

CCI
BIOMASS

ALGORITHM THEORETICAL BASIS DOCUMENT
YEAR 2
VERSION 1.0

DOCUMENT REF:	CCI_BIOMASS_ATBD_V2
DELIVERABLE REF:	D2.2-ATBD
VERSION:	1.0
CREATION DATE:	2020-01-06
LAST MODIFIED	2020-01-06

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	2	14.02.2019	

Document Authorship

	NAME	FUNCTION	ORGANISATION	SIGNATURE	DATE
PREPARED	Maurizio Santoro		GAMMA		
PREPARED	Oliver Cartus		GAMMA		
PREPARED	Heather Kay		Aberystwyth Uni		
PREPARED	Richard Lucas		Aberystwyth Uni		
PREPARED					
VERIFIED	S. Quegan	Science Leader	Sheffield University		
APPROVED					

Document Distribution

ORGANISATION	NAME	QUANTITY
ESA	Frank Seifert	

Document History

VERSION	DATE	DESCRIPTION	APPROVED
0.1	2020-01-06	First draft version	

Document Change Record (from Year 1 to Year 2)

VERSION	DATE	DESCRIPTION	APPROVED

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	3	14.02.2019	

Table of Contents

<i>list of figures</i>	5
<i>List of tables</i>	10
<i>Symbols and acronyms</i>	11
1 Introduction	13
2 Background	15
2.1 Theory behind algorithms for global biomass retrieval	15
2.2 The GlobBiomass biomass dataset	17
2.3 Moving from the GlobBiomass to the CCI Biomass CORE algorithm	19
3 Datasets	20
3.1 Sentinel-1 (C-band, wavelength 5.6 cm)	20
3.2 ALOS-2 PALSAR-2 (L-band, wavelength 23 cm)	28
3.3 Envisat ASAR (C-band, wavelength 5.6 cm)	39
3.4 ALOS-1 PALSAR-1	44
3.5 ICESAT GLAS	46
3.6 Digital Elevation Model	48
3.7 MODIS Vegetation Continuous Fields	48
3.8 Landsat canopy density and density change	49
3.9 CCI Land Cover map	50
3.10 Worldclim Bioclimatic Variables	51
3.11 FAO Global Ecological Zones dataset	51
3.12 Terrestrial Ecoregions of the World	52
3.13 Global forest canopy height	53
3.14 Biomass of dense forest and maximum biomass	53
3.15 Biomass Conversion and Expansion Factor	55
3.16 Allometries	57
3.16.1 Allometry between canopy density and height	57
3.16.2 Allometry between AGB and height.....	63
4 Methods	69
4.1 The GlobBiomass global biomass retrieval algorithm	69
4.2 The CCI Biomass CORE algorithm	71
4.3 The BIOMASAR-C algorithm	74
4.3.1 BIOMASAR-C for GSV retrieval	76
4.3.2 BIOMASAR-C for AGB retrieval.....	83
4.4 The BIOMASAR-L algorithm	89
4.4.1 BIOMASAR-L for GSV retrieval	90
4.4.2 BIOMASAR-L for AGB retrieval.....	98

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	4	14.02.2019	

4.5	Methods to derive a merged biomass dataset.....	102
4.5.1	C- and L-band sensitivity to biomass	103
4.5.2	Number of observations	104
4.5.3	Local errors in the model calibration and inversion.....	104
4.5.4	Topographic effects	107
4.5.5	Total weight for merging.....	108
4.6	GSV to AGB conversion	108
5	Results.....	108
5.1	BIOMASAR-C	108
5.1.1	Comparing the performance of BIOMASAR-C in CCI CORE y1 with respect to GlobBiomass ...	108
5.1.2	Comparing BIOMASAR-C versions in CCI CORE algorithms	112
5.2	BIOMASAR-L.....	115
5.2.1	Comparing the performance of BIOMASAR-L in CCI CORE y1 with respect to GlobBiomass....	115
5.2.2	Comparing BIOMASAR-L versions in CCI CORE algorithms	118
5.3	Merging GSV estimates.....	121
6	<i>Status-quo of the CORE algorithm and potential areas for improvement</i>	122
7	References.....	124
8	Annex A	131
8.1	A.1 Alpine terrain, temperate forests	131
8.2	A2 Flat to hilly terrain, tropical forest.....	136
8.3	Annex B.....	141

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	5	14.02.2019	

LIST OF FIGURES

Figure 1-1: Latitudinal averages of AGB estimates from the GlobBiomass project, Saatchi et al. (2011), Baccini et al. (2012), Avitabile et al. (2016) and GEOCARBON dataset, Hu et al. (2016), Thurner et al. (2014), Liu et al. (2015) and Kindermann et al. (2008).	14
Figure 2-1: The GlobBiomass AGB dataset (Santoro et al., 2018a)	18
Figure 3-1: Observation geometry of the Sentinel-1 mission last accessed on 9 February 2019 (https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario).	21
Figure 3-2: Coverage of the 2017 Sentinel-1 IWS dataset selected to support the estimation of biomass for the 2017-2018 epoch (courtesy J. Kellndorfer, EBD).	23
Figure 3-3: Flowchart of the Sentinel-1 data pre-processing	25
Figure 3-4: Estimates of ENL for 35 polygons distributed over five Sentinel-1 VV-polarized images randomly selected in boreal, temperate and tropical environments.	26
Figure 3-5: Number of observations per pixel for the Sentinel-1 2017 dataset.	27
Figure 3-6: False colour composite of the Sentinel-1 2017 dataset. Red: temporally averaged co-polarized backscatter; green: temporally averaged cross-polarized backscatter; blue: ratio of the temporally averaged cross-polarized and co-polarized backscatter. Pixel size: 150 m × 150 m.	28
Figure 3-7: ALOS-2 FBD mosaics, HV-polarization, for the years 2015 (top), 2016 (middle) and 2017 (bottom).	30
Figure 3-8: ALOS-2 ScanSAR mosaic generated from HV polarization imagery acquired in April 2018 over the Amazon Basin.	31
Figure 3-9: Number of ALOS-2 FBD and ScanSAR observations available in the time frame 2015-2017.	32
Figure 3-10: Geolocation offset between ALOS-2 ScanSAR and ALOS PALSAR FBD mosaic for the year 2010 determined on a 1°x1° tile basis using image cross-correlation.	33
Figure 3-11: Geolocation error of ALOS-2 ScanSAR mosaics as a function of the EGM96-Geoid/WGS84 elevation offset. Only scenes acquired between 20°S and 20°N were considered.	33
Figure 3-12: Effect of co-registration of a 1x1 degree tile of L-band HV backscatter acquired over central Germany with Landsat NDVI imagery.	34
Figure 3-13: Effect of co-registration of a 1x1 degree tile of L-band HV backscatter acquired over the North-western United States with Landsat NDVI imagery. 1.5 pixels (~150) shift in range.	35
Figure 3-14: ALOS-2 L-HV backscatter mosaic for Southern Sweden (top) and Northern Asia (bottom).	36
Figure 3-15: SMAP and ALOS-2 LHV backscatter mosaic for a single pixel line over the Amazon and second order polynomials fitted in moving windows to the observed trend.	37
Figure 3-16: SMAP LHV backscatter mosaic (top), ALOS-2 L-HV mosaic before (top) and after (bottom) normalization with respect to SMAP.	38
Figure 3-17: Map of ASAR acquisitions at 150 m (WSM and IMM combined) and 1,000 m (GMM) per 1°x1° grid-cell for each year between 2009 and 2011. The colour bar has been constrained between 0 and 30 observations to increase the image contrast in regions of poor coverage.	40
Figure 3-18: Flowchart of ASAR pre-processing.	41
Figure 3-19: 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 showed in the panel on the right hand-side.	44
Figure 3-20: Example of the ASAR GM1 used in Figure 3-19 before and after filtering with a moving median filter.	44
Figure 3-21: ALOS-1 PALSAR-1 mosaic of HV images acquired during 2010.	45

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	6	14.02.2019	

Figure 3-22: An ICESAT GLAS waveform showing the vertical distribution of returned energy from a forest (from Los et al., 2012).	46
Figure 3-23: An ICESat GLAS waveform, showing the waveform metrics used to calculate RH100 & CD (Hilbert & Schmullius, 2012)	47
Figure 3-24: Number of ICESAT GLAS footprints after screening of GLA14 product.	47
Figure 3-25: Global seamless DEM based on elevation datasets available at http://www.viewfinderpanoramas.org .	48
Figure 3-26: Canopy density map for the year 2010 produced from Landsat.	49
Figure 3-27: Illustration of CCI Land Cover maps (https://www.esa-landcover-cci.org)	50
Figure 3-28: FAO GEZ dataset	52
Figure 3-29: Terrestrial ecoregions of the world (Olson et al., 2001).	52
Figure 3-30: Grouping of the terrestrial ecoregions into biomes (see legend) and realms (see map). Picture reproduced from Olson et al. (2001).	53
Figure 3-31: Map detailing the origin of the numbers used to generate an estimate of the biomass of dense forests at the spatial scale of 2°.	54
Figure 3-32: Map of the GSV of dense forests with a spatial resolution of 0.2°.	55
Figure 3-33: Global estimates of wood density. Pixel size: 0.0083333°, i.e., 30 arc-seconds.	56
Figure 3-34: Map of BCEF applied to the stem biomass dataset obtained by multiplying the merged GSV dataset from the GlobBiomass project with the map of wood density in Figure 3-29. Pixel size: 0.0083333°, i.e., 30 arc-seconds.	56
Figure 3-35: Map of q values per WWF ecoregion, obtained through least squares regression of available ICESat GLAS footprints within each ecoregion after filtering, ecoregions in white had no footprints	58
Figure 3-36: Scatter plot of summer v winter q values globally, red line indicates 1:1 line	59
Figure 3-37: Examples of least squares regression denoted by blue line (extended to 60m height for comparison). Data was binned to 2m intervals with blue dots indicating mean of each bin and vertical line the interquartile range. Ecoregions showing good fit are Western Congolian forest-savanna mosaic (a) and Central Range montane forests (b). C represents Atlantic dry forests and shows good fit of the means despite wide spread of data and heights below 20m, d shows good fit of the means despite 'noisy' data in Okhotsk-Manchurian taiga.	60
Figure 3-38: Examples of least squares regression denoted by blue line (extended to 60m height for comparison). Data was binned to 2m intervals with blue dots indicating mean of each bin and vertical line the interquartile range. Left hand plot from Trans-Baikal conifer forests demonstrates the sinusoidal pattern found in some ecoregions. The right-hand plot highlights the issue of large variation within the ecoregion, here for Eastern cascades forests.	60
Figure 3-39: Histogram of the distribution of q values in different biomes, biomes were selected and regrouped from the WWF Ecoregions classification with; 'Savanna' representing Temperate, Montane and Flooded Grasslands, Shrublands and Savanna; 'Tropics' representing Tropical & Subtropical Moist & Dry Broadleaf and Coniferous Forests; 'Boreal' representing Boreal & Taiga and 'Tundra' representing the Tundra biome.	61
Figure 3-40: Map showing Patagonian Steppe ecoregion (highlighted in yellow) with ICESat GLAS footprints represented by the dots. This ecoregion has a high q value for a grassland (0.079) obtained from 676 footprints	62
Figure 3-41: Histogram of the difference between ICESat GLAS waveform length and h100 canopy height estimates for filtered footprints	62
Figure 3-42: Map of the coefficient q derived from all ICESAT GLAS measurements.	63
Figure 3-43: Illustrating the relationship between AGB and RH100 using the GlobBiomass AGB dataset and the ICESAT GLAS RH100 measurements averaged to 0.25° (blue crosses). The red curves represent the model fit to the AGB and RH100 values using Equation (3-6). Each panel represents a major FAO ecological zone (rows) and is characterized by a minimum number of	

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	7	14.02.2019	

footprints per 0.25° pixel. The estimates of the coefficients p_1 and p_2 are listed in Table 3-5 and Table 3-6, respectively. 66

Figure 3-44: Model fits of Equation (3-6) per major ecological zone and continent at 0.25° spatial resolution. The dashed lines represent the model fits obtained for an entire ecological zone (number of footprints > 200 except for the temperate dry ecozone with a minimum number of footprints per pixel of at least 100). 68

Figure 3-45: Comparing power-law (red lines) height-to-AGB allometries developed in literature (Santoro et al., 2002; Askne and Santoro, 2012; Asner et al., 2012; Asner and Mascaro, 2014; Coomes et al., 2017; Labriere et al., 2018; Dal Ponte et al., 2019) with own predictions from Equation (3-6) (blue line) on an ecoregion basis. The coefficients of the power-law functions are displayed on the top of each panel. 68

Figure 3-46: Spatially interpolated maps of the coefficients p_1 and p_2 . The maps are here represented in the form of quick-looks with a 0.25° pixel size. 69

Figure 4-1: Flowchart of the GlobBiomass global biomass retrieval algorithm. 70

Figure 4-2: Functional dependencies of datasets and approaches forming the CCI Biomass CORE global biomass retrieval algorithm in year 2 (top) and year 1 (bottom). Text in red visualizes modifications introduced from year 1 to year 2. The shaded part of the flowchart represents potential improvements following the implementation of additional retrieval techniques. 74

Figure 4-3: Flowchart of model training and retrieval implemented in the BIOMASAR-C approach 76

Figure 4-4: Flowchart illustrating the procedure to estimate σ_{gr}^0 and σ_{df}^0 in BIOMASAR-C for the case of a Sentinel-1 backscatter image. 78

Figure 4-5: Illustrating the gap-filling procedure in the case of incomplete fields of σ_{gr}^0 estimates for a Sentinel-1 image. Each pixel represents the estimate for a given tile covered by the Sentinel-1 image. Red is used for pixels for which no estimate has been found. 79

Figure 4-6: Illustrating the gap-filling procedure in the case of incomplete fields of σ_{df}^0 estimates for a Sentinel-1 image. Each pixel represents the estimate for a given tile covered by the Sentinel-1 image. Red is used for pixels for which no estimate has been found. 80

Figure 4-7: Panels with incidence angle range as title show (i) estimates of σ_{0gr} and σ_{0veg} obtained with the combined approach (black asterisks) and with self-calibration only (red asterisks), (ii) Equation (4-8) in the case of the combined approach (black solid curve) and self-calibration only (red dashed curve), and (iii) the median backscatter for a given canopy density (circle). The panel relating backscatter to incidence angle shows the estimates of σ_{gr}^0 (black circles) and σ_{veg}^0 (red asterisks) obtained with the combined approach and their quadratic fits (black curve for σ_{gr}^0 and red curve for σ_{veg}^0) spanning the range of incidence angles between 0° and 90°. Dataset: Sentinel-1, VH-polarization., acquired on 5 April 2017. Tile (top left corner coordinate): 64°N, 30°E (boreal forest). 85

Figure 4-8: Same as in Figure 4-7. Dataset: Sentinel-1, VH-polarization., acquired on 5 July 2017. Tile (top left corner coordinate): 46°N, 11°E (temperate forest). 86

Figure 4-9: Same as in Figure 4-7. Dataset: Sentinel-1, VH-polarization., acquired on 8 July 2017. Tile (top left corner coordinate): 0°N, 11°E (wet tropics). 86

Figure 4-10: Same as in Figure 4-7. Dataset: Sentinel-1, VH-polarization., acquired on 6 July 2017. Tile (top left corner coordinate): 9°S, 17°E (miombo woodlands). 87

Figure 4-11: Illustrating the raster images of the estimates of σ_{gr}^0 and σ_{veg}^0 (bottom row) for the tile used in Figure 4-9. The top row shows the image of the SAR backscatter and the image of the local incidence angle. 88

Figure 4-12: AGB retrieval rules depending on the level of the backscatter measurement. 89

Figure 4-13: 1°×1° 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, grey, and black) have been used. 91

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	8	14.02.2019	

Figure 4-14: Histograms of L-HV backscatter in areas of low (red) and high (green) canopy density according to Landsat. 92

Figure 4-15: 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. 93

Figure 4-16: 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⁻¹ (bottom), respectively. 94

Figure 4-17: 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. L-band derived estimates of the GSV are compared to regional maps (Kellndorfer et al., 2013; Reese et al., 2002; Bouvet et al., 2018) which had been produced using and radar/optical imagery and local forest inventory data for calibrating retrieval models. 95

Figure 4-18: Estimates for β (incl. 95% confidence bounds) obtained for different FAO eco-regions by fitting the model in Equation (4-13) to the observed relationship of forest transmissivity, simulated with the aid of ICESAT GLAS and the corresponding GSV in map products. 97

Figure 4-19: Estimates of the forest transmissivity coefficient β [ha/m³] for L-band. 97

Figure 4-20: Simulated differences in the L-band HV backscatter response to changes in η for forests with canopy heights between 5 and 35 m. 99

Figure 4-21: ALOS-2 LHV backscatter as function of Landsat canopy density (dots) for three acquisitions covering tropical forest in Gabon (top left), the Swiss Alps (top right), and the Pacific Northwest of the United States (bottom left). The curves represent the fit of the model in Equation (4-14) for backscatter observations in different local incidence angle ranges. Derived estimates for α are reported for each region. 100

Figure 4-22: Estimates for the model parameters σ_{gr}^0 (red dots) and σ_{veg}^0 (green dots) as function of the local incidence angle for ALOS-2 orbits covering the Swiss Alps (left) and the Pacific Northwest of the United States (right). The curves represent the corresponding fit of a second order polynomial. 101

Figure 4-23: Spatially explicit estimates for the model parameters σ_{gr}^0 and σ_{veg}^0 for a 1x1° ALOS-2 tile (year 2016) covering the Swiss Alps. 101

Figure 4-24: Forest transmissivity modelled as a function of GSV for two different values of the forest transmissivity coefficient, β (left). Derivative of Equation (4-13) for the two values of β (right). Blue indicates L-band and red C-band. 104

Figure 4-25: Forest transmissivity modelled as a function of canopy cover, forest height vs. BIOMASAR-C and BIOMASAR-L GSV for different FAO ecoregions in South America. The curves represent the fit of Equation (4-10) to the observations. 106

Figure 4-26: Example of weights applied to the BIOMASAR-L dataset, which reflects how consistent the GSV estimates are with the transmissivity modelled with the aid of optical canopy density and height maps. 107

Figure 4-27: Weight for the BIOMASAR-L map reflecting the steepness of the terrain. 108

Figure 5-1: Comparing GSV estimated with BIOMASAR-C using Envisat ASAR data with a 1,000 m pixel size (left) and Sentinel-1 data with a pixel size of 150 m (right). The colour scheme of the GSV estimates is the same as in Figure 5-2. The maps cover an area between 56° and 57° N and 45° and 47° E. 109

Figure 5-2: Example of GSV estimates obtained with BIOMASAR-C applied to Sentinel-1 dataset acquired in 2017. 110

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	9	14.02.2019	

Figure 5-3: Estimates of GSV for an area south of the Baikal Lake, Siberia (left) and optical image from Google Earth (right). The colour scheme of the GSV estimates is the same as in Figure 5-2. 110

Figure 5-4: Map of GSV obtained with the BIOMASAR-C algorithm applied to the multi-temporal dataset of Sentinel-1 backscatter observations of 2017. Pixel size: 150 m. 111

Figure 5-5: Scatterplot comparing provincial averages of GSV from the Forest State Account and from the BIOMASAR-C algorithm applied to the Sentinel-1 dataset of 2017. The vertical bars represent the standard deviation of the Sentinel-1 GSV per province. 111

Figure 5-6: Comparing GSV estimated with BIOMASAR-L applied to ALOS-2 data (left) and with BIOMASAR-C applied to Sentinel-1 data (right). The central part of the map in the left panel shows a checkerboard pattern, corresponding to different ALOS-2 images used in the process of compositing the mosaic. In contrast, the Sentinel-1 map has uniform appearance. 112

Figure 5-7: Maps of AGB obtained with the BIOMASAR-C algorithm described in Section 4.3.1 (y1) and in Section 4.3.2 (y2) from the 2017 dataset of Sentinel-1 images. The scatter plot compared AGB estimates from the two maps for a subset of the pixels. The histograms represent the distribution of AGB for (i) a given version of BIOMASAR-C and (ii) a given quadrant of terrain orientation (see legend). The histograms are limited to pixels with a terrain slope larger than 20° to better appraise the impact of topography on the AGB retrieval. Tile (top left corner coordinate): 2°S, 35°E. 113

Figure 5-8: Same as in Figure 5-7. Tile (top left corner coordinate): 1°N, 29°E. 114

Figure 5-9: Same as in Figure 5-7. Tile (top left corner coordinate): 0°N, 11°E. 115

Figure 5-10: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD and ScanSAR imagery acquired between 2015 and 2018. 116

Figure 5-11: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD imagery acquired over Central Europe between 2015 and 2017. 116

Figure 5-12: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD and ScanSAR imagery acquired over Indonesia and Papua between 2015 and 2018. 116

Figure 5-13: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD imagery acquired over Western Siberia between 2015 and 2017 (top left) and the corresponding subset of the ALOS-2 FBD mosaic for the year 2017 (top right). The bottom plot illustrates the estimates for the model parameter σ_{df}^0 across Siberia for the year 2017. 117

Figure 5-14: Scatterplots of GSV produced from multi-temporal ALOS-2 data and the BIOMASAR-L map for the year 2010 in the frame of the GlobBiomass project. For the comparison, all maps were aggregated to 1 km pixel scale. The examples refer to 5°x5° or 10°x10° large areas in Siberia (top left), Amazon (top right), Central Europe (centre left), South Africa (centre right), Indonesia (bottom left), and Congo Basin (bottom right). 118

Figure 5-15: AGB maps for areas in Central Europe (top row), the Pacific Northwest of the United States (centre row), and Gabon (bottom row) with BIOMASAR-L (y1) i.e., the CORE algorithm deployed in year 1 (left), and with BIOMASAR-L (y2) (right). 120

Figure 5-16: Comparison of AGB maps for areas in the Pacific Northwest of the United States, Central Europe, and Gabon produced with BIOMASAR-L (y1), i.e., the CORE algorithm deployed in year 1, BIOMASAR-L (y2). 121

Figure 5-17: Scatterplots of Sentinel-1 (left), ALOS-2 (centre), and the merged Sentinel-1/ALOS-2 GSV maps against the GlobBiomass GSV map for 2010. 122

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	10	14.02.2019	

LIST OF TABLES

Table 1-1: Reference Documents	13
Table 3-1: Remote sensing data sources, epochs covered and use.	20
Table 3-2: Geographical distribution of the Sentinel-1 data pool used to support estimation of biomass for the 2017-2018 epoch. The coordinates represent the extent of each region.	23
Table 3-3: ALOS-2 acquisition cycles for which mosaics of dual-polarization backscatter observations acquired in ScanSAR mode have been released by JAXA.	31
Table 3-4: Overview of BioClim variables.	51
Table 3-5: Estimates of the coefficient p_1 per ecoregion and minimum number of GLAS footprints within a 0.25° pixel.	66
Table 3-6: Estimates of the coefficient p_2 per ecoregion and minimum number of GLAS footprints within a 0.25° pixel.	66

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	11	14.02.2019	

SYMBOLS AND ACRONYMS

ADP	Algorithm Development Plan
AGB	Above-Ground Biomass
ALOS	Advanced Land Observing Satellite
APM	Alternating Polarization Medium
ASAR	Advanced Synthetic Aperture Radar
ASF	Alaska Satellite Facility
ATBD	Algorithm Theoretical Basis Document
BCEF	Biomass Conversion & Expansion Factor
BEF	Biomass Expansion Factor
CCI	Climate Change Initiative
CCI-Biomass	Climate Change Initiative – Biomass
CCI-LC	Climate Change Initiative – Land Cover
DARD	Data Access Requirements Document
DEM	Digital Elevation Model
E3UB	End to End ECV Uncertainty Budget
ECV	Essential Climate Variables
ENL	Equivalent Number of Looks
ENVISAT	ESA Environmental Satellite
EO	Earth Observation
ESA	European Space Agency
EWS	Extended Wide Swath mode
FAO	Food and Agriculture Organization
FB	Fine Beam
FBD	Fine Beam Dual
FRA	Forest Resources Assessment
G-POD	Grid Processing On Demand
GCOS	Global Climate Observing System
GDAL	Geospatial Data Abstraction Library
GEDI	Global Ecosystems Dynamics Investigation
GEZ	Global Ecological Zones
GLAS	Geoscience Laser Altimeter System
GLCF	Global Land Cover Facility
GMM	Global Monitoring Mode
GRD	Ground Range Detected
GSV	Growing Stock Volume
HOME	Height Of Median Energy
HH	Horizontal-Horizontal

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	12	14.02.2019	

HV	Horizontal-Vertical
ICESAT GLAS	Ice, Cloud, and land Elevation Satellite Geoscience Laser Altimeter System
IIASA	International Institute of Applied Systems Analysis
IMM	Image Mode Medium
IPCC	Intergovernmental Panel on Climate Change
IWS	Interferometric Wide Swath
JAXA	Japan Aerospace Exploration Agency
LUT	Look Up Table
MCH	Mean Canopy Height
MERIS	Medium Resolution Imaging Spectrometer
MLI	Multi Looked Intensity
MODIS	Moderate Resolution Imaging Spectroradiometer
NaN	Not a Number
PALSAR	Phased Array type L-band Synthetic Aperture Radar
PSD	Product Specification Document
PVASR	Product Validation and Algorithm Selection Report
RMSE	Root Mean Squared Error
SAR	Synthetic Aperture Radar
SLC	Single Look Complex
SMAP	Soil Moisture Active Passive
SRTM	Shuttle Radar Topography Mission
URD	User Requirements Document
USGS	United States Geological Survey
VCF	Vegetation Continuous Fields
VOD	Vegetation Optical Depth
WB	Wide Beam
WSM	Wide Swath Mode

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	13	14.02.2019	

Table 1-1: Reference Documents

ID	Title	Issue	Date
RD-1	Users Requirements Document		
RD-2	Product Specification Document		
RD-3	Data Access Requirements Document		
RD-4	Product Validation and Algorithm Selection		
RD-5	End to End ECV Uncertainty Budget		
RD-6	Algorithm Development Plan		
RD-7	Product Validation Plan		
RD-8	Algorithm Theoretical Basis Document of GlobBiomass project		
RD-9	Product Validation Report		

1 Introduction

Above-ground biomass (AGB, units: Mg ha⁻¹) is defined by the Global Carbon Observing System (GCOS) as one of 50 Essential Climate Variables (ECV). For climate science communities, AGB is a pivotal variable of the Earth System, as it impacts the surface energy budget, the land surface water balance, the atmospheric concentration of greenhouse gases and a range of ecosystem services. The requirement is for AGB to be provided wall-to-wall over the entire globe for all major woody biomes, with a spatial resolution between 500 m and 1 km (based on satellite observations of 100-200 m) spatial resolution, with a relative error of less than 20% where AGB exceeds 50 Mg ha⁻¹ and a fixed error of 10 Mg ha⁻¹ where the AGB is below that limit.

The increased availability of remote sensing imagery during the last 20 years allowed the generation of multiple wall-to-wall datasets of AGB. The uncertainty in magnitude and distribution of AGB is illustrated in (Figure 1-1), where each line represents latitudinal averages of AGB estimated with remote sensing data. While the overall gradient in AGB spatial distribution appears to be consistent across the AGB datasets, the variability of AGB among these datasets is, on average, more than 100% (accuracy figures here excluded). While it is acknowledged that remote sensing is the only tool that can provide global spatially explicit estimates of AGB, the large discrepancies observed in Figure 1-1 are due to the fact that AGB can only be inferred from observations since remote sensing instruments do not have the capability to measure the weight of trees. Yet, as remote sensing observations and *in situ* observations increase and improve the characterization of “biomass”, there are substantial margins to improve the accuracy of the estimates.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	14	14.02.2019	

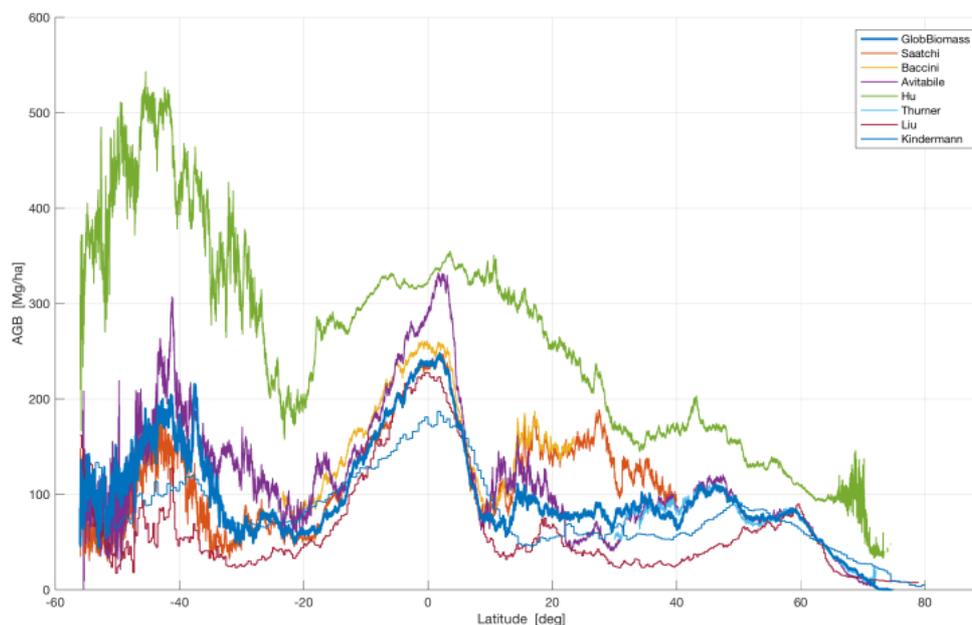


Figure 1-1: Latitudinal averages of AGB estimates from the GlobBiomass project, Saatchi et al. (2011), Baccini et al. (2012), Avitabile et al. (2016) and GEOCARBON dataset, Hu et al. (2016), Thurner et al. (2014), Liu et al. (2015) and Kindermann et al. (2008).

The objectives of the European Space Agency's (ESA) Climate Change Initiative (CCI) Biomass project are to generate global maps of AGB using a variety of Earth Observation (EO) datasets and state-of-the-art models for three epochs (2017-2018, 2018-2019 and 2010) and to assess biomass changes between years and over 10 years. The maps should be spatially and temporally consistent; in addition, they need to be consistent with other data layers thematically similar to the AGB dataset that are produced in the framework of the CCI Programme (e.g., Fire, Land Cover, Snow etc.).

The scope of this document is to present the algorithms implemented to generate the AGB products and the corresponding maps of AGB changes. This Algorithm Theoretical Basis Document (ATBD) relies on indications in the Users Requirements Document (URD) [RD-1], the Product Specifications Document (PSD) [RD-2] and the Data Access Requirements Document (DARD) [RD-3]. In addition, it elaborates on major inputs from the Product Validation and Algorithm Selection (PVASR) document [RD-4], which investigates potential ways to improve the biomass estimated with the algorithms described in this ATBD.

While the ATBD describes the data and algorithms used to generate the global biomass and biomass change products as specified above, the End to End ECV Uncertainty Budget (E3UB) document describes the procedures implemented to quantify the accuracy of the AGB estimates [RD-5]. Future advances that may potentially be implemented in revisions of this ATBD are described in the Algorithm Development Plan (ADP) [RD-6].

During year 1 of the project, methods were developed that led to the generation of a first version of a global AGB product for the epoch 2017-2018. During year 2 of the project, methods developed in year 1 have been refined by taking into account the assessment of the AGB map of 2017-2018 and alternative algorithmic advances documented in the PVASR and in the ADP. The overall objective of the ATBD for year 2 is to document the algorithms implemented to generate AGB estimates for the epochs 2017-2018, 2018-2019 and 2010, with special consideration of

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	15	14.02.2019	

temporal consistency of estimates. An outlook towards potential improvements to be accounted for in the ATBD for the year 3 (e.g., future data, new algorithmic advances, assessment of biomass change across epochs) is provided at the end of this report.

Section 2 provides the background of this ATBD, describing the strategy that underpins the algorithms implemented in CCI Biomass to estimate AGB. Section 3 describes the datasets (EO and auxiliary) used to estimate AGB. The AGB retrieval methods used to generate global maps of AGB for the epochs 2017-2018, 2018-2019 and 2010 are presented in Section 4. An assessment of the retrieval is presented in Section 5 together with a brief outlook on possible advances to be pursued in year 3.

2 Background

2.1 Theory behind algorithms for global biomass retrieval

Thanks to the increasing amount of spaceborne EO data, methods and models that allow estimation of forest variables are being developed with the scope of achieving a global portrait of forest biomass. Below, we briefly outline strengths and weaknesses of algorithms published in scientific journals that led to the generation of a global dataset of a forest variable from remote sensing observations. This list is not meant as an evaluation of the data products but rather to state where past experiences can be of use in enhancing or designing retrieval algorithms based on current EO data.

The availability of global and repeated observations first by the MODIS satellites and more recently by Landsat fostered the estimation of global rasters of canopy height (Lefsky et al., 2010; Simard et al., 2011) and AGB (Baccini et al., 2012; Hu et al., 2016) starting from ICESAT GLAS waveform data. Relationships between ICESAT GLAS waveform metrics were established with respect to *in situ* observations where available and ICESAT GLAS metrics were related to observations by optical sensors (MODIS or Landsat) at pixels corresponding to the ICESAT GLAS footprints. Canopy height and AGB were then extrapolated to the remaining pixels of the optical datasets to obtain wall-to-wall datasets. Even though these methods implement some measurements of the canopy height and AGB (where used), they nonetheless assume that the estimation of canopy height does not require predictors other than MODIS-derived observables, which is questionable since MODIS observables are not a direct measurement of a forest structural parameter. In addition, they rely on a dataset of *in situ* measurements to establish the functional dependency between “true” and “LiDAR-based” height; since such datasets are not available globally, there is a risk that the quality of the estimates is not consistent, being more prone to errors in regions under-represented in the database of *in situ* measurements.

In an attempt to reduce errors in individual maps of AGB, Avitabile et al., (2016) proposed a technique to fuse maps based on the level of agreement of each map with reference AGB measurements. This approach, applied to two pan-tropical maps (Baccini et al., 2012; Saatchi et al., 2011), generated a third pan-tropical map that was then combined with a map of AGB for the boreal and temperate zones (Thurner et al., 2014) to obtain a global map referred to as the GEO-CARBON map. The strength of such an approach is, in our opinion, also its weakness, in the sense that the method is insufficiently constrained in regions where reference datasets (*in situ*, laser-based) are unavailable. Having entered an epoch that can be considered data-rich in terms of spaceborne observations, the demand on reference datasets has also increased and, accordingly, their availability. Hence, efforts should be spent on development of retrieval algorithms that

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	16	14.02.2019	

integrate reference and EO data rather than attempting to fuse estimates from different sources that may ultimately lead to aggregation of errors rather than providing an overall high-quality result.

Compared to optical observations, data acquired in the microwave part of the spectrum contains more information on forest structure because of the weaker attenuation of microwaves by the canopy (passive sensors) or the deeper penetration of microwaves into the canopy (active sensors).

We assume that the potential of such approaches can be enhanced by relating these to additional EO datasets, in particular datasets that contain information about forest structure from active microwave systems (SAR-based observations). Furthermore, the more frequent sampling of the Earth by recent and future LiDAR systems (ICESAT-2, GEDI, MOLI) should allow a larger proportion of the estimated biomass to be explained in terms of the waveform-based measurements, thus in theory leading to a more accurate set of estimates.

When expanding to remote sensing data at coarse resolution, attention needs to be paid to the Vegetation Optical Depth (VOD) observable from passive microwave observations. C-band VOD has been used to generate yearly maps of forest above-ground biomass and carbon over a period of 20 years (Liu et al., 2015) at 25 km spatial resolution. L-band VOD from SMOS has been evaluated for African landscapes and shown to be closely related to AGB (Rodriguez et al., 2018; Fan et al., 2019). The relationship between VOD and biomass is explained in terms of increased attenuation that causes the VOD to increase with AGB. The retrieval algorithms proposed by Liu et al. (2015) and Rodriguez et al. (2018) use empirical functions to link VOD and AGB and are trained with AGB estimates from other AGB maps. This approach can be justified by considering that at the spatial resolution of the passive microwave data (0.25°), a “global” unique trend may characterize the dependence of VOD on biomass. This assumption, however, can be easily challenged by considering that the VOD experiences seasonality and depends on the structural and dielectric properties of a forest. Hence, using an AGB map as a surrogate training set to generate a global map of AGB may introduce errors by distorting trends in correspondence of regions that have not been mapped correctly in the reference dataset.

Apart from the GlobBiomass dataset of forest biomass (Santoro et al., 2018a), which will be addressed later on in this document, data acquired by active microwave sensors have so far remained mostly unexploited, one possible reason being that data are not provided in a ready-to-use format, as in the case of optical and LiDAR measurements.

A common feature of the algorithms listed above is that most emphasise data from a single sensor rather than considering how to exploit the information content in multiple datasets. This can be explained in terms of data availability at the time when the investigations were undertaken. Although not further addressed in this document, EO datasets have also been used to generate national, regional, continental and biome-specific datasets. Unlike global endeavours, the retrievals were built around the availability of reference data and/or multiple EO datasets; in addition, retrieval models could be regionalized by introducing location-specific information on vegetation properties, climate, etc.

Entering a data-rich epoch, these more local approaches may be transferable to the global scale. Any algorithm that aims to estimate AGB should consider exploiting complementary information from multiple sensors and exploit the biomass-related part of the signal. In addition, the uneven distribution of high-quality reference measurements used to train retrieval algorithms should be accounted for by designing the training procedure so that it is unaffected by such a deficiency.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	17	14.02.2019	

The world's forests are not measured evenly in space, which is likely to be a major source of estimation bias at global level, and it is unlikely that a single model realization (i.e., a single set of non-adaptive model parameters) can be applied globally. Using an AGB map as reference dataset could be considered an alternative, as long as the AGB estimates are correctly estimated. This may apply to LiDAR-based maps, which, however, have limited spatial extent. The use of wall-to-wall AGB datasets is discouraged because of local and regional biases (Mitchard et al., 2013; Rodriguez Veiga et al., 2017) that can propagate to the new set of AGB estimates.

In an attempt to generate a global dataset of AGB, the GlobBiomass project (<http://globbiomass.org>) attempted to implement the strategy outlined in the last paragraph and overcome some of the issues listed above by: (i) selecting a well-known modelling framework; (ii) using an adaptive approach to estimating the model parameters in space and time; and (iii) removing the requirement of *in situ* data for training (the model is self-calibrating). Point (i) was justified by the fact that numerous physics-based retrieval models already exist and, in contrast to machine learning algorithms, are transparent. Point (ii) is explained by the fact that remote sensing signals change in space and time, whereas retrieval models typically do not account for such variability. Point (iii) is possibly the most innovative aspect of an algorithm for estimating AGB because it aims to minimise the impact of reference data on the retrieval. Despite this seeming drastic, making a retrieval algorithm independent of reference data allows for a truly independent validation of the retrieval with *in situ* data. On the other hand, it requires profound knowledge of the EO data to be used to avoid macroscopic errors being introduced.

The GlobBiomass retrieval algorithm used state-of-the-art retrieval algorithms with a specific focus on implementing the three criteria discussed in the previous paragraph. However, the design of the algorithm was substantially affected by the EO data available for generating a global map of forest AGB for the epoch 2010. This was a fundamental factor in how the algorithm was designed, in the sense that it was built around globally available EO datasets containing information on "biomass".

2.2 The GlobBiomass biomass dataset

The objective of the GlobBiomass project was to generate a global dataset of forest AGB representative of the year 2010 epoch, satisfying the requirements that the error was at most 30 % and the spatial resolution below 500 m. From a design point of view, the possibility of achieving global coverage was considered to be more important than the requirement on estimation error because the EO data that could support the generation of a global dataset of AGB was sub-optimal. Biomass itself cannot be sensed by any instrument but only inferred with more or less complex mathematical models, from observations that relate to biomass. Such observations for 2010 consisted of wall-to-wall surface reflectance datasets acquired by high and moderate resolution sensors (Landsat, MODIS, MERIS) and SAR backscatter datasets acquired by high-to-coarse resolution sensors at short wavelengths (C- and L-band). In addition, footprints of laser waveforms were available but with too poor spatial sampling for direct ingestion in a biomass retrieval scheme. The outline of the GlobBiomass global retrieval algorithm is provided in Section 3; however, in this Section it is important to realise that the selection of input EO data had a major impact on the estimation results, regardless of how advanced the algorithmic implementation may have been.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	18	14.02.2019	

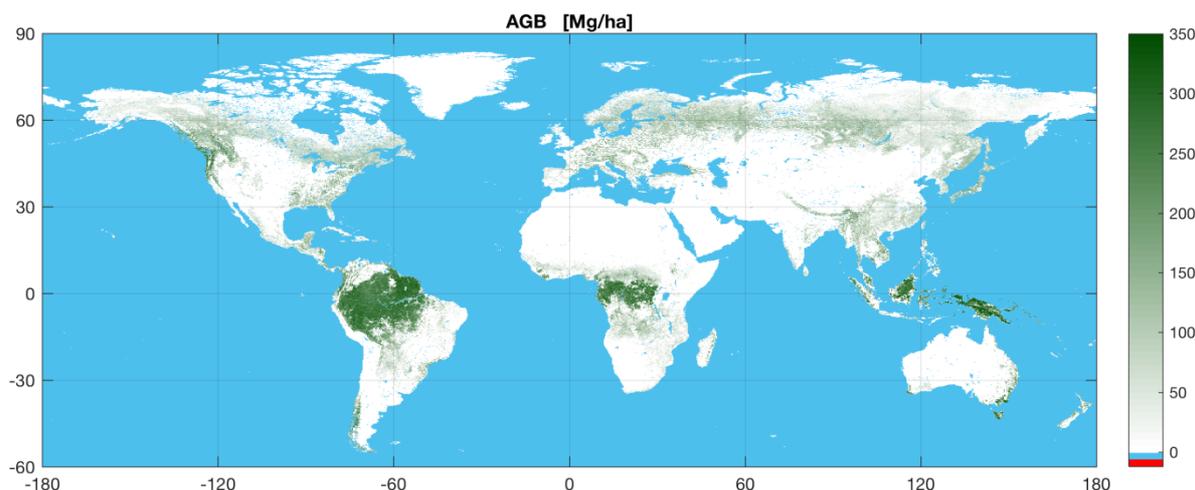


Figure 2-1: The GlobBiomass AGB dataset (Santoro et al., 2018a)

Figure 2-1 shows the GlobBiomass dataset of forest AGB. Validation of the GlobBiomass AGB estimates (Santoro et al., 2018a) indicated the overall reliability of the data product when comparing with AGB derived from inventory measurements at sample plots. While the spatial distribution of AGB appears to be captured, positive biases in the low biomass range (50-100 Mg ha⁻¹) and negative biases in the high biomass range (> 200 Mg ha⁻¹) occurred, although non-systematically. Examination of the spatial distribution of the biases revealed that these were caused by one or more of the factors listed below. These explanations are confirmed by the additional analysis undertaken in version 1.0 of the PVASR [RD-4] where the GlobBiomass map has been screened for structural deficiencies.

- A too conservative constraint on the maximum biomass for a given area (see Section 3.14) causing underestimation in the high biomass range
- A too generic definition of the forest transmissivity term of the models relating SAR backscatter to growing stock volume (GSV) (see Section 4.2.1.3) causing overestimation of biomass in the low-moderate biomass range
- Lack of sensitivity of the SAR backscatter to biomass towards the upper range of biomass.
- Artefacts in analysis-ready EO data (Shimada and Ohtaki, 2010) requiring strong image filtering which cancelled out subtle variations of the SAR backscatter
- Uncorrected effects of sloping terrain on the SAR backscatter (Shimada and Ohtaki, 2010) causing severe under/overestimation of biomass for slopes tilted towards/away from the look direction of the radar.
- Incorrect representation of scattering mechanisms in specific vegetation types where the models used to link SAR backscatter and biomass were not correctly parameterized (e.g., mangroves, flooded forest)
- Coarse representation of the conversion from GSV to AGB (see Section 3.15) causing unwanted local biases.

While the weak sensitivity of the SAR backscatter to AGB is an issue that cannot be compensated for, all other causes of biases can in theory be handled:

- Wider knowledge of the biomass distribution globally allows better characterization of the biomass spatial patterns and hence more realistic constraint of the retrieval models.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	19	14.02.2019	

- Access to unprocessed EO data would allow avoidance of artefacts.
- More precise knowledge of vegetation spatial patterns globally would allow better characterization of models and model parameters that describe the functional dependence of biomass on EO observables.
- Wider knowledge of wood density, biomass allocation to the tree components and allometry linking forest variables could feed back directly to the retrieval models and improve the capability to adapt to the local relationship between biomass and EO observables.

The validation exercise and thorough assessment of the GlobBiomass product also provided some lessons that are of utmost importance when designing a global biomass retrieval algorithm that should potentially solve the question on how uncertain the biomass pool is globally (see Figure 1-1) and overcome issues from GlobBiomass as well as from other endeavours targeting characterization of the world's forest biomass:

- Retrieval of biomass requires multiple data sources, in particular involving EO data not particularly suited to retrieving biomass
- Height information can substantially improve the estimates of biomass where the other EO observables do not present sensitivity to biomass.
- Retrieval of biomass does not necessarily require reference biomass data (e.g., in situ observations of biomass) for training.
- Retrieval should be based on multiple estimates, i.e., multiple models. Each model should allow adaptation of its parameters to cope with spatial variability in the functional relationship between EO data and biomass.

Points 1 and 3 represent two pillars of the GlobBiomass retrieval algorithm. Point 2 was given less importance in the GlobBiomass algorithm than in other approaches described in Section 2, mostly because of the potentially large error introduced by extrapolating relationships between height and biomass developed at sample points using raster datasets only partially sensitive to biomass. Nonetheless, the integration of height information from forthcoming datasets (ICESAT-2, GEDI, MOLI) is mandatory to be able to avoid systematic underestimation in high biomass forests. Point 4 was only touched on in GlobBiomass by pursuing separate retrievals with C- and L-band data and merging them; this needs further development by exploiting other approaches that can compensate for deficiencies in the biomass estimates obtained with the GlobBiomass approach.

2.3 Moving from the GlobBiomass to the CCI Biomass CORE algorithm

Based on the assessment of the GlobBiomass data product above, the global biomass retrieval algorithm implemented in CCI Biomass follows the same rationale as underpinned the development of the GlobBiomass retrieval algorithm. However, it expands and improves the GlobBiomass algorithm to: (i) better represent some vegetation-specific relationships between EO observables and biomass; (ii) account for new EO datasets not available at the time of the GlobBiomass project; and (iii) compensates for systematic errors revealed by the assessment of the GlobBiomass dataset. The current version of the CORE algorithm further expands on the experience of year 1 of this project, when the GlobBiomass algorithm was simply ported to the 2017-2018 epoch of data. In the process of improving the CORE algorithm, changes are applied that account for evidences from an assessment of the year 1 AGB map of 2017 described in the Product Validation Report (PVR, [RD-9]) and the PVASR of year 2 [RD-4].

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	20	14.02.2019	

A further development regards the inter-annual consistency of AGB estimates to be produced starting with year 2 of this project. Since the pool of remote sensing observations used to estimate AGB is not exactly the same for the time interval 2017-2019 and 2010 (see Section 3), retrieval algorithms need to ensure that such diversity is compensated for in order to ensure temporal consistency of the estimates across two decades.

3 Datasets

The remote sensing dataset considered to estimate AGB for the epochs 2017-2018 and 2018-2019 consists of ALOS-2 PALSAR-2 and Sentinel-1 observations of SAR backscatter. Wall-to-wall coverage by other SAR datasets is not accessible whereas passive microwave observations do not match the requirement of the recommended spatial resolution of AGB datasets by GCOS. Optical data are not considered because assumed to provide a negligible contribution to the retrieval. Following the same reasoning, the retrieval of AGB for the epoch 2010 is based on ALOS-1 PALSAR-1 and Envisat ASAR datasets. At the time of compiling this document, we have not explored the recent acquisitions by ICESAT-2 and GEDI data have not been available so far. The only global LiDAR dataset available consisted of the 2003-2009 acquisitions by ICESAT GLAS, which are considered here in the process of calibrating models rather than as predictors of AGB because of the sparse coverage and the part obsolescence in the process of supporting a retrieval for current epochs. The remote sensing datasets used to generate the global maps of AGB in CCI Biomass are summarized in Table 3-1. Scope of this Section is to describe the remote sensing observations used to generate AGB estimated for the three epochs of interest in this project. We also provide a description of auxiliary datasets (e.g., remote sensing data products) used in the process of calibrating AGB retrieval models.

Table 3-1: Remote sensing data sources, epochs covered and use.

Sensor	Epoch	Use
ALOS-2 PALSAR-2	2017-2018, 2018-2019	Predictor
Sentinel-1 (A and B)	2017-2018, 2018-2019	Predictor
ALOS-1 PALSAR-1	2010	Predictor
Envisat ASAR	2010	Predictor
ICESAT GLAS	All epochs	Calibration

3.1 Sentinel-1 (C-band, wavelength 5.6 cm)

Sentinel-1 (S1) is a spaceborne mission operated by the European Commission in the Copernicus framework and consists of two identical units (1A and 1B) flying C-band SARs. Sentinel-1A was launched on April 3, 2014 and became operational in October 2014; after a ramp-up phase, the satellite began routine observations in 2016. Sentinel-1B was launched on February 25, 2016 and became operational at the beginning of 2017. Each unit has a 12-day repeat-pass interval, which halves to 6 days when both units are operating. Thanks to the short repeat-pass interval and the small spatial baselines obtained by constraining the orbital tube within 100 m (Torres et al., 2012), the Sentinel-1 mission is particularly suited for interferometric applications (i.e., displacement monitoring, estimation of elevation, and thematic applications related to land-cover and land-use).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	21	14.02.2019	

Each unit can acquire data at single and dual-polarization (HH+HV or VV+VH) in a number of modes. Over land, the Interferometric Wide Swath (IWS) was selected; using the TOPSAR scanning technique, IWS achieves a spatial resolution of approximately 20 m in range and 5 m in azimuth, covering a swath of approximately 250 km. For remote regions, primarily the interior of polar regions and along its coastlines, S1 is operated in the Extended Wide Swath mode (EWS). Thanks to the ScanSAR observing technique, data acquired in EWS cover a swath of more than 400 km with a spatial resolution of approximately 100 m. Although Sentinel-1 can also acquire using other modes, these are of marginal interest for the scope of this document. Figure 3-1 shows a typical observation scenario of the Sentinel-1 constellation.

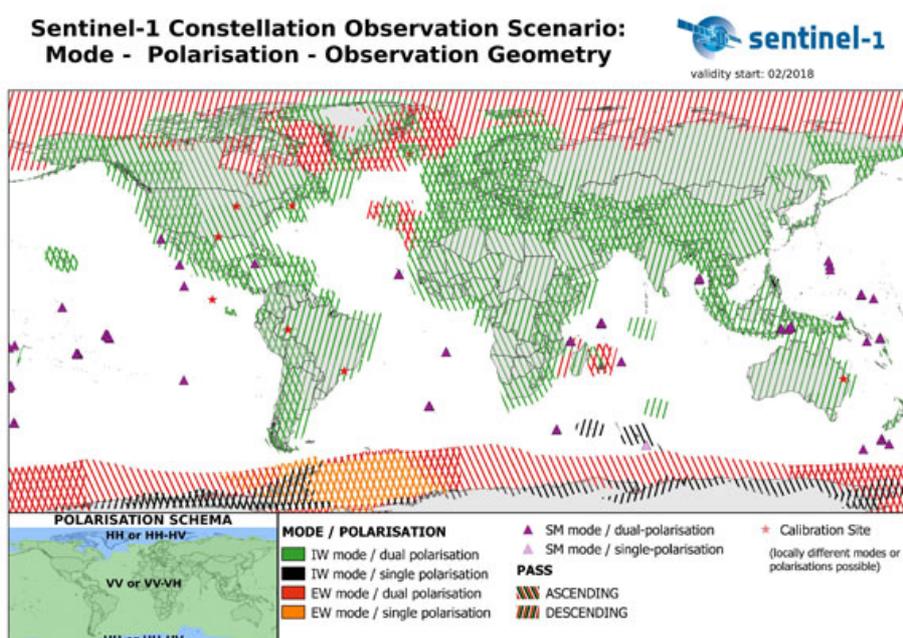


Figure 3-1: Observation geometry of the Sentinel-1 mission last accessed on 9 February 2019 (<https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario>).

Data acquisition by Sentinel-1 in the IWS mode is organized according to a predefined observation scenario with different levels of priority. Being a Copernicus mission, the highest priority is given to acquisitions over Europe, where each unit acquires along both ascending and descending paths (Figure 3-1). The second level of priority is given to areas prone to disasters due to tectonics, volcanic eruptions and earthquakes, as well as to polar and ice-covered regions. The third is given to areas of environmental importance (vegetation; e.g., wall-to-wall coverage of the tropical land surface). Finally, Sentinel-1 operations aim at achieving global coverage every 12 days with each unit. The EWS mode is not used as a complement to IWS but should rather be seen as an independent acquisition mode with specific requirements, i.e., frequent coverage and moderate resolution.

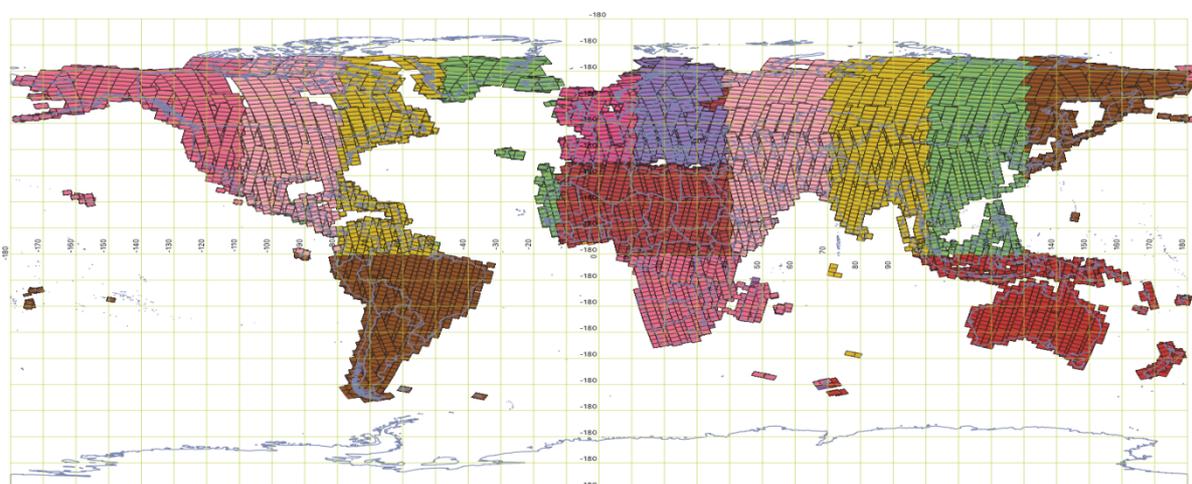
The acquisitions in IWS mode are programmed to allow for minimal overlap of swaths from adjacent orbital tracks at the Equator. The overlap tends to increase towards the poles so that the number of observations within a repeat-pass cycle of 12 days for a given point on the ground increases. Since some regions are observed with both units along both ascending and descending paths, one or more observations per day are possible for some locations. In contrast, the swath overlap of adjacent orbital tracks in EWS mode is large leading to a very high number of observations within the 12-day repeat-pass cycle of one unit (several observations daily are possible at the highest latitudes).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	22	14.02.2019	

It was shown with Envisat ASAR data that the retrieval of biomass benefits from a dense stack of observations of C-band backscatter (Santoro et al., 2011; Santoro et al., 2013; Santoro et al., 2015). However, not all observations in a data stack were found to contribute to the final estimate of biomass. The largest contribution came from images acquired under dry and frozen conditions in the boreal and temperate zone (Santoro et al., 2011). More generally, data acquired under dry conditions appeared to be more suitable than data acquired under wet conditions (Santoro et al., 2015a). In addition, in Santoro et al. (2011) it was concluded that having available at least 20 images with a backscatter contrast between unvegetated terrain and dense forest conditions of more than 0.5 dB allows systematic biases in the retrieved biomass to be reduced. Roughly one third of the C-band backscatter observations investigated in Santoro et al. (2011) fulfilled this requirement.

