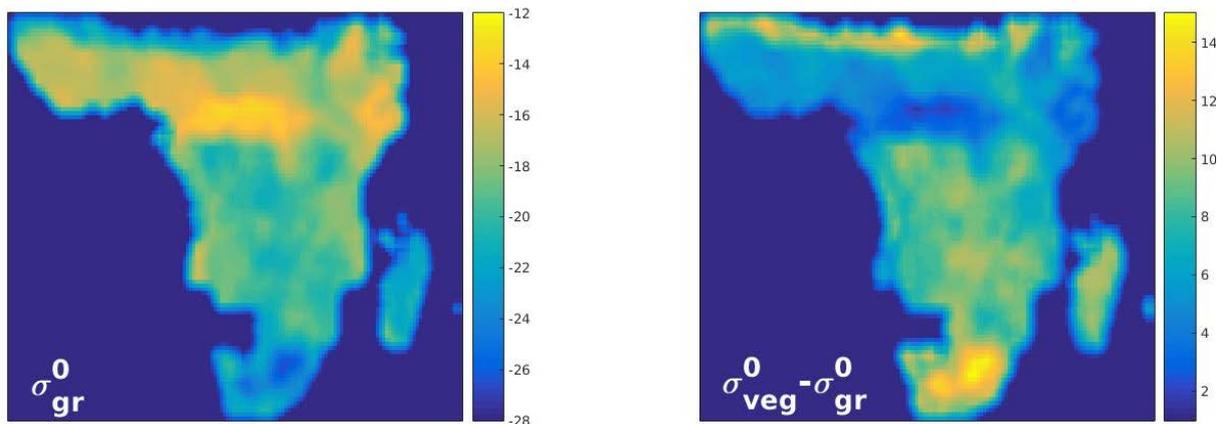


Figure 1-45. BIOMASAR-L estimates for the model parameter σ_{gr}^0 for each $1 \times 1^\circ$ tile of L-HV backscatter from 2010 across Africa.

Even a model that correctly predicts the backscatter to GSV relationship cannot overcome limitations in the sensitivity of the radar imagery used for the retrieval of GSV. The finding that a wide range of values for the forest transmissivity coefficient β between ~ 0.004 and 0.011 ha/m³ needs to be used for estimating GSV with ALOS PALSAR L-band data suggests that significant differences in the sensitivity of L-band backscatter to low and high GSV ranges exist across different forest ecosystems. Figure 1-46 (left column) illustrates for two examples (4 and 10 dB difference between σ_{veg}^0 and σ_{gr}^0) the modelled relationship between L-HV backscatter and GSV for values of β between 0.004 and 0.015 ha/m³. The figure shows that for increasing values of β the sensitivity of backscatter to GSV increases in the lower and decreases in higher ranges of GSV. The decreasing sensitivity for high values of β required in particular in the tropics implies differences in the level of GSV beyond which the signal effectively saturates. Following Hensley et al. (2014), the expected saturation level was calculated with:

$$F(V) = \frac{\sigma(V)}{\sqrt{N}} - \kappa V \frac{\partial \sigma}{\partial V}(V) \quad (1-14)$$

where N is the equivalent number of looks (ENL), κ the target error of the retrieval, $\sigma(V)$ the backscatter modelled as function of GSV, Equation (1-3), and $\partial\sigma/\partial V$ the partial derivative of the model. The saturation level is defined with the GSV beyond which the retrieval error associated with speckle noise exceeds the maximum allowable error (the goal of the GlobBiomass project is an error of 30% or better at a maximum pixel size of 500 m) and calculated with the GSV for which F becomes zero; note that other error sources (e.g., topographic effects, calibration errors, small-scale environmental effects not captured by the BIOMASAR-L model calibration, modelling error) in the relationship between backscatter and GSV may significantly increase the retrieval error but are not considered here. Figure 1-46 (right) shows that the saturation level strongly depends on β . In the case of an ENL of 1000 (i.e., the ENL of the PALSAR mosaics when aggregated to $\sim 500 \times 500$ m² pixel size), speckle alone will prevent the retrieval from reaching a relative error of 30 % beyond a GSV of 200 to 250 m³/ha in the tropics (i.e., high β), regardless of the dynamic range of the backscatter measurements (i.e., $\sigma_{veg}^0 - \sigma_{gr}^0$). In contrast, Figure 1-46 suggests that GSV may be retrievable across the entire range of GSVs that can be observed in the boreal and sub-tropical zones (low β), at least at a spatial scale of 500 m. The results in Figure 1-46 are somewhat supported by the observation that in the Pacific Northwest of the United States (values for β of 0.004-0.05 ha/m³ were used here), GSV could be estimated up to ~ 500 m³/ha.

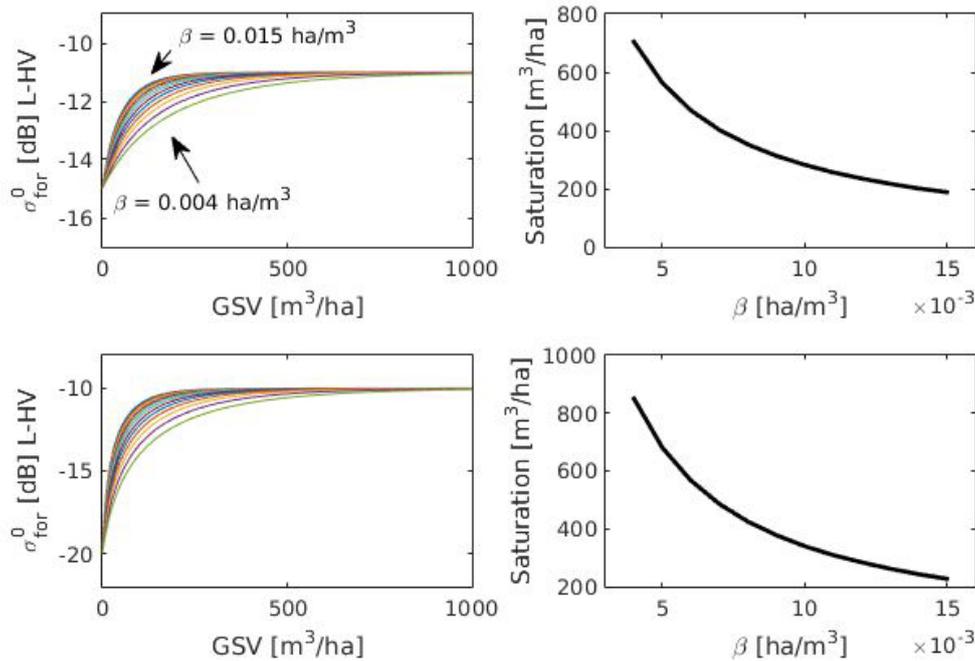


Figure 1-46. L-HV backscatter modelled as function of GSV for different values of the forest transmissivity parameter β given a dynamic range (i.e., $\sigma_{veg}^0 - \sigma_{gr}^0$) of 4 dB (top left) and 10 dB (bottom left) and the corresponding saturation level associated with speckle (ENL = 1000) and a maximum allowable retrieval error of 30 % (right column).

A consequence of the presumed low sensitivity of backscatter to high GSVs in particular in the tropics is that calibration errors, topographic effects, ionospheric effects, etc. in the PALSAR mosaics tend to exceed the sensitivity of backscatter to GSV. Figure 1-47 illustrates for an area in the Colombian Amazon how small-scale topographic effects that were not entirely corrected for in the SAR pre-processing (Shimada, 2010) affect the GSV retrieval with BIOMASAR-L. In the GSV map that was produced with BIOMASAR-L, the estimates present extremely small-scale variations with values of

around 100 m³/ha on slopes tilted away from the sensor and 500 m³/ha (i.e., GSV_{max} , see Equation (1-16)) on slopes facing the sensor.

The PALSAR mosaics also present striping effects over the tropics with in many cases dependence of the backscatter observed over areas of continuous forest cover on the local incidence angle. The PALSAR imagery in the mosaics had to some degree been compensated for the local incidence (θ_i) dependence of the backscatter by converting from σ^0 to γ^0 (i.e., $\sigma^0/\cos(\theta_i)$). However, the decrease of backscatter with increasing incidence angle depends on the cover type and cannot be compensated for completely by using a single correction function for all cover types (Hoekman & Reiche, 2015). The residual dependence on the local incidence angle that is observed over dense tropical forests is generally of the order of a few tenths of a dB but leads to a strong increase of the GSV estimates with increasing local incidence angle, as shown in Figure 1-47. Systematic correction/compensation for incidence angle effects is hindered by the radiometric balancing between neighbouring orbits that JAXA applied regionally to reduce striping, which along the edges of orbital tracks obscures incidence angle effects.

Since the sensitivity of L-band backscatter to GSV over dense tropical forests is expected to be low and distorting effects such as those presented above are observed frequently in all of the four annual PALSAR mosaics, the GSV maps that are produced from the PALSAR mosaics and BIOMASAR-L are expected to present a very high level of uncertainty (and a lot of clearly visible artefacts) over the tropics. It is thus questionable if the GSV retrieval with the PALSAR mosaics will allow improved GSV estimates compared to the 1 km GSV map that will be produced from hyper-temporal ENVISAT ASAR ScanSAR data.

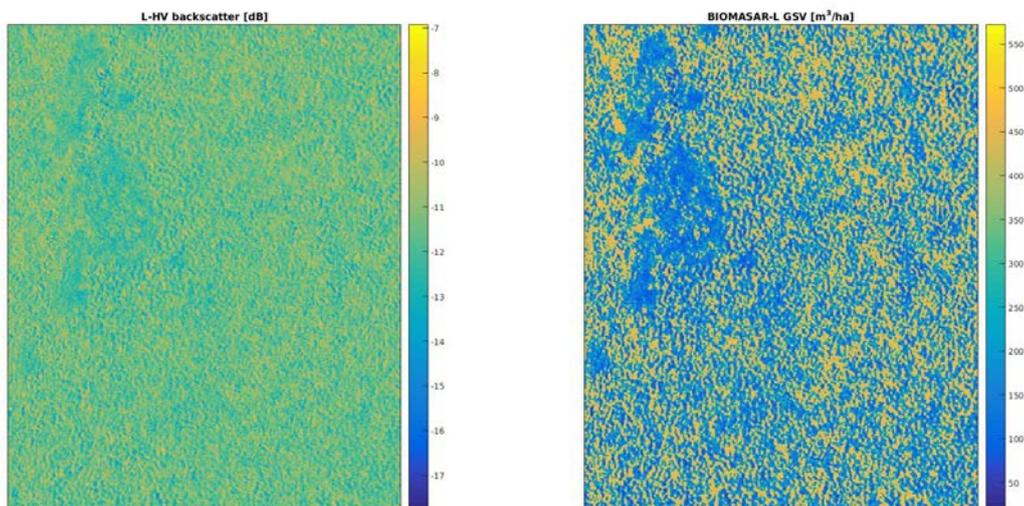


Figure 1-47. Illustration of how small scale topographic effects in the PALSAR mosaic that were not completely removed by topographic corrections (Shimada et al., 2011) affect the retrieval of GSV over dense tropical forest.

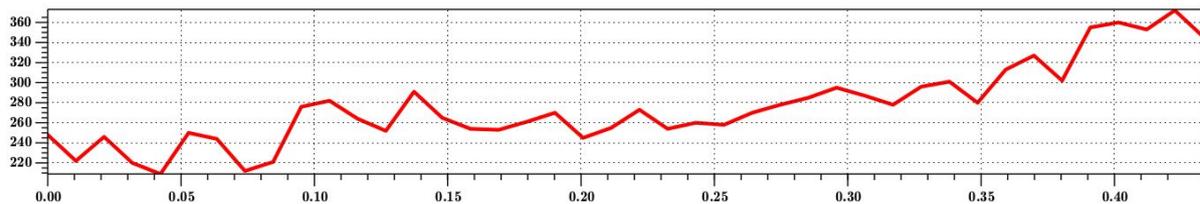
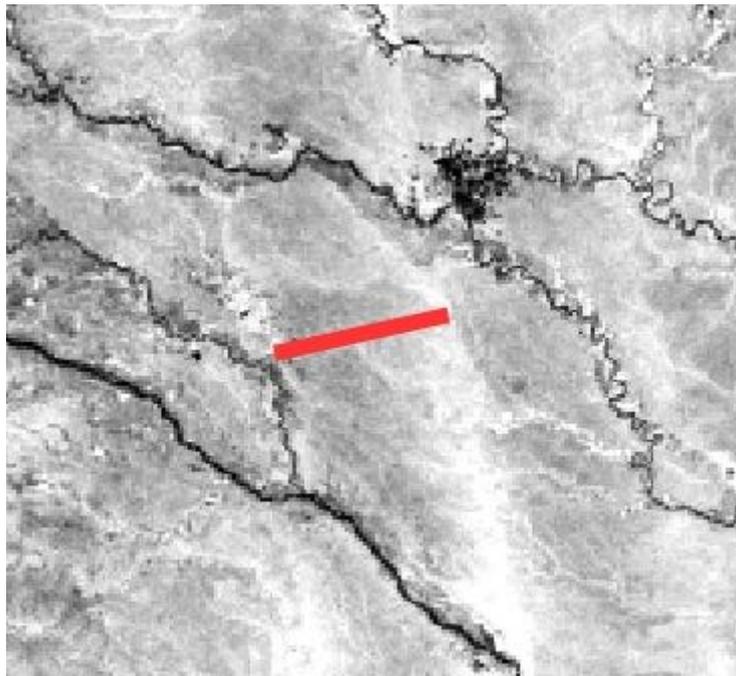


Figure 1-48. BIOMASAR-L GSV map for an area in the Amazon Basin (top). The profile (bottom) across one orbit of PALSAR data in the GSV map (red line) shows a clear dependence of the GSV estimates on the local incidence angle as the GSV systematically increases from near to far range.

1.4.2.3 Upscaling BIOMASAR-C GSV estimates

Tests of the BIOMASAR-C GSV map for the northern hemisphere indicated that the resolution of ASAR limits the retrieval accuracy in very heterogeneous forest ecosystems (Santoro et al., 2013, 2015b). The global retrieval with finer resolution Landsat and PALSAR datasets therefore aims to complement/improve the C-band 1 km GSV map with finer resolution estimates. Here we present two approaches, each converting the ASAR estimates to 30 m using auxiliary high resolution datasets (e.g., Landsat reflectances, ALOS PALSAR backscatter). At this scale, forest heterogeneity should be better captured.

In this project, we first attempted to capitalize on previous upscaling experiences. An approach similar to that in Sexton et al. (2013) for converting 250 m canopy density estimates from MODIS to 30 m using Landsat data was therefore implemented and tested. This approach involves three steps:

- i. The high-resolution datasets from Landsat and PALSAR are aggregated to the 1 km pixel size of the ASAR GSV maps.
- ii. Machine learning algorithms are used to develop models relating Landsat reflectance and PALSAR backscatter at 1 km pixel size to ASAR GSV.
- iii. These models are then used to predict GSV at the full resolution of the Landsat and PALSAR datasets.

	GlobBiomass		Page 65
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

To refine the 1 km GSV maps, we use Cubist (Quinlan, 1993; www.rulequest.com), which has been used to map AGB (Blackard et al., 2008), canopy density (Sexton et al., 2013) and tree height (Walker et al. 2007) using Earth Observation data. Comparisons of different machine learning algorithms suggest that the predictive performance of Cubist is amongst the best, and comparable to that of other popular algorithms (Moisen et al., 2003; Walker et al., 2007). One of the main reasons for choosing Cubist is the scale dependence of GSV, since most machine learning algorithms will not predict outside the range of values in the training dataset. Hence, although higher GSVs will be observed at 30 m resolution, such models will not predict GSVs higher than the maximum value in the 1 km map. Cubist does not have this limitation.

As for BIOMASAR-L, the refinement is carried out in $1^{\circ} \times 1^{\circ}$ tiles of the high-resolution datasets. For each tile, Cubist is trained by aggregating the high-resolution data to 1 km pixel size and developing models using the ASAR GSV map as response variable. To better capture local gradients in GSV, the area used for training is larger than the $1^{\circ} \times 1^{\circ}$ tile in which GSV is to be estimated. For each $1^{\circ} \times 1^{\circ}$ tile, a $3^{\circ} \times 3^{\circ}$ subset of the global mosaics, centred on the $1^{\circ} \times 1^{\circ}$ tile, is extracted from the global Landsat and PALSAR mosaics and aggregated to 1 km pixel size. To evaluate the predictive performance of the Landsat and PALSAR datasets, Landsat (i.e., the four reflectance bands) and PALSAR backscatter intensities in HH and HV polarization were used as predictors. Models for each $1^{\circ} \times 1^{\circ}$ tile were trained on a per PALSAR orbit basis, considering also the area covered by the same orbit in the neighbouring tiles.

Only pixels satisfying the following three conditions are used for model development:

1. in the BIOMASAR-C GSV estimates, only those 10% of pixels per $10 \text{ m}^3/\text{ha}$ wide GSV interval with the largest corresponding number of ENVISAT ASAR observations that were available for producing the map are used for calibrating Cubist;
2. the local incidence angle in the PALSAR mosaics is less than 28° and the layover/shadow mask provided by JAXA does not report layover or shadow;
3. the CCI land cover map reports forest cover.

The retrieval at 30 m pixel size is constrained to the same maximum retrievable GSV as the BIOMASAR-L retrieval, i.e., 30 % higher than GSV_{max} for BIOMASAR-C to account for the scale dependence of the parameter.

A first global GSV map produced by rescaling the ASAR estimates of GSV at 1 km scale to 30 m using Landsat and PALSAR data has been evaluated against the GlobBiomass regional map products as well as third party products.

The comparison of the maps at 1 km scale consistently showed that the rescaled maps gave better agreement with the reference maps than the original BIOMASAR-C map in terms of the RMS difference, coefficient of determination (R^2) and bias (Figures 1-50 – 1-54). The rescaling approach therefore represents a viable option for improving the global 1 km GSV map produced from ENVISAT ASAR C-band data. However, the rescaling approach was not found to improve the estimates of GSV in the higher ranges of GSV. In the case of the map produced for the Colombian Amazon, for instance, the bias in the GSV estimates compared to the reference map produced by Asner et al. (2012) is even larger than the bias of the original BIOMASAR-C product. Also for the United States, improvements in the agreement between the rescaled map and the reference map were limited to GSV ranges up to about $200 \text{ m}^3/\text{ha}$. For higher ranges of the GSV, which in the United States can be found primarily in the Pacific Northwest, the new map strongly underestimated the GSV; the original BIOMASAR-C map tended to overestimate the GSV. This observation indicates that the rescaling with

	GlobBiomass	Page 66	
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

PALSAR and Landsat data may yield improved estimates of GSV only in areas of heterogeneous forest cover but won't help in overcoming potential limitations in the sensitivity of the C-band data used to produce the BIOMASAR-C map at higher GSV/AGB levels.

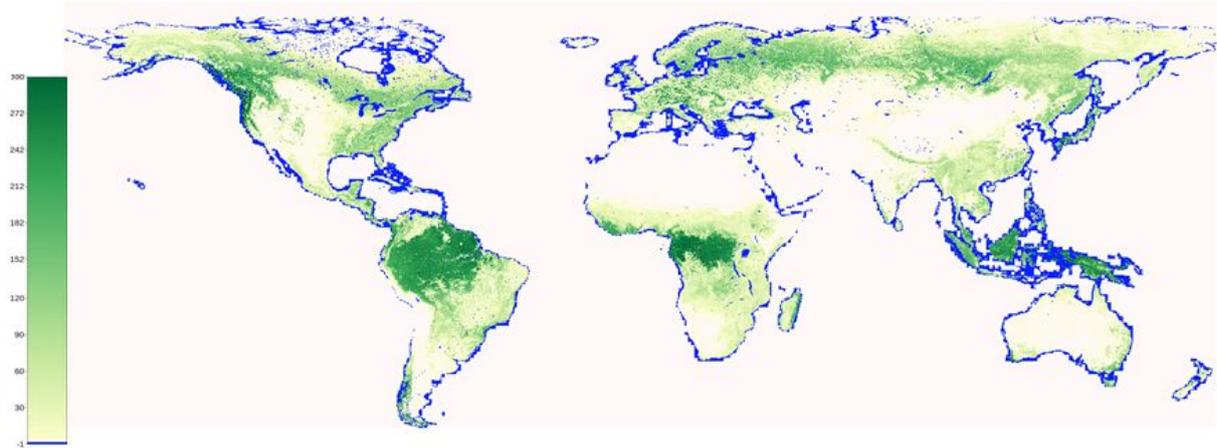


Figure 1-49. GSV map produced by re-scaling the BIOMASAR-C GSV map to 30 m pixel size using ALOS PALSAR and Landsat imagery.

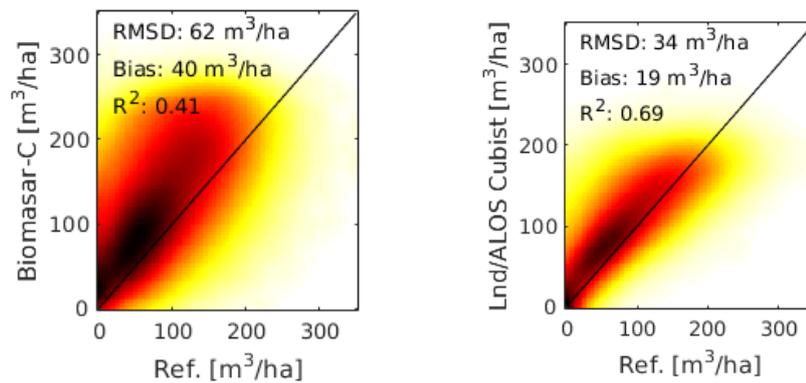


Figure 1-50. Comparison of BIOMASAR-C, Cubist-Landsat/ALOS GSV maps with the kNN GSV dataset for Sweden (SLU).

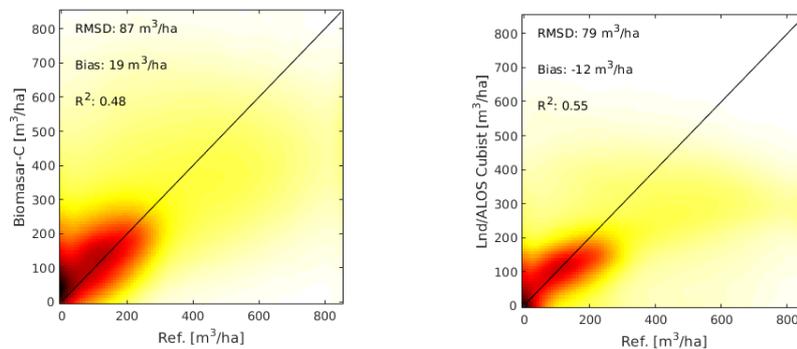


Figure 1-51. Comparison of BIOMASAR-C, Cubist-Landsat/ALOS GSV maps with a reference GSV map for the United States (Kelldorfer et al., 2012).

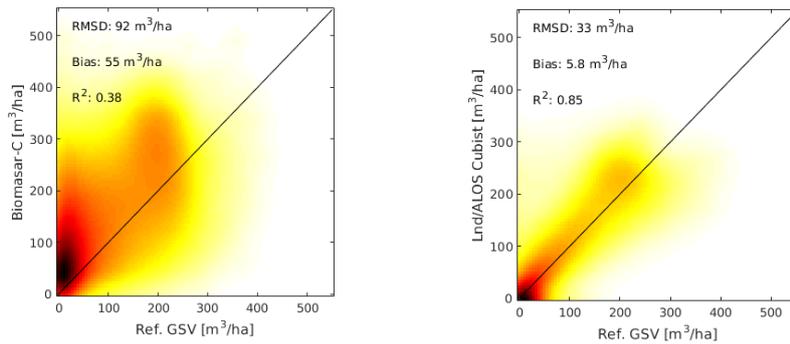


Figure 1-52. Comparison of BIOMASAR-C, Cubist-Landsat/ALOS GSV maps with a reference GSV map for Poland (IGIK).

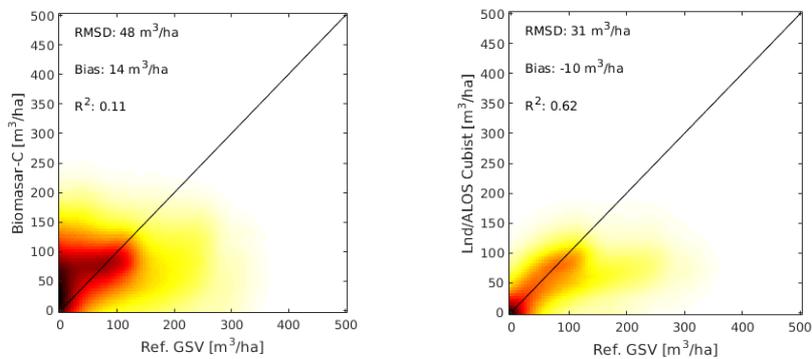


Figure 1-53. Comparison of BIOMASAR-C, Cubist-Landsat/ALOS GSV maps with a reference GSV map for Mexico (UoL).

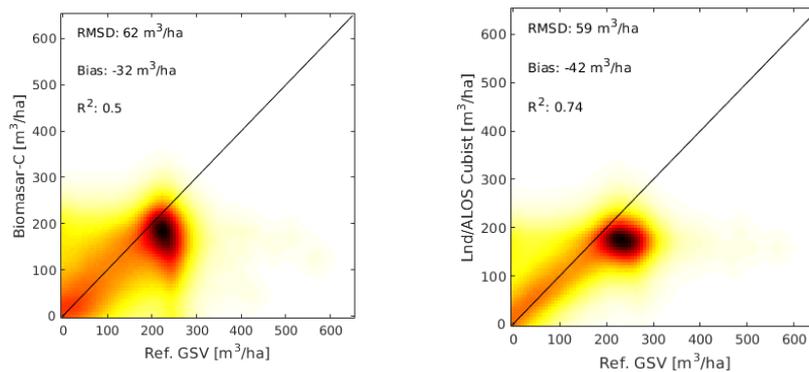


Figure 1-54. Comparison of BIOMASAR-C, Cubist-Landsat/ALOS GSV maps with a reference GSV map for Colombia (Asner et al., 2012).

