CCI
BIOMASS

ALGORITHM THEORETICAL BASIS DOCUMENT
VERSION 6.0

DOCUMENT REF:	CCI_BIOMASS_ATBD_V6
DELIVERABLE REF:	D2.2-ATBD
VERSION:	6.0
CREATION DATE:	2019-02-14
LAST MODIFIED	2024-12-19

Document Authorship

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

Document Distribution

ORGANISATION	NAME	QUANTITY
ESA	Frank Seifert	

Document History

VERSION	DATE	DESCRIPTION	APPROVED
1.0	2019-02-14	ATBD of CCI Biomass year 1	
2.0	2020-01-09	ATBD of CCI Biomass year 2	
3.0	2021-06-14	ATBD of CCI Biomass year 3	
4.0	2023-02-08	ATBD of CCI Biomass year 1, Phase 2	
5.0	2023-11-30	ATBD of CCI Biomass year 2, Phase 2	
6.0	2024-12-19	ATBD of CCI Biomass year 3, Phase 2	

Document Change Record (from Year 1 to Year 2)

VERSION	DATE	DESCRIPTION	APPROVED
2.0	2020-01-09	Introduced description of CORE algorithm used to estimate AGB in year 2	
3.0	2021-06-14	Updated description of CORE algorithm to estimate AGB and introduced the algorithms for AGB change detection	
4.0	2023-02-08	Updated description of SAR datasets and CORE algorithm to estimate AGB. Removed parts related to CORE algorithm developments from previous versions.	
5.0	2023-11-30	Updated description of SAR datasets and CORE algorithm to estimate AGB.	
6.0	2024-12-19	Updated description of SAR datasets and CORE algorithm to estimate AGB.	

Table of Contents

LIST OF FIGURES

LIST OF TABLES

SYMBOLS AND ACRONYMS

ADP	Algorithm Development Plan
AGB	Above-Ground Biomass
ALOS	Advanced Land Observing Satellite
ASAR	Advanced Synthetic Aperture Radar
ASF	Alaska Satellite Facility
ATBD	Algorithm Theoretical Basis Document
ATLAS	Advanced Topographic Laser Altimeter System
BCEF	Biomass Conversion & Expansion Factor
CCI	Climate Change Initiative
CCI-Biomass	Climate Change Initiative – Biomass
CCI-LC	Climate Change Initiative – Land Cover
CF	Cost Function
CRDP	Climate Research Data Package
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
FIA	Forest Inventory Analysis
FRA	Forest Resources Assessment
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
GRD	Ground Range Detected
GSV	Growing Stock Volume
HOME	Height Of Median Energy
HH	Horizontal-Horizontal

HV	Horizontal-Vertical
ICESat	Ice, Cloud, and land Elevation Satellite
IIASA	International Institute of Applied Systems Analysis
IMM	Image Mode Medium
IPCC	Intergovernmental Panel on Climate Change
ISS	International Space Station
IWS	Interferometric Wide Swath
JAXA	Japan Aerospace Exploration Agency
KC	Kyoto & Carbon
LiDAR	Light Detection and Ranging
LUT	Look Up Table
MERIS	Medium Resolution Imaging Spectrometer
MCH	Mean Canopy Height
MLI	Multi-Looked Intensity
MODIS	Moderate Resolution Imaging Spectroradiometer
MOLI	Multi-footprint Observation LiDAR and Imager
MSE	Mean Squared Error
NaN	Not a Number
NESZ	Noise Equivalent Sigma Zero
NFI	National Forest Inventory
PALSAR	Phased Array type L-band Synthetic Aperture Radar
PDP	Partial Dependence Plots
PSD	Product Specification Document
PVASR	Product Validation and Algorithm Selection Report
PVIR	Product Validation and Intercomparison Report
RF	Random Forest
RFI	Radio Frequency Interference
SAR	Synthetic Aperture Radar
SLC	Single Look Complex
SMOS	Soil Moisture and Ocean Salinity
SRTM	Shuttle Radar Topography Mission
TCH	Top of Canopy Height
TEOW	Terrestrial Ecoregions Of the World
URD	User Requirements Document
USGS	United States Geological Survey
VCF	Vegetation Continuous Fields
VIM	Variable Importance Measure
VOD	Vegetation Optical Depth
WB	Wide Beam

WSM	Wide Swath Mode

Reference Documents

ID	Title	Issue	Date
RD-1	Users Requirements Document (URD)	V3	
RD-2	Product Specification Document (PSD)	V6	
RD-3	Data Access Requirements Document (DARD)	V1	
RD-4	Product Validation and Algorithm Selection (PVASR)	V3	
RD-5	End to End ECV Uncertainty Budget (E3UB)	V6	
RD-6	Algorithm Development Plan (ADP)	V6	
RD-7	Product Validation Plan (PVP)	V6	
RD-8	Product Validation Report (PVR)	V5	
RD-9	Product User Guide (PUG)	V5	

All documents available at https://climate.esa.int/en/projects/biomass/key-documents/

1 Introduction

Above-ground biomass density (AGB, units: Mg ha^{-1}) is defined by the Global Carbon Observing System (GCOS) as one of 54 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), 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 has allowed the generation of several wall-to-wall datasets of AGB. The uncertainty in magnitude and distribution of AGB prior to the Climate Change Initiative (CCI) Biomass project is illustrated in (Figure 1-1), where each line represents latitudinal averages of AGB estimated with remote sensing data. While the overall trends in the AGB spatial distribution are consistent across the AGB datasets, the variability of AGB among these datasets is, on average, more than 100% (precision 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 because AGB can only be *inferred* from observations since remote sensing instruments do not have the capability to measure the organic mass stored in vegetation. 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.

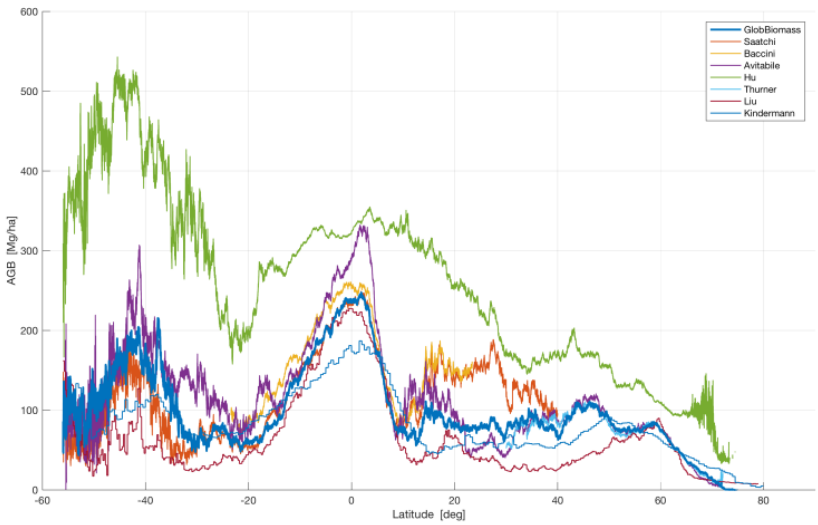


Figure 1-1: Latitudinal averages of AGB estimates from the GlobBiomass dataset (Santoro et al. (2021), Saatchi et al. (2011), Baccini et al. (2012), Avitabile et al. (2016) and GEOCARBON dataset, Hu et al. (2016), Thurner et al. (2014), Liu et al. (2015) and Kindermann et al. (2008).

The objectives of the 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 several epochs and assess biomass changes between epochs (annual and decadal changes). The maps should be spatially and temporally consistent; in addition, they need to be consistent with other thematic data layers 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 that 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 earlier Product Validation and Algorithm Selection (PVASR) documents [RD-4], which investigated 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 precision of the AGB estimates [RD-5]. An estimate of the bias of the maps, assessed with inventory plot data and a modelling framework, is provided in the Product Validation and Intercomparison Report (PVIR) [RD-8]. 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 CCI Biomass (Phase 1), methods were developed that led to the generation of a first version of a global AGB product for the year 2017. During Year 2, these methods were refined by considering the assessment of the AGB map of 2017 and alternative algorithmic advances documented in the PVASR and in the ADP of Year 1. The ATBD was updated in Year 2 to document the algorithms implemented to generate AGB estimates for the epochs 2010, 2017 and 2018. The focus of Year 2 was to generate a first set of three AGB maps. These were generated independently of each other to gather understanding on global AGB mapping in several epochs. The overall spatial distribution was well captured, although the AGB estimates were affected by different biases and errors that were particularly noticeable in the densely forested tropics [RD-8]. As a first approach to quantifying AGB changes at yearly and almost decadal scale, difference maps were also generated. Large scale errors were apparent when comparing the 2010 dataset with the other two datasets. The different set of EO data available for 2010 compared to 2017 and 2018 explained these discrepancies. Based on these conclusions, the work undertaken in Year 3 consisted of improving the accuracy of each of the three individual maps of AGB and allowed a first assessment of AGB change between epochs. At the end of Phase 1 of the CCI Biomass project, the accuracy of the individual maps was improved but correction methods to overcome systematic discrepancies between the maps did not perform sufficiently well to guarantee correct estimates of AGB changes. During the first year of Phase 2 of the CCI Biomass project, the retrieval algorithms were further developed to reduce biases and improve the inter-annual consistency of the AGB estimates and provide reliable values for AGB changes on annual and decadal time scales. Biases were addressed with a novel training of the function relating height to AGB and by using a much wider dataset of spaceborne LiDAR observations. Verification of the AGB maps produced with this algorithm showed that the modifications to the CORE algorithm removed some of the largest biases but also introduced some errors. During year 2 of Phase 2, the CORE algorithm was improved with more extensive datasets of LiDAR observations and subnational statistics of AGB, which supported the calibration of one of the structural functions integrated in the retrieval model. In addition, the temporal adjustment of AGB estimates was introduced to avoid unnatural year-to-year fluctuations due to the weak-to-moderate sensitivity of the SAR observations to AGB. Verification of the AGB estimates confirmed that the algorithm improvements are of benefit but still some issues related to both data and models need to be solved. In this report, we document recent advances of the CORE retrieval algorithm intended to further reduce biases and improve the consistency of the AGB estimates on a decadal scale.

Section 2 provides the background of this ATBD, describing the strategy that underpins the algorithms implemented in CCI Biomass to estimate AGB. The ATBD also describes the datasets (EO and auxiliary) used to estimate AGB (Section 3); the AGB retrieval methods used to generate global maps of AGB (Section 4); and the methods used to quantify AGB changes across epochs (Section 5). Correction of AGB biases is addressed in Section 6. An assessment of the retrieval algorithm and the biomass change algorithm is presented in Section 7.

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 aim of achieving a quantitative global description 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 EO observations up to the start of the CCI Biomass project in 2018. This list is not meant as an evaluation of the data product but rather to state where past experiences can be of use in enhancing or designing AGB retrieval algorithms based on current EO data.

The availability of global and repeated observations, first by the MODIS sensors and more recently by Landsat sensors, fostered the estimation of global raster datasets of canopy height (Lefsky et al., 2010; Simard et al., 2011) and AGB (Saatchi et al., 2011; Baccini et al., 2012; Hu et al., 2016), the latter being the first to utilise the Ice, Cloud and land Elevation (ICESat) GLAS waveform data to calibrate biomass prediction models. 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 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 Light Detection and Ranging (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.

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 (Saatchi et al., 2011; Baccini et al., 2012), generated a new 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 was, in our opinion, also its weakness, in the sense that the method was insufficiently constrained in regions where reference datasets (*in situ*, laser-based) were 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, effort should be spent on developing retrieval algorithms that integrate reference and EO data, rather than attempting to fuse estimates from different sources that may ultimately lead to aggregation of errors rather than provide a 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). Hence, it can be reasonably assumed that an estimation based on microwave data (Synthetic Aperture Radar; SAR-based observations, brightness temperatures) will improve the accuracy of biomass-related variables. Furthermore, the more frequent sampling of the Earth by spaceborne LiDAR, including those that are recent (ICESat-2 and Global Ecosystem Dynamics Investigation; GEDI) or proposed (e.g., the Multi-footprint Observation Lidar and Imager; MOLI) allow a larger proportion of the estimated AGB to be explained in terms of the waveform-based measurements, thus in principle leading to a more accurate set of biomass estimates.

At coarser resolutions, attention must be paid to the Vegetation Optical Depth (VOD) observable from passive microwave observations. X-band VOD has been used to generate yearly maps of forest AGB and carbon over a period of 20 years (Liu et al., 2015) at 25 km spatial resolution. L-band VOD from Soil Moisture and Ocean Salinity (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 AGB 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 easily be challenged by noting that 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 corresponding to regions that have not been mapped correctly in the reference dataset.

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, partly because of data availability 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. 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, if 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.

The GlobBiomass project (https://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 because EO signals change in space and time, whereas retrieval models typically do not account for such variability. Point (iii) was possibly the most innovative aspect of an algorithm for estimating AGB because it aimed to minimise the impact of reference data on the retrieval. Making a retrieval algorithm independent of reference data allows a truly independent validation of the retrieval with *in situ* data. On the other hand, it requires profound knowledge of the EO data 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 set out 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, in this case, 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, global coverage was 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 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 sparse spatial sampling for direct ingestion in a biomass retrieval scheme. It is important to note 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.

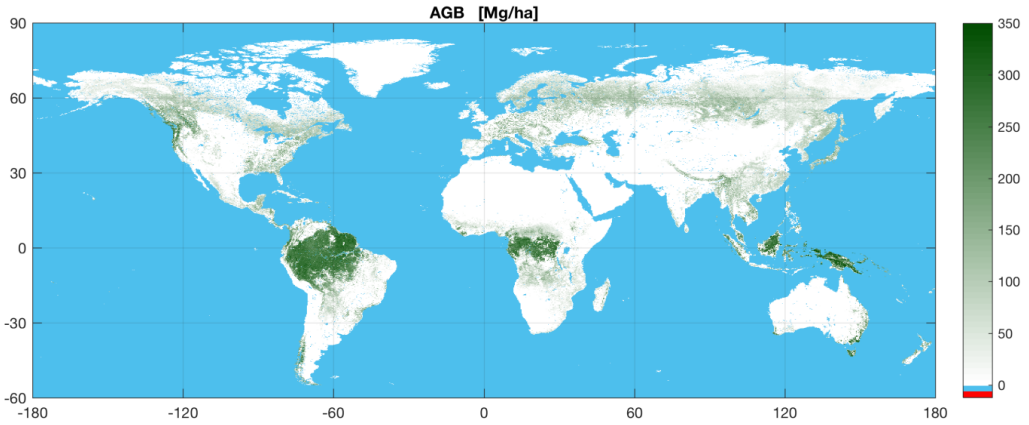


Figure 2-1: The GlobBiomass AGB dataset (Santoro et al., 2021).

Figure 2-1 shows the GlobBiomass dataset of forest AGB. Validation of the GlobBiomass AGB estimates (Santoro et al., 2021) 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 (> 250 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 (Santoro et al., 2021).

- A too conservative constraint on the maximum biomass for a given area, causing underestimation in the high biomass range.
- A too generic definition of the forest transmissivity term in the models relating SAR backscatter to growing stock volume (GSV), 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 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, causing unwanted local biases.

While the weak sensitivity of the SAR backscatter to AGB 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 in the retrieval models.
- 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 can potentially clarify how uncertain the biomass pool is globally (see Figure 1-1) and overcome issues in GlobBiomass and other endeavours aiming to characterize the world's forest biomass.

1. Retrieval of biomass requires multiple data sources, including EO data not particularly suited to retrieving biomass.
2. Height information can substantially improve the estimates of biomass where the other EO observables do not exhibit sensitivity to biomass.
3. Retrieval of biomass does not necessarily require reference biomass data (e.g., in situ observations of biomass) for training.

4. 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, 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 spaceborne LiDAR instruments is mandatory to better characterize the relationship between biomass-related variables in the retrieval models. 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

The global biomass retrieval algorithm implemented in CCI Biomass followed 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) compensate for systematic errors revealed by the assessment of the GlobBiomass dataset. The current version of the CORE algorithm furthermore expands on the experience of previous CCI Biomass project years. In the process of improving the CORE algorithm, changes are applied that account for evidence from assessment of the AGB maps of previous years described in the PVIR [RD-8]) and the PVASR [RD-4].

A further development regards the inter-annual consistency of AGB estimates, which is needed to estimate AGB changes. Since the pool of EO observations used to estimate AGB is not the same throughout the time interval envisaged by the CCI Biomass project (2005-2022), retrieval algorithms need to ensure that such diversity is compensated for to ensure temporal consistency of the estimates across two decades.

3 Datasets and additional material

This section describes the EO datasets used to generate the AGB and AGB change maps to be included in the Climate Research Data Package (CRDP) for the current project year.

Current CRDP: version 6 (v6)
Years: 2007, 2010, 2015-2016, 2017, 2018, 2019, 2020, 2021 and 2022

For 2015-2022, the EO dataset consists of ALOS-2 PALSAR-2 and Sentinel-1 observations of SAR backscatter. Wall-to-wall coverage by other SAR datasets was not accessible and passive microwave observations did not match the requirement of the recommended spatial resolution of AGB datasets by GCOS. Optical data were not considered because they were unlikely to provide a substantial contribution to the retrieval, given global coverage by active microwave data. Microwaves penetrate the forest canopy to a certain extent so that the backscatter recorded by radar sensors is sensitive to forest structural parameters and, therefore, to AGB. Following the same reasoning, the retrieval of AGB for the years 2007 and 2010 was based on ALOS-1 PALSAR-1 and Envisat ASAR datasets.

Spaceborne LiDAR observations from the ICESat (2003-2009), ICESat-2 (2018-ongoing) and GEDI (2019-ongoing) missions are an additional source of observations. LiDAR observations capture vegetation structural features. However, their sampling is still too coarse to use them for wall-to-wall estimates of forest variables. Spaceborne LiDAR observations are, therefore, used here in the process of calibrating models rather than as predictors of AGB. We use height and canopy density metrics because these are directly related to the original observations. LiDAR-based values of AGB, as in Level 4 GEDI data products, are not ingested in the retrieval algorithm because they are themselves an estimate and not a true measurement of biomass.

The EO datasets used to generate or used in producing the global maps of AGB are summarized in Table 3-1. Each set of EO observations is then described in individual sections below.

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

Sensor	Epoch	Use
ALOS-2 PALSAR-2	2015-2022	Predictor
Sentinel-1 (A and B)	2015-2022	Predictor
ALOS-1 PALSAR-1	2007 and 2010	Predictor
Envisat ASAR	2007 and 2010	Predictor
ICESat GLAS	All epochs	Calibration
ICESat-2	All epochs	Calibration
GEDI	All epochs	Calibration

The CORE algorithm also utilises several raster datasets to support the prediction of AGB. These datasets are used either to support the calibration of modules of the retrieval algorithm or as a mask to select EO observations in the process of self-calibration of the algorithm. Table 3-2 summarises the additional datasets and a short description of their use in the context of this project.

Table 3-2: Additional datasets, type and use in the CORE retrieval algorithm.

Dataset	Type	Use
Digital Elevation Model	Surface elevation	Pre-processing of SAR data
Landsat canopy density and density change	Vegetation cover density	Calibration of retrieval model
MODIS Vegetation Continuous Fields	Vegetation cover density	Mask
CCI Land Cover	Land cover	Mask (specific classes)
Copernicus Global Land Operations	Land cover	Mask (specific classes)
FAO Global Ecological Zones	Ecoregions map	Stratification
Ecoregions of the World	Ecoregions map	Stratification

Each of the datasets in Table 3-2 is described in individual Sections. Note that inter-dependency between some auxiliary datasets exist; details are provided in the individual sections.

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. While the Sentinel-1A mission is still operating; a malfunction on Sentinel-1B stopped acquisitions at the end of 2021.

Each unit can acquire data at single and dual polarization (HH+HV or VV+VH) in several 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 their coastlines, S1 is operated in the Extended Wide Swath (EWS) mode. 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 with two satellites in operation. Since Russia and parts of Canada are covered by Sentinel-1B, these regions were not imaged during 2022.

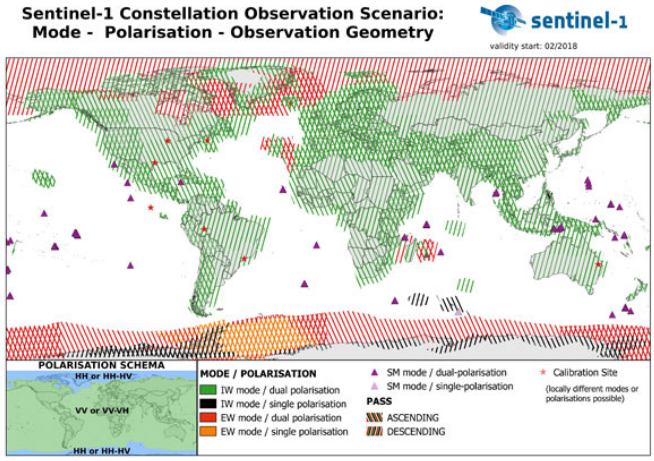


Figure 3-1: Observation geometry of the Sentinel-1 mission last accessed on 9 February 2019 (https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario).

Data acquisition by Sentinel-1 in the IWS mode follows a predefined observation scenario with different levels of priority. 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, volcano 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). Sentinel-1 operations originally aimed at achieving global coverage every 12 days with each unit. However, the duty cycle of each unit allowed global coverage every 12 days only when both satellites were in operation. With the loss of Sentinel-1B, global coverages are not achieved. 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. EWS primarily targets polar regions, with some limited extension over northern regions.

The acquisitions in IWS mode are programmed to give minimal overlap of swaths from adjacent orbital tracks at the Equator. The overlap increases towards the poles so the number of observations within a repeat-pass cycle of 12 days for a given point on the ground increases. For regions observed with both units along both ascending and descending paths, one or more

observations per day were possible. 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).

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., 2015a). 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 at least 20 images available 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, retrieval based on Sentinel-1 images is feasible thanks to the repeated acquisitions, particularly from the start of routine operations by both units in 2017. However, the retrieval does not require the entire archive of data. For regions with almost daily observations, the IWS dataset acquired after 2016 was pruned to exclude observations with a 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, between 2017 and 2021, 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 was unnecessary because it is not covered with woody vegetation.

To demonstrate the wall-to-wall mapping capability of Sentinel-1, we show the data coverage by Sentinel-1 for 2017 in Figure 3-2 after pruning redundant data. The 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 used to fill gaps. Similar EWS imagery was used to fill two gaps smaller than an IWS scene (i.e., 250 x 250 km^2). This gap-filling strategy had no effect on the biomass estimates obtained from Sentinel-1 data since Sentinel-1 imagery was processed to a pixel size of 150 m.