Following these indications, it is clear that retrieval based on Sentinel-1 images benefits from the repeated acquisitions since the start of routine operations by both units in 2017. However, the retrieval does not require the entire archive of data acquired since the start of the mission for retrieving forest biomass. For regions with almost daily observations, the IWS dataset to be used can be pruned to exclude observations with correlation close to 1.

An analysis of the Sentinel-1 data archives in terms of geographical distribution of the imagery at continental scale revealed that since 2017, the image data pool was extremely redundant over Europe. In addition, for the purpose of retrieving biomass, imagery acquired north of 75°N and south of 56°N were considered unnecessary because not covered with woody vegetation. After pruning the 2017 data pool of unnecessary data, the coverage shown in Figure 3-2 was obtained. The 2017 pool of images achieved global coverage of all forests except for a gap in northwest Canada. A more detailed search of the Sentinel-1 archives revealed that for this region there were hardly any acquisitions in IWS mode, whereas several acquisitions were available in the EWS mode. EWS data were therefore selected to fill the gap. Similarly, EWS imagery was used to fill two gaps smaller than an IWS scene (i.e., 250 x 250 km²). This gap-filling strategy had no effect on the biomass estimates to be obtained from Sentinel-1 data since Sentinel-1 imagery was processed to a pixel size of 150 m. For 2018, again pruning over Europe reduced the number of acquisitions to the same level of images available elsewhere. Global coverage was achieved without the need of gap filling with EWS data.



	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	23	14.02.2019	

Figure 3-2: Coverage of the 2017 Sentinel-1 IWS dataset selected to support the estimation of biomass for the 2017-2018 epoch (courtesy J. Kellndorfer, EBD).

Table 3-2 shows the geographic distribution of the Sentinel-1 images of 2017 and 2018 selected to support the estimation of biomass for the 2017-2018 epoch and the 2018-2019 epoch, respectively. Approximately, 252,000 scenes were selected for pre-processing the 2017 year. For 2018, approximately 317,000 images were selected. The reason for the smaller number of images in 2017 is the lack of observations by the 1B unit until April. The search was undertaken on the data repository of the Alaska Satellite Facility (ASF) because it mirrors European data holdings while providing a speedier and more reliable access to the data pool.

Table 3-2: Geographical distribution of the Sentinel-1 data pool used to support estimation of biomass for the 2017-2018 & 2018-19 epoch. The coordinates represent the extent of each region.

Continent	Long W	Long E	Lat N	Lat S	S1 unit	# scenes (K) year: 2017	# scenes (K) year: 2018
Africa	-12	40	34	0	AB	24	32
Africa	-20	70	0	-56	AB	15	19
Total Africa						39	51
Asia	40	70	75	0	AB	21	26
Asia	70	100	75	0	AB	21	24
Asia	100	130	75	0	AB	20	23
Asia	130	180	75	0	AB	9	10
Asia	70	180	0	-56	AB	19	25
Total Asia						90	108
Europe	-12	10	75	34	A	13	17
Europe	10	40	75	34	A	19	27
Europe	30	40	75	50	B	2	2
Total Europe						34	46
N. America	-180	-110	75	0	AB	18	20
N. America	-110	-80	75	0	AB	16	22
N. America	-80	-50	75	0	AB	19	24
N. America	-50	-12	75	0	AB	15	21
S. America	-180	-20	0	-56	AB	21	25
Total America	-180	-20	0	-56	AB	89	112
TOTAL						252	317

The Sentinel-1 IWS data pool consisted of individual images, each covering an area approximately 250 km × 250 km, in the geometry of acquisition of the Sentinel-1 radar. Scope of pre-processing was to generate a stack of terrain geocoded, radiometrically calibrated, speckle-filtered and co-registered Sentinel-1 observations. Having selected the backscattered intensity

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	24	14.02.2019	

as the observable to support the estimation of biomass, the pre-processing was applied to Sentinel-1 images provided in Ground Range Detected (GRD) format. GRD images consist of ground-range projected images of the SAR backscatter intensity. The pixel spacing of a GRD image acquired in the IWS mode is 10 m in both ground range and azimuth. Given that the spatial resolution of the IWS mode in the range direction is slightly less than 20 m (Torres et al., 2012), the images in GRD format are slightly oversampled. For the EWS mode, the same reasoning was applied. Compared to the IWS mode, the pixel spacing of an EWS image in GRD format is 50 m in both range and azimuth, thus oversampled as for IWS data. Although Single Look Complex (SLC) images retain the original spatial resolution of the data, data in GRD format were used for several reasons:

1) SLC images allow the generation of interferometric images such as coherence, and it has been shown that estimation of biomass from C-band coherence is more accurate than when from SAR backscatter (Santoro et al., 2002; Santoro et al., 2018b), but it is unlikely that the 6- and 12-day repeat-pass intervals of the Sentinel-1 constellation will allow coherence to be preserved in vegetated regions. The effort of processing SLC data to coherence globally is therefore likely to be of little value to this project as well as being beyond the available capabilities.

2) A single GRD scene in IWS mode covers an area of 250 x 250 km², corresponding to approximately 1 GB of data. The corresponding SLC image consists of approximately 4 GB. Since SLC data are strongly affected by speckle, multi-looking (i.e., spatial averaging) is required. For a minimal improvement in terms of radiometric resolution, the effort of accessing and managing images in SLC format instead of GRD format is unjustified. This is an important issue given that 252,000 scenes needed to be processed quickly to generate a global map of biomass during year 1.

3) Based on previous experience when using GRD data for large-scale land mapping and monitoring (Santoro et al., 2017), the quality of the data in GRD format was considered to be sufficient to support the retrieval of biomass.

The SAR pre-processing chain is shown in Figure 3-3. Before implementing the pre-processing chain, the output pixel spacing of the Sentinel-1 image data was analysed. The option of pre-processing to preserve the spatial resolution of the data was discarded because of the extremely large amount of data to be handled throughout the phase of retrieving biomass. Since each image file consists of roughly 1 GB, we would have faced a total output of 250-300 TB of backscatter data per year to be used for biomass retrieval. In addition, one would have needed to account for the size of the auxiliary data files that support the retrieval, such as maps of layover/shadow and local incidence angle. In the end, it was decided to spatially average the GRD data files to a pixel size that would preserve spatial details, while effectively removing speckle. It was also taken into account that the purpose of the Sentinel-1 dataset was to support the estimation of biomass in the context of CCI Biomass (i.e., for a community of users that does not require high spatial resolution products). Finally, it was considered that such a dataset should be compatible with other datasets of C-band backscatter measurements, namely from Envisat ASAR (see Section 3.3). It was, therefore, decided to process the Sentinel-1 data to the same geometry as the moderate resolution ASAR dataset (i.e., to a pixel spacing of 150 m). Ultimately, the benefit of working with “clean” SAR backscatter observations appeared to be more important than preserving the high spatial resolution, though a numerical analysis was not undertaken.

The commercial software package by GAMMA Remote Sensing was used to pre-process the Sentinel-1 data. Import of Sentinel-1 SAR images into the GAMMA Software consisted of

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	25	14.02.2019	

reformatting the SAR dataset to the GAMMA Software structure (image dataset and metadata in the image parameter file) (Wegmüller et al., 2016). In addition, while importing, calibration and noise reduction were applied with the calibration gain and the noise factors reported in the original image metadata and auxiliary data files. Precise orbit information was used to replace state vectors provided in the original metadata of each image (https://qc.sentinel1.eo.esa.int/aux_poeorb/).

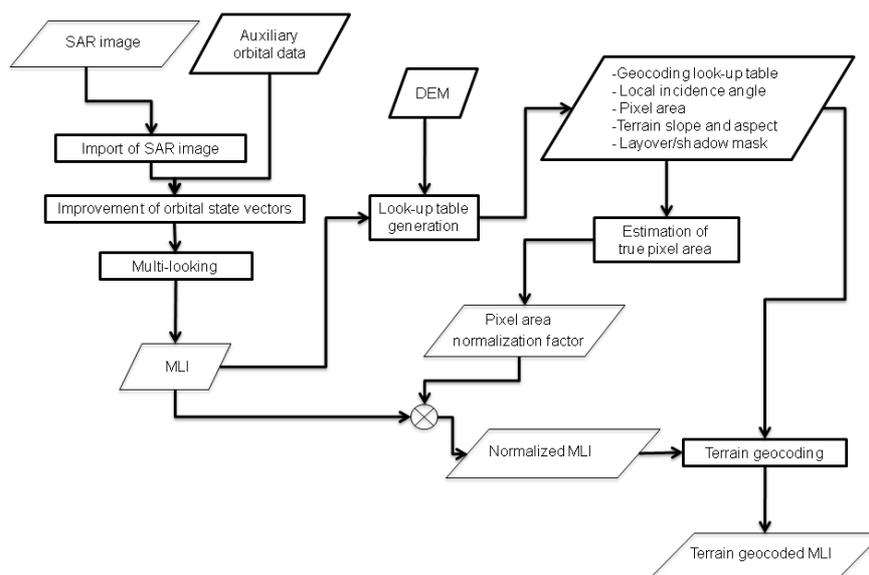


Figure 3-3: Flowchart of the Sentinel-1 data pre-processing

Multi-looking consisted of box-car averaging of the backscatter of contiguous pixels in the averaging window. The averaging window was 15×15 pixels in order to achieve a multi-looked intensity (MLI) image with a pixel spacing of 150 m in both range and azimuth. Because of the strong averaging, no additional speckle filter was applied. To estimate the level of residual speckle noise, the Equivalent Number of Looks (ENL) (Oliver and Quegan, 1998) was computed.

$$ENL = \frac{\mu^2}{\sigma^2} \quad (3-1)$$

The computation of the ENL as in Equation (3-1) was implemented by drawing a polygon that included an area characterized by a homogeneous distribution of features (e.g., a dense forest, a field) and computing the mean and variance of the SAR backscatter within it. This operation was repeated for several polygons spread over the SAR image to obtain a histogram of values in order to better quantify the ENL and avoid having an estimate based on one or a small number of polygons that could be biased because of how these were selected. The computation of the ENL was impossible for all Sentinel-1 images. Since it could be reasonably assumed that ENL should not depend on seasonality or the specific land cover type, we randomly selected a small number of images from the data stack of 2017, then created polygons and finally computed the ENL for each polygon and image. This “global” set of ENL values is displayed in Figure 3-4; the median was 162 and the span was [90, 375] with most values being between 100 and 250. The error statistics derived from this analysis are further discussed in the E3UB document of the CCI Biomass project [RD-5].

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	26	14.02.2019	

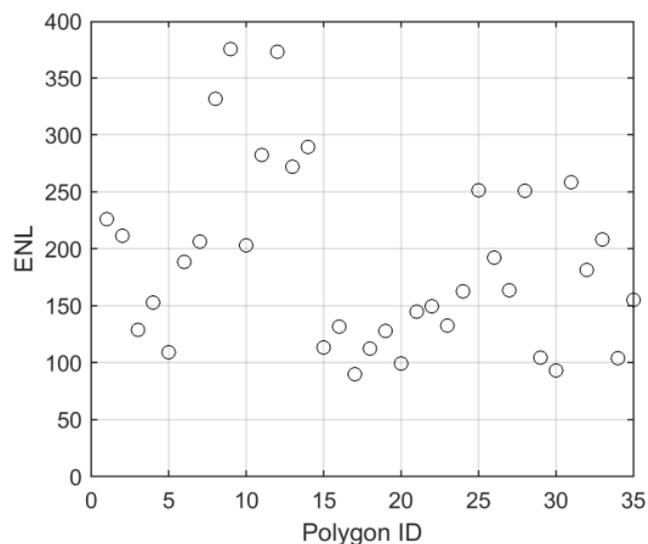


Figure 3-4: Estimates of ENL for 35 polygons distributed over five Sentinel-1 VV-polarized images randomly selected in boreal, temperate and tropical environments.

Since Sentinel-1 images were obtained in radar geometry, they needed to be transformed into the output map geometry. For CCI Biomass, the geographical coordinate system with a pixel spacing of 0.0013888° , corresponding to 150 m at the Equator, was adopted. The transformation of a SAR image from radar to map geometry was implemented in the form of a geocoding look-up table (LUT; Wegmüller, 1999). The LUT reflected the output geometry (map projection in this case); at each pixel, the LUT contained the corresponding x and y coordinates in the SAR image. The LUT was created with the aid of orbital parameters and SAR image processing parameters (e.g., slant-to-ground range polynomials, image start time etc.), and elevation information in a Digital Elevation Model (DEM). Here, we used the global 3 arc-seconds DEM (i.e., roughly 90 m at the Equator) described in Section 3.3. Together with the LUT, data layers directly related to the elevation reported in the DEM were also generated (i.e., the image of the local incidence angle, the image of the pixel area and an image flagging the occurrence of layover or shadow). As the precise orbits were used, there was no need to refine the geocoding LUT. The co-registration error between the DEM and a small number of geocoded Sentinel-1 images was estimated by means of the cross-correlation technique described in Wegmüller et al., (2002). The standard deviation of the co-registration error was below $1/10^{\text{th}}$ of the output pixel size (i.e., less than 15 m). Again, given the impossibility of evaluating the co-registration between DEM and SAR imagery for the entire Sentinel-1 data pool, we assume that the statistics derived here for a small sample of images apply to the entire image dataset. This should be reasonable considering that Sentinel-1 images are well characterized in terms of orbital parameters.

To compensate for distortions of the SAR backscatter due to sloping terrain (foreshortening, shadow and layover), a normalization factor was computed. This accounted for the true size of the pixel instead of the size of the pixel on a flat terrain as assumed when generating the GRD data product (Frey et al., 2013). The area of each pixel in an image was estimated using the DEM and the orbital parameters in the SAR image metadata together with the geocoding LUT. The normalization procedure estimated both the true pixel area and the area of the pixel on the ellipsoid (i.e., for a flat surface); from this, a precise normalization factor was obtained. This factor was applied to each SAR backscatter image to obtain the corresponding image of backscattered intensity with reduced slope-induced distortions. This step was performed in the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	27	14.02.2019	

original Sentinel-1 radar geometry. In addition, the σ^0 measurement was converted to γ^0 (i.e., σ^0 divided by the cosine of the local incidence angle).

The SAR backscatter image MLI normalized for pixel area was finally terrain geocoded with the geocoding LUT.

The pre-processing sequence outlined in Figure 3-3 was repeated for each Sentinel-1 image forming the data pool. To obtain the stack of co-registered observations of the SAR backscatter, each image was tiled to the pre-defined $1^\circ \times 1^\circ$ grid adopted for the pre-processing of the ASAR data in the context of CCI Land Cover. Each tile consisted of 720×720 pixels.

Figure 3-5 shows the number of Sentinel-1 backscatter observations per pixel for the year 2017. The density of observations was highest over Europe, even if we only selected data from one unit. In accordance with the observation priorities of Sentinel-1, outside Europe hazard-prone areas were imaged more frequently than other areas. As a minimum, dual-polarized observations every 12 days were available, resulting in approximately 30 observations per polarization (VV and VH or, primarily in polar regions, HH and HV polarization; i.e., 60 observations per pixel). At the time of compiling this document, the pre-processing of the Sentinel-1 dataset of 2018 was just completed and we could not generate a similar map of observations in time before delivery. It is believed, however, that the frequency of observations in 2018 is of the same order, if not superior, than in 2017.

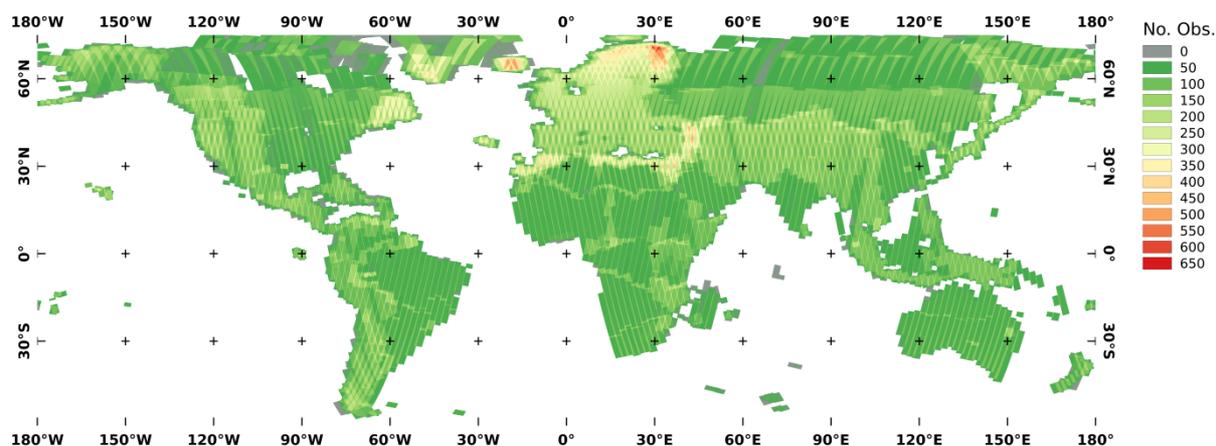


Figure 3-5: Number of observations per pixel for the Sentinel-1 2017 dataset.

To obtain an overall impression of the quality of the pre-processed data, a mosaic of the 2017 Sentinel-1 dataset represented as a false colour composite of temporally averaged backscatter is displayed in Figure 3-6. As there were an extensive number of observations per pixel, the mosaic clearly reveals the features of the land surfaces and highlights that thematic applications based on Sentinel-1 time series are possible globally. The image in Figure 3-6 also shows that the distortions in SAR backscatter due to sloping terrain have been minimized.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	28	14.02.2019	

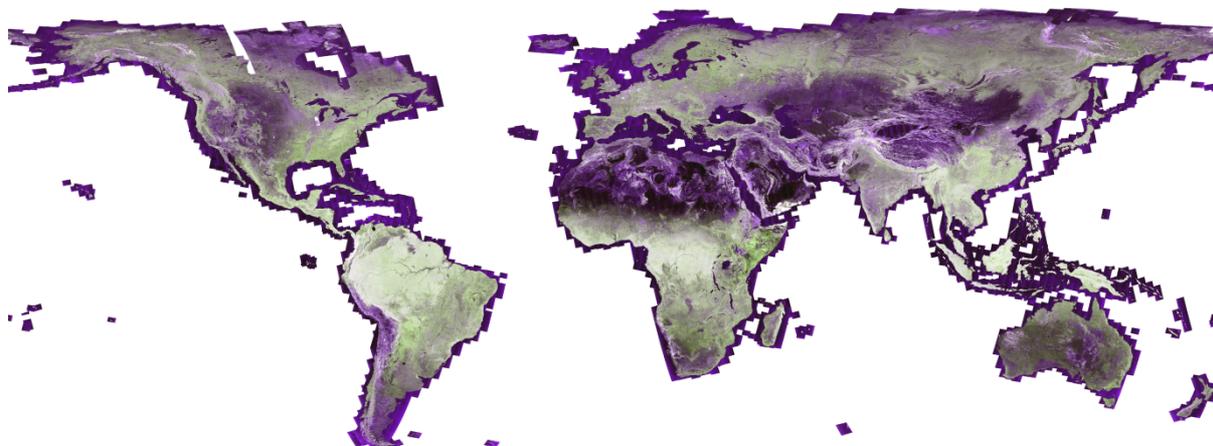


Figure 3-6: False colour composite of the Sentinel-1 2017 dataset. Red: temporally averaged co-polarized backscatter; green: temporally averaged cross-polarized backscatter; blue: ratio of the temporally averaged cross-polarized and co-polarized backscatter. Pixel size: 150 m × 150 m.

The individual images are, however, not free from errors; in particular, images are occasionally affected both by radiometric errors introduced when the Sentinel-1 raw data were processed to GRD format (high backscatter) and residual slope-induced effects in correspondence of errors in the elevation datasets forming the global DEM. A detailed presentation of errors affecting the Sentinel-1 backscatter dataset is given in the E3UB document [RD-5].

3.2 ALOS-2 PALSAR-2 (L-band, wavelength 23 cm)

The ALOS-2 mission started on 24 May 2014 and carries an L-band SAR (PALSAR-2 instrument) with slightly improved performance than its predecessor, ALOS-1 PALSAR-1 (see Section 3.4). ALOS-2 PALSAR-2 operates a high resolution acquisition model (25 m, Fine Beam; FB) and a moderate resolution model (50 m, Wide Beam; WB). Each year global and repeated acquisitions are scheduled during seasons that are known to maximize the information content of the backscattered signal with respect to land surface properties. In both FB and WB mode, PALSAR-2 acquires data in single polarization (HH) and dual polarization (HH and HV), covering swaths of approximately 70 km and 250 km, respectively.

Because of the data policy applied by JAXA to ALOS-1 and ALOS-2 data, only a limited number of images can be obtained free of charge, which hinders any global application. As such, global coverage's of ALOS-2 PALSAR-2 data can only be obtained in the form of yearly backscatter mosaics for the FB mode and per-cycle mosaics (46 days) for the WB mode. The WB mode mosaics are also referred to as ScanSAR mosaics, ScanSAR being the imaging technique used for covering the large swath of 250 km.

The ALOS-2 PALSAR-2 dataset used in the context of CCI Biomass consists of:

- yearly mosaics of HH and HV polarized backscatter acquired in Fine Beam Dual (FBD) mode between 2015 and 2018
- per-cycle mosaics of HH- and HV-polarized backscatter acquired in WB mode between 2016 and 2018

All mosaics were produced by JAXA (Shimada and Ohtaki, 2010; Shimada et al., 2014). While the FBD mosaics are publicly available, the ScanSAR mosaics are available only to a restricted research community (i.e., the Kyoto and Carbon (K&C) Initiative).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	29	14.02.2019	

Each FBD mosaic covers the entire globe and has been generated primarily with ALOS-2 FBD data acquired between May and October of a given year. However, to achieve global land coverage, gaps had to be filled with data acquired in winter throughout the northern hemisphere, and locally also with data from other years. Currently, mosaics for each year between 2015 and 2018 have been released. The annual FBD mosaics (HV-polarization) for 2015-2017 are shown in Figure 3-7. The mosaic of 2018, released at the end of 2019, is currently being downloaded.

The ScanSAR data are primarily acquired over the tropics and therefore the mosaics for each cycle cover only part of the Earth's land surface. An example for a ScanSAR mosaic covering the Amazon basin is shown in Figure 3-8. A list of all ALOS-2 observation cycles for the mosaics released by JAXA can be found in Table 3-3. At the time of writing, it is not foreseen by JAXA to generate mosaics after October 2018.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	30	14.02.2019	

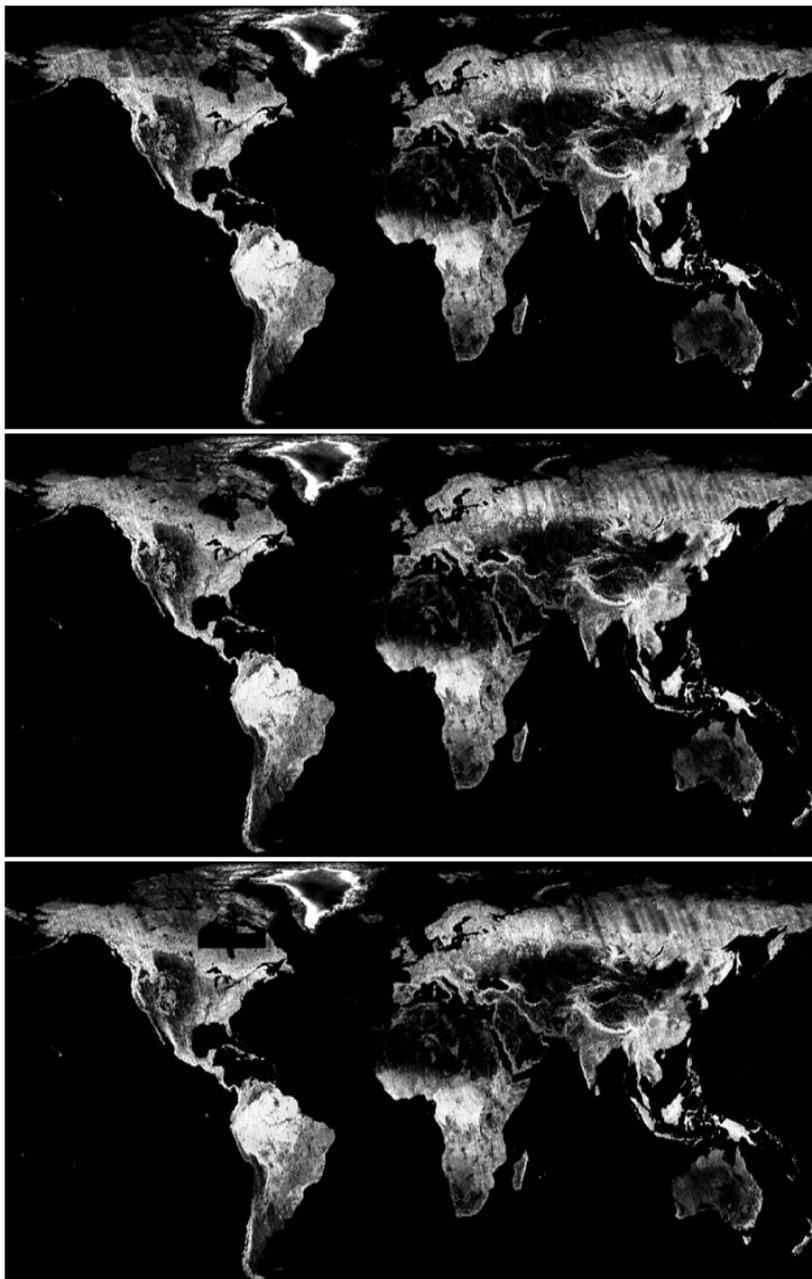


Figure 3-7: ALOS-2 FBD mosaics, HV-polarization, for the years 2015 (top), 2016 (middle) and 2017 (bottom).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	31	14.02.2019	

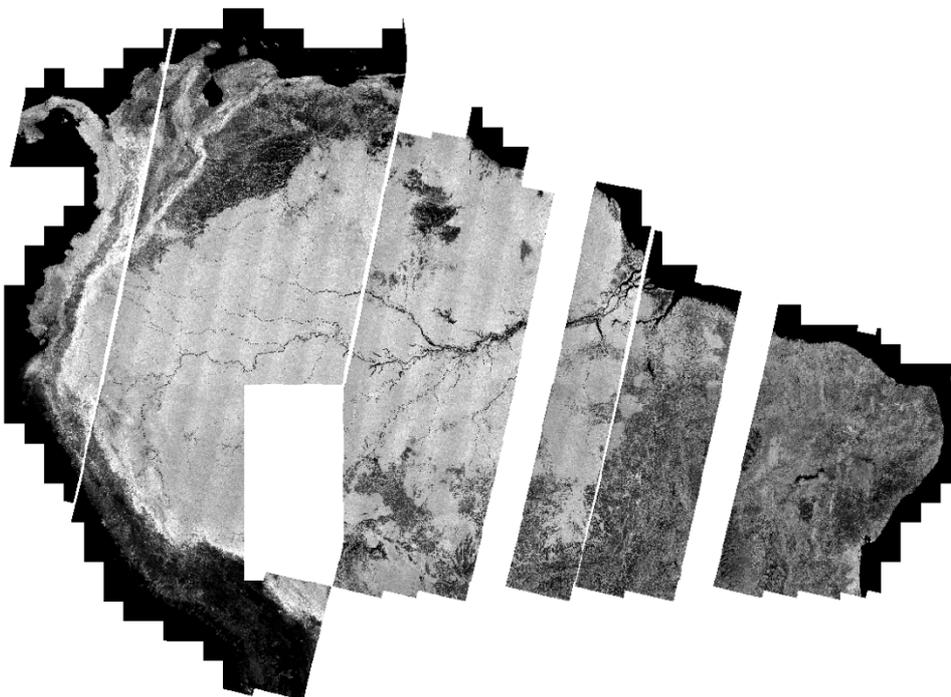


Figure 3-8: ALOS-2 ScanSAR mosaic generated from HV polarization imagery acquired in April 2018 over the Amazon Basin.

Table 3-3: ALOS-2 acquisition cycles for which mosaics of dual-polarization backscatter observations acquired in ScanSAR mode have been released by JAXA.

Cycle	Start date	Cycle	Start date
45	28-Mar-16	91	01-Jan-18
48	09-May-16	93	29-Jan-18
51	20-Jun-16	94	12-Feb-18
53	18-Jul-16	96	12-Mar-18
56	29-Aug-16	97	26-Mar-18
59	10-Oct-16	99	23-Apr-18
62	21-Nov-16	100	07-May-18
65	02-Jan-17	102	04-Jun-18
68	13-Feb-17	103	18-Jun-18
71	27-Mar-17	104	02-Jul-18
74	08-May-17	105	16-Jul-18
77	19-Jun-17	107	13-Aug-18
79	17-Jul-17	108	27-Aug-18
82	28-Aug-17	110	24-Sep-18
85	09-Oct-17	111	08-Oct-18
88	20-Nov-17		

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	32	14.02.2019	

To provide an idea of the disparity of L-band observations in the tropics and elsewhere, Figure 3-9 illustrates the number of ALOS-2 observations available when combining FBD and ScanSAR mode observations from 2015 to 2017. For most of the northern hemisphere only the annual FBD mosaics are available. The number of observations increases in the tropics, Central America and the non-tropical regions of Southern America and Southern Africa, with up to 30 observations where both FBD and ScanSAR imagery are available.

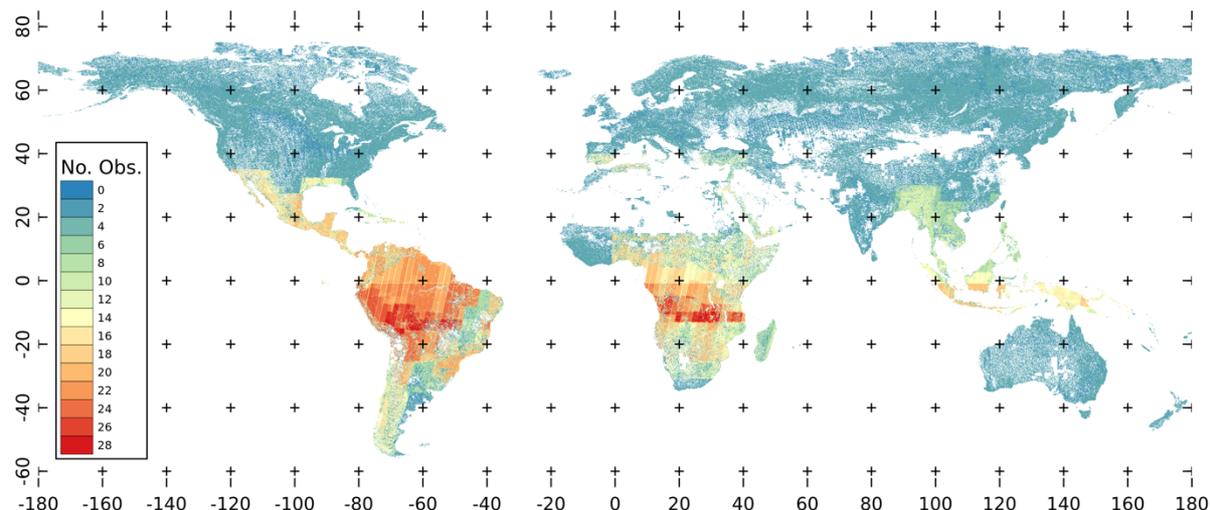


Figure 3-9: Number of ALOS-2 FBD and ScanSAR observations available in the time frame 2015-2017.

Each of the mosaics is provided in the form of 1°x1° tiles and includes the HH and HV backscatter as well as:

- the local incidence angle with respect to the orientation of the pixel, derived from a DEM (3-arcsec Shuttle Radar Topography Mission (SRTM) or 1-arcsec ASTER DEM), as well as layover/shadow masks
- the date of acquisition of the image
- an indication of whether the pixel is land or water

The FBD data were processed to γ^0 (i.e., σ^0 divided by the cosine of the local incidence angle; Shimada, 2010), and resampled to a pixel size of 1/4000th of a degree in both latitude and longitude, corresponding to roughly 25 m at the Equator. The ScanSAR data were instead processed to a pixel size of 1/2000th of a degree, i.e., roughly 50 m at the Equator.

The ALOS-2 datasets were geocoded, orthorectified and calibrated by JAXA. The mosaics 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). However, visual inspection of the imagery available at the time of compiling this document (all except the FBD mosaic of 2018) indicated significant problems with the geolocation accuracy, in particular for the ScanSAR data. The results of the visual inspection were confirmed when using matching techniques based on image cross-correlation to identify systematic linear offsets in Northing and Easting between the ALOS-2 mosaics and an ALOS PALSAR mosaic for the year 2010, which had been used in the GlobBiomass project, with better geolocation accuracy. The per-tile estimates for the offsets in Easting and Northing, which roughly reflect the range and azimuth dimension of the radar acquisitions (at least close to the Equator) are shown in Figure 3-10. On average, the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	33	14.02.2019	

offsets were in the range of 0.5 to 1 pixels (1 pixel corresponds to 50 m). In a few cases, the offsets reached more than 5 pixels in both Easting and Northing. The reason for these offsets is not entirely clear because the details of the SAR processor used by JAXA for geocoding ALOS-2 imagery are not known. One possibility is that the SRTM DEM, which JAXA used for terrain-corrected geocoding and which reports elevations with respect to a geoid instead of an ellipsoid, was used without compensating for the geoid/ellipsoid height difference. This was confirmed when relating the offsets in Figure 3-10 to the EGM96-Geoid to WGS84 ellipsoid height difference (Figure 3-11). For scenes acquired between 20°S and 20°N for which the Easting roughly corresponds to the range dimension of the radar imagery (satellite heading of ~10°), we find a clear relationship. In the North direction (i.e., roughly the radar azimuth direction), the offset is not clearly related to the geoid height offset and the geolocation errors are relatively constant at 0.5 to 1 pixel.

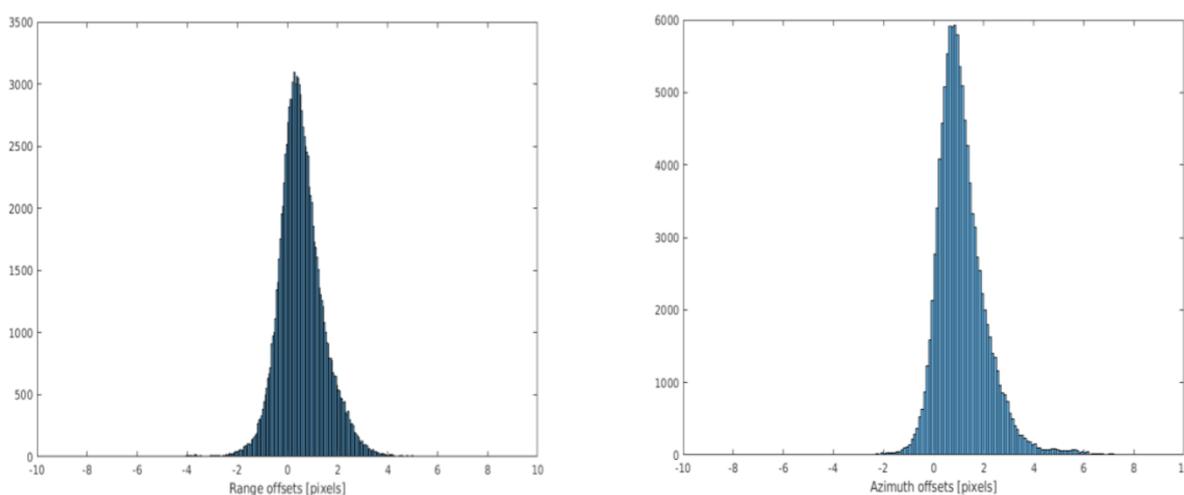


Figure 3-10: Geolocation offset between ALOS-2 ScanSAR and ALOS PALSAR FBD mosaic for the year 2010 determined on a 1°x1° tile basis using image cross-correlation.

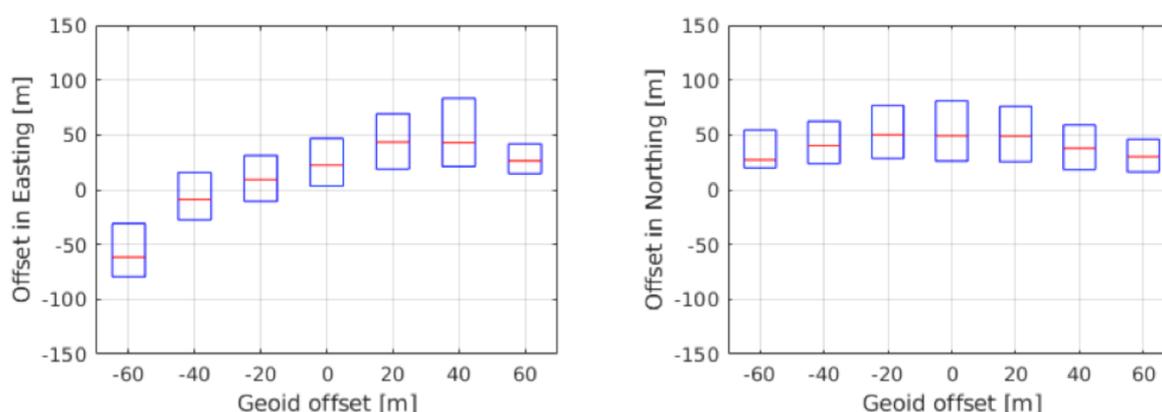


Figure 3-11: Geolocation error of ALOS-2 ScanSAR mosaics as a function of the EGM96-Geoid/WGS84 elevation offset. Only scenes acquired between 20°S and 20°N were considered.

The images were therefore co-registered to a Landsat reflectance mosaic from 2000 (Hansen et al., 2013) assuming a linear offset in Easting and Northing. Different reflectance bands/ratios from Landsat were tested for the co-registration by means of cross-correlating small image chips in the ALOS and Landsat imagery. Tests indicated that the Normalized Difference

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	34	14.02.2019	

Vegetation Index (NDVI) presented the highest correlations with the L-band backscatter imagery, which is why the NDVI was selected as reference for the co-registration. In the co-registration, it was considered that, as a consequence of the height offsets in the DEM that had been used for geocoding the radar imagery, the geometry of the backscatter imagery and the local incidence angle files does not match. As a consequence, the conversion from σ^0 to γ^0 performed sub-optimally. To reduce topographic errors in the backscatter imagery, the γ^0 imagery was reconverted to σ^0 before co-registration. After the co-registration, the backscatter images were converted back to γ^0 using the local incidence angle maps. Figures Figure 3-12 and Figure 3-13 demonstrate that this procedure allows for reducing topographic effects in the backscatter imagery. In the example in Figure 3-12, the co-registration indicated an offset between Landsat and ALOS-2 imagery of 0.5 pixels ($\sim 50\text{m}$) in range. After resampling the backscatter images to better match the geometry of the Landsat NDVI image, the corrected HV backscatter image differed from the uncorrected image for up to 3 dB, mostly in areas of steep terrain. In the example in Figure 3-13, the co-registration indicated a range offset of even 1.5 pixels ($\sim 150\text{m}$) between ALOS-2 and Landsat imagery. Accordingly, the co-registration resulted in backscatter over steep terrain that differed for up to 5 dB from the uncorrected image.

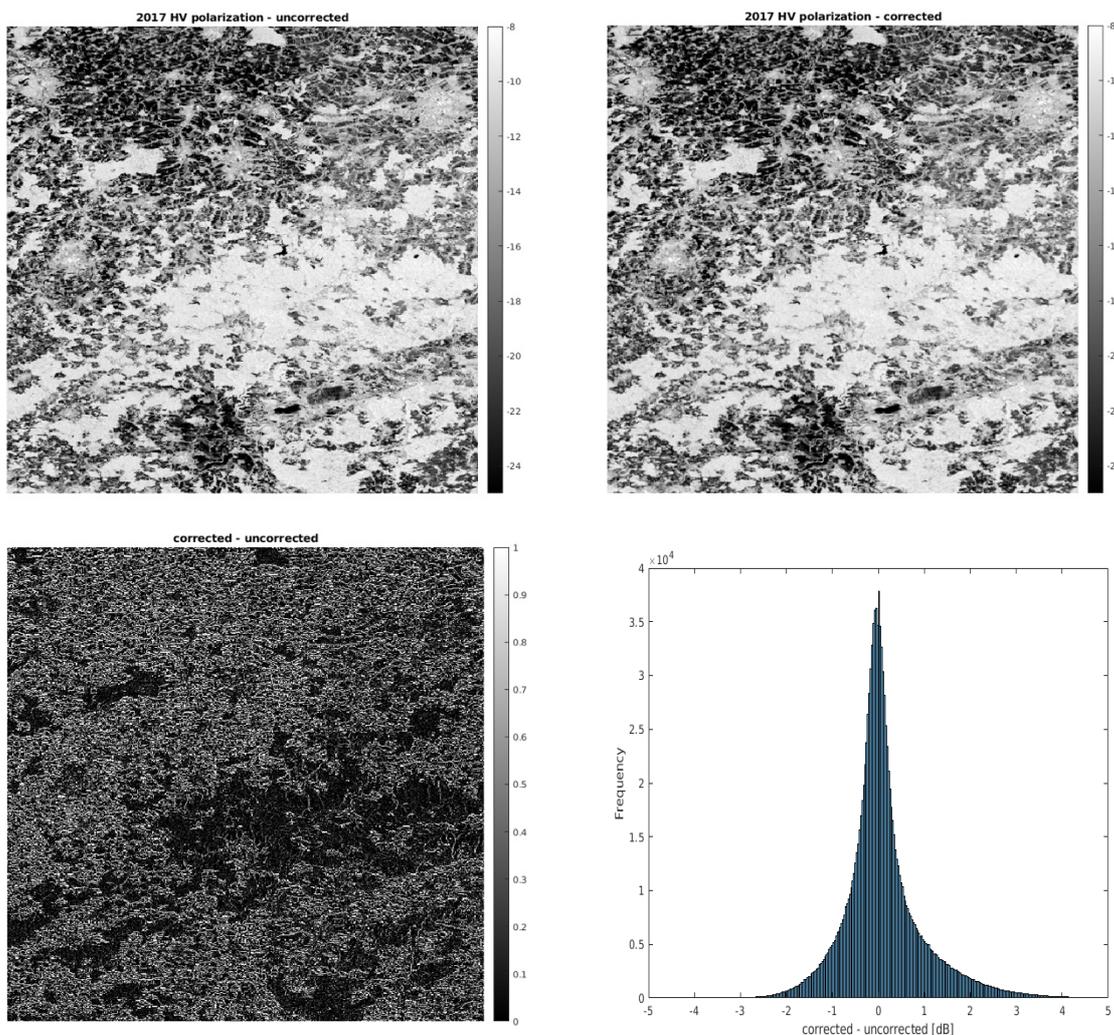


Figure 3-12: Effect of co-registration of a 1x1 degree tile of L-band HV backscatter acquired over central Germany with Landsat NDVI imagery.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	35	14.02.2019	

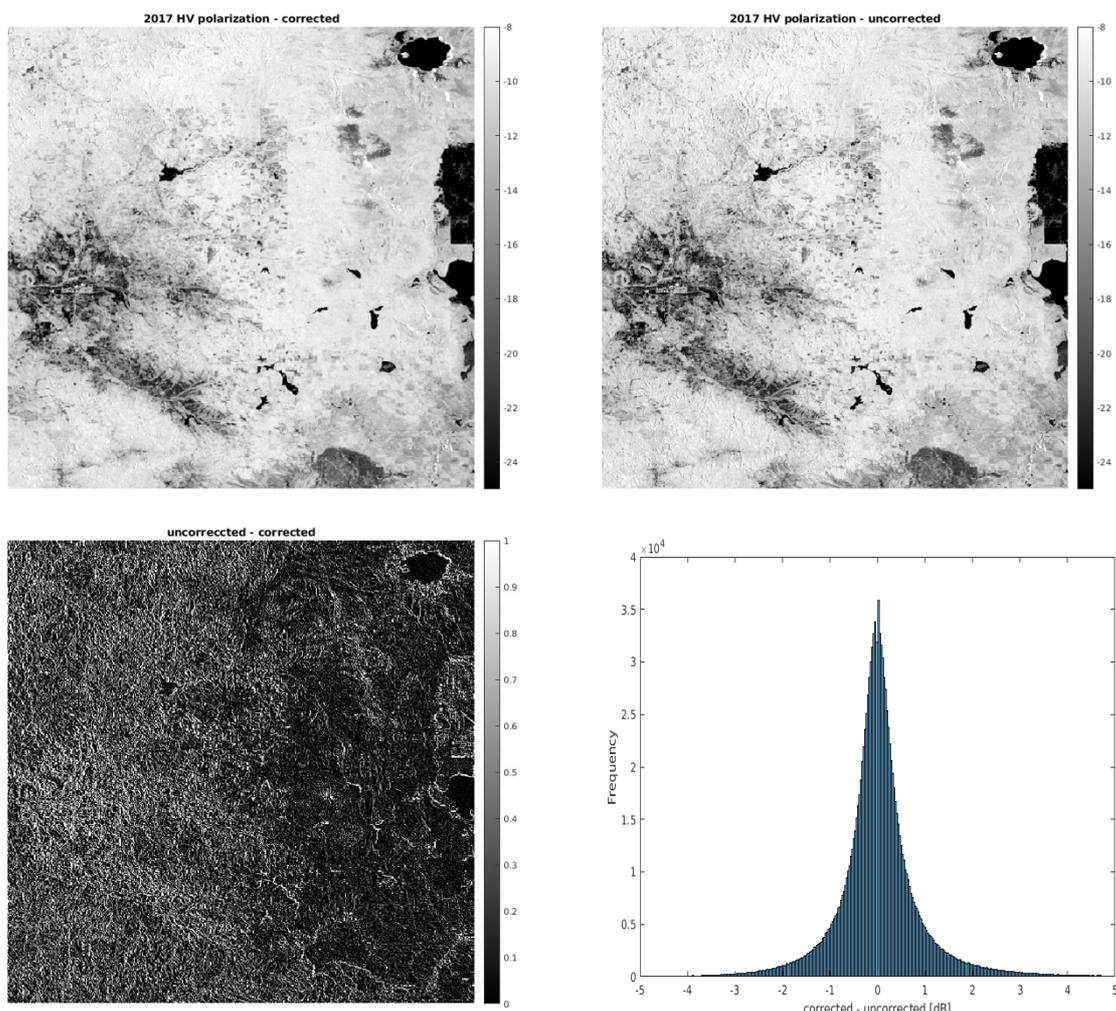


Figure 3-13: Effect of co-registration of a 1x1 degree tile of L-band HV backscatter acquired over the North-western United States with Landsat NDVI imagery. 1.5 pixels (~150) shift in range.

While the co-registration clearly improved the backscatter images in areas of steep terrain, it has to be stressed that a full correction of the geolocation errors and resulting radiometric errors over sloped terrain could not be achieved without detailed information about the imaging geometry. The co-registration represents a temporary workaround until reprocessed ALOS/ALOS-2 mosaics become available. The geometric mismatches were communicated to JAXA and reasons were recently investigated leading to an improved processing sequence that was applied to generate the FBD mosaic of 2018. At the time of writing this additional paragraph, we could not undertake a comparison between older mosaics and the 2018 mosaic, thus being unable to assess improvements in the quality of the data.

In order to reduce the speckle in the ALOS-2 imagery, all images were:

- 1) aggregated to the target pixel size of 100 m (0.000888888°) for the mapping of biomass
- 2) filtered with the multi-temporal filter suggested in Quegan & Yu (2001)

The ENL of the imagery after filtering was assessed for a number of homogenous forest patches, identified by means of visual image interpretation. Since the performance of the multi-temporal filtering depends on the number of images considered in the filtering as well as the level of speckle correlation between images (which given the repeat intervals of ALOS-2 of 14 days

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	36	14.02.2019	

should be low), no global ENL can be specified. In areas where only FBD mosaics were available, we find the ENL to be of the order of 70 to 80. In areas where FBD and ScanSAR imagery could be combined, the ENL was on average of the order of 300.

The mosaics exhibit significant striping, in particular in the boreal zone as well in areas with continuous forest cover, such as the Amazon or Congo Basin. In the boreal zone, the striping is because imagery acquired under winter frozen conditions had to be used by JAXA to achieve global coverage (Figure 3-14). When imagery is acquired under these conditions, the backscatter tends to be several dB lower than under unfrozen conditions and the sensitivity to biomass will also be reduced (Santoro et al., 2015b). Radiometric balancing of the mosaics in the boreal zone was not attempted because when generating the mosaics JAXA already attempted to reduce the differences between adjacent orbits using a weighted feathering approach. However, for adjacent orbits acquired under frozen and unfrozen conditions with backscatter offsets of several dBs, the feathering led to strong artefacts, which cannot simply be undone. It was therefore concluded that it is better to optimize the multi-temporal biomass retrieval algorithm by detecting images affected by freeze/thaw transitions and giving them low weights compared to the other multi-temporal/multi-sensor imagery in the biomass retrieval (i.e., Sentinel-1 and ALOS-2).

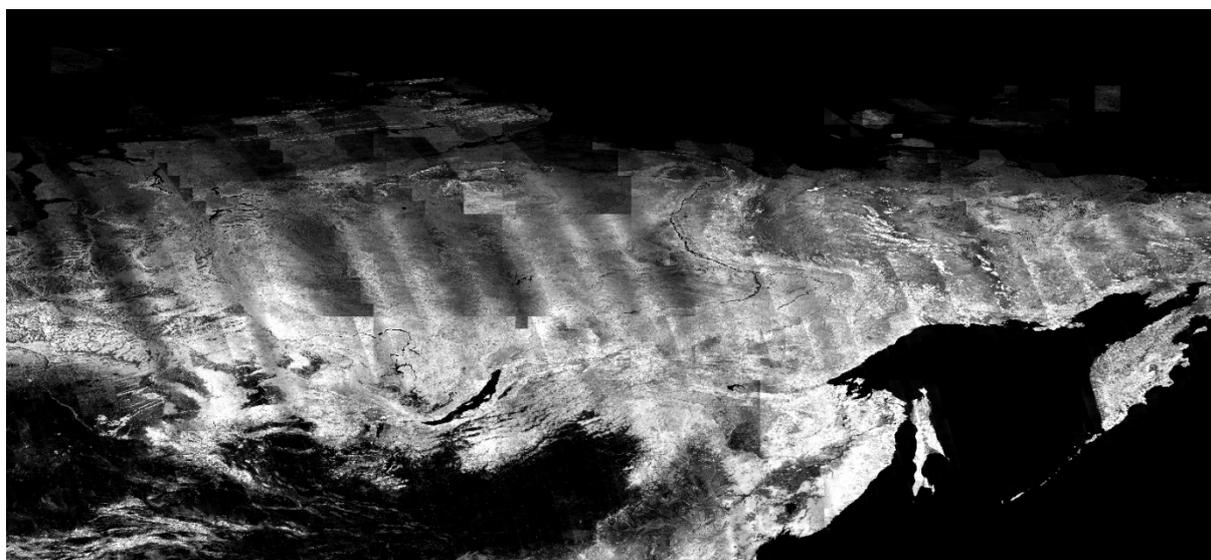
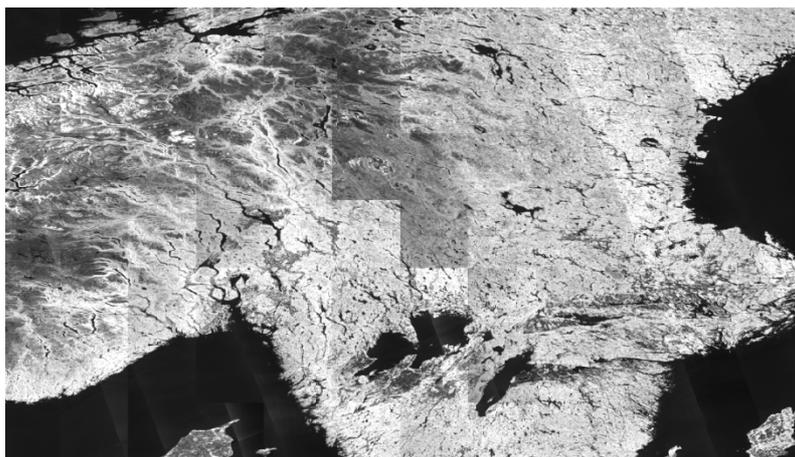


Figure 3-14: ALOS-2 L-HV backscatter mosaic for Southern Sweden (top) and Northern Asia (bottom).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	37	14.02.2019	

The striping visible in the mosaics over continuous tropical rainforest is associated with the incidence angle dependence of backscatter and so is not only a problem for this forest type. In the tropics, however, it is more obvious because of the presence of continuous forest cover and the low sensitivity of backscatter to forest density or biomass. Methods aiming at reducing the incidence angle dependence of backscatter need to consider that the local incidence angle dependence differs between land cover classes which might be considered opaque isotropic volume scatterers (e.g., dense tropical forest) and sparse vegetation classes with increasingly non-isotropic properties. A semi-empirical method aiming at adaptive normalization of backscatter with respect to the incidence angle was presented in Hoekman & Reiche (2015). However, the approach suggested is mostly valid when using radar imagery to map land cover classes. The implementation when aiming at retrieval of a continuous forest variable such as biomass (i.e., when the normalization approach should adapt to the forest variable of interest itself) is not clear and requires further investigation. Efforts on radiometric normalization of the imagery with respect to incidence angle effects are ongoing and not yet included in this document.