Since rescaling with the Cubist method failed to produce results that were consistent with BIOMASAR-C, we resorted to a simpler rescaling algorithm based on multiple linear regression. The input data (PALSAR HH and HV backscatter, Landsat reflectances in Bands 3,4,5,7) and overall workflow remained the same as for Cubist but, for each 1x1 degree tile, linear regression models of the form:

	GlobBiomass		Page 68
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

$$\log(V) = a_1 + a_2\sigma_{HV}^0 + a_3\sigma_{HH}^0 + a_4B_3 + a_5B_4 + a_6B_5 + a_7B_7 \quad (1-15)$$

were now calibrated at 1km scale and then used to predict GSV at 25 m scale. Since the model was calibrated against the logarithm of the BIOMASAR-C GSV estimates, a bias correction had to be performed when retransforming the logarithmic GSV predictions to linear scale. This was done by evaluating the bias between BIOMASAR-C GSV and the new predictions aggregated to 1 km pixel size. The global map of GSV, produced using this approach is shown in Figure 1-55.

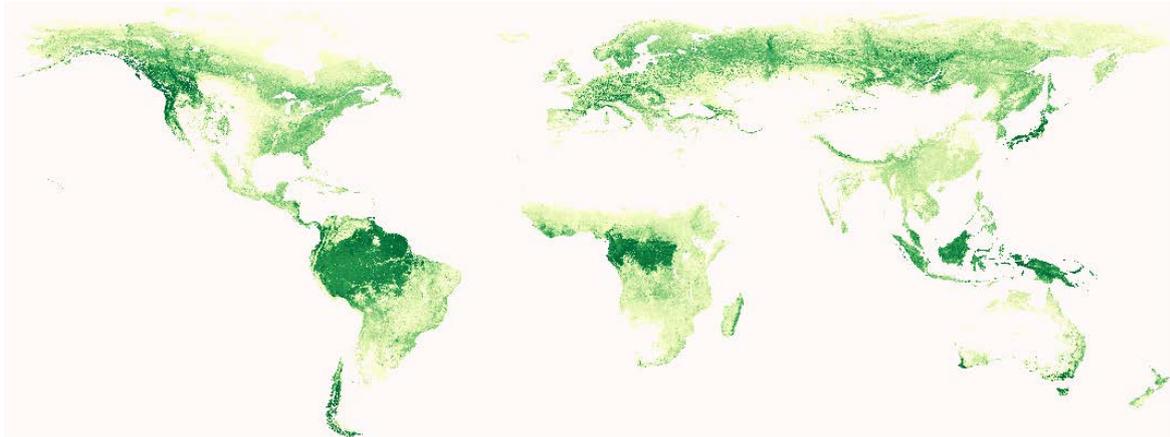


Figure 1-55. GSV map produced by re-scaling the BIOMASAR-C GSV map to 30 m pixel size using ALOS PALSAR and Landsat imagery by means of multiple linear regression models.

As for the evaluation of the Cubist based rescaling method, we compared the new predictions against the reference maps. The comparisons across the different test regions suggested an overall similar performance to Cubist in terms of root mean square differences. However, when evaluating the predictions in intervals of GSV, the results in the higher ranges of GSV tended to be improved and more consistent with the original BIOMASAR-C map.

Figure 1-56, in which the GSV predictions from the Cubist and the linear regression approach are compared, illustrates clearly that the linear regression produces significantly higher estimates across large parts of the high biomass forests in the world. Comparison of the maps with regional forest statistics such as the US FIA inventory (see also Section 1.4.3.1) furthermore showed that the GSV estimates obtained by rescaling BIOMASAR-C to 25 m were i) less affected by underestimation in the higher ranges of GSV and ii) more consistent with BIOMASAR-C. These observations, as well as the much lower computational demands of the simple regression, are the reason why BIOMASAR-C+ was eventually implemented using simple multiple regression models.

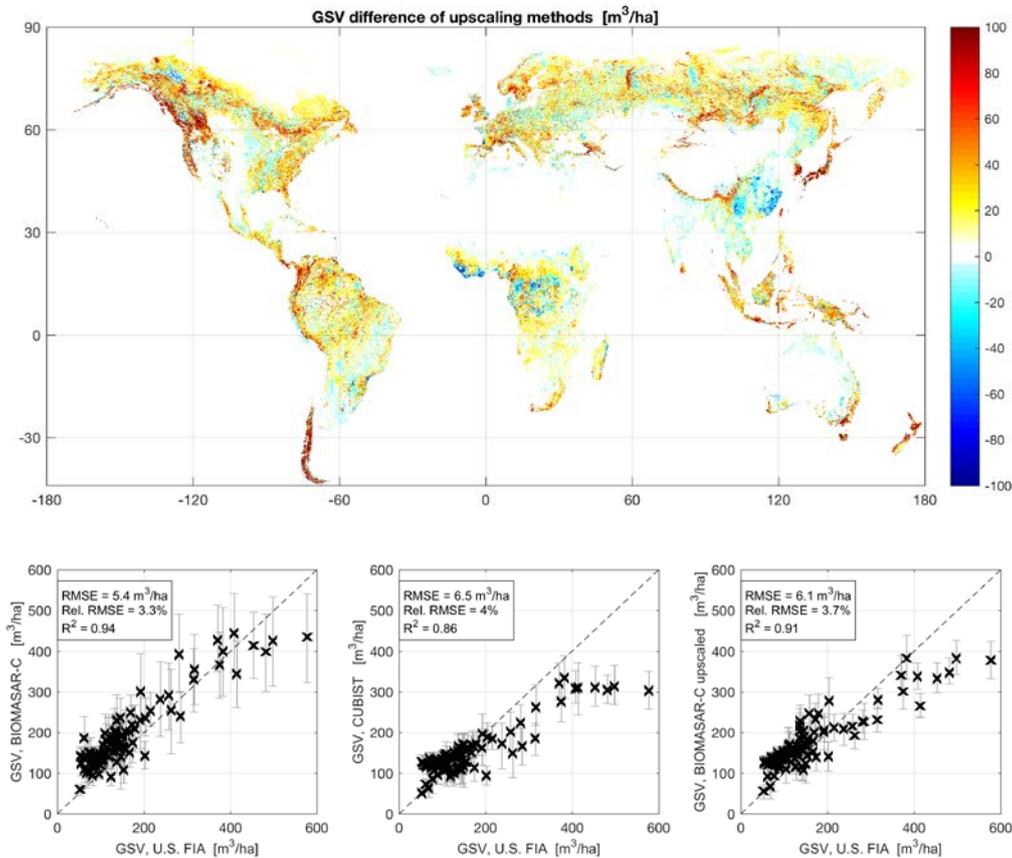


Figure 1-56. Difference in GSV between BIOMASAR-C+ maps produced using linear regression and Cubist (top). Comparison of BIOMASAR-C (bottom left), BIOMASAR-C+ (Cubist model) (bottom centre), and BIOMASAR C+ (linear model, bottom right) with US FIA county statistics on the average GSV.

1.4.2.4 CESBIO

This method has been described in Section 3.5 (ATBD - Regional Biomass Maps) and is a candidate algorithm for the global mapping for the low biomass tropical forest and savannah woodland. The method was developed and validated in Africa and is being validated in Australia and South America. With respect to the other algorithms here proposed for inversion of the high resolution EO data, in this method the coefficient of the forest transmissivity is data-driven and determined by a model-based approach. This aspect is relevant in particular in forests with multiple layers and heterogeneous structures such as in the tropical zone and in savannas where the assumption of a single value for the coefficient of the forest transmissivity might be too simplistic. The major limitation is the need for a set of *in situ* observations, which limits the usefulness of this approach if the training dataset of *in situ* measurements is located away from the area where the retrieval is applied.

1.4.3 Methods to derive a merged GSV data product

1.4.3.1 Inter-comparison of individual biomass datasets

In this Section, the BIOMASAR-C, BIOMASAR-L and upscaled BIOMASAR-C estimates of GSV are compared to understand the performance of each algorithm, identify bottlenecks and derive indications towards the combination of data products. In addition, we present results from the

	GlobBiomass		Page 70
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

comparison of the BIOMASAR-C and –L estimates of GSV with AGB values obtained with the CESBIO method in Africa for forest of biomass < 100 Mg/ha.

The analysis is undertaken at 0.01° pixel size, i.e., the native size of the BIOMASAR-C estimates. The BIOMASAR-L and Cubist GSV 0.01° estimates were instead obtained with spatial averaging (40 x 40 pixels), which reduced the spatial variability of the estimated values, believed to be, at the original scale of the ALOS and Landsat data, mostly noise due to weak sensitivity of biomass to the satellite observables. Similarly, the CESBIO dataset has been averaged but only those aggregated pixels are retained in which AGB was always below 100 Mg/ha. The advantages of working at coarse resolution are easier data handling and the possibility to work with much less noisy data, thus allowing trends, similarities and systematic errors to be identified.

To benchmark the estimates, we also used the global IIASA FAO dataset of GSV estimates (Kindermann et al., 2008) and the FAO FRA 2010 estimates of GSV from the country reports (FAO, 2010). The IIASA FAO GSV dataset is the only global dataset on GSV, being therefore a reference despite its acknowledged weaknesses (see beginning of Section 6). The FAO FRA 2010 country averages are also unique in the sense that for most countries of the world values of GSV have been reported for the epoch 2010. It is understood that the reliability of the GSV averages differs depending on the country. This information, even if not formalized in the form of flags, can be derived from reading the individual country reports. Once the FRA estimates are understood, the values are helpful to assess the quality of the remote sensing GSV estimates.

Since the retrieval algorithms are applied regardless of land cover, the estimates need to be filtered to retain only values in forested areas. A simple forest/non-forest mask was obtained from the CCI Land Cover map of epoch 2010. Classes 40 to 110, 160 and 170 were retained as forest. The remaining classes were associated with the non-forest class. Although the definition of “forest” is simplistic, it is sufficient given that at this stage we are primarily interested in identifying trends and systematic errors.

Working at coarse resolution and applying a forest mask creates, however, some issues that may distort the interpretation of the results. The aggregated estimates of biomass in patchy forest landscapes can be questioned because: (i) the full-resolution estimates over non-forest may be wrong, and (ii) the labelling of the reference land cover map may be erroneous. Scale and labelling issues are considered in this context and referred to when necessary.

To start with, we show in Figures 1-57, 1-58 and 1-59 examples of GSV estimates obtained with BIOMASAR-C, BIOMASAR-L and BIOMASAR-C+, respectively. Here, only pixels belonging to the forest class as defined above have been retained. Water pixels according to the CCI Land Cover product have been coloured in blue. Red is used for pixels labelled as forest according to CCI Land Cover but for which no value was retrieved (e.g., lack of SAR data or backscatter outside the range of modelled backscatter to be inverted).

The BIOMASAR-C dataset exhibits some voids (see e.g., New Zealand) in areas with very few ASAR observations in 2010. This issue may be overcome by including more years in the hyper-temporal dataset or shifting to the year with the largest number of acquisitions. Given that voids are local, it is assumed that void-fills from other years will not affect the overall quality of the GSV dataset. Extending or shifting the time interval of the ASAR data may, however, have some consequences that need to be considered, such as the consistency in time of the estimates, i.e., whether the estimates differ depending on the time interval selected.

BIOMASAR-L exhibits voids mostly in the northern hemisphere because of the backscatter being lower than the range of modelled backscatter values. This is a drawback of having only one SAR image available for the retrieval.

The BIOMASAR-C+ dataset has no voids. Nonetheless, the values attributed to voids in the original BIOMASAR-C map can be questioned as the void-fill was based purely on PALSAR and Landsat observations.

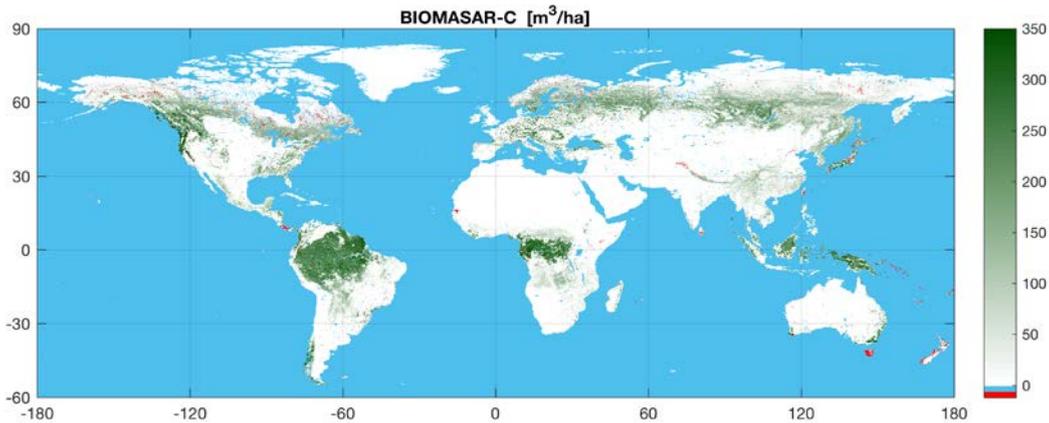


Figure 1-57. Estimates of GSV with a pixel size of 0.01° obtained with the BIOMASAR-C algorithm.

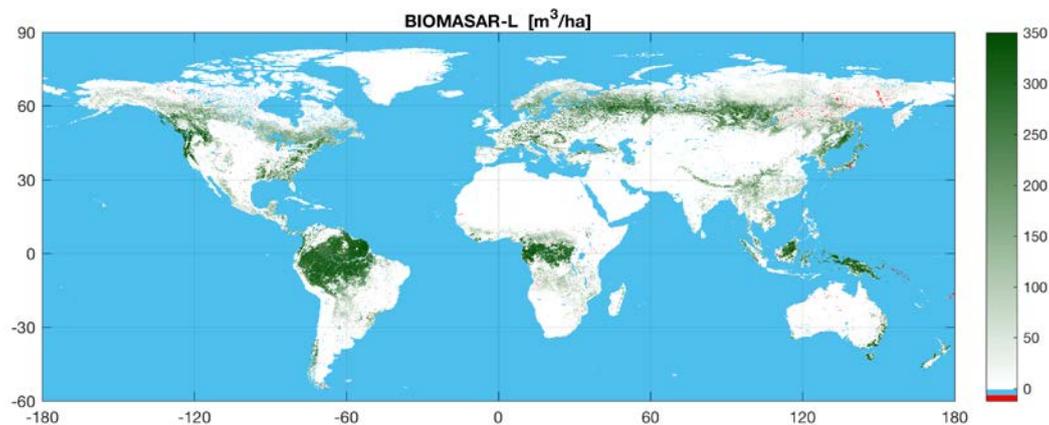


Figure 1-58. Estimates of GSV with a pixel size of 0.01° obtained with the BIOMASAR-L algorithm.

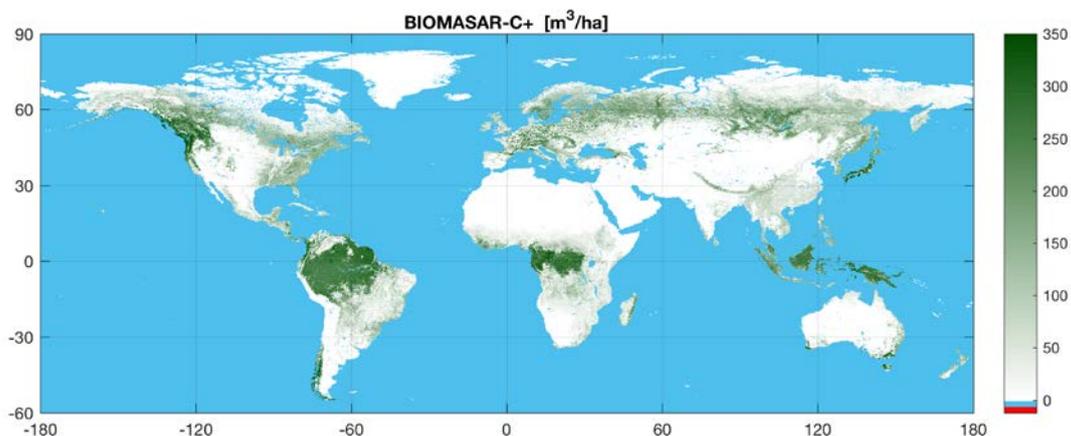


Figure 1-59. Estimates of GSV with a pixel size of 0.01° obtained with the Cubist algorithm.

Figures 1-57, 1-58 and 1-59 show that the GSV distribution was similar; nonetheless, there are some regional differences between BIOMASAR-L and the BIOMASAR-C datasets. To better illustrate the differences between the datasets, we have generated average GSV values per latitude, i.e., for each latitude between 90° N and 60°S we generated the average GSV with a step of 0.01°. Figure 1-60 shows the scatterplots comparing the three datasets. These longitudinal averages did not cluster along the 1:1 line, but we did not observe strongly contrasting results, indicating that our approaches were consistent in the way biomass was estimated. Although the BIOMASAR-C+ dataset is formally an upscaled version of the BIOMASAR-C dataset, some areas exhibited discrepancies, with BIOMASAR-C GSV being lower than Cubist (left panel). The comparison between BIOMASAR-C and –L GSV (central panel) confirms that the L-band estimates were often larger than the corresponding C-band values. In a few cases, the estimates from ASAR were larger than those obtained with PALSAR. As a result, the BIOMASAR-L estimates were slightly higher than the BIOMASAR-C+ estimates, with a few exceptions.

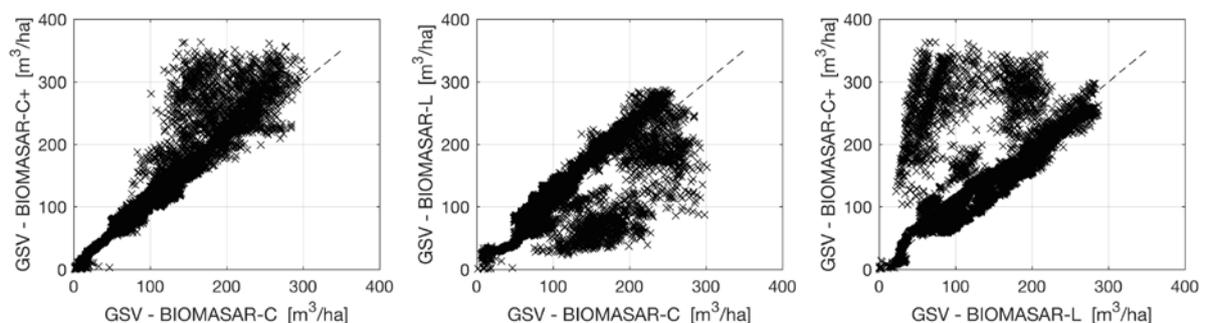


Figure 1-60. Scatter plot comparing averages per latitude (0.01° step) of BIOMASAR-C, BIOMASAR-L and BIOMASAR-C+ GSV at 0.01° resolution.

Figure 1-61 gives more insight into the reasons for the discrepancies. Profiles of the longitudinal GSV averages are shown for each of the three datasets. In addition, we included a profile for the IIASA FAO GSV dataset and a surrogate profile for the FAO FRA 2010 country-wise estimates. We explicitly refer to a “surrogate” profile since this is not a correct representation of longitudinal distribution of GSV. The profile was obtained by associating with each pixel the value the FRA 2010 country reports associate to a specific country. The plot includes also the profile for the merged dataset, which will be referred to in Section 1.4.5.5.

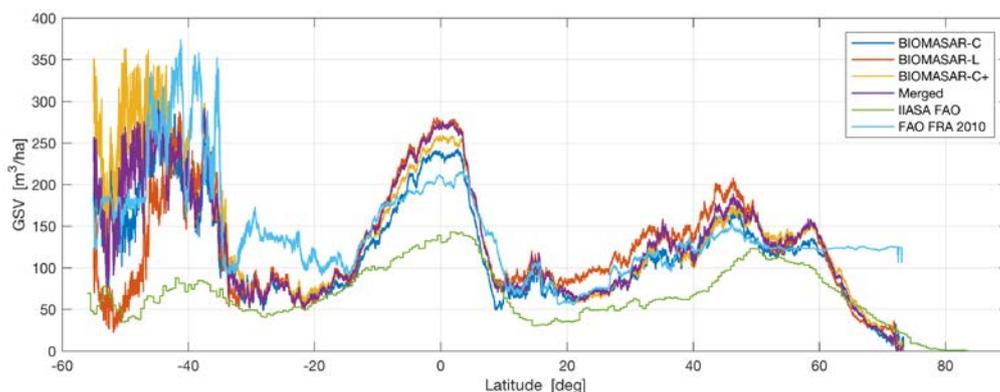


Figure 1-61. Profile of longitudinal averages of GSV from the BIOMASAR-C, BIOMASAR-L, Cubist, IIASA FAO and FRA 2010 datasets.

	GlobBiomass		Page 73
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

The three profiles for the GlobBiomass datasets exhibited similar shapes but were often offset. The FAO IIASA GSV dataset used to benchmark the GlobBiomass retrievals showed the same spatial distribution but, in terms of absolute values, this dataset always exhibited the smallest estimates. Given that the IIASA GSV dataset is currently the only global representation of GSV, the GSV datasets generated within GlobBiomass appear to change substantially the knowledge of spatial distribution of GSV. The latitudinal distribution of GSV from the GlobBiomass datasets agrees with the information derived from the FAO FRA statistics. The interpretation of similarities and differences between the FAO FRA profile and the GlobBiomass profiles is cumbersome and can be misleading. For large countries covering several degrees of latitude, a single value of GSV was associated with its entire extent. The effect is visible at latitudes north of 70°, showing GSV above 100 m³/ha. This is the average of the FRA averages for Canada and Russia, which are obviously not valid at these latitudes. In this paragraph, the evaluation is restricted to the three GlobBiomass datasets:

- In the boreal zone (north of 60°N), all datasets agreed in representing the increase of GSV for decreasing latitude. North of 65°N, the BIOMASAR-C and BIOMASAR-C+ GSV strongly agreed and presented a stronger gradient than BIOMASAR-L. The agreement between the three datasets was strong between 60°N and 65°N.
- In the temperate zone, between 40°N and 60°N, the highest GSV was reached in the northern hemisphere but with remarkable differences in terms of absolute values. North of 50°N, the C-band estimates were higher than the L-band estimates; north of 40°N, the BIOMASAR-L estimates were higher.
- In the subtropical zone between 20°N and 40°N, the difference between BIOMASAR-L and the other datasets was large, especially at latitudes south of 30°N.
- In the tropical zone, between -20°N and 20°N, the agreement between BIOMASAR-L and the other two datasets differed. In humid tropical forest (+/-10°), BIOMASAR-L estimated more GSV. On the contrary, in dry tropical forest (10°/20°N and -20°/-10°N) the estimates agreed.
- For forests of the southern hemisphere south of -20°S (maritime forest, temperate forest, miombo woodland, woody savannah), the BIOMASAR-C, -C+ and -L agreed.
- Large discrepancies occurred in the temperate forest of the southern hemisphere, with BIOMASAR-C+ being much higher than BIOMASAR-C and BIOMASAR-L (see e.g., south of -40°N). At such latitudes, the forest cover is sparse (the Andes between Chile and Argentina, Tasmania, New Zealand) and BIOMASAR-C could not estimate GSV in most regions because of the very small number of ASAR observations in 2010 (see red areas in Figure 1-57). Discrepancies between the two datasets south of -45°S cannot be judged since BIOMASAR-C could retrieve GSV only for part of the forests as a consequence of the paucity of ASAR acquisitions in 2010.

In a more detailed assessment of the agreement between datasets, we compared regional statistics of GSV obtained from forest inventory and corresponding averages from the GlobBiomass product. In (Santoro et al., 2015a), this approach was applied for countries of the northern hemisphere to assess the quality of BIOMASAR-C estimates of GSV. Here, we have extended our analysis to BIOMASAR-L and BIOMASAR-C+ estimates. Comparisons confirmed the indications from the analysis of the latitudinal gradients, revealing at the same time discrepancies at much finer scale between the estimates. A visual comparison of the maps overlaid onto Google Earth imagery helped to explain, at least in part, several discrepancies. The BIOMASAR-L estimates were affected by the approximate

	GlobBiomass		Page 74
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

correction for topography applied by JAXA and by local events in terms of environmental conditions. Another systematic error that was identified was the strong overestimation of biomass in areas affected by recent forest cover changes (fire, deforestation) presenting high soil moisture. Detailed reporting of all such issues is beyond the scope of this document. The BIOMASAR-C+ estimates were in agreement with the original BIOMASAR-C estimates, meaning that they were partly affected by the underestimation of GSV in high biomass forest.

In summary, we identified systematic deficiencies of the BIOMASAR-L and BIOMASAR-C+ datasets, which appeared to be different in nature. This suggested developing a merging scheme that would compensate for such errors and generate improved estimates with respect to the two input datasets. The merging approach is outlined in Section 1.4.3.2.

The global retrieval algorithm includes datasets produced by CESBIO for the dry tropics. The scope was to see whether the more advanced retrieval approach compared to BIOMASAR-L would yield results that could be included in the global estimates of biomass. CESBIO generated a dataset of AGB for African dry forest, validated with 144 forest inventory plots and a small number of Lidar-based AGB dataset in South Africa (Bouvet et al., submitted); in addition, datasets were generated with the same parameters used in Africa for South America and Australia. These have not been validated. Our assumption is that the CESBIO approach estimates biomass more accurately than BIOMASAR because it uses an electromagnetic (EM) model to derive model parameter estimates rather than a statistical approach, but applies a Bayesian combination of the HH and HV backscatter to estimate biomass rather than inverting the EM model. In Figures 1-62 to 1-70, we show scatterplots of AGB (from CESBIO) and GSV (from BIOMASAR-C and BIOMASAR-L) after stratifying the estimates (at 0.01° pixel size) in terms of ecological zone represented in the FAO GEZ dataset. Note that here we compare only those aggregated pixels that contain AGB below 100 Mg/ha, i.e., aggregated pixels that contain AGB estimates above 100 Mg/ha are discarded from the analysis. In each plot, the red dashed line represents the scaling that should be applied between GSV and AGB according to the IPCC BCF.