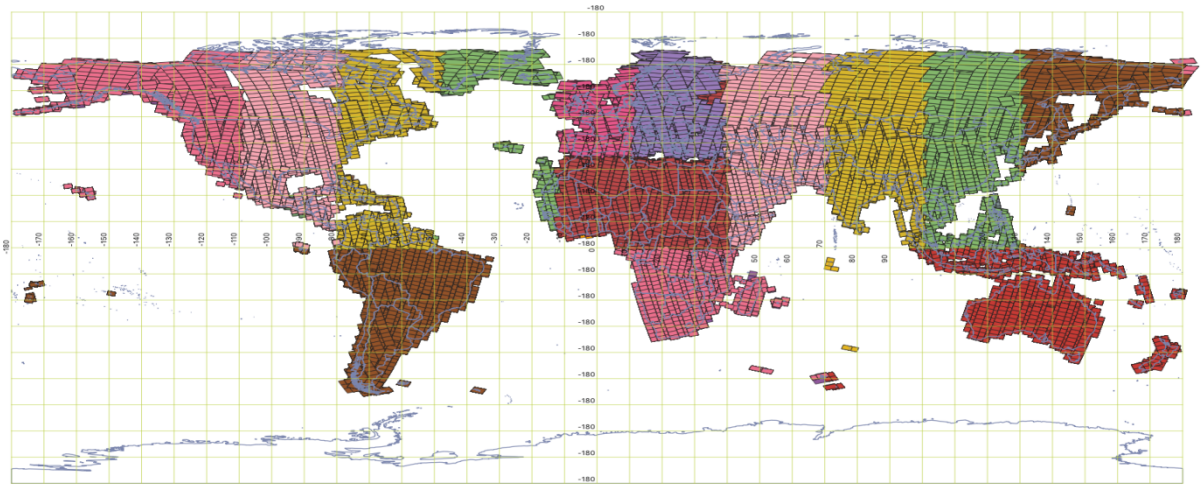


Figure 3-2: Coverage of the 2017 Sentinel-1 IWS dataset selected to support the estimation of biomass (courtesy J. Kellndorfer, EBD).

For 2015 and 2016, about 220,000 scenes acquired by Sentinel-1 A were selected for pre-processing. Approximately, 250,000 scenes were selected for pre-processing for 2017. For 2018 to 2021, the numbers increased to about 300,000 because in 2017 the 1B unit started delivering images only in April. For 2022, the number of Sentinel-1 images decreased to about 200,000 due to the end of operations of the 1B unit. Figure 3-3 shows a comparison in terms of coverage for the years 2021 and 2022. While in 2021 the coverage was almost global, in 2022 large gaps occurred over Russia, South America and Africa. The search was undertaken on the data repository of the Alaska Satellite Facility (ASF) because it mirrors European data holdings while providing speedier and more reliable access to the data.

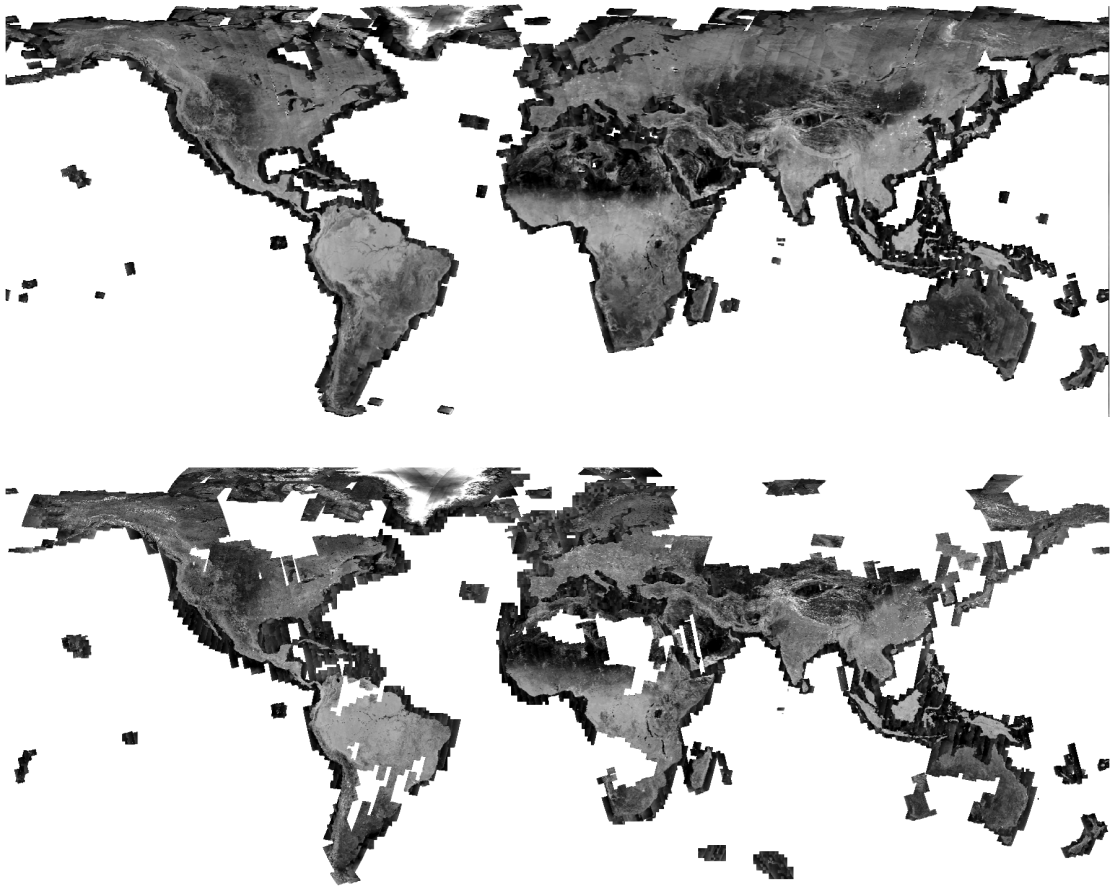


Figure 3-3: Mosaics of Sentinel-1 backscatter images (VH-polarization) based on data acquired in 2021 (top) and 2022 (bottom).

The Sentinel-1 IWS data pool consisted of individual images, each covering an area of approximately 250 km × 250 km, in the acquisition geometry of the Sentinel-1 radar. Pre-processing generated a stack of terrain geocoded, radiometrically calibrated, speckle-filtered and co-registered Sentinel-1 observations 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 azimuth direction is about 14 m and the GRD data has been multi-looked by factor 5 in range to a ground range resolution of ~18 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 variables, such as coherence, and it has been shown that estimation of biomass from C-band coherence is more accurate than from SAR backscatter (Santoro et al., 2002; Santoro et al., 2018), 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 all vegetated regions on Earth. The effort of processing SLC data to coherence globally is therefore likely to be of little value to this project.

2) A single GRD scene in IWS mode covers an area of 250 x 250 km^2, corresponding to approximately 1.6 GB of data. The corresponding SLC image consists of approximately 8 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.

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 sufficient to support the retrieval of biomass.

The SAR pre-processing chain is shown in Figure 3-4. 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 biomass retrieval phase. 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 need 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 considered 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. 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 high spatial resolution, even 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 reformatting the SAR dataset to the GAMMA Software structure (image dataset and metadata in the image parameter file) (Wegmüller et al., 2016). In addition, calibration and noise reduction were applied using 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/).

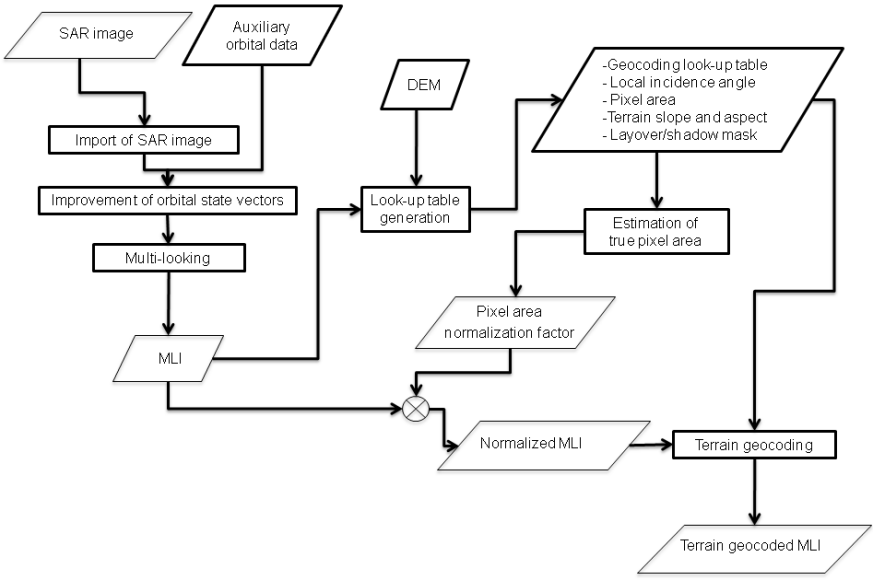


Figure 3-4: Flowchart of the Sentinel-1 data pre-processing

Multi-looking consisted of boxcar averaging of the backscatter of contiguous pixels in an averaging window of 15 × 15 pixels 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.

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

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 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 not attempted 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 3-5; 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].

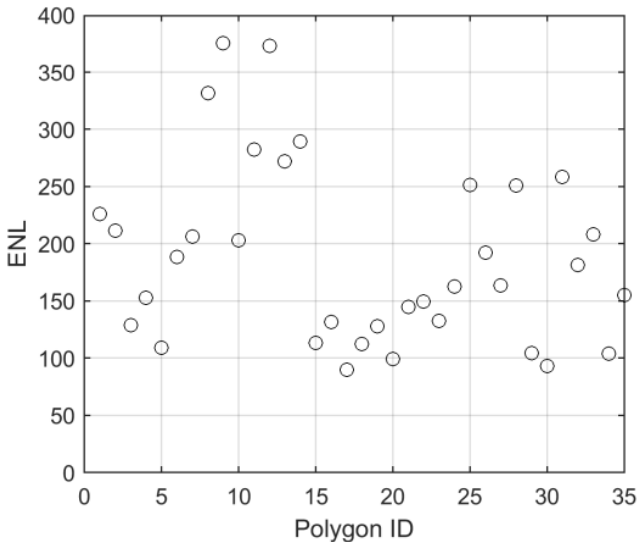


Figure 3-5: Estimates of ENL for 35 polygons distributed over five Sentinel-1 VV-polarized images randomly selected in boreal, temperate, and tropical environments.

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). For the Sentinel-1 data acquired in 2017 and 2018, we used a global 3 arc-seconds DEM (i.e., roughly 90 m at the Equator). For all other years, the newly available Copernicus DEM with 1 arc-second resolution was used. 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 $1/10^{th}$ 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 the high precision and stability of the Sentinel-1 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 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 3-4 was repeated for each Sentinel-1 image part of the data pool. To obtain the stack of co-registered observations of the SAR backscatter, each image was tiled to a pre-defined 1° × 1° grid. Each tile consisted of 720 × 720 pixels.

Figure 3-6 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). For the 2018 dataset, the density map showed the same spatial patterns as Figure 3-6 albeit with an overall larger number of acquisitions per pixel due to the routine operations of both units.

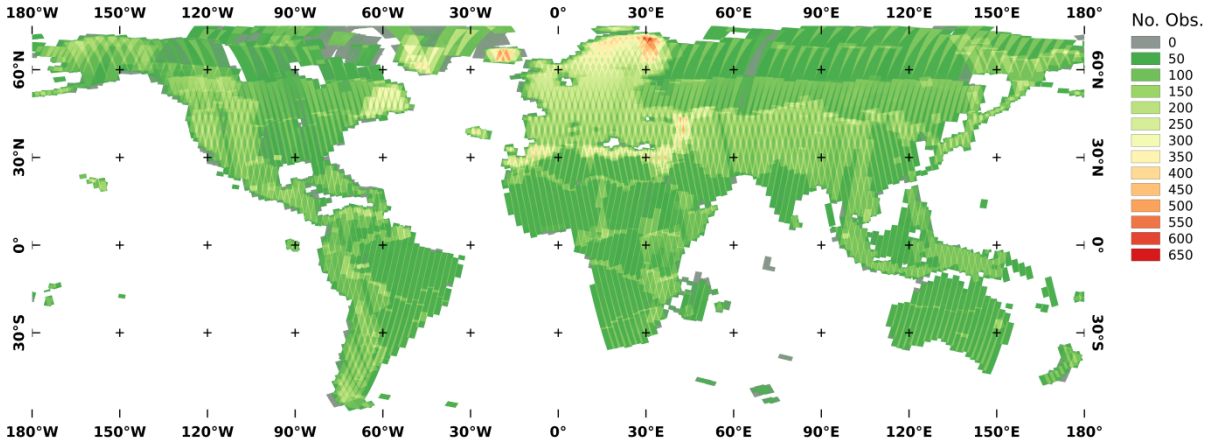


Figure 3-6: Number of observations per pixel for the Sentinel-1 2017 dataset.

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 3-7. As there were many 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 3-7 also shows that the distortions in SAR backscatter due to sloping terrain have been largely minimized.

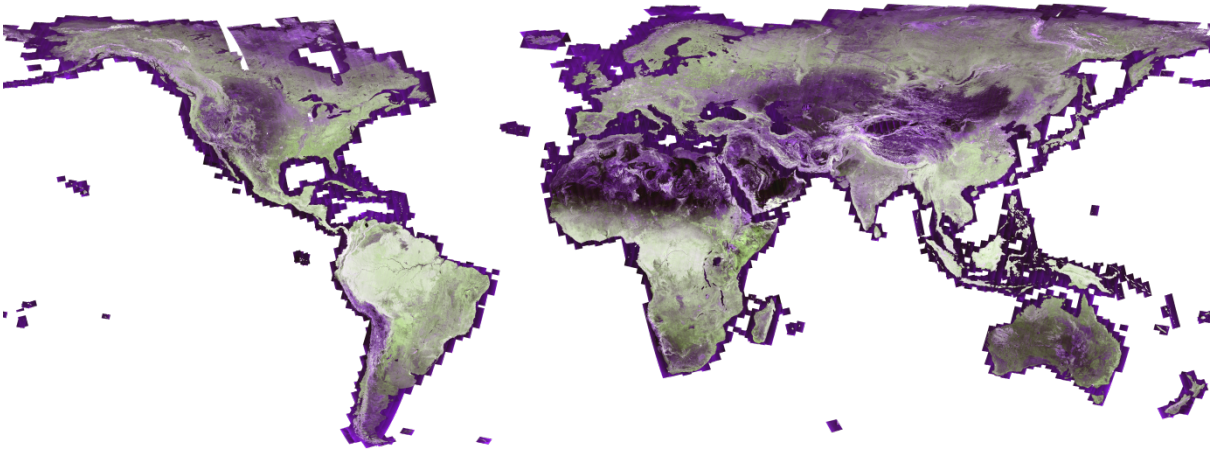


Figure 3-7: False colour composite of the Sentinel-1 2017 dataset. Red: temporally averaged co-polarized backscatter; green: temporally averaged cross-polarized backscatter; blue: ratio of the temporally averaged cross-polarized and co-polarized backscatter. Pixel size: 150 m × 150 m.

The individual Sentinel-1 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 corresponding to errors in the elevation datasets forming the global DEM. Uncompensated Radio Frequency Interference (RFI) affected the measurements locally, along the coast of the Baltic Sea. In addition, seams corresponding to the swath overlap were detected in some regions (southeast China and southwest US) which were due to incorrect representation of the noise in the metadata provided with the image data. A detailed presentation of errors affecting the Sentinel-1 backscatter dataset is given in the E3UB document [RD-5].

The high density of observations by Sentinel-1 and the high resolution of the data implies long processing times to generate biomass estimates. Although the strength of the biomass retrieval with multi-temporal C-band backscatter data is the possibility to maximize the information content on biomass in the signal by filtering out the component related to noise and environmental conditions, the individual backscatter observations are not uncorrelated in time, especially over short periods of time. For this reason, we investigated the correlation of observations acquired within the same month. This exercise was undertaken at global scale. For each 1° × 1° grid cell and for each month of the year, the correlation coefficient (Pearson's) between backscatter observations from the same orbital track was computed. Calculations were undertaken separately for the VH- and the VV-polarization. In Figure 3-8, we display the mean values of all correlation coefficients computed for each grid cell and for VH-polarization. The results described below, however, were not affected by the polarization. The correlation between observations within one month was everywhere very high (> 0.8) except for the dense and intact tropics (Amazon and Congo rainforest). These regions are characterized by a very small range of backscatter values and therefore residual noise dominates the value of the correlation coefficient. The histogram in Figure 3-9 confirms the visual interpretation of the map in Figure 3-8; 90% of the correlation coefficient values were larger than 0.66 and 75% of the values were larger than 0.81.

These results support our assumption that reducing the Sentinel-1 dataset to temporal averages would not result in a loss of information about the estimation of biomass. Eventually, monthly averages were selected because of the 12-day repeat-pass of each Sentinel-1 unit. On one hand, shorter time scales (e.g., weekly, or bi-weekly) would have only marginally reduced the amount of data and data with strong correlation would have been maintained. Longer time scales would

have reduced the seasonal signal in the backscatter data. It is indeed the seasonality of the backscatter that allows improvement of the retrieval accuracy with respect to a retrieval based on a single image. For example, in Santoro et al. (2011), it was shown that images acquired under frozen conditions in winter in the boreal zone performed better than images acquired under unfrozen conditions; nonetheless, these also contributed to reducing the retrieval error compared to an estimate based on winter-time data only.

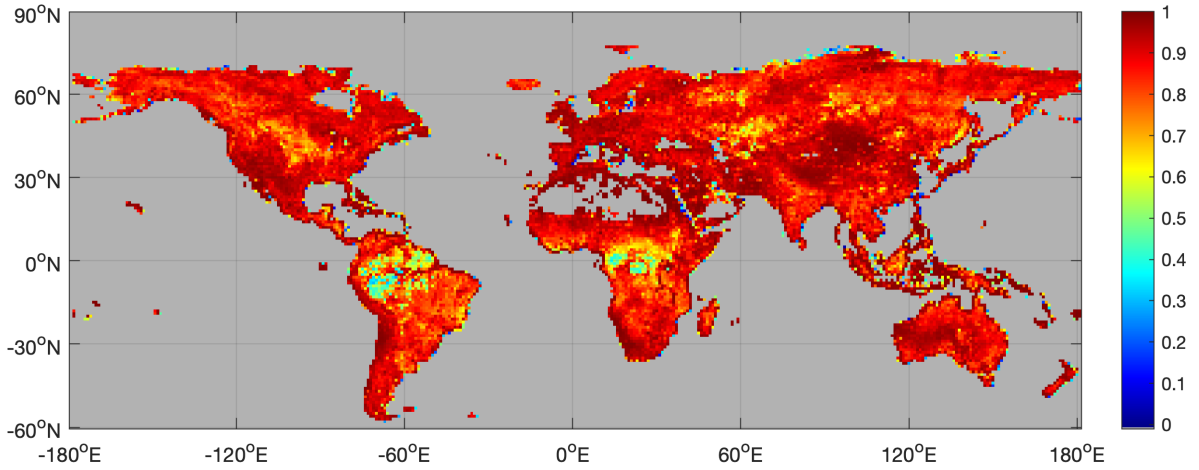


Figure 3-8: Map of the average correlation coefficient for Sentinel-1 backscatter observations (year 2020, VH-polarization) acquired in the same month.

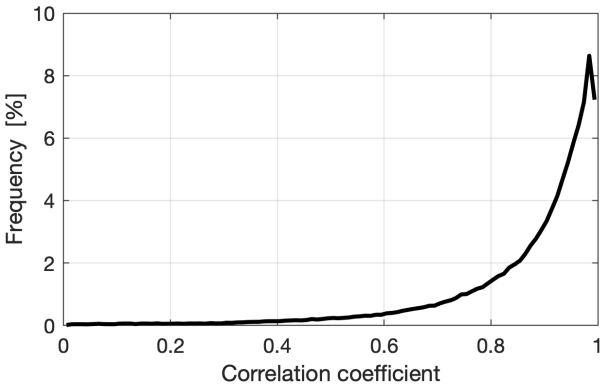


Figure 3-9: Histogram of the correlation coefficients for Sentinel-1 observations (year 2020, VH-polarization) acquired within the same month and grouped for the same orbital track.

The temporal averaging reduced the overall amount of data by a factor of 2 in regions with sparse Sentinel-1 coverage, and up to 5-6 in regions with frequent coverage. In terms of ENL, the improvement was small because the images that were averaged in time were strongly correlated. To obtain some global statistics of ENL, we worked at the level of a 1° × 1° grid cell and for each grid cell, we computed the 90th percentile of the canopy cover from the MODIS Vegetation Continuous Fields dataset. It was assumed that surfaces with a very dense canopy cover represent the most homogeneous type of land surface required to estimate the ENL. The ENL for a given grid cell was then computed with Equation (3-1) by selecting Sentinel-1 pixels corresponding to a canopy cover larger than the predefined threshold. This operation was repeated for each of the 12 monthly Sentinel-1 datasets and for each polarization. Since canopy cover is sparser at

northern and southern latitudes, our definition of the ENL estimate might have reproduced some geographic trends. Low thresholds imply large variability of the backscatter and therefore an ENL estimate that is smaller than values obtained in dense tropical forest. For this reason, we further restricted our computation to grid cells with a VCF threshold of 95%, corresponding to intact tropical forest regions and high-stock boreal forests.

Figure 3-10 shows an example of histograms of ENL for the Sentinel-1 VV- and VH-polarized July 2020 monthly average. The histograms have similar shape but peaks at somewhat different levels. The histograms did not change regardless of the month. The long tail of the histograms agrees with the more local results shown in Figure 3-5 and is probably related to the degree of "scatterer homogeneity" in a grid cell. Taking the mean value of the histograms in Figure 3-10 resulted in a global ENL of 150 for both the VV- and the VH-polarization. For simplicity, these numbers are used in the CCI Biomass CORE algorithm to quantify the precision of the Sentinel-1 images used as predictors. The implications of the ENL on the precision of the biomass estimates is further discussed in [RD-5]. Comparison with the results based on a single image in Figure 3-5 and with the preliminary estimate of 165 indicates that averaging only marginally improved the precision of the SAR backscatter measurements. This is a consequence of the small number of Sentinel-1 backscatter values used to form a monthly average and their strong correlation within one month.