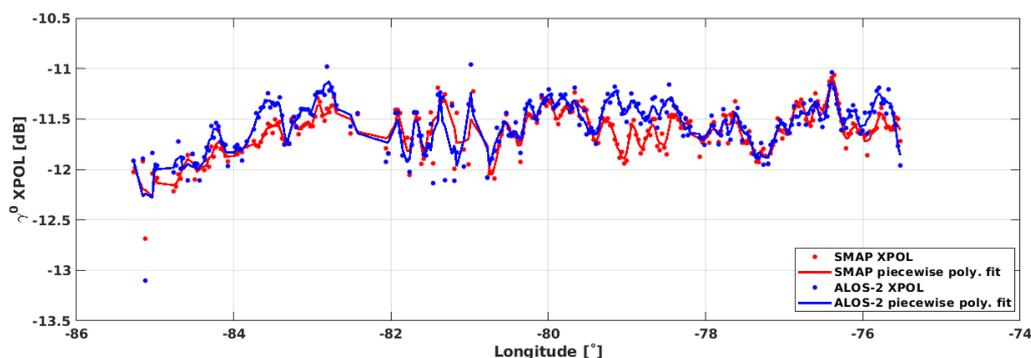


Figure 3-15: SMAP and ALOS-2 LHV backscatter mosaic for a single pixel line over the Amazon and second order polynomials fitted in moving windows to the observed trend.

In the meantime, we opted to minimize the striping effects observed over dense tropical rainforest using an empirical normalization approach, which considers L-band backscatter imagery acquired between April 14th and July 7th 2015 by the SMAP (Soil Moisture Active Passive) SAR mission. While the SMAP backscatter imagery was acquired with low (1-3 km) resolution, a major advantage of the SMAP SAR sensor was that backscatter was acquired with a constant incidence angle of 35° across the entire swath of ~1000 km width; note that SMAP deployed a rotating 6-m-aperture reflector antenna. Geocoded SMAP backscatter data was obtained at 3 km resolution and at HH, VV, and HV polarization from the National Snow and Ice Data Center (NSIDC). The backscatter images were provided in the form of 83 daily mosaics where each mosaic covered about 50 % of the Earth's land surface. In order to obtain a consistent global backscatter mosaic at each of the polarizations, the median backscatter at each pixel location across all acquisitions in the last 30 days of the SMAP SAR mission (29 mosaics) was calculated. At the equator, up to eight observations were considered, at high latitudes up to 25 observations were available at each pixel location. The design of the SAR with a rotating antenna implied a low resolution of the measurements towards Nadir, which is why observations within a range of +/- 100 km around Nadir were not considered. The resulting mosaic is shown in Figure 3-16 (top). Normalization of each HH and HV backscatter mosaic available from ALOS-2 (ScanSAR and FBD) over areas with continuous tropical forest cover entailed:

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	38	14.02.2019	

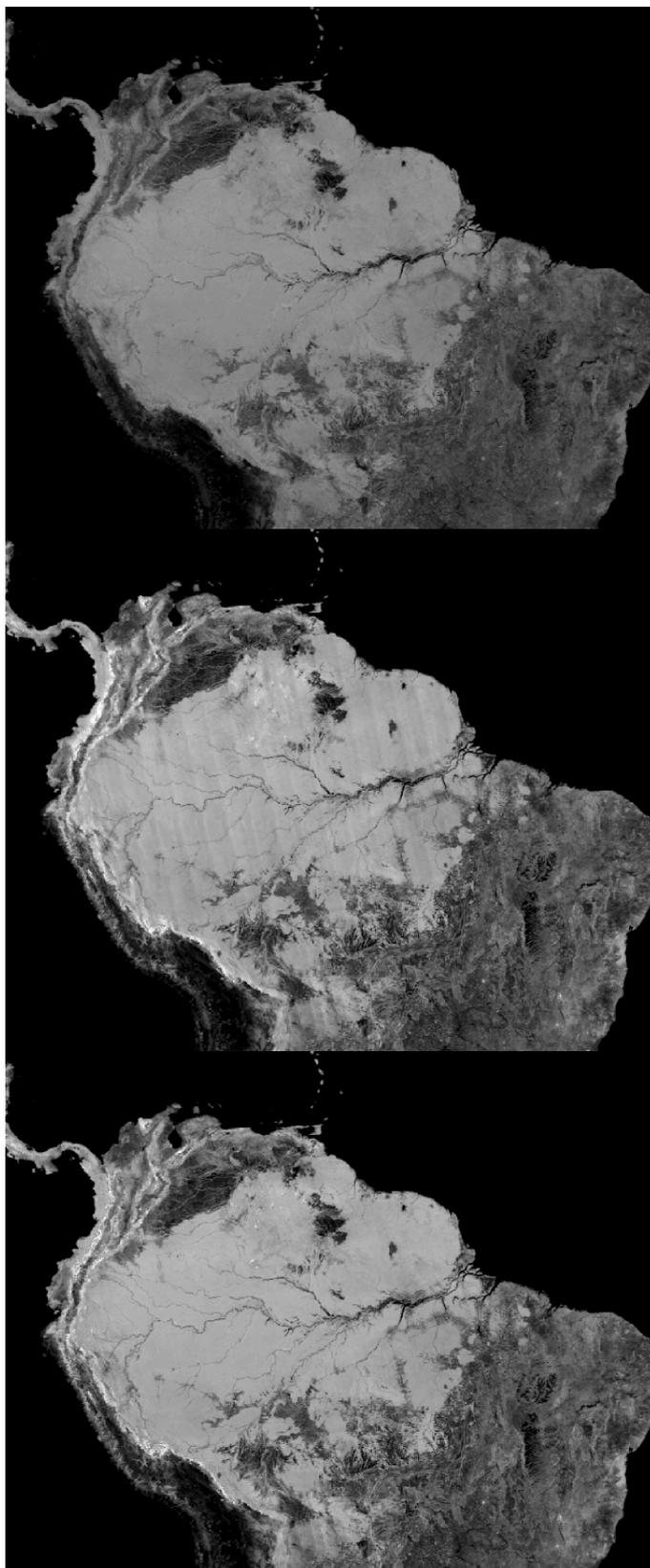


Figure 3-16: SMAP LHV backscatter mosaic (top), ALOS-2 L-HV mosaic before (top) and after (bottom) normalization with respect to SMAP.

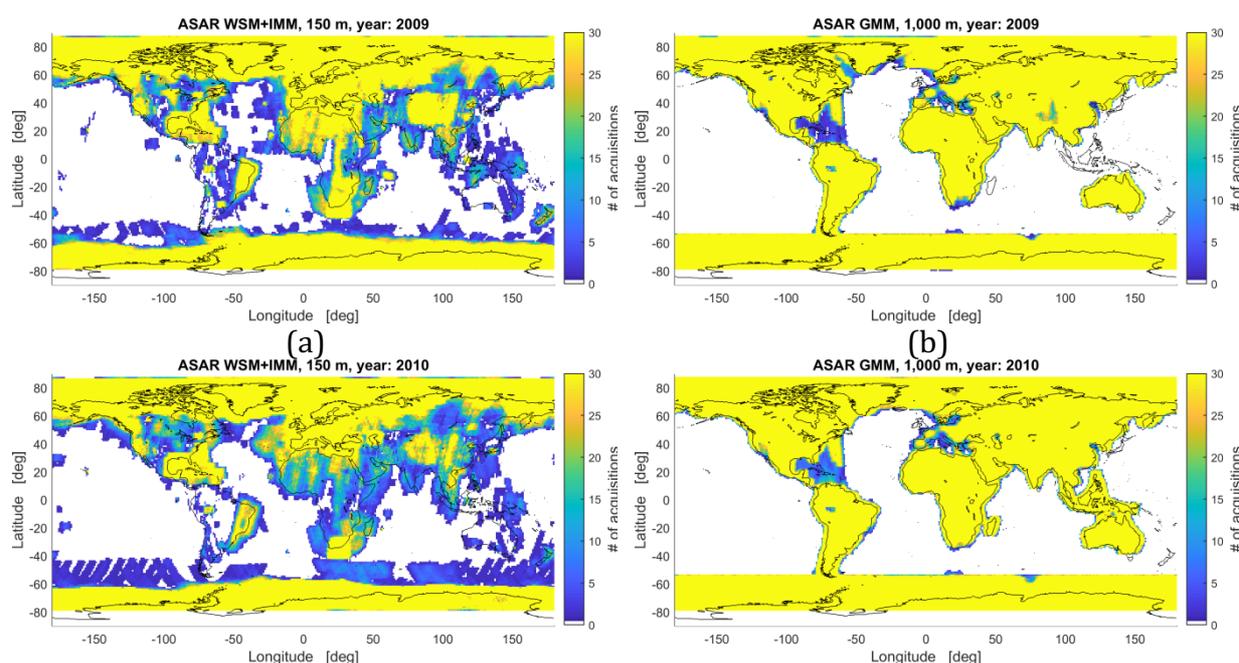
	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	39	14.02.2019	

- 1) aggregation of the ALOS-2 mosaics to the 3 km resolution of the SMAP SAR
- 2) identification of trends in L-HV backscatter (SMAP and ALOS-2) using a line-by-line moving window of 100 pixels window length (Figure 3-15). Within each window second order polynomials are fitted to the observed backscatter from SMAP and ALOS-2
- 3) normalization of the ALOS-2 backscatter (at full resolution) with respect to SMAP by subtracting line-by-line the polynomial trend from the actual observations and adding the residuals back to the polynomial backscatter trend from SMAP.

The normalization leads to more consistent backscatter mosaics over the tropics (Figure 3-16) with mostly removed incidence angle dependent trends. 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).

3.3 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 applications. 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°. From the high-resolution modes, ESA also generated Image Mode Medium (IMM) and Alternating Polarization Medium (APM) datasets at the same resolution as the WSM (Desnos et al., 2000), filling some of the gaps in the WSM coverage. The Global Monitoring mode (GMM), 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.



	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	40	14.02.2019	

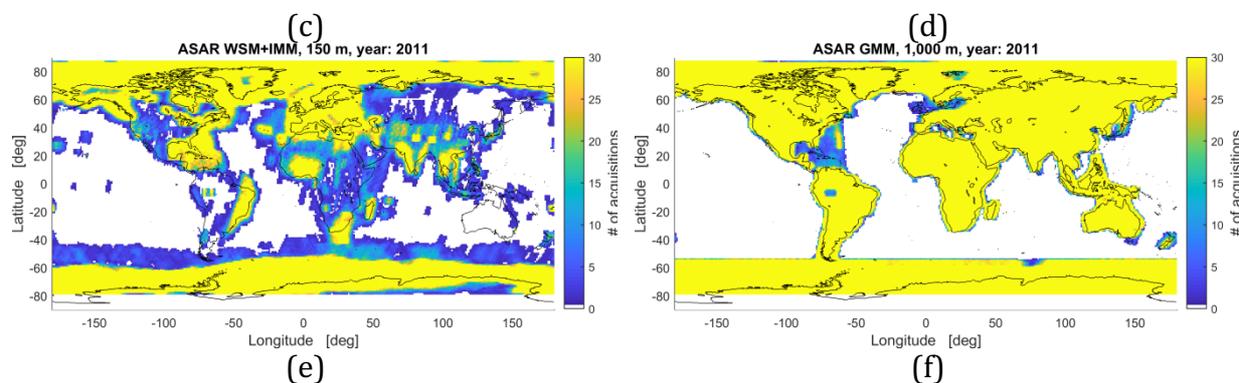


Figure 3-17: Map of ASAR acquisitions at 150 m (WSM and IMM combined) and 1,000 m (GMM) per 1°×1° grid-cell for each year between 2009 and 2011. The colour bar has been constrained between 0 and 30 observations to increase the image contrast in regions of poor coverage.

Figure 3-17 shows the number of ASAR observations over land for acquisitions at moderate resolution (150m) and at coarse resolution (1,000) m for each year between 2009 and 2011. (<https://wiki.services.eoportal.org/rss-storage-data.php>). We omit all other years because of similar spatial patterns. Nearly-global, dense datasets of backscatter were obtained in GMM. On the contrary, the acquisition patterns for modes operating at moderate spatial resolution was patchy, achieving higher frequency in polar regions, Europe and other regions of interest to research communities providing inputs to the acquisition plans by ASAR. Complementing all ASAR acquisitions lead to a global coverage with the exception of some islands in Oceania and isolated islands in the southern hemisphere.

All ASAR data acquired in IMM, WSM and GMM over land have been pre-processed in past research projects to form stacks of co-registered images of SAR backscatter. To support the estimation of AGB in CCI Biomass, we shall rely on ASAR data acquired between 2009 and 2012 to achieve a truly global coverage and a hyper-temporal dataset sufficiently reliable to retrieve biomass. In this way, the amount of images acquired under the most favourable conditions to retrieve biomass is maximized both in cold or dry regions as well as in wet tropical environments (Santoro et al., 2015a).

The pre-processing sequence implemented for the ASAR dataset followed the same strategy applied to process the Sentinel-1 data. The ASAR data were pre-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 3-18) (Santoro et al., 2011; Santoro et al., 2015a). To this end, a global tiling grid tied to the geographic reference system and having its origin at 90° N, 180° W, with tiles of relatively small size was used. The IMM and WSM images (spatial resolution approximately 150m×150m and pixel size 75m×75m) were 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 tiling grid was used for the pre-processing of the Sentinel-1 data as well. The GMM images (spatial resolution approximately 1 km×1 km and pixel size 500 m×500 m) were instead 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).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	41	14.02.2019	

SAR pre-processing

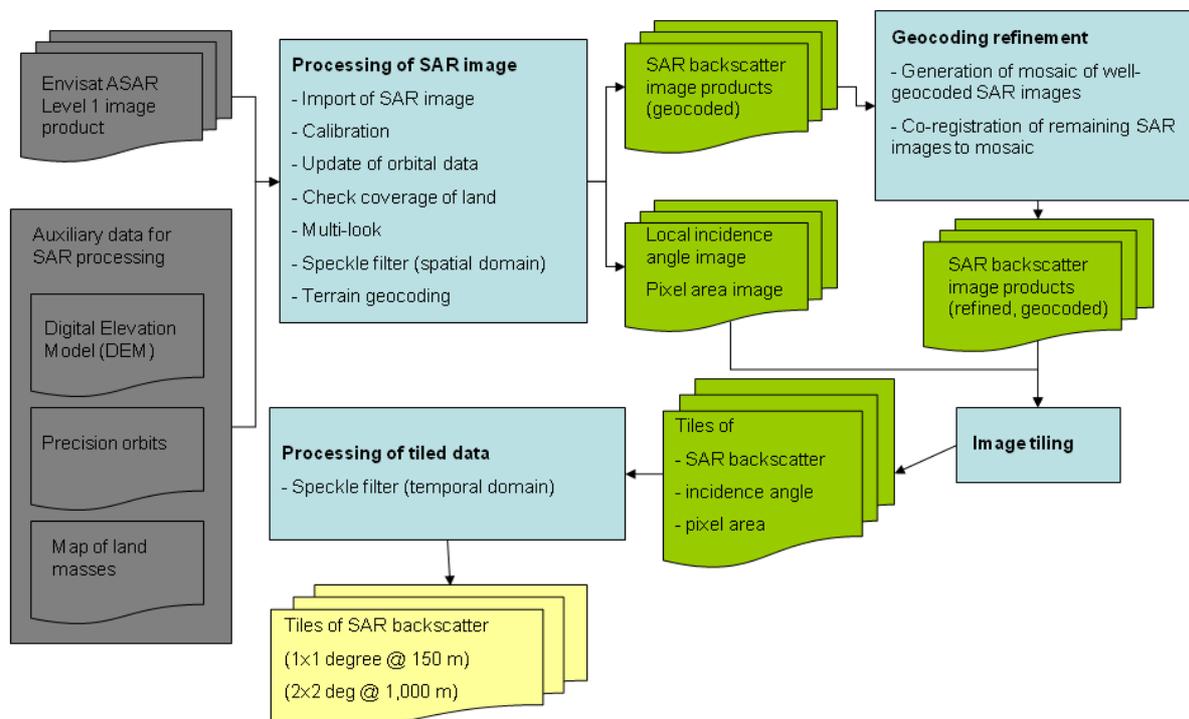


Figure 3-18: Flowchart of ASAR pre-processing.

The ASAR dataset consisted of images of radar backscattered intensity (β^0) in ground range geometry. Each image was first calibrated to σ^0 using factors provided by ESA in the image metadata. Orbital state vectors were 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 was checked against a coarse resolution map of land masses; images acquired over the ocean were discarded. Each image was then multi-looked, i.e., 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) was applied to obtain an estimate of the local average backscatter, which was later used as a weight in multi-channel speckle filtering.

Each ASAR image was geocoded to the geographic projection using the look-up table procedure described for geocoding Sentinel-1 images. For additional information, refer to Santoro et al. (2015b). In addition to the LUT generation, we applied cross-correlation techniques to compensate for offsets between the geometry described in the LUT and the true geometry. This step was necessary for ASAR data even after the correction for precise orbits. Offsets were estimated between the SAR image and an image of the SAR backscatter simulated from the DEM. The map of offsets in range and azimuth direction was then modelled with a four-coefficients polynomial; the resulting 2-D model of offsets was finally applied to the LUT so to obtain a refined version, which was eventually used to terrain geocode the SAR image. For images covering mostly flat terrain and characterized by poor co-registration with the simulated SAR image from 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 ensured that all images were geocoded with sub-pixel accuracy.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	42	14.02.2019	

Each ASAR image and the corresponding images of local incidence angle and pixel area were 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) was 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. Differently than in Quegan & Yu (2001), where the local estimates of the backscattering coefficient were obtained by means of a moving average applied to the intensity images, here we used 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 3-17, a truly global coverage with ASAR data could be obtained only by merging the GMM with the IMM and WSM datasets. Since the amount of observations from the GMM was much larger than those acquired at moderate resolution, it made sense to achieve the truly global coverage at 1,000 m spatial resolution by complementing the GMM pool of observations with WSM and IMM observations multi-looked from 150 m to 1,000 m. While the correct procedure to obtain a data stack at coarse resolution would have required multi-looked each image acquired in IMM and WSM to 1 km and then terrain geocoding to 1,000 m, we decided to simply multi-look and resample the already geocoded WSM and IMM images to avoid substantial pre-processing activities that could not be accommodated at the time of the ASAR image processing. As a result of such a decision, the multi-channel filtered geocoded WSM and IMM images and the corresponding images of incidence angle and pixel area were multi-looked with factor 7×7 and resampled to the 0.01° pixel size of the GMM images using bi-cubic interpolation.

This merged dataset was used in the GlobBiomass project as predictor in the BIOMASAR-C algorithm in order to guarantee wall-to-wall coverage and reliability of the estimates. The drawback was loss of spatial detail and approximations in regions with patchy landscapes. Although we could not explicitly assess the impact of spatial resolution on the AGB estimates, visual assessments of the GlobBiomass map for 2010 and the CCI Biomass map for 2017 in regions where the contribution of C-band was strong indicated higher quality and level of detail in the latter. For this reason, we are re-considering the use of ASAR data in the context of CCI Biomass to map AGB in 2010. Although the coverage at 150 m is not optimal to allow for unbiased estimates of AGB in large parts of the world, the 1,000 m dataset would probably have an impact on the temporal consistency of AGB estimates across the last decade. At the time of compiling the document, the use of both the merged dataset at 1,000 m and the 150 m dataset are considered with the scope of generating two sets of AGB estimates and assess the temporal consistency with respect to the estimate obtained from Sentinel-1 in 2017 and 2018.

Regardless of the final spatial resolution, slope-induced distortions need to be accounted for to avoid artefacts in the AGB maps in consequence of too low or too high backscatter. Because of the moderate-to-coarse spatial resolution, the procedure proposed by Frey et al. (2013) and applied to the Sentinel-1 data could be relaxed. The terrain geocoded backscatter was normalized by compensating for the effective pixel scattering area and local incidence angle as follows (Wiesmann et al.; 2004; Ulander, 1996; Castel et al., 2001).

$$\gamma^0 = \sigma^0 \frac{A_{flat}}{A_{slope}} \left(\frac{\cos\theta_{ref}}{\cos\theta_{loc}} \right)^n \quad (3-2)$$

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	43	14.02.2019	

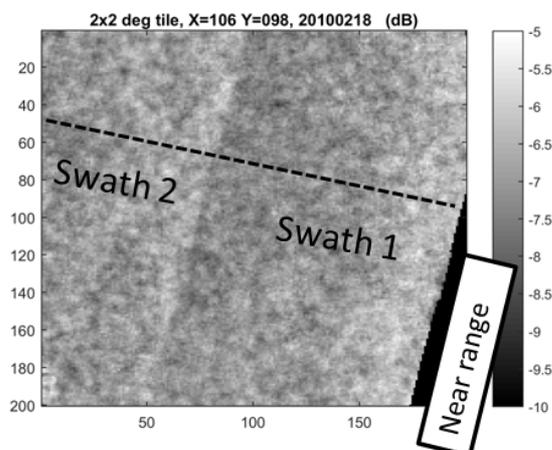
In Equation (3-2), θ_{loc} and θ_{ref} represent the local incidence angle and a reference incidence angle (e.g., the incidence angle at mid-swath) respectively. A_{slope} and A_{flat} represent the true pixel area and the local pixel area for theoretically flat terrain respectively. The images of the area normalization factor (A_{flat}/A_{slope}) and the local incidence angle were 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). It is noted that $n=1$ was also applied when correcting the Sentinel-1 data for slope-induced terrain and by JAXA in their mosaics processing sequence. The impact of local incidence angle on the SAR backscatter and on biomass retrieval is further addressed in Section 4. Again, the two-step approach proposed by Hoekman & Reiche (2015) according to which the correction of slope-induced effects is tuned with land-cover based empirical functions is not considered to be feasible in this context as it would require the biomass to be known a priori to select the appropriate correcting function.

One major issue with the ASAR GMM and WSM 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. This issue is critical in environments where the backscatter difference between unvegetated terrain and dense forest is 1-2 dB. Figure 3-19 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 drawn on the ASAR image shows the clear offset at the swath intersection.

Since the calibration of the ASAR data cannot be reversed, the only procedure to avoid that radiometric offsets become biomass offsets is to apply a crude 1-D moving median filter (length: 11 pixels) on each line of a geocoded image prior to biomass retrieval. The filtering is limited to dense tropical rainforest and moist forests where such offsets would lead to offsets of 100 Mg ha⁻¹ or more. Currently, the filter is applied to pixels labelled as tropical wet and tropical moist by the Food & Agriculture Organisation (FAO) Global Ecological Zones (GEZ) dataset (Section 3.11). Figure 3-20 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.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	44	14.02.2019	

Example for an ASAR GM1 image covering dense forest (VCF: 80%)



Profile of SAR backscatter along transect

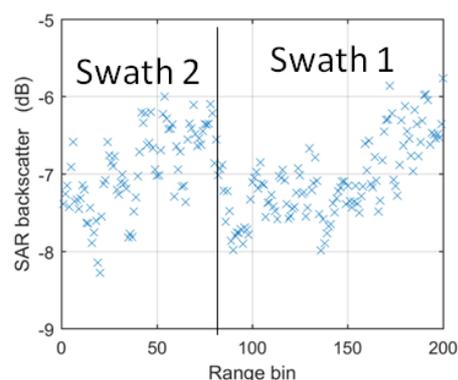


Figure 3-19: Illustration of backscatter offset along the seam between two adjacent swaths of an ASAR GM1 image covering dense tropical forest in the Amazon (left panel). The profile of the backscatter along the dashed line superimposed to the ASAR image is showed in the panel on the right hand-side.

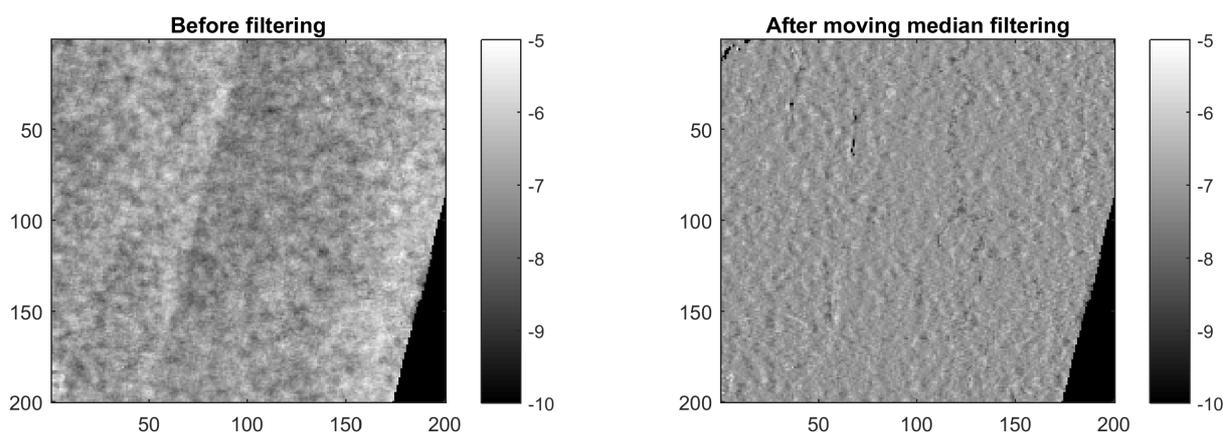


Figure 3-20: Example of the ASAR GM1 used in Figure 3-19 before and after filtering with a moving median filter.

3.4 ALOS-1 PALSAR-1

The ALOS-1 mission operated between 2006 and 2011 and operated the PALSAR-1 radar instrument. Differently than any other sensor on a spaceborne platform, PALSAR-1 was used to generate consistent, large-area coverage each year at high-resolution in the Fine Beam mode (25 m) and moderate resolution with the Wide Beam mode (100 m). Each year global and repeated acquisitions were scheduled during seasons that were known to maximize the information content of the backscattered signal with respect to land surface properties. In FB mode, PALSAR-1 could acquire data in single polarization (HH) and dual polarization (HH and HV), covering a swath of approximately 70 km. In the WB mode, data could be acquired only in HH-polarization but covering a swath of approximately 250 km.

Because of the data policy applied by JAXA to ALOS-1 and ALOS-2 data, only a limited number of images can be obtained free of charge, which hinders any global application. As for ALOS-2 (see

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	45	14.02.2019	

Section 3.2), global coverages of ALOS-1 PALSAR-1 data can only be obtained in the form of yearly backscatter mosaics for the FB mode and per-cycle mosaics (46 days) for the WB mode.

The ALOS-1 PALSAR-1 dataset used in the context of CCI Biomass consists of

- yearly mosaics of HH and HV polarized backscatter acquired in FBD mode between 2007 and 2010
- per-cycle mosaics of HH-polarized backscatter acquired in WB mode between 2007 and 2010

All mosaics were generated and distributed by JAXA (Shimada & Ohtaki, 2010; Shimada et al., 2014). Each FBD mosaic includes data acquired between May and October of the given year. To achieve global land coverage, gaps were occasionally 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, any of the mosaic datasets provides:

- the HH (FBD and WB) and HV (FBD only) 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
- indication of whether the pixel is land or water

Figure 3-21 shows the HV backscatter for 2010; bright tones correspond to forested areas, while dark tones correspond to bare areas and water surfaces. With respect to the ALOS-2 yearly mosaics, the ALOS-1 datasets present less striping due to the higher success rate of ALOS-1 acquisitions and the less frequent necessity of gap filling with data acquired in other years or from multiple seasons (personal communication, A. Rosenqvist).

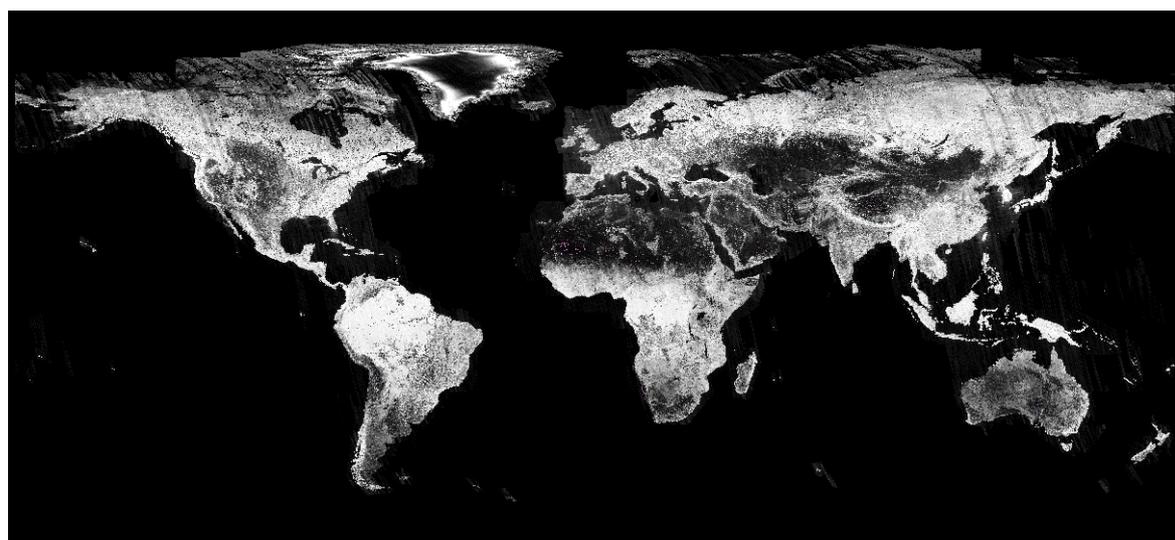


Figure 3-21: ALOS-1 PALSAR-1 mosaic of HV images acquired during 2010.

As for the ALOS-2 mosaics, the ALOS-1 mosaics 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.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	46	14.02.2019	

The weighted feathering approach applied by JAXA to smooth the backscatter differences between adjacent orbital tracks distorted 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.

3.5 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 3-22. 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.

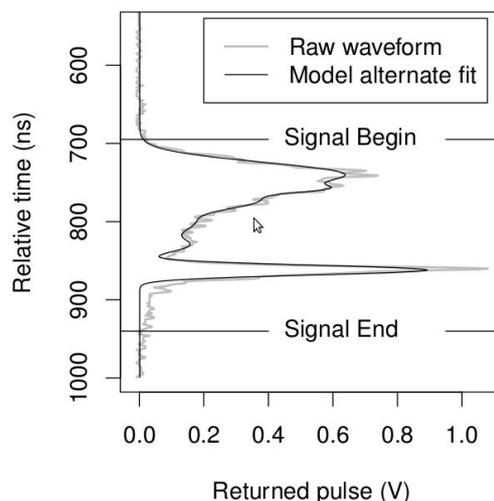


Figure 3-22: An ICESAT GLAS waveform showing the vertical distribution of returned energy from a forest (from Los et al., 2012).

In CCI Biomass, we use the dataset prepared in the GlobBiomass project (refer to the ATBD of the GlobBiomass project for details on the GLA14 product; RD-8). The GLA14 product (version 34), which provides altimetry data for land surfaces only to which geodetic, instrument and atmospheric corrections have already been applied. For each footprint, up to six Gaussians were used to model the raw waveform, as described in Hofton et al. (2000).

The GLAS data were used to estimate canopy density (CD) estimated as the ratio of energy received from the canopy (returns above the ground peak) to the total energy received and the height (h) as the distance between the ground peak and signal beginning ($RH100$) (Figure 3-23).

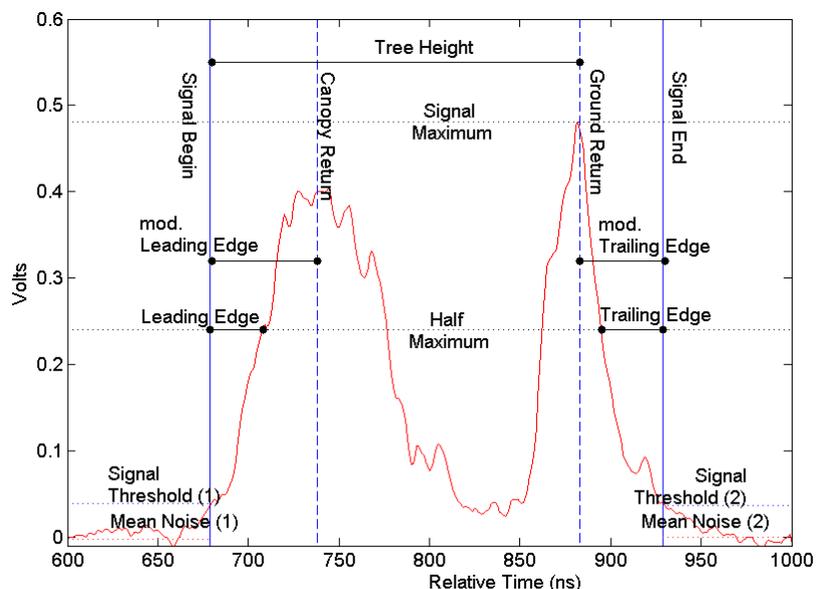


Figure 3-23: An ICESat GLAS waveform, showing the waveform metrics used to calculate RH100 & CD (Hilbert & Schullius, 2012)

Forest height was computed 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 [RD-8]. The remaining GLAS database contained estimates of RH100 for ca. 26.5 million footprints; their distribution is illustrated in Figure 3-24. 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.

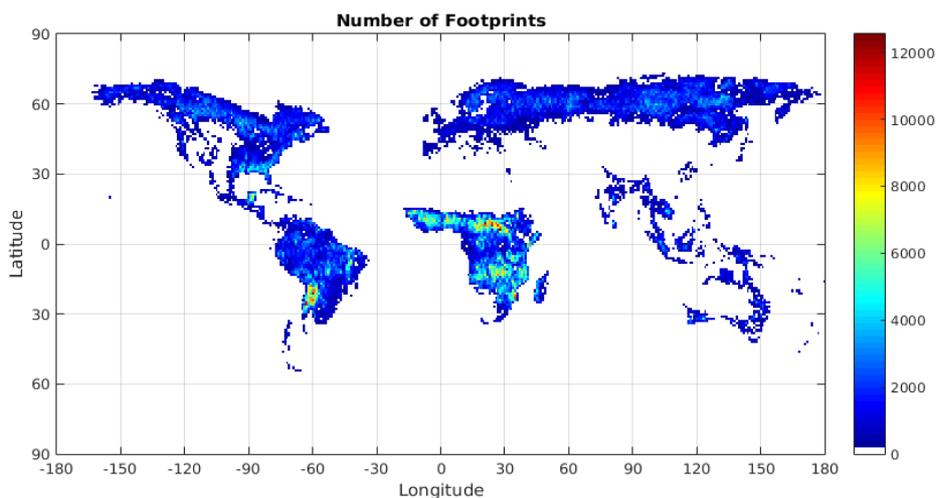


Figure 3-24: Number of ICESAT GLAS footprints after screening of GLA14 product.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	48	14.02.2019	

3.6 Digital Elevation Model

The DEM (Figure 3-25) used in the pre-processing phase of the SAR data and to analyse the quality of the retrieved biomass consists of a global dataset of $1^\circ \times 1^\circ$ tiles collated from various sources to form a seamless and gap-free dataset of surface elevation (de Ferranti, 2009). The dataset has a pixel spacing of 3 arc-seconds, corresponding to 90 m at the Equator. For regions between 60°N and 56°S , the DEM consists of gap-filled 3 arc-seconds SRTM elevations; refer to de Ferranti (2009) for details on the gap-filling approach. For latitudes north of 60°N , the elevations consist of a blend of datasets (topographic maps, coarse and high resolution DEMs, optical imagery) selected according to which has the best quality in a given region (de Ferranti, 2009).

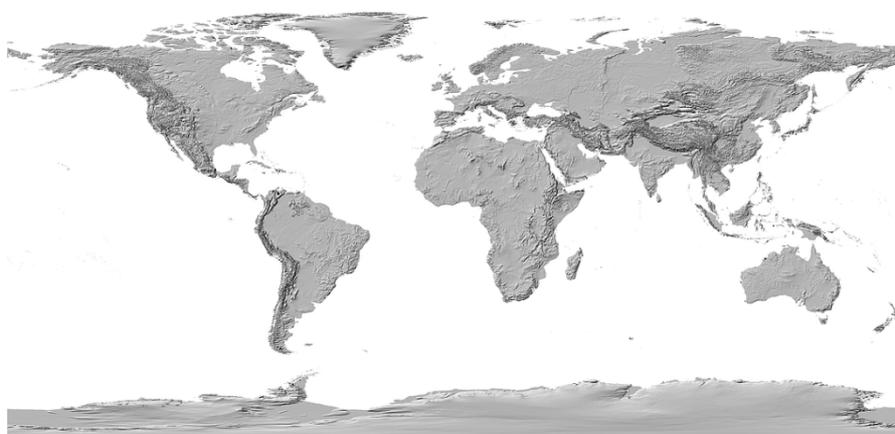


Figure 3-25: Global seamless DEM based on elevation datasets available at <http://www.viewfinderpanoramas.org>.

To support the terrain geocoding of the Sentinel-1 images to 150 m pixel size, the DEM was resampled using cubic resampling, as implemented in the `gdalwarp` tool of the Geospatial Data Abstraction Library (GDAL). The same procedure was applied to the DEM to terrain geocode the ASAR images to 150 m and 1,000 m.

Following indications by de Ferranti (2009), the consistency of elevations reported in this global dataset is superior to other global elevation datasets. In spite of some inaccuracy, we still find this DEM to be the most reliable option for pre-processing and analysis. An analysis of recently released global WorldDEM based on TanDEM-X interferometric data (Rizzoli et al., 2017) revealed somewhat higher thematic accuracy and spatial details than our global DEM but also a range of artefacts of various types (wrong height in correspondence of water bodies, voids etc.). As artefacts would propagate to the retrieved biomass, we still prefer using a void-free and clean DEM.

3.7 MODIS Vegetation Continuous Fields

The MODIS Vegetation Continuous Fields (VCF) product (MOD44B) (DiMiceli et al., 2015) 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 between 2000 and 2017 with a spatial resolution of 250 m. Version 6 is the current version. Data are provided in an sinusoidal projection in tiles of 4800×4800 pixels from the Global Land Cover Facility (GLCF) and can be accessed via the following website

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	49	14.02.2019	

(https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod44b_V006). The dataset has been resampled from 250 m to 300 m to match the CCI Land Cover dataset and allows nested gridding with the Sentinel-1 and ASAR datasets with a pixel size of 150 m. The use is foreseen in the training phase of the models relating C-band backscatter to biomass (see Section 4).

3.8 Landsat canopy density and density change

A global Landsat-based canopy density map representing forest state in 2000 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 and used in regression tree models, trained with the aid of very high resolution imagery (e.g., Quickbird) classified to forest/non-forest classes. 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 2017, as well as gain throughout the entire period. The forest cover change database includes a 30 m water body map, but no information was available about how this map was generated.

Canopy density and related changes are mapped globally at 1 arc-second pixel posting. Data are available at <https://earthenginepartners.appspot.com/science-2013-global-forest>

Using the same algorithm, the United States Geological Survey (USGS) and the University of Maryland, Department of Geographical Sciences, released 30 m resolution raster data layers for circa 2010 of canopy density and bare ground from Landsat 7 ETM+ data (Figure 3-26). The canopy density and bare ground data are per pixel estimates, 1 to 100% (given as integers values 1-100). Data description and access information can be found under <https://glad.umd.edu/dataset/global-2010-tree-cover-30-m> (last consulted on 13 January 2020)

The use is foreseen in the training phase of the models relating C-band and L-band backscatter observations to biomass (see Section 4).

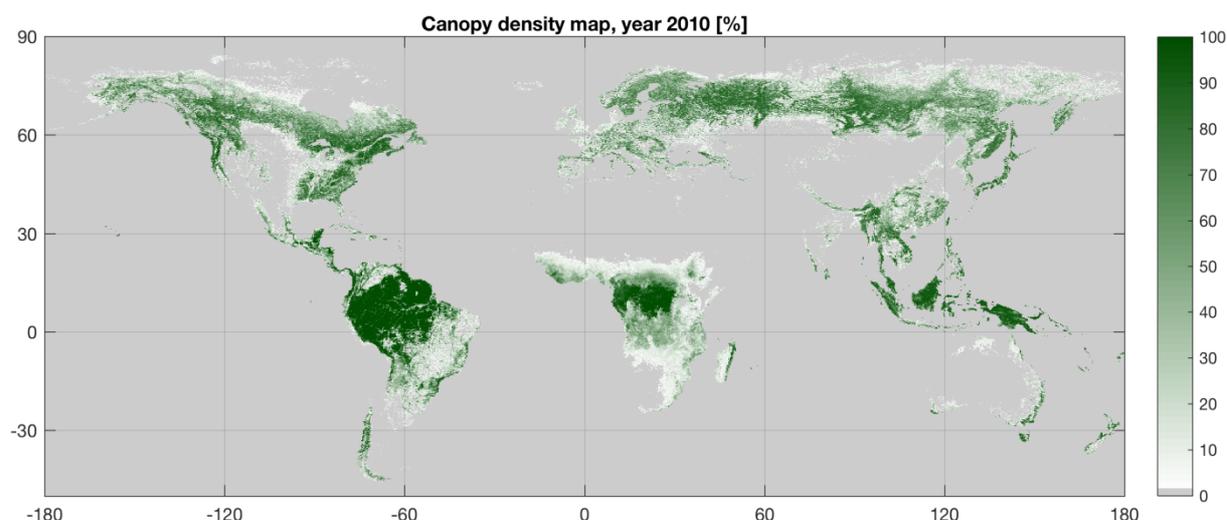


Figure 3-26: Canopy density map for the year 2010 produced from Landsat.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	50	14.02.2019	

3.9 CCI Land Cover map

Land cover information is used during training of the biomass retrieval models to mask out specific classes. More specifically, the BIOMASAR approach requires the identification of areas with low canopy density and only belonging to a vegetation type of class. Water bodies, urban areas and permanently snow-covered areas are characterized by very low canopy density. If not flagged, their backscatter values can distort the histograms used to estimate the parameters of the model relating SAR backscatter to biomass (see Section 4). At C-band, the SAR backscatter of water bodies presents some of the highest and lowest backscatter values. 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. At L-band, the backscatter of water bodies and ice-covered terrain is much lower compared to the backscatter of other types of unvegetated terrain. Built-up areas have high co-polarized backscatter and very low cross-polarized backscatter.

The Climate Change Initiative Land Cover (CCI-LC) project has released annual land cover maps between 1992 and 2015 (<https://maps.elie.ucl.ac.be/CCI/viewer/>) based on optical spaceborne datasets (Figure 3-27). The land cover maps are provided in equiangular projection with a pixel size of 1/360th of a degree in latitude and longitude. For CCI Biomass, the data layers of 2010 and 2015 were resampled to the geometries of the SAR datasets. For this, the GDAL gdalwarp tool was used with nearest neighbour resampling.



Figure 3-27: Illustration of CCI Land Cover maps (<https://www.esa-landcover-cci.org>)

The overall accuracy of the yearly land cover dataset was reported to be slightly above 70% (Product User's Guide under <https://www.esa-landcover-cci.org>). 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.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	51	14.02.2019	

3.10 Worldclim Bioclimatic Variables

The Worldclim data base (www.worldclim.org) developed by Fick and Hijmans (2017) includes a set of global maps of 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 3-4). Similarly to the GlobBiomass project, the maps are used to support the estimation of the maximum biomass across forests in different ecosystems (see Section 3.11). The nineteen maps depicting annual trends in temperature and precipitation have been aggregated to a $2^{\circ} \times 2^{\circ}$ grid.

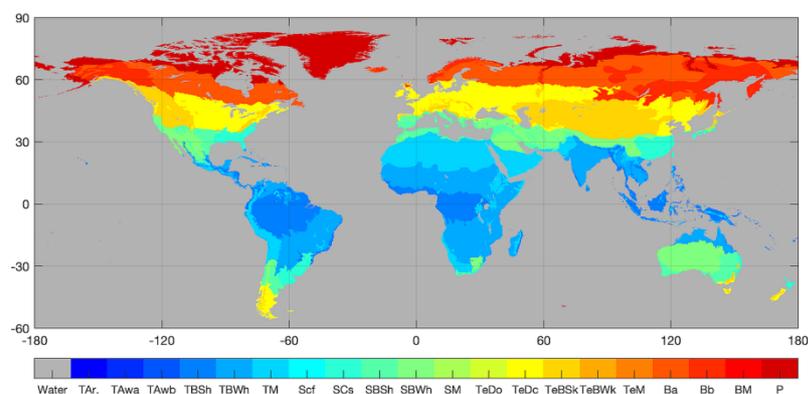
Table 3-4: Overview of BioClim variables.

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
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

3.11 FAO Global Ecological Zones dataset

The Global Ecological Zones (GEZ) dataset produced by the FAO (Simons, 2001) divides the land surface into 20 zones with “broad yet relatively homogeneous natural vegetation formations, similar (but not necessarily identical) in physiognomy (Figure 3-28). 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.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	52	14.02.2019	



- TAr = Tropical rainforest
- TAwa = Tropical moist deciduous forest
- TAwb = Tropical dry forest
- TBSh = Tropical shrubland
- TBWh = Tropical desert
- TM = Tropical mountain
- SCf = Subtropical humid
- SCs = Subtropical dry
- SBSH = Subtropical steppe
- SBWh = Subtropical desert
- SM = Subtropical mountain
- TeDo = Temperate oceanic
- TeDc = Temperate continental
- TeBSk = Temperate steppe
- TeBwk = Temperate desert
- TeM = Temperate mountain
- Ba = Boreal coniferous
- Bb = Boreal tundra woodland
- BM = Boreal mountain
- P = Polar

Figure 3-28: FAO GEZ dataset

The GEZ dataset is used to stratify the estimation of some of the model parameters of the retrieval algorithms (see Sections 3.11 and 4.1.2.3). To this scope, the GEZ dataset is rasterized on-the-fly to the geometry of the SAR images used to retrieve biomass. The re-projection is done with the program `gdal_rasterize` of GDAL.

3.12 Terrestrial Ecoregions of the World

The very broad definition of ecoregions in the GEZ dataset does not allow for fine tuning of retrieval algorithms. To this scope, we also consider the better delineation of vegetation in the Terrestrial Ecoregions of the World (TEOW) dataset (Olson et al., 2001). Compared to the GEZ dataset, the TEO dataset also brings in ecological properties of the landscape. The TEO dataset divides the Earth land surfaces into 825 ecoregions (Figure 3-29). These are categorized within 14 biomes and eight biogeographic realms (Figure 3-30). The boundaries of each ecoregion, biome and realm correspond to the original extent of natural communities prior to major land-use change.

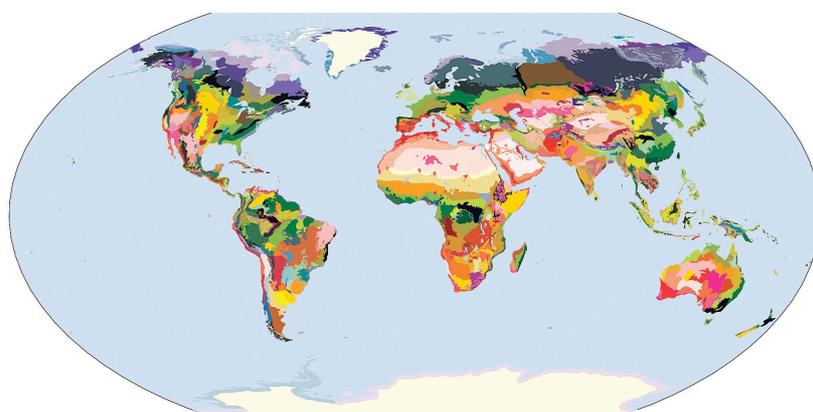


Figure 3-29: Terrestrial ecoregions of the world (Olson et al., 2001).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	53	14.02.2019	

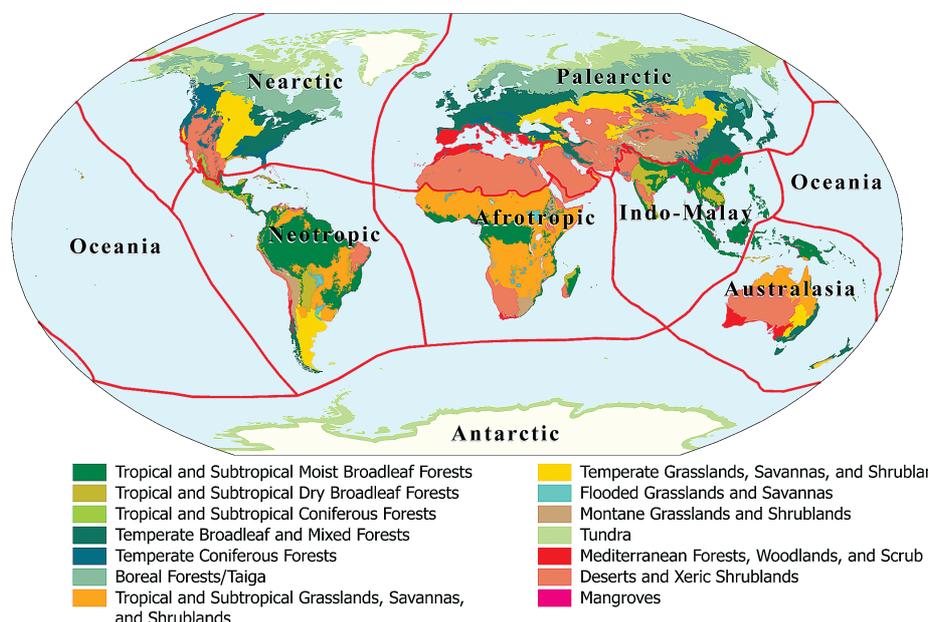


Figure 3-30: Grouping of the terrestrial ecoregions into biomes (see legend) and realms (see map). Picture reproduced from Olson et al. (2001).

As for the GEZ dataset, the TEOW is rasterized on-the-fly to the geometry of the SAR images used to retrieve biomass. The re-projection is done with the program `gdal_rasterize` of GDAL. This dataset is also used to segment the ICESAT GLAS dataset in the process of estimating the model coefficient relating canopy density and RH100 measurements from the individual waveforms (see Section 3.16.1).

3.13 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 6.1 m with underestimation in dense, tall, broadleaf forest (e.g., the Amazon). Spatially explicit estimates of forest canopy height are here used to support the merging of biomass estimates obtained with the individual retrieval approaches described in Section 4.

3.14 Biomass of dense forest and maximum biomass

The retrieval algorithms aim to minimize the dependence on *in situ* information about biomass to train the models relating the SAR backscatter to biomass. However, knowledge of the maximum biomass in a certain (eco)region is required to constrain the retrieval based on spaceborne SAR imagery within a realistic range of biomass values. To this end, we use a spatial database reporting estimates of biomass assumed to be representative for the densest forests. This value is assigned to the centre of each tile in a regular $2^{\circ} \times 2^{\circ}$ grid. Where available, *in situ* measurements from field plots or spatially explicit datasets of GSV or AGB were used. The biomass of dense forests was then defined as the 90th percentile of the histogram within the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	54	14.02.2019	

2°×2° area (Santoro et al., 2011). Interestingly, we identified a rather robust scaling factor between this parameter and the maximum biomass of 1.2. Elsewhere, it was estimated with an empirical piece-wise linear function (Santoro et al., 2015a) starting from values of the average biomass reported at provincial or national level. For tiles including several provinces or nations, the average biomass representative for the tile was obtained by weighting the individual averages by the area of each within the tile. In regions where numbers based on in situ measurements were unavailable, but we could gather more than one map of AGB (preferably based on laser scanning observations), we estimated the biomass of dense forest as the joint 90th percentile of the histogram of the map values. For areas lacking any form of biomass estimates, the IIASA FAO 0.5° dataset of global biomass was used. The value for the biomass of dense forests was then set equal to the maximum of the 16 values within the 2°×2° large tile. Figure 3-28 shows the origin of the biomass values used to compute the biomass of dense forests. Given that the database contained both GSV and AGB observations, we applied the Biomass Conversion and Expansion Factor (BCEF) computed in the GlobBiomass project (Section 3.15) to convert between the two variables (AGB = BCEF * GSV). Figure 3-31 shows the origin of the estimate of the biomass of dense forests.

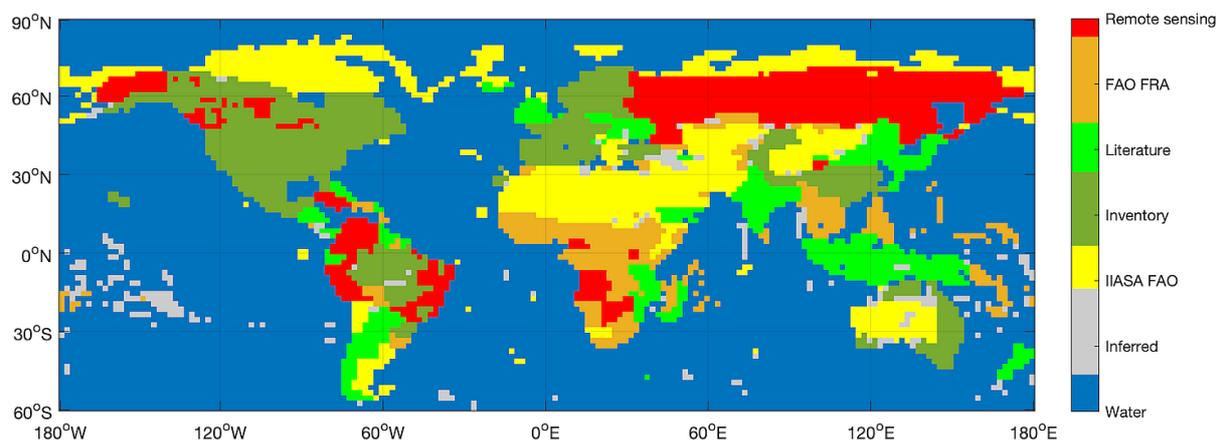


Figure 3-31: Map detailing the origin of the numbers used to generate an estimate of the biomass of dense forests at the spatial scale of 2°.

The procedure implemented to characterize the maximum biomass globally was very crude in its essence and should be considered a first-order approximation. Attaching a single value to areas covering thousands of km² was assumed to be sufficient to characterize it in regions including a dominant type of vegetation. In fact, we assumed that the spatial variability of the maximum biomass would be small. This approximation fails in regions with transitions between ecosystems (e.g., tropical forest and savannah). Ultimately, ensuring spatial consistency of the estimates could not avoid that the dataset has errors and uncertainties propagating from the input datasets, the rules implemented to estimate the biomass of dense forests from the data available in a given tile, and the BCEF values.

To reduce some of the weaknesses in the dataset, additional processing steps were, therefore, carried out, whose aim was to:

- Improve the reliability of the database, 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 maximum exist
- Fill gaps in the 2-degree data base
- Increase the resolution to better depict smaller scale variations in the maximum GSV, such as are expected for transition areas between tropical rainforest and savannah

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	55	14.02.2019	

The maximum biomass 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 of forest management. In order to verify/improve the initial estimates and to fill gaps, a database of predictor layers that are expected to have predictive power for maximum biomass was compiled at $2^\circ \times 2^\circ$ resolution, including the nineteen WorldClim bioclimatic variables and the ICESAT GLAS observations of canopy density and height. For each $2^\circ \times 2^\circ$ 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 the Height Of Median Energy; HOME). RandomForest (Breiman, 2001) models were then developed for each FAO ecoregion using the initial estimates in our database as response and the WorldClim and ICESAT GLAS layers as predictors. Once calibrated, the models were then used to predict the GSV of dense forests globally at a resolution of $0.2^\circ \times 0.2^\circ$ (Figure 3-32). The assessment of this dataset is reported in the ATBD of the GlobBiomass project [RD-8].