For forest within the tropical rainforest and tropical moist regions (GEZ codes 11 and 12, respectively), the difference between the biomass estimated with BIOMASAR-L and the estimates of AGB from the CESBIO dataset increased for increasing biomass (Figure 1-62 and 1-63). The red line represents the 1:1 line for biomass, assuming a hypothetical BCF of 1.0 for tropical humid regions. It is surprising to observe that a substantial number of aggregated pixels exhibited large GSVs despite our having selected only aggregated pixels containing AGBs below 100 Mg/ha. This deserves further investigation with in situ data to understand the reason for the discrepancy.

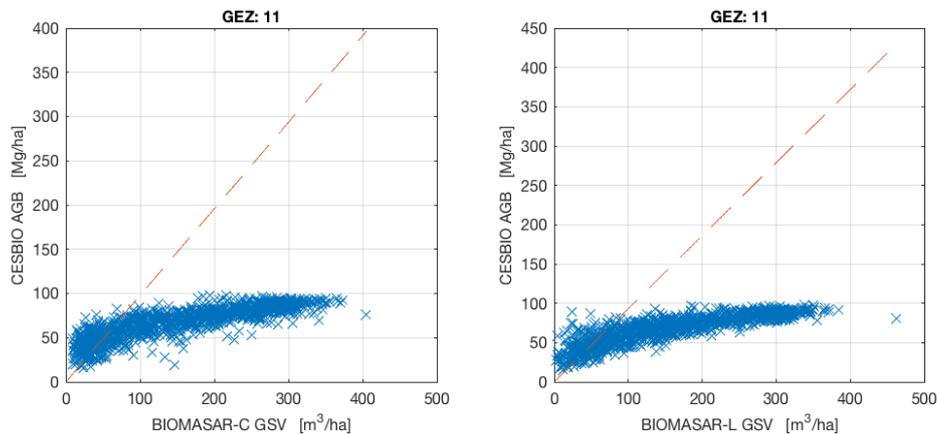


Figure 1-62. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: tropical rainforest.

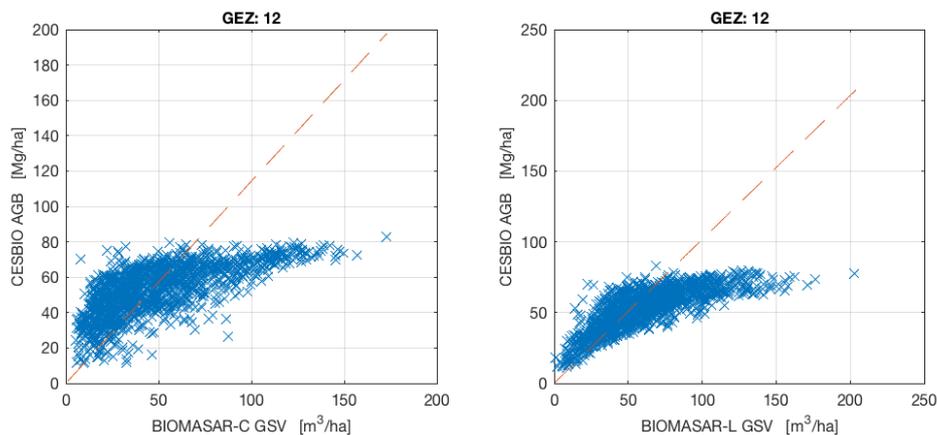


Figure 1-63. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: tropical moist forest.

For forests in the tropical dry (GEZ = 13, Figure 1-64), tropical steppe (GEZ = 14, Figure 1-65) and tropical desert zones (GEZ = 15, Figure 1-66), we found reasonable agreement between AGB and GSV estimates, in particular for BIOMASAR-L. The AGB by CESBIO tends to saturate with respect to the corresponding values by BIOMASAR-L when reaching the 100 Mg/ha level. Nonetheless, the slope of the scatterplots between BIOMASAR-L GSV and CESBIO AGB, which can be thought of as representing the BCEF, was in good agreement with values reported by the IPCC.

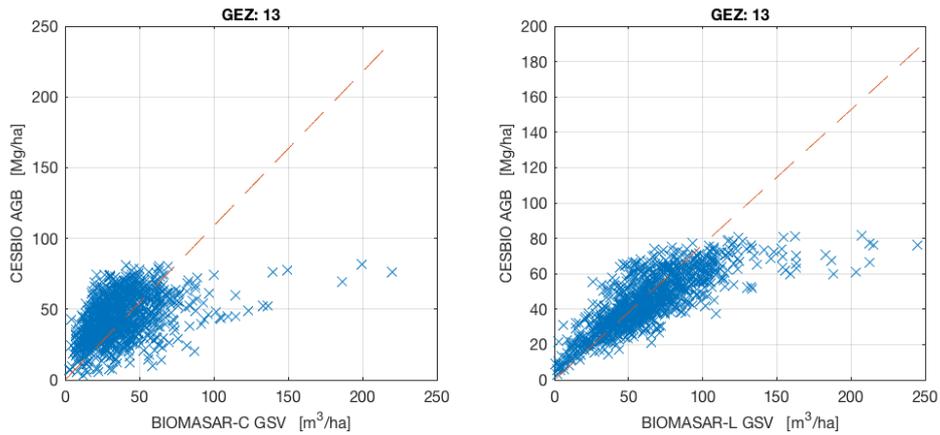


Figure 1-64. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: tropical dry.

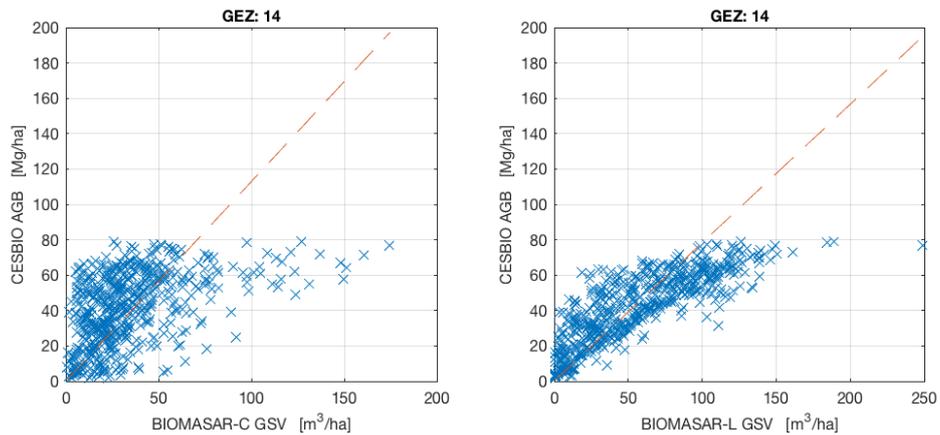


Figure 1-65. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: tropical steppe.

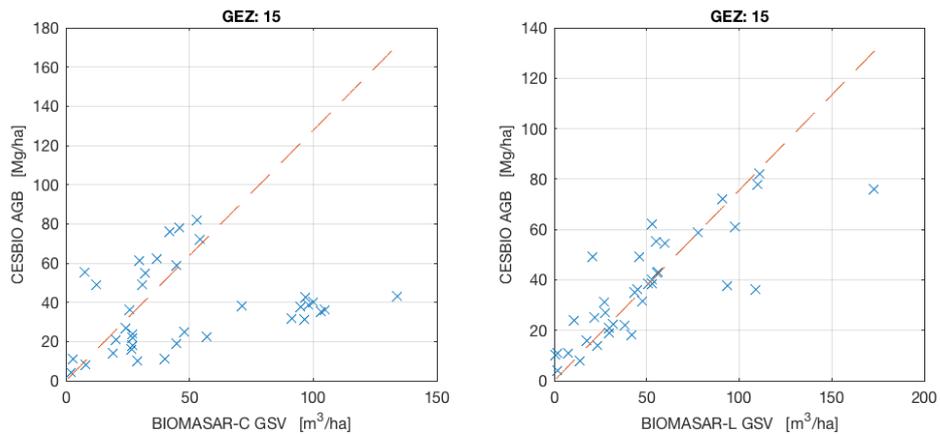


Figure 1-66. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: tropical desert.

For tropical mountain forest (GEZ = 16, Figure 1-67), we observed the same pattern as for other tropical forest types, i.e., agreement in low biomass forest and a tendency to saturation of the AGB values compared to the BIOMASAR GSV values. The red line indicating perfect agreement between AGB and GSV under the assumption that the BCEF reported by the IPCC is correct is far from the observations, in particular in dense forest.

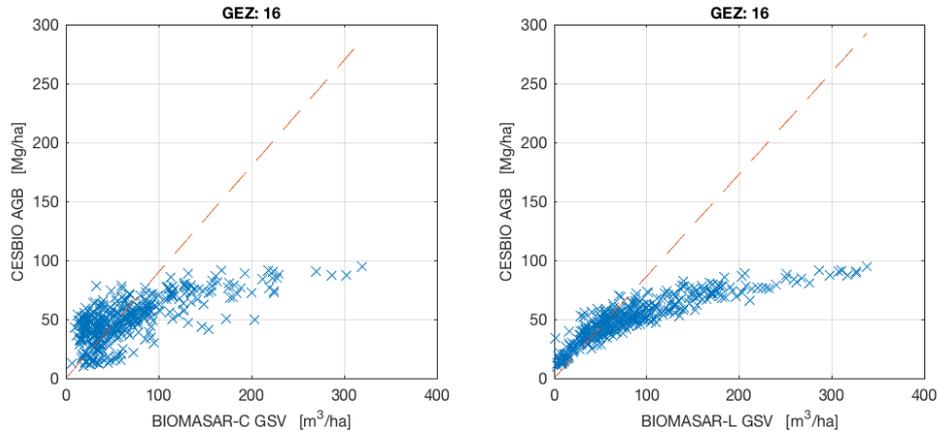


Figure 1-67. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: tropical mountain system.

Also in sub-tropical dry, steppe and mountain forests (GEZ = 22, 23 and 25, respectively), the relationship between CESBIO AGB and BIOMASAR-L estimates showed a tendency for the former to saturate (Figures 1-68, 1-69 and 1-70). Nonetheless, the difference was not as remarkable as in tropical forest. The agreement between BIOMASAR-C and CESBIO was instead rather weak, which we assume to be caused by the patchiness of the land cover (shrubland, grassland, cropland and woody vegetation) in this region.

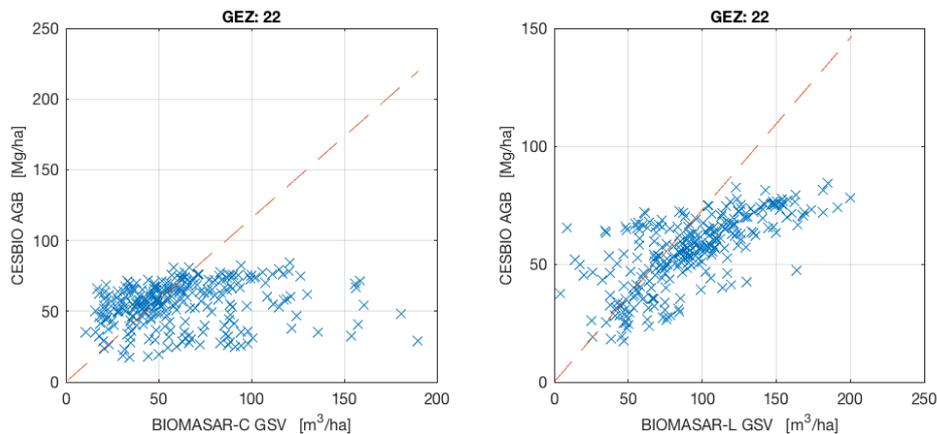


Figure 1-68. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: sub-tropical dry.

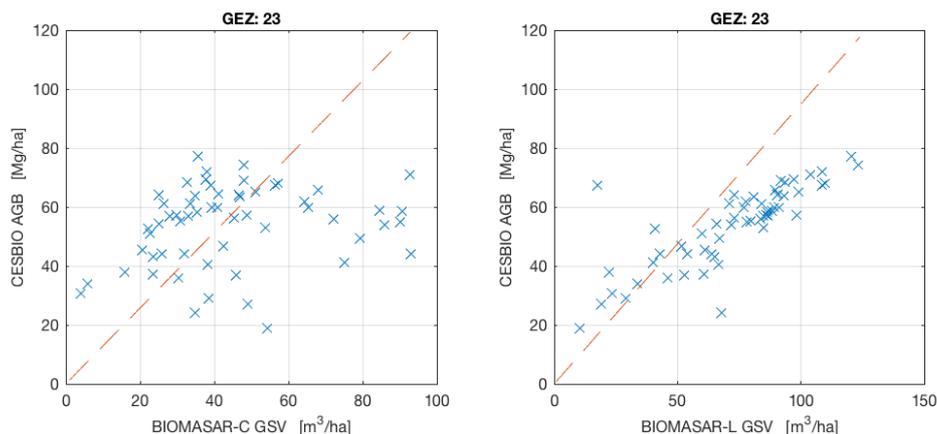


Figure 1-69. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: sub-tropical steppe.

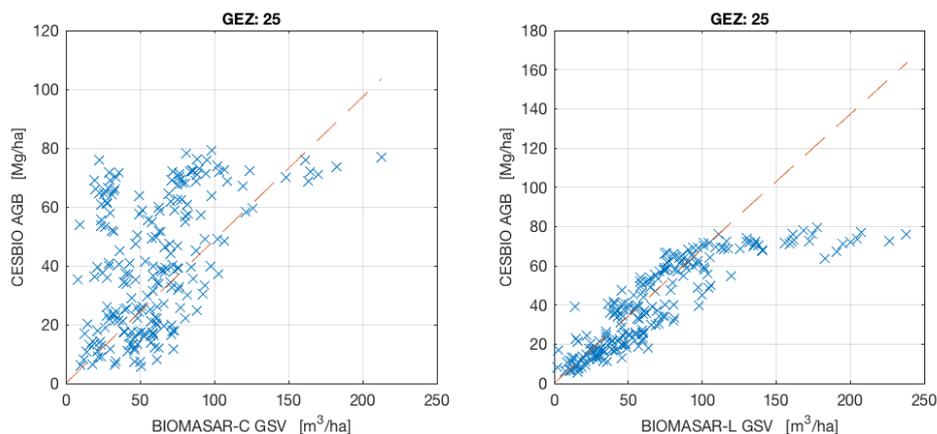


Figure 1-70. Scatterplot of 0.01° averages of AGB obtained with the CESBIO approach in Africa vs. averages of GSV obtained with the BIOMASAR-C (left) and BIOMASAR-L (right) approaches, respectively. Ecological zone: sub-tropical mountain.

It is clearly necessary to compare CESBIO and BIOMASAR data products with in situ observations to clarify the ambiguous results in high biomass forest. It is inevitable that the CESBIO approach will underestimate AGB because of the constraint of mapping AGB below 100 Mg/ha imposed in the retrieval, while the assessment of the BIOMASAR-L values often showed overestimation in very high biomass forest.

From the perspective of generating a global biomass dataset, the inclusion of estimates obtained with the CESBIO approach has potential, but the CESBIO approach is not of sufficient maturity to be included in forming the global dataset because of the issues in high biomass forest (voids or underestimation). For low biomass, it is assumed that complementing the upscaled BIOMASAR-C estimates with the BIOMASAR-L estimates is sufficient. This enables the generation of a GSV dataset, from which AGB can be computed with a conversion factor, derived independently from the remote sensing data used to generate the GSV layer.

	GlobBiomass		Page 79
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

1.4.3.2 Merging biomass estimates

As a result of the investigations reported in this section, we conclude that none of the three datasets represent GSV (and therefore AGB) in a consistent manner globally. Combination of estimates from the GlobBiomass algorithms is necessary.

Combination of estimates has been proposed in the past (Avitabile et al., 2015) and is applied for the regional exercise in Sweden. However, we believe that such approaches are not viable in this case because both require a spatially well-distributed set of reference data. Although the ground reference database generated within GlobBiomass includes a large number of plot measurements, their distribution is uneven. Hence areas not well covered by in situ measurements might be scaled differently than if appropriate measurements had been available.

In general, the purpose of integrating datasets is to reduce the error and increase the accuracy of each estimate. The simplest approach is to take the average. This is, however, useful only when the estimates being combined do not have biases, which is not true in our case. A more suitable approach is to consider a weighted average of the kind proposed in the multi-temporal combination of BIOMASAR-C when combining single-image GSVs. Here, the weighted average would be applied to two values only, from BIOMASAR-L and BIOMASAR-C+.

An elegant way of defining the weights is to use a measure of how well estimates and reference values agree. If both estimates of GSV have the same error the weights would be similar. If one of the estimates is characterized by a larger error, the weight associated with this estimate would be smaller. The definition of the weights relies on a measure of the error of the estimate. Quantifying the error at the level of a single pixel is likely to generate weights with strong fluctuations in space because the estimates of GSV have large uncertainties and they typically do not match with the biomass estimated within a plot at the spatial resolution of the BIOMASAR-L and BIOMASAR-C+ maps (25 m). In addition, the weights could only be defined in areas where reference data are available, causing the weights to be highly uncertain in other areas.

Guided by the results of the map inter-comparison in the previous section, we therefore implement a weighting scheme that does not rely on the availability of GSV/AGB reference data and accounts for:

- 1) differences in the sensitivity of C- and L-band data to GSV
- 2) the number of observations used for estimating GSV
- 3) local errors in the model calibration and inversion, which for both BIOMASAR-C(+) and BIOMASAR-L, are carried out locally adaptive
- 4) topographic effects in the ALOS PALSAR mosaic.

A weighting scheme focussing on full resolution, i.e., weights that are defined at the 25 m pixel size of the maps, entails the risk of strong fluctuations in the weights between adjacent pixels and the generation of artefacts. With the exception of topographic effects, weights are therefore produced at a scale of 0.1x0.1 degree and then oversampled to the full resolution of the GSV maps. In that way the weighting is more geared towards identification of systematic “regional” differences in the GSV maps.

C- and L-band sensitivity to GSV

The modelling of C- and L-band backscatter suggests that the respective sensitivity to GSV changes with increasing GSV. In the lower ranges of GSV, we expect a higher sensitivity at C-band, whereas in

the higher ranges, L-band is expected to present better sensitivity. In the semi-empirical model used in BIOMASAR-C and BIOMASAR-L, this is reflected in the forest transmissivity, which is considered a function of the canopy density, η , the canopy height, h , and the two-way attenuation, α :

$$T_{for} = 1 - \eta + \eta(1 - \exp(-\alpha h)) \quad (1-16)$$

and approximated as function of GSV with:

$$T_{for} = \exp(-\beta V) \quad (1-17)$$

Because of stronger attenuation of the C-band signal in the canopy, and thus a faster decrease of the transmissivity with increasing GSV, higher values for the empirical transmissivity coefficient, β , are generally required to describe the transmissivity as function of GSV at C-band than at L-band. This is exemplified in Figure 1-71 which shows the transmissivity as a function of GSV for two different values of β (left); the lower value represents L-band and the higher value C-band. The derivative of Equation (1-17), which may be considered an indicator of the sensitivity of the signals to GSV, is therefore more negative for the higher value of β in the lower ranges of GSV. In the higher ranges of GSV, instead, the derivative is higher for the lower values of β .

A simple weighting scheme that reflects the difference in sensitivity between C- and L-band may therefore be defined by calculating the difference in the derivatives of Equation (1-17):

$$w_s = \left| \frac{\partial T_L}{\partial V} \right| - \left| \frac{\partial T_C}{\partial V} \right| \quad (1-18)$$

where the transmissivities at L- and C-band, T_L and T_C , are determined using: 1) ecoregion-specific values for β that were presented in previous sections for the retrieval with C- and L-band data, and 2) the 1km BIOMASAR-C GSV map as reference GSV. Initial validation demonstrated that among the three maps that were produced (BIOMASAR-C, BIOMASAR-C+, BIOMASAR-L), BIOMASAR-C best depicted the distribution of GSV globally and may well serve such a purpose.

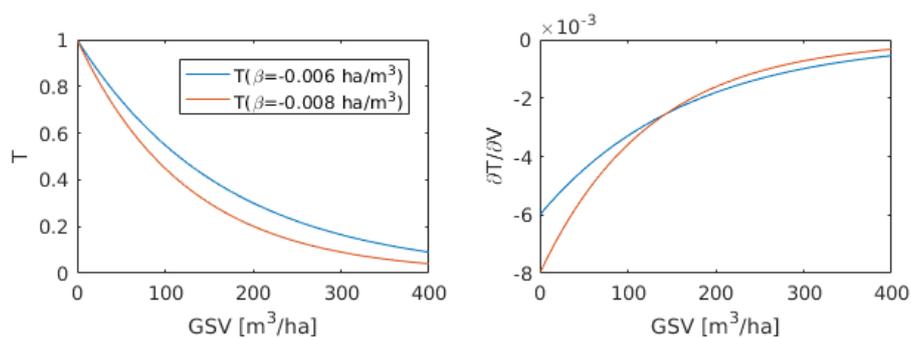


Figure 1-71. Forest transmissivity modelled as a function of GSV for two different values of the forest transmissivity coefficient, β (left). Derivative of Equation (1-17) for the two values of β (right).

Number of observations

The performance of the retrieval of GSV with C- or L-band imagery depends on the number of backscatter observations available. In the case of BIOMASAR-L, a single L-band cross-polarization backscatter observation (i.e., one observation per pixel location) was used to estimate GSV globally, which is a significant limitation on the performance of the approach, since locally adverse imaging conditions or a failed model calibration cannot be compensated by combining estimates from a

multi-temporal stack of observations. For BIOMASAR-C, instead, a large number of images was available at each pixel location, with some regional exceptions. The number of images used in BIOMASAR-L, N_L , and BIOMASAR-C(+), N_C , is therefore considered in the weighting by scaling the weight defined in (1-18) with the square root of the number of observations available at C- and L-band at a given pixel location:

$$w_s = \left| \frac{\partial T_L}{\partial V} \sqrt{N_L} \right| - \left| \frac{\partial T_C}{\partial V} \sqrt{N_C} \right| \quad (1-19)$$

This weight can then be rescaled to the range 0 to 1 to obtain normalized weights:

$$w_{s, norm} = (w_s - \min(w_s)) / \max(w_s - \min(w_s)) \quad (1-20)$$

The rescaling is done separately for each FAO ecoregion. Given the way Equations (1-19) and (1-20) are defined, w_s will be high in higher ranges of GSV and low in lower ranges of GSV (see Figure 1-72), i.e., $w_{s, norm}$ serves as a weight for BIOMASAR-L. Since only two maps are to be combined, the corresponding weight for BIOMASAR-C+ can simply be defined as $1-w_{s, norm}$. Depending on the number of images used in the retrieval of the original BIOMASAR-C map, N_C , and the FAO ecoregion, a weight of 0.5 (when BIOMASAR-L and BIOMASAR-C+ are simply averaged) is applied for GSVs of the order of $\sim 100 \text{ m}^3/\text{ha}$. Figure 1-72 (bottom) illustrates the weight for BIOMASAR-L in form of a map at 0.1° pixel size.

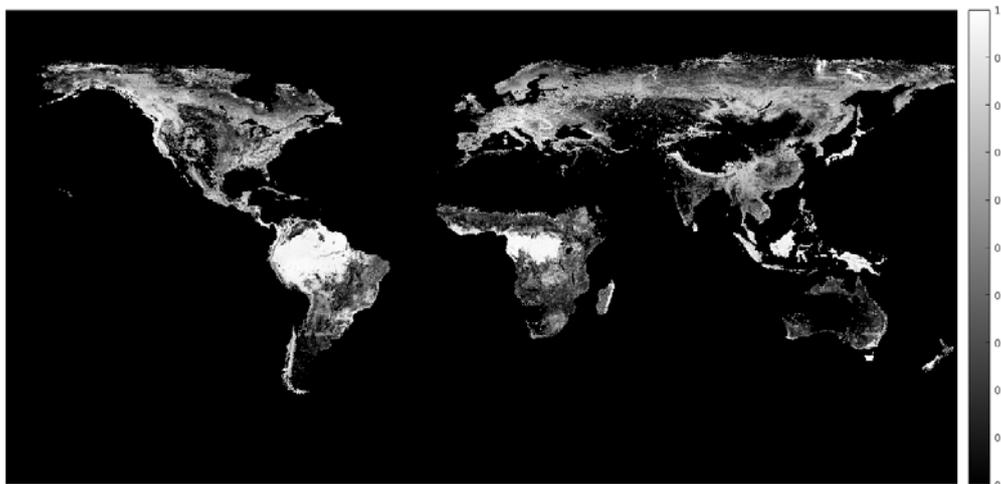
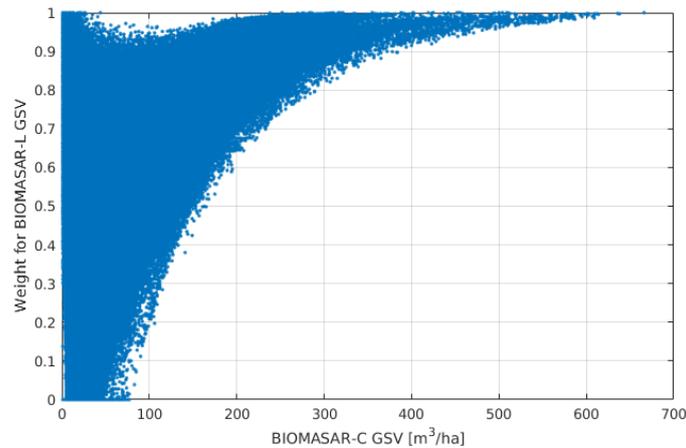


Figure 1-72. Weight for BIOMASAR-L defined with Equations (1-19) and (1-20) as a function of BIOMASAR-C GSV at 0.1° resolution (top); map of weights for BIOMASAR-L (bottom).

Local errors in the model calibration and inversion

Visual inspection of the GSV maps produced using BIOMASAR-L and BIOMASAR-C+ and comparisons with regional reference datasets (inventory, maps) revealed local discrepancies between the maps that could not be explained by differences in the sensitivity of C- and L-band to GSV, but with locally failed model calibration and inversion in either of the mapping approaches. We therefore implemented an additional weighting scheme that identifies and mitigates such effects.