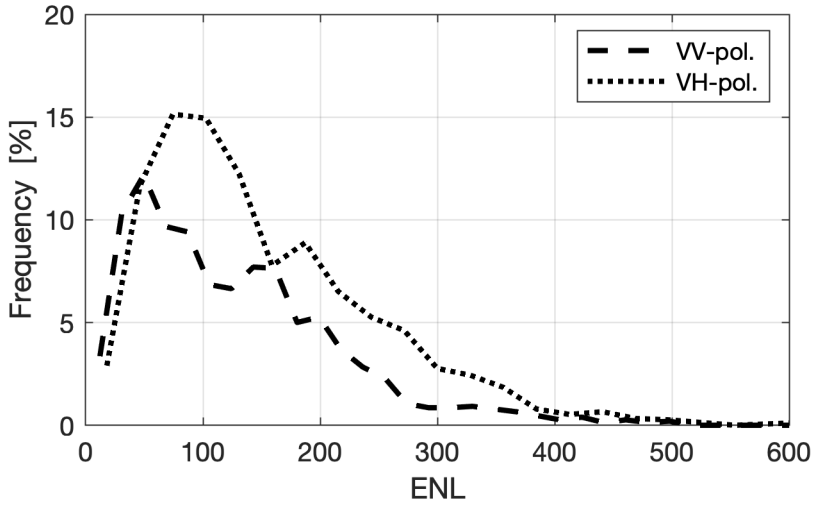


Figure 3-10: Histograms of the ENL estimates for the Sentinel-1 July 2020 monthly average.

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

The ALOS-2 mission started on May 24th, 2014, and carries an L-band SAR (PALSAR-2 instrument) with slightly better performance than its predecessor, ALOS-1 PALSAR-1. ALOS-2 PALSAR-2 operates in a high-resolution acquisition mode (25 m, Fine Beam, FB) and a moderate resolution mode (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 modes, PALSAR-2 acquires data in single polarization (HH) and dual polarization (HH and HV, or VV and VH over Japan), covering swaths of approximately 70 km in FB mode and 250 km in WB mode.

Because of the data policy applied by JAXA to ALOS-2 data, only a limited number of images can normally be obtained free of charge, which hinders global applications. Up to v3 of the CCI Biomass CRDP, global coverages of ALOS-2 PALSAR-2 data could 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 latter could be accessed thanks to the involvement of CCI Biomass team members in JAXA's Kyoto and Carbon Initiative (KC) (https://www.eorc.jaxa.jp/ALOS/en/kyoto/kyoto_link.htm). 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 WB mode is, however, operated at regional level, primarily across the tropics, thus not allowing annual global coverages. In support of the CCI Biomass project, JAXA provided access to all individual observations of ALOS-2 in Fine-Beam dual-polarization (HH&HV) mode. The data are provided in the form of ca. 3-degree long subsets of ALOS-2 orbits in detected format, i.e., backscatter only, and slant range geometry, a format referred to as "KC strip" because such data are prepared uniquely for members of the KC Initiative, aiming at mapping and monitoring forests and wetlands with L-band data. These data are used in the mapping of AGB starting with v4 of the CCI Biomass products.

The ALOS-2 PALSAR-2 dataset used by the current version of CCI Biomass CORE algorithm thus consists of

- Yearly mosaics of HH and HV polarized backscatter acquired in Fine Beam Dual (FBD) mode, years 2015- 2022
- Per-cycle mosaics of HH- and HV-polarized backscatter acquired in WB mode between 2015 and 2022
- Individual observations of ALOS-2 HH and HV polarized backscatter acquired in FBD mode, 2015- 2022 (referred to as KC strip data)

The mosaics were produced by JAXA (Shimada and Ohtaki, 2010; Shimada et al., 2014). 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. The annual FBD mosaics (HV-polarization) for 2015-2017 are shown in Figure 3-11. Both, the FBD and WB mosaics, initially produced by JAXA exhibited geolocation errors locally and were thus reprocessed in 2022. For a description of these errors, the reader is referred to the ATBD of the CCI Biomass v3 data products available at https://climate.esa.int/en/projects/biomass/key-documents/ (last accessed on 25 January 2023). For the CRDP based on this version of the ATBD, we use v2.1 of the ALOS-2 mosaic products.

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 3-12. In total, ScanSAR data acquired between 2015 and 2022 in 120 cycles have been released.

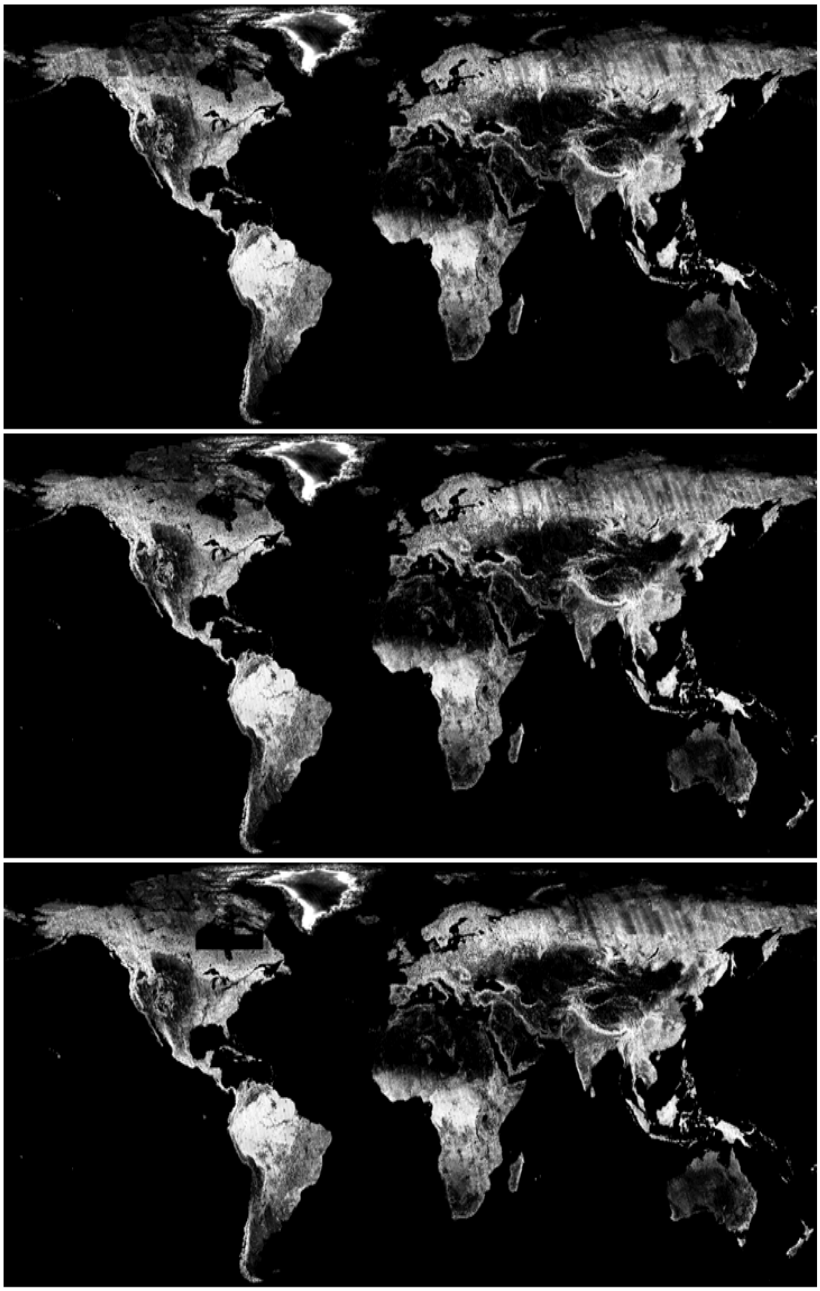


Figure 3-11: ALOS-2 FBD mosaics, HV-polarization, for the years 2015 (top), 2016 (middle) and 2017 (bottom).

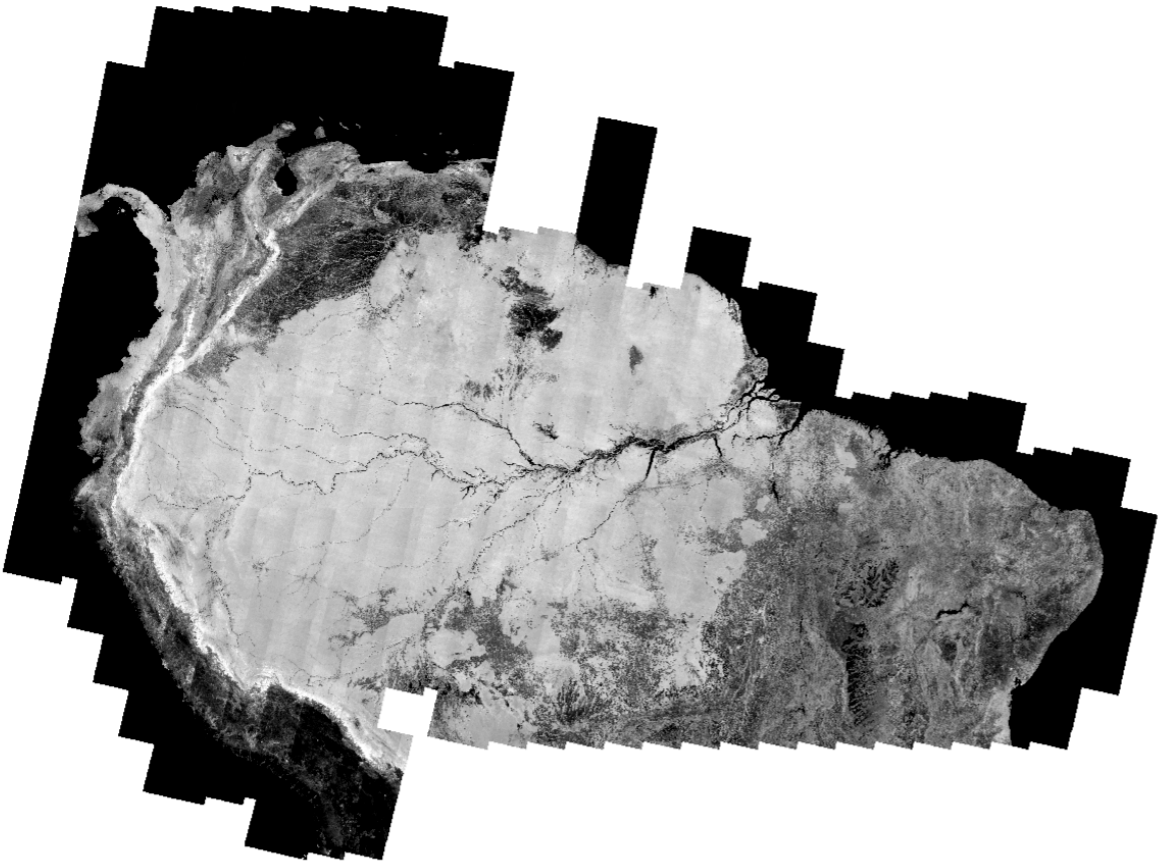


Figure 3-12: ALOS-2 ScanSAR mosaic generated from HV polarization imagery acquired in April 2018 over the Amazon Basin.

Each of the mosaics is provided in the form of 1°x1° tiles and includes the HH and HV backscatter (VV and VH over Japan) as well as:

- The local incidence angle with respect to the orientation of the pixel, derived from the DEM used by JAXA for the pre-processing (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 FBD and ScanSAR datasets were geocoded, orthorectified and calibrated by JAXA. The mosaics were also 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).

To reduce the speckle in the ALOS-2 PALSAR-2 imagery processed by JAXA to a resolution higher than the target resolution for the mapping of biomass, all images were aggregated to the pixel size of 100 m (0.00088888°). A first estimate of the ENL of the imagery was assessed for several homogenous forest patches, identified by means of visual image interpretation. We found the ENL to be of the order of 70 to 80.

For the ALOS-2 KC strip data, a pre-processing chain was implemented based on the GAMMA software (Figure 3-13). Pre-processing aimed at producing fully calibrated and terrain-corrected backscatter data at 000088888° resolution (EPSG:4326), i.e., the target resolution of CCI Biomass products. The processing comprises compensation for the Noise Equivalent Sigma Zero (NESZ) based on range gradients of backscatter identified over smooth water surfaces (Figure 3-14), topographic corrections according to Frey et al. (2013), and a verification of the geocoding accuracy for each individual scene (on average a few tenths of the pixel size). Examples of annual composites (produced for illustration purposes only) for the years 2019 and 2020 are shown in Figure 3-15, together with maps depicting the number of individual backscatter observations available from the KC strip datasets annually.

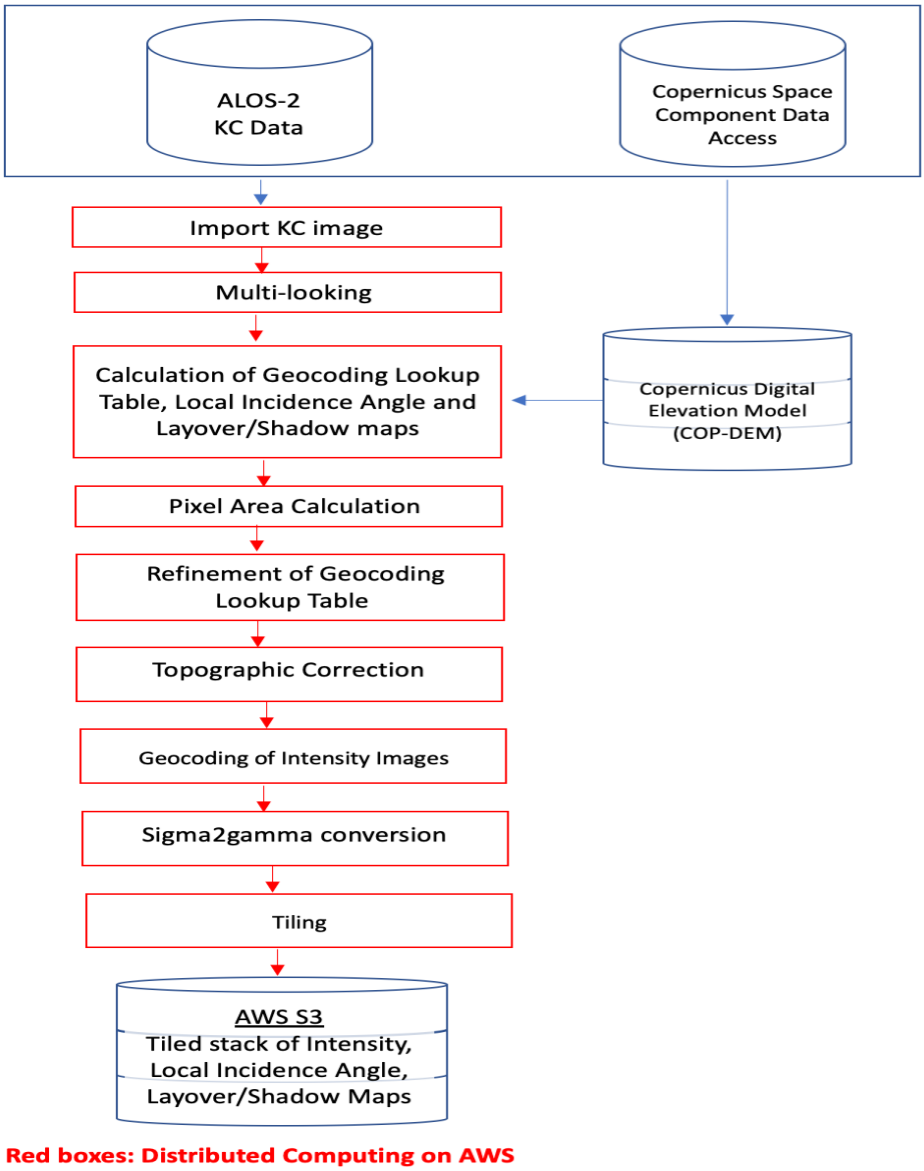


Figure 3-13: *Processing workflow for the ALOS-2 KC strip data.*

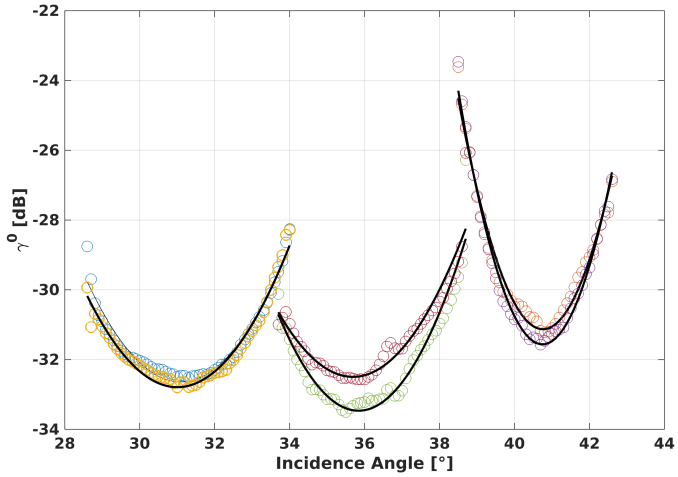


Figure 3-14: *Noise floor of ALOS-2 Fine-Beam HV polarization backscatter imagery as observed in several images acquired over smooth water surfaces.*

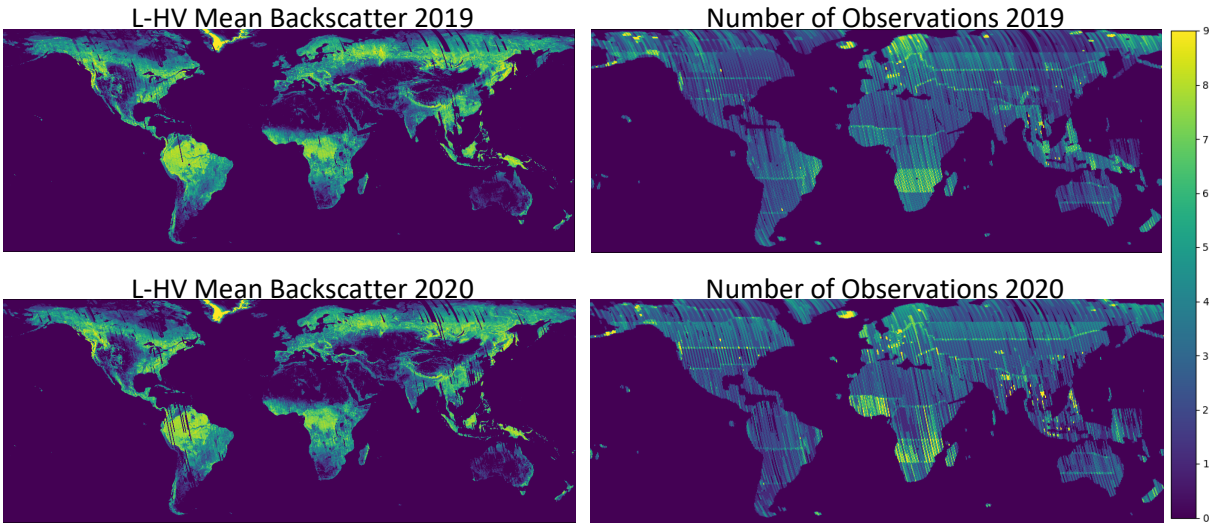


Figure 3-15: *Annual mean L-HV backscatter calculated from all ALOS-2 Fine Beam dual-polarization KC strip data acquired in 2019 (top left) and 2020 (bottom left).*

The ALOS-2 PALSAR-2 backscatter data exhibit significant striping, particularly in areas with continuous forest cover, such as the Amazon or Congo Basin. The striping, which is clearly visible in the mosaics over continuous tropical rainforest (e.g., Figure 3-12), is associated with the incidence angle dependence of backscatter as well as uncertainties in the calibration of the backscatter imagery. In the tropics, the striping is more apparent than elsewhere because of the continuous forest cover and the low sensitivity of backscatter to forest density or biomass. Methods aiming at reducing the incidence angle dependence need to consider that locally it differs between land cover classes that 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 was 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.

As an interim approach, we opted to minimize the striping effects observed over dense tropical rainforest using an empirical normalization approach in which trends of the backscatter over closed tropical forest canopies (as indicated by a Landsat canopy density map in the Easting direction, i.e., roughly the range dimension in the SAR imaging geometry) were compensated by:

1) Identification of trends in L-HV backscatter using a line-by-line moving window median filter of 100 pixels window length,
2) Normalization of the ALOS-2 PALSAR-2 backscatter by subtracting line-by-line the median trend from the actual observations and adding the residuals back to the average backscatter observed over closed tropical forest canopies.

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

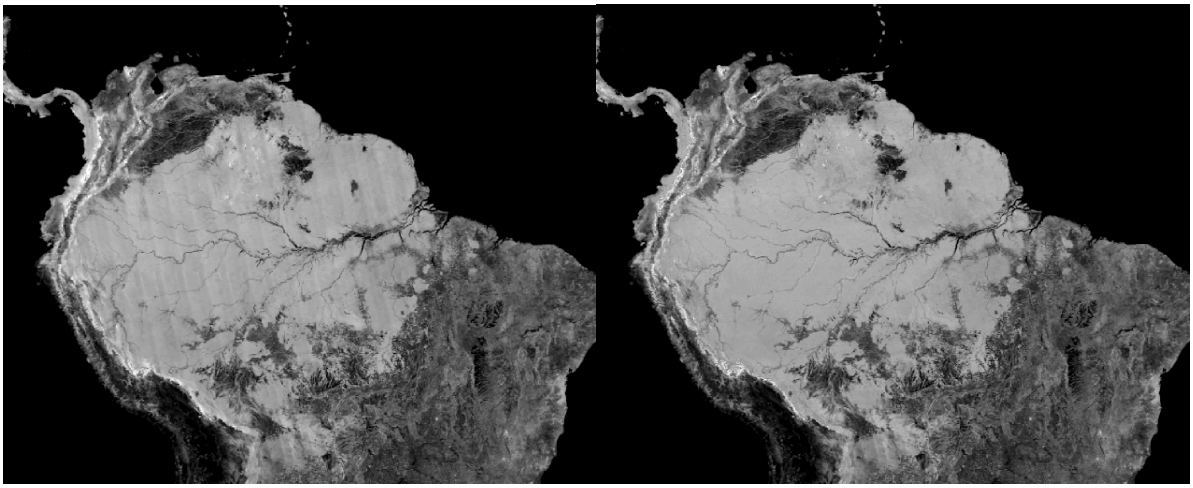


Figure 3-16: ALOS-2 L-HV mosaic before (left) and after (right) normalization.