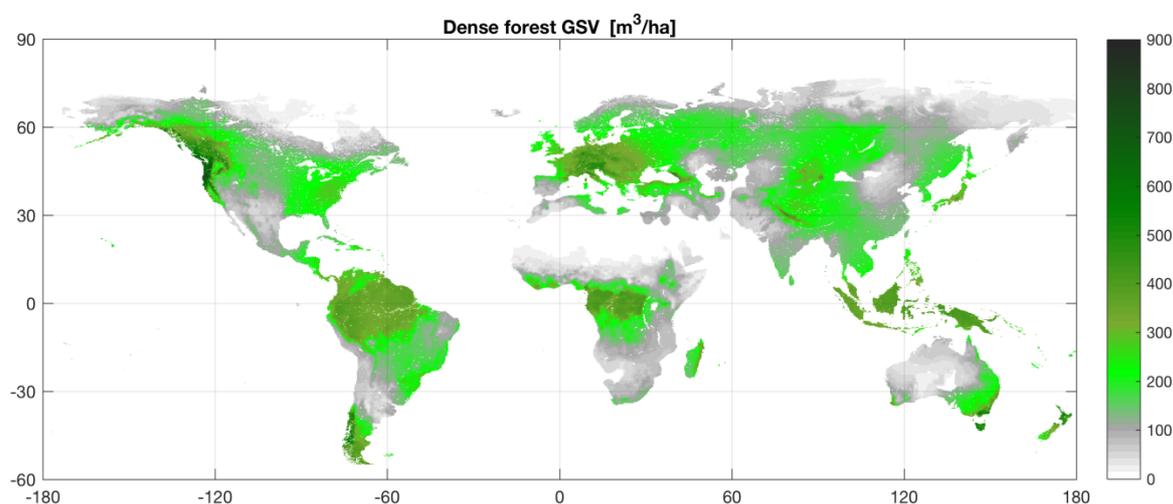


Figure 3-32: Map of the GSV of dense forests with a spatial resolution of 0.2° .

Using the scaling factor of 1.2 between GSV of dense forests and maximum GSV as well as the BCEF relating GSV and AGB (Section 3.15), we generated a global layer of maximum AGB.

3.15 Biomass Conversion and Expansion Factor

The BCEF represents a scaling factor that allows the conversion of GSV to AGB. It is the product of wood density and a factor representing the fraction of total biomass with respect to stem biomass. Characterization of the BCEF therefore requires knowledge of both wood density and the allometric relationship between stem mass and whole above-ground mass, including branches and leaves.

Traditionally, BCEF values provided by the Intergovernmental Panel on Climate Change (IPCC) are used whenever detailed measurements and equations for the estimation of wood density and biomass expansion factors are lacking (IPCC, 2006). Values of BCEF have been published for biomes and for GSV ranges based on a number of studies where the relationship between wood volume and woody biomass was investigated. This representation is, however, too coarse since both wood density and allometric relationships result from biological processes that respond strongly to variations in environmental conditions. Hence there is strong variation of wood

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	56	14.02.2019	

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.

In the GlobBiomass project, modelling of wood density and biomass expansion factors was pursued leading to global wall-to-wall datasets based on extensive sets of measurements published in ecological databases [RD-8]. The overall estimates of wood density were unbiased but tended somewhat to the mean and did not correctly represent extremes. The biomass expansion factors were based on highly generalized models relating total to stem biomass because the characterization of this relationship by measurement is very inhomogeneous in global terms. Figure 3-33 Figure 3-34 show the maps of wood density and biomass expansion factors, respectively.

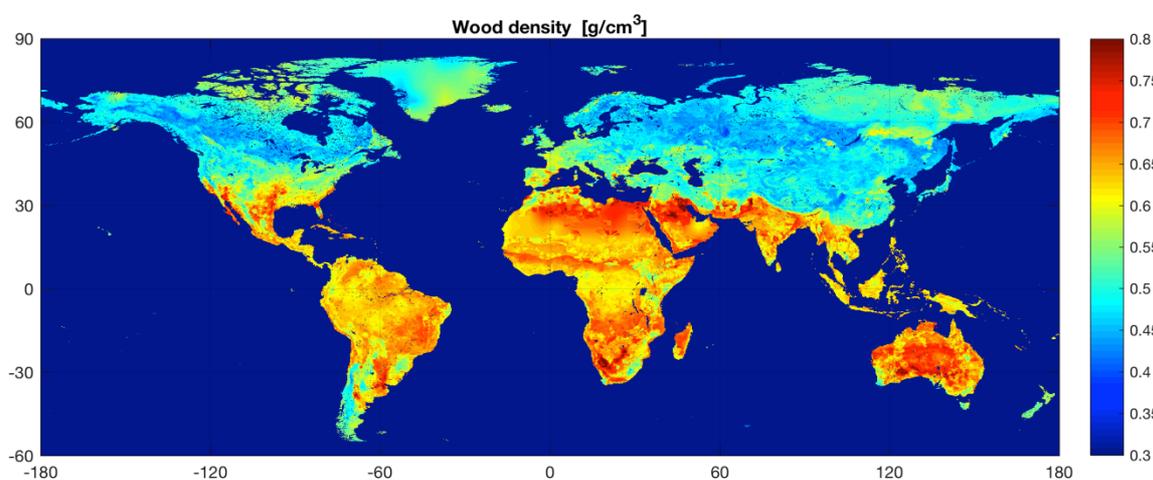


Figure 3-33: Global estimates of wood density. Pixel size: 0.0083333°, i.e., 30 arc-seconds.

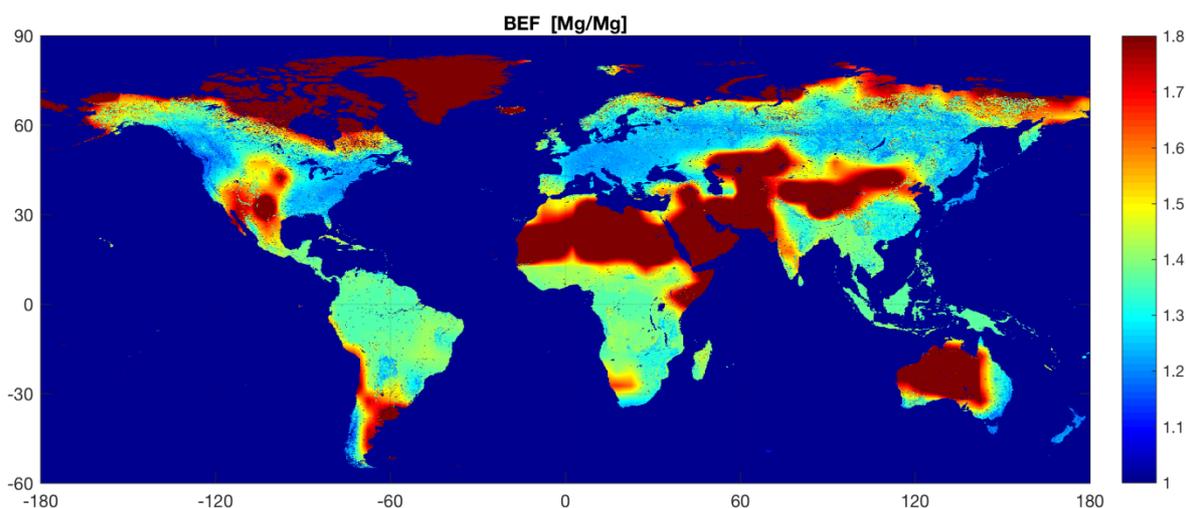


Figure 3-34: Map of BCEF applied to the stem biomass dataset obtained by multiplying the merged GSV dataset from the GlobBiomass project with the map of wood density in Figure 3-29. Pixel size: 0.0083333°, i.e., 30 arc-seconds.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	57	14.02.2019	

3.16 Allometries

Forest variables, such as diameter at breast height, tree height, canopy density, crown diameter, growing stock volume and above-ground density, are more or less correlated in the sense that much of the variability of one variable (e.g., biomass) can be explained in terms of the variability of one or a few other variables (e.g., tree height and diameter at breast height). Models can therefore be created that aim at generalizing the functional relationship between forest variables and allow estimation in regions where the output variable of interest is poorly characterized by in situ observations (Chave et al., 2005).

In Section 4, we present models relating the SAR backscatter to a number of forest variables. As our interest is to estimate AGB, we need to establish allometric functions that may (i) reduce the number of forest variables in the models and (ii) relate those forest variables to AGB. The models in Section 4, in particular, consider canopy density, tree height and AGB. Hence, we are here interested in two sets of functions:

Allometry between canopy density and height is here introduced to reduce the number of forest variables in the Water Cloud Model to height only.

$$\eta = f(h) = g(f(B)) \quad (3-3)$$

Allometry linking the AGB to height allows for an explicit formulation of the retrieval model in terms of AGB.

$$h = f(B) \quad (3-4)$$

The two sets of allometries are further described in this Section and are applied in the retrieval methods described in Section 4. The concept is described in the Algorithm Development Plan (ADP) of this project [RD-6].

This advance is possible thanks to the extensive dataset of ICESAT GLAS waveforms for each of which several height metrics and canopy density were estimated. Explaining AGB as a function of LiDAR height metrics is, however, a rather novel topic of investigation, with substantial potential, as demonstrated by a number of studies at several sites (Lefsky et al., 2005; Saatchi et al., 2011; Asner et al., 2012; Asner and Mascaro, 2014; Coomes et al., 2017; Labriere et al., 2018). However, some approximations are required, as AGB measurements are not available for ICESAT GLAS footprints and only global AGB maps can be used as proxy references.

3.16.1 Allometry between canopy density and height

In this study, we used a generic model explaining canopy density, CD , as a function of the height metric corresponding to the start of a GLAS waveform (RH100)

$$CD = 1 - e^{-qh} \quad (3-5)$$

The coefficient q in Equation (3-5) is empirical and describes the rate of change of canopy density for a change of tree height. Higher q means that canopy tends to close faster when trees grow.

To investigate the variation of q across the globe, the prepared ICESAT-GLAS dataset described in section 3.6 was used. Any remaining footprints in unvegetated areas were removed using the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	58	14.02.2019	

CCI-LC product for 2010 (<https://maps.elie.ucl.ac.be/CCI/viewer/>), the ICESat GLAS data were then divided according to the different WWF ecoregions (Section 3.12). Outliers were removed by calculating the logarithm of RH100 for each ecoregion and removing the upper and lower 5%. A least squares regression, using the scipy optimize curve fit function, was then undertaken to obtain q (Figure 3-35) for each ecoregion. The coefficient of determination (r^2) and root mean square error (RMSE) were also calculated for each regression. The variability in q as a function of leaf on and leaf off conditions was also investigated by dividing the ICESat GLAS dataset into classes of Northern Hemisphere ‘summer’ (April - September) and ‘winter’ (October – March).

The coefficient q shows a variation (0.032 – 0.175) across biomes with the lowest values found in the Tropical & Subtropical Savanna or Boreal biomes and the highest in Tropical & Subtropical Moist Broadleaf Forest and Mangrove biomes (Figure 3-35). In general, q values were higher in summer compared to winter although there are exceptions (Figure 3-36).

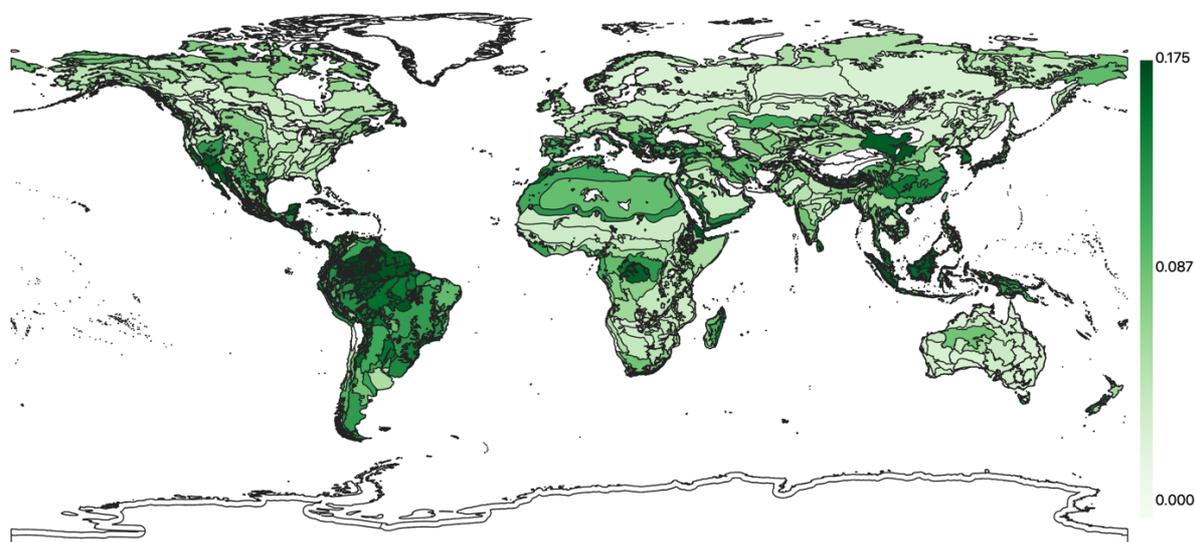


Figure 3-35: Map of q values per WWF ecoregion, obtained through least squares regression of available ICESat GLAS footprints within each ecoregion after filtering, ecoregions in white had no footprints

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	59	14.02.2019	

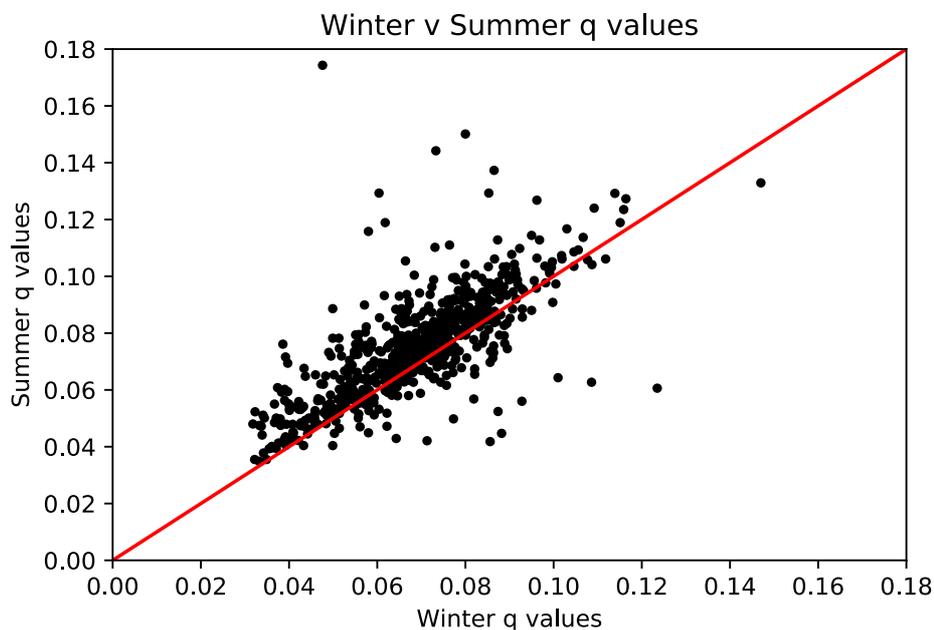
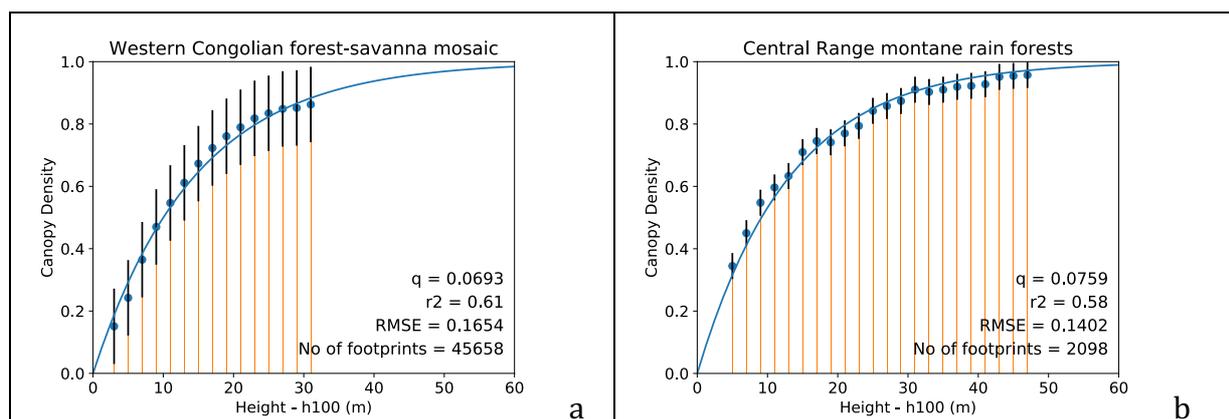


Figure 3-36: Scatterplot of N hemisphere summer v winter q values globally; the red line is the 1:1 line

In many ecoregions, canopy height and canopy density increases were related and reflected the different environmental settings: higher q values were found in wet tropical regions characterised by dense forest (e.g. Amazon, Congo, SE Asia) and lower values in sparser, more open savannas and grasslands (e.g. Sub-Saharan Africa, Mongolia; Figure 3-37). Even where the height range of the forests was relatively limited (e.g. Figure 3-37c) or a wide variety of forest structures occurred (e.g. Figure 3-37d) the same relationship was observed despite a larger spread of data around the regression line.



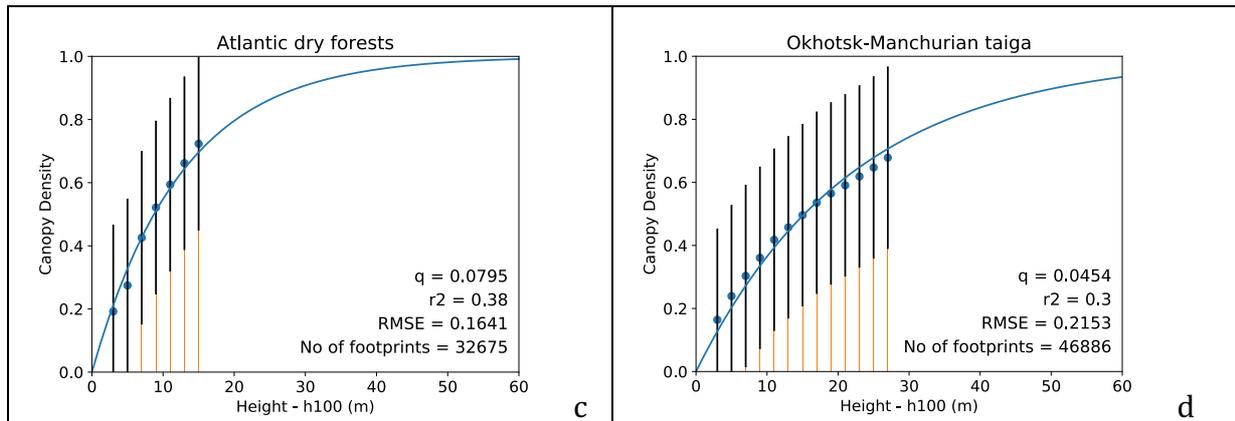


Figure 3-37: Examples of least squares regression denoted by blue line (extended to 60m height for comparison). Data was binned to 2m intervals with blue dots indicating mean of each bin and vertical line the interquartile range. Ecoregions showing good fit are Western Congolian forest-savanna mosaic (a) and Central Range montane forests (b). C represents Atlantic dry forests and shows good fit of the means despite wide spread of data and heights below 20m, d shows good fit of the means despite ‘noisy’ data in Okhotsk-Manchurian taiga.

However, some ecoregions require further investigation in order to better understand the relationship. For example, some boreal or sub-arctic zones present a more sinusoidal pattern imposed on the overall trend, potentially as a result of forest management (Figure 3-38a). Other ecoregions contain many types of forest and show large variation in the data. For example, the Eastern Cascades forests have high variation in vegetation structures, primarily influenced by edaphic processes and disturbance regimes (WWF ecoregions; Figure 3-38b).

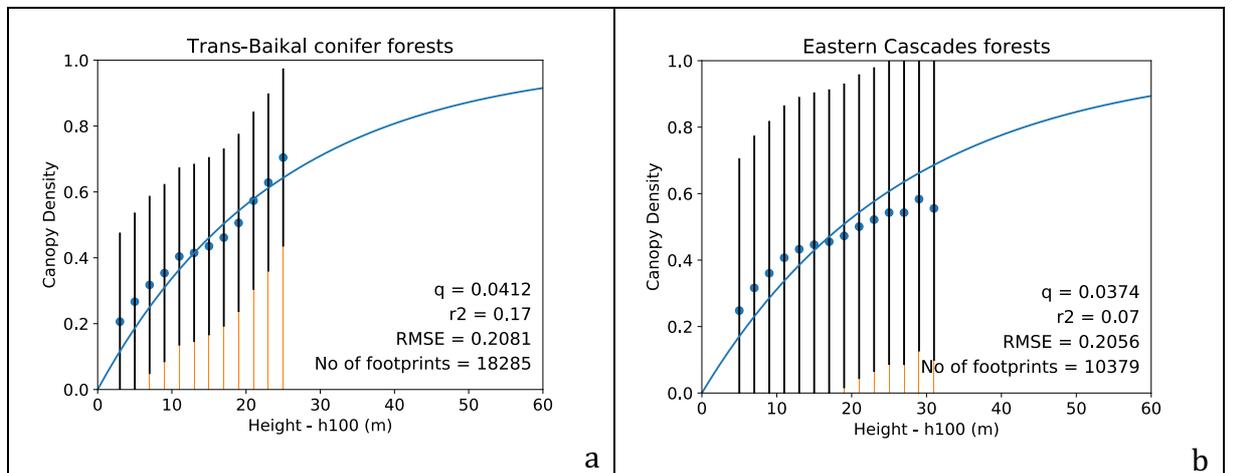


Figure 3-38: Examples of least squares regression denoted by blue line (extended to 60m height for comparison). Data was binned to 2m intervals with blue dots indicating mean of each bin and vertical line the interquartile range. Left hand plot from Trans-Baikal conifer forests demonstrates the sinusoidal pattern found in some ecoregions. The right-hand plot highlights the issue of large variation within the ecoregion, here for Eastern cascades forests.

Comparison of q values by biome indicates that it seems under-estimated in the boreal/taiga biome as it has a lower mean q value than both the Savanna and Tundra biomes which are generally characterised by sparser vegetation (Figure 3-39). A bimodal distribution of q values can also be seen in Savannas & Grasslands. This could be due to factors such as:

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	61	14.02.2019	

- High levels of agriculture in ecoregions such as the California Central Valley Grasslands, where only 1% of grasslands remain. This ecoregion is now the most productive agricultural region of the USA with a wide range of fruit and nut trees, with these creating a higher q value than expected for a grassland.
- Grasslands in proximity to rainforest having very few ICESAT GLAS footprints due to a lack of forest vegetation ($VCF < 4$): the footprints used to calculate the q values were situated on the edge of the ecoregion and therefore are more similar to the neighbouring ecoregion vegetation than the savanna. Figure 3-40 demonstrates this for Patagonian Steppe.

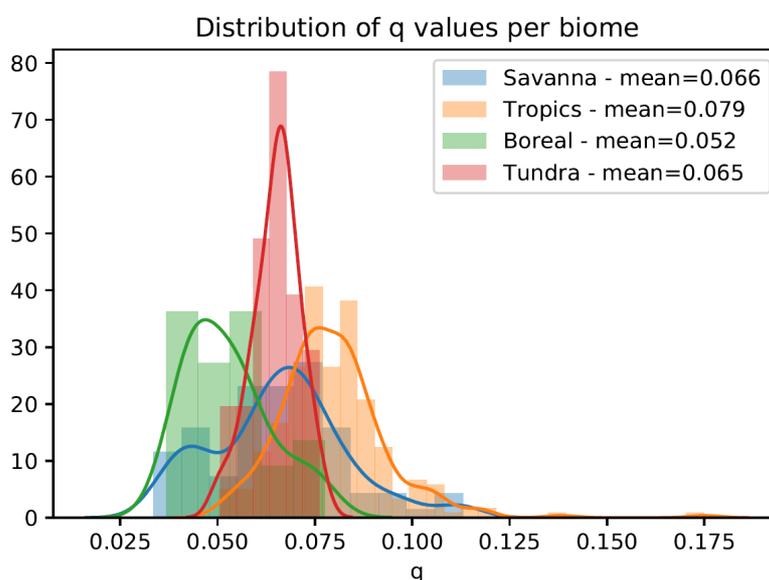


Figure 3-39: Histogram of the distribution of q values in different biomes, biomes were selected and regrouped from the WWF Ecoregions classification with; 'Savanna' representing Temperate, Montane and Flooded Grasslands, Shrublands and Savanna; 'Tropics' representing Tropical & Subtropical Moist & Dry Broadleaf and Coniferous Forests; 'Boreal' representing Boreal & Taiga and 'Tundra' representing the Tundra biome.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	62	14.02.2019	

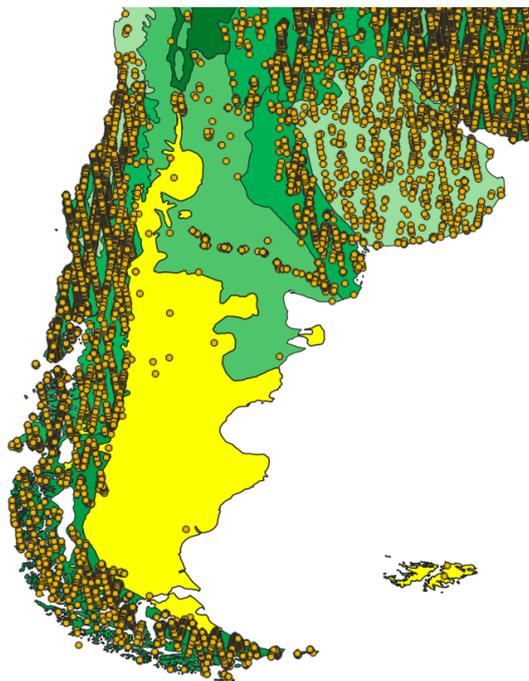


Figure 3-40: Map showing Patagonian Steppe ecoregion (highlighted in yellow) with ICESat GLAS footprints represented by the dots. This ecoregion has a high q value for a grassland (0.079) obtained from 676 footprints

Further consideration should also be made for regions where topographic variation occurs: topographic filtering was computed with slope from a 1 arcsec DEM, but this may not be sufficient to identify all footprints affected by topography. When comparing RH100 to the full waveform length outliers can be identified, which could be removed in future analysis, for example where the difference between the waveform length and estimated canopy height is above a selected value (Figure 3-41), or a proportion of total waveform length.

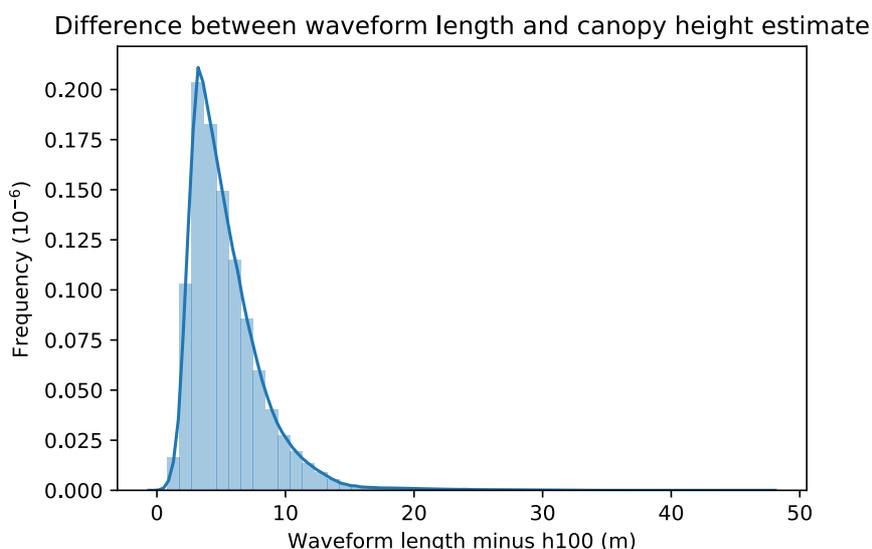


Figure 3-41: Histogram of the difference between ICESat GLAS waveform length and h100 canopy height estimates for filtered footprints

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	63	14.02.2019	

The experimental results indicate that the functional dependencies described in Equation (3-5) are maintained across the globe for the majority of ecoregions, with a rapid increase in canopy density with increasing height for dense tropical forests and a flatter slope for the sparser savanna and boreal regions. Further investigation will explore the environmental variables that affect these functional dependencies, in order to improve the estimation of the spatial variation in q across the globe.

From the set of estimates of the coefficient q for each ecoregion, we generated a spatially explicit raster map with a pixel size of 1,000 m by rasterizing the WWF dataset (Figure 3-42). Inpainting was applied to extrapolate values to surfaces not represented in the WWF dataset. At this stage, we preferred not to consider seasonal datasets of q as these are not yet fully understood. The raster dataset is then further resampled to the geometry and spatial resolution of the SAR data being used to retrieve AGB.

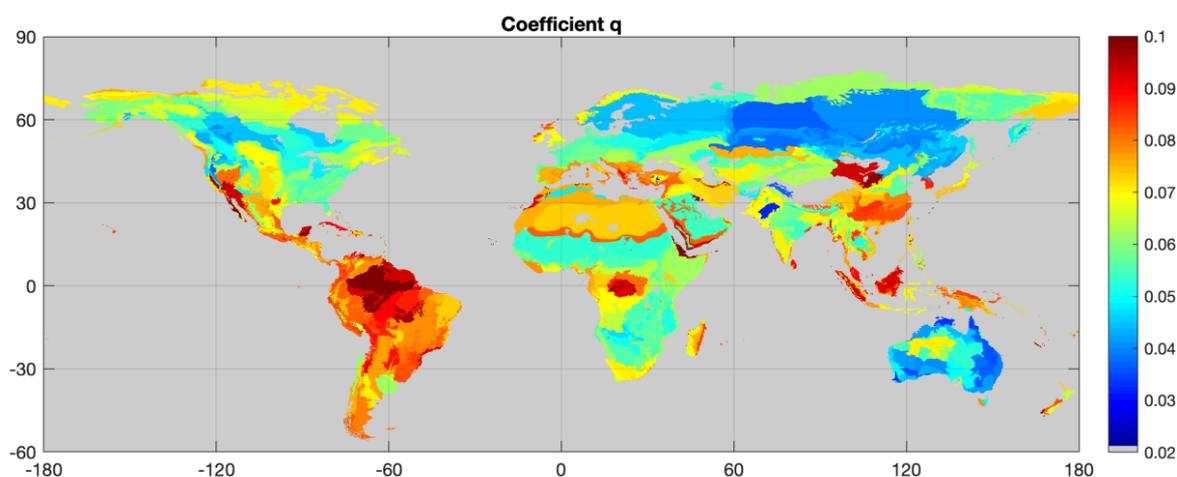


Figure 3-42: Map of the coefficient q derived from all ICESAT GLAS measurements.

3.16.2 Allometry between AGB and height

In practice, AGB at plot level is obtained from measurements of the trees' stem diameter, height, form factor, wood density and number of trees per unit area. An inventory of trees to compute AGB can be complicated from an economic and logistical point of view, especially in remote areas. For this reason, remote sensing measurement techniques have been developed in recent times (e.g., terrestrial laser scanners) that can allow for a more rapid quantification of a number of such parameters (Disney et al., 2019). Whenever the measurements of forest variables are "outsourced" to remote sensing, however, allometries (based on destructive sampling) start playing a role. Airborne laser scanning has been proven to be a reliable sensing technique, allowing for an accurate measurement of a number of parameters related to forest structure. Accordingly, studies are developing in the direction of characterizing the relationship between AGB and laser-based metrics. In particular, the relationship between canopy height and AGB has been investigated at several locations because of the direct sensing of the vertical structure of trees while other variables such as diameter of trees or the wood density cannot be measured. In this sense, efforts are being spent to set up models that are either able to predict these from the observations or to embed them somehow into the height-to-AGB models.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	64	14.02.2019	

Lefsky et al. (2005) found a linear relationship linking AGB to the square of maximum canopy height estimated from ICESAT GLAS waveforms in tropical forest of the Amazon, then showing good agreement between field measurements and predictions. Asner et al. (2012) proposed a generic power-law model relating above-ground carbon density, i.e., roughly half of the AGB, and a LiDAR metric referred to as mean canopy height (MCH) for tropical forests. The method appeared to be robust across four sites although the model had to be trained separately at each region to capture the region-specific forest structural properties in the power-law model. Saatchi et al. (2011) proposed a similar power-law function relating basal area weighted tree height, referred to as Lorey's height, and AGB to estimate AGB in tropical forest. The models were trained on a continent-by-continent basis using *in situ* observations and applied to ICESAT GLAS measurements and a number of image datasets to generate a map of AGB for the tropical regions. Asner and Mascaro (2014) proposed a set of global and regional equations relating the LiDAR metric top-of-canopy height (TCH) obtained from airborne observations to above-ground carbon density, in tropical forest of the Neotropics. Their conclusion was that global models can explain the variability of AGB with TCH but they are not able to characterize the variability at the level of single sites. They also observed that Lorey's height is flawed in open canopy forest and can potentially generate wrong biomass estimates. Coomes et al. (2017) expanded the work by Asner and Mascaro to Southeast Asian forest and demonstrated that the power-law function is still applicable but needs further adaptation to the site conditions. They also showed that an explicit use of canopy gap information derived from the laser measurements improves the retrieval of carbon density. Labriere et al. (2018) used airborne laser scanned data and *in situ* observations to test a power-law function, expressed as a log-log linear function, in tropical forests of Gabon and French Guyana, relating AGB to several height metrics, including TCH and MCH.

While these studies could profit from dedicated campaigns where *in situ* data and laser measurements were collected, an attempt to use an allometry relating height metrics and AGB at the global scale can build upon such local studies but also requires simplifying assumptions and a great deal of generalization. The ICESAT GLAS dataset provides a decent sampling of canopy height globally; unfortunately, there are no measurements available at the footprint locations nor are there alternative databases containing height and AGB from the same location that would allow the characterization of the functional dependency between height and AGB at the global scale. The major limitation to establishing a functional dependency between AGB and height metrics globally, and thereby characterizing its spatial variability, can be overcome by relating the height metrics to spatially explicit estimates of AGB derived from remote sensing observations. We attempted this in our study using the GlobBiomass AGB dataset as reference. By doing so, we were fully aware that the errors of the GlobBiomass dataset might affect the estimates of the model relating AGB and canopy height, which would then propagate to the estimates of AGB. Nonetheless, our profound knowledge of the GlobBiomass AGB dataset and its errors were exploited when generating the allometric models here proposed to relate AGB and height.

The survey of literature dealing with biomass estimation based on LiDAR observations indicated that a power-law function relating AGB to the LiDAR height metric RH100 is a sensible way to proceed:

$$AGB = p_1 H^{p_2} \quad (3-6)$$

where p_1 and p_2 are regression coefficients and are estimated by means of non-linear least squares.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	65	14.02.2019	

Here, H represents RH100 derived from the ICESAT GLAS waveform. For each RH100 measurement, we extracted the AGB estimates from the GlobBiomass dataset corresponding to the laser footprint and computed an area-weighted mean of the estimates falling within the footprint. To avoid the temporal mismatch between a GLAS footprint and the corresponding AGB estimate affecting the regression, we discarded observations acquired before 2006. Indeed, restricting to the year closest to the GlobBiomass AGB dataset (2009) would have caused serious data gaps in the global coverage. Including GLAS observations from the last four years of the mission (2006-2009) meant that the coverage was global.

Our analysis was undertaken at a spatial resolution of 0.25° because of the reduced noise in the GlobBiomass AGB compared to the full resolution value and the reliable spatial patterns at such scale. The major caveat of the GlobBiomass dataset of increasing underestimation for AGB above 250 Mg ha^{-1} was handled by excluding any such estimate from the computations. For each 0.25° pixel, we computed the mean RH100 and the mean AGB. This set of “observations” was then used in Equation (3-6) to estimate the two unknown model parameters. To understand the spatial patterns of the height-to-AGB allometry, we divided the RH100 and AGB values according to several criteria, as discussed below.

We first split the data according to five major global ecological zones classes (tropical wet, tropical dry, temperate wet, temperate dry and boreal). For each of these five ecoregions, Equation (3-6) was fitted to value pairs of RH100 and AGB. To understand the impact of the number of GLAS footprints within a 0.25° pixel on the relationship between RH100 and AGB, we set a threshold on the number of footprints per pixel and repeated the regression each time (Figure 3-43). When considering all pixels, the spread of the observations is large and hardly any relationship between the two variables can be appreciated. Already with a threshold of 100 footprints, a functional dependency became visible. By increasing the threshold on the number of footprints within a pixel, the shape of the model fit changed only marginally unless the number of pixels used to estimate the model parameters was small and not uniformly distributed across the range of heights and AGBs. For tropical wet forests, the RH100 dataset was clipped at 25 m, roughly corresponding to an AGB of 250 Mg ha^{-1} to avoid the allometric function being distorted by underestimates of AGB in the high range of values. The high density of footprints in the wet tropics implied that the sensitivity analysis of the model fit upon number of footprints could be undertaken up to 500 footprints per pixel and showed strong stability of the estimates of p_1 and p_2 (see also Table 3-5 and Table 3-6). For the tropical dry forest, we identify two clusters of observations so that the model fit represents an average relationship between AGB and RH100, with an almost constant AGB for increasing RH100 after an initial rapid increase. For temperate wet forests, the shape of the model fit changed with the threshold on the minimum number of footprints due to the scarcity of pixels including more than 200 footprints (see also estimates of p_1 and p_2 in Table 3-5 and Table 3-6). Unlike the other ecological zones, tropical dry forests and boreal forests displayed a fairly linear relationship between AGB and RH100 (see also values of p_2 close to 1 in Table 3-6). For boreal forests, the result is somewhat in contradiction with previous studies (Santoro et al., 2002; Santoro et al., 2007; Askne and Santoro, 2012) that showed a non-linear relationship between height and biomass (in the form of stem volume). Our interpretation of the linear trend is that overestimates of AGB in the GlobBiomass dataset at around $50\text{-}100 \text{ Mg ha}^{-1}$ (Santoro et al., in preparation) has linearized a trend that otherwise would have been non-linear. In the absence of an alternative wall-to-wall dataset at high resolution that is unbiased in the boreal zone, we preferred using the estimates of p_1 and p_2 obtained with the model fit to the GlobBiomass dataset rather than extrapolating local functions, such as those obtained in Sweden (Askne and Santoro, 2012), to the entire boreal zone.

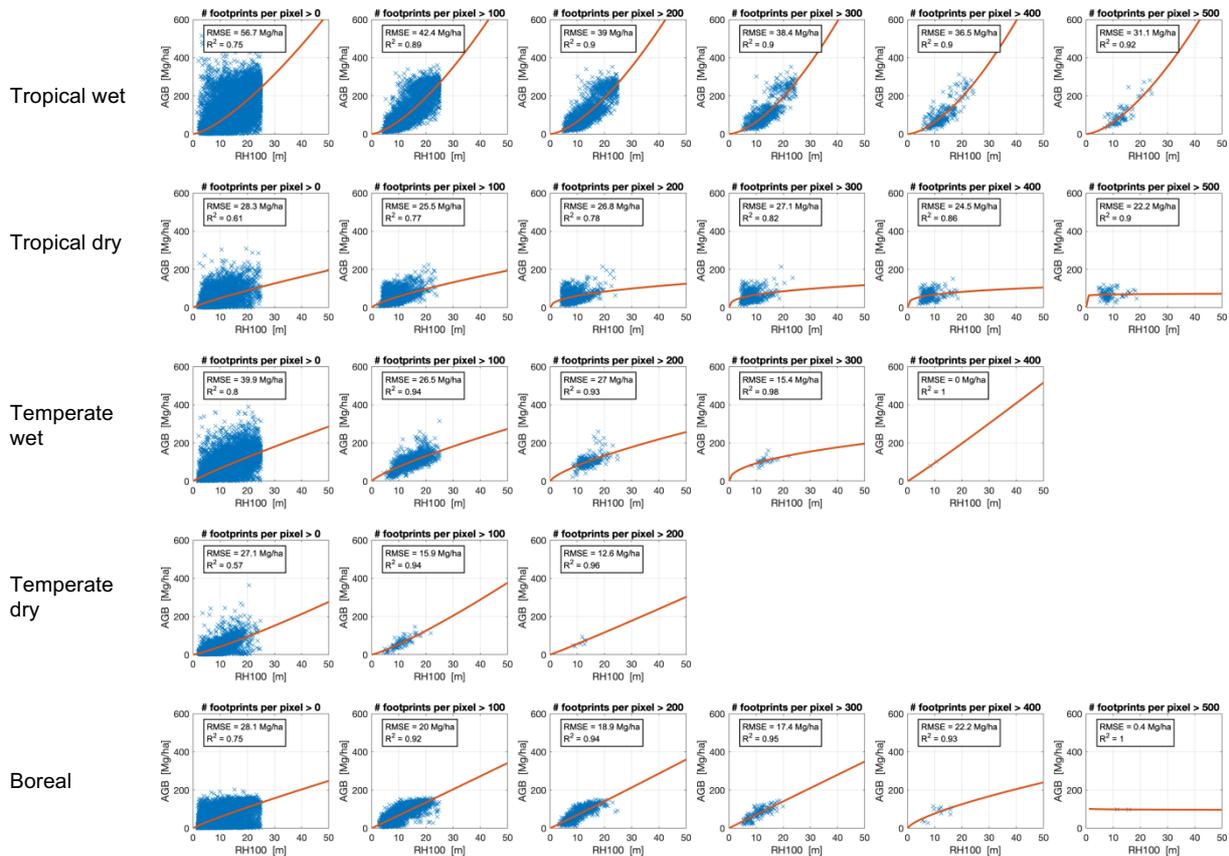


Figure 3-43: Illustrating the relationship between AGB and RH100 using the GlobBiomass AGB dataset and the ICESAT GLAS RH100 measurements averaged to 0.25° (blue crosses). The red curves represent the model fit to the AGB and RH100 values using Equation (3-6). Each panel represents a major FAO ecological zone (rows) and is characterized by a minimum number of footprints per 0.25° pixel. The estimates of the coefficients p_1 and p_2 are listed in Table 3-5 and Table 3-6, respectively.

Table 3-5: Estimates of the coefficient p_1 per ecoregion and minimum number of GLAS footprints within a 0.25° pixel.

		Number of GLAS footprints per 0.25° pixel					
		> 0	> 100	> 200	> 300	> 400	> 500
Ecological zone	Tropical wet	2.7544	2.2233	1.9446	1.5801	1.4957	1.6651
	Tropical dry	6.9638	10.5090	22.6443	30.3126	36.8574	63.9148
	Temperate wet	8.4097	11.9814	15.5787	32.2020	8.5486	-
	Temperate dry	2.8010	3.4957	5.89	-	-	-
	Boreal	7.4667	6.5502	6.5873	7.1631	15.0861	101.9097

Table 3-6: Estimates of the coefficient p_2 per ecoregion and minimum number of GLAS footprints within a 0.25° pixel.

	Number of GLAS footprints per 0.25° pixel
--	---

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	67	14.02.2019	

		> 0	> 100	> 200	> 300	> 400	>500
Ecological zone	Tropical wet	1.3914	1.4880	1.5296	1.6063	1.6222	1.5761
	Tropical dry	0.8524	0.7450	0.4364	0.3463	0.2686	0.0325
	Temperate wet	0.9015	0.7998	0.7169	0.4621	1.0482	-
	Temperate dry	1.1740	1.1963	1.0412	-	-	-
	Boreal	0.8956	1.0107	1.0226	0.9937	0.7072	-0.0140

Figure 3-43 suggested that the model fit obtained with at least 200 footprints per pixel is the most reliable except for temperate dry forests, which are well characterized with at least 100 footprints per pixel. To understand more about regional differences in a given ecological zone, we further fitted Equation (3-6) to values of RH100 and AGB per major ecological zone as well as per continent (Africa, America, Eurasia, Southeast Asia and Australia). Figure 3-44 shows a comparison of model fits for the different continents. The model fits are presented for forest heights up to 50 m to reveal possible continent-specific trends. However, when interpreting the model fits in the different panels, one should consider that tree heights above 30 m are unusual in dry and boreal forests whereas trees can grow as tall as 100 m in wet regions.

In the tropical wet ecozone, we observe a slightly lower rate of increase of AGB with RH100 in Southeast Asia and Australia. In tropical dry forest, we observe two separate curves for Africa and other continents, which explains the two clusters of observations illustrated in Figure 3-43. In the temperate zone, we observe different trends in America and Eurasia, whereas the temperate wet forests of Southeast Asia and Australia do not present any similarity with the other continents probably due to the small number of pixels belonging to this ecozone (but well clustered along the predicted line). In boreal forests, the model fits obtained for Eurasia and America do not seem to present any significant difference. An explanation of the different trends is beyond the scope of this project also bearing in mind that some of the trends may have been affected by deficiencies of the GlobBiomass dataset. However, it is worth investigating in future studies the global aspects of the height-to-AGB allometry; in this sense, an important analysis would be to repeat this study with an unbiased AGB dataset at high resolution.

The results illustrated in Figure 3-43 and Figure 3-44 indicate that characterizing model fits on an ecoregion basis at the level of continents can be of help but does not seem a robust approach to be considered in the retrieval models developed here. We also attempted fitting Equation (3-6) at the level of smaller blocks (between $10^\circ \times 10^\circ$ and $40^\circ \times 40^\circ$) but the results were ambiguous, probably due to the too small number of 0.25° pixels within each block, thus not allowing an accurate estimate of the coefficients p_1 and p_2 . For this reason, we preferred, at this stage, a conservative approach and used the estimates of p_1 and p_2 determined at the ecozone level without further geographic constraints.

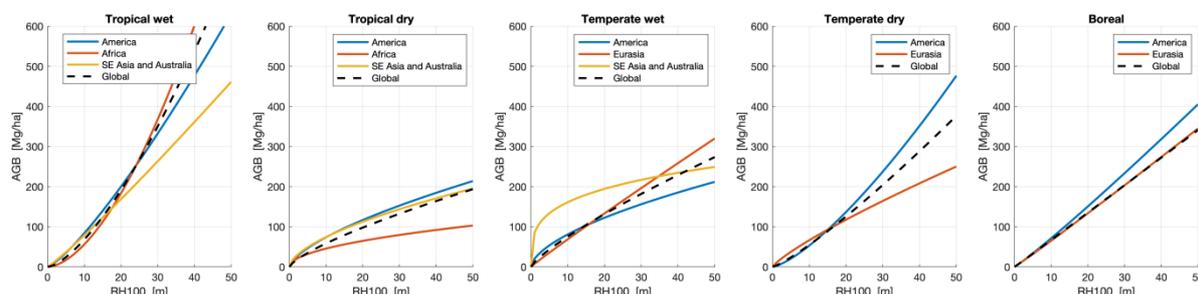


Figure 3-44: Model fits of Equation (3-6) per major ecological zone and continent at 0.25° spatial resolution. The dashed lines represent the model fits obtained for an entire ecological zone (number of footprints > 200 except for the temperate dry ecozone with a minimum number of footprints per pixel of at least 100).

To understand the reliability of our predictions, we compared them with allometries relating RH100 (or TCH) with AGB published in literature. Figure 3-45 shows our prediction (blue line) and predicted models from local studies and based on *in situ* data (red lines) for each major ecoregion. Most published allometries have been developed in tropical wet forests (Lefsky et al., 2005; Asner et al., 2012; Asner and Mascaro, 2014; Coomes et al., 2017; Labriere et al., 2018) and Figure 3-45 shows that our generalized allometry is well within the range of predictions from local studies. The panel in Figure 3-45, however, shows the diversity of allometries and points at detailing the relationship between LiDAR height metrics and AGB if used in accurate models to predict AGB with other remote sensing observations. For other zones, it is difficult to provide any conclusions about the quality of our predicted power-law function because studies relating height metrics to AGB are lacking. Still, Figure 3-45 indicates a rough agreement between our models and models proposed in the literature but also an increasing discrepancy towards the taller forests (> 30 m in temperate wet and > 20 m in boreal forest).

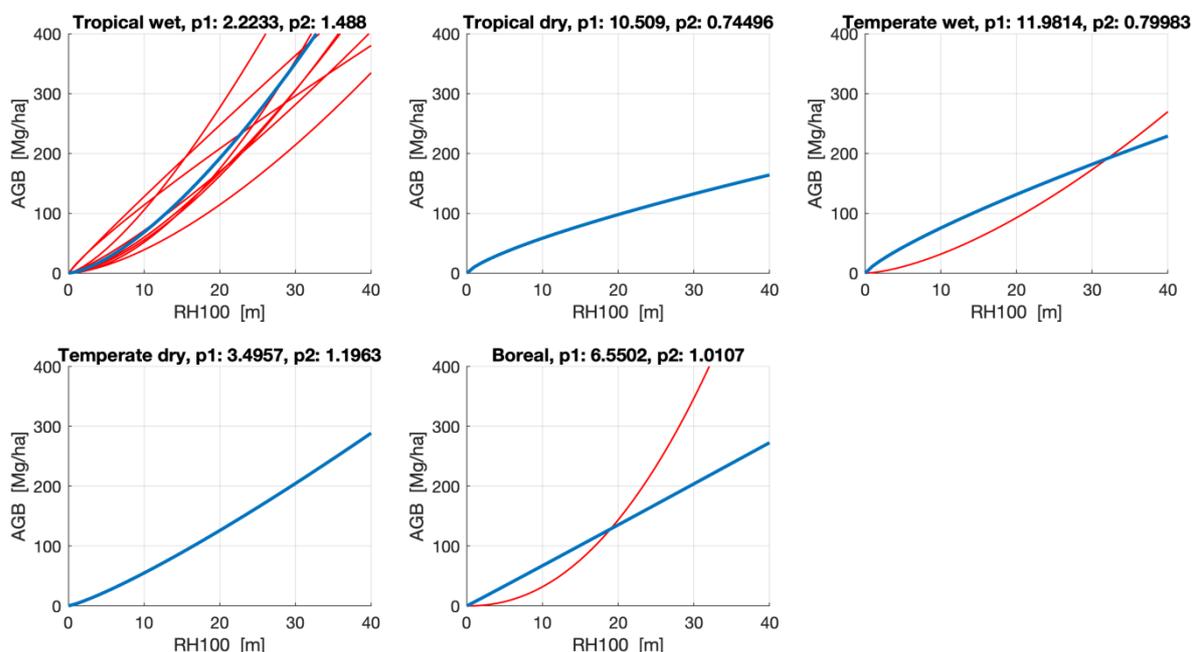


Figure 3-45: Comparing power-law (red lines) height-to-AGB allometries developed in literature (Santoro et al., 2002; Askne and Santoro, 2012; Asner et al., 2012; Asner and Mascaro, 2014; Coomes et al., 2017; Labriere et al., 2018; Dal Ponte et al., 2019) with own predictions from Equation (3-6) (blue line) on an ecoregion basis. The coefficients of the power-law functions are displayed on the top of each panel.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	69	14.02.2019	

Figure 3-46 shows global rasters with the estimates of p_1 and p_2 . Since we allocated a value of each coefficient to a major ecological zone, the spatial variability of p_1 and p_2 is restricted to transitions at the border of ecozones. We are aware of the fact that the GlobBiomass estimates may have introduced errors in these estimates and that we are not capturing spatial variability or more complex relationships between height and biomass. Still, this should be considered as a starting point towards a more precise modelling of backscatter as a function of AGB and a field of future investigation linking remote sensing and ecological communities.

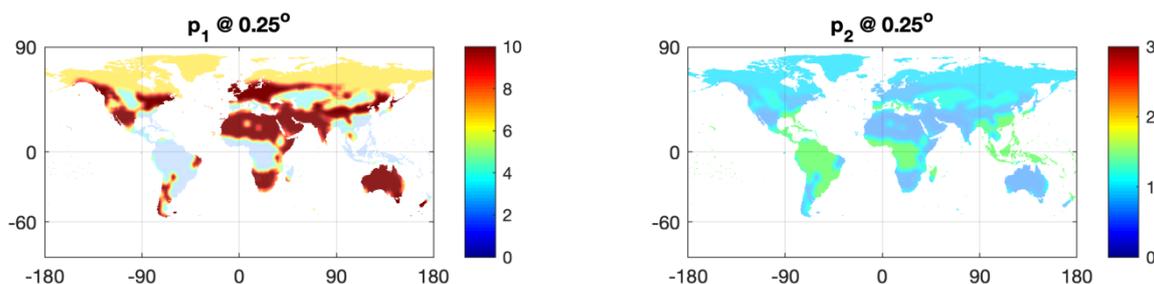


Figure 3-46: Spatially interpolated maps of the coefficients p_1 and p_2 . The maps are here represented in the form of quick-looks with a 0.25° pixel size.

4 Methods

4.1 The GlobBiomass global biomass retrieval algorithm

This Section starts with an overview of the GlobBiomass retrieval algorithm because it served as starting point for the development of the CORE algorithm used in CCI Biomass.

The GlobBiomass algorithm consisted of a three-stage approach that exploits a simple Water Cloud Model (WCM) to obtain two independent estimates of biomass from multi-temporal C-band SAR backscatter observations and from a single observation of L-band backscatter. The estimates were then combined with the intention to compensate for systematic errors in one or the other dataset. Because the WCM was expressed in a form relating SAR backscatter to the GSV ($\text{m}^3 \text{ha}^{-1}$), AGB was estimated from GSV by scaling using the BCEF.

Although there is no experimental evidence that estimating AGB from GSV is more accurate than estimating AGB directly, it was believed that this approach is more robust than a direct retrieval of AGB from the SAR backscatter for the following reasons:

- The SAR backscatter at C- and L-band is affected by the forest structure and dielectric. Here, we initially discard the contribution from the forest floor. For such wavelengths, the major component of the backscatter from a forest is expected to originate in the upper part of the canopy, thus explaining the limited sensitivity of the backscatter to forest structural parameters for increasing density. However, it was also reported that under dry or frozen conditions, the sensitivity of the SAR backscatter to forest structural variables increases (Santoro et al., 2011; Santoro et al., 2015a), which can be explained by increased penetration of the wavelength into the canopy so that major elements of the forest are being sensed. If a retrieval algorithm is designed to give more weight to such observations than to other observations (if available), it is reasonable to assume

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	70	14.02.2019	

that GSV can be retrieved from SAR backscatter (as implemented in the GlobBiomass algorithm).

- By relating to a structural parameter such as GSV, one can base the retrieval on a physically-based model, with parameters that can be predicted by using observations. If the variable of interest were AGB, it is unclear how such models could be parameterized given that the weight of the trees cannot be inferred from measurements of structural parameters only.
- AGB requires knowledge of the wood density. It remains undemonstrated that the SAR backscatter at C- and L-band is sensitive to the specific wood gravity of trees. Lacking such evidence, it is preferable to proceed with the estimation of forest structural parameters from the SAR backscatter and convert to AGB using a separate layer (the BCEF) that does not depend on remote sensing observations.
- Volume is the major predictor of biomass. The use of volume as the major predictor of AGB is clearly evidenced by the country reports to the FAO 2010 Forest Resources Assessment (FRA). More than 80% of the 171 countries reporting their biomass and carbon resources to the FRA based their numbers on estimates of volume (and not vice versa).