One of the assumptions in the modelling of C- and L-band backscatter as a function of GSV is that there is a well-defined exponential relationship between the forest transmissivity and GSV, Equation (1-17). We thus define an additional weighting scheme that evaluates if the BIOMASAR-C+ and BIOMASAR-L maps are consistent with the underlying transmissivity concept on a per-ecoregion basis with the aid of globally available canopy density (Hansen et al., 2013) and height (Simard et al., 2011) optical maps.

The forest transmissivity at C- and L-band can be simulated as a function of canopy density and GLAS/MODIS forest height (Simard et al., 2011) based on Equation (1-16), assuming a two-way tree attenuation of 0.5 dB/m for L- and 2 dB/m for C-band. When plotting the simulated transmissivity at C- and L-band against the BIOMASAR-C+ and BIOMASAR-L maps respectively (Figure 1-73) and fitting Equation (1-17) to the observed trend (i.e., by estimating β) between (a) the simulated C-band transmissivity and BIOMASAR-C+ GSV and (b) the trend between the simulated L-band transmissivity and BIOMASAR-L GSV, we can evaluate if the GSV estimates are consistent with the expected modelled trend of transmissivity as a function of GSV. In the examples shown in Fig. 1-73, we see, for instance, that in subtropical dry forests in South America, the observed trend of simulated transmissivity at C-band and the BIOMASAR-C+ GSV estimates (green error bars) deviates less from the fit of Equation (1-17) (green curve) than the observed trend of simulated transmissivity at L-band and the BIOMASAR-L GSV estimates (red error bars) (red curve), in particular in higher GSV ranges. Although errors in the optical canopy density and forest height maps used to simulate the transmissivity may have an influence, we take the fact that one of the maps is less consistent with the assumed relationship between transmissivity and GSV as indicating that, locally, the more consistent map should be preferred, i.e., should be given more weight in the map merging.

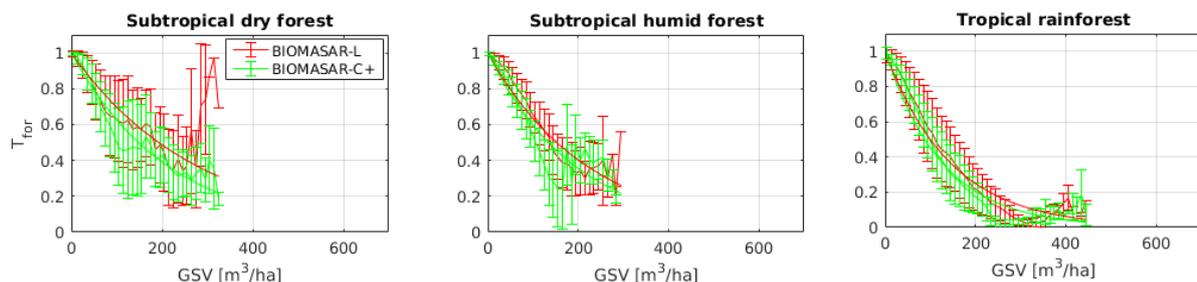


Figure 1-73. Forest transmissivity modelled as a function of Landsat canopy cover, GLAS/MODIS forest height vs. BIOMASAR-C+ and BIOMASAR-L GSV for different FAO ecoregions in South America. The curves represent the fit of Equation (1-17) to the observations.

The scheme for evaluating if the observed trend of GSV in the maps is consistent with the transmissivity concept and to create an additional weight for merging the BIOMASAR-C+ and BIOMASAR-L maps is:

	GlobBiomass		Page 83
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- 1) Simulate the transmissivity at C- and L-band, $T_{sim}(C)$ and $T_{sim}(L)$, using global maps of canopy density and height in Equation (1-16).
- 2) Fit Equation (1-17) to the observed trend between the simulated C-band transmissivity and BIOMASAR-C+ GSV as well as the trend between the simulated L-band transmissivity and BIOMASAR-L GSV on a per ecoregion basis. The resulting fit allows the creation of new maps of the transmissivity at C- and L-band, $T_{mod}(C)$ and $T_{mod}(L)$, by applying the fitted models to the BIOMASAR-C+ and BIOMASAR-L maps.
- 3) Weights are then created by evaluating at pixel level for which map (i.e., BIOMASAR-C+ or BIOMASAR-L) the difference between T_{sim} and T_{mod} is smaller:

$$w_t = \left| T_{sim}(C) - T_{mod}(C) \right| - \left| T_{sim}(L) - T_{mod}(L) \right| \quad (1-21)$$

and normalizing w_t to the range 0 to 1 for each ecoregion separately with:

$$w_{t,norm} = (w_t - \min(w_t)) / \max((w_t - \min(w_t))) \quad (1-22)$$

Given how Equation (1-21) has been defined, the weights will be high where the BIOMASAR-L map is more consistent. Again, weights for BIOMASAR-C+ may simply be defined with $1-w_t$.

The resulting map of w_t is illustrated in Figure 1-74. The map shows that for large parts of the world, the weights are ~ 0.5 . Only very locally will one of the two maps be given more weight. For some of the areas for which the weights differ significantly from 0.5 (e.g., Central British Columbia and the Russian Far East), we found evidence that the map which is assigned the lower weight actually exhibits systematic biases in the GSV estimates. For other areas with weights significantly different from 0.5, such evidence was not possible to obtain and the benefit of applying the proposed weighting remains uncertain.

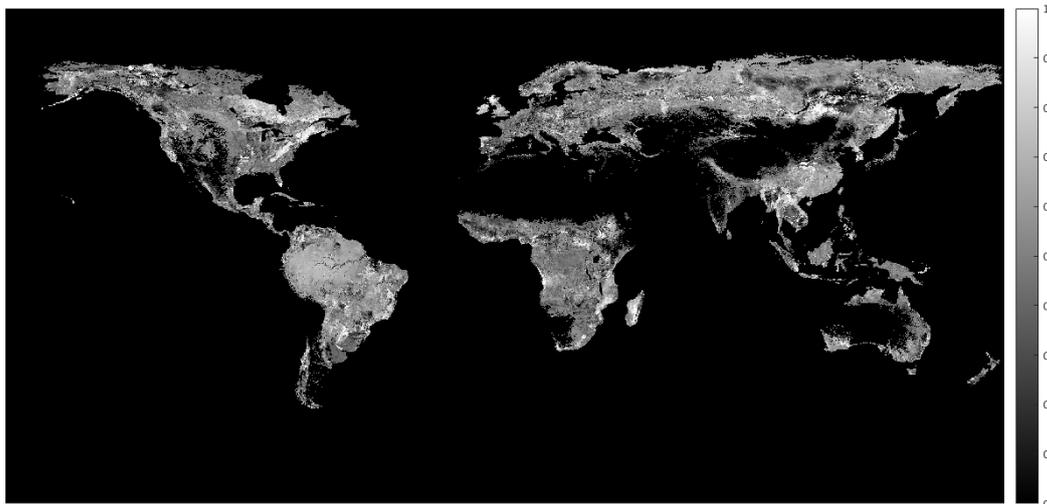


Figure 1-74. Weight for BIOMASAR-L which reflects how consistent the GSV estimates are with the transmissivity modelled with the aid of optical canopy density and height maps.

Topographic effects

The BIOMASAR-L map was produced using only the single observation (~ 2010) ALOS PALSAR mosaic. BIOMASAR-C+, instead, also considered Landsat reflectance mosaics for mapping GSV. A drawback of

	GlobBiomass		Page 84
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

working with JAXA's ALOS PALSAR mosaic is that the mosaic exhibits strong topography-related distortions, geometric as well as radiometric, which could not be reduced/corrected without re-processing the original PALSAR SLC data. Visual comparison of the BIOMASAR-L and BIOMASAR-C+ GSV maps revealed that, because it was created also from Landsat imagery with less pronounced topography-related errors, the BIOMASAR-C+ map was less affected by topography than the BIOMASAR-L map. In order to reduce the effect of topography, an additional weight was defined that successively gives more weight to the BIOMASAR-C+ map the steeper the terrain is.

Topographic weights were defined using the local incidence angle map provided by JAXA together with the backscatter mosaics. It was assumed that radiometric terrain effects in the mosaics are primarily due to inaccurate estimates of the pixel scattering area (Ulander, 1996; Small, 2011), which scales approximately with:

$$A \sim 1 / \sin(\theta_i) \quad (1-23)$$

Radiometric variations due to changes in azimuth slope (Ulander, 1996) as well as the dependence of backscatter on the local incidence angle are here considered minor and not considered. Weights were defined by calculating the percent difference in pixel scattering area between flat terrain (38° incidence angle) and the pixel area estimated from the local incidence angle with:

$$\Delta_{area} = 100 \cdot |1 - \sin(38^\circ) / \sin(\theta_i)| \quad (1-24)$$

Weights, w_{topo} , for the BIOMASAR-L map were then defined by linearly scaling Δ_{area} from 0 to 1 with 1 representing flat terrain (38° incidence angle) and 0 reflecting Δ_{area} values beyond 30%, i.e., BIOMASAR-L will be given little weight in areas of steep terrain. The weight map is shown in Figure 1-75.

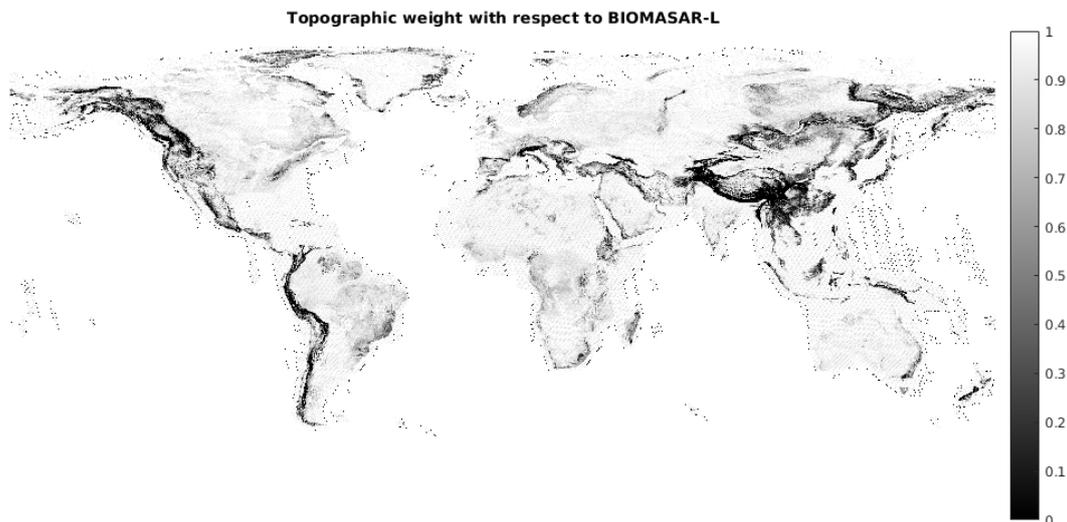


Figure 1-75. Weight for the BIOMASAR-L map reflecting the steepness of the terrain.

Final weight calculation

The three different calculated weights need to be combined into one. For BIOMASAR-L and BIOMASAR-C+, the final normalized weights were given by:

$$w(L) = 0.5 * (w_s + w_t) * w_{topo}$$

$$w(C) = 1 - w(L) \quad (1-25)$$

For BIOMASAR-L w_s and w_t were simply averaged, i.e., considered equally important. In contrast, w_{topo} was applied as a factor that lowers the weight for BIOMASAR-L everywhere over steep terrain.

The final weight maps are shown below in Figure 1-76. The maps clearly show that BIOMASAR-L is given more weight in areas of high GSV, unless there is steep topography. BIOMASAR-C+ will be given more weight in areas of low GSV and steep terrain.

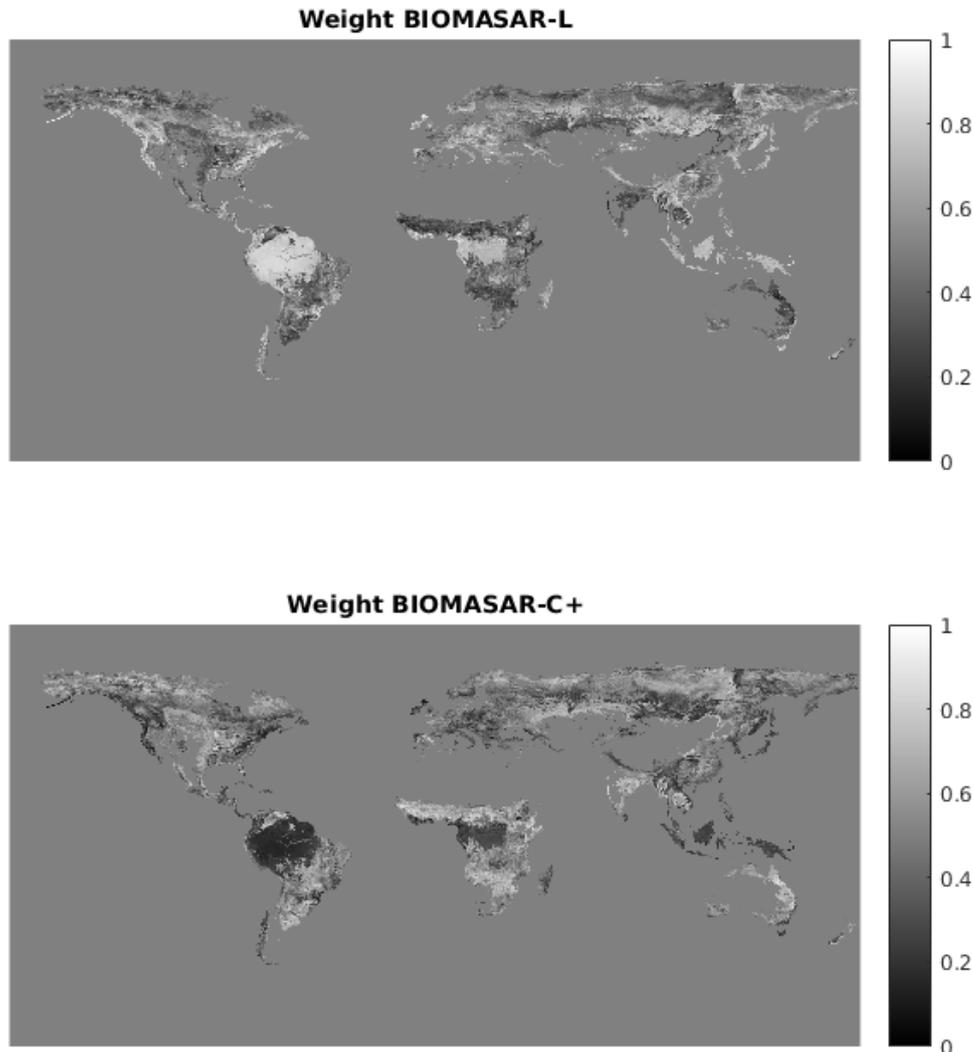


Figure 1-76. Weights for the BIOMASAR-L and BIOMASAR-C+ maps used in producing the final merged map.

1.4.4 Conversion of GSV to AGB

The conversion of GSV to AGB requires knowledge on both wood density and the allometric relationship between stem mass and whole above ground mass, including branches and leaves. Both wood density and allometric relationships result from biological processes that respond strongly to variations in environmental conditions. As such, there is a strong variation of wood density between species that share the same environmental space, but also within species that differ in their

geographic locations, and consequently experience different environmental conditions during plant growth (see e.g., Figure 1-77). Furthermore, the responses to changes in environmental conditions vary with species plasticity. Previously, regional studies have shown evidence of environmental controls on wood density [Chave *et al.*, 2009; Thurner *et al.*, 2014] and allometric relationships [Thurner *et al.*, 2014]. The species is systematically observed as the main controlling factor. However, given reduced diversity in species dominance, the aggregations to leaf habits/family still provide an appropriate conversion strategy to derive biomass as Thurner *et al.* [2014] demonstrated for the Northern Hemisphere (>30°N). The problem of Thurner’s approach, e.g., for the tropics is the wide range in variability of wood density, and likely allometric relationships, within family/leaf habits. Here, we seek to extend the previous regional efforts by expanding the databases to be more widely representable, and exploring machine learning methods that explore the information content in other environmental covariates to provide a global picture on the GSV to AGB relationships. In particular, it is necessary to test the contribution of other covariates related to, e.g., climate, topography and soil properties, to explain the spatial variability in wood density and allometry. This section reports on current provisional results with a likely final update until the end of the project duration.

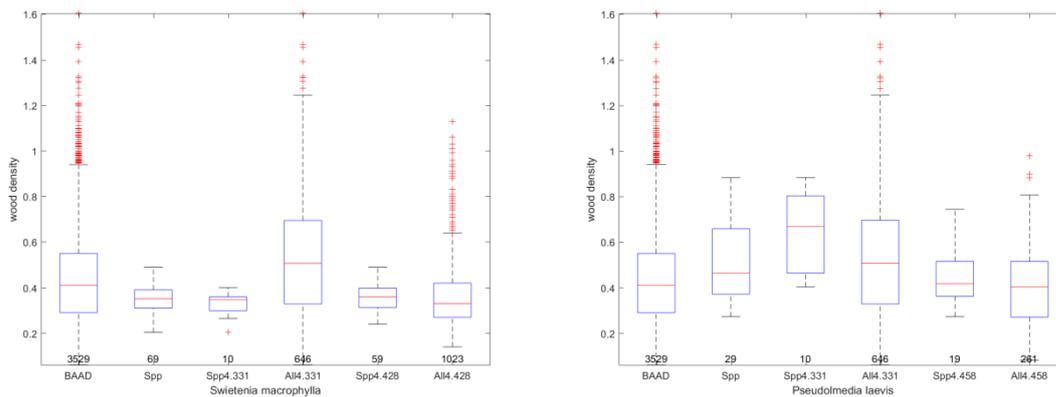


Figure 1-77. Variations in wood density in the BAAD database for two different species *Swietenia macrophylla* (left) and *Pseudolmedia laevis* (right). See that for *S. macrophylla*, the variability within species is quite constrained (second column, label Spp) and for both locations in the database (Spp4.331 and Spp4.428) the distributions in wood density are rather similar, despite a strong variability in both sites between species (All4.331 and All4.428). Conversely, *P. laevis* reveals a strong variation in wood density between sites (see on the right, the boxplots for Spp4.331 is substantially different from Spp4.458) revealing a higher plasticity in the wood density trait when compared to *S. macrophylla*.

Data

The *in situ* data sources for wood density and allometric relationships stand on a set of published databases, or collections of databases [e.g. TRY, Kattge *et al.*, 2011], that are based on inventories where at least one of the variables is reported. For wood density we relied on the following databases: BAAD [Falster, 2015]; TRY [Kattge *et al.*, 2011] and TRY gap-filled [Schrodt *et al.*, 2015]; Philips *et al.* [2003]; and Schepaschenko *et al.* [2017] (see Figure 1-78). For the relationship between

	GlobBiomass		Page 87
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

stem mass and AGB the following databases were explored: BAAD [Falster, 2015]; AFOLU¹; and Schepaschenko *et al.* [2017] (see Figure 1-79). These databases are not necessarily temporally or spatially synchronous, but allow exploring the wood density and allometric relationships to a set of independent covariates that are globally available, and hence support its upscaling.

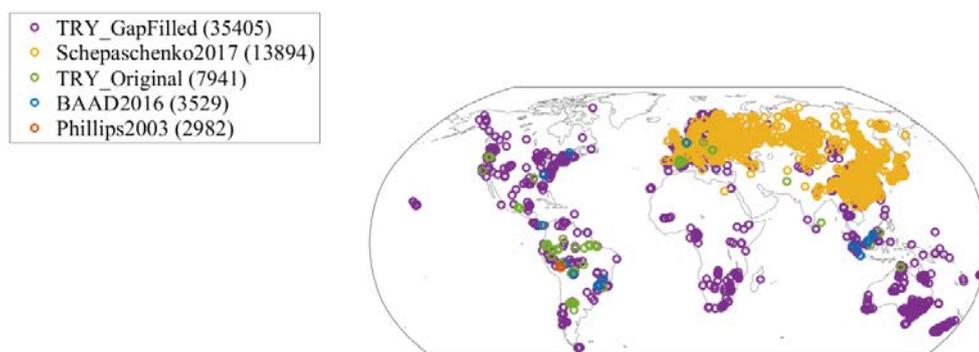


Figure 1-78. Distribution of wood density datasets according to the different sources. Names in parenthesis report the number of records of wood density in each database. Overall, these databases report wood density for a total of 4952 species, showing a median number of 2 records per species.

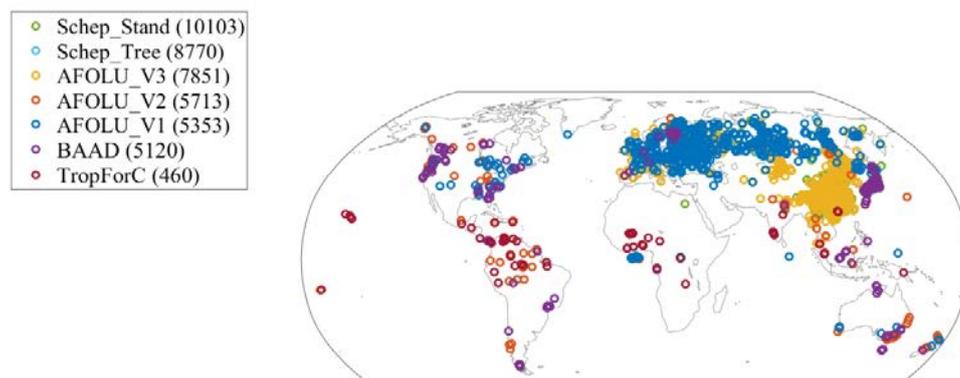


Figure 1-79. Distribution of allometric measurements according to the different sources used to estimate AGB as a function of stem mass and other covariates. Values in parenthesis report the number of records in each database.

For independent covariates, climatological (e.g. WorldClim² [Fick, 2017]), atmospheric [MODIS cloud cover metrics, Wilson and Jetz, 2016] and soil properties [e.g. Harmonized World Soil Database,

¹ ftp://mars.jrc.ec.europa.eu/Afoludata/Public/all_datasets.html

² <http://www.worldclim.org/>

	GlobBiomass		Page 88
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

Batjes et al., 2012] data were extracted from global databases. Further, land surface properties such as vegetation indexes (e.g. NDVI [*Pinzon and Tucker*, 2014], EVI [*Huete et al.*, 2002], FPAR [*Myneni et al.*, 2002]), productivity [*Tramontana et al.*, 2016], vegetation cover [*Sexton et al.*, 2013], height [*Simard et al.*, 2011] and land cover³, were also extracted since these can be informative on the development stage of forests which correlate with resource allocation strategies, hence also with wood density and allometric relationships in forests. Derived statics such as: (i) annual means, standard deviations on spatial variability were computed for the datasets with finer spatial resolution; and (ii) temporal trends, standard deviation and percentiles we estimated for datasets that reported time series.

Methods

Machine learning methods

We used the random forests (RFs) approach here given its ability to generalize and extract nonlinear relationships between target and predictor variables [*Breiman*, 2001]. The application of RFs has been widely used in ecology and biogeochemistry [e.g. *Jung et al.*, 2011; *Tramontana et al.*, 2016] and recently, in the study by *Moreno-Martinez et al.* [2017, in review] a comparison to different machine learning approaches showed the robustness of RFs in predicting the spatial patterns in plant traits using remote sensing observations and climatological data.

In the current exercise, the challenge with including a high number of covariates in such an approach is to select the important variables to explain wood density and allometry in order to maximize the predictive power of the final statistical model. Two approaches are followed: (1) evaluation of the variable importance based on permutation tests for each variable at the time; and (2) a feature selection algorithm [*Jung and Zscheischler*, 2013].

Given that different databases provide different information to the conversion problem, the current strategy relies on evaluating independently how which of the following approaches behave in predicting wood density and allometry independently and computing AGB as the product between these two terms and GSV.

Predicting wood density at tree/plot level

The set of covariates were extracted for every valid record of wood density, yielding a total of approximately 435 potential predictors (excluding MODIS). The feature selection [*Jung and Zscheischler*, 2013] was performed and the final models were trained for the set of covariates yielding the best Nash Sutcliffe Efficiency [*Janssen and Heuberger*, 1995]. This was repeated for a set of two factorial experiments to evaluate the importance of: (1) including or removing the gap-filled records in the TRY database [*Schrodt et al.*, 2015]; and (2) including or not classification information about leaf type and leaf phenology. The first experiment was particularly relevant given that the distribution of wood density data in these global databases had very few observations in Oceania and Africa, and few in the Americas except if the gap-filled TRY database was included in the training set. However, the gap-filled TRY is not completely observation-based and much of its records are estimated based on the trait-to-trait covariation structure [*Schrodt et al.*, 2015]. Despite being strongly driven by species and genus, the variation in wood density hence cannot be fully considered

³ <https://www.esa-landcover-cci.org/>

	GlobBiomass		Page 89
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

an observation and is relevant to evaluate how sensitive are our predictions to the inclusion of such dataset. The second experiment is related to the findings of Thurner *et al.* [Thurner et al., 2014], which also appreciated the strong within PFT variations in wood density. Here, we explore how much does PFT, translated by leaf type and phenology, contributes to the prediction of wood density, in addition to environmental factors expressed in climatological and soil property conditions, and to vegetation productivity and biophysical variables.

Predicting wood density at landscape scale

Moreno-Martinez *et al.* [2017, in review] propose a different way of looking at the prediction of global spatial patterns of wood density. In short, they propose a landscape level wood density estimated as $WD_j = \sum_{i=1}^M WD_i \cdot f_{j,i}$, where the wood density in a given MODIS gridcell j at 500m resolution (WD_j) is estimated as the weighted average of all the WD values representing the abundance of different PFTs⁴ (i) in that same gridcell j ($f_{j,i}$). Grid cells are selected based on a high resolution land cover map (30 m resolution), and WD_j computed based on in situ close observations to the corresponding MODIS pixel taken with a maximum 100km range distance. Grid cells were only used if more than 50% of the PFT composition was represented by nearby WD observations [see Moreno-Martinez *et al.*, 2017, in review for details]. As above, this approach also allows including and removing the gap-filled records in the TRY database to test for their statistical importance in predicting wood density. This resulted in a dataset for training with 5000 data points (grid cells) for the gap-filled example, and 4000 data points (grid cells) for the non-gap-filled dataset.