For both FBD and ScanSAR datasets, the ENL was computed in a more systematic manner following the approach applied to the Sentinel-1 data. The box plot in Figure 3-17 shows that the ENLs for the ALOS-2 PALSAR-2 datasets were lower because of the smaller multi-look factors and the lower radiometric quality. For the FBD dataset, the median value of the ENL was 47 (HH-pol.) and 50 (HV-pol.) whereas for the ScanSAR data we obtained a value of 19 regardless of the polarization. The impact of these estimates of the ENL on the precision of the AGB obtained from the ALOS-2 data is discussed in the E3UB report [RD-5].

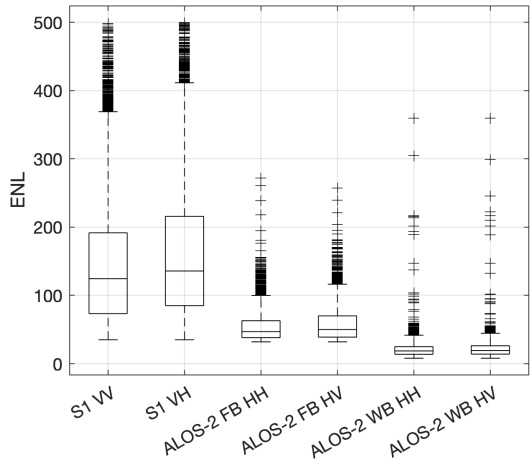


Figure 3-17: Box plots representing the distribution of the ENL estimates from 1° × 1° grid cells per SAR sensor, mode and polarization. Each box shows the median value (central mark), the interquartile range (edges of the box), the most extreme data points not considered outliers (whiskers) and the outliers (plus markers).

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 (GMM) mode, which was operated as a background mission when the instrument was not meeting other requests, also had a 405 km swath but with 1 km resolution. From the 10 years of acquisitions, the ASAR archive contains acquisitions in one specific mode for virtually anywhere on Earth.

Figure 3-18 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. We omit all other years because of similar spatial patterns. Near-global, dense datasets of backscatter were obtained in GMM. In contrast, 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. Aggregating all ASAR acquisitions led to global coverage except for some islands in Oceania and isolated islands in the southern hemisphere.

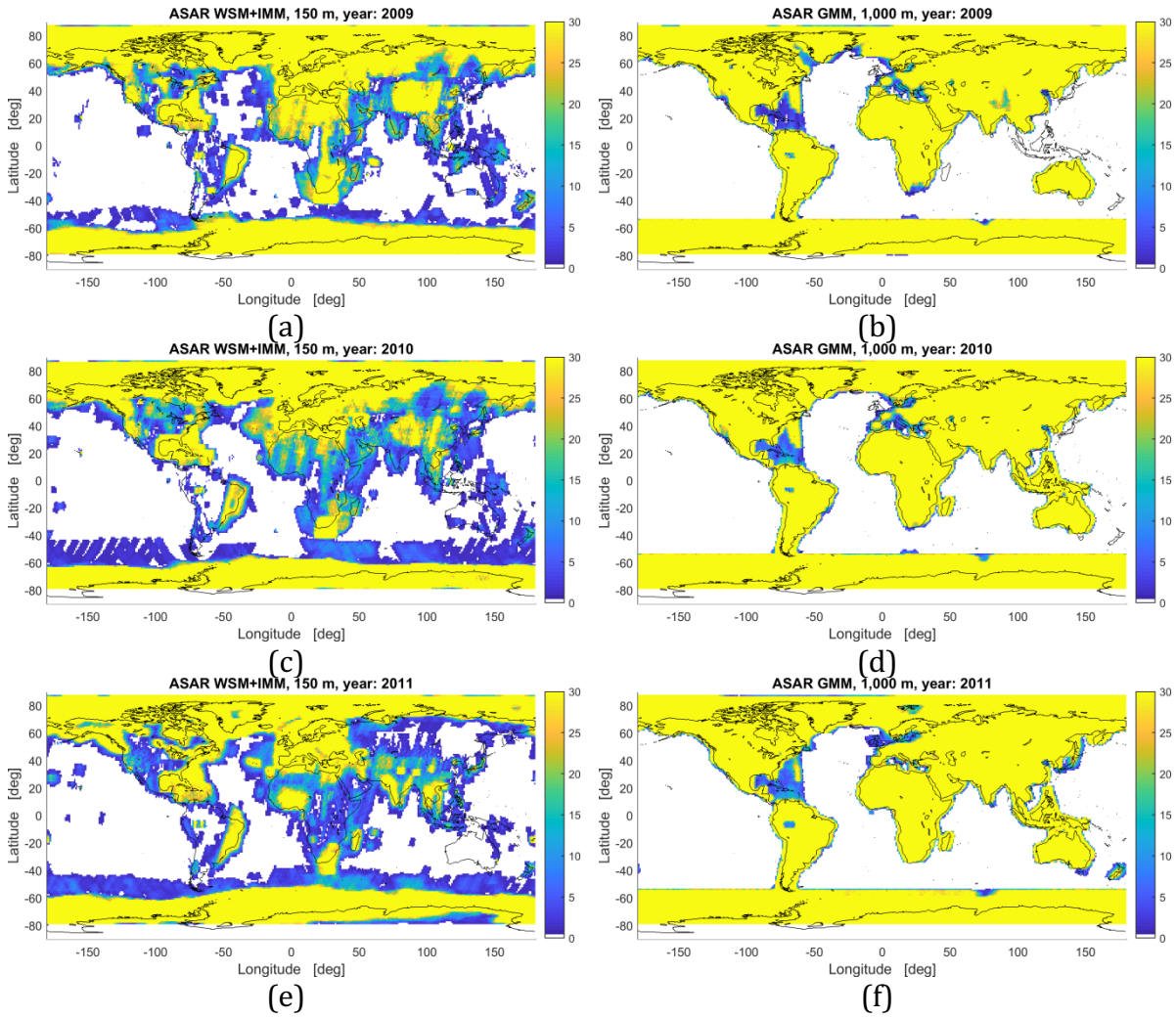


Figure 3-18: Map of ASAR acquisitions at 150 m (WSM and IMM combined) and 1,000 m (GMM) per 1°×1° grid-cell for each year between 2009 and 2011. The colour bar has been constrained between 0 and 30 observations to increase the image contrast in regions of poor coverage.

All ASAR data acquired in IMM, WSM and GMM over land were pre-processed in past research projects by Gamma Remote Sensing to form stacks of co-registered images of SAR backscatter (Santoro et al., 2015a and 2015c). 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 3-19) (Santoro et al., 2011; Santoro et al., 2015a). 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 same tiling grid was used for the pre-processing of the Sentinel-1 data. The GMM images (spatial resolution approximately 1 km×1 km and pixel size 500m×500m) 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).

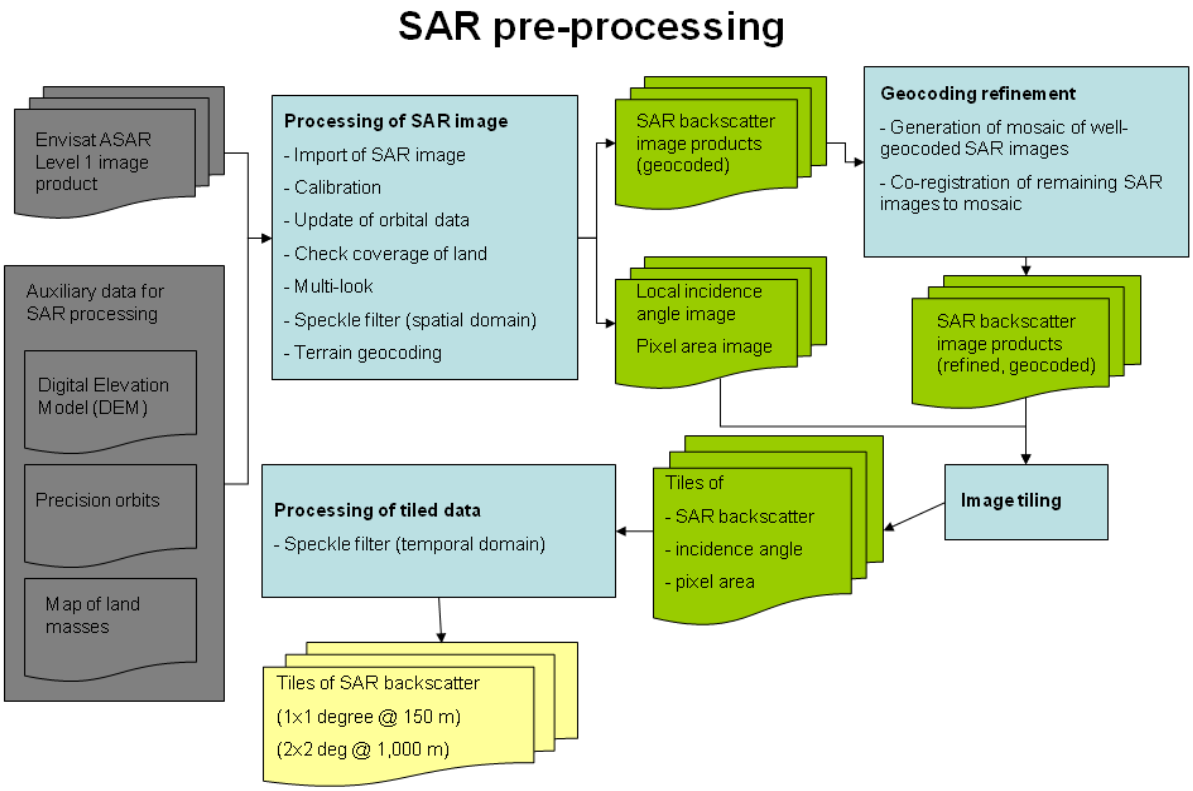


Figure 3-19: Flowchart of ASAR pre-processing.

The ASAR dataset consisted of images of 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, last accessed on 2 October 2024). At this stage, the coverage of the ASAR image was checked against a coarse resolution map of landmasses; images acquired over the ocean were discarded. Each remaining image was 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 (Santoro et al., 2015c). 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-coefficient polynomial; the resulting 2-D model of offsets was finally applied to the LUT 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, e.g., corresponding to predominantly flat regions, 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.

The ASAR backscatter images 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 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 eventually 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 originated 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 was not affected by the different size of the pixel for different latitudes.

As shown in Figure 3-18, full global coverage with ASAR data could be obtained only by merging the GMM with the IMM and WSM datasets. Since the number of observations from the GMM was much larger than those acquired at moderate resolution, it made sense to achieve the full 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-looking 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 that could not be accommodated at the time of the ASAR image processing. As a result, 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 a predictor in the BIOMASAR-C algorithm to guarantee wall-to-wall coverage and increase the reliability of the biomass 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 version 1 of the CCI Biomass CRDP in regions where the contribution of C-band was strong (e.g., tundra regions of Central Asia) indicated higher quality and level of detail in the latter.

For this reason, we reconsidered the use of ASAR data in the context of CCI Biomass to map AGB around 2010. Although the coverage at 150 m was not optimal for unbiased estimates of AGB in large parts of the world, the 1,000 m dataset would probably have decreased the effective resolution of the map product, introducing local biases due to the rescaling from 1,000 m to the target pixel size of the CCI Biomass maps of 100 m.

Regardless of the final spatial resolution, slope-induced distortions need to be accounted for to avoid artefacts in the AGB maps due to 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):

$$\gamma^0 = \sigma^0 \frac{A_{flat}}{A_{slope}} cos\theta_{loc}{}^{-n} \qquad (3\text{-}2)$$

In Equation (3-2), θ_{loc} represents the local incidence angle. 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). The exponent n expresses the variation of scattering mechanisms 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 mosaic processing sequence. 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 ScanSAR data (GMM and WSM) is the sub-optimal inter-calibration of the swaths forming an image. This results in an offset of the backscatter across the seam between two adjacent swaths, typically of the order of a few tenths of a dB. This issue is critical in environments where the backscatter difference between unvegetated terrain and dense forest is 1-2 dB. Figure 3-20 shows an example for an ASAR GM1 image covering the rainforest of the Brazilian Amazon. The area covered by the image corresponds to very dense tropical forest (percent tree cover from the MODIS Vegetation Continuous Fields dataset > 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 radiometric offsets becoming 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 FAO GEZ dataset. Figure 3-21 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 of minor importance compared to producing a map with artefacts, given that C-band should not be able to estimate biomass with high accuracy in high biomass tropical forest.

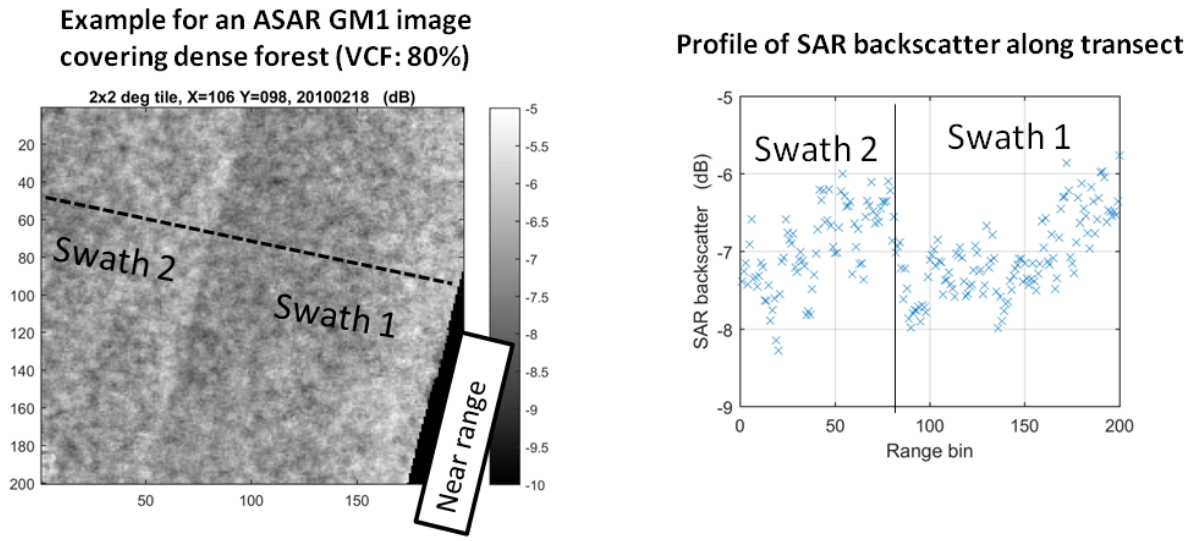


Figure 3-20: Illustration of backscatter offset along the seam between two adjacent swaths of an ASAR GM1 image covering dense tropical forest in the Amazon (left panel). The profile of the backscatter along the dashed line superimposed to the ASAR image is showed in the panel on the right hand-side.

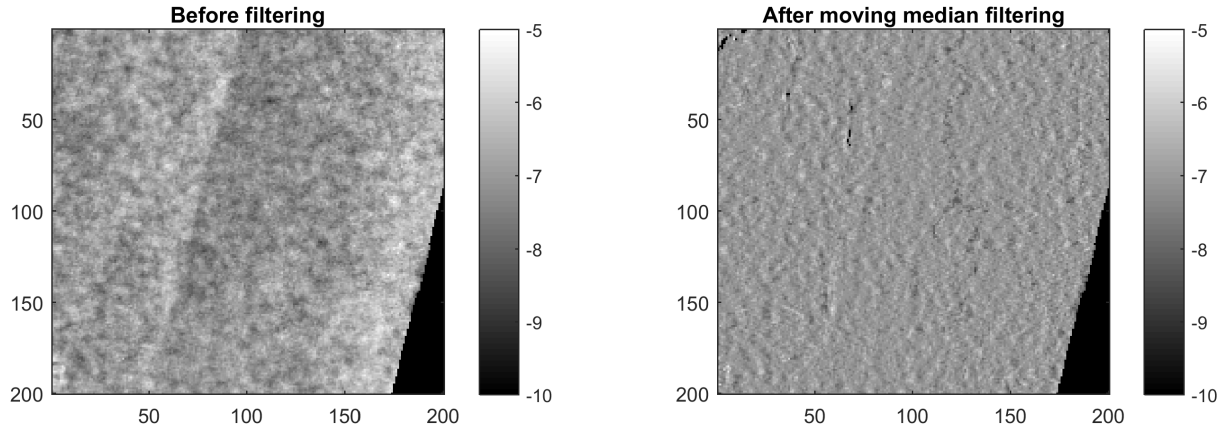


Figure 3-21: Example of the ASAR GM1 used in Figure 3-20 before and after filtering with a moving median filter.

3.4 ALOS-1 PALSAR-1

The ALOS-1 mission operated between 2006 and 2011 with the PALSAR-1 radar instrument onboard. Coverages were achieved both at high resolution in the Fine Beam mode (FB, 25 m) and at moderate resolution with the Wide Beam mode (WB, 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.

Until 2022, the data policy applied by JAXA to ALOS-1 and ALOS-2 data allowed a limited number of images to be obtained free of charge, which hindered global applications. As for ALOS-2, global coverages of ALOS-1 PALSAR-1 data could therefore only be obtained in the form of yearly backscatter mosaics for the FB mode and per-cycle mosaics (46 days) for the WB mode. Starting with v4, access to the archive of ALOS-1 PALSAR-1 images was possible.

For the processing of the data until CRDP v3, it is referred to previous versions of this document. Beginning with v4 of the CCI Biomass CRDP, the release of all available ALOS-1 PALSAR observations acquired in FB single- and dual-polarization modes by JAXA to the public in Level 2.2 format allowed the multi-temporal coverage acquired by ALOS PALSAR to be explored (Figure 3-22). The data comprises fully calibrated, terrain-corrected, and geocoded backscatter images in UTM projection with 12.5 m resolution. Since already fully pre-processed, post-processing of the data for CCI Biomass was limited to reprojection to the geographic coordinate system (EPSG: 4326) and aggregation to the target resolution of 0.00088888°. The ENL of the backscatter imagery was estimated to be of the order of 80.

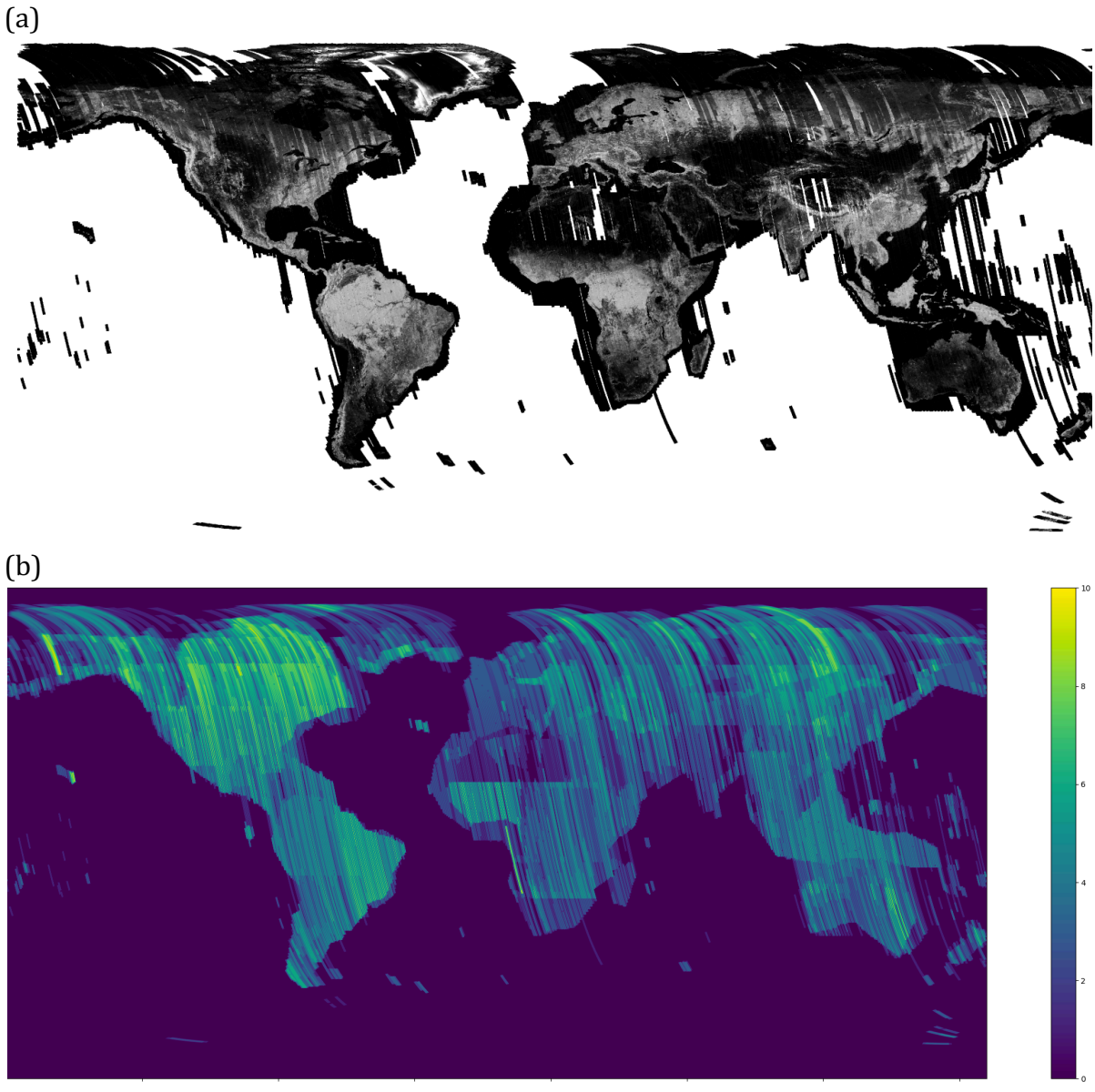


Figure 3-22: ALOS-1 PALSAR-1 HV backscatter mosaic produced from images acquired in 2010 and released by JAXA in Level 2.2. format.

3.5 Spaceborne LiDAR

3.5.1 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 3-23. Depending on the height and structure of the forest, the waveform will exhibit 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.

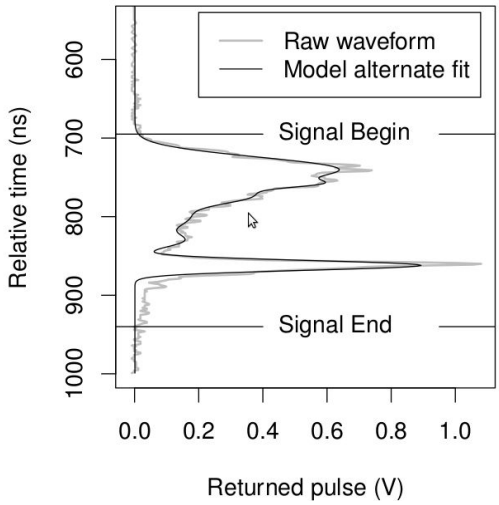


Figure 3-23: An ICESat GLAS waveform showing the vertical distribution of returned energy from a forest (from Los et al., 2012).