Given that there are hardly any datasets reporting measurements of GSV and AGB, it is currently not possible to go beyond these statements, but some of them clearly require being addressed in the future at the level of prototyping studies.

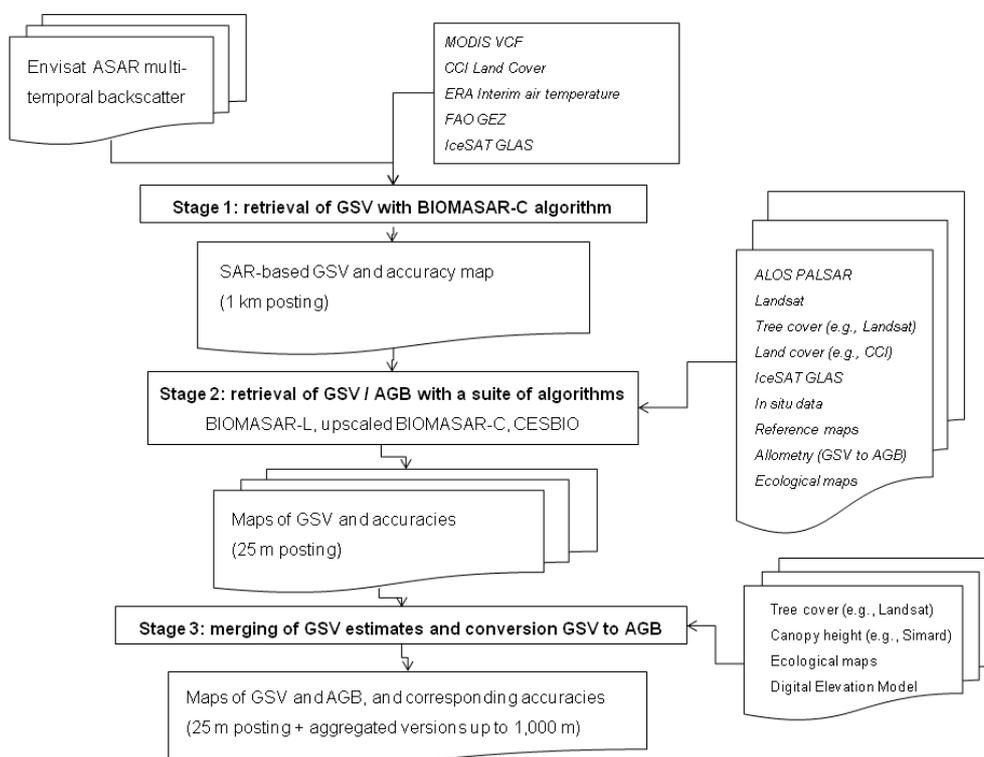


Figure 4-1: Flowchart of the GlobBiomass global biomass retrieval algorithm.

The three stages of the GlobBiomass algorithm were structured as follows (Figure 4-1).

- A global dataset of GSV was derived from hyper-temporal observations of C-band backscatter using the BIOMASAR-C algorithm (Santoro et al., 2011; Santoro et al.,

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	71	14.02.2019	

2015a). BIOMASAR-C was trained without *in situ* measurements and retrieved GSV at the spatial resolution of the input EO data. While BIOMASAR-C was found to capture the spatial distribution of GSV, even under unfavourable conditions such as in the wet tropics [RD-8], there were evident limitations of C-band SAR for estimating GSV in dense forests or in patchy landscapes with a mix of forest and other land cover types. In addition, cropland was often associated with non-zero biomass because of the seasonal increase of the backscatter to levels observed in young forests. A global GSV map obtained with BIOMASAR-C was 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 bulk of the GlobBiomass retrieval corresponded to the second stage, which included several retrieval approaches applied to high-resolution SAR data. Multiple approaches were considered to reduce potential flaws in each single approach due either to the input dataset or the simplifying assumptions used to model the relationship between SAR backscatter and GSV. Here, the L-band backscatter was used as a predictor in a model-based approach mimicking BIOMASAR-C (hence referred to as BIOMASAR-L) and in a re-scaling approach of the moderate resolution BIOMASAR-C estimates together with other high resolution datasets (e.g., Landsat reflectances). The re-scaling approach was referred to as BIOMASAR-C+ and was developed to complement the retrieval with BIOMASAR-L in areas of poor performance of the retrieval based on a single L-band observation (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). Similarly to BIOMASAR-C, both approaches were designed to be calibrated without *in situ* measurements and retrieve GSV at the spatial resolution of the input EO data.
- For each pixel, the final estimate of biomass was obtained by linearly combining the BIOMASAR-L and the BIOMASAR-C+ estimates with weights defined by their similarity to theoretical behaviour when related to forest canopy height, canopy density and forest transmissivity observations. The integration of the biomass estimates was parameterized at the ecozone level. The integration was implemented to produce a combined biomass expressed as GSV.
- An additional step was then required to estimate AGB. AGB was estimated from the GSV dataset using spatially explicit estimates of wood density and stem-to-total biomass expansion factors, which were derived from an extrapolation of *in situ* observations and modelling.
- At this stage, additional approaches were considered to complement structural deficiencies of the BIOMASAR-type of estimations and/or the conversion of GSV to AGB. In GlobBiomass, several data products were evaluated but none could compensate for the deficiencies of the GlobBiomass data product.
- Each map of GSV and AGB has a corresponding map of accuracy.
- To account for different user needs, the high-resolution biomass (GSV and AGB) estimates can be aggregated (by spatial averaging) to form new estimates at moderate and coarse scale. Accordingly, estimates of the estimation accuracy for the averaged biomass are obtained.

4.2 The CCI Biomass CORE algorithm

Before cloning the three-stage approach developed in the GlobBiomass project for the satellite data of 2010 to represent the CORE algorithm of CCI Biomass, it was necessary to understand if the same conditions apply for the satellite data of 2017-2018.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	72	14.02.2019	

Unlike in the GlobBiomass project, the spatial resolution of the hyper-temporal dataset of Sentinel-1 C-band observations is substantially higher (150 m vs. 1000 m) and comparable to the spatial resolution of the AGB dataset obtained in GlobBiomass (150 m vs. 100 m). In addition, the availability of the cross-polarized backscatter and the possibility to compensate for topographic effects on the backscatter allows us to assume that the C-band estimates of biomass are sufficiently reliable to be used without refinements or rescaling. In other words, the rescaling applied at stage 2 in Figure 4-1 with the BIOMASAR-C+ algorithm becomes redundant. Another reason for avoiding the rescaling is the poorer quality of the EO datasets used in the rescaling process. Both the ALOS-2 and the Landsat surface reflectances for the 2017-2018 epoch exhibit artefacts (see Section 3.2 for ALOS-2). Rescaling would have meant introducing such artefacts into the final output.

We also take into account that the ALOS-2 datasets available to this project are of poorer quality than the ALOS-1 datasets used in GlobBiomass but we have an additional layer (ScanSAR acquisitions) at 50 m spatial resolution that contributes substantially to the retrieval in the tropics. As a result, it is preferred to estimate biomass from the L-band data at 100 m pixel size to reduce the effect of artefacts in the L-band data but still preserve details to a level comparable to the details reproduced in the GlobBiomass dataset.

Merging according to predefined rules that prefer one or other dataset based on the plausibility of the estimates is maintained. For this, the C-band estimates of biomass are simply resampled to the geometry of the L-band estimates.

While in year 1 of the project, the objective was to establish the CORE algorithm by adapting the GlobBiomass algorithm to the 2017 dataset of spaceborne observations, in year 2, we accounted for recent advances towards the generation of global biomass estimates with methods complementing those pursued in year 1 (PVASR and ADP). One of the major deficits of the models relating SAR backscatter to biomass in the previous version of the CORE algorithm was the non-explicit contribution of vegetation height to the retrieval. In year 2, we explicitly included height in the retrieval models and considered allometries between vegetation height and biomass. By doing so, we bypass the estimation of GSV, which was understood to be a source of disagreement when interpreting the past mapping results in the tropics. Also, a more direct retrieval of AGB implies that BCEFs are not used in the retrieval algorithms; this is a consequence of the reduced activity on the characterization between vegetation volume and organic mass in recent years.

The biomass estimation procedure described in Figure 4-1 was further detailed in the ATBD of the GlobBiomass project [RD-8] by showing the relationship between EO data, methods and outputs. In this document, we follow the same procedure and present in Figure 4-2 a more functional flowchart focusing on the interdependencies of datasets and algorithms. The CCI Biomass CORE global biomass retrieval algorithm used in year 1 is shown next to the CORE implemented in year 2. Text in red visualizes the updates introduced between the two versions. Below we summarize the main aspects of the two versions of the CORE retrieval algorithm, with emphasis on the current version used in year 2 of the CCI Biomass project to generate three maps of forest AGB representative of the 2017-2018, the 2018-2019 and the 2010 epochs. For this version of the ATBD, both implementations of the CORE are described because their evaluation globally is not concluded at the time of compiling this document.

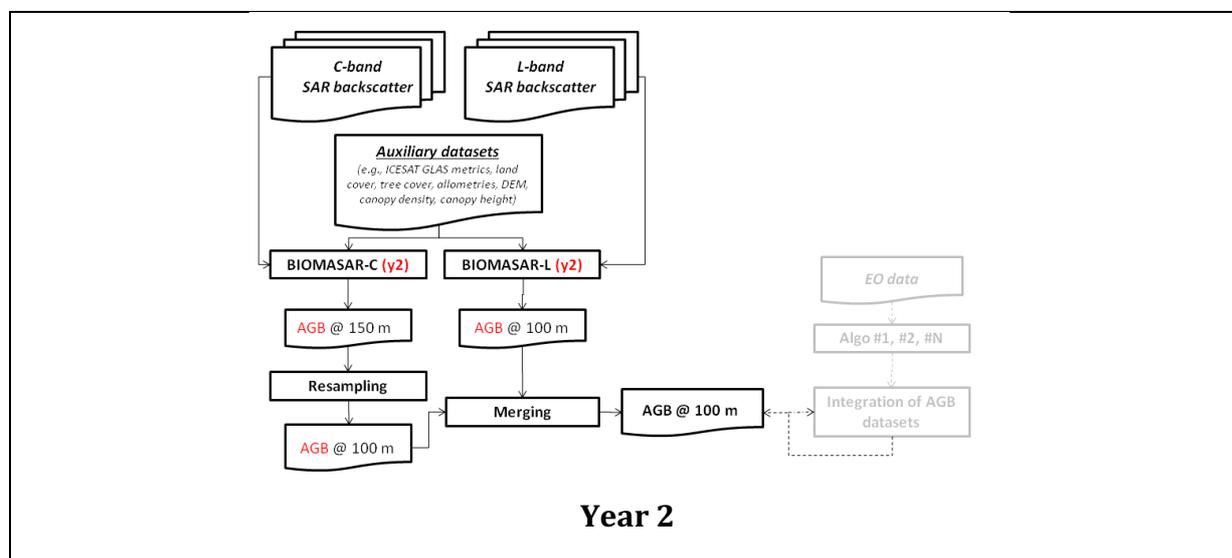
- Multi-temporal C-band SAR backscatter data are used to generate a global estimate of biomass with the BIOMASAR-C algorithm. The estimates are produced at 150 m pixel

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	73	14.02.2019	

size and then resampled to 100 m. The y2 version of BIOMASAR-C foresees the direct estimation of AGB instead of GSV as in year 1.

- Multi-temporal mosaics of L-band SAR backscatter are used to generate a global estimate of biomass with the BIOMASAR-L algorithm. The estimates are produced at 100 m pixel size. The y2 version of BIOMASAR-L foresees the direct estimation of AGB instead of GSV as in year 1.
- A set of auxiliary datasets is used to calibrate the models relating observables and biomass since the models are implemented so that they do not need to rely on *in situ* measurements for training.
- The biomass estimates of BIOMASAR-L and BIOMASAR-C are merged to allow reduction of systematic errors in one or the other dataset.
- In year 1, a conversion from GSV to AGB was implemented at the end of the retrieval as a separate step. Characterization of the AGB errors can therefore rely on separate estimates related to the retrieval algorithm and the conversion factors. Because of the direct retrieval of AGB implemented in year 2, this part does not apply in the current version of the CORE retrieval algorithm.
- Accuracy is characterized at each step shown in Figure 4-2. An estimate of the GSV and AGB accuracy is attached to each pixel at 100 m.
- The CORE algorithm can be expanded by linking with additional datasets produced with algorithms that are proven to perform better than those proposed here (see PVASR)

Spatial averaging would be applied to reduce pixel-wise retrieval errors and increase the accuracy. This final step is done “on demand” where a user can specify the target spatial resolution.



	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	74	14.02.2019	

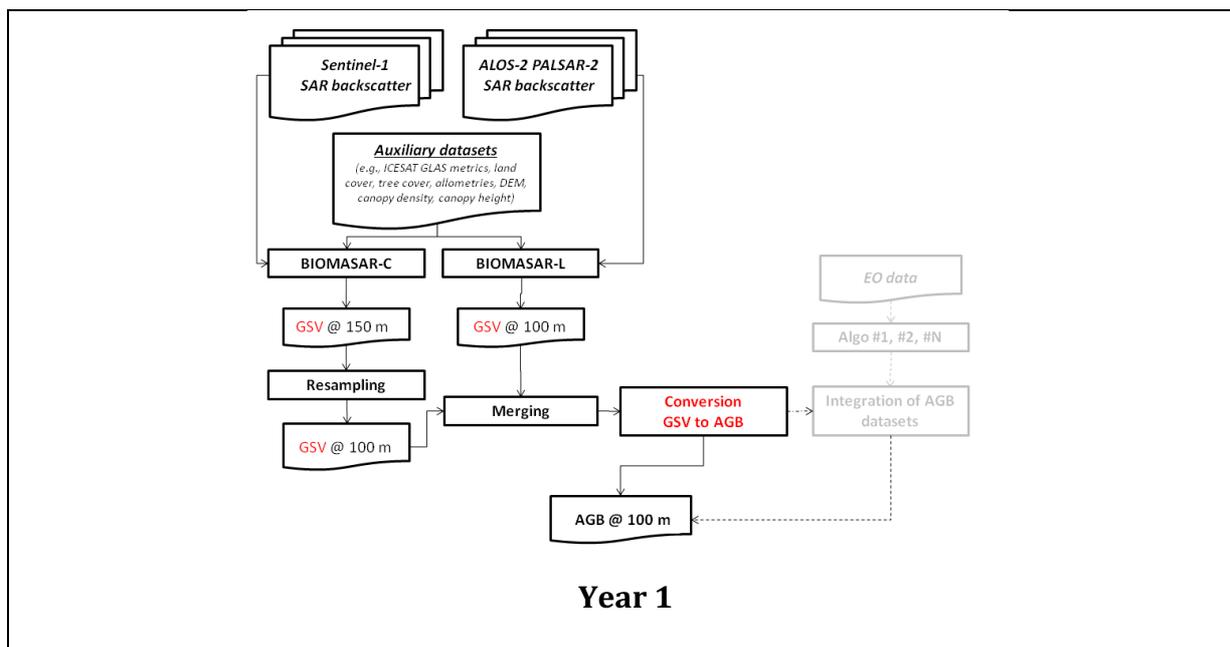


Figure 4-2: Functional dependencies of datasets and approaches forming the CCI Biomass CORE global biomass retrieval algorithm in year 2 (top) and year 1 (bottom). Text in red visualizes modifications introduced from year 1 to year 2. The shaded part of the flowchart represents potential improvements following the implementation of additional retrieval techniques.

In the following Sections, we describe the individual components of the CORE algorithm:

- BIOMASAR-C
- BIOMASAR-L
- Merging

Because the advances in BIOMASAR-C and BIOMASAR-L from year 1 to year 2 modified parts of the model training procedure, we split the description of each into separate Sections, allowing for some redundancy in the text to provide as complete as possible reporting.

The GSV to AGB conversion is also described for completeness.

4.3 The BIOMASAR-C algorithm

The theoretical basis of the BIOMASAR-C algorithm has been presented in Santoro et al. (2011) and (2015a) and was extensively presented in the ATBD of the GlobBiomass project [RD-8]. Here, the major components are summarized and a reference is given to previous documentation for technical aspects that do not need to be repeated in this context.

At C-band, spatial and temporal variability of the backscatter make empirical modelling of biomass (regardless of GSV or AGB) derived using *in situ* measurements almost useless if the aim is to produce large-scale estimates based on a small set of reference measurements, which is very often the reality. Hence robust retrieval of biomass 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. We opted for the semi-empirical Water Cloud Model (Attema & Ulaby, 1978) with gaps based on the formulation reported by Askne et al. (1997). The reason for this choice is the demonstrated reliability of such a type of model in the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	75	14.02.2019	

retrieval of forest biomass (Santoro and Cartus, 2018) and the extensive knowledge gathered by the team developing the retrieval algorithm in this project with such a modelling framework.

The WCM with gaps, given as Equation (4-1), was derived from the original WCM presented by Attema & Ulaby, 1978) to express the total forest backscatter of a forest as the sum of direct scattering from the ground through gaps in the canopy, ground scattering attenuated by the canopy and direct scattering from vegetation:

$$\sigma_{for}^0 = (1 - \eta)\sigma_{gr}^0 + \eta\sigma_{gr}^0 T_{tree} + \eta\sigma_{veg}^0(1 - T_{tree}) \quad (4-1)$$

Here, η is the area-fill or canopy density factor, representing the fraction of the area covered by vegetation, σ_{gr}^0 and σ_{veg}^0 , 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 $e^{-\alpha h}$, 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 (4-1) is not useful since the area-fill factor is not a parameter of interest to foresters and it relates two forest variables (height and canopy density, i.e., area-fill factor) to a single observation. For retrieval purposes, it is more convenient to describe the backscatter as a function of biomass. Leaving aside all possible formulations of the WCM expressing the SAR backscatter as a function of AGB with empirical coefficients (Santoro & Cartus, 2018), the semi-empirical equation proposed by Pulliainen et al. (1994) rewrote the original WCM to a similar relationship as in Equation (4-1), relating the SAR backscatter to stem volume, V , which can be considered a synonym of GSV:

$$\sigma_{for}^0 = \sigma_{gr}^0 e^{-\beta V} + \sigma_{veg}^0(1 - e^{-\beta V}) \quad (4-2)$$

In Equation (4-2), β is an empirically defined coefficient expressed in ha m^{-3} . However, this coefficient has physical meaning taking into account that by comparing Equations (4-1) and (4-2) the link between β , η and α is given by (Santoro et al., 2002):

$$e^{-\beta V} = 1 - \eta(1 - e^{-\alpha h}) \quad (4-3)$$

The major assumption when rewriting the exponent of the original WCM as a simple scaling of GSV was that height and volume are linearly correlated. This may apply to mature forests but is not correct throughout the range of heights and volumes in boreal forests, where Equation (4-3) was developed. Similarly, the non-linear relationship between height and AGB suggests that the inversion of Equation (4-3) to estimate biomass from the SAR backscatter may introduce systematic biases in the retrieval regardless of the forest structure.

Equation (4-3) was used in year 1 of CCI Biomass to estimate GSV. The implementation of this version of the WCM in BIOMASAR-C is described in Section 4.3.1. In year 2, we consider the original WCM with gaps in Equation (4-1) and a set of functional dependencies between canopy density, tree height and AGB to express the WCM as a function of AGB. The implementation of the WCM and the allometric functions to retrieve AGB are described in Section 4.3.2.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	76	14.02.2019	

4.3.1 BIOMASAR-C for GSV retrieval

Equation (4-2) contains three unknowns that need to be estimated: β , σ_{veg}^0 and σ_{gr}^0 . If a set of reference GSV values is available, the model parameters can be estimated with a least squares regression to the reference GSV and the measurements of the SAR backscatter corresponding to the GSV observations. This approach is, however, unfeasible for large areas because it requires a dense network of training sites to correctly capture the spatial variability of the SAR backscatter of the forest and therefore of its ground and vegetation component as well as of the attenuation in the canopy. For this reason, in BIOMASAR-C a method was developed that can provide estimates of σ_{gr}^0 and σ_{veg}^0 by means of statistics of the backscatter for certain types of forest cover whereas the parameter β is estimated from metrics of the ICESAT GLAS waveforms.

To estimate σ_{gr}^0 and σ_{veg}^0 , a dataset of the canopy density is used in order 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}^0 is obtained after correcting the values of the backscatter for dense forest pixels, σ_{df}^0 , 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 4-3. The estimation of the two parameters σ_{gr}^0 and σ_{df}^0 as well as the procedure to derive estimates of the forest transmissivity is described below. Once the model is trained, it is inverted to estimate GSV from corresponding observations of the SAR backscatter. If multiple observations of the backscatter are available, GSV can be estimated from each observation. Noise can then be reduced with a linear combination, to which we refer as multi-temporal combination. The individual steps are described below.

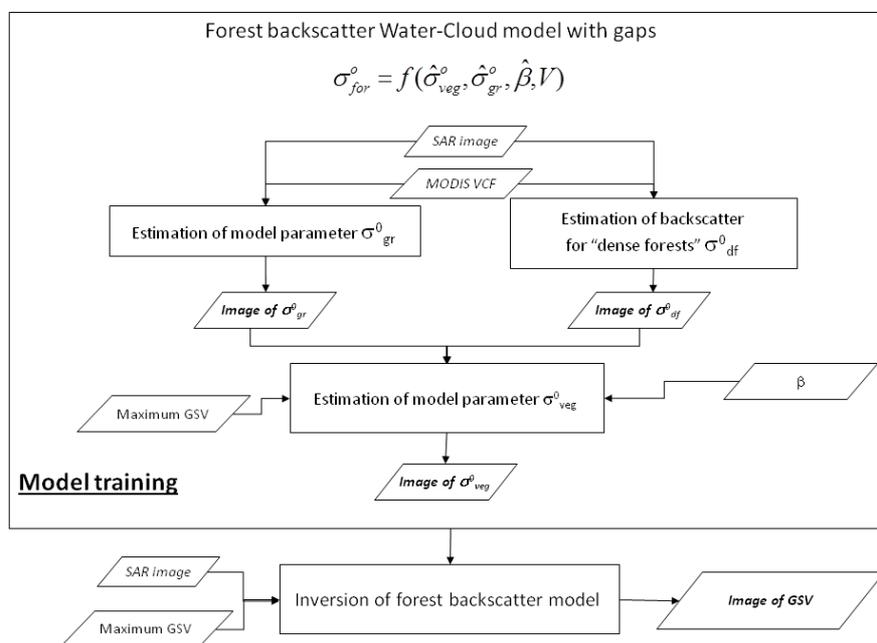


Figure 4-3: Flowchart of model training and retrieval implemented in the BIOMASAR-C approach

4.3.1.1 Estimation of σ_{gr}^0

σ_{gr}^0 is estimated as the median backscatter of pixels labelled as ground within an estimation window centred on the pixel of interest (Santoro et al., 2011). A pixel is labelled as “ground” when the canopy density of that pixel is below a certain threshold. Canopy density information

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	77	14.02.2019	

is currently available from the yearly MODIS VCF dataset at moderate resolution (Di Miceli et al., 2015) or from periodic canopy density datasets at high resolution based on Landsat observations (Hansen et al., 2013).

Originally, BIOMASAR-C was written to retrieve GSV from coarse resolution 1 km C-band Envisat ASAR data. The low resolution implied that the estimation of σ_{gr}^0 could be undertaken on a pixel-by-pixel basis with an estimation window that would slide along and across the ASAR image. The window size grew from a minimum size until a certain percentage of pixels that could be labelled as “ground” was identified. The requirement on a minimum number of pixels labelled as “ground” was often met by relaxing the threshold according to which a pixel would be labelled as “ground” (e.g., increase of the maximum canopy density from 15% to 30%).

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 thematic data product giving information on the land cover types that should be excluded.

This approach is not viable with SAR data processed to 150 m pixel spacing (and in general for any high resolution SAR dataset) for two reasons:

- The estimation window can become extremely large in areas where “ground” or “dense forest” pixels cannot be identified (e.g., “ground” pixels in the rainforest of the Amazon). This can cause severe out-of-memory errors or large gaps where σ_{gr}^0 is not estimated.
- The computation of per-pixel estimates of the backscatter model parameters at 150 m would increase the computational load by a factor 50 compared to the 1000 m case.
- To circumvent such issues, the estimation window is here set equal to the size of a tile (i.e., $1^\circ \times 1^\circ$). “Ground” pixels are selected in this window based on a fixed threshold on maximum canopy density characterizing ground pixels. Given a Sentinel-1 image, the average of the histogram of the selected ground pixels is computed unless the fraction of ground pixels with respect to all pixels in the estimation window is less than a certain threshold (see below). The average backscatter represents the estimate of σ_{gr}^0 for the specific SAR image in the tile and is associated with the centre of the tile. Figure 4-4 illustrates the procedure to estimate σ_{gr}^0 .

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	78	14.02.2019	

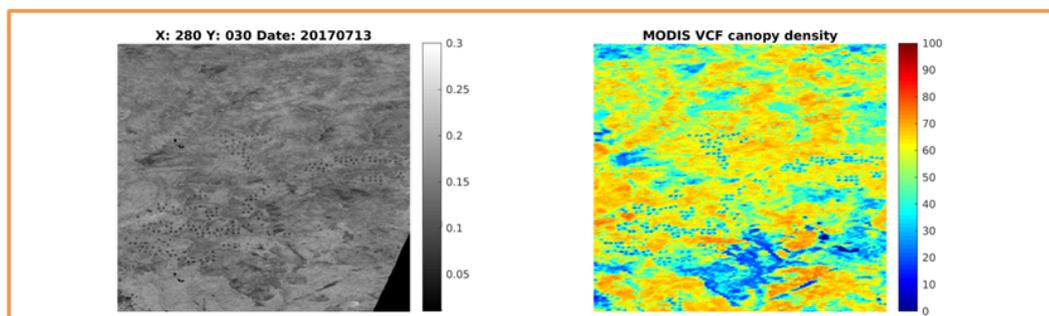


Figure 4-4: Flowchart illustrating the procedure to estimate σ_{gr}^0 and σ_{df}^0 in BIOMASAR-C for the case of a Sentinel-1 backscatter image.

The technical specifications regarding the estimation of σ_{gr}^0 are as follows:

- Average the backscatter image from 150 m to 300m (multi-look: 2 x 2)
- Canopy density: MODIS VCF of the year closest to the epoch of acquisition of the SAR data, resampled from 250 m to 150 m (bilinear interpolation)
- Thematic layer to identify non-vegetated pixels: CCI Land Cover of the year closest to the epoch of acquisition of the SAR data, resampled from 300 m to 150 m (nearest neighbour)
- Minimum fraction of ground pixels: 1%
- Maximum tree canopy cover to label pixels as ground: 30%

The selection of the canopy density dataset and the thematic layer to exclude non-vegetated pixels from estimation of σ_{gr}^0 prioritised the use of data layers as contemporaneous as possible with the SAR dataset. The minimum fraction of ground pixels was hard-coded after investigating the impact of this threshold on the number of selected ground pixels, the corresponding median of the histogram and the value predicted by a least-square regression of Equation (4-2) to a set of training samples in different regions (Santoro et al., 2011).

Figure 4-5 shows the set of estimates of σ_{gr}^0 for a Sentinel-1 image. Each pixel in Figure 4-5 represents a specific tile. The estimation procedure could not identify a valid value for σ_{gr}^0 in every tile. Voids occurred for tiles where the requirement of a minimum fraction of ground pixels was not satisfied. To overcome the issue of having a gap in the fields of σ_{gr}^0 estimates, the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	79	14.02.2019	

not-a-number (NaN) values are replaced by extrapolating from valid estimates for the specific Sentinel-1 image (<https://ch.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans>). This allows an estimate of σ_{gr}^0 to be obtained in regions with unbroken dense forest cover (e.g., Amazon, Congo Basin) or at the edge of an image when only a minor fraction of the tile's area is covered by the SAR image.

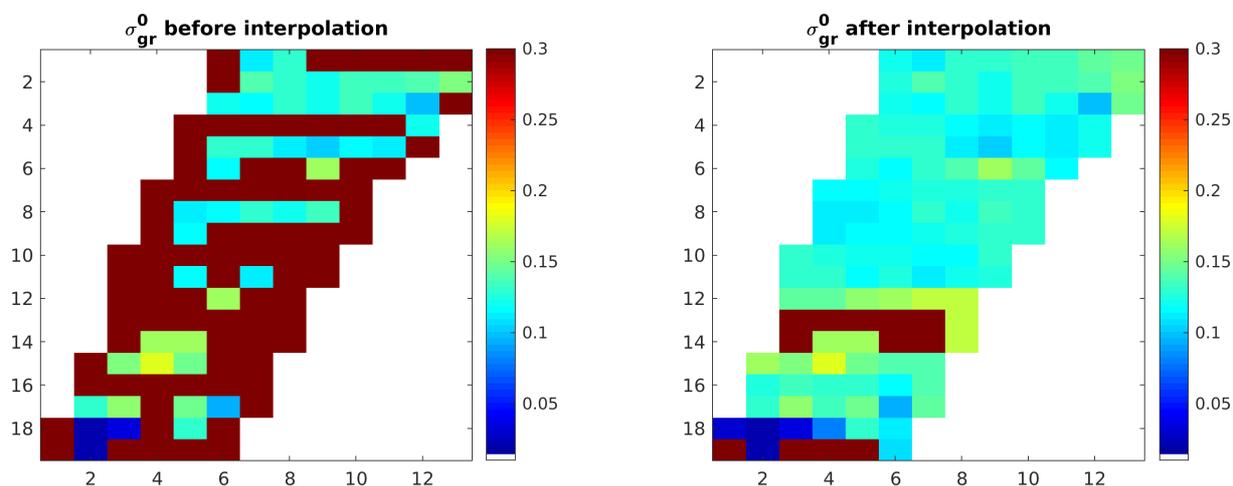


Figure 4-5: Illustrating the gap-filling procedure in the case of incomplete fields of σ_{gr}^0 estimates for a Sentinel-1 image. Each pixel represents the estimate for a given tile covered by the Sentinel-1 image. Red is used for pixels for which no estimate has been found.

While recovering values everywhere, the procedure for estimating σ_{gr}^0 from high-resolution SAR data misses local patterns of the backscatter from the forest floor, e.g., because of spatially variable soil moisture. 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 forests, except for young regrowing trees.

Because of the strong requirement on the minimum number of SAR observations to retain estimate of σ_{gr}^0 , we neglect the dependency of the backscatter on local incidence angle. Although the SAR backscatter is corrected for true pixel size and local incidence angle, the correction does not take into account scatterer-specific dependencies of the backscatter on the local orientation and position of the scatterer. While the C-band backscatter from a dense forest tends to be rather constant for varying incidence angle, the backscatter of a bare surface differ by several dB when going from near range (steep look angle) to far range (shallow look angle) (Santoro 2015c). The impact of neglecting such variability on modelling and retrieval has not been quantified.

4.3.1.2 Estimation of σ_{df}^0

The procedure for estimation σ_{df}^0 , which is defined as the average SAR backscatter for dense forests, is illustrated in Figure 4-4. As with the estimation of σ_{gr}^0 , an estimation window is defined and pixels within it with a canopy density above a pre-defined threshold are selected to form the “dense forest” class. The median of the backscatter values for the pixels forming this class is associated with σ_{df}^0 . Similar to σ_{gr}^0 , the threshold on canopy density used to identify dense forest pixels is predefined, being set equal to 75 % of the maximum value. All pixels satisfying this requirement within the $1^\circ \times 1^\circ$ tile are then selected to be included in the dense forest class.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	80	14.02.2019	

The technical specifications regarding the estimation of σ_{df}^0 are as follows:

- Average backscatter image from 150 m to 300m (multi-look: 2 x 2)
- Canopy density: MODIS VCF of the year closest to the epoch of acquisition of the SAR data, resampled from 250 m to 150 m (bilinear interpolation)
- Thematic layer to identify non-vegetated pixels: CCI Land Cover of the year closest to the epoch of acquisition of the SAR data, resampled from 300 m to 150 m (nearest neighbour)
- Minimum fraction of dense forest: 0.1%
- Minimum tree canopy cover to label pixels as dense forest: 0.75 of the maximum canopy density in the tile of interest.

As in the case of σ_{gr}^0 , the canopy density dataset and the thematic layer to exclude non-vegetated pixels from the estimation were chosen primarily to prioritise data layers as contemporaneous as possible to the SAR dataset. The minimum fraction of dense forest pixels was hard-coded after investigating the impact of this threshold on the number of selected dense forest pixels, the corresponding average of the histogram and the value predicted by a least-square regression of Equation (4-2) to a set of training samples in different regions (Santoro et al., 2011).

A set of σ_{df}^0 for a Sentinel-1 image, where each pixel represents a given tile, is given in Figure 4-6. Gaps occurred for tiles where the requirement on a minimum fraction of dense forest pixels was not satisfied. To overcome the issue of having a gap in the fields of the σ_{df}^0 estimates, the NaN values are replaced by extrapolating from the valid estimates for the specific Sentinel-1 image (<https://ch.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans>). This allows an estimate of σ_{df}^0 to be obtained in regions with sparse forest cover (e.g., tundra-taiga border) or at the edge of an image when only a minor fraction of the tile's area is covered by the SAR image.

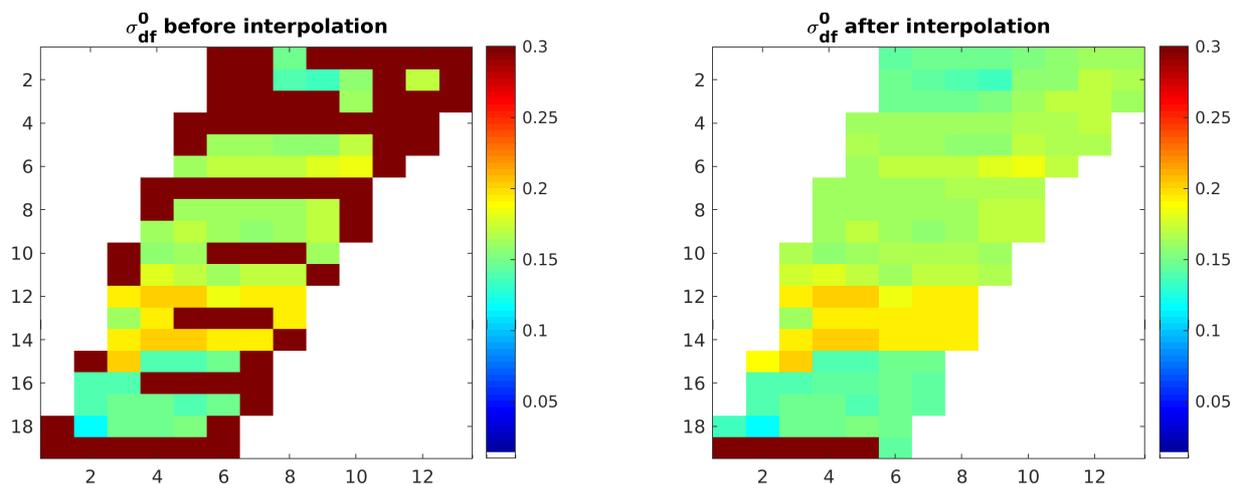


Figure 4-6: Illustrating the gap-filling procedure in the case of incomplete fields of σ_{df}^0 estimates for a Sentinel-1 image. Each pixel represents the estimate for a given tile covered by the Sentinel-1 image. Red is used for pixels for which no estimate has been found.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	81	14.02.2019	

4.3.1.3 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 should be adapted to local environmental and forest conditions. Using Equation (4-3), 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 (4-3) 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 applied to all forests globally and stratification in terms of ecological zone was applied in order to provide a realistic representation of the spatial distribution of β . Simulations of area-fill from Equation (4-3) were generated using β values between 0.004 and 0.014 ha m⁻³ and α values between 0.2 and 2 dB m⁻¹. 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 in the biomass estimates or the conversion factors. Nonetheless, it was found to be the only practical 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 and area-fill, i.e., canopy density. For additional details on the estimation procedure refer to the ATBD of the GlobBiomass project [RD-8].

The estimates of β were obtained at the level of individual Global Ecological Zones [RD-8]. β varied between 0.012 ha m⁻³ in tropical rainforest and 0.004 ha m⁻³ in arid environments. In general, higher β values were estimated in forest with a more closed canopy, which agrees with theory.

The estimation of β was 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. Nonetheless, the evidence that the relationship in Equation (4-3) assumes a linear relationship between height and GSV undermines the reliability of such functional dependence.

The VCF product was derived for leaf-on conditions. The canopy cover in leaf-off conditions, however, can differ substantially from leaf-on 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.

4.3.1.4 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 and σ_{df}^0 are known at

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	82	14.02.2019	

this stage and all other parameters in Equation (4-2) are known, Equation (4-2) can be inverted to obtain an estimate of σ_{veg}^0 :

$$\sigma_{veg}^0 = \frac{\sigma_{df}^0 - \sigma_{gr}^0 e^{-\beta V_{df}}}{1 - e^{-\beta V_{df}}} \quad (4-4)$$

In Equation (4-4), V_{df} is the GSV value representative of dense forest (Section 3.14).

4.3.1.5 Inverting the forest backscatter model

Once the model parameters have been estimated, the model in Equation (4-2) can be inverted to derive an estimate of GSV from a measurement of the backscatter, σ_{meas}^0 :

$$\hat{V} = -\frac{1}{\beta} \ln \left(\frac{\sigma_{meas}^0 - \sigma_{veg}^0}{\sigma_{gr}^0 - \sigma_{veg}^0} \right) \quad (4-5)$$

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⁻¹ or the maximum retrievable GSV when the measurement falls within a buffer zone either below or above the modelled backscatter. The buffer zone corresponds to 3 times the uncertainty of the backscatter measurements (i.e., roughly 1 dB assuming an ENL of 163; see Section 3.1). No GSV is retrieved for backscatter outside this range. The maximum retrievable GSV is defined as 1.2 times V_{df} (Section 3.14). The same procedure applies if the backscatter decreases for increasing GSV with the only difference that this time the maximum GSV is associated with pixels where the measurements are below the minimum modelled backscatter, whereas 0 is associated with pixels for which the backscatter is above the maximum modelled backscatter.

4.3.1.6 Multi-temporal combination of GSV estimates

Given N individual estimates of GSV from Equation (4-2), a weighted linear combination of the estimates is used to obtain the final estimate of GSV, V_{mt} (Equation (4-6)). 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).

$$V_{mt} = \frac{\sum_{i=1}^N w_i \hat{V}_i}{\sum_{i=1}^N w_i} \quad (4-6)$$

The weights, w_i , in Equation (4-6) are defined as the vegetation-to-ground backscatter difference in dB, $\sigma_{veg}^0 - \sigma_{gr}^0$, normalized by the maximum backscatter difference:

$$w_i = \frac{\sigma_{veg,i}^0 - \sigma_{gr,i}^0}{\max(\sigma_{veg,i}^0 - \sigma_{gr,i}^0)} \quad (4-7)$$

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. However, this rule is not applied here because it would discard a large number of scenes in the wet tropics, resulting in an undetermined estimate of GSV.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	83	14.02.2019	

4.3.2 BIOMASAR-C for AGB retrieval

While the outline of BIOMASAR-C for AGB retrieval follows the implementation of BIOMASAR-C to retrieve GSV, there are three main differences in terms of

- The Water Cloud Model (Equation (4-1) instead of Equation (4-2))
- The model training procedure (regression combined with self-calibration instead of self-calibration only)
- Adaptation of the model training to local incidence angle (not applied in BIOMASAR-C for GSV)

4.3.2.1 Estimation of σ_{gr}^0 and σ_{veg}^0

To estimate AGB from a measurement of the backscatter, Equations (3-5) and (3-6) are inserted into Equation (4-1). Similarly to the GSV retrieval, the retrieval model used to relate AGB to the SAR backscatter has three unknowns, namely σ_{gr}^0 , σ_{veg}^0 and α . For the latter, we are currently assuming 2 dB/m following our own synthesis of studies dealing with attenuation of C-band microwaves in tree canopies; an attempt to distinguish between unfrozen conditions and leaf-on conditions from frozen or leaf-off conditions has not been attempted yet.

For the estimation of σ_{gr}^0 and σ_{veg}^0 , we applied some of the self-calibration concepts developed in the framework of the BIOMASAR-C approach to retrieve GSV but tried to increase robustness because we identified several locations where the modelled backscatter with such estimates of the parameters was often above the measured backscatter, even if slightly, causing severe underestimation of biomass (Section 5).

Self-calibration is used to obtain an initial estimate of σ_{gr}^0 and σ_{veg}^0 . The two model parameters are associated with the median value of the backscatter for pixels belonging to a class of low vegetation cover density ("ground" pixels) and high vegetation cover density ("dense forest" pixels), respectively, within an estimation window of finite size. As self-calibration is only used to generate an initial set of values and we want to make sure that initial estimates are obtained to then drive the actual estimation of the two parameters, we apply a single canopy density threshold (i.e., pixels are labelled either as "ground" or as "dense forest" depending whether the canopy density is below or above 30%). The estimation window corresponds to the size of a tile (i.e., $1^\circ \times 1^\circ$). It is assumed that enough samples are captured in each class with such a window size; the drawback can be the inclusion of pixels being characterized by different properties of the backscatter as a consequence of different environmental conditions within the area covered by the tile.

In contrast to BIOMASAR-C for GSV retrieval, the final estimates of σ_{gr}^0 and σ_{veg}^0 are here obtained by means of least squares regression between SAR backscatter observations and corresponding canopy density values with a simplified version of Equation (4-1) within a window of finite size. Assuming strong tree attenuation, Equation (4-1) can be simplified to Equation (4-8), which expresses the forest backscatter as a weighted average of the backscatter from the ground and the canopy through its canopy density. This assumption can be considered valid because we are interested in estimating the backscatter of a completely bare surface (σ_{gr}^0), and a completely opaque vegetation layer (σ_{veg}^0), but not the trend between 0 and 100% canopy density. It should be noted that this approach does not require knowledge of forest transmissivity (i.e., the coefficient β in Equation (4-2)) nor information on a certain level of biomass (e.g., biomass of dense forests, maximum biomass etc.)

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	84	14.02.2019	

$$\sigma_{for}^0 = (1 - \eta)\sigma_{gr}^0 + \eta\sigma_{veg}^0 \quad (4-8)$$

Here, we used the $1^\circ \times 1^\circ$ area of a tile as the estimation window and associated at this stage a single value of σ_{gr}^0 and a single value of σ_{veg}^0 to the area covered by the SAR image within the $1^\circ \times 1^\circ$ tile. This assumption contributed to increased computation speeds introduced a certain degree of generalization, cancelling small-scale variations of the ground and canopy backscatter due to different forest structures or environmental conditions within the tile.

The tree cover density dataset used in the self-calibration and the regression steps was the percent tree cover dataset of 2010 because it is more reliable in terms of spatial consistency than other high-to-moderate resolution datasets currently available (see Annex B). In addition, to avoid distorting estimates of σ_{gr}^0 , areas with very low canopy density not corresponding to natural vegetation had to be masked *a priori* because of their different features in terms of C-band backscatter. For this, the CCI Land Cover dataset of the year closest in terms of availability to the year of the C-band data was used and pixels labelled as cropland, urban areas, bare soil, permanent snow and ice, and water bodies were masked out.

To account for the different levels of SAR backscatter depending on local incidence angle, the procedure was applied separately to each set of backscatter measurements and percent tree cover values were characterized by a specific range of incidence angles. As a trade-off between precision, speed of computation and representativeness of the estimates, we divided the observations into five 10° wide intervals of local incidence angle, starting with 20° and ending with 70° . Because of the rather large spread of backscatter observations for any given level of canopy density (see vertical bars in Figures A5 and A11), we opted for regressing the median values of the SAR backscatter rather than the original SAR backscatter observations to canopy density (see line connecting data points in Figures A5 and A11). The median backscatter was computed for each integer value of canopy density. Prior to this, the backscatter dataset had been filtered for outliers, appearing for example as a consequence of forest cover changes between 2010 and the year of acquisition of the SAR image.

The estimates of σ_{gr}^0 and σ_{veg}^0 obtained for a given image in a tile and for a given range of incidence angles were then retained if the following two requirements were met by the observational dataset: (i) at least 3 valid median values of the SAR backscatter and (ii) the correlation coefficient between the median backscatter values and the corresponding tree cover values was greater than 0.3. This was a safety measure to avoid cases when only a small range of canopy densities was represented (e.g., very high tree cover or very low tree cover) thus causing potentially erroneous estimates of one or the other parameter. When the requirements were not met, no estimate was assigned to σ_{gr}^0 and σ_{veg}^0 for the specific image and range of local incidence angle.

In our interpretation of the results in Figures A11-A14, we understood that the relationship between the percent tree cover dataset by (Hansen et al., 2013) and SAR backscatter is somewhat distorted above 80% canopy density. A consequence is that σ_{veg}^0 obtained from the regression between SAR backscatter and percent tree cover is overestimated because the WCM in Equation (4-8) assumes a steady increase of the backscatter even for canopy density close to 100%, which is not found in the observations. For this reason, we implemented a different approach to estimate σ_{veg}^0 , in case the tile contains pixels with a canopy density higher than 80%, and relied on the self-calibration step based on Equation (4-10) by BIOMASAR-L.

This Equation was obtained after inverting the WCM to express σ_{veg}^0 as a function of the other parameters where σ_{df}^0 represents the median of the backscatter observations for canopy density

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	85	14.02.2019	

greater than 80%, σ_{gr}^0 is equal to the value estimated from the least squares regression of Equation (4-8) and η_{df} represents the canopy density of dense forests, which has been estimated from the ICESAT GLAS observations for each tile.

$$\sigma_{veg}^0 = \frac{\sigma_{df}^0 - \sigma_{gr}^0 e^{-\eta_{df}}}{\eta_{df}} \quad (4-9)$$

This step ensured that σ_{veg}^0 was estimated at the level of backscatter for the pixels with the highest canopy density to avoid significant under-estimation when estimating biomass.

To show the performance of the combined approach (i.e., self-calibration and regression) with respect to self-calibration only in this report, we selected four tiles along a north-south transect in Europe and Africa containing a Sentinel-1 image. Figure 4-7 to Figure 4-10 show estimates of the two model parameters obtained with the combined approach and from the self-calibration at each location. The estimates with the combined approach appear to well represent the levels of the backscatter of an unvegetated surface and of the densest canopies. On the contrary, with self-calibration, there are cases when the estimates of one or both parameters are affected by the distribution of backscatter values, often following the choice of the threshold to define the "ground" and the "dense forest" classes. We also show the modelled canopy density with Equation (4-8) to confirm the suitability of the model proposed in Equation (4-8) to relate canopy density and SAR backscatter. It is, however, remarked that these curves should not be interpreted as having a predicting meaning since the model behind them was here introduced only to estimate model parameters. Being a linear approximation of the relationship between canopy density and SAR backscatter, which indeed is non-linear because of the tree transmissivity component (see Equation (4-1)), the curves have only a qualitative meaning.

Figure 4-7 to Figure 4-10 also show examples of σ_{gr}^0 and σ_{veg}^0 as a function of the five intervals of local incidence angle considered in this work. A quadratic model appeared to be a reliable description of the relationship between the model parameters and incidence angle and performed robustly across forest types and forest landscapes of the world.

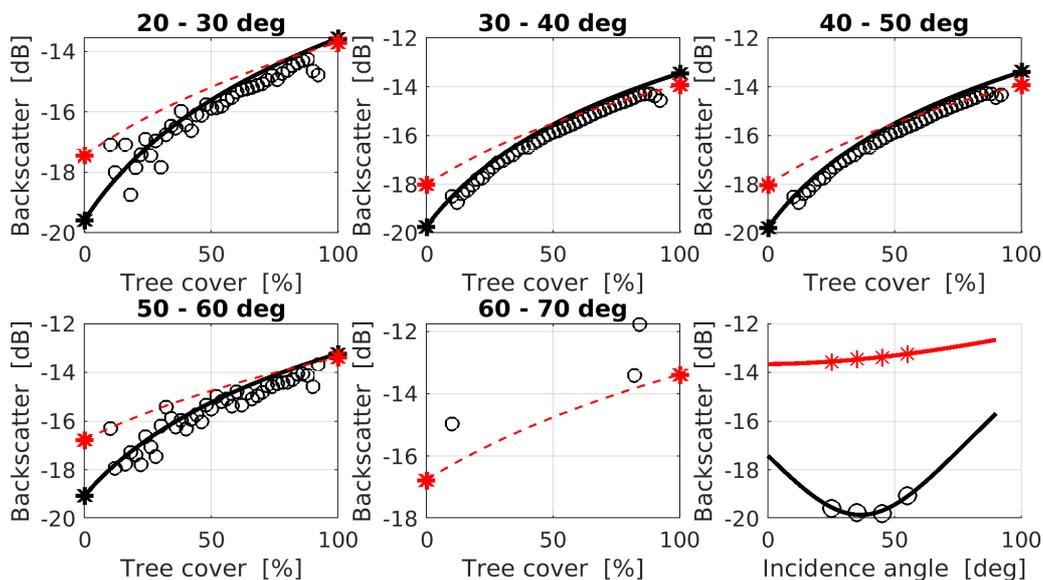


Figure 4-7: Panels with incidence angle range as title show (i) estimates of σ_{gr}^0 and σ_{veg}^0 obtained with the combined approach (black asterisks) and with self-calibration only (red asterisks), (ii) Equation (4-8) in the case

of the combined approach (black solid curve) and self-calibration only (red dashed curve), and (iii) the median backscatter for a given canopy density (circle). The panel relating backscatter to incidence angle shows the estimates of σ_{gr}^0 (black circles) and σ_{veg}^0 (red asterisks) obtained with the combined approach and their quadratic fits (black curve for σ_{gr}^0 and red curve for σ_{veg}^0) spanning the range of incidence angles between 0° and 90°. Dataset: Sentinel-1, VH-polarization., acquired on 5 April 2017. Tile (top left corner coordinate): 64°N, 30°E (boreal forest).

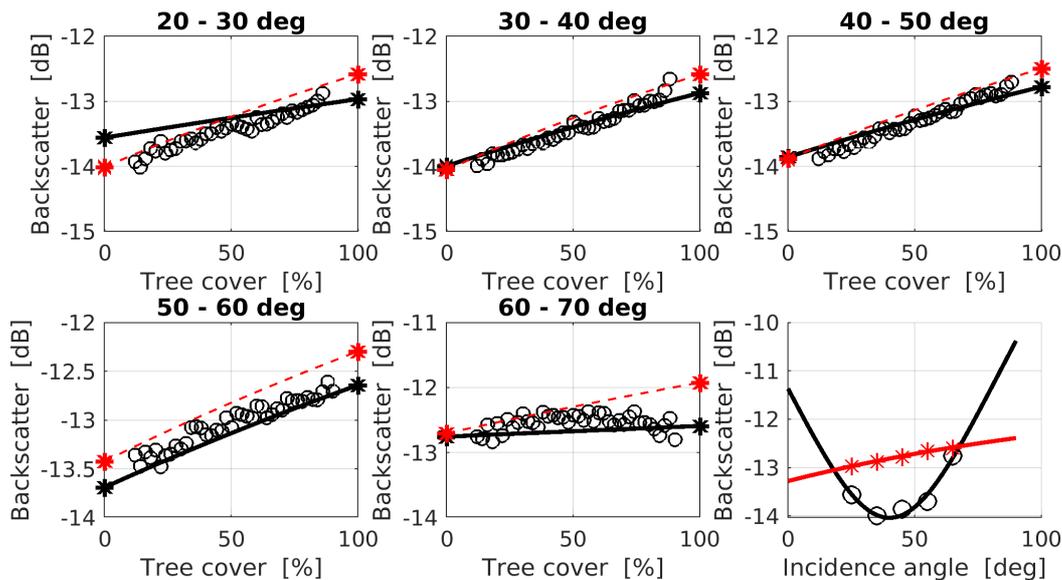


Figure 4-8: Same as in Figure 4-7. Dataset: Sentinel-1, VH-polarization., acquired on 5 July 2017. Tile (top left corner coordinate): 46°N, 11°E (temperate forest).

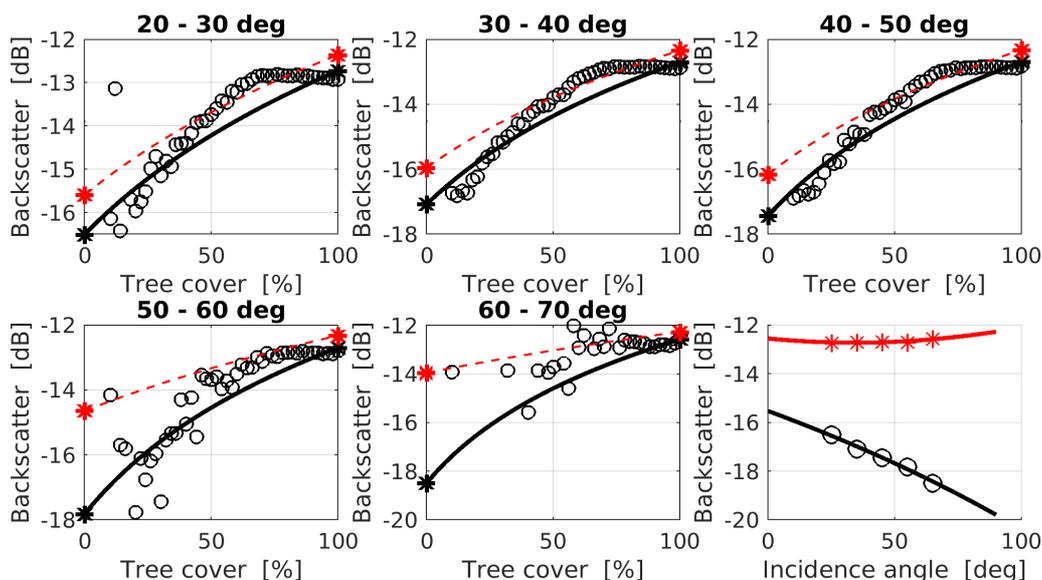


Figure 4-9: Same as in Figure 4-7. Dataset: Sentinel-1, VH-polarization., acquired on 8 July 2017. Tile (top left corner coordinate): 0°N, 11°E (wet tropics).

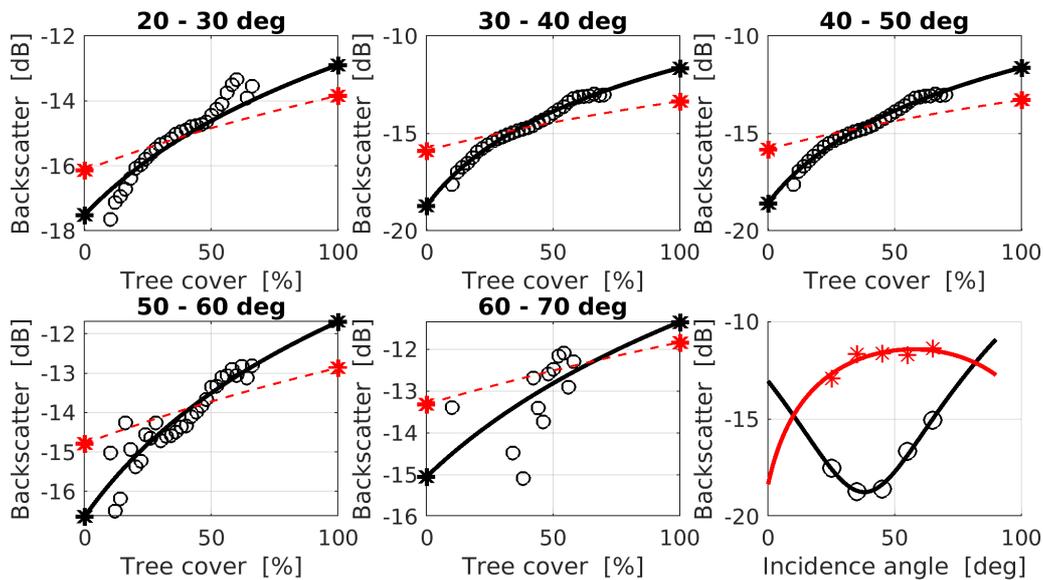


Figure 4-10: Same as in Figure 4-7. Dataset: Sentinel-1, VH-polarization, acquired on 6 July 2017. Tile (top left corner coordinate): 9°S, 17°E (miombo woodlands).