Stem mass to Above Ground Biomass

The correlation between stem mass and AGB surpasses 97% (Pearson squared correlation) for the datasets used here (Figure 1-80), translating that most of the AGB variations emerge from variations in stem mass, and that once stem mass is known the determination of AGB should be straightforward. However, given the importance of the plant functional type and of local conditions in shaping the AGB-to-stem relationship, we followed a similar approach and predicted AGB based on stem mass and a selected set of covariates.

⁴ 1) Evergreen Needleleaf forest, 2) Evergreen Broadleaf forest, 3) Deciduous Needleleaf forest, 4) Deciduous Broadleaf forest and 6) Shrublands

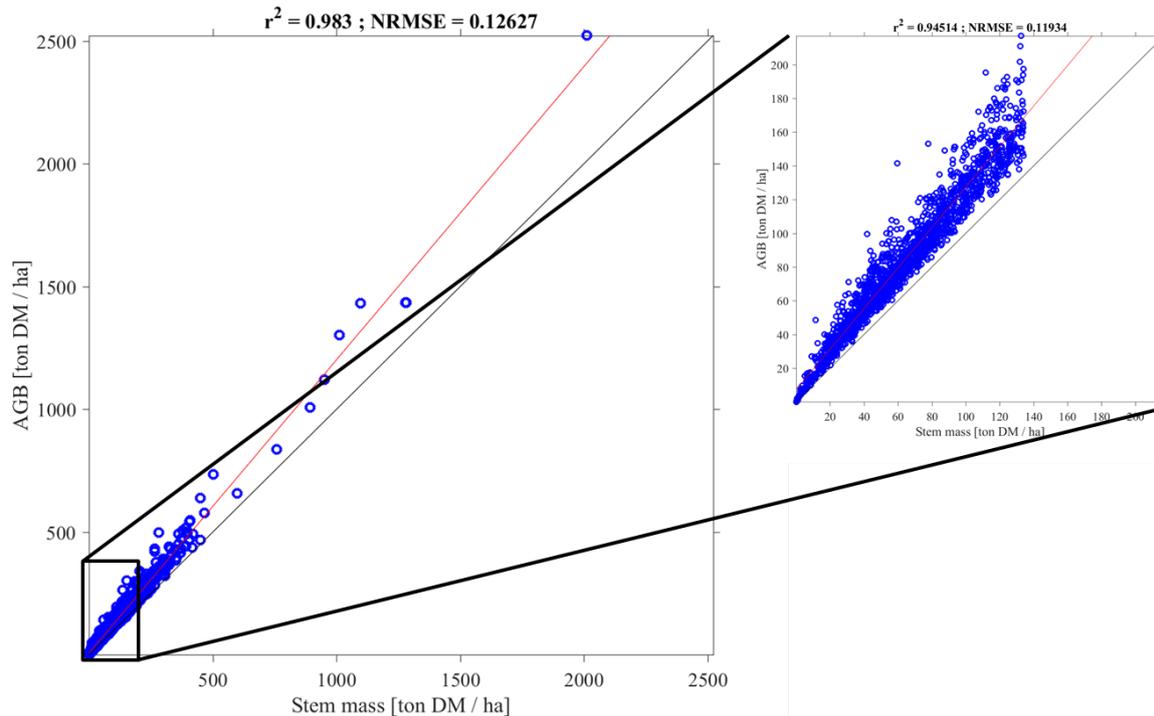


Figure 1-80. Relationship between stem mass and aboveground biomass.

Results

Wood density

Overall, throughout the different experiments, the tree/plot level results show a substantial fraction of unexplained variance (Figure 1-81) although the predictions show to be unbiased (Figure 1-82). The inclusion of information on leaf properties substantially improves the performance of the RF results in cross validation up to a maximum of ~60% of explained variance when not considering gap-filled data in the training set (compare top to bottom left plots in Figure 1-81). This improvement is consistent with the results of an experiment which, for a reduced set of predictors, introduced vegetation type, family, genus and species information progressively in the training set, yielding NSE values of 0.16, 0.53, 0.68 and 0.73, respectively. The main challenge is that though at site level the species information can be obtained, such is not the case for global prediction purposes. In addition, much of the inability to explain the variability in the observations may stem from the fact that many observations share very proximal locations for the same plant functional types, implying that different plot level estimates will show exactly the same input covariates (categorical and climatological). This means that the variation in the predictors will be lower than the variation in the wood density observations and the overall explanatory power is reduced. This is especially relevant for covariates that exhibit coarser spatial resolutions. The variability in landscape level wood density is well explained using a set of bioclimatic variables (WorldClim⁵) and MODIS NDVI and EVI as covariates (NSE \geq 0.7, Figure 1-83). These results rely on the strength of the assumptions underlying the estimate of a landscape WD based on PFT, which reduce the variability but also the noise in

⁵ <http://www.worldclim.org/>

observables, and increase the representativeness consistency between the predictor variables and wood density. However, in both cases, tree/plot and landscape level, the predictions of wood density seem virtually unbiased (see Figure 1-82 and 1-83). We observe that for a progressive aggregation of wood density observations, the normalized root mean square errors reduce substantially, suggesting a higher reliability in the predictions at coarser scales and the role of natural variability and mismatch in scales between covariates and in situ observations in the unexplained variance.

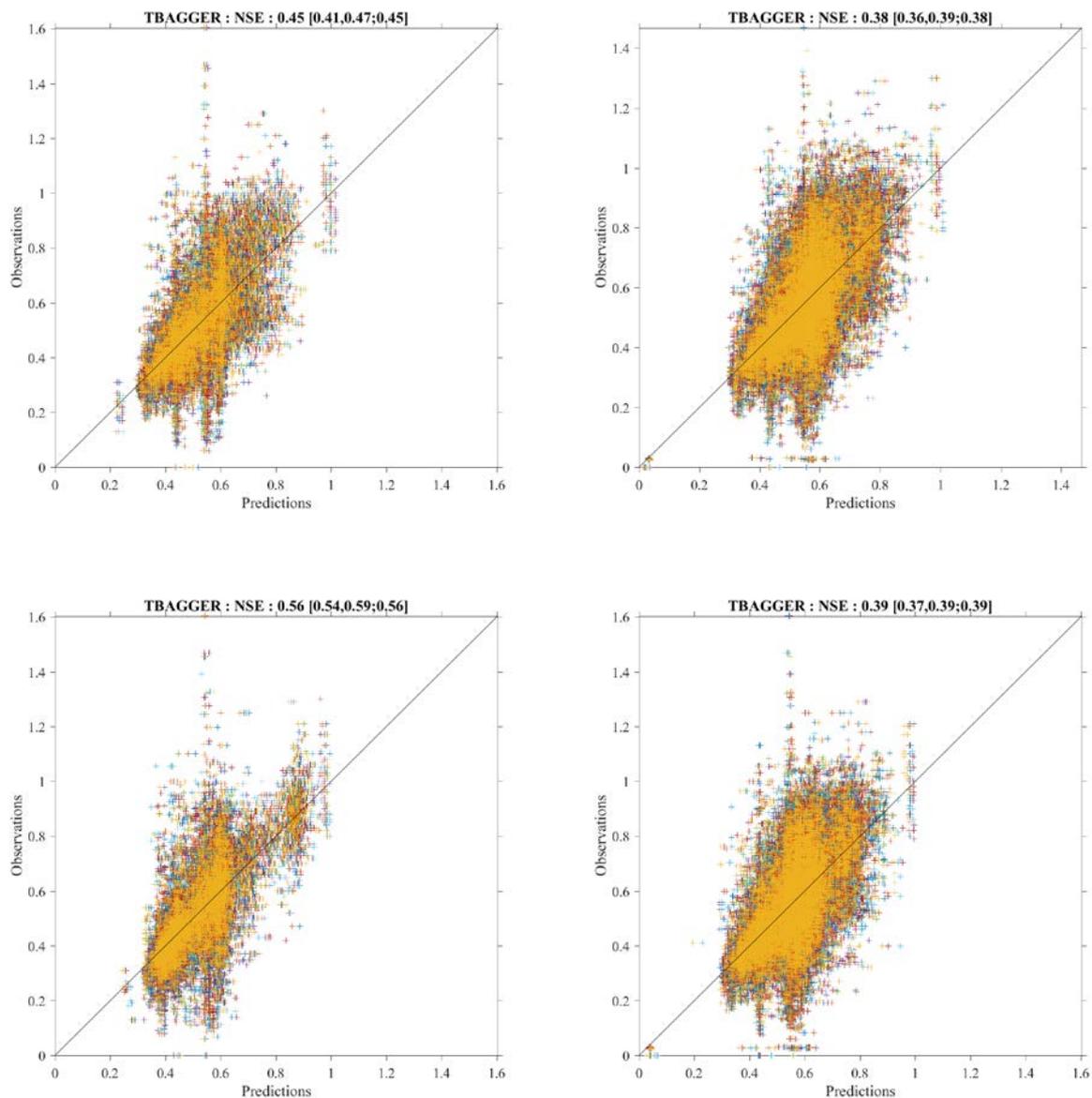


Figure 1-81. Goodness-of-fit of different random forests in cross validation. Every colour represents a random sample removed from the training set (80% to 20% split). On the top the resulting models do not include leaf type/phenology in the potential predictor variables. On the left, the gap-filled data was removed from the training and testing sets. The numbers in the titles report the Nash Sutcliffe Efficiency [Janssen and Heuberger, 1995] and in parenthesis the ranges in NSE and the median for the different cross validation folds tested.

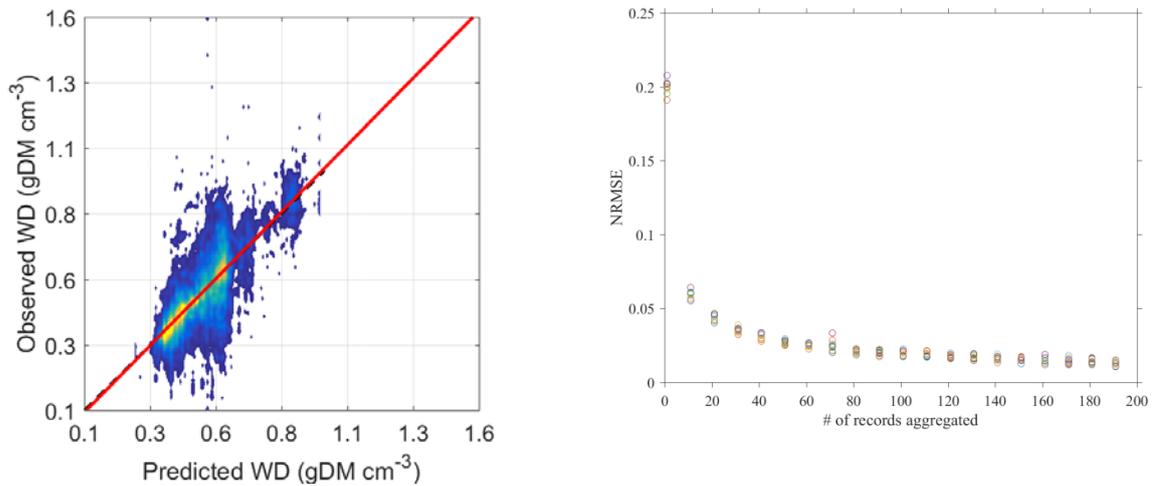


Figure 1-82. Density scatter plot for the best statistical model found (not gap-filled data and using leaf type/phenology as a covariate). The regression line (red) almost overlaps the 1:1 line (dashed black line). On the right, changes in NRMSE with increasing number of aggregated records. This behaviour is observed in all the statistical models that were experimented. All of the above results are on cross-validated predictions.

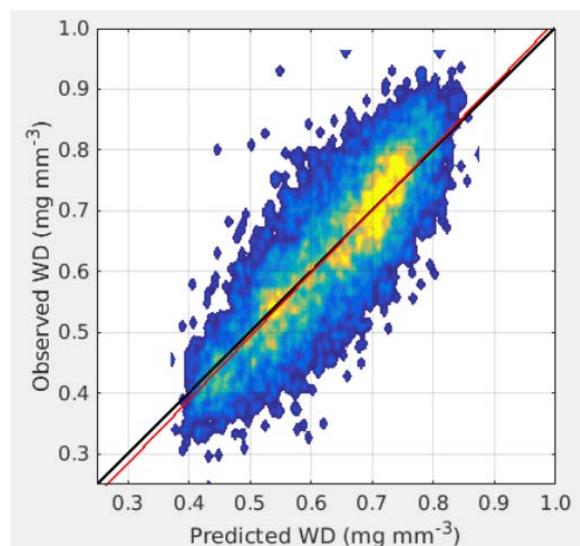
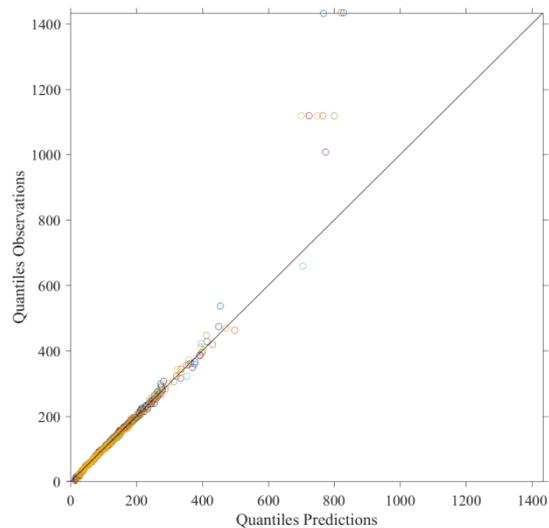
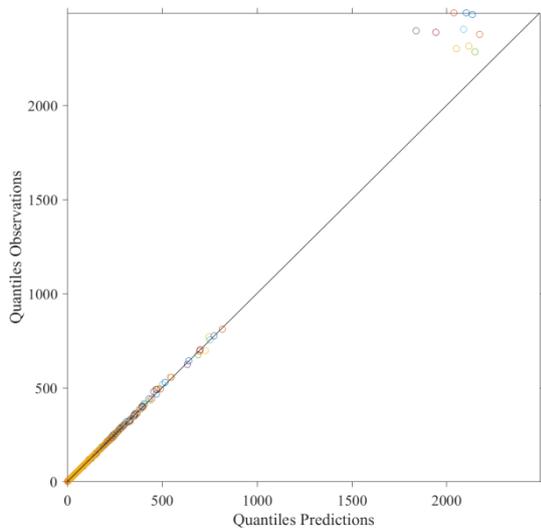
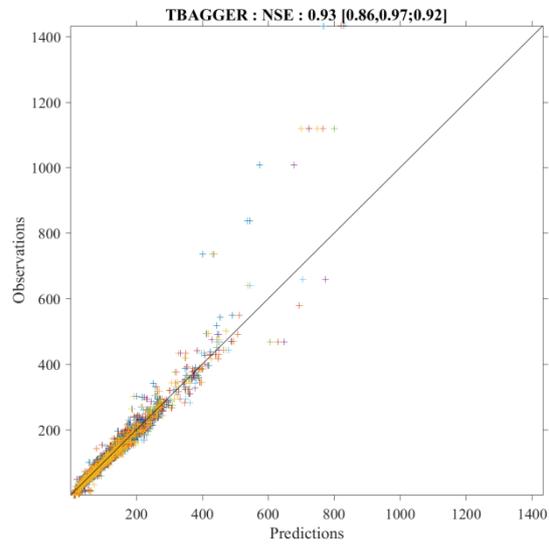
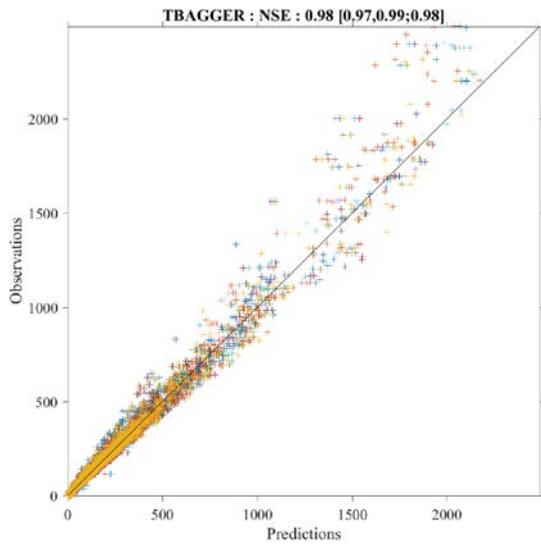


Figure 1-83. Density scatter plot of predictions of landscape scale wood density

Both methods show the same qualitative differences when considering gap-filled data, reflecting a loss in predictive ability (compare left to right columns in Figure 1-81). In this case, it seems that even considering PFT information does not improve significantly the tree/plot wood density predictions. This difference is observed but less substantial in the landscape scale estimates, although, for instance, for the Siberian region, there is a change in correlation (R) from 0.6 to 0.8 by removing the gap-filled data. The reason(s) behind these differences are still unclear and are subject of further investigations.

Stem to aboveground biomass predictions

Estimating AGB as a function of stem mass and an index of water availability (WAI, calculated as the mass balance between precipitation and evapotranspiration) based on a RF model explains more than 90% of the variance in observations globally (Figure 1-84). This is strongly driven by the already high correlation between AGB and stem mass. The ratio between stem to AGB can also be estimated with an NSE up to 63%, but including 10 to 12 predictors including stem, tree cover and ALOS backscatter retrievals, as well as MODIS cloud cover metrics and vegetation indexes, soil properties and bioclimatic variables. For simplicity and traceability, at this point we focus on the first model to predict the AGB from stem mass.



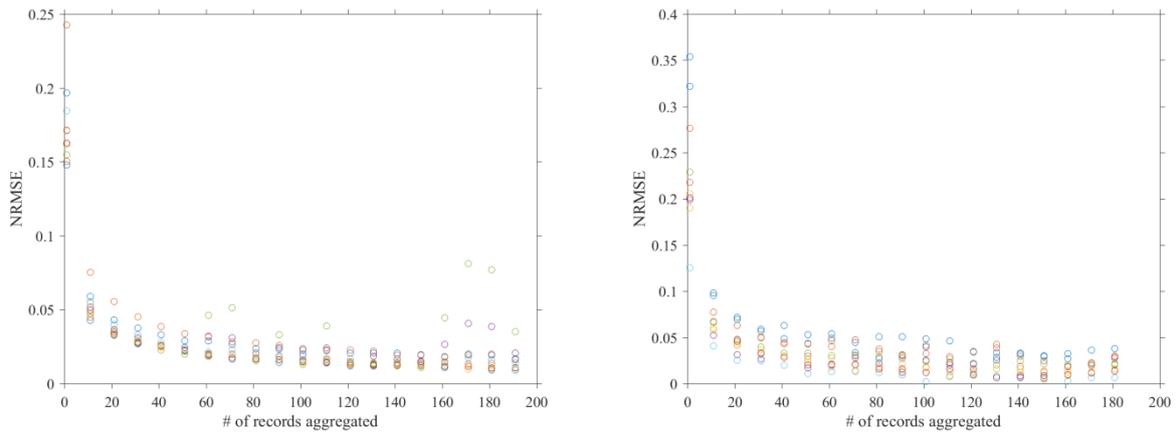


Figure 1-84. Cross validation results for training the stem to AGB conversions for all data points (left) averaging of similar locations (right) in terms of scatter (top rows) but also reporting the match between distributions (QQ plots on the 2nd row) and the effect of aggregation on AGB estimates.

Global distribution of wood density and AGB estimates

Based on the trained RFs, and using spatially distributed predictors, we estimate the global distributions of wood density and of AGB (at 1km²), based on stem mass estimated as the product between wood density and growing stock volume, GSV. The global means differ little between approaches, and the relative bias (RB) between them is also low (ratio close to 1). However, the global correlations are usually statistically significant, but low, reflecting that the patterns in wood density can be very different depending: on the approach (tree/plot and landscape level); on the inclusion, or not, of plant functional type; and by considering (or not) gap-filled data in the training of the different algorithms (see Figure 1-85 and 1-86). The spread in wood density estimates between the different approaches can be substantial (Figure 1-86). The inclusion of gap-filled data reduces the spread amongst approaches in 64% of the considered land area, and in these regions the spread amongst approaches is reduced by 57%. Globally, this translates into a reduction in the wood density ranges between approaches of ~30%.

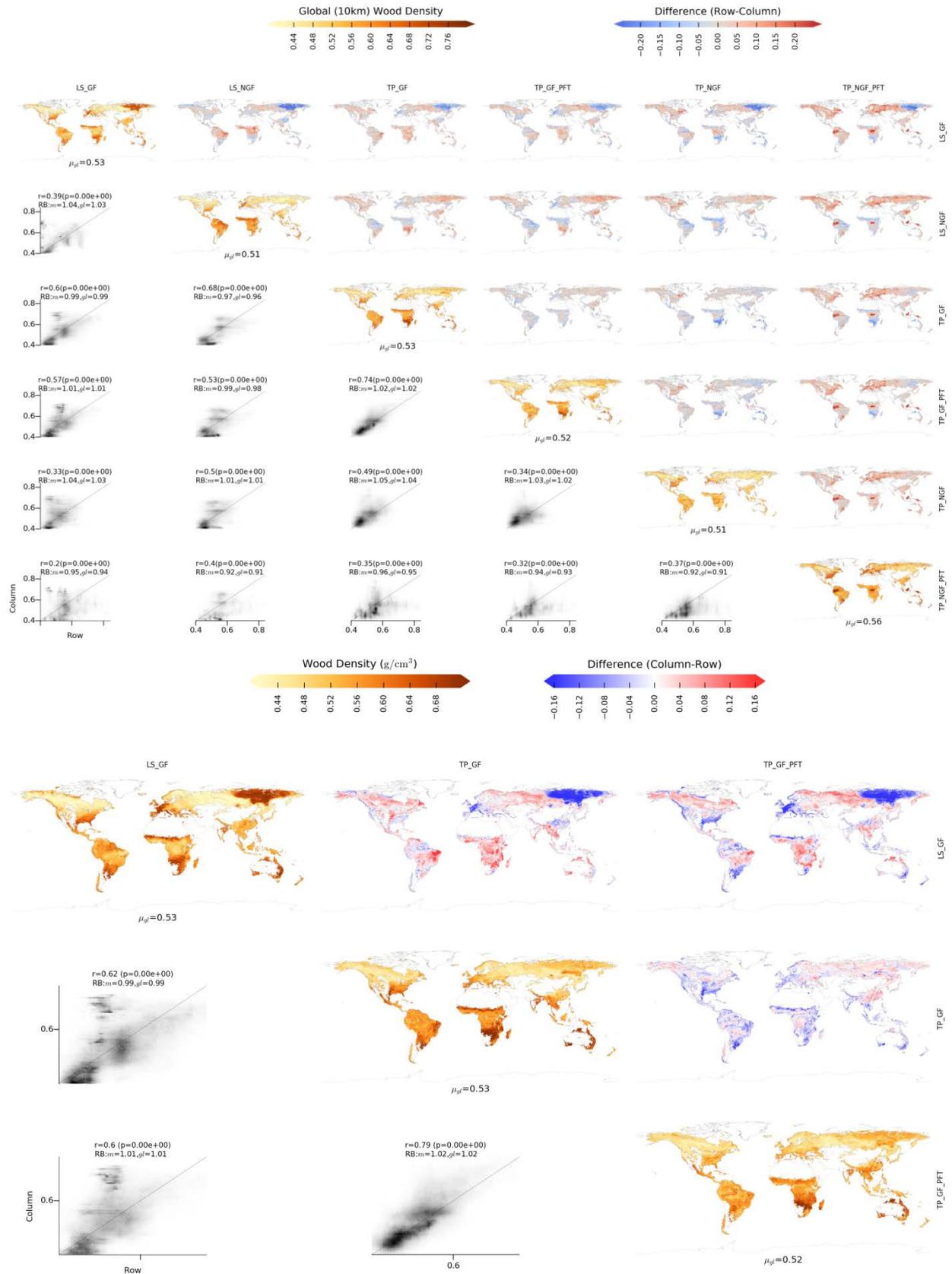


Figure 1-85. Comparison between the different methods used to estimate global wood density using the gap-filled datasets of in situ observations. The methods include landscape scale (LS_GF) and

tree/plot scale including, or not, PFT information (TP_GF_PFT and TP_GF , respectively). The maps in the diagonals report the the wood density, while the off-diagonal maps show the differences between approaches and the scatter plots, reporting Pearson correlations (r) and significance level (p), relative bias (RB , as the ratio between the global mean of two maps ($RB:gl$) or the median of the RB per gridcell ($RB:m$)). The global mean of wood density is reported under each map (μ_{gl}). Gridcells with GSV lower than $2m^3/ha$ are excluded from this comparison.