In CCI Biomass, we use the dataset prepared in the GlobBiomass project (Santoro et al., 2021). We used 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 3-24).

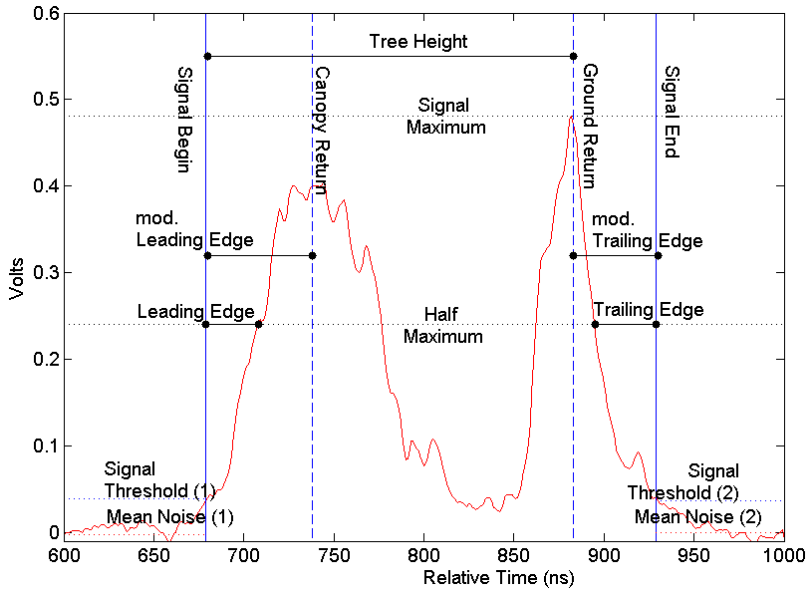


Figure 3-24: An ICESat GLAS waveform, showing the waveform metrics used to calculate RH100 & CD (Hilbert &Schmullius, 2012)

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. The remaining GLAS database contained estimates of RH100 for ca. 26.5 million footprints.

Figure 3-25 shows the spatial coverage of canopy height and canopy density expressed as average values from the footprint-level data in 0.125° large grid cells. Because of the rather small number of footprints retained after filtering, we created one single layer for each metric from all the footprint level data. The number of footprint-level observations used to compute the grid cell average values is also shown in Figure 3-25. Because of the large distance between the orbital tracks and the strong filtering applied to the GLA14 dataset, the coverage was not complete. The density of observations presented some spatial variability; larger gaps corresponding to forests occurred in South China and the southernmost regions of South America, Africa and Australia.

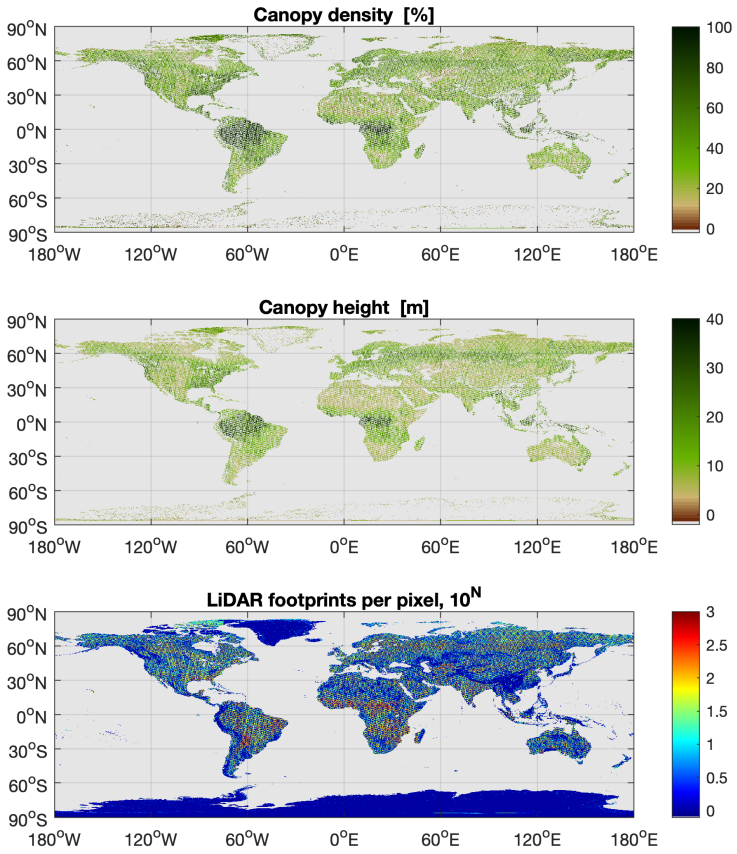


Figure 3-25: Maps of canopy density, canopy height (RH100) and corresponding number of GLAS footprints within each grid cell. For the display, the original heights were averaged to a pixel size of 0.125°.

3.5.2 ICESat-2

Unlike the GLAS sensor, the Advanced Topographic Laser Altimeter System (ATLAS) onboard the ICESat-2 satellite uses photon counting to retrieve elevation. ATLAS sends pulses of laser light to the ground, collects photons reflected by the surface underneath and times each photon return. With a frequency of 10,000 pulses per second, ATLAS achieves a much denser portrait of the surface compared to the 40 pulses used by GLAS. Consecutive shots are separated by 70 cm, which reduces the ambiguity of the surface vs. vegetation reflections and the impact of topography on the reflected signal (Neuenschwander and Pitts, 2019). Together with very accurate timing of the photons, these two features enable accurate mapping of the Earth's topography. In addition, it enables profiling of vegetation, even though the measurement technique is strongly affected by the power recorded by the instrument. ATLAS splits the laser into six beams arranged as three pairs of beams approximately 3.3 km apart (Figure 3-26) (Markus et al., 2017). Each pair consists of a strong and weak energy beam (4:1 ratio). Combination of the strong and weak beam returns allows better characterization of surface topography (Neuenschwander and Pitts, 2019). However, for vegetation studies, it is advised to avoid measurements corresponding to weak beams because of the partly undetected vegetation layering in the returned signals.

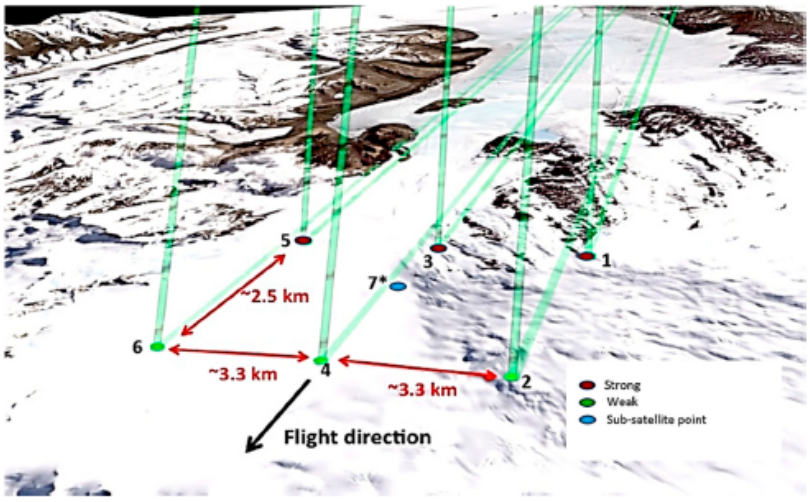


Figure 3-26: Configuration of the ICESat-2 observations (Neuenschwander and Pitts, 2019).

For land and vegetation, a specific product has been developed (ATL08) (Neuenschwander and Pitts, 2019) that contains geophysical parameters related to vegetation and terrain heights. The ATL08 algorithm estimates the ground surface and top of canopy surface elevations from the photons, from which several parameters of relative height are computed. From an investigation in boreal forests, it was understood that the RH98 (relative height, 98 percentile) corresponds to canopy height and that seasonal conditions (e.g., snow on the ground) influence the height estimates. In addition, the properties of the canopy height retrieved in the dense tropics may be erroneous, in particular when using data from the weak beam, because of the very small number of photons recorded from the forest floor (Neuenschwander and Pitts, 2019). Also, persistent cloud cover hinders the acquisition of a sufficient number of photons to pass the set of quality filters implemented in the data processing algorithms.

The ATL08 product provides the parameters with a 100 m step size along the flight direction. Currently version 6 of the product is available from the National Snow and Ice Data Center (NSIDC) (https://nsidc.org/data/atl08 , last access on 02 October 2024) in the form of strips of photons collected along one orbit. ICESat-2 data have been available since 14 October 2018. To obtain segments from the original photon data, the original files are reformatted with the pysl4land Tool, a set of Python tools to process spaceborne lidar (GEDI and ICESat2) for land (pySL4Land) applications (https://github.com/remotesensinginfo/pysl4land, last access on 02 October 2024). Herewith, the original photons are grouped into segments of 100 m length and 25 m width.

For this version of the ATBD, ICESat-2 files covering the years 2019-2022 were used. Data were grouped on an annual basis and treated separately. It is foreseen to extend the time series of observations in the next versions of the document to reinforce the statistics derived from the individual measurements. From the segment-wise data, we generated maps of canopy height (RH98) with several grid spacings depending on the use (inter-comparison of LiDAR-based canopy heights, prediction of AGB from canopy height, etc.). For each grid cell, the canopy height was obtained by taking the mean value of the original values. Prior to averaging, the segments were filtered following indications by the producers of the ATL08 data product. Segments were discarded that were (i) acquired by the weak beam, (ii) characterized by less than three photons reflected by the canopy, (iii) flagged as not belonging to natural vegetation in the ATL08 metadata

and (iv) exhibited an elevation that differed by more than 25 m from the reference DEM used in all ATL products.

A detailed analysis of the spatial distribution of canopy heights and their fluctuations in time revealed several locations with unrealistically high values in one or other of the three years of data. Starting with this version of the ATBD, segments corresponding to a terrain slope steeper than 10° were discarded because they were often characterized by unnatural values.

After such filtering, we still identified several locations with unusually high values. These occurred in areas with a relatively low number of photons per segment, as shown by the example in Figure 3-27. Setting a threshold on the minimum number of photons reduced outliers. However, too strong filtering caused gaps and strong underestimation of canopy height (see cases for >40 and >50 photons). This is because tree canopies scatter fewer photons than the surface underneath. By deselecting segments with a small number of photons, we discard segments where photons come exclusively from the canopy and privilege ground returns. Visual investigation of the canopy height maps based on different number thresholds initially indicated that a threshold of 20 photons per segment was a reasonable compromise between outlier reduction and accuracy (used for CRPD v5). Considerable loss of segments in tropical rainforest suggested a correction of the threshold to 10 photons per segment for the CRDP v6. Outliers were filtered when generating the spatially explicit dataset of canopy height metrics.

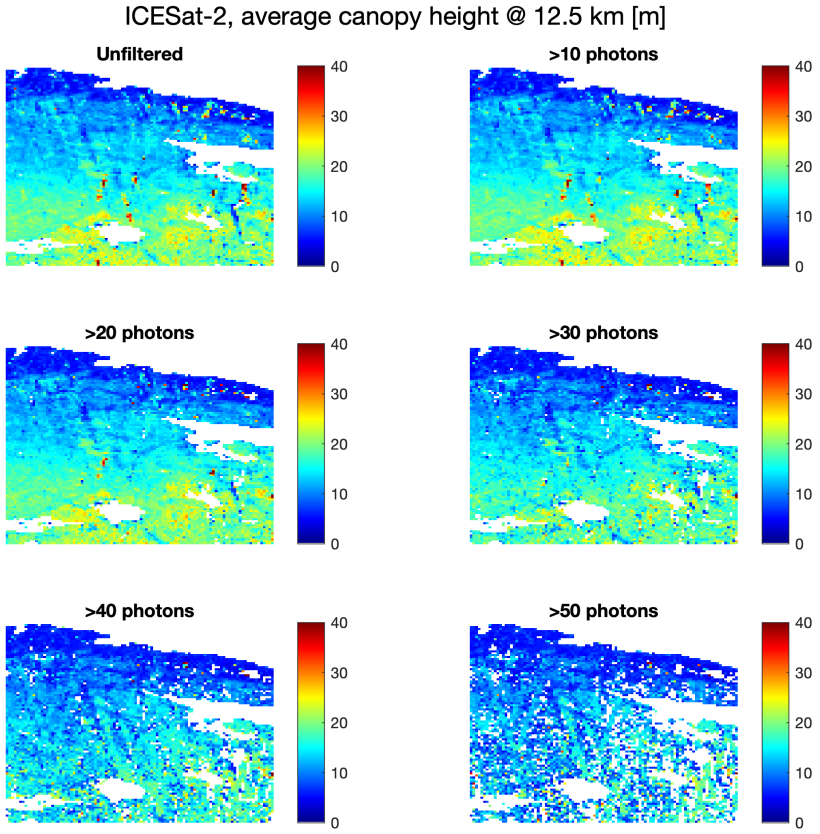


Figure 3-27: Canopy height maps from ICESat-2 segments averaged at 12.5 km and filtered for different thresholds on the minimum number of photons per segment. The example covers the Karelian region between Finland and Russia.

Figure 3-28 shows an example of average canopy height values based on averaging at 0.1° and the corresponding number of segments used to estimate the average height per grid cell. With three years of data gave homogeneous coverage of the world's forests. The spatial patterns correspond to the known distribution of forest height, with taller forests in the wet tropics along the Equator (Amazon basin, Congo, and Southeast Asia) and in temperate forests (Pacific Northwest, Southeast U.S., Central Europe, Andes, and Southeast Australia). In boreal forests (north of 60°N), the canopy height decreased with increasing latitude. The number of segments per grid cell was high for boreal and temperate forests. In tropical forests, the persistent presence of clouds and the strong filtering on the minimum number of photons explain the sparser coverage. With the availability of a few more years of observations, it likely that one will be able to relax on the condition applied to the number of photons segments since point-wise noise will be more automatically flagged.

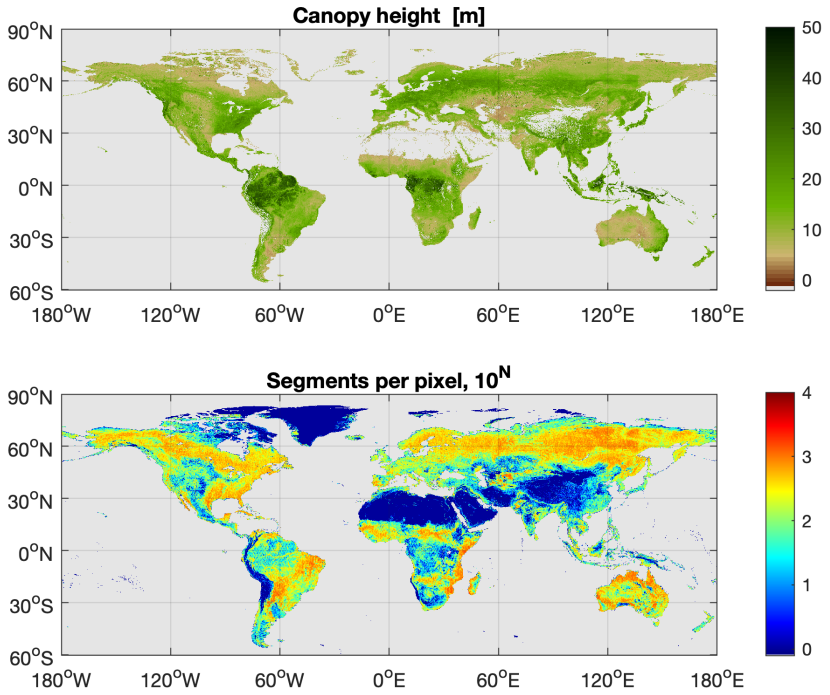


Figure 3-28: Global distribution of canopy height estimated from the ICESat-2 ATL08 dataset of 2019-2022 (top panel), and corresponding number of segments retained to form the average canopy height (bottom panel). Dark blue areas in this panel correspond to data gaps. For the display, the original heights were averaged to a pixel size of 0.1°.

Comparison of canopy height values for two consecutive years (e.g., 2019 and 2020), shows very high consistency (Figure 3-29). The correlation coefficient between the two datasets was 0.79 and the mean difference between the two years was less than 0.1 m. Nonetheless, Figure 3-29 shows some spread, which indicates temporal variability of the average values (standard deviation of 3-4 m) due to the sampling.

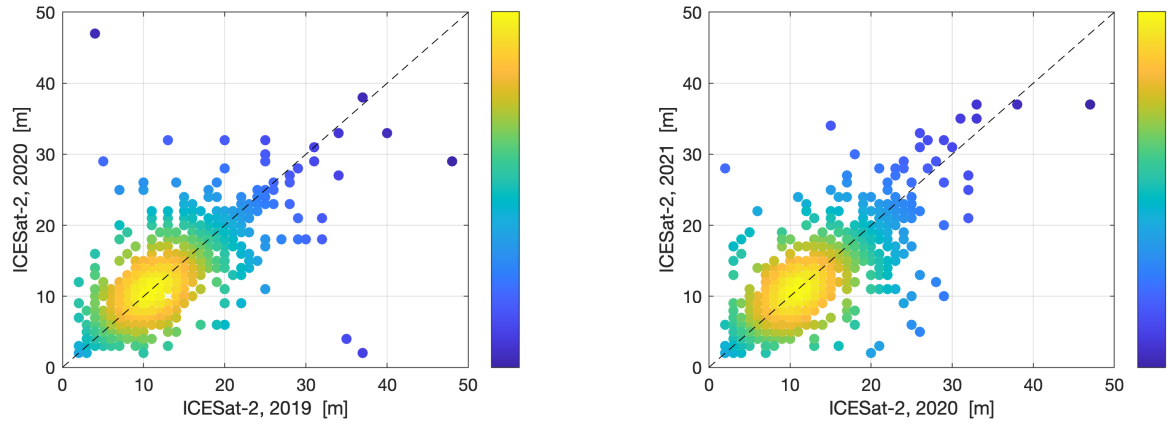


Figure 3-29: Scatterplot comparing canopy height averages at 0.1° for ATL08 data acquired by ICESat-2 in 2019, 2020 and 2021. The density plot is colour coded, with colour changing from blue to yellow for increasing density of data points.

Although not shown here, we also created canopy height rasters by splitting the data for day- and night-time acquisitions and/or summer- and winter-time acquisitions. The purpose of using a dataset acquired in different seasons was to understand whether seasonal conditions affect the values of the canopy height metrics. For leaf-on conditions (summertime in the northern hemisphere and wintertime in the southern hemisphere), there was a slight tendency towards higher values, the difference being however only a fraction of a meter. Separating canopy heights based on daytime or night-time acquisitions did not reveal any difference, regardless of the period of the year considered.

3.5.3 GEDI

The GEDI instrument (Dubayah et al., 2020) is a full waveform LiDAR installed on the International Space Station (ISS) and observes land masses between +/-52° latitude. The size of the footprint is smaller than for ICESat GLAS (25 m vs. 70 m diameter) and the density of observations is higher. The system consists of one laser split into two beams ("coverage" beams), and two lasers operating at full power ("power" beams). Operation of GEDI leads to 8 parallel tracks, separated by about 600 m across track. Along each track, footprint centres are separated by 60 m. The distance between adjacent orbital tracks was about 1 km until January 2020 (Figure 3-30), after which it was increased to 70 km resulting in sparser sampling.

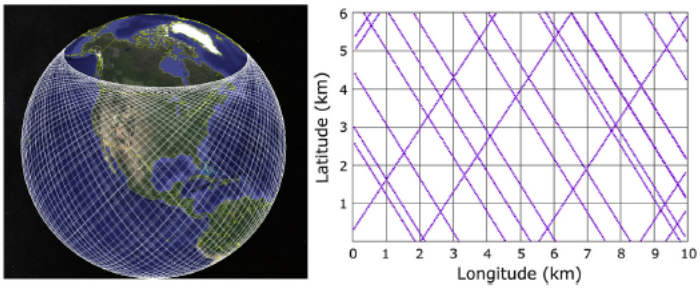


Figure 3-30: GEDI orbital tracks (2 weeks) as illustrated in (Dubayah et al., 2020).

From the waveform data, several height metrics, including canopy height (defined as Relative Height 100) and canopy density are obtained. These level 2A (height metrics) and 2B (canopy density) data are provided at the level of individual footprints. Currently version 2, with data starting on 25 March 2019, is available (https://lpdaac.usgs.gov/products/gedi02_av002/, last access on 2 October 2024). To be able to ingest the GEDI data into the environment of the CORE algorithm, data from individual orbital files were reformatted with the pysl4land Tool. For this version of the ATBD, all data acquired in 2019 and 2020 were considered. The original footprint-level data were reformatted with the pysl4land Tool and filtered for the quality flags of the level 2A and 2B products reported in the metadata. In total, 357 and 470 million footprints were retained for 2019 and 2020, respectively. From these, we have generated yearly averages, as well as wintertime and summertime averages per year.

Figure 3-31 shows the spatial distribution of canopy density, canopy height and number of footprints per 0.1° grid cell for 2020. The spatial distribution of canopy height and canopy density is similar to those obtained for the ICESat GLAS and the ICESat-2 data. The coverage presents systematic gaps due to the large spacing between orbits although in regions covered by GEDI footprints the number of footprint-level data is higher. This is a consequence of the finer resolution of GEDI compared to the other spaceborne LiDAR instruments.

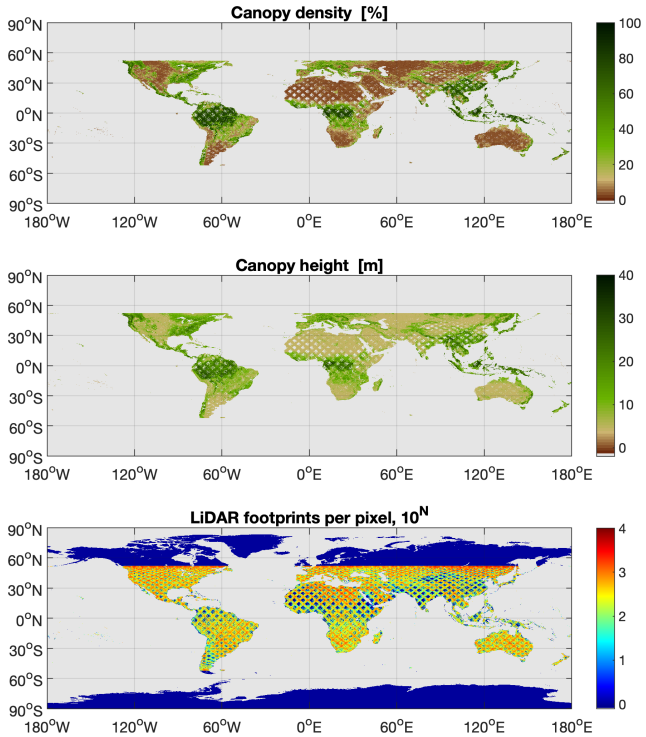


Figure 3-31: Maps of average canopy density, canopy height (RH98) and corresponding number of GEDI footprints at 0.1° for 2020. Dark blue areas in the bottom panel correspond to data gaps.