As a result of the multiple requirements imposed to retain an estimate of σ_{gr}^0 and σ_{veg}^0 , it is likely that the $1^\circ \times 1^\circ$ grid of estimates for a certain SAR image is incomplete (see e.g., Figure 46). Similar to the post-processing implemented in BIOMASAR-C for GSV retrieval, we interpolate by interpolating over valid estimates to fill gaps. In contrast to the GSV retrieval case, interpolation was applied separately for each of the five intervals of incidence angles in which the SAR backscatter and canopy density were stratified. The quadratic model was then applied to obtain a raster for each of the two parameters with these corresponding to the raster of local incidence angles for a given SAR backscatter image. Figure 4-11 shows an example of σ_{gr}^0 and σ_{veg}^0 rasters for the Sentinel-1 image used in Figure 4-9. The moderate topography in the $1^\circ \times 1^\circ$ area covered by the SAR image is visible in the local incidence angle image. It is also clear that incidence angle increased from east to west (i.e., the image was acquired along a descending orbit). The raster image of σ_{gr}^0 shows a decrease of the backscatter for increasing incidence angle, as shown by the panel relating incidence angle and σ_{gr}^0 estimates in Figure 4-9. For σ_{veg}^0 , we can hardly see any variability as shown by the same panel in Figure 4-9. Both images however show that the model parameter estimates follow the patterns of the local incidence angles. Figure 4-11 also shows that while the SAR backscatter image is affected by a seam corresponding to adjacent Sentinel-1 sub-swaths, the seam does not appear in the σ_{gr}^0 and σ_{veg}^0 images because of the interpolation used to generate these images. The consequence is the presence of a seam in the AGB map obtained from the Sentinel-1 image.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	88	14.02.2019	

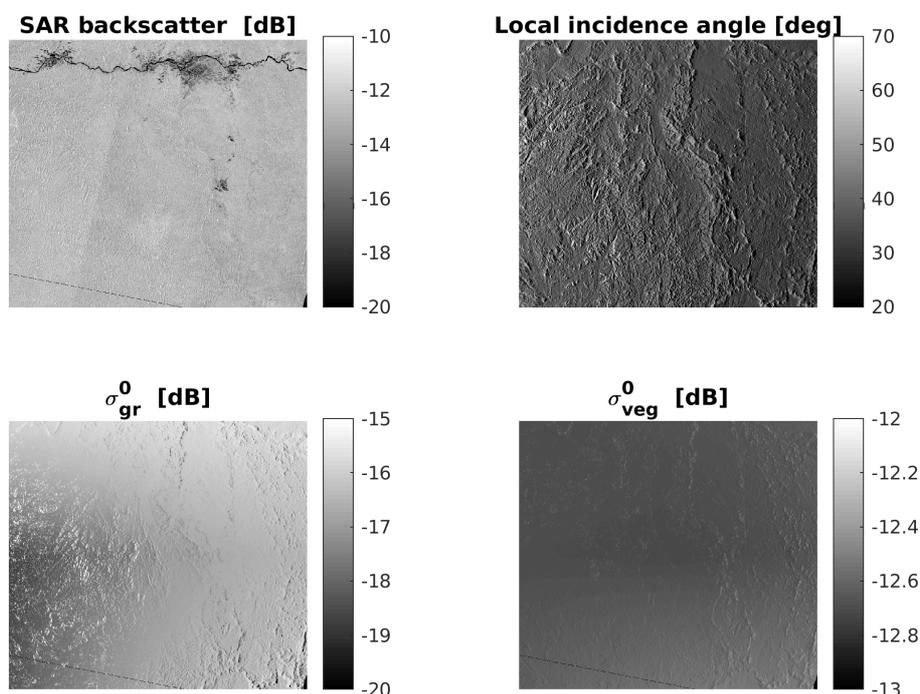


Figure 4-11: Illustrating the raster images of the estimates of σ_{gr}^0 and σ_{veg}^0 (bottom row) for the tile used in Figure 4-9. The top row shows the image of the SAR backscatter and the image of the local incidence angle.

4.3.2.2 Inverting the forest backscatter model

The estimation of AGB using the combined set of Equations (4-1, 3-5 and 3-6) requires a numerical minimization technique and a constraint on the maximum retrievable AGB. In addition, estimates of AGB obtained for a backscatter measurement not within the range of modelled backscatter values need to be corrected for.

Figure 4-12 shows the simulation of the Water Cloud Model in Equation (4-1) for a pixel with a maximum AGB of 362 Mg/ha. The backscatter first increases rapidly with increasing AGB; then, the sensitivity of the backscatter to AGB reduces. The range of backscatter values covered by the model is about 2 dB. The estimation of AGB for a backscatter measurement falling within this range returns realistic numbers. If a measurement of backscatter falls outside this range, the inversion rule overrides the estimate because it is either negative or unrealistic. We define two intervals of backscatter to which the inversion is either associated with the maximum AGB or 0 Mg ha⁻¹ depending on whether the measurement is above or below the range of modelled backscatter. The intervals account for the uncertainty of the backscatter measurement. Measurements that are within an interval between the minimum or maximum modelled backscatter and this value plus at least 3 times the standard error of the measurement are associated with 0 and the maximum AGB, respectively. Otherwise they are assumed to be unrealistic and discarded from the retrieval.

To visualize the difference between this retrieval approach and the approach that retrieves GSV, we illustrate the model fit obtained with Equation (4-2) (i.e., not exploiting height and canopy density allometries, as in Figure 4-12). For a given measurement of the backscatter, the Water Cloud Model based on allometries allows higher AGBs to be estimated in the case of high AGB. This is an improvement on the use of BIOMASAR-C for GSV retrieval which tends to locally underestimate AGB in high biomass forests.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	89	14.02.2019	

Because of the limited sensitivity of C-band SAR backscatter to biomass, the biomass map obtained from a single SAR backscatter image is often characterized by the extreme values 0 and maximum biomass. For this reason, the performance of the two BIOMASAR-C implementations described in this document cannot be compared.

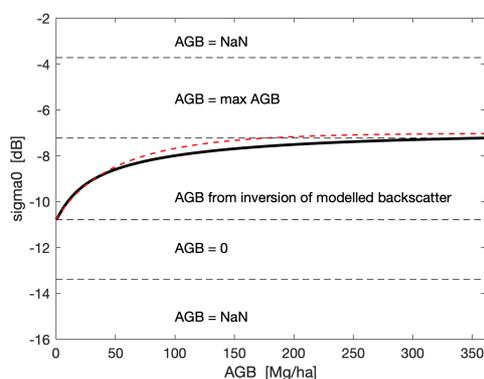


Figure 4-12: AGB retrieval rules depending on the level of the backscatter measurement.

4.3.2.3 Multi-temporal combination of AGB estimates

Similar to the retrieval of GSV, the final estimate of AGB is obtained with the linear combination in Equation (4-7) of the AGB estimates from the individual SAR backscatter images.

4.4 The BIOMASAR-L algorithm

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. The highest sensitivity was usually reported for the L-band cross-polarized intensity. 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 decreases and the volume scattering contribution from the canopy increases, and forest structural differences have been shown to affect the functional relationship between backscatter and GSV or AGB. Although not as evident as in the case of C-band, L-band backscatter is affected by environmental conditions at the time of acquisition as demonstrated in Section 3.

Algorithms aiming to exploit the sensitivity of L-band backscatter to biomass 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, and particularly the tropics, the number of available plots is very limited so that models may only be calibrated using reference information collected over large areas (Bouvet et al., 2018). Conventional approaches generally necessitate working with mosaics of L-band backscatter imagery that are compiled in such 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, mosaicking entails careful selection of imagery to minimize between-orbit radiometric differences. Alternatively, empirical

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	90	14.02.2019	

inter-orbit normalization techniques may be applied (e.g., De Grandi et al., 2011; Shimada et al., 2010).

BIOMASAR-L is here used to retrieve biomass globally from the L-band backscatter mosaics. As with BIOMASAR-C, multi-temporal stacks of SAR backscatter observations are modelled individually and biomass is obtained by integrating the estimates from the individual mosaics. Major improvements in retrieval 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.

To model the relationship between L-band backscatter observations and biomass, we use the same Water Cloud type of models as for C-band shown in Section 4.3. 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 $\sim 35^\circ$ 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 biomass containing only first order scattering depicted well the observed relationship of SIR-C L-HV intensity and biomass.

As in the case of BIOMASAR-C, Equation (4-3) was used in year 1 of CCI Biomass to estimate GSV from L-band backscatter observations. The implementation of this version of the WCM in BIOMASAR-L is described in Section 4.4.1. In year 2, we consider the original WCM with gaps in Equation (4-1) and the same set of functional dependencies between canopy density, tree height and above-ground biomass to express the WCM as a function of AGB. The implementation of the WCM and the allometric functions to retrieve AGB are described in Section 4.4.2.

4.4.1 BIOMASAR-L for GSV retrieval

The BIOMASAR-L algorithm mimics BIOMASAR-C with a few adaptations. However, there are different scattering mechanisms occurring at C- and L-band. While the concept of the BIOMASAR-L approach is essentially the same as for BIOMASAR-C, some modifications were necessary as the data were delivered as mosaics, with each containing imagery acquired from up to four different orbits. Figure 4-13 shows an image 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 4-13). The acquisition date map allows 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.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	91	14.02.2019	

4.4.1.1 Estimation of 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). This is achieved 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.

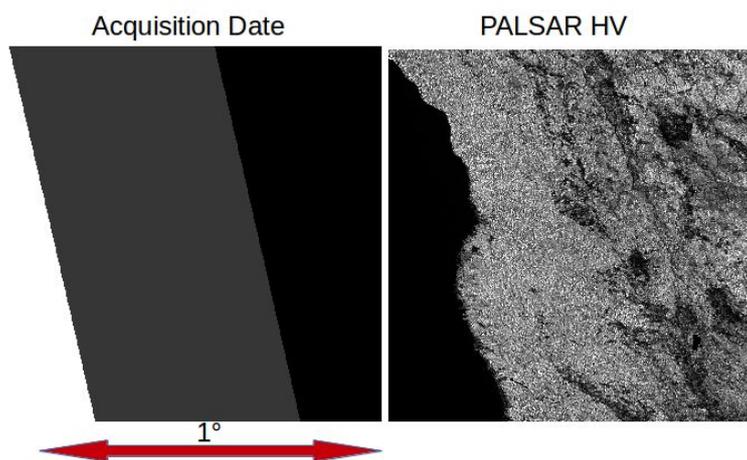


Figure 4-13: $1^\circ \times 1^\circ$ tile of ALOS PALSAR HV backscatter (right) and the acquisition dates of the images used to create the mosaic (left). The acquisition date map allows the acquisition date for each pixel in the tile to be identified. In this example, images from three different acquisition dates (illustrated in white, grey, and black) have been used.

The workflow to estimate σ_{gr}^0 and σ_{df}^0 for each orbit in a $1^\circ \times 1^\circ$ ALOS-2 PALSAR-2 image tile comprises the following three steps.

Data preparation

The $1^\circ \times 1^\circ$ 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 maps are resampled to the target $\sim 100 \times 100 \text{ m}^2$ pixel size ($0.00888^\circ \times 0.00888^\circ$). Only HV polarization imagery is considered because the results of GlobBiomass suggested that the integration of HH polarization does not improve the retrieval accuracy and locally (e.g., flooded forests) leads to systematic biases in the estimates. After aggregating the datasets, the CCI-LC map is used to mask out cropland, urban areas, impervious surfaces, and permanently and temporarily flooded land cover types. Layover/shadow areas are masked out based on maps provided by JAXA. Although topographic corrections had been applied by JAXA, the L-band mosaics still exhibit significant topographic effects. Areas imaged with a local incidence angle of less than 28° , as well as areas of layover and shadow, are therefore masked.

Identification of areas of low and high canopy density

Using the optical canopy density maps, areas of sparse and dense forest cover in each orbit and tile are identified to infer the backscattered power arising from the forest floor and dense forest canopies. As a compromise between characterising the spatial variability of scattering from the forest floor and canopies due to changes in, for instance, soil/canopy moisture, freeze/thaw

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	92	14.02.2019	

transitions, or the local incidence angle (primarily relevant for the moderate resolution Wide Beam data), and the fact that a certain number of image pixels need to be considered to reliably estimate the model parameters σ_{gr}^0 and σ_{df}^0 , the imagery from a single orbit in each $1^\circ \times 1^\circ$ tile is divided into $15 \times 15 \text{ km}^2$ larger blocks. For each block, the two parameters are then estimated using locally variable canopy density thresholds for delineating areas of sparse and dense forest cover in the radar imagery. 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 (100 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 (100) with a canopy density above the selected threshold is found. The minimum threshold is set to 70%. If this requirement cannot be fulfilled, a direct estimate of the respective parameter is not considered possible and interpolations based on adjacent areas with valid estimates of the model parameters is required. Interpolation is primarily required in areas of continuous dense forest cover, such as the Amazon and Congo basins, where the number of pixels corresponding to sparse forest cover is limited.

Model parameter estimation

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 4-14). For the estimation of σ_{df}^0 , we use the mode of the backscatter distribution in areas of high canopy density once at least 100 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 the estimation of the mode. The parameter σ_{gr}^0 is, instead, estimated with the mode of the backscatter distribution in areas of low canopy density.

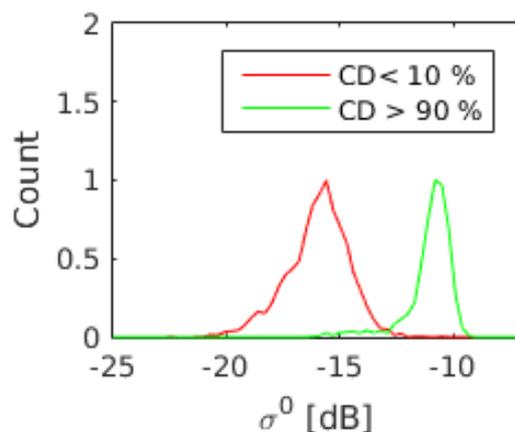


Figure 4-14: Histograms of L-HV backscatter in areas of low (red) and high (green) canopy density according to Landsat.

4.4.1.1.1 Estimation of 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 (4-2) describing backscatter as a function of the forest transmissivity of the densest forests in the area of interest, which itself is considered a

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	93	14.02.2019	

function of residual canopy gap fraction, η_{df} , and canopy height, h_{df} , when inverting the model for σ_{veg}^0 :

$$\sigma_{veg}^0 = \frac{\sigma_{df}^0 - \sigma_{gr}^0 T_{df}}{1 - T_{df}} \quad (4-10)$$

with:

$$T_{df} = (1 - \eta_{df}) + \eta_{df} e^{-\alpha h_{df}} \quad (4-11)$$

Figure 4-15 illustrates the concept for estimating 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 4-15, the compensation for the residual transmissivity leads to an estimate for σ_{veg}^0 which is ~ 1.5 dB higher than σ_{df}^0 .

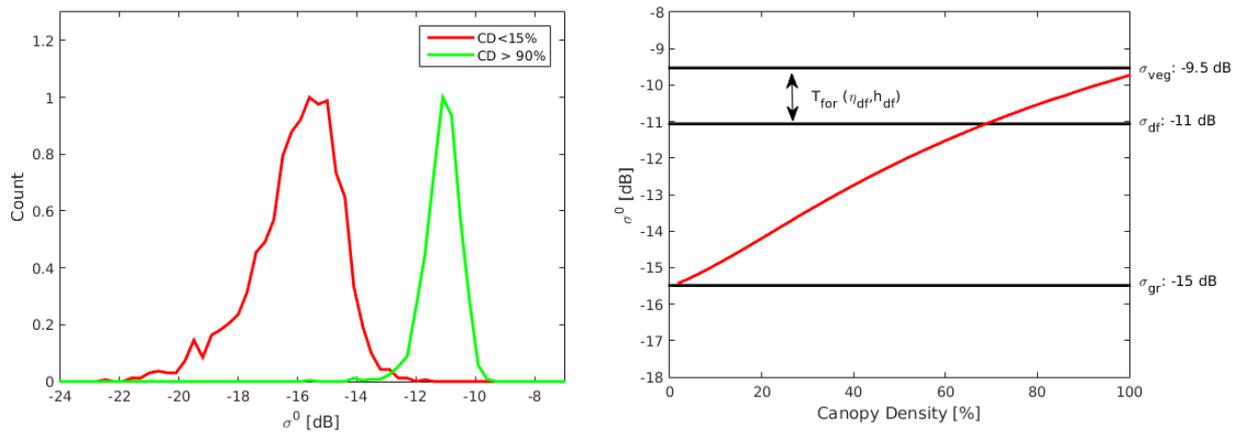


Figure 4-15: Left: Histogram of L-HV backscatter in areas of low (red) and high (green) canopy density. Right: Modelled relationship of L-HV backscatter as function of canopy density (red line) with σ_{gr}^0 and σ_{df}^0 derived from the histograms and σ_{veg}^0 derived from σ_{df}^0 with the aid of 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 , the average height of dense forest, and the residual canopy gap fraction are required. The height of dense forests is characterized for each $1^\circ \times 1^\circ$ tile from local forest height estimates derived from ICESAT GLAS waveforms with the height difference between the signal beginning and the ground peak in the waveform, which is generally identified by fitting multi-gaussian models to the waveforms (Hofton et al., 2000). The residual canopy gap fraction in areas for which the optical canopy density map suggests dense forest canopies is estimated with the ratio of signal energy returned from the canopy and the ground after normalizing the ratio with respect to differences in reflectivity of canopies and the forest floor at 1000 nm wavelength. 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^{-1} (see [RD-8]). 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^{-1} instead of 0.5 dB m^{-1} , σ_{veg}^0 changed by only up

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	94	14.02.2019	

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⁻¹ which, based on our current understanding, may be associated with images acquired under frozen conditions. The L-band 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 appeared to be justified. For images acquired under frozen conditions, which had to be considered by JAXA when generating the annual mosaics to obtain complete coverage at high latitudes (boreal North America and Eurasia), the modelling of backscatter as a function of GSV may therefore not be optimal. It has to be considered, however, that images acquired under frozen conditions generally present weak sensitivity to GSV (Santoro et al., 2015b). The reduced sensitivity is then accounted for in the multi-temporal combination of GSV estimates derived from single acquisitions so that such observations should, ideally, not result in systematic biases in the multi-temporal GSV estimates.

When estimating the transmissivity of the densest forests at L-band globally (i.e., with > 90 % density according to Landsat) with the aid of ICESAT GLAS (Figure 4-16), 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 canopy 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. With the exception of imagery acquired under frozen conditions, differences in the transmissivity due to varying two-way canopy attenuation ($\alpha = 0.5 - 1$ dB/m) can be expected to be minor.

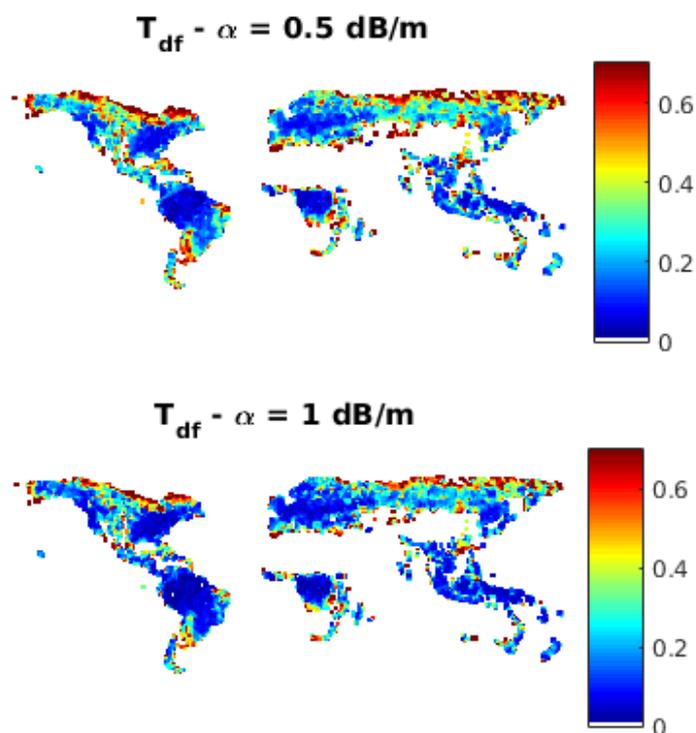


Figure 4-16: L-band transmissivity of dense forests modelled based on ICESAT GLAS based estimates of canopy density and height when assuming a two-way tree attenuation of 0.5 (top) and 1 dB m⁻¹ (bottom), respectively.

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 4-17 for 1° x 1° PALSAR tiles covering forest areas in Montana (USA), Lapland (Sweden), and the Eastern Cape Province (South Africa). The retrieval of GSV was undertaken in one case considering the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	95	14.02.2019	

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 a function of GSV and as a result leads to systematic overestimation of GSV.

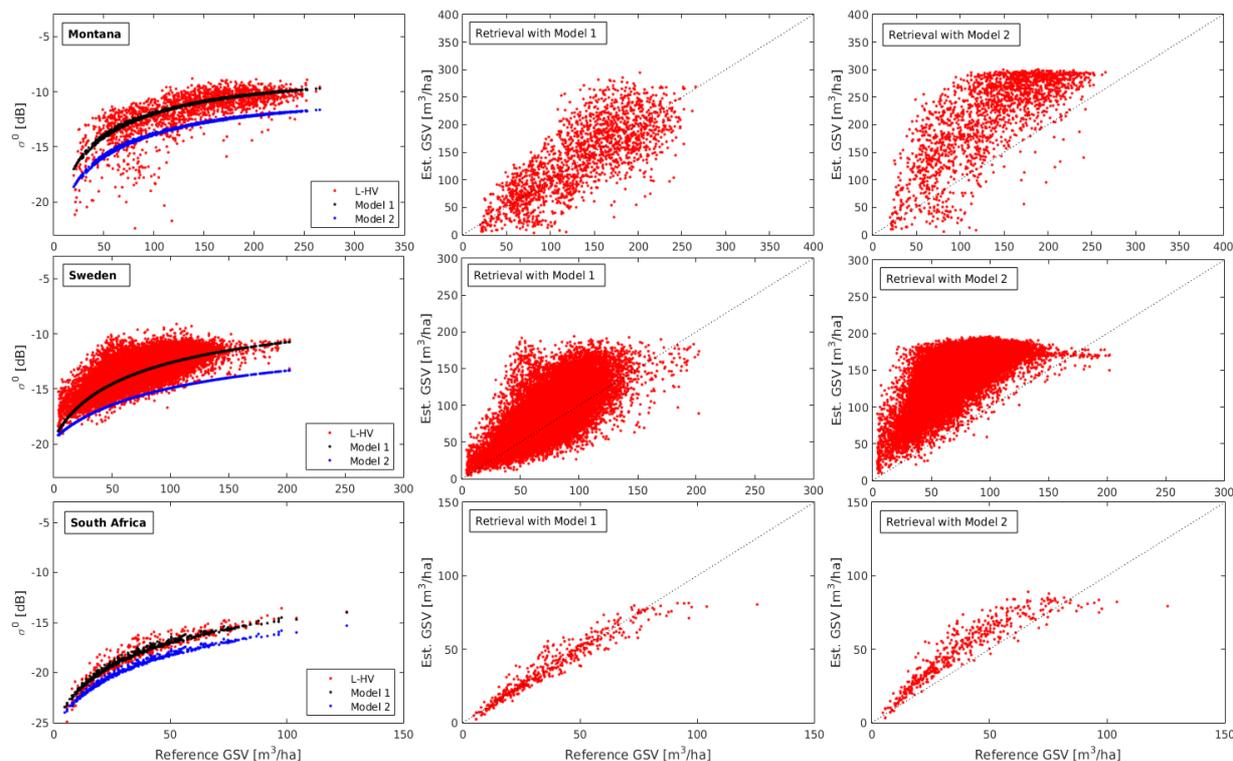


Figure 4-17: Observed and modelled relationship of L-HV backscatter and GSV for a forest area in Montana (top row), Sweden (centre row), and South Africa (bottom row) using BIOMASAR-L for model calibration (left). Model 1 accounts for the transmissivity of dense forests when estimating σ_{veg}^0 whereas Model 2 does not. The effect on the GSV retrieval is demonstrated in the centre and right plots. L-band derived estimates of the GSV are compared to regional maps (Kellndorfer et al., 2013; Reese et al., 2002; Bouvet et al., 2018) which had been produced using radar/optical imagery and local forest inventory data for calibrating retrieval models.

4.4.1.1.2 Estimation of β

As in the case of C-band, while the model parameters σ_{gr}^0 and σ_{veg}^0 need to be determined as locally as possible because the backscatter to GSV relationship may present a strong (spatial and temporal) dependence on the environmental imaging conditions, the forest transmissivity parameter β , which characterizes how quickly backscatter increases with increasing GSV, needs to be adjusted primarily to capture differences in the backscatter to GSV relationship associated with forest structural differences. Nonetheless, a seasonal dependence may also exist as, for instance, freeze/thaw transitions, wet/dry season differences, or the defoliation state of forests potentially alter the transmissivity of forests at L-band.

Studies concerned with modelling L-band backscatter as a function of GSV focussed primarily on boreal forests in Scandinavia and Russia. These studies suggest that β was generally of the order of 0.004 ha m^{-3} (Askne et al., 2003; Santoro et al., 2006; Santoro et al., 2015b) with no apparent dependence on the forest type (i.e., deciduous vs. evergreen). However, recent studies using C- and L-band imagery in Sweden suggest that the parameter is influenced by forest structural differences associated with different forest management (Cartus et al., 2017). Limited

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	96	14.02.2019	

experience has so far been gathered on the transmissivity to biomass relationship at C- and L-band in the temperate zone (Cartus et al., 2012), or subtropics or tropics.

In order to be able to consider the effect of forest structural differences on the L-band backscatter to biomass relationship, we therefore opted to model the relationship of forest transmissivity at L-band and biomass based on the information on forest structure provided by ICESAT GLAS. Rewriting Equation (4-3) to express the transmissivity as a function of canopy density and height yields:

$$T_{for} = (1 - \eta) + \eta e^{-\alpha h} \quad (4-12)$$

where the GLAS height estimates can be used to characterize the canopy height, h , and the ratio of LiDAR energy received from the canopy and ground (normalized for reflectivity differences between canopy and ground) in each 65 m footprint can be used to estimate η . The two-way tree transmissivity α was again assumed to be 0.5 dB m^{-1} . For each GLAS footprint covering one of the available GSV reference maps, the corresponding estimate of GSV was extracted to analyse the trend of T_{for} , simulated with the aid of GLAS as a function of GSV, and to estimate the transmissivity parameter β by fitting Equation (4-13):

$$T_{for} = e^{-\beta V} \quad (4-13)$$

to the observed trend by means of non-linear regression. Regional maps which were considered in this context included maps produced for Sweden, South Africa, and Mexico in the GlobBiomass project (Rodríguez-Veiga et al., 2019) and other existing products such as the National Biomass and Carbon Dataset for the US (Kellndorfer et al., 2012), or maps produced by fusing airborne LiDAR and optical imagery. Note that when regional map products reported AGB, AGB was converted to GSV with the aid of the BCEF layer described in Section 3.16.

Figure 4-18 illustrates the relationship of T_{for} and GSV as well as the fit of the model in Equation (4-13) for seven FAO eco-regions. The trend of the simulated T_{for} as a function of GSV and the derived estimates for β confirmed the previous observation that significant differences between the eco-regions exist. The comparison indicated as well that low values for β of $\sim 0.004\text{-}0.005 \text{ ha m}^{-3}$ 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 m^{-3} are needed. In the analysis, it was assumed that a two-way tree transmissivity α of 0.5 dB m^{-1} 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^{-1} , the derived estimates for β changed only by about 0.001 to 0.002 ha m^{-3} . Larger differences of up to 0.004 ha m^{-3} were observed only in the tropics (i.e., the estimates for β reached values of up to 0.015 ha m^{-3}), which implies a loss of sensitivity of backscatter to GSV in very low GSV ranges. An evaluation of L-band backscatter to GSV relationships with the aid of existing map products in the tropics did not, however, suggest that such high values for β are needed for modelling backscatter.

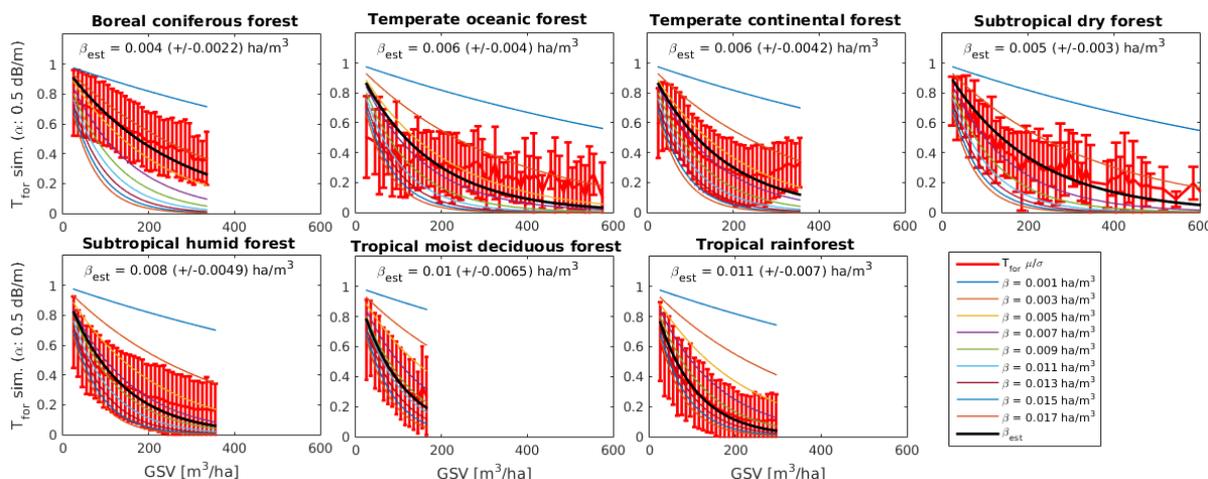


Figure 4-18: Estimates for β (incl. 95% confidence bounds) obtained for different FAO eco-regions by fitting the model in Equation (4-13) to the observed relationship of forest transmissivity, simulated with the aid of ICESAT GLAS and the corresponding GSV in map products.

Based on the results presented above, a global LUT for the forest transmissivity coefficient β at L-band was compiled. The LUT is visualized in Figure 4-19. For boreal and dry sub-tropical forests, β in the range of 0.004 to 0.005 ha/m^3 was obtained. In temperate and sub-tropical humid forest regions, higher values in the range of 0.006-0.008 ha/m^3 are used and in the tropics values of 0.011 ha/m^3 .

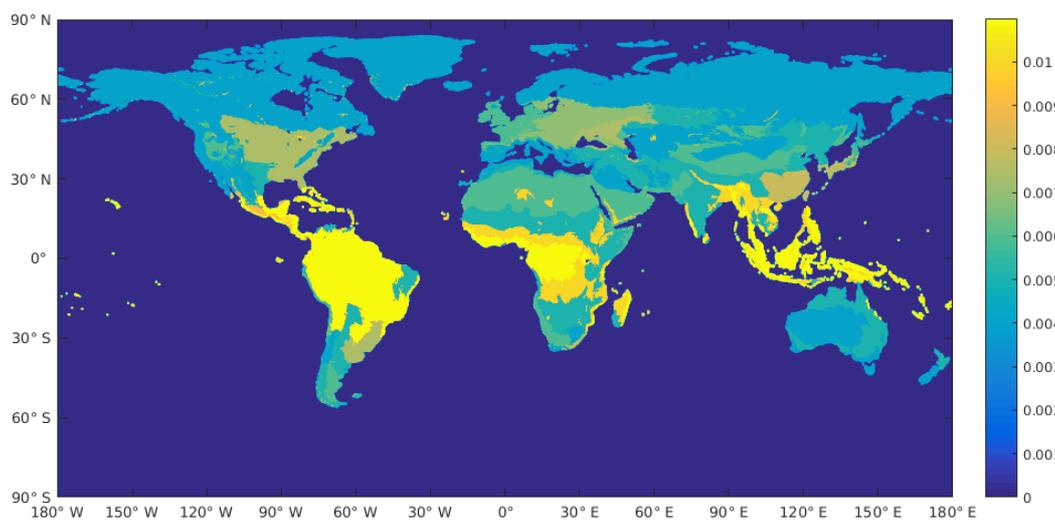


Figure 4-19: Estimates of the forest transmissivity coefficient β [ha/m^3] for L-band.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	98	14.02.2019	

4.4.2 BIOMASAR-L for AGB retrieval

As in the case of BIOMASAR-C, we seek to adapt the modelling and model calibration framework so that:

- AGB can be retrieved directly,
- incidence angle effects in the L-band backscatter to AGB relationship can be accounted for.

The adapted BIOMASAR-L retrieval approach generally complies with the modelling basis and workflow presented in Section 4.3.2 for the case of Sentinel-1 C-band data. In this section, we therefore focus on differences between BIOMASAR-C and BIOMASAR-L which are a consequence of differences in the response of L-band backscatter to canopy density, height, and eventually AGB, and present initial results achieved with the new approach.

The retrieval approach comprises three major steps:

- Calibration of the model in Equation (4-1) with the aid of a Landsat canopy density map, considering the dependence of the relationship between backscatter and canopy density on incidence angle,
- Retrieval of AGB for each single acquisition in a multi-temporal stack of L-band observations based on a new formulation of the WCM which now relates backscatter to AGB,
- Weighted multi-temporal combination of single image AGB estimates.

Model calibration

The model, relating L-band backscatter to AGB, is based on that in Equation (4-1). The model relates backscatter to canopy density, η , while considering that backscatter may not only be affected by the level of canopy closure but also the depth (i.e., height) of the canopy and the strength of the attenuation while passing through the canopy. The model is reformulated so that backscatter can be modelled as a function of AGB by:

- Expressing η as function of canopy height, Equation (3-5),
- Exploiting allometric relationships between height and AGB, Equation (3-6),

so that:

$$\sigma_{for}^0 = (1 - \eta(h(B)))\sigma_{gr}^0 + \eta(h(B))\sigma_{gr}^0 e^{-\alpha h(B)} + \eta(h(B))\sigma_{veg}^0 (1 - e^{-\alpha h(B)}) \quad (4-14)$$

When assuming that the allometric relationships between η and height and between height and AGB are known for a given type of forest, three model parameters remain unknown: σ_{gr}^0 , σ_{veg}^0 , and α . The unknown parameters may be estimated with the aid of existing global maps of canopy density (Hansen et al., 2013) by means of least-squares regression using the original model in Equation (4-1). Because no spatially explicit and up-to-date maps of canopy height exist at global scale, the model in Equation (4-1) needs to be reformulated to:

$$\sigma_{for}^0 = (1 - \eta)\sigma_{gr}^0 + \eta\sigma_{gr}^0 e^{-\alpha h(\eta)} + \eta\sigma_{veg}^0 (1 - e^{-\alpha h(\eta)}) \quad (4-15)$$

In the modified model in Equation (4-15), height is expressed as a function of η by:

$$h = -\frac{\log(1-\eta)}{q} \quad (4-16)$$

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	99	14.02.2019	

where q characterizes the allometric relationships between η and height, Equation (3-5).

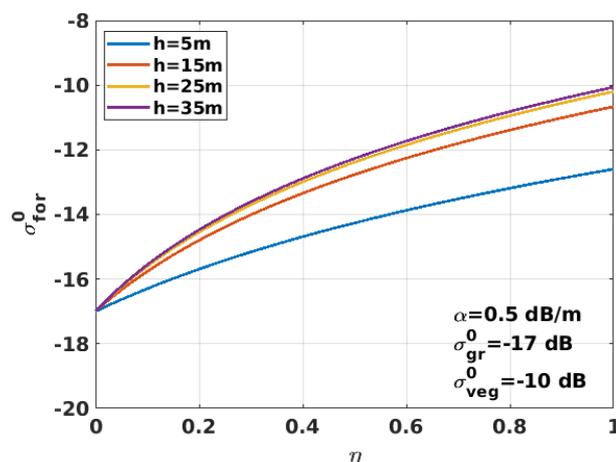


Figure 4-20: Simulated differences in the L-band HV backscatter response to changes in η for forests with canopy heights between 5 and 35 m.

In contrast to C-band, the two-way transmissivity term in the model ($e^{(-\alpha h)}$) cannot be neglected because it significantly affects the backscatter to η relationship (Cartus et al., 2018). The relevance of the two-way transmissivity is illustrated in Figure 4-20. Assuming a two-way attenuation of 0.5 dB/m (a value considered realistic for boreal forests; Praks et al., 2012; Askne et al., 2005), the simulations based on Equation (4-1) demonstrate that the two-way transmissivity term cannot be neglected at L-band since the response of backscatter to changes in η depends strongly on the canopy height, particularly for low canopy height ranges (<15m). These simulations furthermore advocate the use of the model in Equation (4-14) to relate L-band backscatter to AGB, since the model allows regional differences in the interrelationships between canopy density, height, and AGB to be captured.

The model in Equation (4-15) is calibrated for each orbit contained in a $1^\circ \times 1^\circ$ tile of the ALOS-2 PALSAR-2 backscatter mosaics provided by JAXA by least-squares regression, resulting in estimates for the unknown parameters σ_{gr}^0 , σ_{veg}^0 , and α . In order to capture the dependence of the parameters on the local incidence angle, models are fitted to the observed trend of L-band backscatter as a function of Landsat canopy density and separately for 10° wide ranges of the local incidence angle (i.e., 20-30°, 30-40°, ..., 80-90°).

To increase the robustness of the model fit and the obtained model parameter estimates:

- Areas for which the CCI land cover map reports cropland, bare ground, wetland or built-up areas are masked,
- Canopy densities below 10% are not considered because backscatter to canopy density relationships present high uncertainty in sparsely vegetated areas,
- Areas for which backscatter is affected by layover/shadow are not considered,
- Models are fitted to the observed backscatter to canopy density relationship after binning backscatter in 2% canopy density intervals. For each canopy density interval, a trimmed mean of the corresponding backscatter observations is calculated. This is only if at least 5 backscatter observations for a given interval are available (if not, the respective canopy density interval is not considered) to account for the fact that a

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	100	14.02.2019	

canopy density map for the year 2010 is to be used for calibrating models for radar data acquired in 2017,

- The estimation of the two-way attenuation parameter α is constrained to a range of 0.2 to 1.5 dB/m to avoid “unrealistic” model fits that may be associated with local errors/artefacts in the Landsat canopy density maps.

Examples of the fit of the model in Equation (4-15) in three different areas are presented in Figure 4-21. For the model calibration, values of the parameter q in Equation (4-16) were selected. In general, a well-defined relationship between L-band backscatter and Landsat canopy density is observed, as well as a reasonable fit of the model up to the maximum canopy density of 100%. In areas characterized by steep topography (Swiss Alps, Pacific Northwest in the United States), we find that the relationships between backscatter and canopy density differ significantly depending on the local incidence angle. Tests of the model calibration across different forest sites in the wet tropics indicated, however, that above a canopy density of 60 to 70%, L-band backscatter no longer presents any sensitivity to canopy density (see example in Figure 4-21, top left). In our understanding, a lack of sensitivity of L-band backscatter to canopy densities beyond 60% is rather unrealistic. The observations indicate that for dense tropical forest the Landsat canopy density product appears to systematically underestimate high canopy densities. As a consequence of these biases in the Landsat-derived maps, estimates for the model parameter σ_{veg}^0 are also systematically too large. For estimating the model parameter in the wet tropics, we therefore follow the approach in Section 4.4.1 where σ_{veg}^0 is estimated based on the observed median backscatter in areas where Landsat reports forests with canopy densities close to 100%.

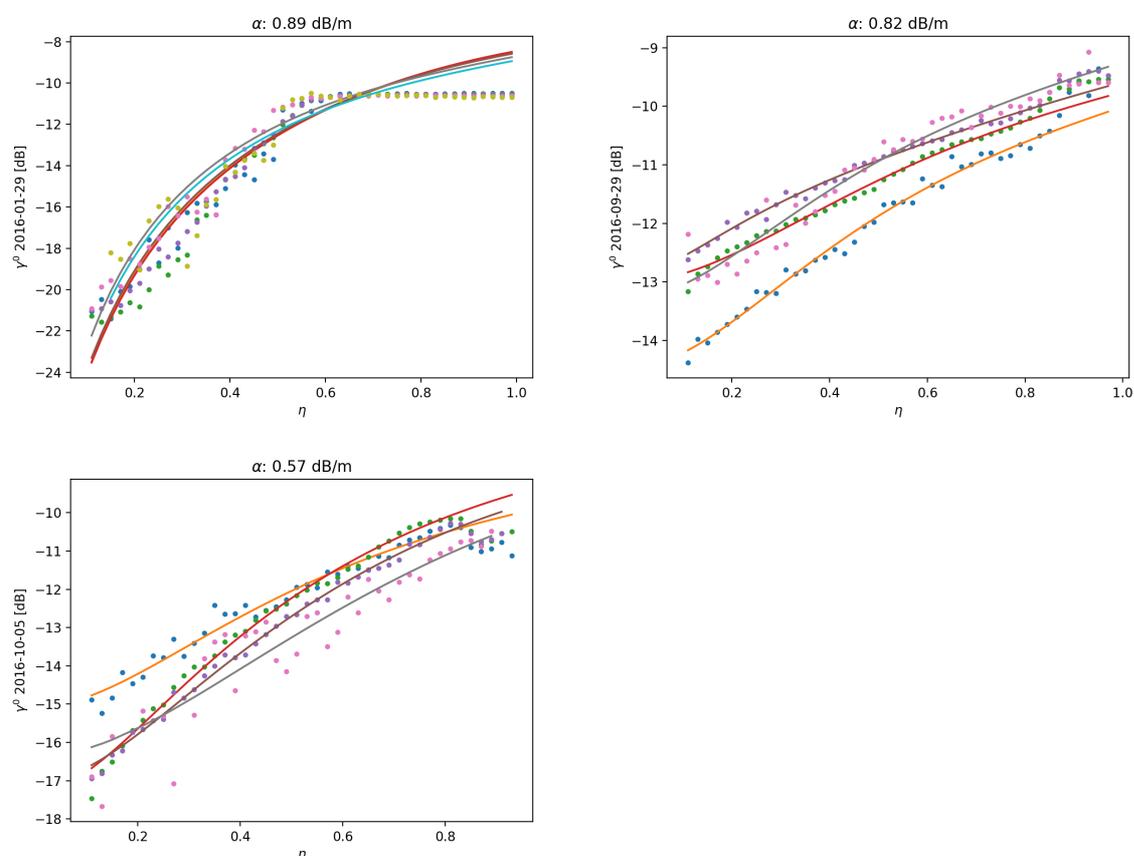


Figure 4-21: ALOS-2 LHV backscatter as function of Landsat canopy density (dots) for three acquisitions covering tropical forest in Gabon (top left), the Swiss Alps (top right), and the Pacific Northwest of the United

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	101	14.02.2019	

States (bottom left). The curves represent the fit of the model in Equation (4-14) for backscatter observations in different local incidence angle ranges. Derived estimates for α are reported for each region.

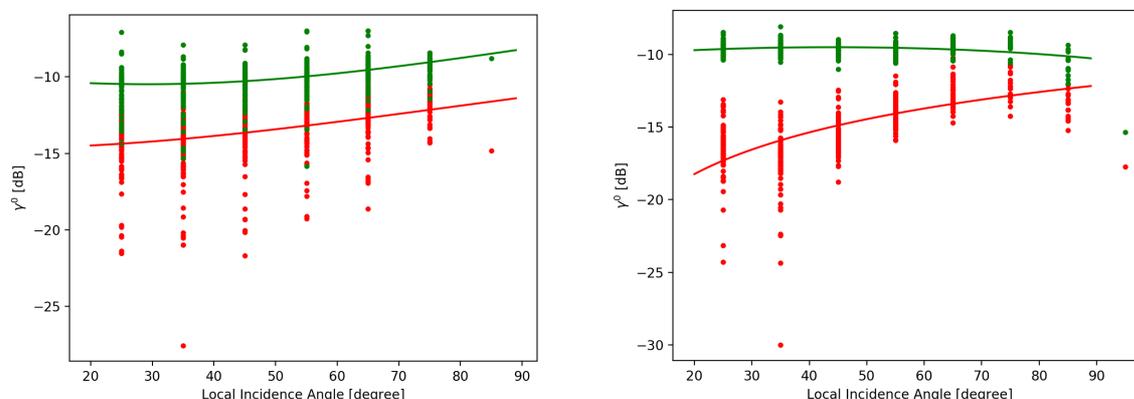


Figure 4-22: Estimates for the model parameters σ_{gr}^0 (red dots) and σ_{veg}^0 (green dots) as function of the local incidence angle for ALOS-2 orbits covering the Swiss Alps (left) and the Pacific Northwest of the United States (right). The curves represent the corresponding fit of a second order polynomial.

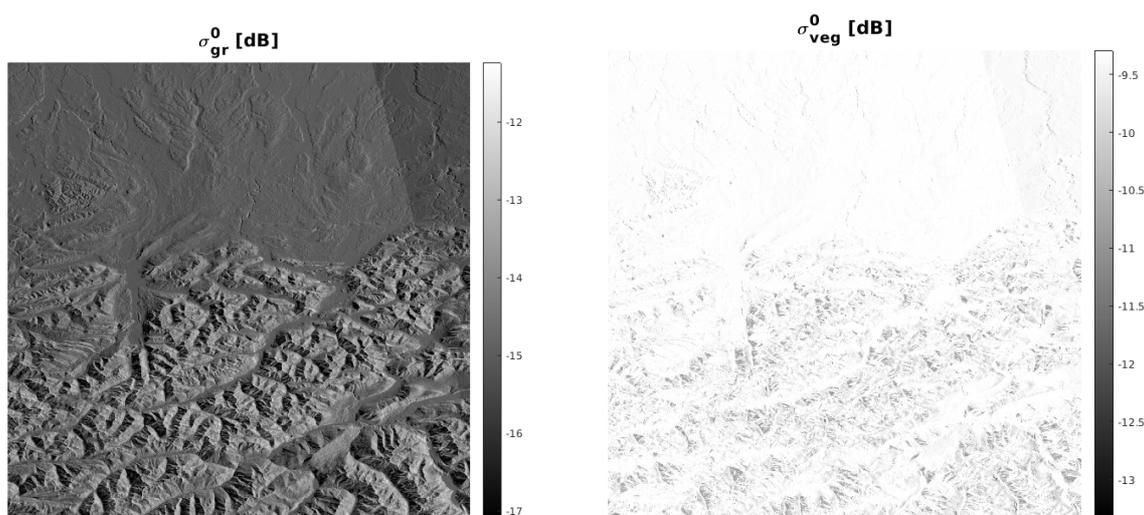


Figure 4-23: Spatially explicit estimates for the model parameters σ_{gr}^0 and σ_{veg}^0 for a $1^\circ \times 1^\circ$ ALOS-2 tile (year 2016) covering the Swiss Alps.

Once σ_{gr}^0 and σ_{veg}^0 have been estimated for each local incidence angle range, second order polynomials are fitted to the observed relationships between the parameters and the local incidence angle (Figure 4-22). These polynomials are subsequently used to obtain spatially explicit estimates of the parameters using the local incidence angle maps (Figure 4-23). The initial model calibration is performed for each orbit on a $1^\circ \times 1^\circ$ tile-by-tile basis. Since within a given tile only part of the incidence angle range in an ALOS-2 PALSAR-2 image may be represented (in particular in the case of ALOS-2 ScanSAR imagery with a swath width much larger than 1°), we here also consider derived estimates for the model parameters obtained for the backscatter images acquired from same orbit in adjacent tiles (in a $5^\circ \times 5^\circ$ window). In order to increase the robustness of the polynomial fit, the polynomial coefficients are estimated by means of a weighted regression with weights calculated based on the inverse of the standard

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	102	14.02.2019	

error of the fit of the model in Equations (4-15) and (4-16) to the observed relationship between L-band backscatter and Landsat canopy density.

Single image retrieval

Once the parameters σ_{gr}^0 and σ_{veg}^0 are known, estimates for the parameter q , relating canopy density to height, and allometric equations relating height to AGB for a given WWF ecoregion may be used to retrieve AGB for each L-band backscatter image in the multi-temporal stack of observations using Equation (4-14). However, a direct inversion of the model is not possible. In order to retrieve AGB, a numerical minimization approach is required. As in the case of the retrieval of GSV (Section 4.4.1), the retrieval is constrained to a maximum AGB.

Multi-temporal retrieval

Single image derived estimates are subsequently combined in a weighted multi-temporal combination. The weighted combination and the definition of weights have already been described in Section 4.3.

4.5 Methods to derive a merged biomass dataset

C- and L-band derived estimates of biomass are expected to present specific advantages and disadvantages in terms of the sensitivity to biomass, the number of images that are available for multi-temporal combination, or the quality of the pre-processing (calibration, topographic correction, geocoding). It is therefore advised to combine the estimates that were obtained independently but taking into account the respective strengths and weaknesses of each dataset.

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, however, is 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 estimates. Here, the weighted average would be applied to two values only, from BIOMASAR-L and BIOMASAR-C. This still retains the bias, though possibly weakened.

An elegant way of defining the weights is to use a measure of how well estimates and reference values agree (Avitabile et al., 2016). If all biomass estimates 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 biomass 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 (~100 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.

Based on a systematic assessment of differences in global biomass estimates derived from C- (ENVISAT ASAR) and L-band (ALOS PALSAR) backscatter data in the GlobBiomass project, a merging scheme was developed that accounted for:

- Systematic differences in the sensitivity of C- and L-band data to biomass
- The number of observations used for estimating biomass

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	103	14.02.2019	

- Local errors in the model calibration and inversion which, for both BIOMASAR-C and BIOMASAR-L, which are locally adaptive
- Topographic effects, particularly in the L-band mosaics.

A weighting scheme focusing on full resolution (i.e., weights defined at the ~100 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 0.1° and then oversampled to the full resolution of the biomass maps. In that way, the weighting is more geared towards identification of systematic regional differences in the biomass maps.

4.5.1 C- and L-band sensitivity to biomass

The modelling of C- and L-band backscatter suggests that their respective sensitivities to biomass changes with increasing biomass. In the lower ranges of biomass, we expect a higher sensitivity at C-band, whereas in the higher ranges, L-band is expected to present higher 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 (see Equation (4-12) and (4-13)).

Because of stronger attenuation of the C-band signal in the canopy, and thus a faster decrease of the transmissivity with increasing biomass, higher values for the empirical transmissivity coefficient, β , are generally required to describe the transmissivity as a function of biomass at C-band than at L-band. This is exemplified in Figure 4-24, which shows the transmissivity as a function of GSV for two different values of β (left); the lower curve represents L-band and the higher curve C-band. The derivative of Equation (4-13), 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 (4-13):

$$w_s = \left| \frac{\partial T_L}{\partial v} \right| - \left| \frac{\partial T_C}{\partial v} \right| \quad (4-17)$$

where the transmissivity 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 GlobBiomass GSV map as reference GSV. Because the weights are determined at coarse resolution, changes in biomass between 2010 and 2017 are not expected to significantly affect the merging of GSV maps from Sentinel-1 and ALOS-2. In addition, this scheme can be preserved for the case of expressing the WCM as a function of AGB.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	104	14.02.2019	

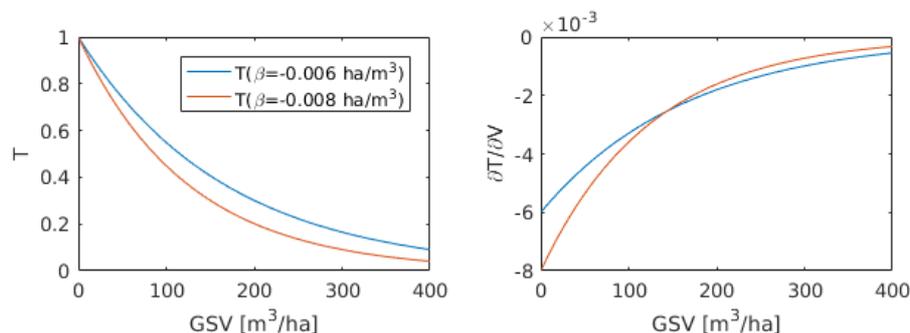


Figure 4-24: Forest transmissivity modelled as a function of GSV for two different values of the forest transmissivity coefficient, β (left). Derivative of Equation (4-13) for the two values of β (right). Blue indicates L-band and red C-band.

4.5.2 Number of observations

The performance of the retrieval of biomass with C- or L-band imagery depends on the number of backscatter observations available. In the case of BIOMASAR-L, only three L-band cross-polarized observations were available for most of the Northern hemisphere, 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 large multi-temporal stack of observations. Only for the tropics and sub-tropics could a larger stack of images be considered because of the availability of ALOS-2 ScanSAR observations. For BIOMASAR-C, instead, a larger number of Sentinel-1 images were available across most forest areas worldwide. The largest number of observations (>100) was generally available for Europe whereas for large areas in boreal Asia, most of Brazil, most of Africa, and Australia, the number of observations was rather in the range of 25 to 30 images; note that most of the acquisitions were in dual-polarization mode so that, for instance, 100 observations correspond to 200 backscatter images. Previous experiences with the retrieval of biomass using multi-temporal C-band data acquired by Envisat ASAR (Santoro et al., 2011) suggested that, ideally, dozens of observations should be used to achieve a high retrieval performance. Although the number of images required for the retrieval also depends on the ecosystem, the coverage available from Sentinel-1 might be close to the required minimum locally.

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 Equation (4-18) by the square root of the number of observations available at C- and L-band at a given pixel location:

$$w_s = \left| \frac{\partial T_L}{\partial V} \sqrt{N_L} \right| - \left| \frac{\partial T_C}{\partial V} \sqrt{N_C} \right| \quad (4-18)$$

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

$$w_{s,norm} = \frac{(w_s - \min(w_s))}{\max(w_s - \min(w_s))} \quad (4-19)$$

The rescaling is undertaken separately for each FAO ecoregion. Given the way Equations (4-18) and (4-19) are defined, w_s will be greater in higher ranges of biomass and low in lower ranges of biomass (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 $1 - w_{s,norm}$.