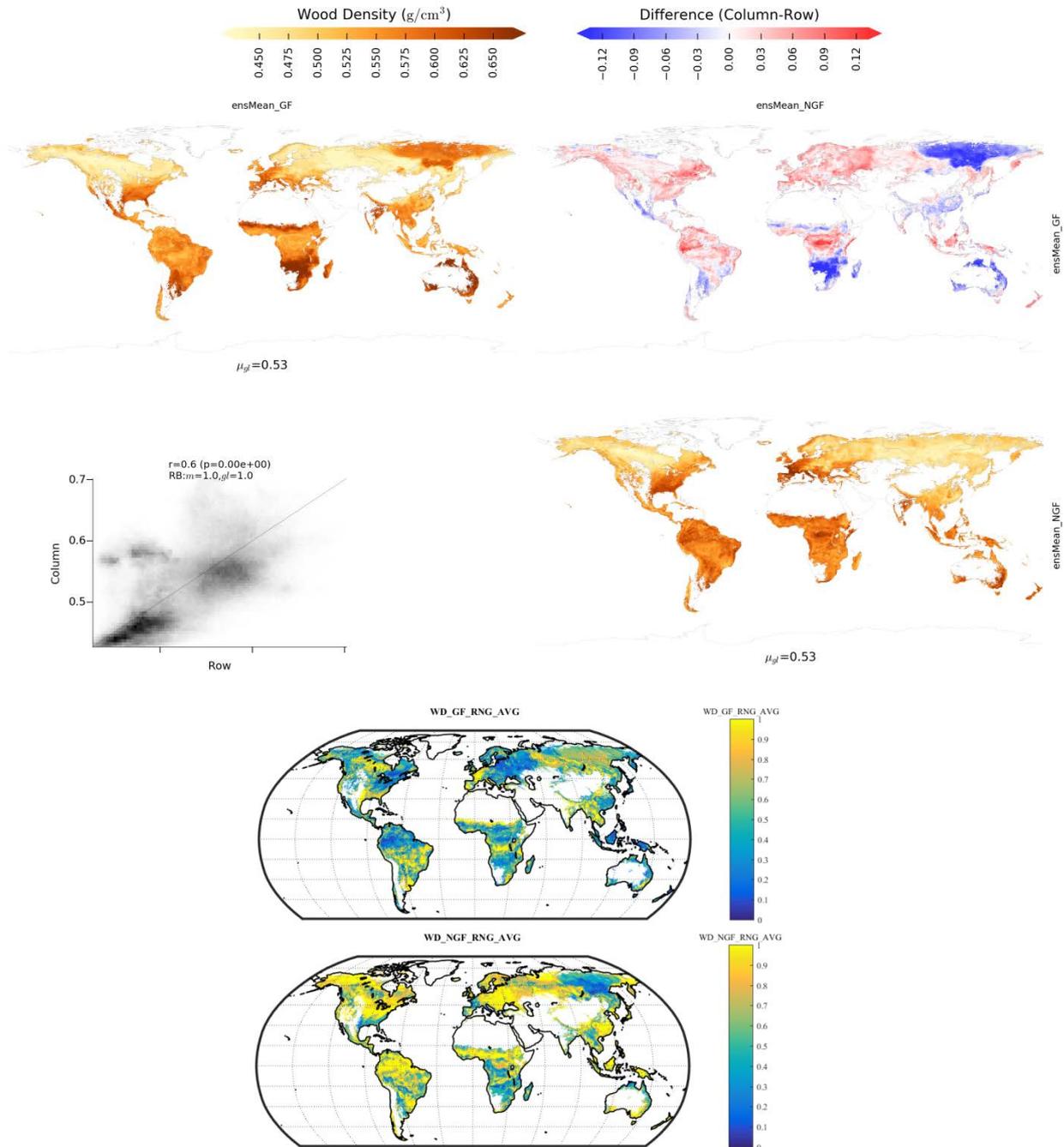


Figure 1-86. Comparison of wood density estimates using ($ensMean_GF$) and not using ($ensMean_NGF$) gapfilled data (top, colour scale binned to the first and 99th percentiles). Although the mean is robust between the two ($\mu_{gl}=0.53$), and 95% of the absolute differences are below 0.12,

the reduction in the spread of different methods is substantial by including gap-filled data. In the bottom figures is shown the ratio between the across-method spread only using gap-filled divided by the full ensemble spread (left) and the ratio between not using gap-filled divided by the full ensemble spread (right). The closer the ratio is to 0 the lowest is the spread between methods for a given data source (gap-filled or not) compared to the whole ensemble.

Despite these differences, the comparison between the different AGB predictions shows a strong consistence between approaches (Figure 1-85): overall the global AGB ranges ~10% of the median; the global correlations between the different estimates are above 0.95; and the global relative biases are marginal. Despite so, locally, or for some regions, like the Amazon, the differences between approaches can range beyond 80 ton/ha. The Amazon, and the tropical regions at large, are regions where considering different data sources mostly impact the spread between the different AGB estimates, where considering the gap-filled TRY dataset reduces substantially the spread between the different approaches (Figure 1-87). At the global scales, including gap-filled data reduces the spread in ~57% of the land area considered, and results in an overall reduction of 43% in the ranges between the different approaches (Figure 1-88). In general, the local ranges in AGB driven by the ranges in the wood density ranges show a median of 15%, and the highest ranges are located in western Amazon, central Congo, tropical South East Asia and central Siberian region (Figure 1-89). The substantial correlation in the AGB estimates between methods suggests that the global patterns in AGB are strongly dominated by the variations in GSV, although the wood density variations can have local or regional implications in estimating the mean and ranges of AGB.

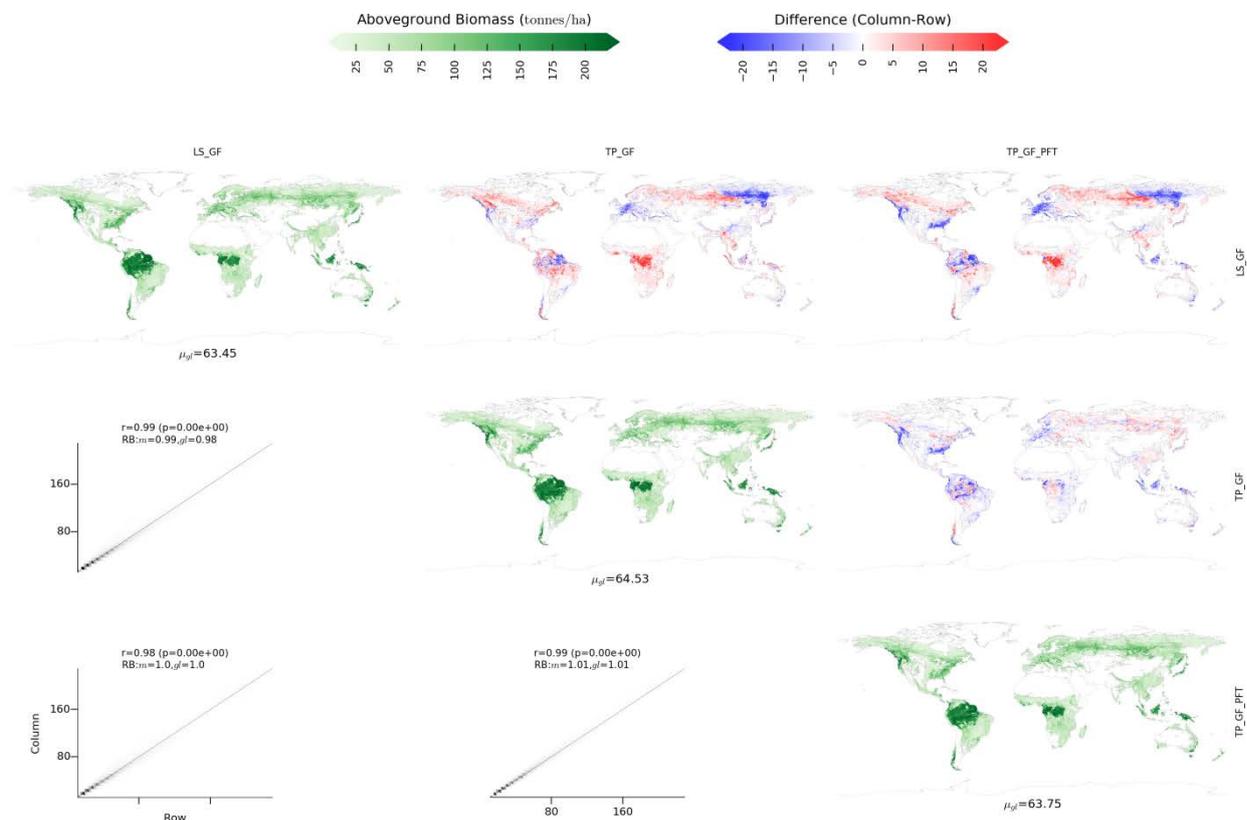


Figure 1-87. Same as Figure 1-84, but for AGB (colour scale binned to the 98th percentile).

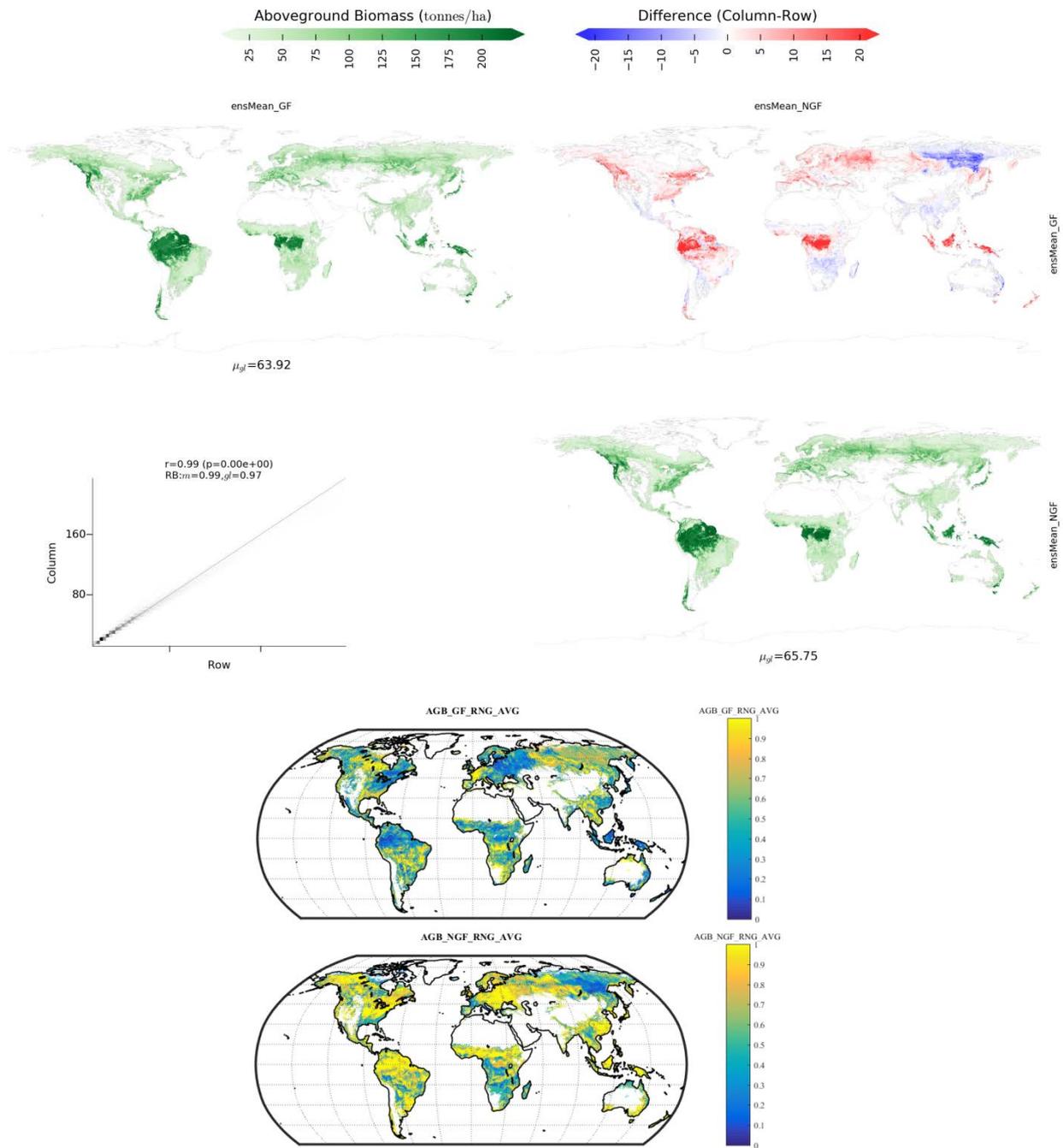


Figure 1-88. Same as Figure 1-85, but for AGB. Here, 95% of the differences between gap-filled and non-gap-filled data are below 12 ton/ha.

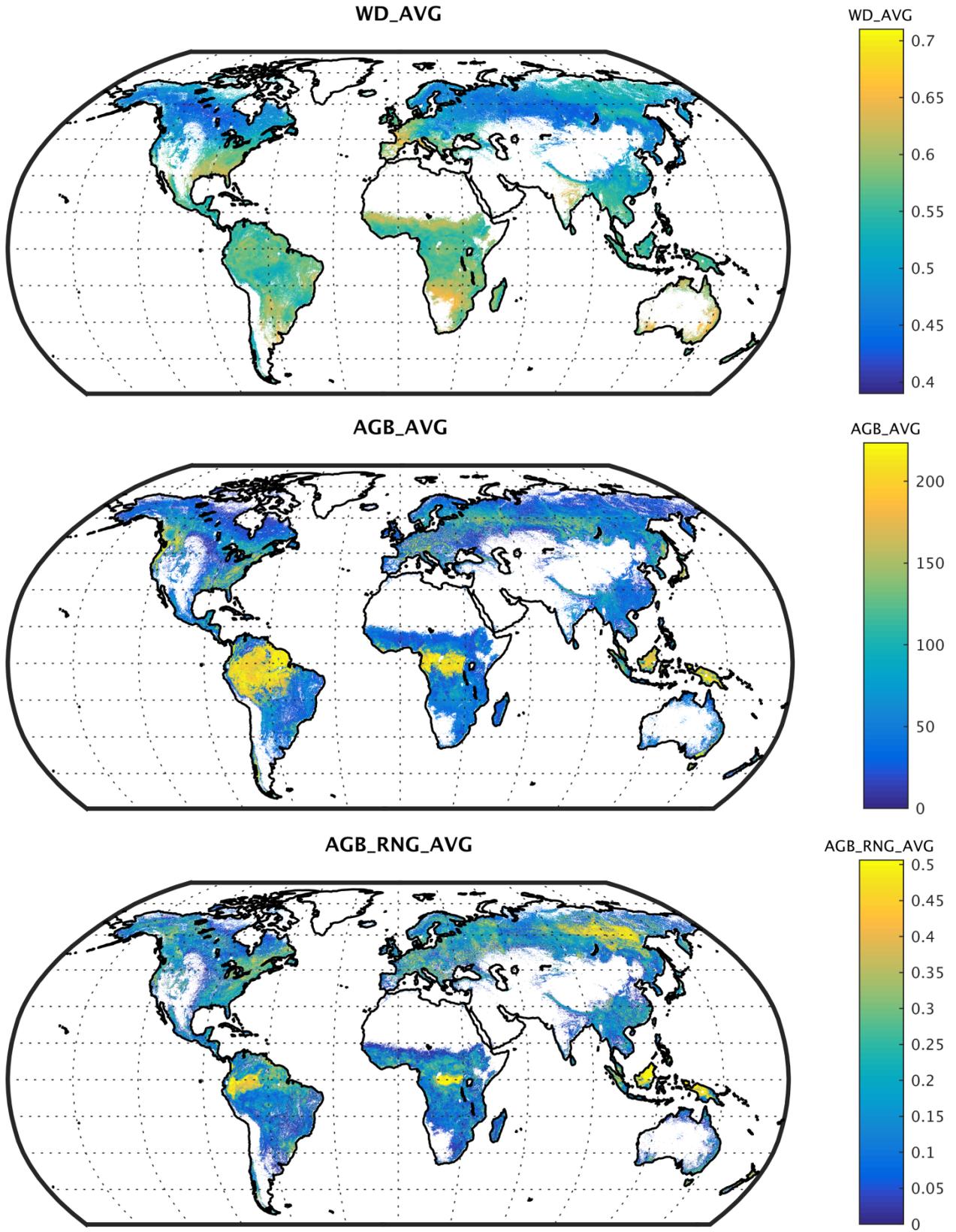


Figure 1-89. Ensemble mean wood density (top, g/cm³), AGB (ton/ha, maximum set to the global 98th percentile) and the ratio between the AGB ensemble range and ensemble mean (bottom).

	GlobBiomass		Page 100
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

Overall, the current results show that wood density can be approximated using information on plant functional type, or leaf type and habits, vegetation properties and environmental conditions. The final results are strongly dependent on the approach, where the landscape level approaches show a better statistical fitness when compared to the tree/plot level predictions. The latter approaches show a proportion of explained variance in wood density that seems still low for a robust estimate at a spatial resolution of tens to a few hundred meters. However, since the estimates seem unbiased, coarser scale estimates may be more trustworthy. The main limiting factors at this point seem to be (1) the lack of information on species and forest development stage worldwide, which is understood to be the most significant contributor to the variability of wood density globally. However, this information is not available globally. Another relevant aspect is (2) the mismatch in scale between in situ observations of wood density and allometry, and the predictors extracted from global databases, as well as (3) the fact that wood density measurements may be asynchronous to the representative dates of some covariates, a factor that we tried to circumvent by working with climatology, and long term average, datasets. Another key aspect is (4) the inclusion of additional independent observations that would allow a better coverage of the Americas, Africa and Oceania, in order to evaluate the robustness of the gap-filling approach, but more importantly, for overcoming the limitations in these under-represented regions. Methodologically, the statistics also suggest that the landscape scale approach seems to point in the direction that abstracting from the tree/plot level estimates to a community level wood density brings substantial advantages in explaining its spatial variability. These results seem to point in the direction to (i) develop further approaches based on the principles of the landscape, or community, level wood density based on species abundance information for plot measurements; and, that in this regard, (ii) it is likely advantageous to explore time series of Landsat data – spanning now more than three decades – as a higher resolution data-stream informative on the development stage of forests; and not forgetting the (iii) need for a better representation of wood density variations in tropical and subtropical regions [e.g. *Lewis et al.*, 2013; *Mitchard et al.*, 2014].

Ultimately, one relevant aspect relates to the importance of benchmarking the global patterns given the differences in the predictions by the different methods, and the difficulty in confidently rule out any of them. Likely, the evaluation must focus on the spatial variability of AGB, which at global scale is mostly dominated by variations in GSV, but at regional level can help separating better wood density predictions.

1.4.5 Methods to assign accuracies to the retrieved AGB

Accuracy is defined as the difference between the retrieved and true GSV. For an ensemble of data, this gives rise to two statistical measures of accuracy: the average of this difference (bias) and a term indicating the variability of the estimate (standard error). The accuracy of the retrieved biomass depends on the accuracy of the input data and the estimation procedure. Here we detail methods to quantify the standard error for each of the individual global inversion methods and for the final product. The methods presented reflect our understanding of the standard errors embedded in the retrieval algorithms.

1.4.5.1 BIOMASAR-C

The standard error of the individual GSV estimates obtained with the BIOMASAR-C approach is quantified by propagating the standard error of (i) the measured SAR backscatter, σ_{meas}^0 , and (ii) the estimates of the forest backscatter model parameters σ_{gr}^0 , σ_{df}^0 , β and V_{df} . Since the standard error of

	GlobBiomass		Page 101
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

the five random variables listed above are uncorrelated and small (see below), the standard error of the estimate of GSV obtained from a single backscatter observation, σ_v , is given by:

$$(1-26)$$

where the partial derivatives are derived from Equation (1-3) (Annex B). The uncertainty in a backscatter measurement is affected by the radiometric and calibration accuracies, thermal noise and speckle. The SAR pre-processing also introduced additional uncertainty related to: (i) the accuracy of the geocoding transformation and resampling between radar and map geometries; (ii) the horizontal and vertical accuracy of the DEM used as reference for the map geometry, and (iii) the accuracy of the pixel area and local incidence angle used to normalize the backscatter for slope-induced effects on the backscatter. Since the pixel level uncertainties in the DEMs used in this study are unavailable, we cannot estimate the variance of a backscatter measurement from the individual variances of the terms listed above. We therefore estimate it empirically by equating it to the ENL. The ENL was estimated in hundreds of polygons, each covering an area of visually homogeneous backscatter and land cover. Polygons were drawn manually in randomly selected images from randomly selected tiles. This led to estimates of the backscatter variance given by:

$$0.6 \text{ dB for } N \leq 50; 0.5 \text{ dB for } 51 \leq N \leq 150; \text{ and } 0.4 \text{ dB for } N > 150,$$

where N is the number of backscatter observations at a pixel.

The standard error of the model parameters expressing the backscatter from the ground and dense forest combines the standard error in the observations of the pixels labelled as unvegetated or dense forest, respectively, and the standard error of the backscatter for the percent tree cover values representing the “ground” and “dense forest” classes. Here σ_{gr}^o and σ_{df}^o are estimated from the histograms of the SAR backscatter measurements for unvegetated and dense forest pixels (see Figure 9 in Santoro et al., 2011). The uncertainties can be set equal to the standard error of each histogram minus the standard error associated with each backscatter measurement.

In previous studies on boreal and temperate forest, we concluded that a plausible range of values for the coefficient β is between 0.003 and 0.007 ha/m³; hence, we assumed that a Gaussian distribution with standard deviation of 0.002 ha/m³ can be used to describe the variability associated with the model parameter β . Having revisited the estimation of β to account for all ecological zones of the world, this assumption has become obsolete. Here we define the standard error of the coefficient β as the range of values for which the model in Equation (1-4) fits observations of MODIS VCF and GSV. It is important to remark that this is an educated guess due to the lack of observations of forest transmissivity from different biomes at the spatial resolution of ASAR.

The standard error of the V_{df} parameter has been discussed in Section 1.4.1.10. The standard error in percent values has been obtained for each class of the FAO GEZ dataset. This approximation might be too coarse but in our opinion is currently the best estimate that could be achieved.

The standard error in the multi-temporal GSV estimate is modelled as a linear combination of the single-image GSV standard errors given by Equation (1-26). Here it is assumed that the original weights, w_i , are the best estimate of the individual variances of the individual estimates of GSV, so that the standard error of the multi-temporal GSV estimate is defined as:

	GlobBiomass		Page 102
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

$$\delta_{V_{mi}} = \frac{\sum_{i=1}^N w_i \delta_{V_i}}{\sum_{i=1}^N w_i} \quad (1-27)$$

The covariance between different observations is neglected. The strong sensitivity of the C-band SAR backscatter to environmental conditions typically results in very low correlation between observations over forests, e.g., Askne & Santoro (2009). In addition, the speckle patterns are uncorrelated because of the different viewing geometries of observations within the repeat-pass cycle and because of temporal and spatial decorrelation of images along the same orbital track.

1.4.5.2 BIOMASAR-L

The BIOMASAR-C error model used to quantify the standard error, δ_V , Equation (1-26), of GSV estimates from single backscatter observations will also be applied in the case of BIOMASAR-L. However, some modifications are required for BIOMASAR-L because of the differences in the estimation of the parameter σ_{veg}^0 .

In the model inverted for GSV, σ_{veg}^0 is expressed as a function of the average backscatter observed over dense forests, σ_{df}^0 , the average canopy density, η_{df} , the average height of dense forests, h_{df} , and the two-way attenuation coefficient, α . The error model therefore needs to be reformulated to consider the error associated with η_{df} , α , and h_{df} :

$$\delta(V) = \sqrt{\left(\frac{\partial V}{\partial \sigma_m^0}\right)^2 \delta(\sigma_m^0)^2 + \left(\frac{\partial V}{\partial \eta_{df}}\right)^2 \delta(\eta_{df})^2 + \left(\frac{\partial V}{\partial \alpha}\right)^2 \delta(\alpha)^2 + \left(\frac{\partial V}{\partial h_{df}}\right)^2 \delta(h_{df})^2 + \left(\frac{\partial V}{\partial \sigma_{df}^0}\right)^2 \delta(\sigma_{df}^0)^2 + \left(\frac{\partial V}{\partial \sigma_{gr}^0}\right)^2 \delta(\sigma_{gr}^0)^2} \quad (1-28)$$

As for BIOMASAR-C, the uncertainty in the estimates of σ_{df}^0 and σ_{gr}^0 is estimated as the standard deviation of the histograms of backscatter observations in areas of low and high canopy density, since the histograms summarize the uncertainties associated with estimating the parameters due to spatially variable imaging conditions, uncompensated topographic effects, etc. (e.g., variable soil/canopy moisture). Reported accuracies of height and canopy density estimates derived from ICESAT GLAS are used to determine the uncertainty associated with estimating h_{df} and η_{df} . Following the results in Los et al. (2012) and Simard et al. (2011), who validated GLAS-based height estimates at boreal, temperate, sub-tropical, and tropical forest sites, we assumed standard errors for height estimates at the GLAS footprint-level between 4 (boreal) and 10 m (tropics). While a large number of studies on the estimation of canopy cover and closely related variables such as fractional cover, gap probability or transmittance from Lidar exist, only a few have presented comprehensive validation. It is thus not possible to provide forest type-specific numbers for the error in the BIOMASAR-L retrieval associated with errors of the parameter η_{df} . As indicated by Garcia et al. (2012), the estimation of canopy cover from ICESAT GLAS with the ratio of energy returned from the canopy to the total energy returned may be of the order of 15 to 20 %. We therefore assume a global error of 20 %. Note that the parameters h_{df} and η_{df} are estimated with the average GLAS height and density across all footprints covering a PALSAR 1x1 degree tile that cover dense forest according to Landsat. It is therefore assumed that the standard error for the two parameters reduces as the square root of the number of GLAS footprints used in the estimation.

The uncertainties in the forest transmissivity coefficient β and the related two-way attenuation coefficient α are the most difficult to specify and it is only possible to provide a best guess, as for BIOMASAR-C. In the case of β , the associated uncertainty may be inferred from the relationship of

the forest transmissivity, simulated with the aid of GLAS height and optical canopy density estimates, and GSV (see Figure 1-34). The results presented in Figure 1-34 suggest that the uncertainty associated with the forest transmissivity parameter increases with increasing β . The 95% bounds of the estimate for β increased from ± 0.002 ha/m³ in the case of low values of β that are valid in boreal and subtropical dry forests to ± 0.007 ha/m³ for the highest values of β that are applied in the tropics. For the two-way attenuation coefficient α , we assume a standard error of 0.25 dB/m, which is roughly consistent with the range of values reported in the literature (Ulaby et al., 1990; Chauhan et al., 1991; Shinohara et al., 1992; Sheen et al., 1994; Kurum et al., 2009; Praks et al., 2012).

The plots in Figure 1-90 show the standard error of the BIOMASAR-L GSV estimates at 25 m resolution as a function of the estimated GSV for two forest sites in Colombia and South Africa. The plots suggest that the standard error at full resolution is large, in particular for dense tropical forests, and that aggregation of the maps to coarser pixel sizes will be required to significantly reduce the error.