The analysis of canopy density and canopy height split by seasons reveals higher values of canopy density during leaf-on conditions (April/September in the northern hemisphere and October/March in the southern hemisphere) in the intermediate range and comparable values of canopy height (Figure 3-32). We do not discuss these results further as the GEDI canopy cover observations have not yet been implemented in the CORE algorithm.

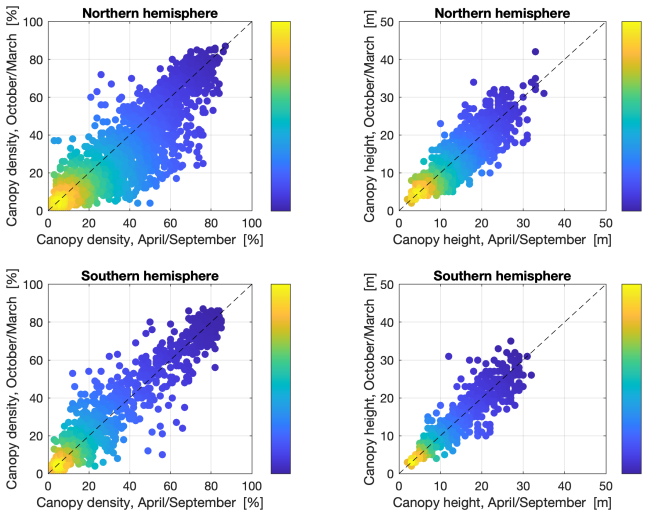


Figure 3-32: Scatterplots comparing canopy density (left panels) and canopy height (right panels) estimated from GEDI data during summer (April/September) and winter (October/March) months for the northern (top panels) and southern hemisphere (bottom panels). Each density plot is colour-coded, with colour changing from blue to yellow for increasing density of data points.

3.5.4 Comparison of LiDAR-based metrics of canopy height

Individual footprint-level data are not used in the CORE algorithm because there is hardly any coincidence of the footprints with the area covered by forest inventory plots where biomass is measured. For this reason, it is not possible to construct reliable models that would capture the spatial variability of the association between canopy height and biomass. The best spatial coverage of AGB can either be obtained from maps or from statistics published by National Forest Inventories (NFI) at sub-national or national level. Although each of these AGB sources has its own errors, they can be used to characterize the shape of the curve relating AGB and canopy height. It is therefore of interest to relate statistics of AGB with similar statistics derived from spaceborne LiDAR data.

While the datasets produced with observations by the new spaceborne LiDAR missions are more complete in terms of spatial coverage compared to ICESat GLAS, they are undersampled in several regions of the world. For this reason, it is necessary to understand the reliability of the GEDI and ICESat-2 canopy height averages. In this exercise, we compare 0.1° averages of canopy height from the three datasets, noting the ICESat GLAS averages are not contemporary with the GEDI and ICESat-2 based values (epoch 2005 vs. 2020). Figure 3-33 shows density plots of ICESat-2 and GEDI with respect to ICESat GLAS canopy heights. While the ICESat-2 values appear to be slightly overestimated in low canopies, the GEDI values are systematically underestimated regardless of the canopy height level.

While we cannot exclude forest cover changes as one of the reasons for the discrepancy between the three datasets of canopy heights, it is puzzling to observe divergent trends for the recent spaceborne LiDAR missions. Although we applied all filters suggested by the production teams to reduce the number of incorrect observations and further kept footprints that were within areas labelled as forest according to the CCI Land Cover dataset, there may be an aspect that we have overlooked in our comparison, given that our results differ from published literature suggesting superior accuracy of the GEDI data compared to ICESat-2. To help clarify this issue, the results of the inter-comparison have been shared with the GEDI and the ICESat-2 teams.

Given the global coverage and the strong agreement of the ICESat-2 canopy height dataset with the ICESat GLAS data, which was used in previous versions of the CCI Biomass CRDP, we decided to use the ICESat-2 dataset for this version of the ATBD and for the CRDP version 5.

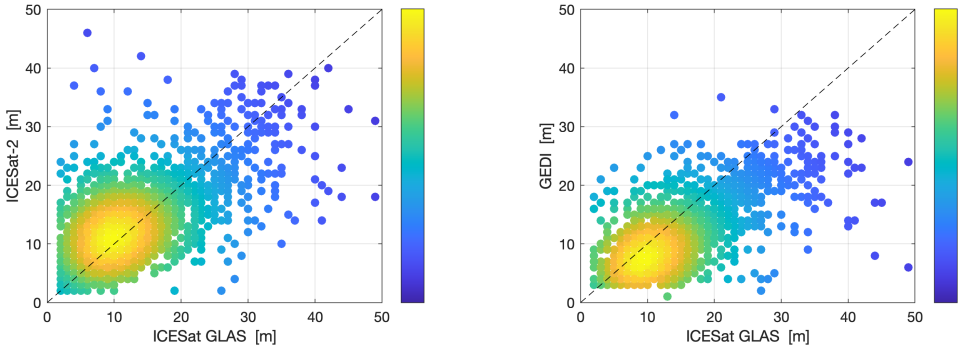


Figure 3-33: Density plots comparing 0.1° averages of canopy height from the ICESat GLAS (2003-2009, horizontal axis), ICESat-2 (2020, vertical axis, left plot) and GEDI (2020, vertical axis, right plot).

3.6 Digital Elevation Model

3.6.1 SRTM-based DEM

Until v3, we used an SRTM-based DEM (Figure 3-34) in the pre-processing phase of the SAR data (ENVISAT ASAR, Sentinel-1 acquired in 2017 and 2018) and the Lidar data (ICESAT GLAS) and to analyse the quality of the retrieved biomass. The SRTM-based DEM consisted of a global dataset of 1° × 1° 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 each region (de Ferranti, 2009).

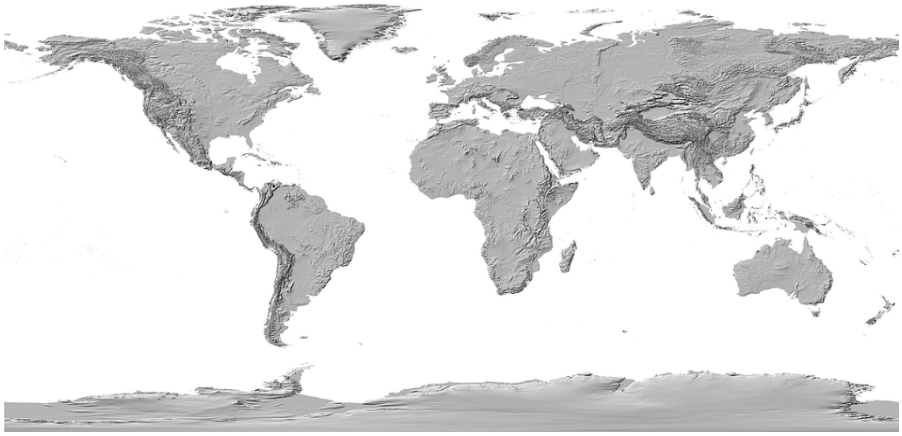


Figure 3-34: Global seamless DEM based on elevation datasets available at http://www.viewfinderpanoramas.org.

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. Despite some inaccuracy, this DEM was the most reliable option for pre-processing and analysis until the release of the Copernicus DEM.

3.6.2 Copernicus DEM

While the Sentinel-1 data acquired in 2017 and 2018 had been processed with the SRTM-based DEM described above, the Sentinel-1 data for the years 2019 and 2020 as well as the ALOS-2 KC strip data have been processed (ortho-rectification, geocoding, topographic corrections) with the more recent Copernicus Digital Surface Model.

The Copernicus DEM is derived from the WorldDEM produced from interferometric X-band radar observations of the German TandemX satellite mission. The Copernicus DEM differs from the WorldDEM in that water bodies have been flattened. The elevations reported over water in the original radar product presented a high level of uncertainty (noise, offsets) due to low interferometric coherence of single-pass interferometric X-band observations over water surfaces. The absolute vertical and horizontal accuracy of the DEM has been reported to be better than 4 m and 6 m, respectively (Fahrland et al., 2022).

The Copernicus DEM is provided in three different resolutions with either regional or global coverage and in three different data formats (DGED, DTED, INSPIRE). The highest resolution of 0.3-arc-second is available only for Europe (EEA-10), whereas the 1- (GLO-30) and 3-arc-second (GLO-90) versions are available globally. All DEM versions are provided in form of 1° x 1° tiles in geographic coordinates with World Geodetic System 1984 (WGS84; EPSG 4326) and Earth Gravitational Model 2008 (EGM2008; EPSG 3855) horizontal and vertical reference, respectively. All versions comprise a set of quality layers indicating where the original WorldDEM had been edited, the water body mask that was applied, and a layer indicating the error of elevation estimates at pixel level.

To evaluate which version of the Copernicus DEM should be used, the performance of the S1 pre-processing with each of the different DEM versions was tested for different regions (Van De Kerchove et al., 2021). The tests of the pre-processing of Sentinel-1 data with all different versions of the Copernicus DEM indicated that the differences in the accuracy of geocoding and topographic corrections between the EEA-10 and GLO-30 DEM (regardless of the format) are small whereas the use of the 3-arc-second version would have resulted in a significantly reduced quality of the topographic corrections. It was therefore decided to use the GLO-30 DEM for the global processing of Sentinel-1 data.

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 was derived from all seven bands of MODIS (Hansen et al., 2003) and is available annually since 2000 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 (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod44b_V006).

In this project, the dataset was 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. Its use was originally foreseen in the training phase of the models relating C-band backscatter to biomass.

Due to macroscopic artefacts especially in the tropics because of permanent cloud cover, the model training does not rely on the dataset any longer. Instead, the MODIS VCF dataset is currently used to support the estimation of the ENL for the Sentinel-1 data.

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 since the year 2000, as well as gains. 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 3-35). 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 2 October 2024). The dataset is used in the training phase of the models relating C-band and L-band backscatter observations to biomass.

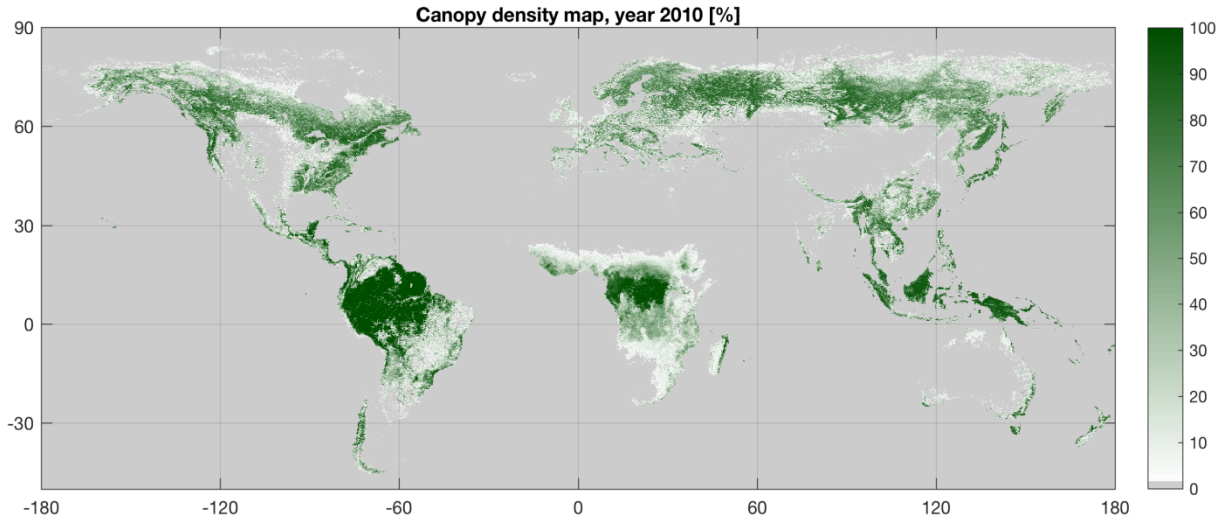


Figure 3-35: Canopy density map for the year 2010 produced from Landsat.

3.9 CCI Land Cover

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. Cropland is also characterized by low canopy density. If not flagged, the backscatter values from such types of land surfaces can distort the histograms used to estimate the parameters of the model relating SAR backscatter to biomass. 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. Cropland can be assumed to be an unvegetated surface when plants are not growing; however, when plants grow, they have a distinct signal that can substantially differ from the backscatter of an unvegetated surface. It is therefore preferable to discard observations in correspondence of cropland even if this may reduce the number of samples used to estimate the backscatter model parameters at the lower end. The Climate Change Initiative Land Cover (CCI-LC) project has released annual land cover maps between 1992 and 2020 (https://maps.elie.ucl.ac.be/CCI/viewer/ , last access on 2 October 2024) based on optical spaceborne datasets (Figure 3-36). The land cover maps are provided in equiangular projection with a pixel size of 1/360th of a degree in latitude and longitude. The two land cover maps were resampled to the geometries of the SAR datasets in support of the biomass retrieval procedure using nearest neighbour resampling.

Figure 3-36: 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, last access on 2 October 2024). Commission and omission errors occur, particularly in mixed classes or areas of strongly heterogeneous land cover. The classes of interest to CCI Biomass have among the highest classification accuracy though, typically above the overall accuracy of 70%. A data layer giving the classification certainty and a set of quality flags are provided.

3.10 Copernicus Global Land Operations

Retrieval of forest above-ground biomass with C- and L-band SAR data implies that non-forest land cover types, such as urban areas or agricultural land, may be assigned AGB values if their backscatter is in the range of values observed over forest. In addition, the CCI Biomass map products report AGB values for vegetated land that may not be considered forest according to the forest definition being applied. The general approach followed in CCI Biomass is not to apply any masks over vegetated land to allow for users to apply their own forest definitions. However, obvious erroneous biomass assignments over built-up areas, bare ground, snow/ice, or water are masked out using the Copernicus Global Operations Land Cover product (CGLS-LC100) version 3.0.1 (doi=10.5281/zenodo.3939050). The land cover maps are produced from PROBA-V data at 100 m spatial resolution at three different levels of detail according to the Land Cover Classification System (LCCS). The number of classes range from 12 at level 1 up to 23 classes at level 3. Version 3 of CGLOPS provides annual maps for the years 2015 to 2019 with an accuracy of the order of 80 %. In the CCI Biomass AGB maps, the CGLOPS land cover classes urban (code=50), water (code=80), open sea (code=200), bare (code=60), and snow/ice (code=70) have been masked out.

3.11 FAO Global Ecological Zones

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 3-37). 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. The GEZ dataset is used to stratify the estimation of some of the model parameters of the retrieval algorithms. 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 GDAL.

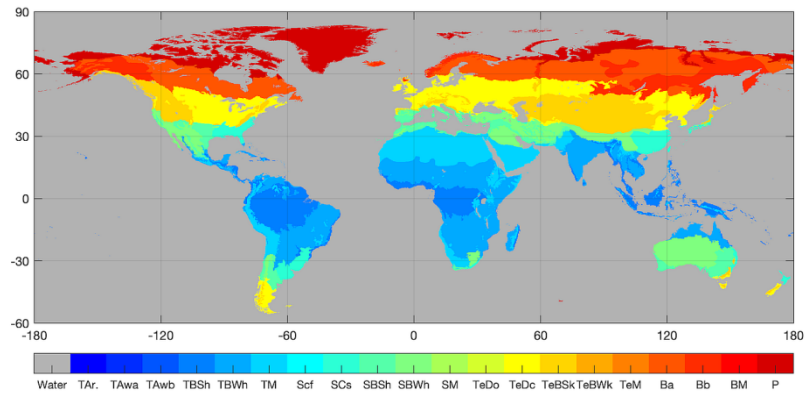


Figure 3-37: FAO GEZ dataset

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

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 TEOW dataset also brings in ecological properties of the landscape. The TEOW dataset divides the Earth land surfaces into 825 ecoregions (Figure 3-38**Error! Reference source not found.**). These are categorized within 14 biomes and eight biogeographic realms (Figure 3-39). The boundaries of each ecoregion, biome and realm correspond to the original extent of natural communities prior to major land-use change.

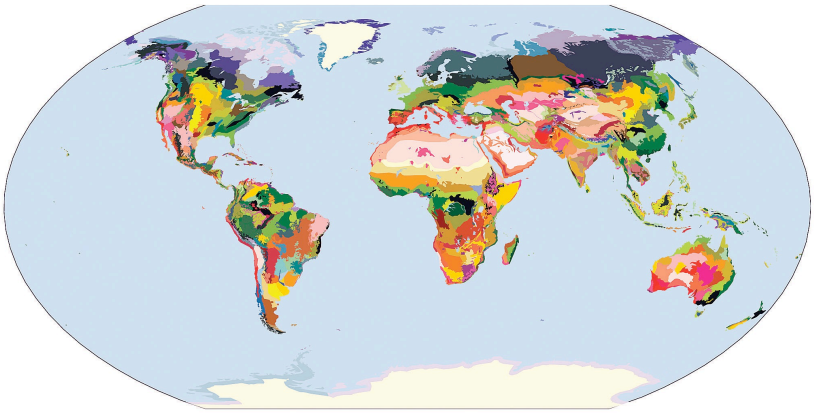


Figure 3-38: Terrestrial ecoregions of the world (Olson et al., 2001).

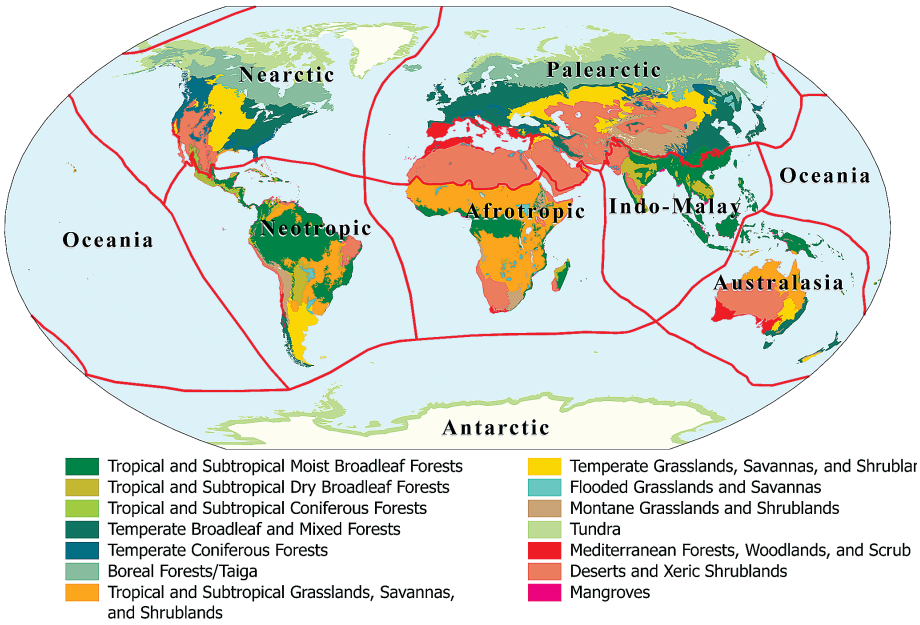


Figure 3-39: Grouping of the terrestrial ecoregions into biomes (see legend) and realms (see map). Picture reproduced from Olson et al. (2001).

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.

3.13 Relating forest variables

Tree variables such as diameter at breast height, tree height, crown diameter and tree biomass, are correlated in the sense that much of the variability of one variable (e.g., AGB) 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). Stand-level variables such as basal area, canopy height, canopy density, crown diameter, growing stock volume and above-ground biomass density are also related to each other, although the functional relationships might be different between the tree and stand level.

Since the retrieval model implemented in the CORE algorithm originally relates the SAR backscatter to variables that express the horizontal and vertical structural properties of a forest (canopy density and canopy height), it requires an additional link to AGB. As our interest is to estimate AGB, we then need to establish functions that may (i) reduce the number of forest variables in the models and (ii) relate those forest variables to AGB. Functional dependencies between forest variables at the level of forest stands or even coarser resolution are therefore of interest.

The CORE retrieval algorithm implements two sets of such functions:

1) A function between canopy density (CD) and canopy height (h) reduces the number of forest variables to height only:

$$CD = \mathrm{f}(h) \tag{3-2}$$

2) A second function linking AGB to canopy height allows explicit formulation of the retrieval model in terms of AGB:

$$h = f(B) \tag{3-3}$$

These two sets of functions are described in this Section and applied in the retrieval methods. Equation (3-3) is also used to derive the layer of maximum AGB from the LiDAR data.

The coefficients of the functions in Equations (3-2) and (3-3) represent auxiliary datasets ingested by the CORE algorithm. The coefficients are introduced, presented, and discussed below.

3.13.1 Linking LiDAR canopy density and canopy height

We use a generic model explaining canopy density as a function of the height metric corresponding to the start of a GLAS waveform (RH100). Canopy density was estimated as the ratio between the energy received from the canopy and the total energy returned for each footprint. Here, canopy density is considered as a normalized value ranging between 0 (unvegetated surface) and 1 (gapless canopy):

$$CD = 1 - e^{-qh} \quad (3\text{-}4)$$

The coefficient q in Equation (3-4) is empirical and describes variation in the relationship between canopy density and canopy height. Higher q means that canopy tends to close faster when trees grow.

To investigate the variation of q across the globe, the ICESat-GLAS dataset described in Section 3.6 was used. Any remaining footprints in unvegetated areas were removed using the CCI-LC product for 2010. The ICESat GLAS data were then divided according to a combination of the different TEOW ecoregions and a 1° x 1° grid. Each grid cell was divided by the ecoregions within it and *vice versa*, with 100 footprints being a minimum requirement for a polygon to be retained for the regression, otherwise the regression was undertaken with all the footprints available for the underlying ecoregion. The upper and lower 5 % were removed to account for potential outliers. A least squares regression, using the scipy optimize curve fit function, was then undertaken to obtain q (Figure 3-41) for each polygon. The mean square error (MSE) was also calculated for each regression.