4.5.3 Local errors in the model calibration and inversion

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	105	14.02.2019	

Visual inspection of the biomass maps produced using BIOMASAR-L and BIOMASAR-C and comparisons with regional reference datasets (inventory, maps) at the time of the GlobBiomass project revealed local discrepancies between the maps that could not be explained by differences in the sensitivity of C- and L-band to biomass, but instead by 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. This weighting scheme can be avoided in the case of a direct AGB retrieval because of the more advanced calibration and inversion procedures.

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. 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 (4-12), assuming a two-way tree attenuation of 0.5 dB m^{-1} for L- and 2 dB m^{-1} for C-band. When plotting the simulated transmissivity at C- and L-band against the BIOMASAR-C and BIOMASAR-L maps respectively and fitting Equation (4-12) to the observed trend (i.e., by estimating β) between (a) the simulated C-band transmissivity and BIOMASAR-C GSV and (b) 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 Figure 4-25, we see, for instance, that in boreal forest, the observed trend of simulated transmissivity at C-band and the BIOMASAR-C GSV estimates (green error bars) deviates more from the fit of Equation (4-12) (green curve) than the observed trend of simulated transmissivity at L-band and the BIOMASAR-L GSV estimates (red error bars) (red curve), particularly 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 to indicate that, locally, the more consistent map should be preferred (i.e., should be given more weight in the map merging). Once GEDI LiDAR data become available, it is foreseen that height and canopy density calculated from LiDAR waveforms will be used instead of the MODIS-based data products.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	106	14.02.2019	

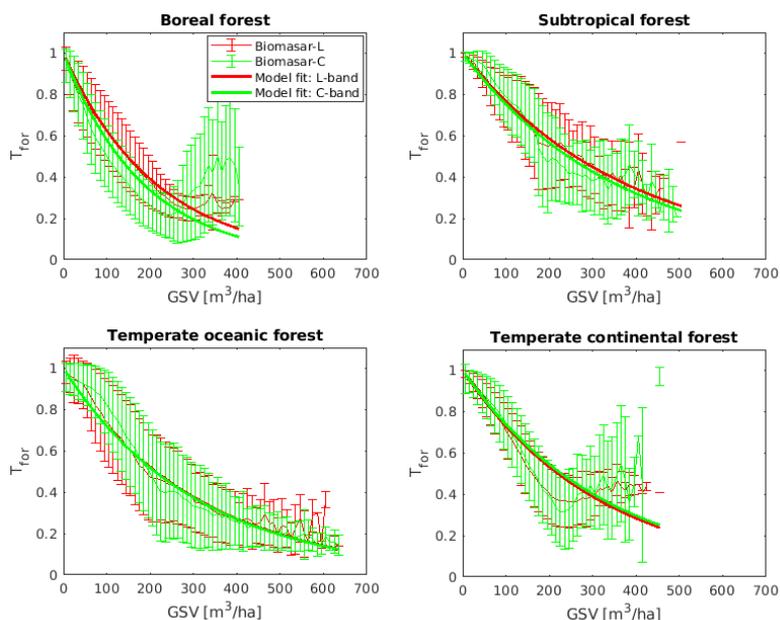


Figure 4-25: Forest transmissivity modelled as a function of canopy cover, forest height vs. BIOMASAR-C and BIOMASAR-L GSV for different FAO ecoregions in South America. The curves represent the fit of Equation (4-10) to the observations.

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:

- 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 (4-12).
- Fit Equation (4-11) 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.

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:

$$w_t = |T_{sim}(C) - T_{mod}(C)| - |T_{sim}(L) - T_{mod}(L)| \quad (4-20)$$

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

$$w_{t, norm} = (w_t - \min(w_t)) / \max((w_t - \min(w_t))) \quad (4-21)$$

Given how Equation (4-21) 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 as $1 - w_t$.

The resulting map of w_t is illustrated in Figure 4-26 for latitudes above 50° North. The map shows that for large parts of the boreal zone, the weights are ~ 0.5 . Only very locally will one of the two maps be given more weight.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	107	14.02.2019	

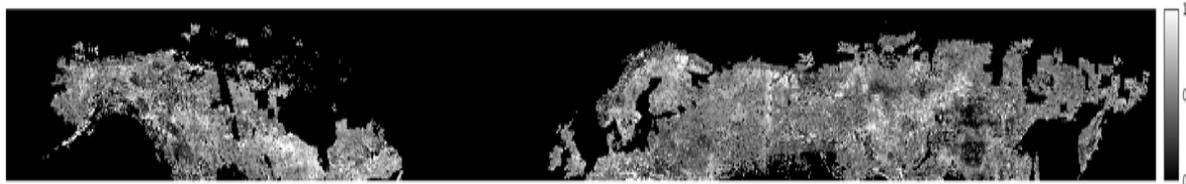


Figure 4-26: Example of weights applied to the BIOMASAR-L dataset, which reflects how consistent the GSV estimates are with the transmissivity modelled with the aid of optical canopy density and height maps.

4.5.4 Topographic effects

A drawback of working with JAXA's ALOS mosaics is that the mosaics exhibit strong topography-related distortions, geometric as well as radiometric, which could not be reduced/corrected without re-processing the original SLC data. The C-band data instead were pre-processed using state-of-the-art algorithms for geocoding as well as for compensation for topographic effects related to varying pixel scattering areas over sloped terrain (Section 3). An initial comparison of BIOMASAR-L and BIOMASAR-C biomass maps confirmed that the BIOMASAR-C maps were less affected by topography than the BIOMASAR-L map. In order to reduce the effect of topography in the merged biomass maps, an additional weight is therefore defined which progressively gives more weight to the BIOMASAR-C map as the terrain becomes steeper.

Topographic weights are defined using the local incidence angle maps 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 as:

$$A \sim 1 / \sin(\theta_i) \quad (4-22)$$

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 negligible. Weights were defined by calculating the percentage difference in pixel scattering area between flat terrain (38° incidence angle) and the pixel area estimated from the local incidence angle with:

$$\Delta_{area} = 100 \cdot |1 - \sin(38^\circ) / \sin(\theta_i)| \quad (4-23)$$

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 4-27.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	108	14.02.2019	

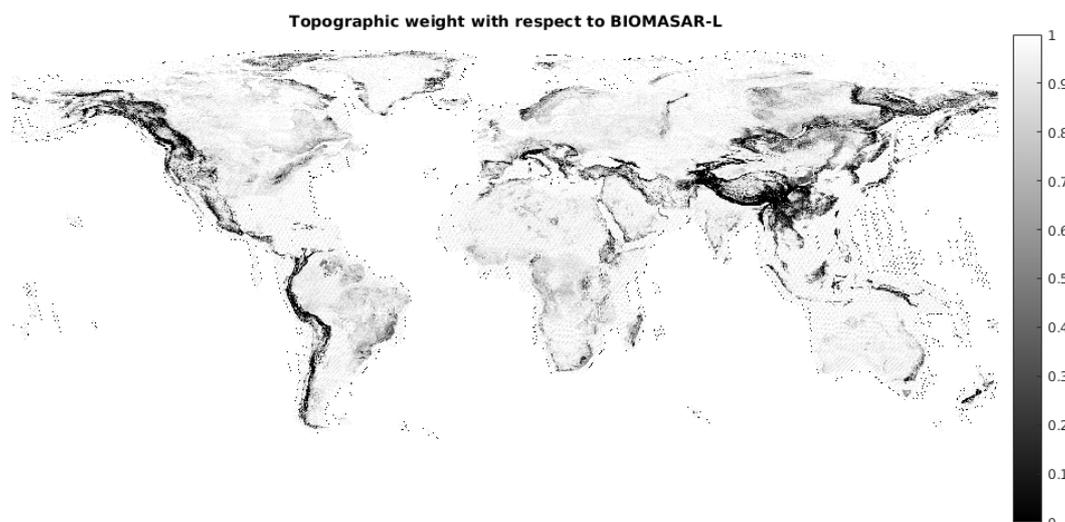


Figure 4-27: Weight for the BIOMASAR-L map reflecting the steepness of the terrain.

4.5.5 Total weight for merging

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

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

$$w(C) = 1 - w(L) \tag{4-24}$$

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

4.6 GSV to AGB conversion

Knowledge of wood density and biomass expansion factors allows conversion from GSV to AGB:

$$AGB = GSV * WD * BEF \tag{4-25}$$

In Equation (4-25), WD represents the wood density and BEF is the stem-to-total biomass expansion factor. The product between WD and BEF is also referred to as Biomass Conversion and Expansion Factor (BCEF). To compute the AGB in year 1, the data layers produced in the GlobBiomass dataset were used [RD-8].

5 Results

5.1 BIOMASAR-C

5.1.1 Comparing the performance of BIOMASAR-C in CCI CORE y1 with respect to GlobBiomass

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	109	14.02.2019	

Compared to previous experiences with BIOMASAR-C, where Envisat ASAR data with 1000 m spatial resolution were used, implementation of Sentinel-1 represented substantial novelty in many senses. The higher resolution compared to the Envisat ASAR 1000 m imagery required some adaptations on how the model parameters are computed. Furthermore, it was unclear at the beginning whether the cross-polarized channel would outperform the co-polarized channel. Finally, the moderate resolution of the Sentinel-1 data allows appreciating the compensation of the SAR backscatter after having accounted for true pixel area. In this Section, we present an assessment of such aspects.

We first compare the performance of BIOMASAR-C applied to Envisat ASAR backscatter with a pixel size of 1000 m (year 2010) and Sentinel-1 backscatter with a pixel size of 150 m. The maps in Figure 5-1 show that the spatial distribution is well reproduced in both datasets. The Sentinel-1 based estimates, however, identify more details and the range of GSV mapped is wider than the ASAR-based results. This indicates that relying on ASAR 150 m data would be more efficient, still the coverage is the main issue if the aim is to complement L-band retrievals of biomass with C-band retrievals globally.

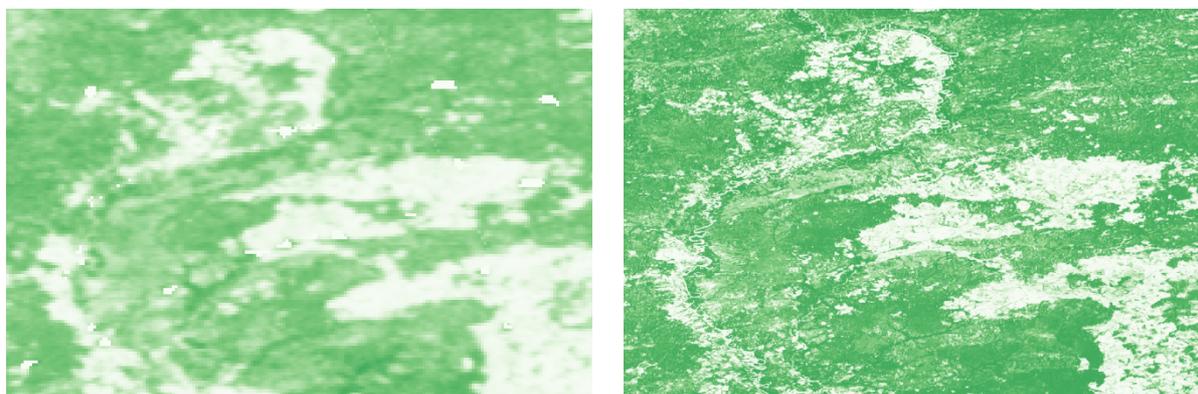


Figure 5-1: Comparing GSV estimated with BIOMASAR-C using Envisat ASAR data with a 1,000 m pixel size (left) and Sentinel-1 data with a pixel size of 150 m (right). The colour scheme of the GSV estimates is the same as in Figure 5-2. The maps cover an area between 56° and 57° N and 45° and 47° E.

The performance of BIOMASAR-C using Sentinel-1 data is confirmed by the example in Figure 5-2. The area is intensively managed boreal forest with frequent clear-cuts (rectangular shapes) and regenerating forests (light green areas). The level of detail is very high, as demonstrated by the linear features (deforested corridors), despite the moderate resolution thanks to the multi-temporal features of the Sentinel-1 dataset. Neither of the examples in Figures 5-1 and 5-2 show apparent artefacts, confirming the impression given by the Sentinel-1 colour composite in Figure 3-2 that suggested a spatially consistent representation of the SAR backscatter globally. This impression is further confirmed when analysing a map of GSV estimates for areas characterized by strong topography (Figure 5-3). Thanks to the normalization of the SAR backscatter for true pixel area, the GSV estimates on sloped terrain are uniform. It should be noted that this assessment does not include the advances in year 2, related to the compensation of incidence angle on the retrieval.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	110	14.02.2019	

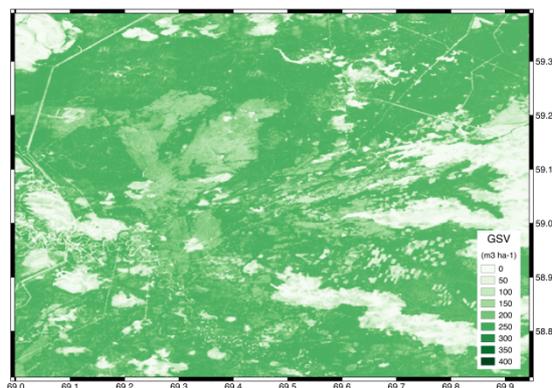


Figure 5-2: Example of GSV estimates obtained with BIOMASAR-C applied to Sentinel-1 dataset acquired in 2017.

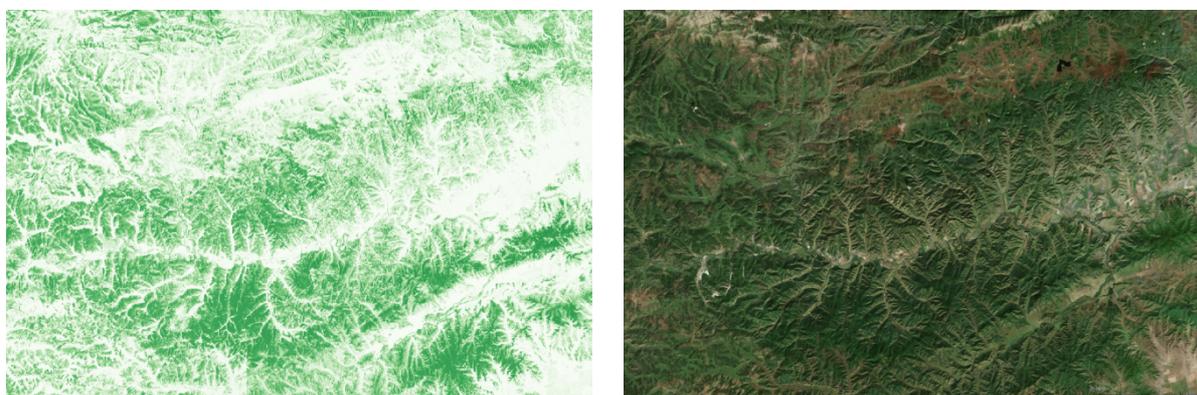


Figure 5-3: Estimates of GSV for an area south of the Baikal Lake, Siberia (left) and optical image from Google Earth (right). The colour scheme of the GSV estimates is the same as in Figure 5-2.

The GSV map from BIOMASAR-C obtained in year 1 is shown in Figure 5-4. The spatial distribution of GSV fits well with the known patterns, with the highest GSV occurring in the Pacific Northwest, followed by Central Siberia and European Russia. However, evident gaps are evident in the wet tropics (Gabon, central Amazonia) where we were not able to train the WCM because of too few ground pixels. Overall, the estimates were apparently too low in the wet tropics, similarly to the retrieval obtained in GlobBiomass which points at something structural in which the retrieval is failing beyond the weak sensitivity of C-band to biomass in these regions.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	111	14.02.2019	

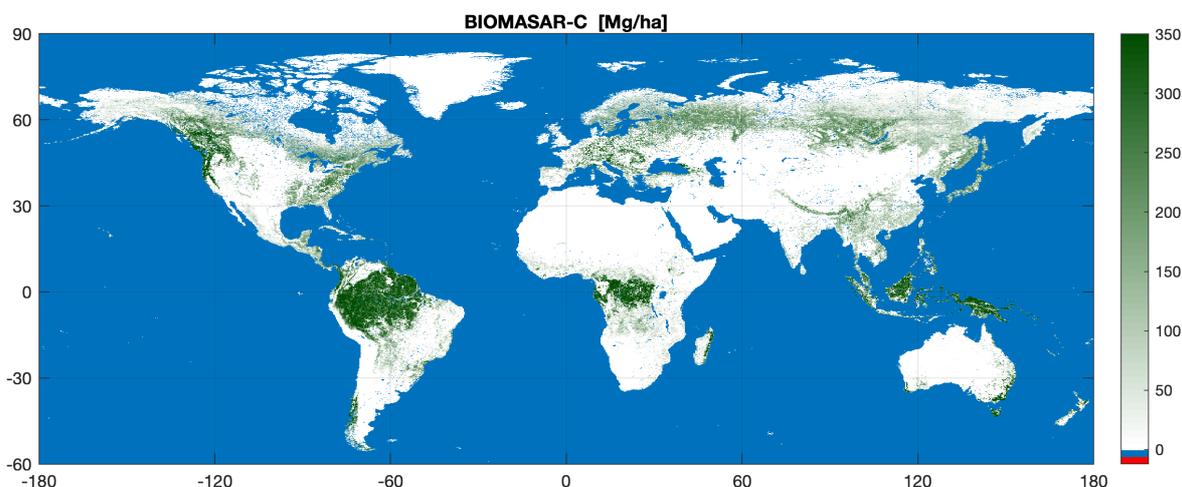


Figure 5-4: Map of GSV obtained with the BIOMASAR-C algorithm applied to the multi-temporal dataset of Sentinel-1 backscatter observations of 2017. Pixel size: 150 m.

Figure 5-5 shows a quantitative assessment of the GSV estimates for Russia in which averages per province are compared to values obtained by the Forest State Account (courtesy of Dmitry Schepaschenko, IIASA). The averages are assumed to be only marginally affected by the nearly 10-year time difference between the two datasets. Overall, the provincial averages appear to be unbiased, with some minor discrepancies of less than 10% that, however, need to be understood. This result indicates the reliability of C-band estimates in extra-tropical regions.

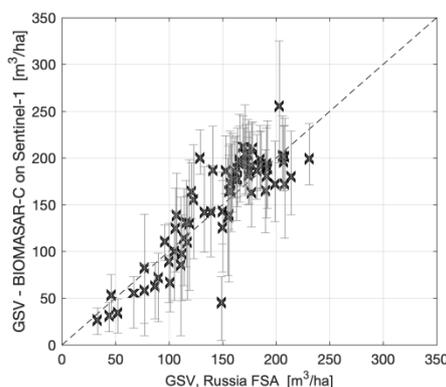


Figure 5-5: Scatterplot comparing provincial averages of GSV from the Forest State Account and from the BIOMASAR-C algorithm applied to the Sentinel-1 dataset of 2017. The vertical bars represent the standard deviation of the Sentinel-1 GSV per province.

Having assessed the quality of the GSV estimates from Sentinel-1, it is important to compare the quality of the retrieval with the results from the L-band ALOS-2 data. Figure 5-6 shows the BIOMASAR-L and BIOMASAR-C maps for an area where the ALOS-2 backscatter mosaic was characterized by JAXA's manual compositing of images from different dates. Such compositing resulted in a checkerboard pattern across the GSV map whereas the Sentinel-1 estimates did not present any offset. What is, however, also important to remark is that the ALOS-2 estimates outside of the area affected by offsets present higher GSV than the Sentinel-1 estimates. Assuming that these areas are correctly estimated, the higher GSV from ALOS-2 than from Sentinel-1 is consistent with previous results in the GlobBiomass project that biomass estimates

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	112	14.02.2019	

from L-band are usually higher than from C-band. The reason is the weaker sensitivity of C-band backscatter to biomass so that an even slightly erroneous estimates of the WCM model parameters translates into substantial biases. These biases are usually negative because of the way the model parameters in BIOMASAR-C are computed, as shown by the examples in Section 4.3.2.

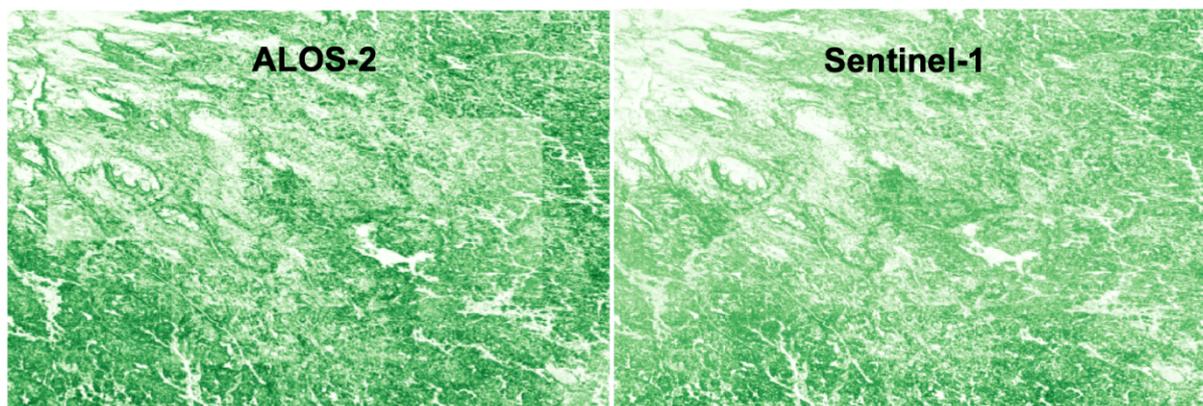


Figure 5-6: Comparing GSV estimated with BIOMASAR-L applied to ALOS-2 data (left) and with BIOMASAR-C applied to Sentinel-1 data (right). The central part of the map in the left panel shows a checkerboard pattern, corresponding to different ALOS-2 images used in the process of compositing the mosaic. In contrast, the Sentinel-1 map has uniform appearance.

5.1.2 Comparing BIOMASAR-C versions in CCI CORE algorithms

The performance of BIOMASAR-C as implemented in the CORE algorithm of year 1 and in the updated version of the CORE is illustrated by three examples characterized by different AGB levels and topography.

The first example illustrates a tile that straddles Kenya and Tanzania, dominated primarily by East African montane forests and savannas (Figure 5-7). The spatial distribution of AGB predicted by BIOMASAR-C does not show major macroscopic differences. When looking in detail, however, the map obtained with the updated algorithm (y2) shows greater contrast. This is confirmed in Figure 57, which reveals that the updated BIOMASAR-C estimates are smaller than those in the previous version in the low AGB range. This is significant because one of the issues with BIOMASAR-C was the frequent overestimation of biomass in this range. The improved behaviour is a consequence of expressing the Water Cloud Model as a function of AGB (through height, canopy density and related allometries), which has a steeper increase for increasing biomass in the low biomass range compared to the same model expressed as a function of GSV. The impact of an explicit use of topographic information in the model training is illustrated by the histograms in Figure 5-7. With BIOMASAR-C y1, the AGB estimates were systematically higher for slopes toward the East. With BIOMASAR-C y2, however, the distribution of AGB appears to be more uniform, regardless of terrain slope orientation, with somewhat higher AGB for slopes toward the SW and lower AGB for slopes toward the NE, both of which being expected.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	113	14.02.2019	

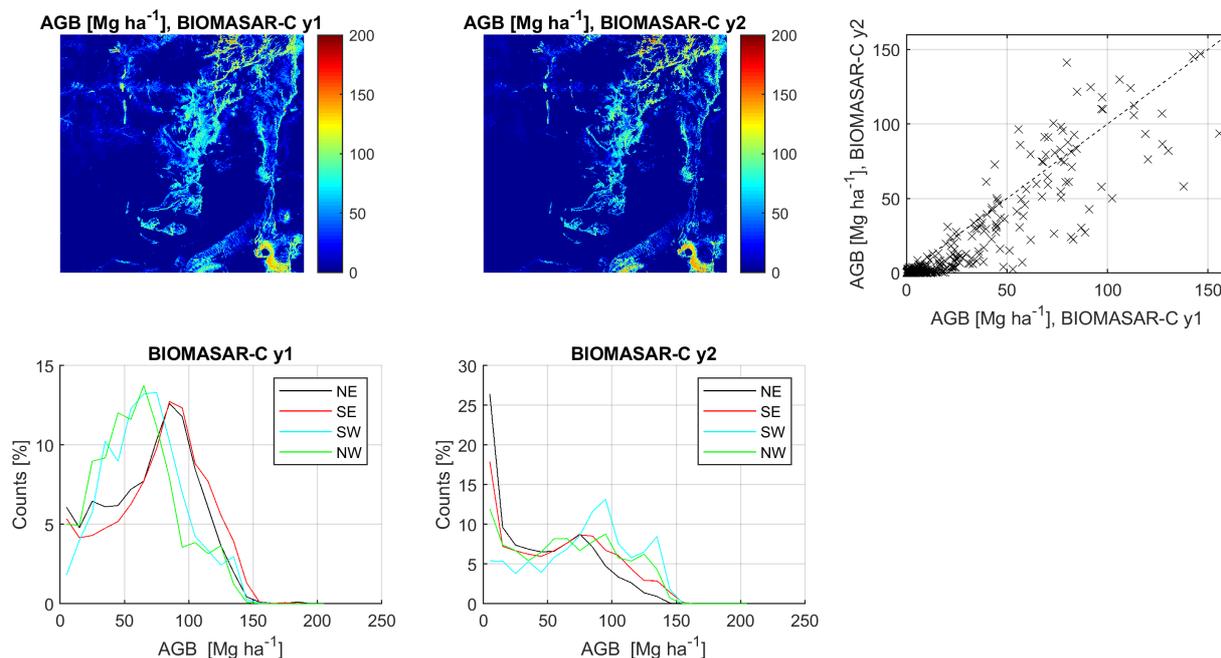


Figure 5-7: Maps of AGB obtained with the BIOMASAR-C algorithm described in Section 4.3.1 (y1) and in Section 4.3.2 (y2) from the 2017 dataset of Sentinel-1 images. The scatterplot compared AGB estimates from the two maps for a subset of the pixels. The histograms represent the distribution of AGB for (i) a given version of BIOMASAR-C and (ii) a given quadrant of terrain orientation (see legend). The histograms are limited to pixels with a terrain slope larger than 20° to better appraise the impact of topography on the AGB retrieval. Tile (top left corner coordinate): 2°S, 35°E.

The second example illustrates a tile corresponding to the borders of the Democratic Republic of Congo and Uganda, at the border between moist and dry tropical ecotones (Figure 5-8). Using the updated algorithm (y2), estimates of AGB were generally higher, as is confirmed in Figure 5-8. This is likely to be due to the effect of the more explicit use of allometry in the retrieval, which leads to an overall higher level of AGB (Figure 5-8). The impact of modelling the backscatter taking into account terrain slope is not as clear as in the previous example although, again, we observe two distinct histograms for eastward and westward slopes. This difference tends to be less clear with BIOMASAR-C y2.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	114	14.02.2019	

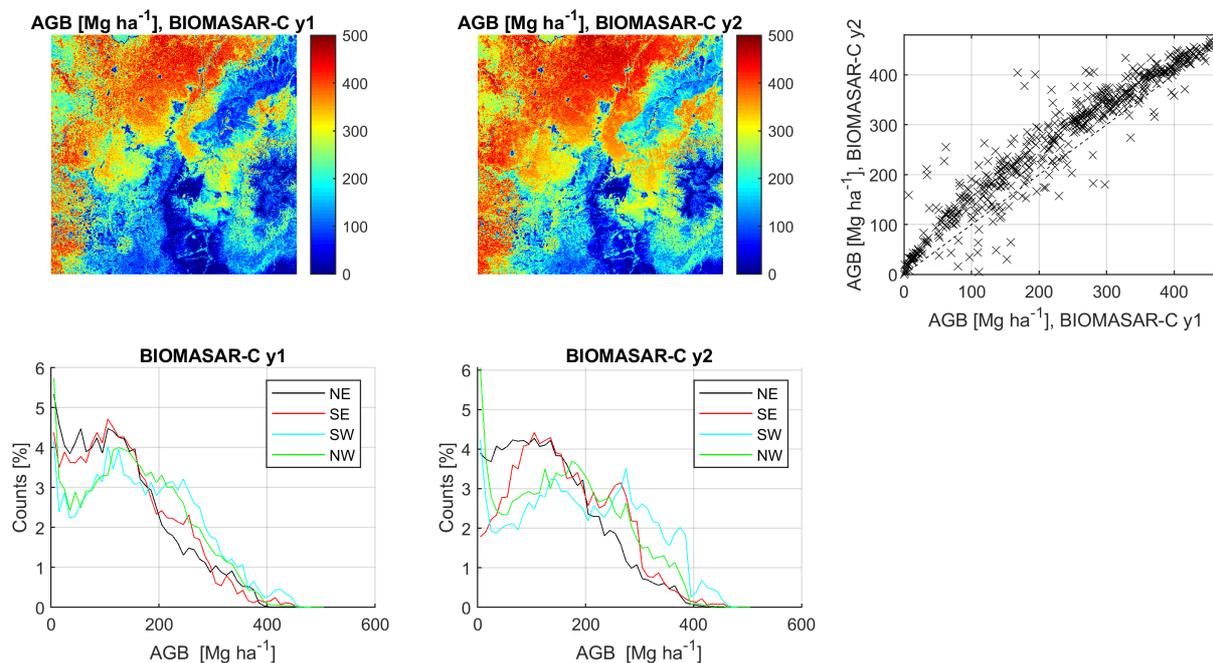


Figure 5-8: Same as in Figure 5-7. Tile (top left corner coordinate): 1°N, 29°E.

The third example illustrates a tile located in Gabon, which is an area of both tropical dry and wet forests but overall characterized by high AGB (Labriere et al., 2018) (Figure 5-9). This is possibly the most relevant example of the improvements introduced in year 2. The map produced with the previous version of BIOMASAR-C could not achieve full coverage of the tile because it was not able to generate an estimate for σ_{gr} due to an insufficient number of pixels that could be labelled as "ground". The use of regression in the updated version of BIOMASAR-C overcomes this issue. Secondly, the new map gives higher AGB, which was one of the issues with the previous version of the BIOMASAR algorithms. By expressing the Water Cloud Model as a function of height and then inserting an allometry between height and biomass, we are now able to better resolve high levels of AGB. However, a seam remains visible because the slight offset between sub-swaths of 0.2 dB had a direct impact on the retrieved AGB. The interpretation of the histograms of AGB for steep slopes is not addressed here since the major differences between maps are related to the factors described above.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	115	14.02.2019	

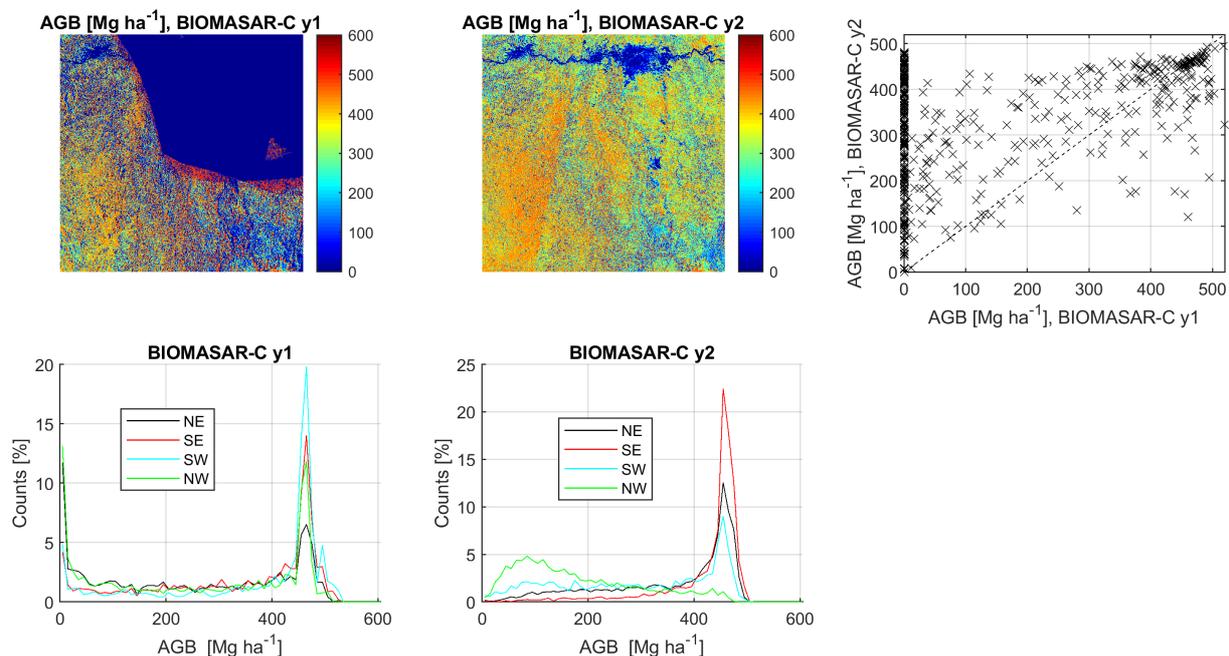


Figure 5-9: Same as in Figure 5-7. Tile (top left corner coordinate): 0°N, 11°E.

5.2 BIOMASAR-L

5.2.1 Comparing the performance of BIOMASAR-L in CCI CORE y1 with respect to GlobBiomass

In this Section, we present an assessment of the potential of the available ALOS-2 PALSAR-2 datasets for the global retrieval of biomass. In year 1, a global GSV map was produced using all available ALOS-2 FBD and ScanSAR imagery (Figure 5-10). The potential of the ALOS-2 data for global mapping of GSV was assessed by comparing the map to the BIOMASAR-L GSV map produced in the frame of the GlobBiomass project, which was based on a single ALOS PALSAR FBD backscatter mosaic for the year 2010. Given the lack of a more recent forest cover map, the Landsat canopy density product for the year 2010 (Hansen et al., 2013) was applied to mask out unvegetated terrain (canopy density of 0%) in the ALOS-2 PALSAR-2 product.

When evaluating the GSV obtained from the ALOS-2 PALSAR-2 map (examples for different regions are shown in Figure 5-11Figure 5-12), we find the map to be spatially consistent across most areas in the world. Striping effects between adjacent orbits, which are visible in the ALOS-2 PALSAR-2 mosaics in many areas in the world, are generally not visible, indicating that the BIOMASAR-L model calibration approach, which estimates model parameters adaptively in 15x15 km² large windows, is capable of capturing differences in the backscatter to biomass relationship associated with changing environmental imaging conditions.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	116	14.02.2019	

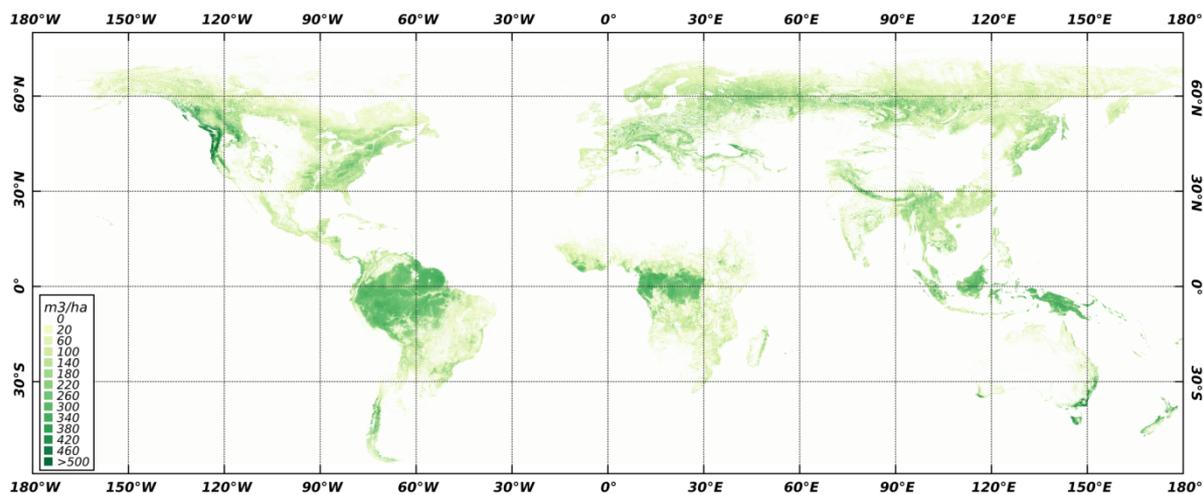


Figure 5-10: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD and ScansAR imagery acquired between 2015 and 2018.



Figure 5-11: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD imagery acquired over Central Europe between 2015 and 2017.

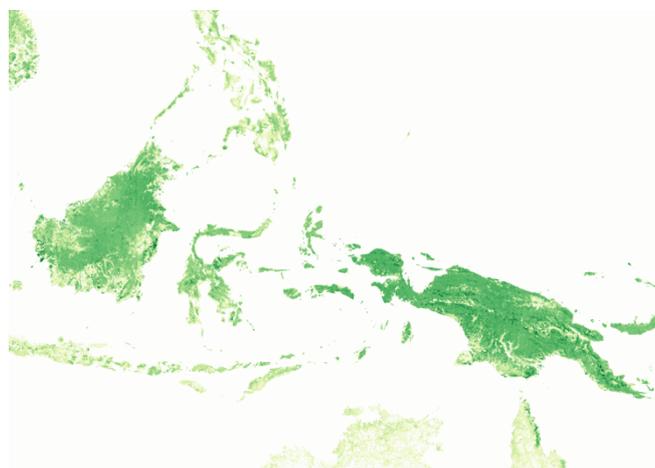


Figure 5-12: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD and ScansAR imagery acquired over Indonesia and Papua between 2015 and 2018.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	117	14.02.2019	

A clear exception is boreal forest regions where striping associated with images acquired under winter frozen conditions is visible when all images in the FBD mosaics were acquired in winter. In areas where only one of the three mosaics included imagery from winter, the weighted multi-temporal combination successfully eliminated the impact of winter imagery on the final retrieval. An example is shown in Figure 5-13 for Western Siberia where all three FD mosaics included winter acquisitions. The corresponding subset of the ALOS-2 PALSAR-2 mosaic for the year 2017 is also shown. The striping is not as pronounced as in the backscatter mosaics because the adaptive model calibration approach is generally capable of capturing the backscatter changes due to frozen conditions. Figure 5-13 (bottom) shows a map of the model parameter σ_{df}^0 , which describes the backscatter from closed forest canopies for an area in Siberia. It can clearly be seen that distinct differences in the backscatter statistics associated with frozen/unfrozen conditions are reflected in the model parameter estimates. Nonetheless, residual artefacts persist, particularly towards the edges of such orbits.

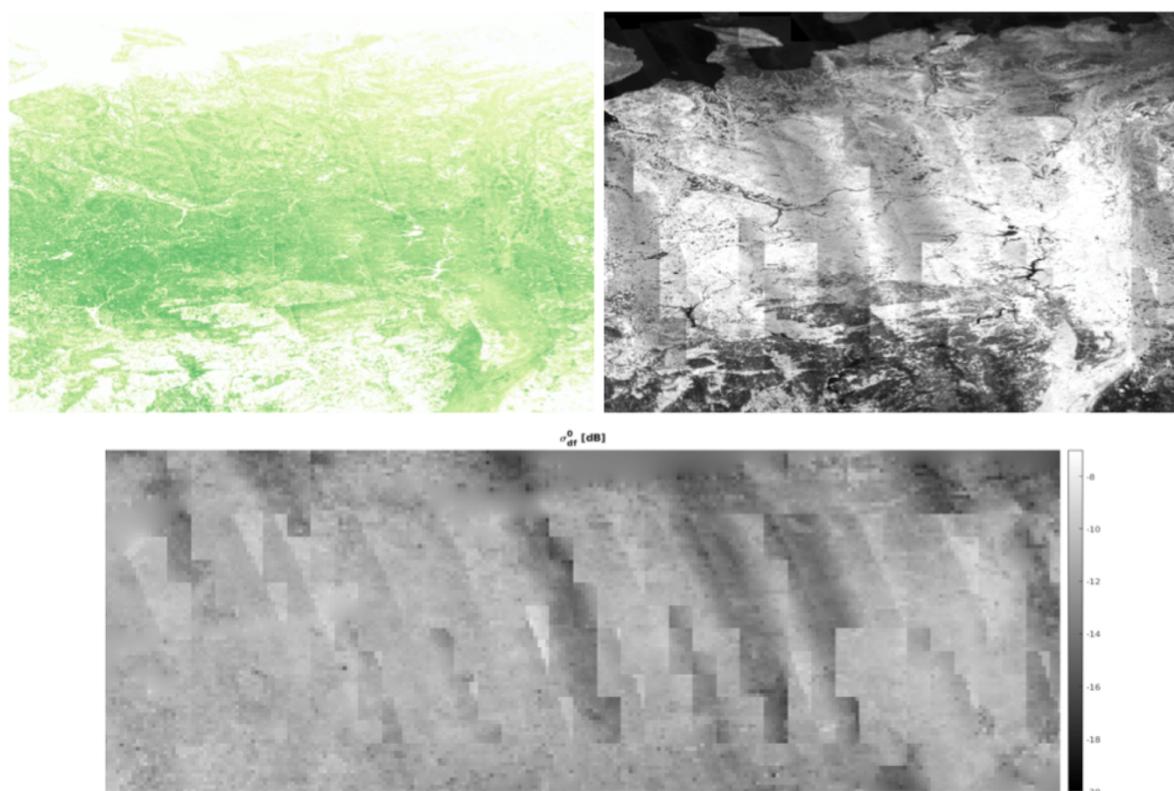


Figure 5-13: BIOMASAR-L GSV map with 100 m pixel size derived from ALOS-2 FBD imagery acquired over Western Siberia between 2015 and 2017 (top left) and the corresponding subset of the ALOS-2 FBD mosaic for the year 2017 (top right). The bottom plot illustrates the estimates for the model parameter σ_{df}^0 across Siberia for the year 2017.

A comparison of the GSV estimates in the ALOS-2 PALSAR-2 GSV map and the ALOS PALSAR GSV 2010 map (Figure 514) indicated that, despite local inconsistencies in the ALOS-2 PALSAR-2 data and hence the derived maps, overall these are consistent. Aspects which require further investigation, however, are the tendency in the new GSV maps to estimate somewhat lower GSVs in high GSV ranges, topographic effects, and specific forest types such as flooded forests and mangroves for which the BIOMASAR type of modelling and model calibration approach have been found to underestimate GSV systematically. In a few areas, such as Indonesia (Figure 5-14,

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	118	14.02.2019	

bottom left), the comparison of the GlobBiomass and the new GSV maps suggests more pronounced differences between the maps, the reason for which is not clear.

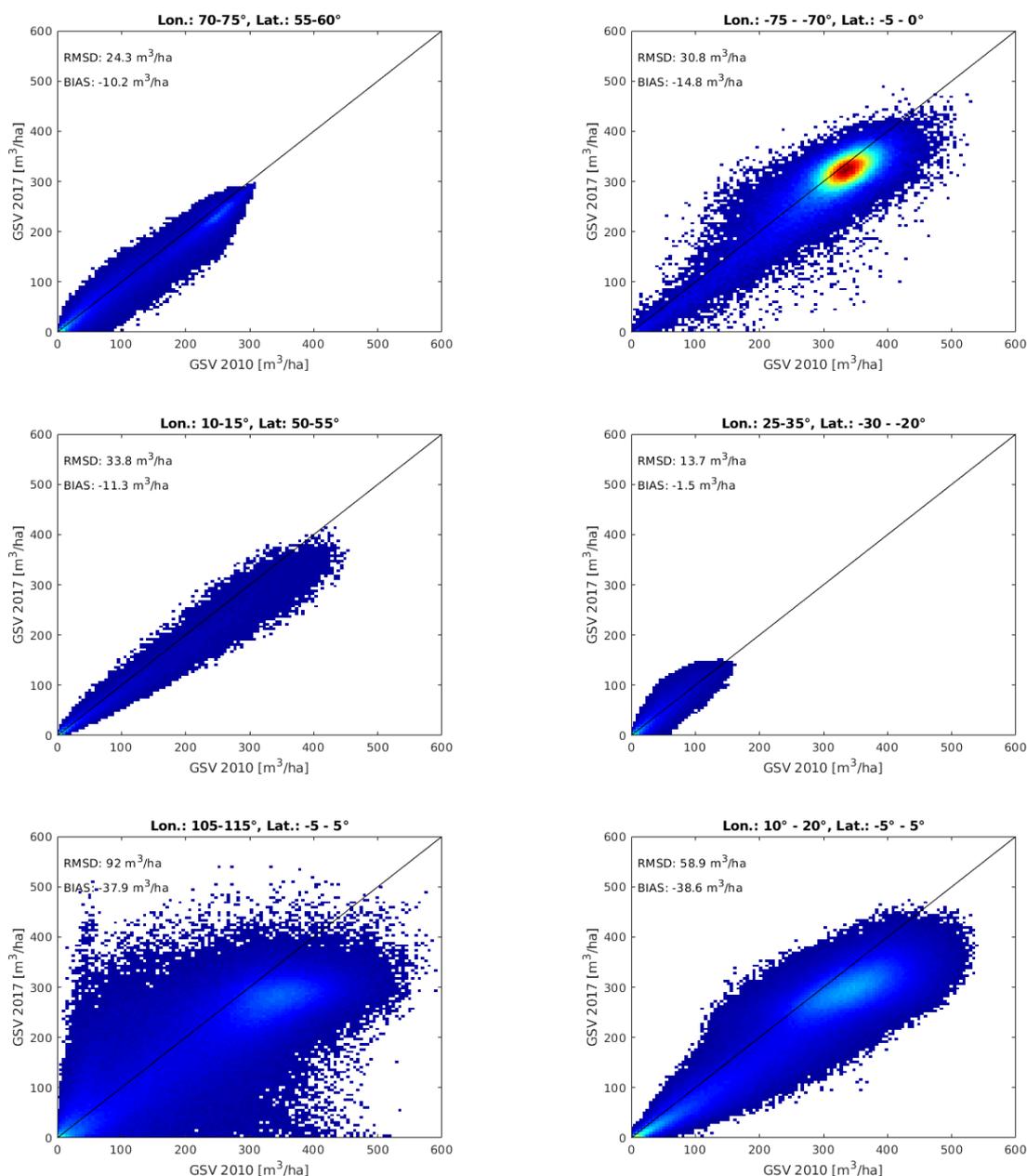


Figure 5-14: Scatterplots of GSV produced from multi-temporal ALOS-2 data and the BIOMASAR-L map for the year 2010 in the frame of the GlobBiomass project. For the comparison, all maps were aggregated to 1 km pixel scale. The examples refer to 5°x5° or 10°x10° large areas in Siberia (top left), Amazon (top right), Central Europe (centre left), South Africa (centre right), Indonesia (bottom left), and Congo Basin (bottom right).

5.2.2 Comparing BIOMASAR-L versions in CCI CORE algorithms

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	119	14.02.2019	

Differences in the performance of the retrieval of AGB with the updated BIOMASAR-L approach and the CORE algorithms implemented in year 1 are illustrated for three different regions in Figure 5-15 and Figure 5-16.

In the first, a 10°x10° large area (5-15° E, 46-56° N) ranging from Eastern France in the West to Poland in the East and from the Alps in the South to Denmark and Southern Sweden in the North was considered. About 30% of the land surface considered is covered by mixed broadleaf and coniferous temperate forests. Here, the BIOMASAR-L y1 and y2 maps exhibit an overall similar distribution of AGB (Figure 5-15, top row) and agree well, particularly in AGB ranges below 100 t/ha (Figure 5-16, top left) but differences are more pronounced in high AGB ranges. The scatterplot reveals no systematic bias between the two maps. However, the spread of the estimates is large. A visual comparison of the maps in high AGB regions showed that differences are most pronounced in areas of steep terrain, such as the Alps or the secondary mountain ranges in Southern Germany. The maps produced with BIOMASAR-L y1 present strong topography related artefacts with large difference in AGB estimates for slopes facing and those tilted away from the radar. In the BIOMASAR-L y2 map, such differences in areas of steep terrain are clearly reduced and the local distribution of AGB is more uniform. The results therefore suggest that model calibration considering the dependence of the backscatter-AGB relationship on local incidence angle improved the AGB retrieval with the L-band data.

Similar observations can be made when comparing the AGB maps in a 5°x5° large area (118-123° W, 45-50° N) in the Pacific Northwest of the United States and Southern British Columbia Canada (Figure 5-15, centre row). The temperate forests in the Western part of the region from the Pacific coast to the Cascade Mountain range present some of the highest AGB densities worldwide. Further East, where vegetation is dominated by dry forests and steppe vegetation, the AGB decreases. While the general spatial distribution of AGB is captured similarly in both maps, major differences can be observed for dense forests in the Cascade Mountain Range with AGBs beyond 100 Mg ha⁻¹ (Figure 5-16, bottom left). In this area, the new AGB map reports significantly higher AGB values and presents a more uniform spatial distribution with reduced topographic effects. These observations suggest that use of the new WCM model, in which differences in the interrelationships between canopy density, height and AGB are accounted for, as well as the model calibration considering incidence angles, allow improved AGB mapping in high AGB forest regions.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	120	14.02.2019	

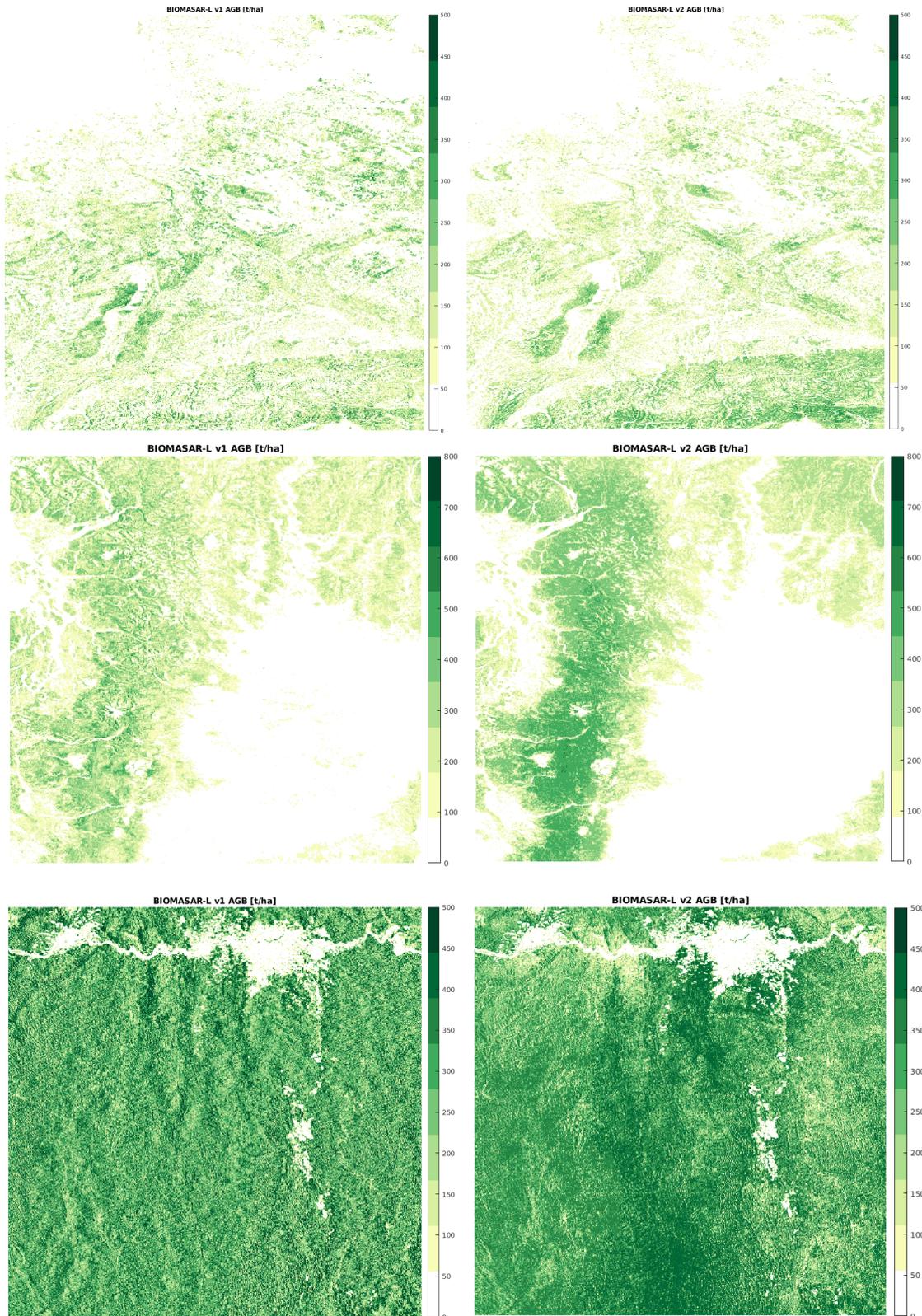


Figure 5-15: AGB maps for areas in Central Europe (top row), the Pacific Northwest of the United States (centre row), and Gabon (bottom row) with BIOMASAR-L (y1) i.e., the CORE algorithm deployed in year 1 (left), and with BIOMASAR-L (y2) (right).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	121	14.02.2019	

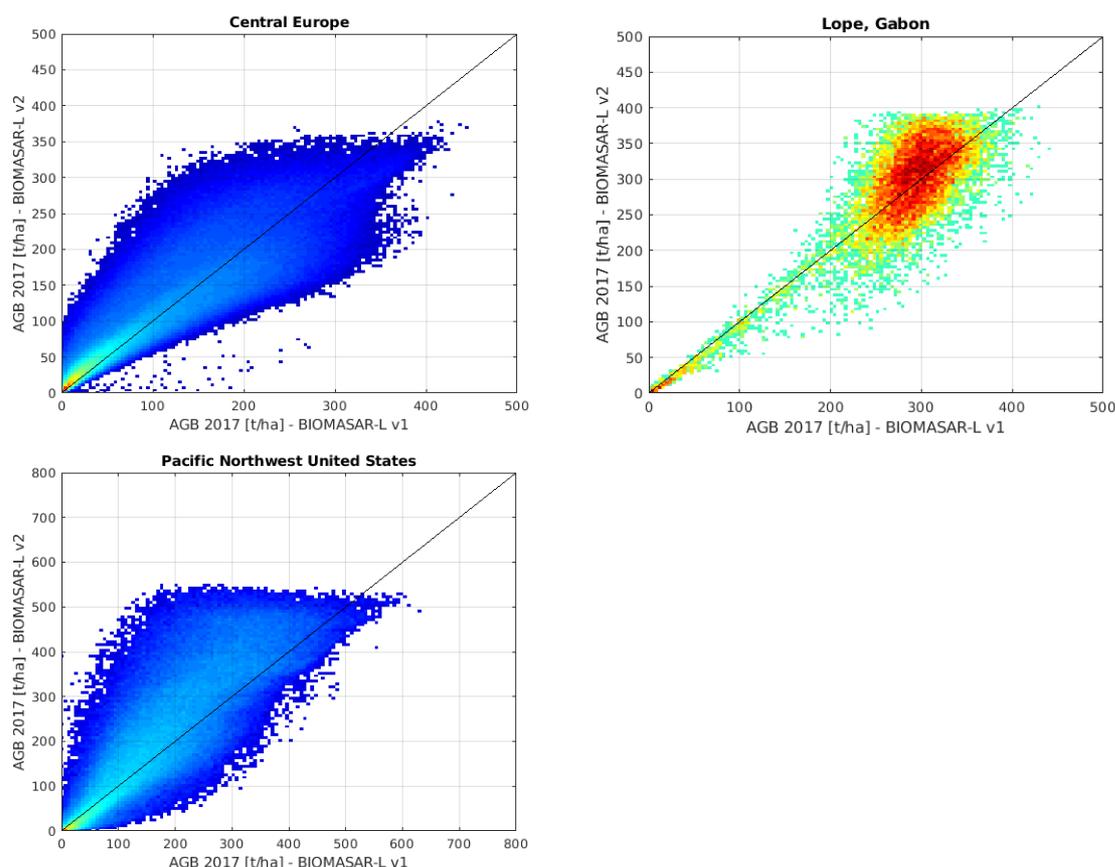


Figure 5-16: Comparison of AGB maps for areas in the Pacific Northwest of the United States, Central Europe, and Gabon produced with BIOMASAR-L (y1), i.e., the CORE algorithm deployed in year 1, BIOMASAR-L (y2).