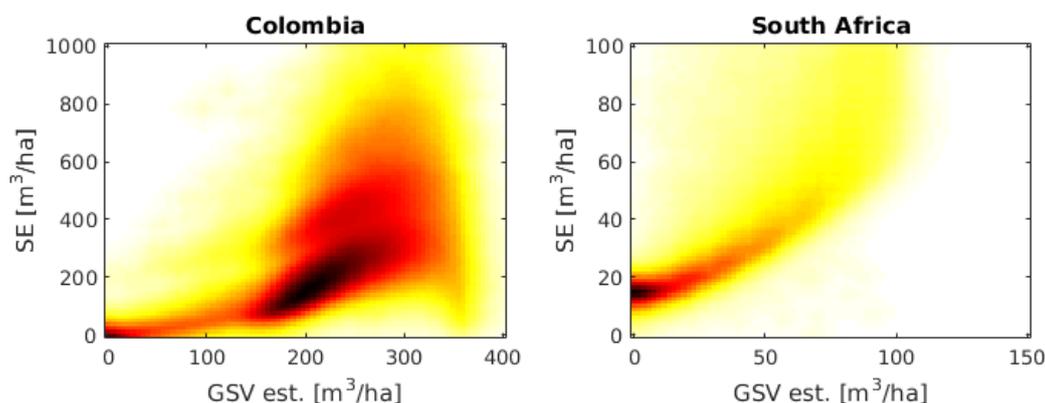


Figure 1-90. Standard error of BIOMASAR-L GSV estimates at 25 m resolution for a forest site in the Colombian Amazon (left) and South Africa (right).

1.4.5.3 BIOMASAR-C+

The BIOMASAR-C+ map at full 25m pixel size has been produced by rescaling the BIOMASAR-C 1 km GSV estimates derived from multi-temporal ENVISAT ASAR data to 25 m using a simple multiple linear regression model with ALOS PALSAR co- and cross-polarized backscatter as well as Landsat reflectances in bands 3, 4, 5, and 7 as predictor layers. To quantify the error associated with the 25 m prediction of GSV, we consider three error sources:

- 1) The error in the 1km BIOMASAR-C map
- 2) Modelling
- 3) Scaling

1) The error in the 1km BIOMASAR-C map, which has been determined globally as described in Section 1.4.5.1, will propagate to the 25 m pixel size map. Figure 1-91 illustrates the relative standard error of the BIOMASAR-C GSV estimates as a function of the GSV estimates. The Figure shows for one example that the relative error (i.e. SE/GSV in percent), decreases with increasing GSV and may be described by an exponential model of the form:

$$\delta V_{BIO-C} / V_{BIO-C} = ae^{bV_{BIO-C}} + c \quad (1-29)$$

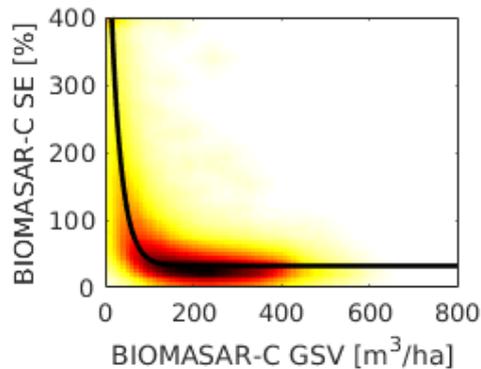


Figure 1-91. Relative standard error of BIOMASAR-C GSV estimates as a function of GSV.

2) Modelling. When calibrating models for rescaling the 1km BIOMASAR-C GSV estimates to higher resolution, additional error will be introduced. We here characterize this error for each FAO ecoregion by looking at the root mean square differences between BIOMASAR-C GSV and BIOMASAR-C+ GSV (aggregated back to 1km pixel size) in narrow intervals of GSV. This allows us to describe differences in the error across the range of GSVs, e.g., by fitting an empirical model such as a polynomial to the observed trend (Figure 1-92).

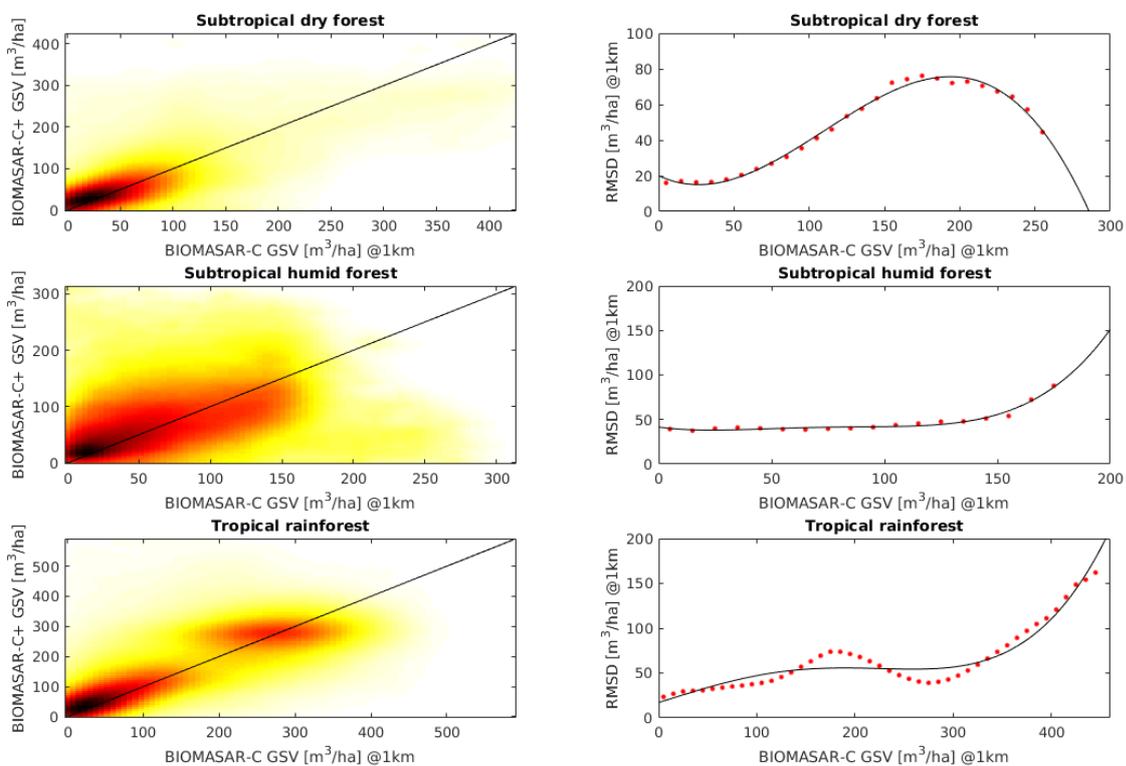


Figure 1-92. Left: Comparison of BIOMASAR-C and BIOMASAR-C+ GSV estimates aggregated to 1 km pixel size. Right: Root mean square difference between BIOMASAR-C and BIOMASAR-C+ GSV in 10 m³/ha intervals of GSV including the fit by a fourth order polynomial.

	GlobBiomass		Page 105
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

3) Scaling. We here assume that the standard error when rescaling from 1 km to a finer resolution scales with the pixel area of the GSV predictions, i.e.:

$$\delta V_{A1} = \delta V_{A2} \sqrt{\frac{A2}{A1}} \quad (1-30)$$

where A1 represents the pixel area of the 1km products and A2 the pixel area of the rescaled product.

Equations (1-29) and (1-30) are used to quantify the standard error of GSV estimates at high resolution in the BIOMASAR-C+ map.

1.4.5.4 CESBIO

See Section 3.5 (ATBD – Regional Biomass Maps).

1.4.5.5 Integrated GSV data product

The accuracy of the integrated GSV product is given by the weighted mean of the accuracies of the BIOMASAR-L and the BIOMASAR-C+ datasets.

1.4.5.6 Conversion of GSV to AGB

The accuracy of the conversions will be provided by the cross-validated prediction differences (differences in estimates not used in the algorithm training). Given that the estimates seem unbiased through the wood density ranges, the accuracy can be estimated as the prediction error. The uncertainty in predictions will be reported as the 95% confidence intervals across the different approaches. At this point, the ranges in the methodological ensemble are shown here for a qualitative assessment of regions that seem more affected by the spread in the different wood density estimates.

1.5 Products

The most recent version of the GSV dataset obtained at 25 m by merging the BIOMASAR-L and BIOMASAR-C+ estimates is shown in Figure 1-93. Boxcar averaging is applied to generate aggregated versions. The latitudinal gradient displayed in Figure 1-61 shows that the merging algorithm has adapted depending on the region, preferring the input estimate closest to the theoretical behaviour when compared with other datasets of forest variables. Accordingly, spatially explicit accuracies are produced with the methodology described above. The dataset is currently being computed; unfortunately, it could not be completed before submission of this document. The AGB estimates and the corresponding accuracy dataset will be generated on completion of the conversion from GSV to AGB by MPI.

	GlobBiomass		Page 106
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

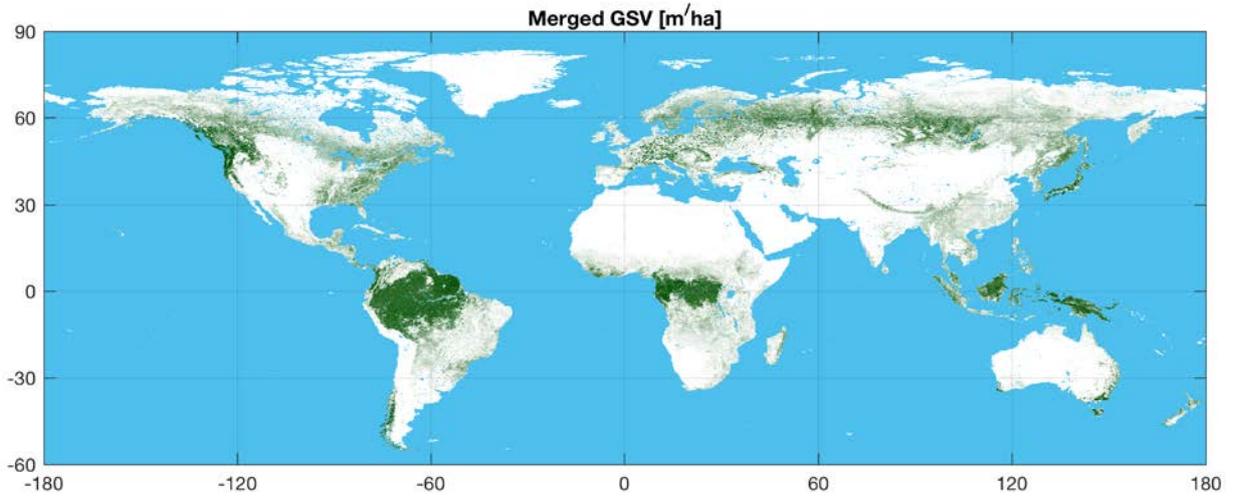


Figure 1-93. Estimates of GSV with a pixel size of 0.01° obtained after merging the BIOMASAR-L and BIOMASAR-C+ datasets (version of 2017-08-15).

	GlobBiomass		Page 107
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

2 Annex A

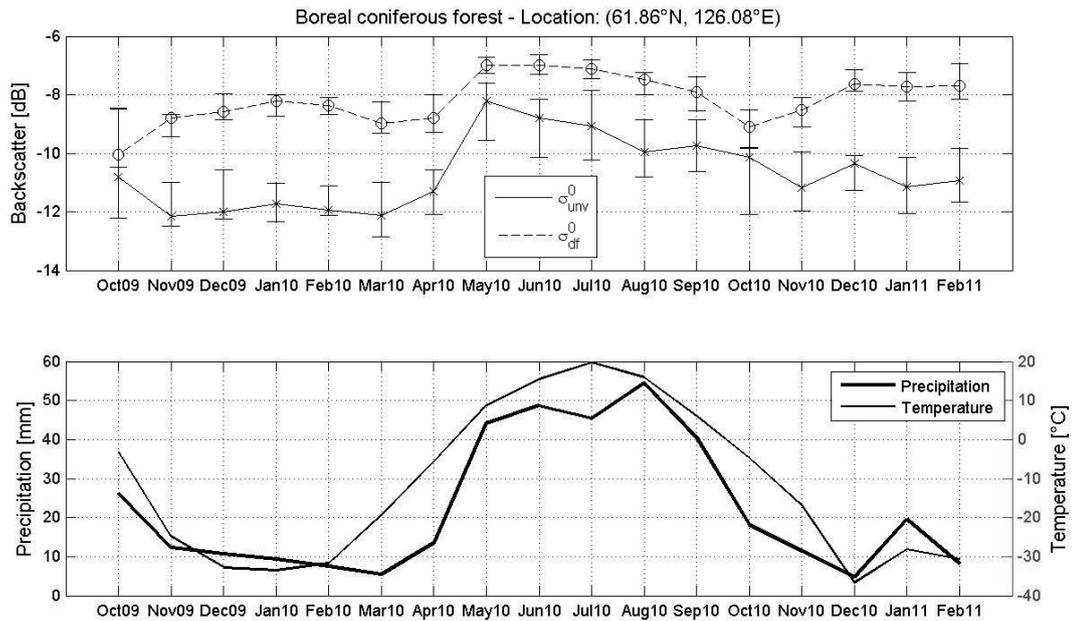
This Annex describes the behaviour of SAR backscatter over time for sparsely vegetated and dense forest areas. The locations reported in Figure S1 were part of a large set of points randomly chosen within each FAO ecological zone for which there were (i) at least four backscatter observations each month and (ii) a weather station with continuous records of monthly total precipitation and monthly average temperature within a radius of 100 km. A 100 km window size was used to derive the statistics of the backscatter for the given location using Version 5 of the MODIS VCF dataset in support of the selection of sparsely vegetated areas (VCF < 30%) and dense forest areas (VCF > 65%). The plots in Figure S1 are representative of average behaviour of the backscatter for a given ecological zone with respect to precipitation and temperature. Values inferred from these measurements are summarised in Table A1.

In boreal forests, the largest backscatter difference between dense forest (σ_{df}^0) and sparsely vegetated areas (σ_{unv}^0) was obtained under frozen conditions during winter-time (typically 3-4 dB) (Figure S1a). Observations under wet and unfrozen conditions had the smallest backscatter difference (1-2 dB). In temperate forests, the backscatter difference was at most 2 dB and occurred under frozen conditions (Figure S1b). During unfrozen conditions, the backscatter difference was on average 1 dB. Almost the same backscatter level occurred in months with the largest amount of total precipitation and in the following month (Figure S1b). For subtropical dry forest, the backscatter difference, with an average of approximately 2 dB, was largest during the summer months with long periods of dry conditions (Figure S1c). It was smallest (around 1 dB) in months with the largest amount of precipitation or periods immediately after. In subtropical humid forest, the seasonality trends were not as clear as in tropical dry forest. The backscatter difference was usually larger during drier periods, but always less than 2 dB on average (Figure S1d). Wet conditions were characterized by less than 1 dB difference. In tropical dry forest, the backscatter from sparsely vegetated areas depended on the amount of precipitation (Figure S1e), thus affecting the backscatter difference. Before the rainy season, the backscatter difference increased from 1 to 3 dB, then dropped to almost 0 dB during the rainy season and increased again following periods of intense precipitation. The impact of seasonal precipitation was also visible for tropical moist deciduous forest and tropical rain forest (Figures S1f and S1g), but the backscatter difference was smaller than for tropical dry forest. It was at most 2 dB before the rainy period, and was negligible during the rainy period, with occasional higher backscatter difference in sparsely vegetated areas than in dense forest (Figure S1f). The backscatter difference started increasing after the rainy season, but was below 1 dB on average during several months.

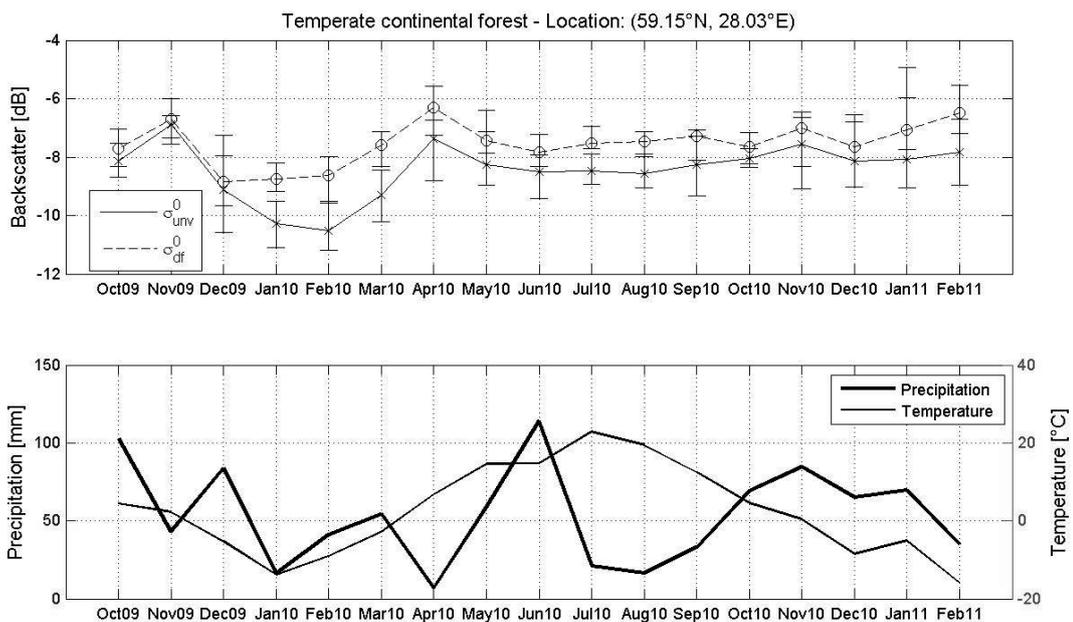
Table A1. Largest and smallest ASAR backscatter difference between dense forest and unvegetated areas, by ecological zone.

Ecological zone	Largest difference	Smallest difference
Polar	Frozen conditions, 3-4 dB	Wet/unfrozen conditions, 1-2 dB
Boreal	Frozen conditions, 3-4 dB	Wet/unfrozen conditions, 1-2 dB
Temperate	Frozen conditions, 2 dB	Wet/unfrozen conditions, 0-1 dB
Subtropical dry	Dry/summer conditions, 2 dB	Wet conditions, 1 dB
Subtropical humid	Dry conditions, 2 dB	Wet conditions, < 1 dB

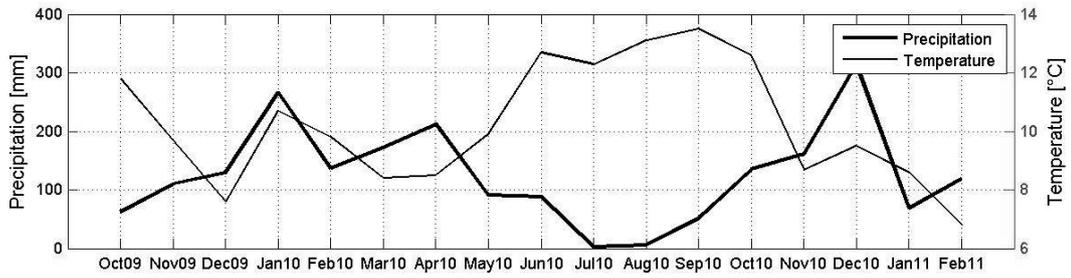
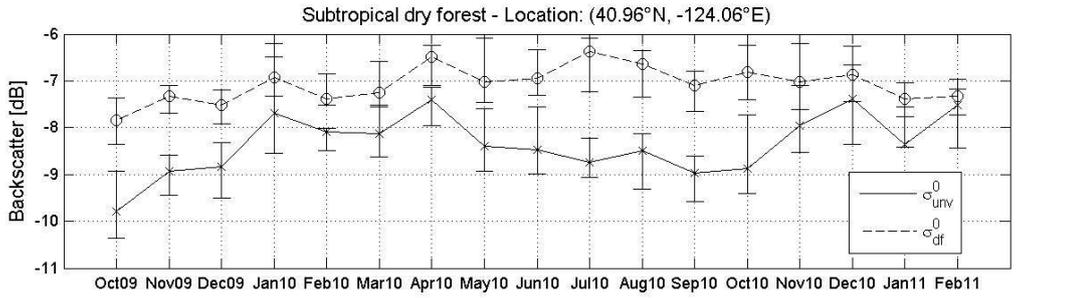
Tropical dry	Before the rainy season (dry), 1-3 dB	During and after the rain season, 0 dB
Tropical moist	Before the rain season (dry), 2 dB	During and after the rain season, 0-1 dB
Tropical rain forest	Before the rain season (dry), 2 dB	During and after the rain season, 0-1 dB



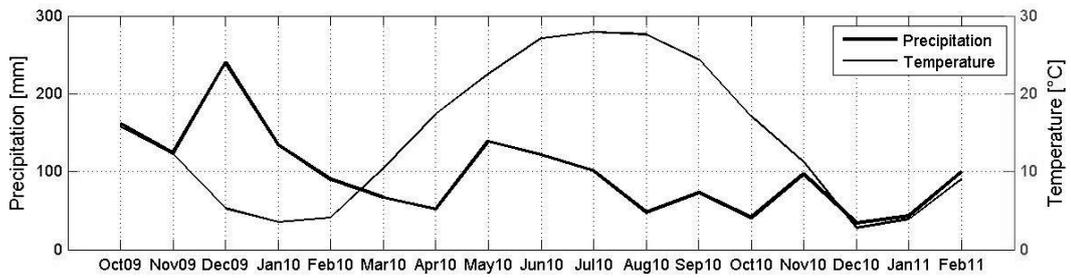
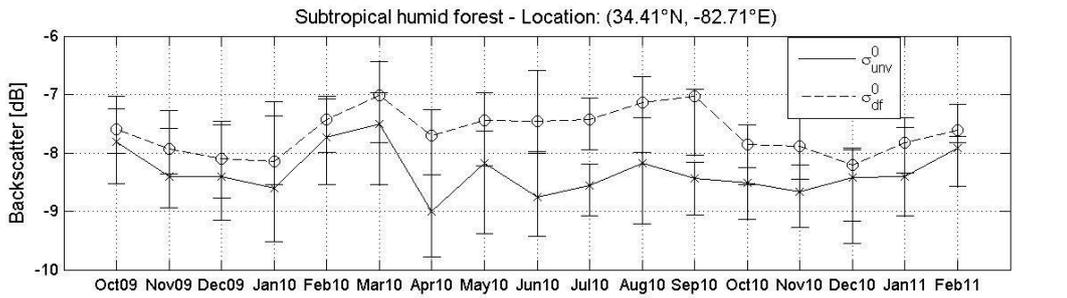
(a)



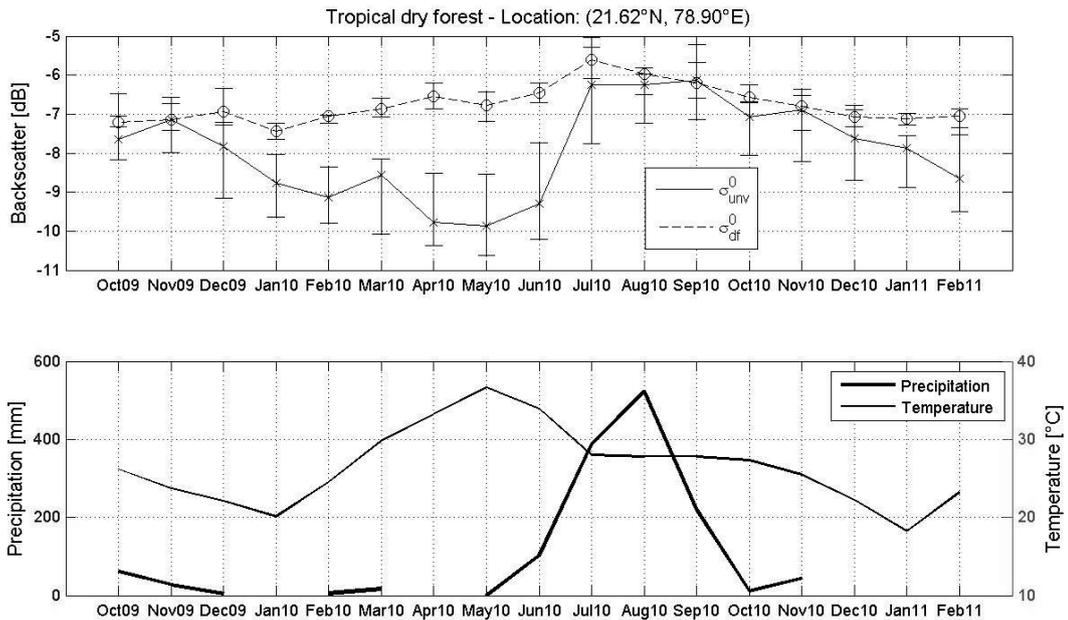
(b)



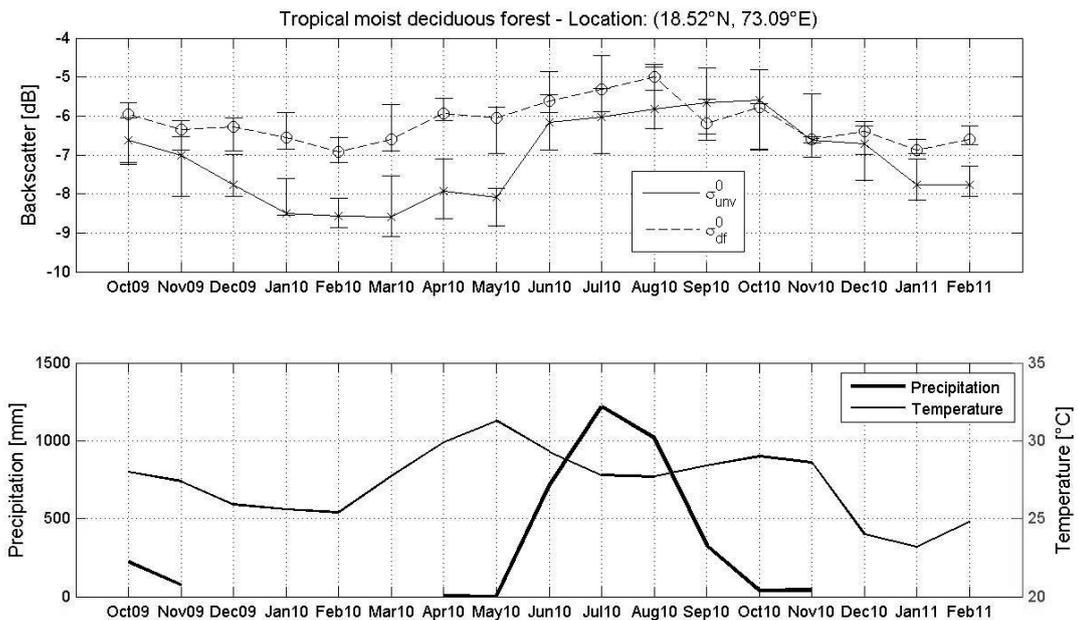
(c)



(d)



(e)



(f)

Figure Annex S1. Pairs of panels illustrating (top) the monthly distribution of backscatter for sparsely vegetated (σ_{unv}^0) areas and dense forest (σ_{df}^0), and (bottom) total precipitation and monthly average temperature throughout the time span of the ASAR dataset. Circles and crosses represent the median backscatter for the dense forest and the sparsely vegetated classes, respectively. The vertical bars represent the interquartile range of the backscatter.