The model described in Equation (3-4) corresponded to the ICESat GLAS metrics despite varying patterns of the canopy density to canopy height relationship (Figure 3-40; Kay et al., 2021). The varying dispersion of the data resulted in relative MSE values with a median of 39% and inter-quartile range of 27-60% relative to the q values obtained. The coefficient q varies from 0.019 to 0.153 across the globe, with the lowest values found in the sparser forests of the savanna or boreal regions and the highest in tropical regions (Figure 3-41 and Figure 3-42).

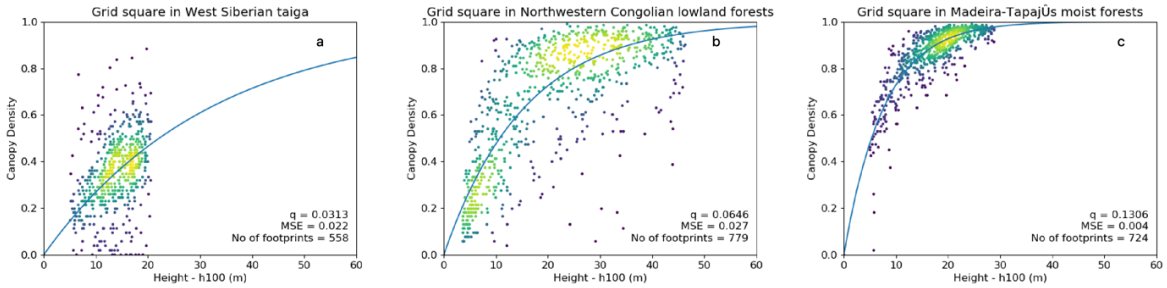


Figure 3-40: Least squares regression curves denoted by blue line (extended to 60m canopy height for comparison) for example polygons. With {a} a low q value (0.031), {b} a q value close to the global mean (0.064) and {c} with a high q value (0.131)

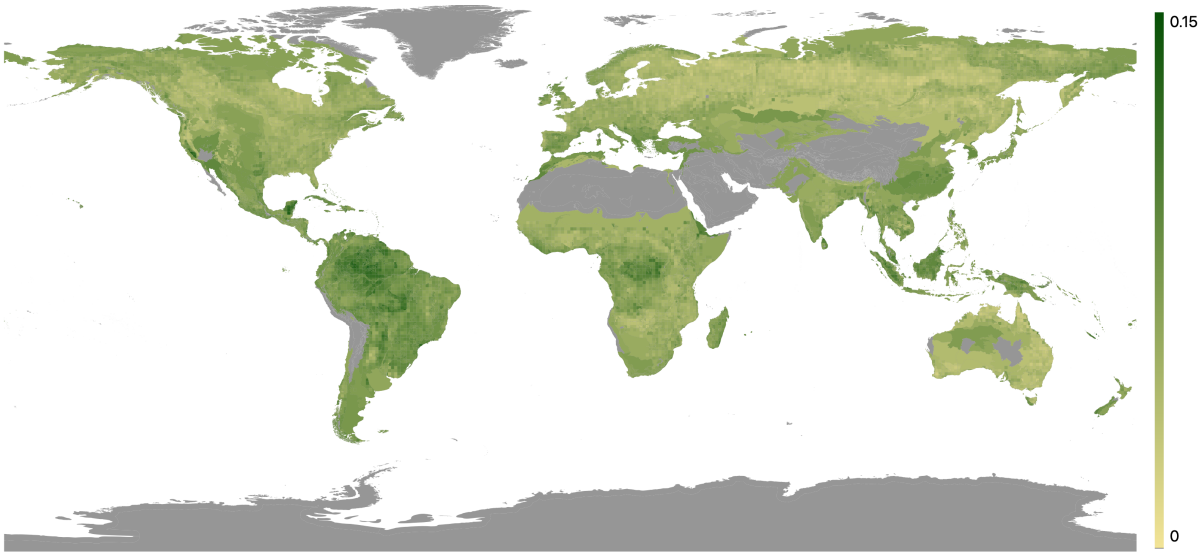


Figure 3-41: Map of q values per combination of TEOW ecoregion and 1° x 1° grid, obtained through least squares regression of available ICESat GLAS footprints within each ecoregion after filtering, ecoregions in white had no footprints

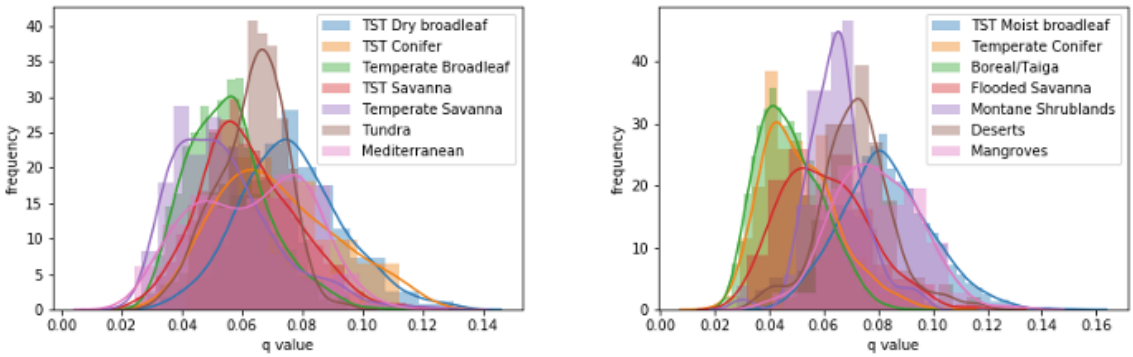


Figure 3-42: Histograms of q values per biome, split into two figures for clarity

These observations indicate that the functional dependencies in Equation (3-4) are maintained across the globe for most ecoregions. A rapid increase in canopy density with increasing height for dense tropical forests and a shallower slope for the sparser savanna and boreal regions was observed. The TEOW ecoregions were selected as a broad characterisation of multiple biophysical variables that may influence forest structure. However, localized variations of these variables within ecoregions are not available on a global scale. To capture some of these, a 1° grid was applied as a compromise to the number of footprints available for a regression. Analysis of the factors potentially contributing to these localised variations indicate that they may be better captured by using additional information such as altitude, temperature, precipitation, geology, or a wilderness layer. This was not possible in this analysis due to the relatively sparse sampling of ICESat GLAS and the additional filtering applied.

From the set of estimates of the coefficient q for each ecoregion, we generated a raster map with a pixel size of 1,000 m by rasterizing the TEOW dataset. Inpainting was applied to extrapolate values to surfaces not represented in the TEOW dataset.

GEDI LiDAR data were processed using the same methods, with non-vegetated footprints removed by applying the PROBA-V land cover product (Buchhorn et al., 2020). In addition, only the night-time full power beams were used, and these were filtered for l2a and l2b quality flags. To match the ICESat-GLAS dataset, all orbits from the 1st and 4th quarter of the year (e.g., data from October to March) were also removed.

ICESat GLAS is the only satellite which provides global coverage, even though the sampling is less dense than GEDI. In addition, the MSE values in relation to q are considerably lower for ICESat GLAS (median of 39% and interquartile range of 27-60%) compared to GEDI (median of 69% and interquartile range of 44-91%). Visual analysis of the GEDI dataset has also shown several discrepancies which include: concentrations of footprints with very low canopy density values (Figure 3-43), a step in height to canopy density values around 8 m height, and regions which have some orbits with lower canopy density values (personal communication John Armston). The above functional relationships are based on those calculated with ICESat GLAS data. At the time of writing, there is no equivalent dataset released from ICESat-2 data.

The single parameter (q) varied depending on which satellite data were applied, with GEDI having consistently lower q values across the majority of polygons (Figure 3-44,d). This variation is due to differences in the canopy density values from each satellite (assessed by comparing the means per polygon; Figure 3-44,b), with mean canopy density values being consistently lower for GEDI except for those from polygons with high mean density (canopy density values of ~0.8).

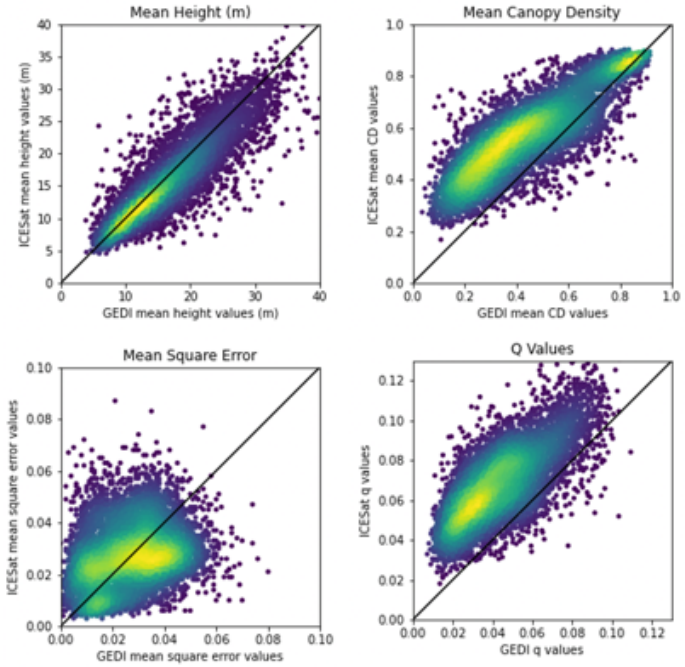


Figure 3-43: Density scatterplots with viridis colour scale comparing GEDI data on the x axis and ICESat GLAS data on the y axis. Comparing mean canopy height per polygon (top left), mean canopy density per polygon (top right), mean square error of the q regression per polygon (bottom left) and derived q values per polygon (bottom right). A 1:1 line has been added to each blot in black.

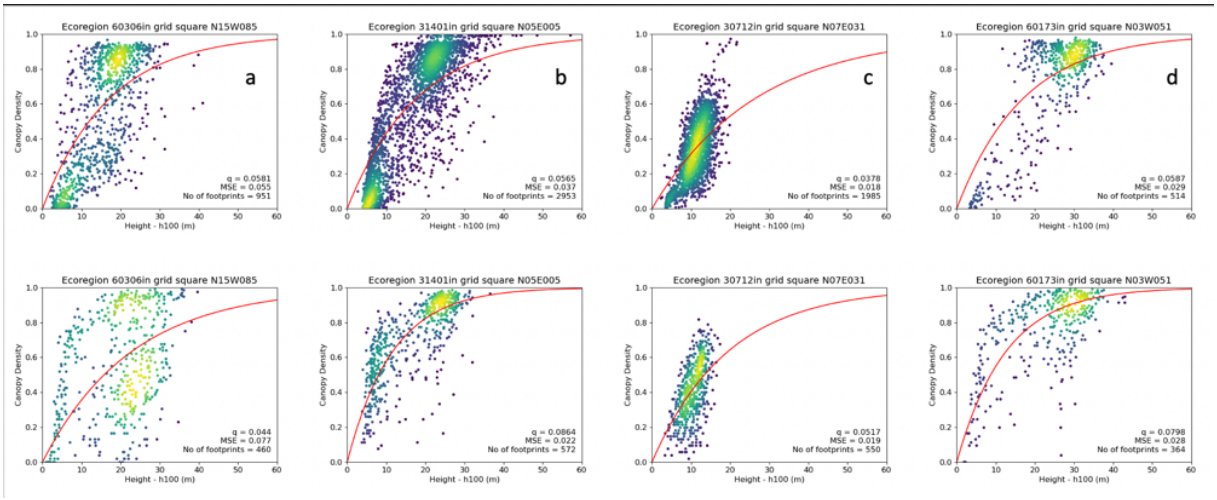


Figure 3-44: Least squares regression curves denoted by red line (extended to 60 m canopy height for comparison) for example polygons. The regression of GEDI data is on the top row with ICESat GLAS data on the bottom row. With (a) varied concentrations of the data distribution, (b) concentration of low canopy density values for GEDI, (c) a jump in GEDI at around 8 m height and 0.18 canopy density and (d) a shift in canopy density values, in particular in intermediate canopy density ranges.

3.13.2 Linking AGB and canopy height

AGB is obtained from measurements of 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, close-range EO measurement techniques have been developed (e.g., terrestrial laser scanners) that can allow more rapid quantification of several such parameters. However, even such techniques do not allow mapping of large areas, which instead can only be achieved with far-range EO instruments (i.e., mounted on airborne or spaceborne platforms). Whenever the measurements of forest variables are "outsourced" to far-range EO, tree allometries (based on destructive sampling) start to play a role. Airborne laser scanning has proved to be a reliable sensing technique, allowing accurate measurement of variables related to forest structure. Accordingly, studies have been developing in the direction of characterizing the relationship between AGB and laser-based metrics at stand level to mimic the relationship between tree variables in tree-level allometric functions. In particular, the relationship between canopy height and AGB has been investigated at several locations because both horizontal and vertical structures of the forest are sensed by LiDAR.

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 in the Amazon and showed 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 valid across four sites, although the model had to be trained separately at each to capture the region-specific forest structural properties in the power-law model. Saatchi et al. (2011) proposed a similar power-law function relating AGB to basal area weighted tree height (Lorey's height) 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 several 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 LiDAR metric top-of-canopy height (TCH) obtained from airborne

observations to above-ground carbon density in neotropical forest. 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 therefore can potentially generate incorrect 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 site conditions. They also showed that an explicit use of canopy gap information derived from laser measurements improves the retrieval of carbon density. Labriere et al. (2018) used airborne laser scanned data and in situ observations in tropical forests in Gabon and French Guiana to test a power-law function relating AGB to several height metrics, including TCH and MCH.

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

$$AGB = p_1 h^{p_2} \tag{3-5}$$

where p_1 and p_2 are regression coefficients estimated by non-linear least squares. Here h represents the canopy height of a forest.

An attempt to use a function relating a forest height metric and AGB at the global scale can build on such local studies but requires simplifying assumptions and a great deal of generalization. The spaceborne LiDAR datasets provide a reasonable sampling of canopy height globally. Unfortunately, there are no measurements available at the footprint level nor are there alternative databases containing height and AGB from the same location that would allow spatial characterization of the functional dependence between height and AGB at the global scale and at the hectare scale spatial resolution envisaged in CCI Biomass.

The major limitation to estimating the coefficients in Equation (3-5) and characterizing its spatial variability, may be overcome in two ways:

1) By exploiting spatially explicit estimates of AGB derived from EO observations
2) By exploiting statistics obtained from inventory-based measurements of AGB at the level of administrative or ecological level

The first approach was pursued for previous versions of the CCI Biomass CRDP. In a first attempt, the first version of the CCI Biomass dataset was used as reference. This version was obtained with the approach developed for the GlobBiomass dataset and did not include an explicit function that related forest variables (Santoro et al., 2021). While we were fully aware that errors in the map dataset might have affected the estimates of the model relating AGB and canopy height, our deep knowledge of this AGB dataset and its errors was exploited to reduce the impact of source errors on the models proposed to relate AGB and height.

To estimate the regression coefficients, we computed the AGB value from the map dataset corresponding to the area of the LiDAR footprint. An assessment of the relationship between canopy height from spaceborne LiDAR and AGB from the map revealed low correlation and very large variance. However, after averaging to coarse resolution grid cells, the error was found to decrease and the coefficient of the equation relating height to AGB became significant. The average values of canopy height and AGB were then stratified according to different criteria to capture the spatial variability of the association between height and AGB. We refer to previous versions of the CCI Biomass ATBD for a description of the stratification and the outcome of the investigations (https://climate.esa.int/en/projects/biomass/key-documents/, last access on 30 October 2023).

From the extensive analysis undertaken in previous years, we concluded that the most reliable set of coefficients corresponded to average values in 1° × 1° grid cells and a window size of 10° × 10°.

The map with the estimates of the coefficients p_1 and p_2 used until v3 of the CRDP, i.e., based on a combination of data from ICESat GLAS and ICESat-2, is shown in Figure 3-45. While this set of estimates may be considered a realistic approximation, the evaluation of the CCI Biomass AGB maps showed that they caused biases.

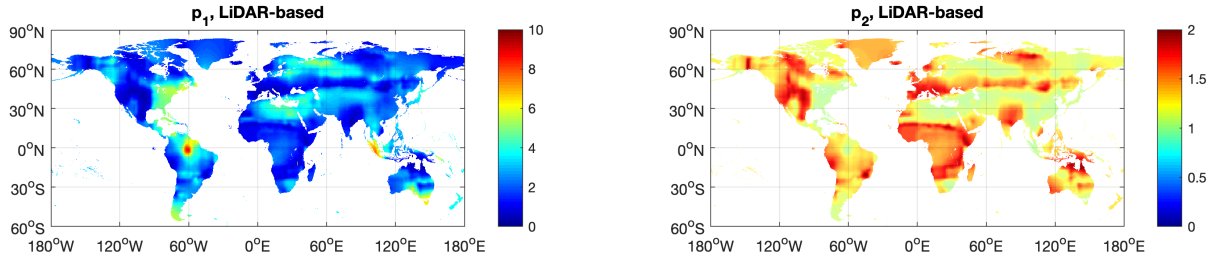


Figure 3-45: Maps of the coefficients p_1 and p_2 used until v3 of the CRDP. The maps were based on canopy height from ICESat GLAS and ICESat-2, and AGB from the GlobBiomass and CCI Biomass v1 datasets.

Starting with v4 of the CORE retrieval algorithm, we pursued the second approach, and linked averages of spaceborne LiDAR canopy height values to AGB averages derived from national forest inventory (NFI) data per administrative unit (e.g., provinces, states, counties) or ecological unit (broadleaf forest, coniferous forest).

The ICESat-2 dataset was used to compute the averages of canopy height per administrative or ecological unit to associate with the values of AGB. The AGB averages were either computed directly when data were available to us or, in most cases, were extracted from inventory reports published by the NFIs. We were able to identify NFI-based data or statistics for 106 countries. In addition, we used national averages of AGB reported in the FAO Forest Resources Assessment for 2020 (https://fra-data.fao.org/WO/fra2020/home/, last access on 1 October 2024) to complement our database over 94 countries for which we could not identify NFI measurements or statistical data. For 50 remaining countries, we could neither identify NFI-based statistics nor the FRA 2020 database reported AGB. Figure 3-46 shows the spatial coverage of our database of NFI AGB averages in the form of maps. We split the averages according to whether they were obtained from (i) NFI measurements either by ourselves or from the reports and (ii) FAO FRA 2020 country reports. Except for large parts of Africa, the Middle East, parts of Southeast Asia and the Caribbean, we were able to identify NFI-based values. Gaps could, however, be filled with the FAO FRA 2020 data. The 50 countries for which we could not identify any value on AGB were either very small countries or countries without forest cover.

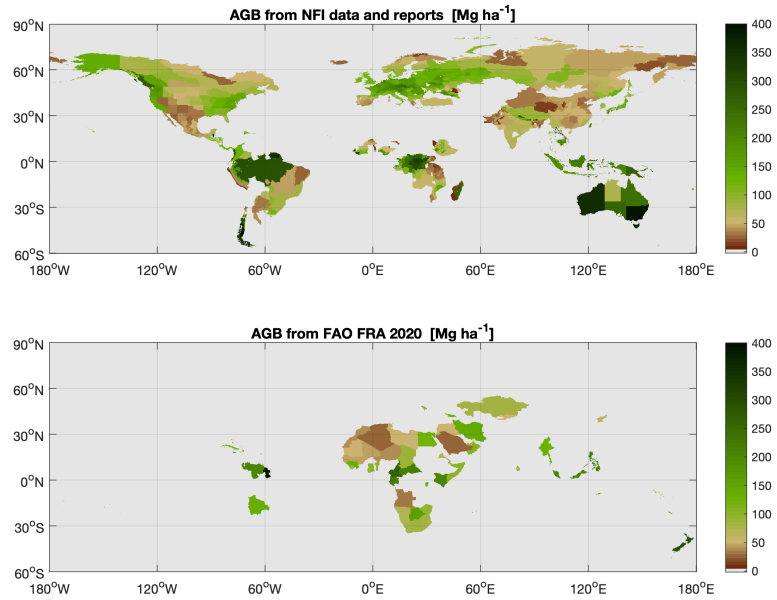


Figure 3-46: Map illustrating the coverage of AGB values per administrative or ecological unit derived from NFI datasets and statistical reports.

Our database of AGB averages (see Annex D) currently consists of 761 values for sub-national administrative or ecological units and 94 values at national level from the FRA. To obtain some sort of spatial characterization of the height-to-AGB function, the data were stratified by continent and major ecological traits. The definition of the 18 strata in Figure 3-47 followed an additional set of criteria based on number of data values, availability of NFI-based values and uniform representation of the range of AGB. It is understood that such a stratification does not allow the capture of small-scale variability of the height-to-AGB relationship. Nonetheless, the focus was on establishing valid and reliable predictions of AGB everywhere, trying at the same time to capture macroecological patterns. Our overall goal was to minimize the risk that large regions would have been characterized by AGB biases attributed to an imperfect function.

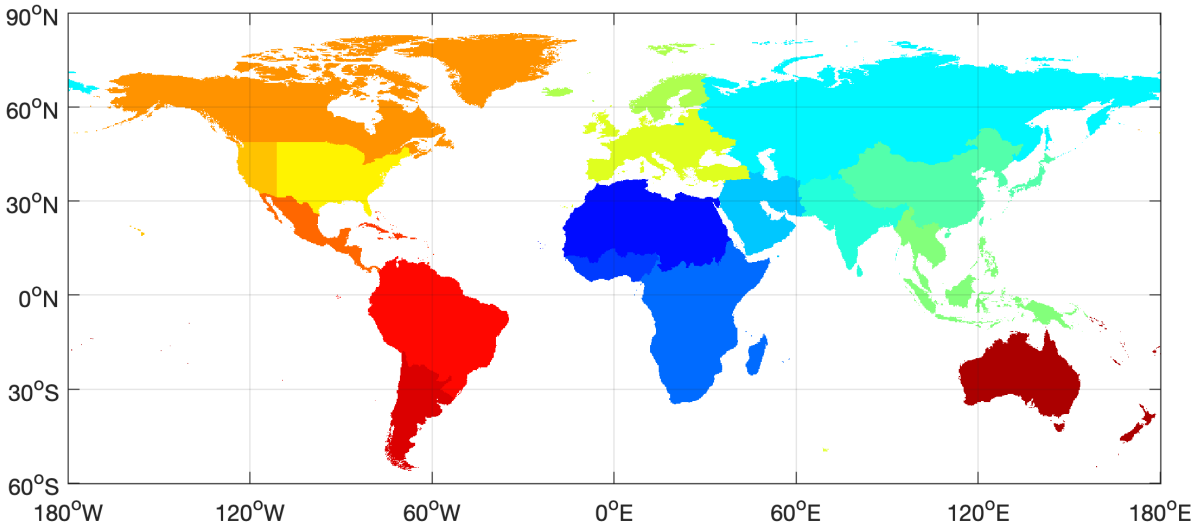


Figure 3-47: Strata used to split the database of AGB and LiDAR canopy height statistics.