The third example shown in Figure 5-15 (bottom row) is a tropical forest site in Lope, Central Gabon (11-12° E, 0-1° S). The area comprises a mosaic of savannah and colonizing forest with low AGB and high AGB *Okoumé* and *Marantaceae* forest. The two maps agree well for AGB ranges below 200 Mg ha⁻¹ (Figure 5-16, top right), but are poorly correlated for AGB above 200 Mg ha⁻¹. Accordingly, AGB patterns over closed forest canopies are different. In part, this may be associated with the different modelling approach. The handling of incidence angle effects in the production of the year 1 map was also different. A benefit of the modelling with the new WCM may be seen is that the new map reports somewhat higher AGBs over closed canopies than the year 1 map, for which the validation suggested underestimation of AGB. The topography in the area is moderate which is why improvements from the model calibration per local incidence class interval are minor.

5.3 Merging GSV estimates

The merging approach, which had already been implemented for generation of the GlobBiomass product, confirm that the approach is generally suitable for improving the estimates of biomass derived from ALOS-2 and Sentinel-1 as well. Figure 5-17 illustrates the improved agreement of the year 2017 GSV maps produced from Sentinel-1 and ALOS-2 data with the GlobBiomass GSV product for 2010. For additional background to merging results, refer to the ATBD of the GlobBiomass project. At the time of compiling this document for year 2, we have not tested the

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	122	14.02.2019	

performance of merging with the updated CORE algorithm. It is, however believed, that the performance does not differ compared to results presented in this Section.

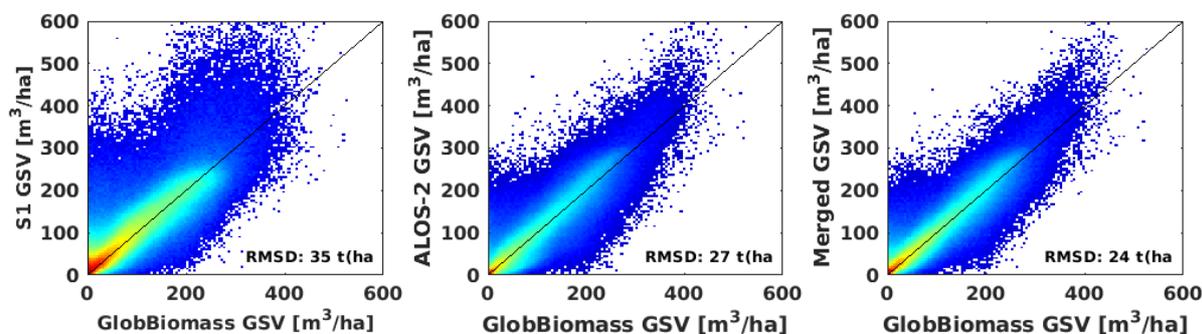


Figure 5-17: Scatterplots of Sentinel-1 (left), ALOS-2 (centre), and the merged Sentinel-1/ALOS-2 GSV maps against the GlobBiomass GSV map for 2010.

6 Status of the CORE algorithm and potential areas for improvement

The most important aspect when setting up the CORE retrieval algorithm was that it should perform similarly to the GlobBiomass algorithm while taking into account that the algorithm ingests new datasets (Sentinel-1, ALOS-2 PALSAR-2) with different coverage, spatial and temporal resolution, and caveats compared to ALOS PALSAR and Envisat ASAR. The refinement of the CORE algorithm in the second year of the CCI Biomass project was focused on better representing vertical properties of the canopy in the retrieval models as well as tackling one of the major issues in radar remote sensing of forests, namely the impact of topography on the retrieval.

Strengths and weaknesses of the current implementation of the CORE algorithm have been reported in this document. The retrieval of AGB with the algorithm developed in year 1 appeared to perform similarly to the results obtained with the GlobBiomass algorithm, although most comparisons reported here are more qualitative than quantitative (for the latter refer to the Product Validation Report [RD-9]). Nonetheless, there are indications on how the CORE algorithm should develop beyond adapting the GlobBiomass approach to the current suite of EO data suitable for biomass estimation. In year 2, the algorithms were further developed on the basis of the GlobBiomass algorithm but we revised the way AGB is estimated by taking into account two very promising fields of investigation, namely the relationship between LiDAR metrics on one hand and a LiDAR metric and AGB on the other. These, in turn, were inserted in the CORE algorithm in the form of allometric equations. In this way, we bypass the estimation of GSV via BCEF, which has not developed in recent years as fast as the LiDAR-based research. Also, we characterized the impact of topography on the retrieval by using experimental relationships between topographic index (incidence angle) and backscatter rather than developing models that would have probably failed due to the subtle difference in backscatter as landscape and topography change.

The retrieval of biomass implemented in y1 is rather conservative (i.e., misses the extreme values of AGB, because of the loss of sensitivity of the major biomass predictors in high biomass forest). Having explicit tree height information is extremely relevant and has been implemented in year 2 with interesting preliminary results. The unavailability of GEDI and ICESAT-2 datasets

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	123	14.02.2019	

at the time of writing implied that the allometries derived so far and based on the sparser sampling of ICESAT GLAS could only be characterized at the level of ecoregions. With the denser coverage of GEDI and ICESAT-2, the allometry between canopy density and tree height will be characterized globally. In addition, we foresee the use of more specific allometries linking height (from LiDAR measurements) and AGB thanks to the increased attention being paid to laser observations as predictors for biomass in different regions of the world.

The retrieval of biomass is still based on simplifying assumptions that cause the retrieval models to be too general to capture the spatial variability of the relationship between observables and vegetation properties. Vegetation structural information, as developed in the Data Access Requirement Document [RD-3], should provide the backbone for a more targeted estimation of model parameters. Along the same line, knowledge gathered by investigating the relationship between EO observables and biomass in specific forest classes should be exploited. Unfortunately, most EO-based datasets that could complement a retrieval do not have a full error characterization so that the impact of a direct implementation in our retrieval schemes may not be controllable. As for alternative approaches to retrieve AGB from the set of observations currently available from spaceborne sensors, we have not identified any ground-breaking approaches that may improve our retrievals while fulfilling at the same time the requirements in terms of spatial resolution and temporal coverage of CCI biomass maps.

Alternative solutions to estimating AGB, which would overcome the systematic over- and underestimation issues of the CORE algorithm, have been explored in year 2. Still, it is likely that including a wider spectrum of observations and algorithms in the CORE would further improve AGB estimation. In this sense, we are particularly interested in the potential offered by coarse resolution observations from radiometers and scatterometers. Although they do not match the requirement on spatial resolution of the CCI biomass maps, these multi-decadal sets of observations and the annual maps derived from these in recent times (VOD-based e.g. by Fan et al., (2019), or scatterometer-based, e.g., Santoro et al., (2020)) may serve as guidelines in establishing rules to guarantee the temporal consistency of AGB estimates. This issue has not been discussed in this version of the ATBD since we are still lacking the 2018 dataset and discussions are ongoing whether the mosaics of 2010 may be replaced with individual strips given the commitment by JAXA to release these at the time of writing.

Moving from a GSV-centric to an AGB-centric retrieval implies that the BCEF is no longer a crucial variable in the process of biomass estimation. However, only once global maps of AGB with both methods here presented are compared will it be possible to understand whether efforts should be dedicated to characterization of wood density and expansion factors beyond the results obtained in the GlobBiomass project. It should, however, be noted that the GSV data product should not be discarded from the portfolio of CCI Biomass data products because for regions of the northern hemisphere, GSV represents the major variable used in the process of forest resources assessment. Being relevant to GSV, BCEF should continue to receive significant attention.

Finally, regardless of the procedures here developed to estimate biomass, the accuracy of the retrieval strongly depends on the quality of the EO data used as predictors. We have identified a number of systematic issues in the SAR data that hinder obtaining the highest quality of results. It is likely that being able to pre-process the EO data would allow attainment of such quality. In this sense, continual exchange with data providers is needed.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	124	14.02.2019	

7 References

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 (7), 1540-1550.

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., Mascaró, J., Galindo García, G.A., Chadwick, K.D., Navarrete Encinales, D.A., Paez-Acosta, G., Cabrera Montenegro, E., Kennedy-Bowdoin, T., Duque, Á., Balaji, A., Von Hildebrand, P., Maatoug, L., Phillips Bernal, J.F., Yepes Quintero, A.P., Knapp, D.E., García Dávila, M.C., Jacobson and J., Ordóñez, M.F., (2012). High-resolution mapping of forest carbon stocks in the Colombian Amazon. *Biogeosciences* 9, 2683–2696.

Asner, G.P. and Mascaró, J., (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment* 140, 614–624.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Asthon, 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. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology*, 22 (4), 1406-1420.

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.

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L. and Asner, G. (2018). An above-ground biomass map of African savannahs and woodlands at 25 meters resolution derived from ALOS PALSAR. *Remote Sensing of Environment*, 206, 156-173.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.

Cartus, O., Santoro, M. and Kelldorfer, J. (2012). Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sensing of Environment*, 124, 466-478.

Cartus, O., Santoro, M., Wegmuller, U. and Rommen, B. (2017). Estimating total aboveground, stem and branch biomass using multi-frequency SAR. In: *9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*. City: IEEE Publications, Piscataway, NJ

Castel, T., Beaudoin, A., Stach, N., Stussi, N., Le Toan, T. and Durand, P. (2001). Sensitivity of space-borne SAR data to forest parameters over sloping terrain. Theory and experiment. *International Journal of Remote Sensing* 22, 2351–2376.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	125	14.02.2019	

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B. and Yamakura, T., (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145, 87–99.

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Phillips, O.L., Phua, M.-H. and Qie, L., (2017). Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sensing of Environment* 194, 77–88.

de Ferranti, J. (2009), <http://www.viewfinderpanoramas.org>.

De Grandi, G. D., Bouvet, A., Lucas, R., Shimada, M., Monaco, S. and Rosenquist, Å. (2011). The K&C PALSAR mosaic of the African continent: Processing issues and first thematic results. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (19), 3593-3610.

DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D. H., Kelly, M. and Townshend, J. R. G. (2015). MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.

Desnos, Y.L., Buck, C., Guijarro, J., Suchail, J.L., Torres, R. and Attema, E., (2000). ASAR - Envisat's Advanced Synthetic Aperture Radar - Building on ERS achievements towards future earth watch missions. *ESA Bulletin* 91–100.

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.

Fan, L., Wigneron, J.-P., Ciais, P., Chave, J., Brandt, M., Fensholt, R., Saatchi, S.S., Bastos, A., Al-Yaari, A., Hufkens, K., Qin, Y., Xiao, X., Chen, C., Myneni, R.B., Fernandez-Moran, R., Mialon, A., Rodriguez-Fernandez, N.J., Kerr, Y., Tian, F. and Peñuelas, J. (2019). Satellite-observed pantropical carbon dynamics. *Nat. Plants* 5, 944–951.

Fick, S. E. and Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37 (12), 4302-4315.

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 (1), 48-52.

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.

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., Minster, J. B. and Blair, J. B. (2000). Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38 (4), 1989-1996.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	126	14.02.2019	

Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J. and Guo, Q. (2016). Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery and forest inventory data. *Remote Sensing*, 8 (565), doi:10.3390/rs8070565.

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 (4), 767-784.

Kellndorfer, J. M., Dobson, M. C., Vona, J. D. and Clutter, M. (2003). Toward precision forestry: Plot-level parameter retrieval for slash pine plantations with JPL AIRSAR. *IEEE Transactions on Geoscience and Remote Sensing*, 41 (7), 1571-1582.

Kellndorfer, J., Walker, W., LaPoint, E., Bishop, J., Cormier, T., Fiske, G., Hoppus, M., Kirsch, K. and Westfall, J. (2013). NACP Aboveground Biomass and Carbon Baseline Data (NBCD 2000) Woods Hole Research Centre. 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>

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 (1), 198-205.

Labriere, N., Tao, S., Chave, J., Scipal, K., Toan, T.L., Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P., Casal, T., Davies, S.J., Ferraz, A., Herault, B., Jaouen, G., Jeffery, K.J., Kenfack, D., Korte, L., Lewis, S.L., Malhi, Y., Memiaghe, H.R., Poulsen, J.R., Rejou-Mechain, M., Villard, L., Vincent, G., White, L.J.T. and Saatchi, S. (2018). In Situ Reference Datasets From the TropiSAR and AfriSAR Campaigns in Support of Upcoming Spaceborne Biomass Missions. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 11, 3617-3627.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O. and de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters* 32, 1-4.

Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, 37, L15401.

Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M., Canadell, J. G., McCabe, M. F., Evans, J. P. and Wang, G. (2015). Recent reversal in loss of global terrestrial biomass. *Nature Climate Change*, 5, 470-474.

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.

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, 1.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	127	14.02.2019	

Oliver, C. and Quegan, S. (1998) Understanding Synthetic Aperture Radar Images, Boston, Artech House, 479.

Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'Amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, Y. Kura, J.F. Lamoreux, W.W. Wettengel, P. Hedao, and K.R. Kassem. Terrestrial Ecoregions of the World: A New Map of Life on Earth, *BioScience* 51:933-938.

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 (8), 1720-1730.

Pulliainen, J. T., Heiska, K., Hyypä, 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 (5), 1041-1050.

Pulliainen, J. T., Kurvonen, L. and Hallikainen, M. T. (1999). Multitemporal behavior of L- and C-band SAR observations of boreal forests. *IEEE Transactions on Geoscience and Remote Sensing*, 37 (2), 927-937.

Reese, H., Nilsson, M., Sandström, P. and Olsson, H. (2002). Applications using estimates of forest parameters derived from satellite and forest inventory data. *Computers and Electronics in Agriculture*, 37 (1-3), 37-55.

Rizzoli, P., Martone, M., Gonzalez, C., Wecklich, C., Borla Tridon, D., Bräutigam, B., Bachmann, M., Schulze, D., Fritz, T., Huber, M., Wessel, B., Krieger, G., Zink, M. and Moreira, A. (2017). Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS Journal of Photogrammetry and Remote Sensing* 132, 119–139.

Rodríguez-Fernández, N.J., Mialon, A., Mermoz, S., Bouvet, A., Richaume, P., Al Bitar, A., Al-Yaari, A., Brandt, M., Kaminski, T., Le Toan, T., Kerr, Y.H. and Wigneron, J.-P. (2018), The high sensitivity of SMOS L-Band vegetation optical depth to biomass. *Biogeosciences Discussions* 1–20.

Rodríguez-Veiga, P., Wheeler, J., Louis, V., Tansey, K. and Balzter, H. (2017). Quantifying Forest Biomass Carbon Stocks From Space. *Current Forestry Reports* 3, 1–18.

Rodríguez-Veiga, P., Quegan, S., Carreiras, J., Persson, H. J., Fransson, J. E. S., Hoscilo, A., Ziolkowski, D., Sterenczak, K., Lohberger, S., Stängel, M., Berninger, A., Siegert, F., Avitabile, V., Herold, M., Mermoz, S., Bouvet, A., Le Toan, T., Carvalhais, N., Santoro, M., Cartus, O., Rauste, Y., Mathieu, R., Asner, G. P., Thiel, C., Pathe, C., Schmullius, C., Seifert, F. M., Tansey, K. and Balzter, H. (2019). Forest biomass retrieval approaches from earth observation in different biomes. *International Journal of Applied Earth Observation and Geoinformation*, 77 (53-68).

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 (24), 9899-9904.

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 (1), 19-35.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	128	14.02.2019	

Santoro, M., Eriksson, L., Askne, J. and Schmullius, 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 (16), 3425-3454.

Santoro, M., Shvidenko, A., McCallum, I., Askne, J. and Schmullius, C. (2007), Properties of ERS-1/2 coherence in the Siberian boreal forest and implications for stem volume retrieval. *Remote Sensing of Environment*, 106, 154-172.

Santoro, M., Beer, C., Cartus, O., Schmullius, 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 (2), 490-507.

Santoro, M., Cartus, O., Fransson, J. E. S., Shvidenko, A., McCallum, I., Hall, R. J., Beaudoin, A., Beer, C. and Schmullius, 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 (9), 4503-4532.

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 (4), 4290-4317.

Santoro, M., Wegmüller, 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. <https://doi.org/10.1016/j.rse.2015.10.031>

Santoro, M., Cartus, O., Wiesmann, A., Wegmüller, U., Kelldorfer, J., Schmullius, C., Defourny, P., Arino, O., Engdahl, M. and Seifert, F. M. (2017). Exploitation of Envisat ASAR and Sentinel-1 SAR data in support of carbon and water cycle studies. In: *Big Data from Space (BiDS'17)*. City: European Commission, Joint Research Centre, Ispra, Italy, 157-160.

Santoro, M. et al. (2018a): GlobBiomass - global datasets of forest biomass. PANGAEA, <https://doi.org/10.1594/PANGAEA.894711>

Santoro, M., Wegmüller, U. and Askne, J. (2018b). Forest stem volume estimation using C-band interferometric SAR coherence data of the ERS-1 mission 3-days repeat-interval phase. *Remote Sensing of Environment*, 216, 684-696.

Santoro, M., Cartus, O., Carvalhais, N., Besnard, S. and Fan, N. (2020), Forest above-ground biomass estimates across three decades from spaceborne scatterometer observations, EGU General Assembly.

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.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	129	14.02.2019	

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 (4), 657-671.

Shimada, M. and Ohtaki, T. (2010). Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3 (4), 637-656.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, 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.

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. FAO, Rome.

Small, D. (2011). Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (8), 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.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Navas Traver, I., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, C., Croci, R., Pietropaolo, A., Huchler, M. and Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9-24.

Ulander, L. M. H. (1996). Radiometric slope correction of synthetic-aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 34 (5), 1115-1122.

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.

Wegmüller, U. (1999). Automated terrain corrected SAR geocoding. In: *IGARSS'99*. City: IEEE Publications, Piscataway, NJ, 1712-1714.

Wegmüller, U., Werner, C., Strozzi, T. and Wiesmann, A. (2002) In *Analysis of Multi-temporal remote sensing images*, Vol. 2 (Eds, Bruzzone and Smits) World Scientific, Trento, 13-14 September 2001, pp. 37-49.

Wegmüller, U., Werner, C., Strozzi, T., Wiesmann, A., Frey, O. and Santoro, M. (2016). Sentinel-1 support in the GAMMA software. *Procedia Computer Science*, 100, 1305-1312.

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, in: *4th International*

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	130	14.02.2019	

Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	131	14.02.2019	

8 Annex A

To assess the dependency of SAR backscatter on local orientation of terrain, observations at C-band (Sentinel-1) and L-band (ALOS PALSAR) were stratified in terms of canopy density and aspect angle or incidence angle. In addition, we analysed the same sets of observations as a function of canopy density, stratifying in terms of local incidence angle. Here, we show two examples. The first example corresponds to a $1^\circ \times 1^\circ$ large tile characterized by high biomass forest, pasture, agriculture and mountainous terrain with steep slopes between Austria and Slovenia. This is representative for the behavior of the backscatter in case of strong topography. The second corresponds to a tropical landscape in Gabon, including both dry and wet forest types, with hilly terrain. This example is representative for the behavior of the backscatter in case of moderate topography. In both examples, we use the Hansen percent tree cover data product (Hansen et al., 2013) as reference for the canopy density.

8.1 A.1 Alpine terrain, temperate forests

Figure A1 shows the relationship between Sentinel-1 SAR backscatter (VV- and VH-polarization) and terrain aspect angle for the $1^\circ \times 1^\circ$ tile between Austria and Slovenia and characterized by local terrain slopes well above 40° . Observations are stratified by canopy density to better understand scattering patterns. The backscatter presents a quasi-sinusoidal pattern, being more accentuated in the case of sparsely vegetated terrain. The results are in line with other experimental results shown in Hoekman and Reiche, 2015. The highest backscatter was obtained for slopes facing the radar. A slightly different pattern was obtained for the ALOS SAR backscatter (Figure A2). Although we observe more sensitivity to the orientation of the terrain in case of unvegetated or sparsely vegetated terrain, we do not observe the same sinusoidal pattern but rather a peak at about 150° . Interestingly the range of aspect angles at which we observe a maximum backscatter corresponds to slopes facing away from both ALOS PALSAR and Sentinel-1. It can be assumed that the topographic correction introduced during pre-processing over-corrected the backscatter, introducing artefacts in the final image product.

The plot in Figure A1 and A2 do not distinguish between pixels located on steep terrain or flat terrain. A 3-dimensional plot of observations where backscatter is plotted as a function of terrain slope and terrain orientation is difficult to interpret. A more straightforward visualisation that combines slope and aspect and can still be considered to well synthesize landscape-specific scattering patterns and terrain conditions is provided in Figures A3 and A4. There, the C- and L-band backscatter, respectively, is plotted as a function of local incidence angle. The plots show a minimum in correspondence of the incidence angle for flat terrain conditions and higher backscatter for sloped terrain, following a quadratic pattern. The sensitivity of the backscatter upon incidence angle differed depending on canopy density and, in addition, to polarization and frequency. Although one would have expected that the highest backscatter corresponded to the smallest incidence angles (i.e., slopes facing the radar), we see high backscatter also in correspondence of the largest incidence angles for steep slopes facing away from the radar. This is another way of showing the over-correction applied during pre-processing, i.e., a "global" correction for pixel area and incidence angle that did not account for the specific landscape.

To then understand the impact of terrain slope on a retrieval of a forest variable, we plotted the same observations of Figures A3-A4, as a function of canopy density after stratifying by local

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	132	14.02.2019	

incidence angle. Figures A5 and A6 show that there is a clear difference in terms of functional dependence between backscatter and canopy density when considering observations for different incidence angles. The strongest sensitivity was obtained for flat terrain (incidence angles between 30° and 40°); the weakest for steep slopes with large incidence angles.

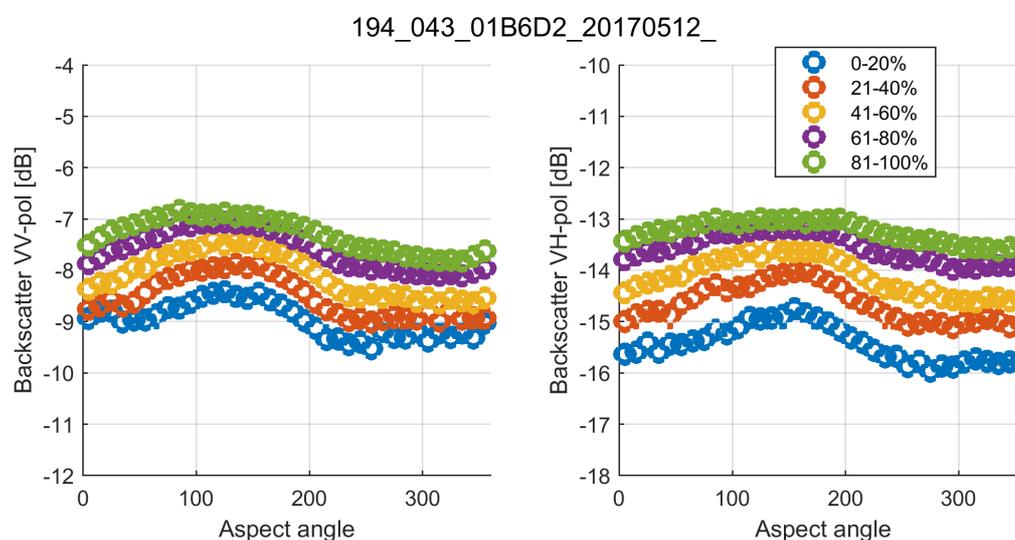


Figure A1. Observations of Sentinel-1 backscatter as a function of terrain aspect angle for the $1^\circ \times 1^\circ$ tile with top left coordinate: 14°E , 47°N . The tile includes temperate forests and alpine terrain between Austria and Slovenia. Aspect angle is defined as being zero when the normal to the surface is oriented northwards. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and aspect angle interval (10° interval).

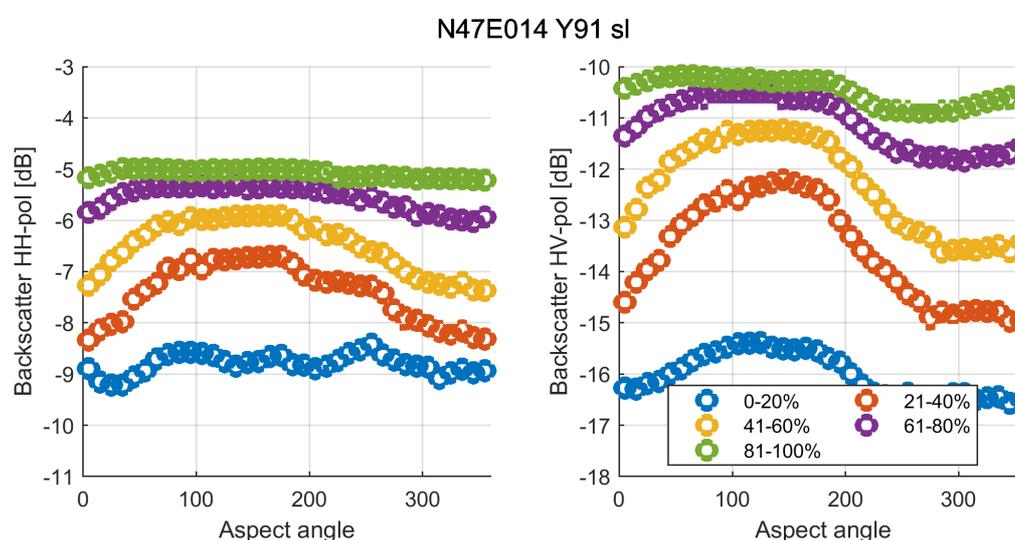


Figure A2. Observations of ALOS PALSAR backscatter as a function of terrain aspect angle for the same $1^\circ \times 1^\circ$ tile as in Figure A1. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and aspect angle interval (10° interval).

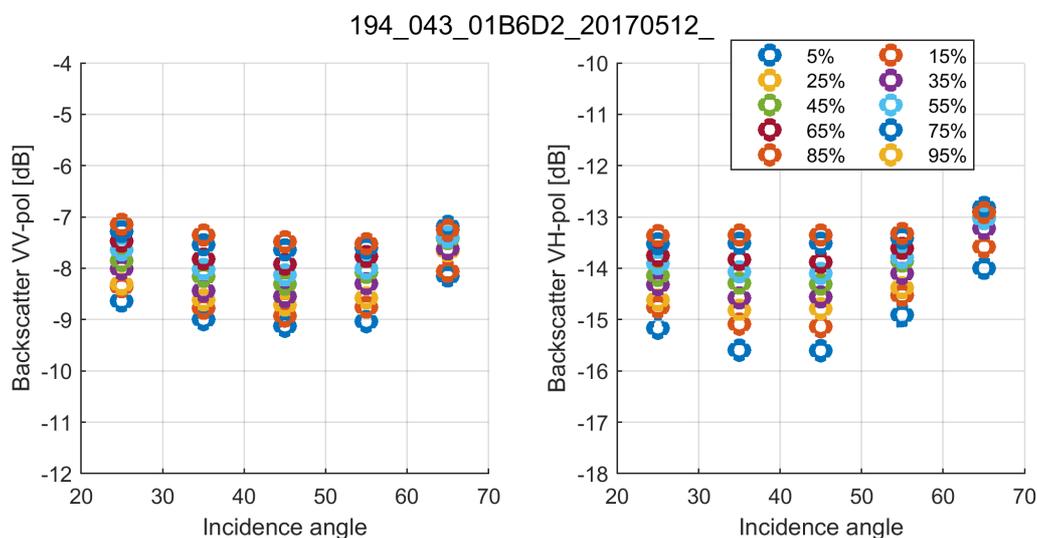


Figure A3. Observations of Sentinel-1 backscatter as a function of local incidence angle for the same dataset and tile in Figure A1. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

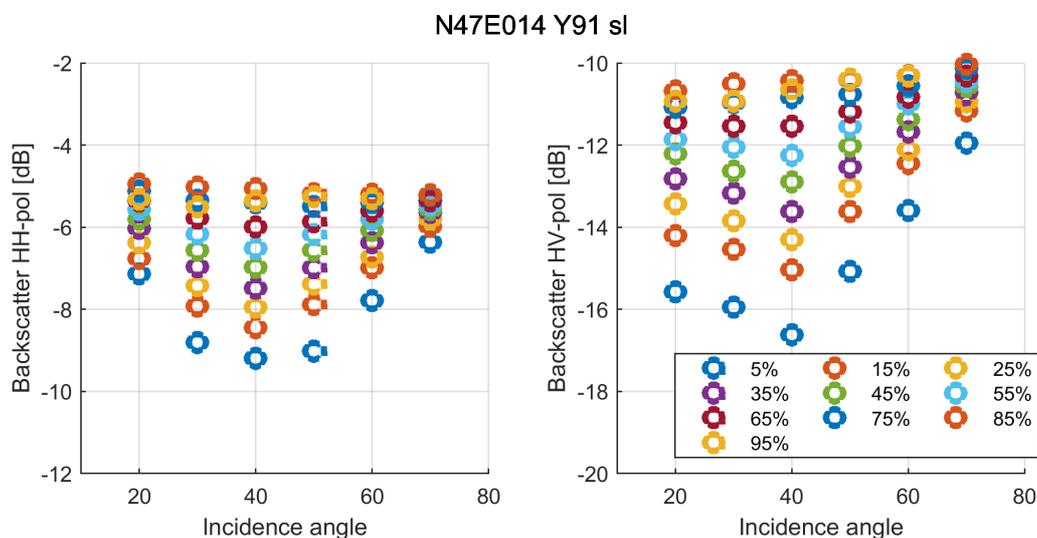


Figure A4. Observations of ALOS PALSAR backscatter as a function of local incidence angle for the same dataset and tile in Figure A2. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

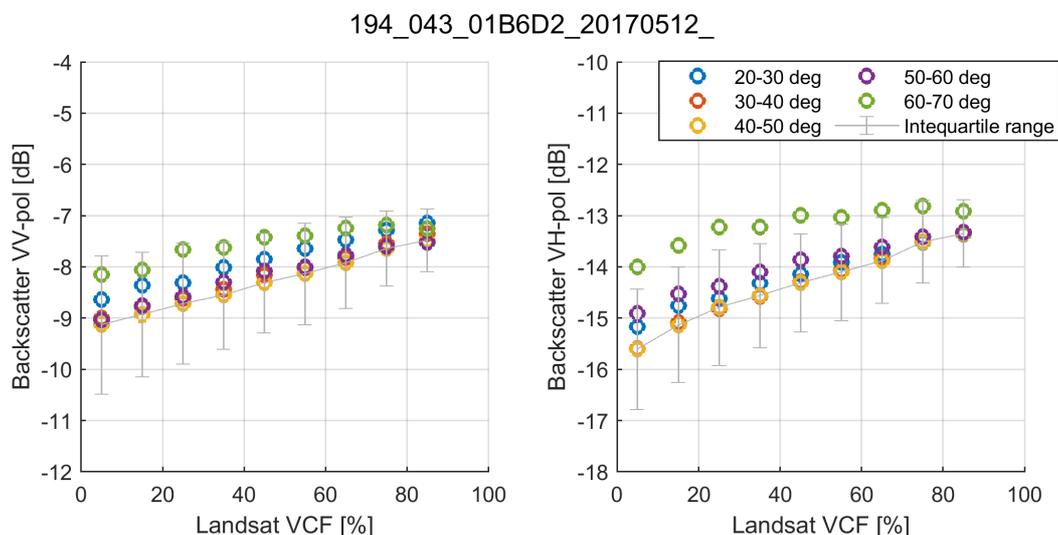


Figure A5. Observations of Sentinel-1 backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A1. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval). The vertical bars represent the interquartile range of backscatter measurements in a given VCF range and incidence angle interval of flat terrain.

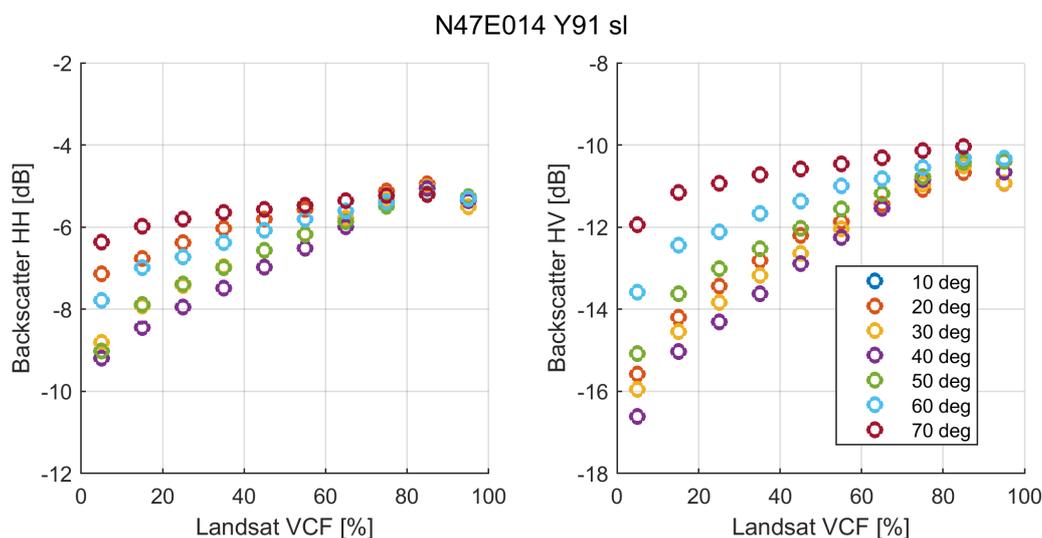


Figure A6. Observations of ALOS PALSAR backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A2. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	136	14.02.2019	

8.2 A2 Flat to hilly terrain, tropical forest

The analysis in Section A1 was repeated for a more moderate topography and a different forest type. The results are in line with observations in Section A1, showing less variability though because of the smaller range of incidence angles. The relationship between aspect angle and SAR backscatter (Figures A7 and A8) show the same sinusoidal pattern as in Section A1. For canopy density ranges where this is not as clear, the reason is the small number of pixels. Similarly, plotting the SAR backscatter as a function of local incidence angle and stratifying by canopy density (Figures A9 and A10) confirms that the smallest backscatter is obtained in correspondence of flat terrain whereas steep slopes are characterized by higher values. When reversing the plots by expressing the SAR backscatter as a function of canopy density after stratifying by local incidence angle (Figures A11 and A12), we see different backscatter levels for different ranges of incidence angle for canopy density up to 70%. Thereafter the almost negligible sensitivity of short-wavelength backscatter to forest variable appears to set in and the differences are of the order of a fraction of dB. Still, there appear to be systematic differences related to local incidence angle. This is confirmed when looking at Figures A13 and A14, where we plotted the SAR backscatter as a function of AGB for a subset of the $1^\circ \times 1^\circ$ corresponding to the area covered by the LiDAR-based dataset of AGB at the test site of Lope, Gabon (Labriere et al., 2018), Interestingly, the SAR backscatter presented some variation for increasing AGB even at the highest levels of AGB whereas there was hardly any variability when the same observations were plotted against canopy density. This seems to indicate that the reliability of the percent tree cover values above 80% can be questioned as no sensitivity of the backscatter to canopy density was evident above 80%.

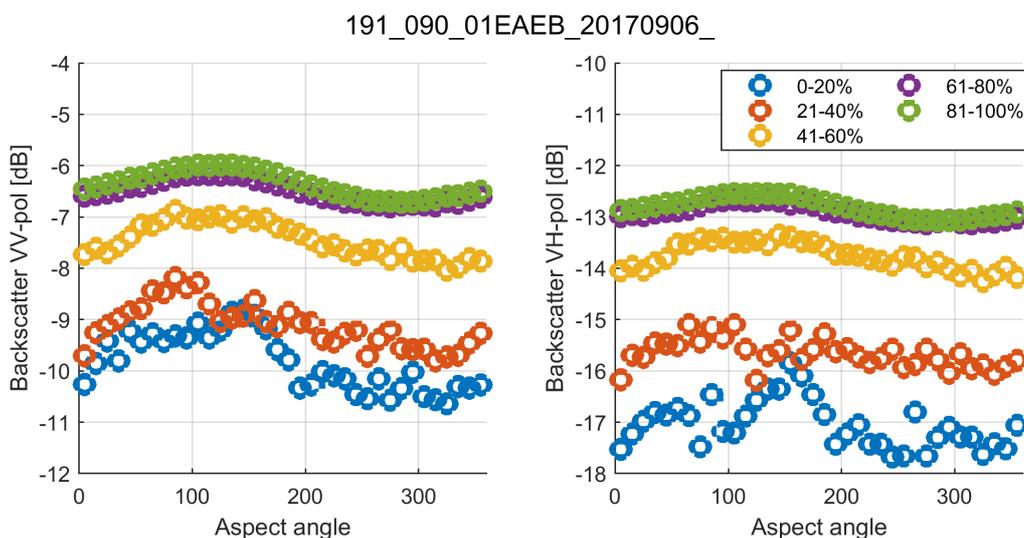


Figure A7. Observations of Sentinel-1 backscatter as a function of terrain aspect angle for the 1° x 1° tile with top left coordinate: 11°E, 0°N. The tile includes the forest site of Lope, Gabon. Aspect angle is defined as being zero when the normal to the surface is oriented northwards. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and aspect angle interval (10° interval).

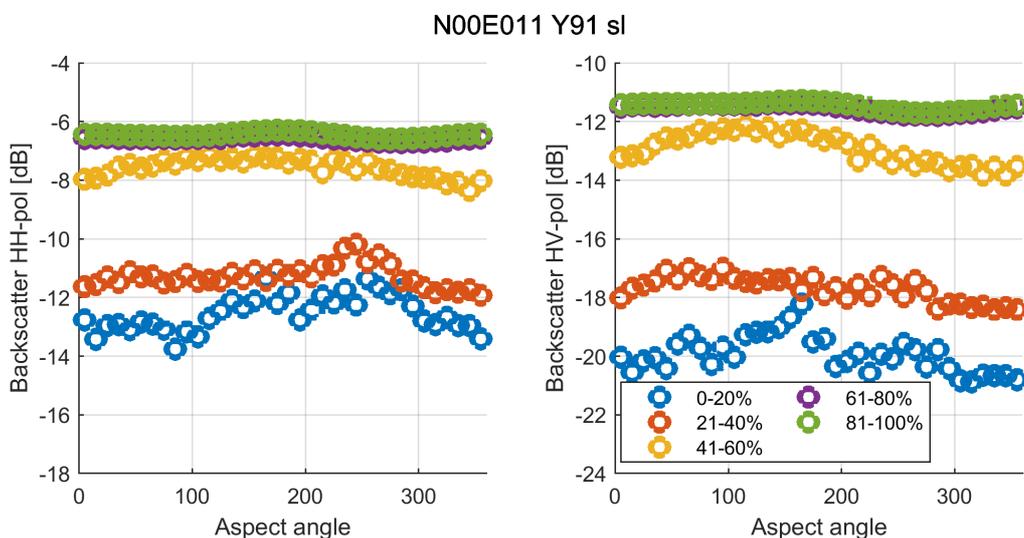


Figure A8. Observations of ALOS PALSAR backscatter as a function of terrain aspect angle for the same 1° x 1° tile as in Figure A7. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and aspect angle interval (10° interval).

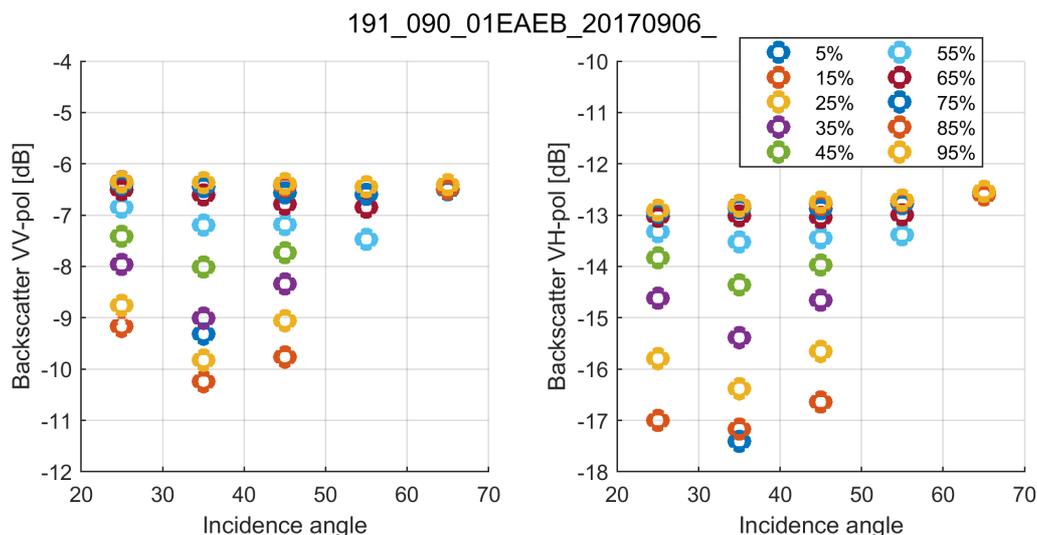


Figure A9. Observations of Sentinel-1 backscatter as a function of local incidence angle for the same dataset and tile in Figure A7. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

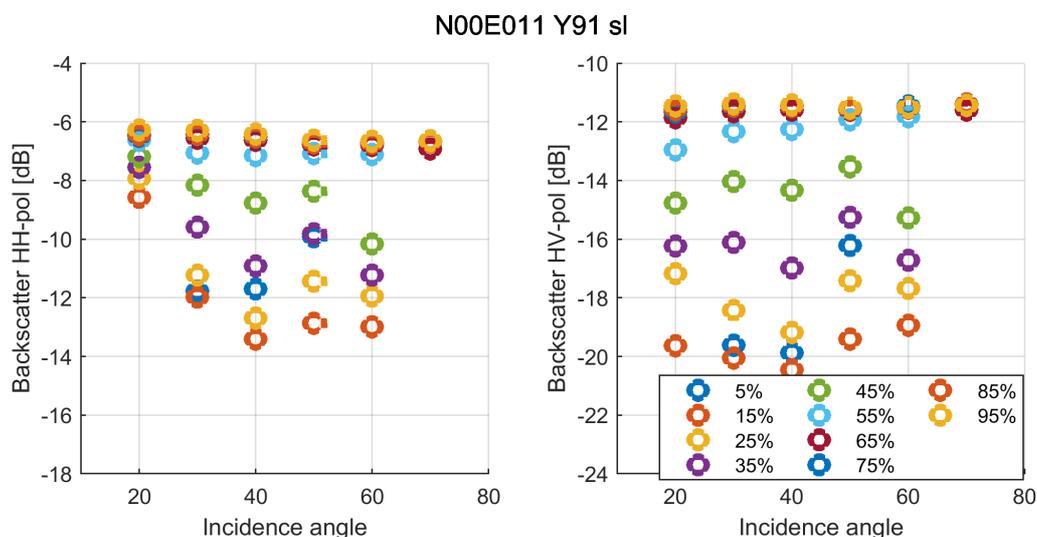


Figure A10. Observations of ALOS PALSAR backscatter as a function of local incidence angle for the same dataset and tile in Figure A8. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

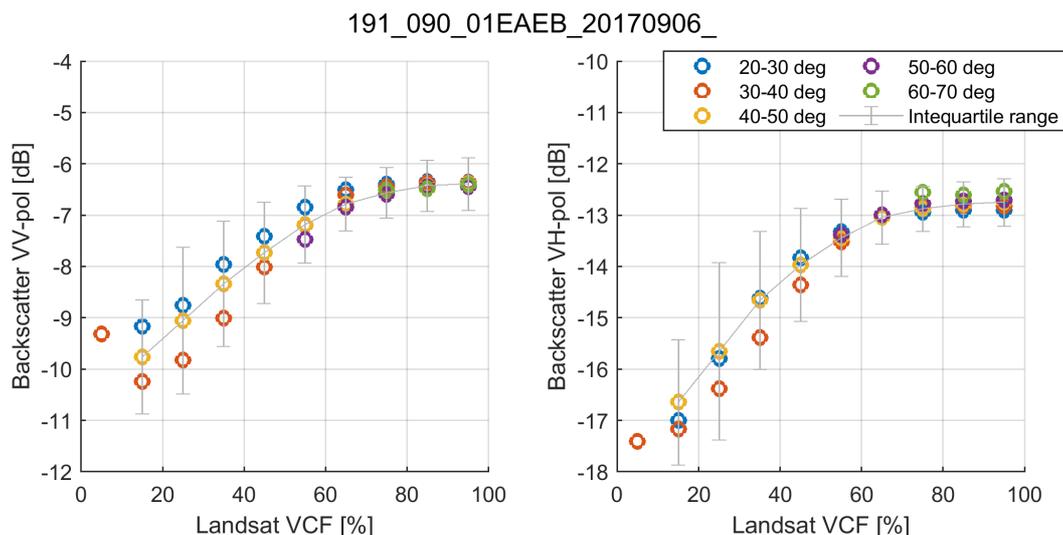


Figure A11. Observations of Sentinel-1 backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A7. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval). The vertical bars represent the interquartile range of backscatter measurements in a given VCF range and incidence angle interval of flat terrain.

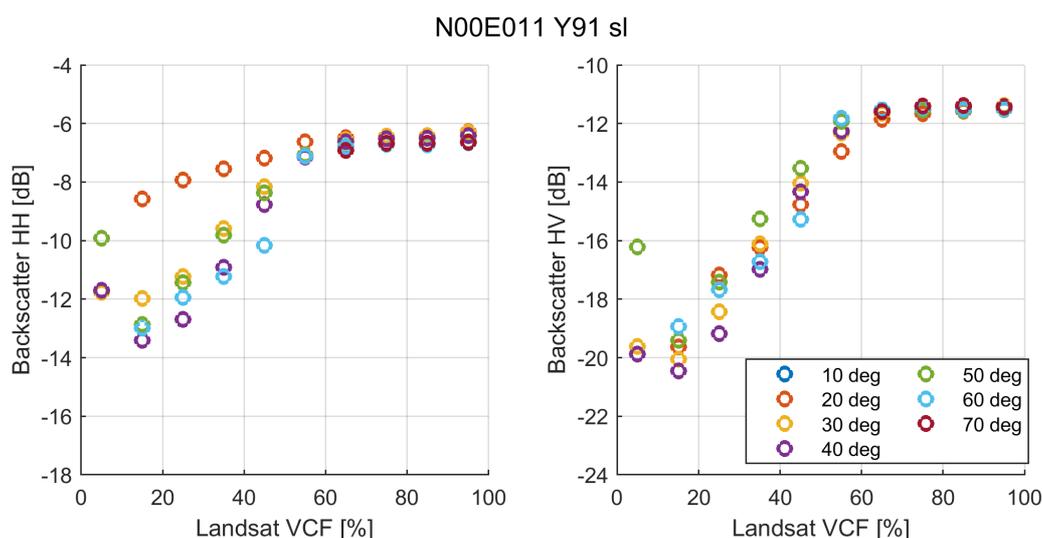


Figure A12. Observations of ALOS PALSAR backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A8. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval).

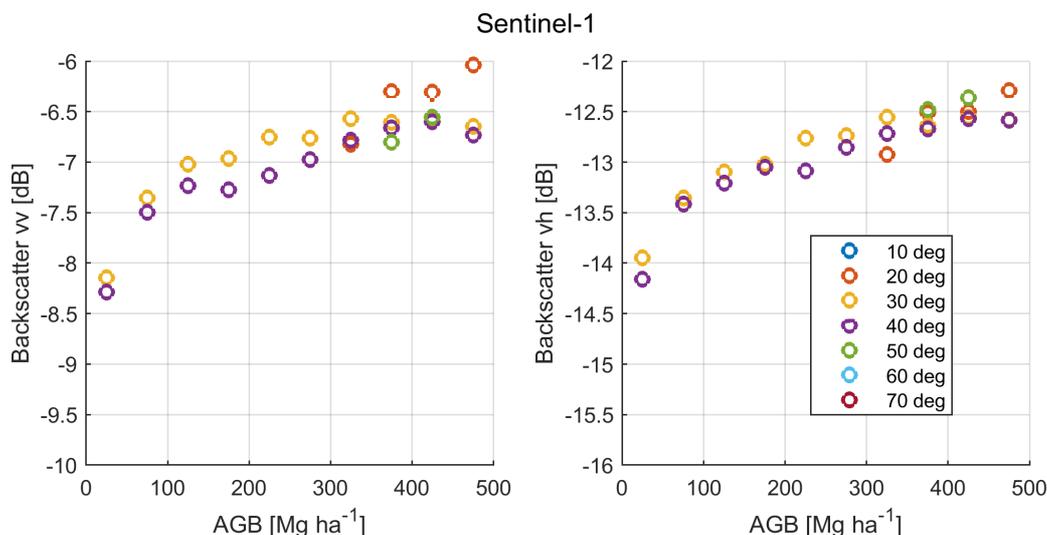


Figure A13. Observations of Sentinel-1 backscatter as a function of AGB over Lope, Gabon, (Labriere et al., 2018) for the same dataset in Figure A7. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and AGB interval (50 Mg ha⁻¹).

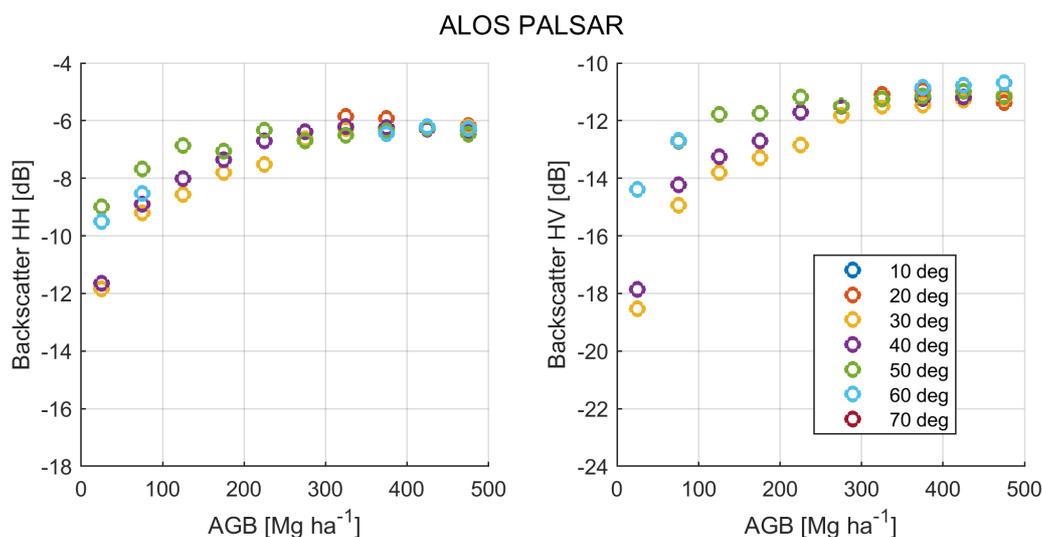


Figure A14. Observations of ALOS PALSAR backscatter as a function of AGB over Lope, Gabon, (Labriere et al., 2018) for the same dataset in Figure A8. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and AGB interval (50 Mg ha⁻¹).

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	141	14.02.2019	

8.3 Annex B

In this Annex, we present a comparison of percent tree cover datasets with the objective of drawing some conclusion on the quality of canopy density datasets to be used in the model training phase of the BIOMASAR algorithm. Several regions of the Earth were analyzed; the examples below are indicative for an area of frequent cloud cover throughout the year. Cloud cover was indeed identified the major factor to cause artefacts in global maps of canopy density.

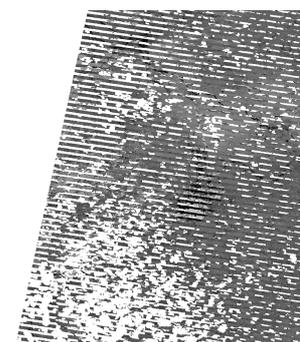
Figure B1 shows four percent tree cover datasets with spatial resolution between 30 m and 1,000 m of potential interest to support the model training. For reference, we included the optical image in Google Earth. This area(100 km × 60 km) is located in Equatorial Guinea. The nearly horizontal, repeated segments in the two Landsat canopy density datasets are a consequence of the SLC-off artefacts in Landsat imagery due to scanning. Because of the poor Landsat coverage in recent years, the annual Landsat VCF datasets by Sexton et al. (2013), are often characterized by SLC-off effects and missing data because of cloud cover. The MODIS VCF dataset shows strong variability as a consequence of frequent cloud cover, which introduced a number of artefacts and reduced the overall quality of the canopy density estimates. The Proba-V tree cover fraction appears to be reliable in terms of artefacts but seems to classify tree cover as being 100% everywhere there is vegetation. This feature was noticed in boreal forests as well, where forests are sparser.

As a result, the percent Landsat-based tree cover described in Section 3 appears to be the most reliable dataset globally whereas the others suffer from artefacts and missing values due to cloud cover and haze with some quite destructive impact on model training and, thereof, biomass retrieval.

	Ref	CCI Biomass Algorithm Theoretical Basis Document v1		
	Issue	Page	Date	
	1.0	142	14.02.2019	



Landsat percent tree cover, 30 m, 2010.



Landsat Vegetation Continuous Fields (tree cover), 30 m, 2015.



MODIS Vegetation Continuous Fields v6, Proba-V tree cover fraction, 100 m, 2015. 250 m, 2017.

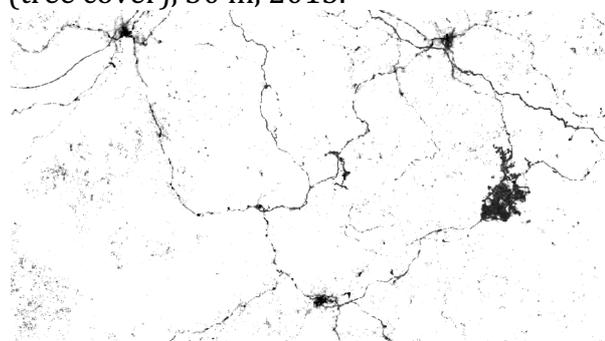


Figure B1. Comparing four percent tree cover datasets for a 100 km × 60 km large area in Equatorial Guinea. The image at the top of this Figure was taken from Google Earth. Each percent tree cover image is scaled between 0 and 100. The white features in the Landsat VCF dataset by Sexton et al. (2013) represent invalid locations and are caused either by SLC-off artefacts (segments) or clouds.