	GlobBiomass		Page 111
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

3 Annex B

Partial derivatives:

$$\left(\frac{\partial V}{\partial \sigma_{meas}^0} \right)_{\sigma_{gr}^0, \sigma_{df}^0, \beta, V_{df}} = \frac{e^{-\beta V_{df}} - 1}{\beta \left[(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}} - \sigma_{meas}^0 + \sigma_{df}^0 \right]} \quad (S1)$$

$$\left(\frac{\partial V}{\partial \sigma_{gr}^0} \right)_{\sigma_{meas}^0, \sigma_{df}^0, \beta, V_{df}} = \frac{1}{\beta} \left[\frac{1}{\sigma_{df}^0 - \sigma_{gr}^0} - \frac{e^{-\beta V_{df}}}{(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}} - \sigma_{meas}^0 + \sigma_{df}^0} \right] \quad (S2)$$

$$\left(\frac{\partial V}{\partial \sigma_{df}^0} \right)_{\sigma_{meas}^0, \sigma_{gr}^0, \beta, V_{df}} = \frac{1}{\beta} \left[\frac{e^{-\beta V_{df}}}{(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}} - \sigma_{meas}^0 + \sigma_{df}^0} - \frac{1}{\sigma_{df}^0 - \sigma_{gr}^0} \right] \quad (S3)$$

$$\left(\frac{\partial V}{\partial \beta} \right)_{\sigma_{meas}^0, \sigma_{gr}^0, \sigma_{df}^0, V_{df}} = \left\{ \frac{\ln \left[(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}} - \sigma_{meas}^0 + \sigma_{df}^0 \right] - \ln(\sigma_{df}^0 - \sigma_{gr}^0)}{\beta^2} + \frac{(\sigma_{meas}^0 - \sigma_{gr}^0) V_{df} e^{-\beta V_{df}}}{\beta \left[(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}} - \sigma_{meas}^0 + \sigma_{df}^0 \right]} \right\} \quad (S4)$$

$$\left(\frac{\partial V}{\partial V_{df}} \right)_{\sigma_{meas}^0, \sigma_{gr}^0, \sigma_{df}^0, \beta} = \frac{(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}}}{(\sigma_{meas}^0 - \sigma_{gr}^0) e^{-\beta V_{df}} - \sigma_{meas}^0 + \sigma_{df}^0} \quad (S5)$$

	GlobBiomass		Page 112
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

4 References

- Askne, J., Dammert, P. B. G., Ulander, L. M. H. and Smith, G. (1997), C-band repeat-pass interferometric SAR observations of the forest. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 25-35.
- Askne, J., Santoro, M., Smith, G. and Fransson, J. E. S. (2003), Multitemporal repeat-pass SAR interferometry of boreal forests. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1540-1550.
- Askne, J. and Santoro, M. (2009), Automatic model-based estimation of boreal forest stem volume from repeat pass C-band InSAR coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 513-516.
- Askne, J. and Santoro, M. (2012). Experiences in boreal forest stem volume estimation from multitemporal C-band InSAR. In: I. Padron (Editor), *Recent Interferometry Applications in Topography and Astronomy*. InTech.
- Asner, G. P., Clark, J. K., Mascaro, J., Galindo Garcia, G. A., Chadwick, K. D., Navarrete Encinales, D. A., Paez-Acosta, G., Cabrera Montenegro, E., Kennedy-Bowdoin, T., Duque, A., Balaji, A., von Hildebrand, P., Maatoug, L., Phillips Bernal, J. F., Yepes Quintero, A. P., Knapp, D. E., Garcia Davila, M. C., Jacobson, J. and Ordóñez, M. F. (2012), High-resolution mapping of forest carbon stocks in the Colombian Amazon. *Biogeosciences*, 9, 2683-2696.
- Asner, G. P., Knapp, D. E., Martin, R. E., Tupayachi, R., Anderson, C. B., Mascaro, J., Sinca, F., Chadwick, K. D., Higgins, M., Farfan, W., Llactayo, W. and Silman, M. R. (2014), Targeted carbon conservation at national scales with high-resolution monitoring. *Proceedings of the National Academy of Sciences*, 111, E5016-E5022.
- Attema, E. P. W. and Ulaby, F. T. (1978), Vegetation modeled as a water cloud. *Radio Science*, 13, 357-364.
- Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Athon, P., Banin, L., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C., Kearsley, E., Lindsell, J., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M., Ryan, C. M., Slik, F., Sunderland, T. C. H., Vaglio Laurin, G., Valentini, R., Verbeeck, H., Wijaya, A. and Willcock, S. (2015), An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology*.
- Avtar, R., Suzuki, R., Takeuchi, W. and Sawada, H. (2013), PALSAR 50 m mosaic data based national level biomass estimation in Cambodia for implementation of REDD+ mechanism. *PLoS ONE*, 8, e74807.
- Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A., Samanta, S. and Houghton, R. A. (2012), Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2, 182-185.
- Baker, J. R. and Luckman, A. J. (1999), Microwave observations of boreal forests in the NOPEX area of Sweden and a comparison with observations of a temperate plantation in the United Kingdom. *Agricultural and Forest Meteorology*, 98-99, 389-416.
- Beaudoin, A., Bernier, P. Y., Guindon, L., Villemaire, P., Guo, X. J., Stinson, G., Bergeron, T., Magnussen, S. and Hall, R. J. (2014), Mapping attributes of Canada's forests at moderate resolution through kNN and MODIS imagery. *Canadian Journal of Forest Research*, 44, 521-532.
- Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., Lister, A. J.,

	GlobBiomass		Page 113
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- Moisen, G. G., Nelson, M. D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D. L., Winterberger, K. C., Brandeis, T. J., Czaplowski, R. L., McRoberts, R. E., Patterson, P. L. and Tymcio, R. P. (2008), Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, 112, 1658-1677.
- Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L. and Kimes, D. S. (2008), Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112, 3876-3890. Breiman, L. (2001), Random Forests. *Machine Learning*, 45, 5-32.
- Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L. and Asner, G. (submitted), An above-ground biomass map of African savannahs and woodlands at 25 meters resolution derived from ALOS PALSAR. *Remote Sensing of Environment*.
- Bruniquel, J. and Lopes, A. (1997), Multi-variate optimal speckle reduction in SAR imagery. *International Journal of Remote Sensing*, 18, 603-627.
- Carreiras, J. M. B., Melo, J. B. and Vasconcelos, M. J. (2013), Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data. *Remote Sensing*, 5, 1524-1548.
- Cartus, O., Santoro, M., Schmullius, C. and Li, Z. (2011), Large area forest stem volume mapping in the boreal zone using synergy of ERS-1/2 tandem coherence and MODIS vegetation continuous fields. *Remote Sensing of Environment*, 115, 931-943.
- Cartus, O., Santoro, M., Kelldorfer, J., "Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band," *Remote Sensing of Environment*, vol. 124, pp. 466-478, 2012.
- Cartus, O., Kelldorfer, J., Walker, W., Bishop, J., Franco, C., Santos, L. and Michel Fuentes, J. M. (2014), A national, detailed map of forest aboveground carbon stocks in Mexico. *Remote Sensing*, 6, 5559-5588.
- Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T. and Durand, P. (2001), Sensitivity of spaceborne SAR data to forest parameters over sloping terrain. Theory and experiment. *International Journal of Remote Sensing*, 22, 2351-2376.
- CCG (2007), Canadian Digital Elevation Data, GeoBase, available at <http://www.geobase.ca/geobase/en/index.html>, last accessed December 2014.
- Chauhan, N. S., Lang, R. H. and Ranson, K. J. (1991), Radar modeling of a boreal forest. *IEEE Transactions on Geoscience and Remote Sensing*, 29, 627-638.
- de Ferranti, J., Digital Elevation Data, available at <http://www.viewfinderpanoramas.org/dem3.html>, last accessed December 2014.
- Desnos, Y.-L., Buck, C., Guijarro, J., Suchail, J.-L., Torres, R. and Attema, E. (2000), ASAR - Envisat's Advanced Synthetic Aperture Radar. *ESA Bulletin*, 102, 91-100.
- DiMiceli, C. M., Carroll, M. L., Sohlberg, R. A., Huang, C., Hansen, M. C. and Townshend, J. R. G. (2011), Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5 Percent Tree Cover, University of Maryland, College Park, MD, USA, available at <http://glcf.umd.edu/data/vcf/>, last accessed December 2014.
- Dobson, M. C., Ulaby, F. T., Le Toan, T., Beaudoin, A., Kasichke, E. S. and Christensen, N. (1992), Dependence of radar backscatter on coniferous forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 412-416.

	GlobBiomass		Page 114
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- FAO (2001), FRA 2000 - Global ecological zoning for the Global Forest Resources Assessment 2000. Final Report, Working Paper 56, Rome.
- FAO (2010), Global Forest Resources Assessment 2010. FAO Forestry Paper, 163, Rome.
- Fransson, J. E. S. and Israelsson, H. (1999), Estimation of stem volume in boreal forests using ERS-1 C- and JERS-1 L-band SAR data. *International Journal of Remote Sensing*, 20, 123-137.
- Frey, O., Santoro, M., Werner, C. and Wegmuller, U. (2013), DEM-based SAR pixel-area estimation for enhanced geocoding refinement and radiometric normalization. *IEEE Geoscience and Remote Sensing Letters*, 10, 48-52.
- Gallaun, H., Zanchi, G., Nabuurs, G. J., Hengeveld, G., Schardt, M. and Verkerk, P. J. (2010), EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*, 260, 252-261.
- Garcia, M., Popescu, S., Riano, D., Zhao, K., Neuenschwander, A., Agca, M., and Chuvieco, E. (2012). Characterization of canopy fuels using ICESAT/GLAS data. *Remote Sensing of Environment*, 123, 81-89.
- Hansen, M. C., De Fries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C. and Sohlberg, R. A. (2003), Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS Vegetation Continuous Field algorithm. *Earth Interactions*, 7, 1-15.
- Hansen, M. C., Egorov, A., & Roy, D. (2011). Continuous fields of land cover for the conterminous United States using Landsat data : first results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters*, (915935429).
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G. (2013), High-resolution global maps of 21-st century forest cover change. *Science*, 342, 850-853.
- Harrell, P. A., Bourgeau-Chavez, L. L., Kasischke, E. S., French, N. H. F. and Christensen Jr., N. L. (1995), Sensitivity of ERS-1 and JERS-1 radar data to biomass and stand structure in Alaskan boreal forest. *Remote Sensing of Environment*, 54, 247-260.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.
- Hoekman, D. and Quiñones, M. J. (2000), Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 685-696.
- Hoekman, D. H. and Reiche, J. (2015), Multi-model radiometric slope correction of SAR image of complex terrain using a two-stage semi-empirical approach. *Remote Sensing of Environment*, 156, 1-10.
- Hofton, M. A., Minister, J. B., and Blair, J. B.(2000). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 1989-1996.
- Houghton, R. A., Butman, D., Bunn, A. G., Krankina, O. N., Schlesinger, P. and Stone, T. A. (2007), Mapping Russian forest biomass with data from satellites and forest inventories. *Environmental Research Letters*, 2, 045032.
- Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J. and Guo, Q. (2016), Mapping global forest aboveground

	GlobBiomass		Page 115
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

biomass with spaceborne LiDAR, optical imagery and forest inventory data. *Remote Sensing*, 8, doi:10.3390/rs8070565.

Imhoff, M. L. (1995), Radar backscatter and biomass saturation: Ramifications for global biomass inventory. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 511-518.

IPCC (2006), *IPCC Guidelines for National Greenhouse Gas Inventories*: Institute for Global Environmental Strategies, Japan.

Karam, M. A., Fung, A. K., Lang, R. H. and Chauhan, N. S. (1992), A microwave scattering model for layered vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 767-784.

Kellndorfer, J., Walker, W., LaPoint, E., Bishop, J., Cormier, T., Fiske, G., Hoppus, M., Kirsch, K. and Westfall, J., 2012. NACP Aboveground Biomass and Carbon Baseline Data (NBCD 2000), available on-line at <http://daac.ornl.gov> from ORNL DAAC, Oak Ridge, Tennessee, U.S.A. <http://dx.doi.org/10.3334/ORNLDAAC/1081>,

Kindermann, G. E., McCallum, I., Fritz, S. and Obersteiner, M. (2008), A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*, 42, 387-396.

Kurum, M., Lang, R. H., O'Neill, P. E., Joseph, A. T., Jackson, T. J. and Cosh, M. H. (2009), L-Band radar estimation of forest attenuation for active/passive soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3026-3040.

Kurvonen, L., Pulliainen, J. and Hallikainen, M. (1999), Retrieval of biomass in boreal forests from multitemporal ERS-1 and JERS-1 SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 198-205.

Le Toan, T., Beaudoin, A., Riou, J. and Guyon, D. (1992), Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 403-411.

Lopes, A., Nezry, E. and Touzi, R. (1990), Adaptive filters and scene heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 992-1000.

Los, S. O., Rosette, J., Kljun, N., North, P. R. J., Chasmer, L., Suárez, J. C., Hopkinson, C., Hill, R. A., van Gorsel, E., Mahoney, C. and Berni, J. A. J. (2012), Vegetation height and cover fraction between 60° S and 60° N from ICESat GLAS data. *Geoscientific Model Development*, 5, 413-432.

Lucas, R. M., Cronin, N., Lee, A., Moghaddam, M., Witte, C. and Tickle, P. (2006), Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. *Remote Sensing of Environment*, 100, 407-425.

Lucas, R., Armston, J., Fairfax, R., Fensham, R., Accad, A., Carreiras, J., Kelley, J., Bunting, P., Clewley, D., Bray, S., Metcalf, D., Dwyer, J., Bowen, M., Eyre, T., Laidlaw, M. and Shimada, M. (2010), An evaluation of the ALOS PALSAR L-band backscatter - Above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3, 576-593.

Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., & Seifert-Granzin, J. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sensing of Environment*, 155, 109-119.

Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment*, 159, 307-317.

	GlobBiomass		Page 116
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- Mitchard, E. T. A., Saatchi, S. S., Woodhouse, I. H., Nangendo, G., Ribeiro, N. S., Williams, M., Ryan, C. M., Lewis, S. L., Feldpausch, T. R. and Meir, P. (2009), Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters*, 36, L23401, doi:10.1029/2009GL040692.
- Mitchard, E. T. A., Saatchi, S. S., Baccini, A., Asner, G. P., Goetz, S. J., Harris, N. L. and Brown, S. (2013), Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. *Carbon Balance and Management*, 8:10.
- Moisen, G. G., Frescino, T. S., Huang, C., Vogelmann, J. E. and Zhu, Z. (2003), Predictive modeling of forest cover and tree canopy height in the central Rocky Mountains of Utah. *Proceedings of 2003 Meeting of the American Society of Photogrammetry and Remote Sensing, Anchorage*.
- Nelson, R., Boudreau, J., Gregoire, T. G., Margolis, H., Næsset, E., Gobakken, T. and Ståhl, G. (2009), Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, 39, 862-881.
- Padilla, M., Stehman, S. V., Chuvieco, E. (2014). Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. *Remote Sensing of Environment*, 144, 187-196.
- Picard, G., Le Toan, T., Quegan, S., Caraglio, Y. and Castel, T. (2004), Radiative transfer modeling of cross-polarized backscatter from a pine forest using the discrete ordinate and eigenvalue method. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1720-1730.
- Pierce, L. E., Bergen, K., Dobson, M. C., & Ulaby, F. T. (1998). Multitemporal land-cover classification using SIR-C/X-SAR Imagery. *Remote Sensing of Environment* 64, 20–33.
- Potapov, P., Turubanova, S. A., Hansen, M. C., Adusei, B., Broich, M., Altstatt, A., Mane, L., Justice, C. O. (2012). Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. *Remote Sensing of Environment*, 122, 106-116.
- Praks, J., Antropov, O. and Hallikainen, M. (2012), LIDAR-aided SAR interferometry studies in boreal forest: Scattering phase center and extinction coefficient at X- and L-band. *Remote Sensing*, 50, 3831-3843.
- Pulliainen, J. T., Mikkilä, P. J., Hallikainen, M. T. and Ikonen, J.-P. (1996), Seasonal dynamics of C-band backscatter of boreal forests with applications to biomass and soil moisture estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 34, 758-770.
- Pulliainen, J. T., Heiska, K., Hyyppä, J. and Hallikainen, M. T. (1994), Backscattering properties of boreal forests at the C- and X-bands. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 1041-1050.
- Pulliainen, J. T., Kurvonen, L., Hallikainen, M. T., "Multitemporal behavior of L- and C-band SAR observations of boreal forests," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 927-937, 1999.
- Quegan, S., Yu, J.J. 2001, Filtering of Multichannel SAR Images, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, 2001
- Quinlan, J. R. (1993), *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufmann Publishers Inc., pp.
- Rabus, B., Eineder, M., Roth, A. and Bamler, R. (2003), The Shuttle Radar Topography Mission - A new class of digital elevation models acquired by spaceborne SAR. *ISPRS Journal of Photogrammetry & Remote Sensing*, 57, 241-262.
- Ranson, K. J. and Sun, G. (1994), Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 388-396.

	GlobBiomass		Page 117
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- Ranson, K. J., Saatchi, S. and Sun, G. (1995), Boreal forest ecosystem characterization with SIR-C/XSAR. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 867-876.
- Ranson, K. J. and Sun, G. (2000), Effects of environmental conditions on boreal forest classification and biomass estimates with SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 1242-1252.
- Reese, H., Nilsson, M., Granqvist Pahlén, T., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M. and Olsson, H. (2003), Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. *Ambio*, 32, 542-548.
- Rosenqvist, Å., Shimada, M., Ito, N. and Watanabe, M. (2007), ALOS PALSAR: a pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3307-3316.
- Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S. and Aoki, E. (2014), Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 155, 3-12.
- Saatchi, S. S. and Moghaddam, M. (2000), Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 697-709.
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M. and Morel, A. (2011), Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108, 9899-9904.
- Salas, W. A., Ducey, M. J., Rignot, E. and Skole, D. (2002), Assessment of JERS-1 SAR for monitoring secondary vegetation in Amazonia: I. Spatial and temporal variability in backscatter across chrono-sequence of secondary vegetation stands in Rondonia. *International Journal of Remote Sensing*, 23, 1357-1379.
- Santoro, M., Askne, J., Smith, G. and Fransson, J. E. S. (2002), Stem volume retrieval in boreal forests from ERS-1/2 interferometry. *Remote Sensing of Environment*, 81, 19-35.
- Santoro, M., Eriksson, L., Askne, J. and Schmillius, C. (2006), Assessment of stand-wise stem volume retrieval in boreal forest from JERS-1 L-band SAR backscatter. *International Journal of Remote Sensing*, 27, 3425-3454.
- Santoro, M., Fransson, J. E. S., Eriksson, L. E. B., Magnusson, M., Ulander, L. M. H. and Olsson, H. (2009), Signatures of ALOS PALSAR L-band backscatter in Swedish forest. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 4001-4019.
- Santoro, M. and Cartus, O. (2010), STSE-BIOMASAR: Validating a novel biomass retrieval algorithm based on hyper-temporal Wide-Swath and Global Monitoring Envisat ASAR datasets. Final Report, ESA ESRIN contract No. 21892/08/I-EC, 2010.
- Santoro, M., Beer, C., Cartus, O., Schmillius, C., Shvidenko, A., McCallum, I., Wegmüller, U. and Wiesmann, A. (2011), Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sensing of Environment*, 115, 490-507.
- Santoro, M., Cartus, O., Fransson, J. E. S., Shvidenko, A., McCallum, I., Hall, R. J., Beaudoin, A., Beer, C. and Schmillius, C. (2013), Estimates of forest growing stock volume for Sweden, Central Siberia and Québec using Envisat Advanced Synthetic Aperture Radar backscatter data. *Remote Sensing*, 5, 4503-4532.

	GlobBiomass		Page 118
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- Santoro, M., Wegmuller, U., Fransson, J. E. S. and Schmullius, C. (2014), Regional mapping of forest growing stock volume with multitemporal ALOS PALSAR backscatter. *Proceedings of IGARSS'14, Quebec City, 13-18 July*, (pp. 2313-2316), IEEE Publications, Piscataway, NJ.
- Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J. E. S., Hall, R. J., Pathe, C., Schepaschenko, D., Schmullius, C., Shvidenko, A., Thurner, M. and Wegmüller, U. (2015a), Forest growing stock volume of the northern hemisphere: spatially explicit estimates for 2010 derived from Envisat ASAR data. *Remote Sensing of Environment*, 168, 316-334.
- Santoro, M., Eriksson, L. E. B. and Fransson, J. E. S. (2015b), Reviewing ALOS PALSAR backscatter observations for stem volume retrieval in Swedish forest. *Remote Sensing*, 7, 4290-4317.
- Santoro, M., Wegmuller, U., Lamarche, C., Bontemps, S., Defourny, P. and Arino, O. (2015c), Strengths and weaknesses of multi-year Envisat ASAR backscatter measurements to map permanent open water bodies at global scale. *Remote Sensing of Environment*, 171, 185-201.
- Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C. and Townshend, J. R. (2013), Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6, 427-448.
- Sheen, D. R., Malinas, N. P., Kletzli, D. W., Lewis, T. B. and Roman, J. F. (1994), Foliage transmission measurements using a ground-based ultrawide band (300-1300 MHz) SAR system. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 118-130.
- Shimada, M. (2010), Ortho-rectification and slope correction of SAR data Using DEM and its accuracy evaluation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3, 657-671.
- Shimada, M., & Ohtaki, T. (2010). Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. *Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE Journal of, 3(4), 637-656.
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraiishi, T., Thapa, R. and Lucas, R. (2014), New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155, 13-31.
- Shinohara, H., Homma, T., Nohmi, H., Hirose, H. and Tagawa, T. (1992), Relation between L-band microwave penetration / backscattering characteristics and state of trees. *Proceedings of IGARSS'92, Houston, 26-29 May*, (pp. 539-541), IEEE Publications, Piscataway, NJ.
- Simard, M., Pinto, N., Fisher, J. B. and Baccini, A. (2011), Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research - Biogeosciences*, 116, G04021.
- Simons, H. (2001), FRA 2000. Global ecological zoning for the Global Forest Resources Assessment 2000. FRA Working Paper 56, FAO, Rome.
- Small, D. (2011), Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3081-3093.
- Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. and Schmullius, C. (2014), Carbon stock and density of northern boreal and temperate forests. *Global Ecology and Biogeography*, 23 (3), 297-310.
- Tomppo, E., Katila, M., Mäkisara, K. and Peräsaari, J. (2013), The Multi-source National Forest Inventory of Finland – methods and results 2009. *Metlan työraportteja / Working Papers of the Finnish Forest Research Institute*, 273, 216 pages. ISBN 978-951-40-2428-3 (PDF), available at <http://www.metla.fi/julkaisut/workingpapers/2013/mwp273.htm>.

	GlobBiomass		Page 119
	v 03		
	ATBD / DJF	Global Biomass Map	Date 16-Oct-17

- Twedt, D. J., Ayala, A. and Shickel, M. R. (2015), Leaf-on canopy closure in broadleaf deciduous forests predicted during winter. *Forest Science*, 61, 926-931.
- Ulaby, F. T., Whitt, M. W. and Dobson, M. C. (1990), Measuring the propagation properties of a forest canopy using a polarimetric scatterometer. *IEEE Transactions on Antennas and Propagation*, 38, 251-258.
- Ulander, L. M. (1996). Radiometric slope correction of synthetic-aperture radar images. *Geoscience and Remote Sensing, IEEE Transactions on*, 34(5), 1115-1122.
- USGS, 2001. 7.5-Min DEM Native Format of the United States, available at <http://www.webgis.com/>, last accessed December 2014.
- Villard, L. (2009). *Forward & Inverse Modeling for Synthetic Aperture Radar Observables in Bistatic Configuration. Applications in Forest Remote Sensing* (Doctoral dissertation, Institut Supérieur de l'Aéronautique et de l'Espace).
- Walker, W. S., Kelndorfer, J. M., LaPoint, E., Hoppus, M. and Westfall, J. (2007), An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, 109, 482-499.
- Wang, Y., Day, J. L. and Davis, F. W. (1998), Sensitivity of modeled C- and L-band radar backscatter to ground surface parameters in loblolly pine forest. *Remote Sensing of Environment*, 66, 331-342.
- Way, J., Rignot, E. J. M., McDonald, K. C., Oren, R., Kwok, R., Bonan, G., Dobson, M. C., Viereck, L. A. and Roth, J. E. (1994), Evaluating the type and state of Alaska taiga forests with imaging radar for use in ecosystem models. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 353-370.
- Wegmüller, U. (1999), Automated terrain corrected SAR geocoding. Proceedings of IGARSS'99, Hamburg, 28 June - 2 July, (pp. 1712-1714), IEEE Publications, Piscataway, NJ.
- Wiesmann, A., Wegmüller, U., Santoro, M., Strozzi, T. and Werner, C. (2004), Multi-temporal and multi-incidence angle ASAR Wide Swath data for land cover information. Proceedings of 4th International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, Innsbruck, 16-19 November, CD-ROM, (pp.