The data values of average canopy height and AGB, and the corresponding fit of the power-law function in Equation (3-5) to the data are illustrated in Figure 3-48. Strata characterized by many sample points, which were, furthermore, based on data from well-established NFIs (e.g., Europe, Russia, or North America) demonstrated the capability of the power-law model to reproduce the trend between AGB and canopy height. In such cases, the coefficient p_2 was always between 1 and 2.5. This can be considered a range of realistic values for the coefficient, even if the spatial scale at which we are fitting Equation (3-5) is much coarser than the sub-hectare scale of other studies. Indeed, AGB must increase with height and the rate of the biomass increase must be larger than the rate of height increase (i.e., p_2 must be larger than 1). A value larger than 2.5 is unrealistic because it would imply an extraordinarily rapid increase of biomass for a small increase of height. Originally (v4), our approach included strata that were characterized by a small number of data points or by substantial spread of the observations, when the estimates of the two coefficients did not have physical meaning. In this case, we opted for fitting a linear function, i.e., a power-law model with p_2 = 1 (e.g., Oceania, North Africa, Middle East). Such cases were mostly characterized by lack of sub-national values of AGB by NFIs so that the models had to be based on FAO FRA national values. We preferred associating a set of coefficient estimates to each stratum, even if approximated, rather than expanding estimates from neighbouring strata to preserve, even if minimally, the spatial variability of the relationship between height and AGB. Validation of the CRDP v4 revealed that this linear assumption caused significant biases. For version 5, we therefore revisited the definition of the strata, reducing it from 20 to 17, and revisited some of the AGB statistics, which were apparently incorrect and distorted the estimates of the coefficients in Equation (3-5). Validation of the CRDP v5 revealed strong biases in Southeast Asia and Oceania, which could be attributed to the imperfect structural function. For version 6, we revisited the AGB statistics for these regions and updated the structural functions.

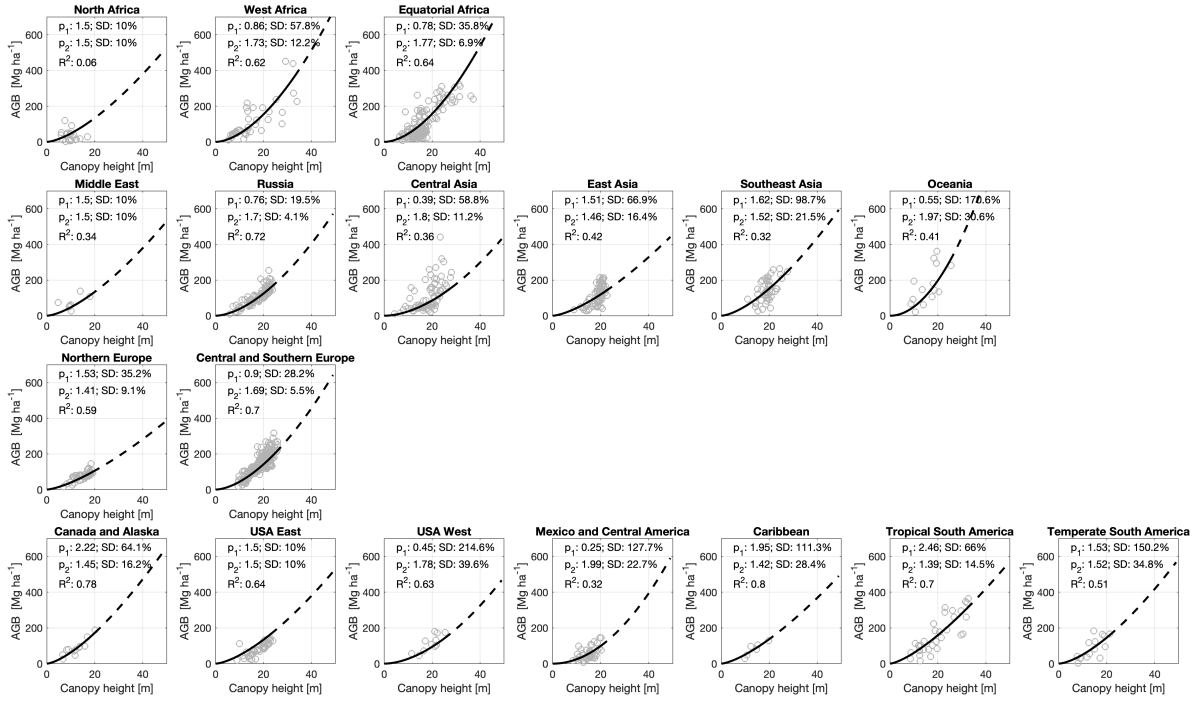


Figure 3-48: Observations of average canopy height and average AGB at national and sub-national level (circles) and corresponding model fit based on Equation (3-7) for each of the 18 strata in Figure 3-47. The 18 panels are here arranged by continent. Each panel reports the estimate of the model's coefficients and their SDs. The model fitted to the observations (solid curve) is extended up to 50 m (dashed curve) in accordance with the canopy height range shown in Figure 3-28.

The stratum-wise estimates of the two coefficients of Equation (3-5) were finally converted to the spatially explicit dataset shown in Figure 3-49. Compared to the previous version of the estimates of the coefficients in Figure 3-45, the range of values estimated for the two coefficients p_1 and p_2 is smaller and we obtained a realistic model fit in each stratum. With this update in the stratification, we could solve one issue that assumed a linear function instead of a power-law if the estimates of p_2 had been smaller than 1, i.e., for increasing height the biomass accumulation would have reduced.

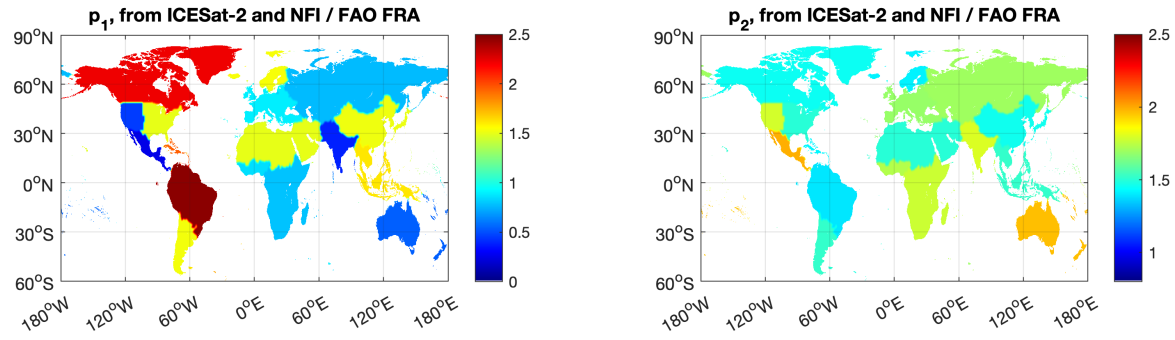


Figure 3-49: Maps of the coefficients p_1 and p_2 based on canopy height from ICESat-2, and AGB averages obtained from inventory data by NFIs or the FAO FRA 2020 country reports.

To verify the reliability of AGB predicted with the function in Equation (3-5), we compared the AGB estimates based on Equation (3-5) and the set of coefficients obtained for this version of the ATBD in Figure 3-49 with AGB values obtained from plot inventory measurements at the 0.1° large grid cells used in the Product Validation Report of v4 [RD-8]. We also compared the same AGB values from the plot inventory data with AGB estimates derived from previous versions of the estimates of p_1 and p_2 in Equation (3-5). For this, we used grid cell averages of ICESat-2 canopy heights from the segments for the years 2019-2022. We note that the AGB predictions from ICESat-2 and the AGB values based on forest inventory data are independent of each other.

Figure 3-50 shows that for each version of the coefficients of the function relating LiDAR-based canopy height to AGB, the data points clustered more tightly along the identity line. Although we are not able to model correctly the highest AGBs, which occurred primarily in a few regions of Australia, the improvement since v3 is remarkable, indicating that the modifications leading to the current version of the CORE algorithm point in the right direction. v4, v5 and v6 share the same predictor, i.e., ICESat-2 observations and inventory-based sub-national averages, but different stratification of the world and updated support datasets. The function obtained for v6 is closer to those obtained for v3 (based on a blend of spaceborne LiDAR and the first CCI Biomass AGB map) and v2 (based on ICESat GLAS only and the GlobBiomass AGB dataset). Dispersion and bias are however smaller, particularly in the low biomass range.

This analysis was undertaken to understand the reliability of relating canopy height from spaceborne LiDAR to AGB from a global dataset of observations. Results from this analysis impact the CORE retrieval algorithm but it is remarked that the AGB values from the database of plot inventory measurements used to validate the AGB maps were not used to calibrate or fine tune the height-to-AGB relationship. Indeed, the overall assumption that LiDAR-based canopy height from space and inventory-based statistics of AGB could be used as a surrogate to predict the relationship between height and AGB is confirmed.

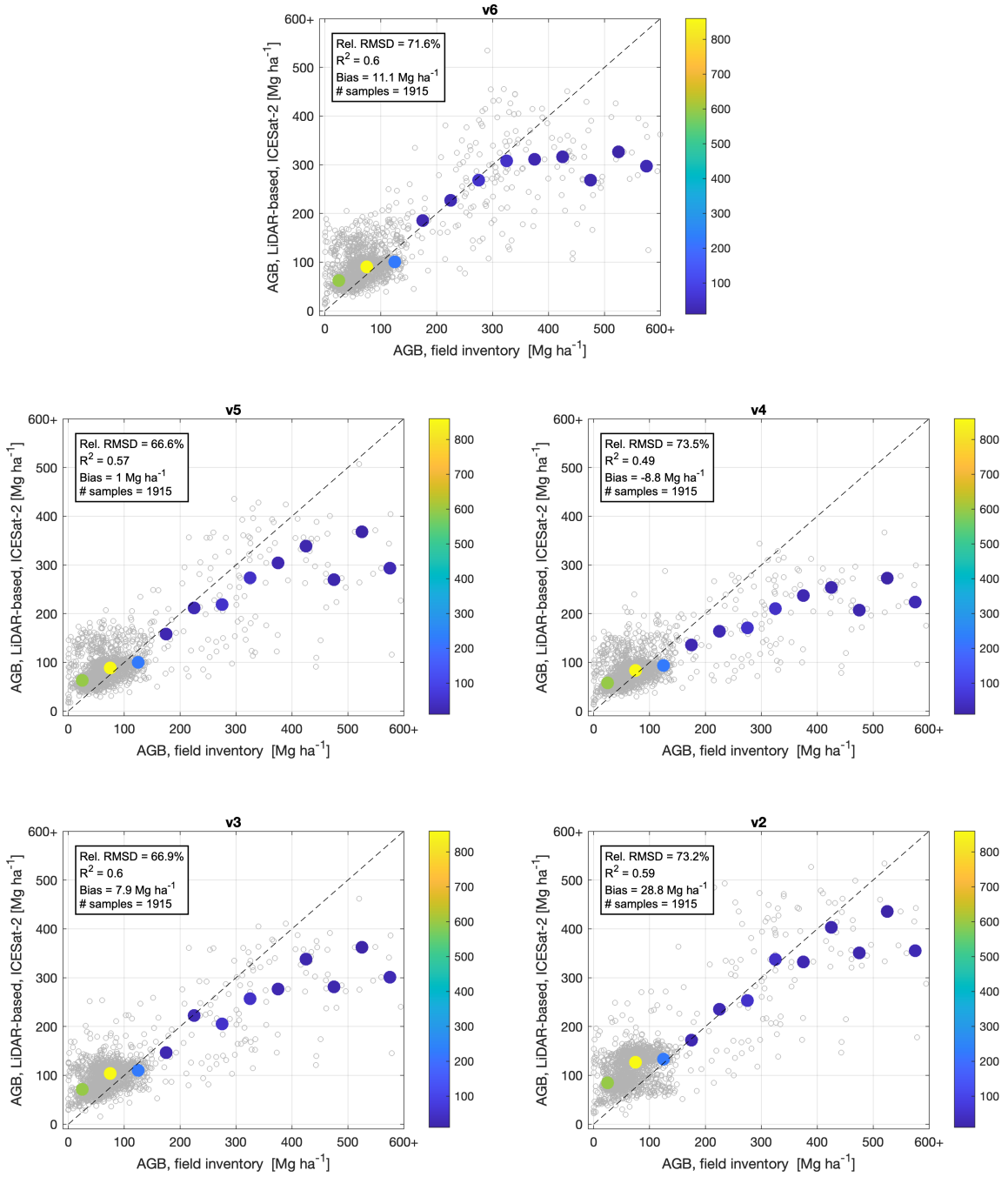


Figure 3-50. Comparison of grid cell average values of AGB from field inventory data and from Equation (3-7) applied to ICESat-2 canopy height measurements (circles) for each version of the ATBD. The filled circles represent the median value of AGB from the LiDAR-based predictions in the corresponding 50 Mg ha^{-1} wide bin of AGB from the inventory-based values. The colour bar refers to the number of 0.1° grid cell observations in each AGB bin.

Regardless of the approach developed in CCI Biomass to train Equation (3-5), the use of average values at coarse spatial resolution implies that the set of estimates for the equation's coefficients is not valid for the hectare scale because of the non-linear nature of the model. In addition, average values of AGB and canopy height cover a smaller range of values than the original data at full

resolution, which is at hectare scale. By missing extreme values in the training dataset, the model in Equation (3-5) is not well constrained for forests with the highest biomass densities. The two sources of potential bias are here acknowledged, but it is not possible to quantify the bias introduced by using an "average" model for the height-to-AGB relationship except at the local scale where retrievals with the CORE algorithm implementing the "average" model and a "local" model (e.g., Labriere et al., 2018) are compared to each other. Possible endeavours to bridge the gap between data availability and spatial scale are discussed in the Algorithm Development Plan report [RD-6]. Nonetheless, we believe that our approach using coarse resolution but equally distributed measurements of AGB and canopy height as being currently the only option to characterize the height-to-AGB relationship globally.

3.13.3 Raster dataset of 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 AGB. However, the models built in the retrieval algorithms need to be constrained. On one hand, the spatially explicit datasets of canopy density and canopy height derived from optical and LiDAR observations are used to derive estimates of the retrieval model parameters. On the other hand, an estimate of the maximum biomass in a region is required to constrain the retrieval within a realistic range of biomass values. This is necessary to avoid values of biomass that are outside their known range, which is likely to occur with the EO data here selected as predictors because of their increasingly weaker sensitivity to biomass. To this end, we first developed a spatial database reporting estimates of AGB assumed to be representative of dense forests, i.e., with high biomass density, in a certain area. These estimates were then converted to maximum AGB. Later, we developed an alternative framework based on LiDAR metrics of canopy height and spatially-explicit explicit estimates of AGB from inventory or map data through the height-to-AGB function in Equation (3-5).

3.13.3.1 Inventory and map-based maximum biomass

Initially, a value was assigned to the centre of each tile in a regular 2°×2° grid. Where available, *in situ* measurements from field plots or spatially explicit datasets of Growing Stock Volume (GSV) or AGB were used. The AGB of dense forests, i.e., a parameter used in the retrieval model was then defined as the 90th percentile of the histogram within the 2°×2° area (Santoro et al., 2011). Interestingly, we identified a rather robust scaling factor between this parameter and the maximum AGB of 1.2. Elsewhere, it was estimated with an empirical piece-wise linear function (Santoro et al., 2015a) starting from values of the average AGB reported at a provincial or national level. For tiles including several provinces or nations, the average AGB 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 AGB of dense forest as the joint 90th percentile of the histogram of the map values. For areas lacking any form of AGB estimates, the IIASA FAO 0.5° dataset of global AGB was used. The value for the AGB of dense forests was then set equal to the maximum of the 16 values within the 2°×2° large tile. Given that the database contained both GSV and AGB observations, we applied the Biomass Conversion and Expansion Factor **(**BCEF) computed in the GlobBiomass project (Santoro et al., 2021) to convert between the two variables (AGB = BCEF * GSV). Figure 3-51 shows the origin of the estimate of the biomass of dense forests.

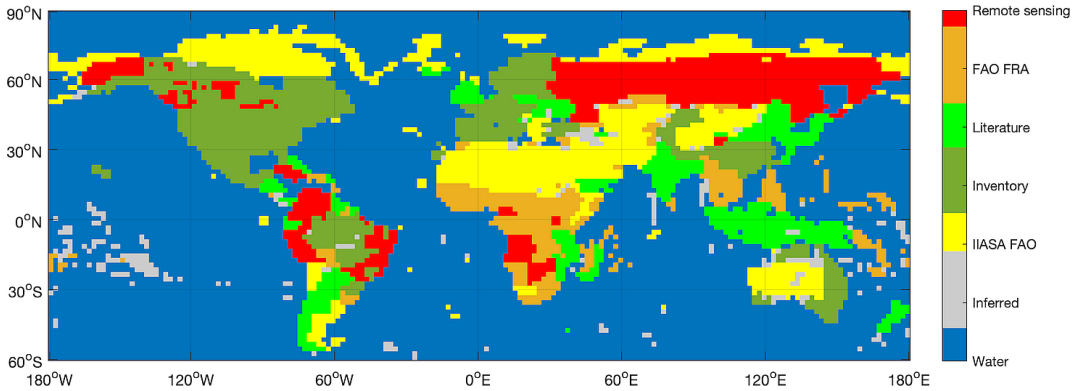


Figure 3-51: Map detailing the origin of the numbers used to generate an estimate of the biomass of dense forests at the spatial scale of 2°.

The procedure implemented to characterize the maximum AGB globally was crude and should be considered a first-order approximation. Attaching a single value to areas covering thousands of km^2 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 AGB 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 prevent the dataset having errors and uncertainties propagating from the input datasets, the rules implemented to estimate the AGB of dense forests from the data available in a tile, and the BCEF values.

To reduce some of the weaknesses in the dataset, additional processing steps were 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 database.
- Increase the spatial resolution to better depict smaller scale variations in the maximum AGB, such as are expected for transition areas between tropical rainforest and savannah.

The maximum AGB 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. To verify/improve the initial estimates and to fill gaps, a database of predictor layers that are expected to have predictive power for maximum AGB was compiled at 2° × 2° resolution, including the bioclimatic variables and ICESat GLAS observations of canopy density and height. For each 2° × 2° 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 ancillary datasets as predictors. Once calibrated, the models were then used to predict the GSV of dense forests globally at a resolution of 0.2° × 0.2° (Santoro et al., 2021).

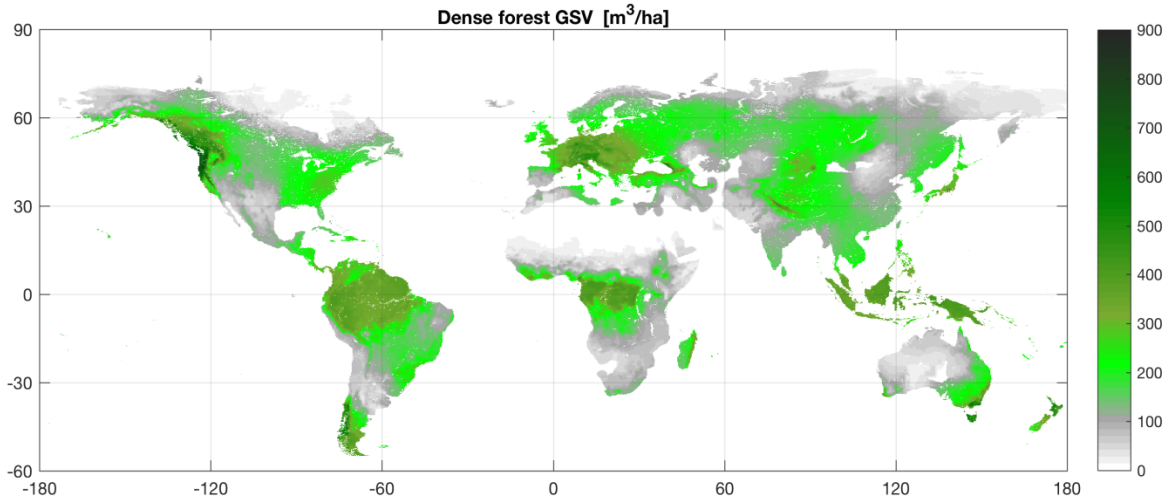


Figure 3-52: Map of the GSV of dense forests with a spatial resolution of 0.2°.

Using the scaling factor of 1.2 between GSV of dense forests and maximum GSV as well as the BCEF relating GSV and AGB (Santoro et al., 2021), we generated a global layer of maximum AGB. The layer of maximum AGB obtained from the dataset of maximum GSV and scaled with the BCEF is displayed in Figure 3-53.

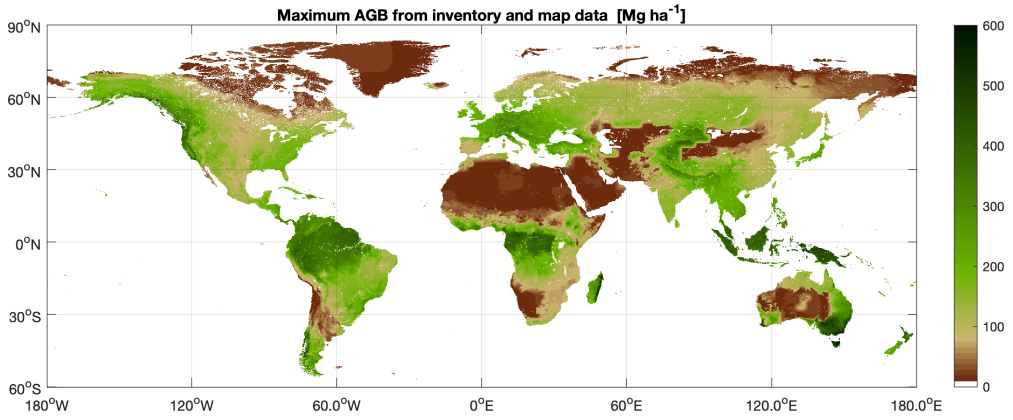


Figure 3-53: Map of maximum AGB derived from inventory and map-based datasets and the BCEF.

These datasets of dense forest GSV, maximum GSV and maximum AGB were used to generate the v1 of the CRDP. Validation of the CCI Biomass maps based on the initial map of maximum AGB revealed locations affected by underestimation of the map-based values. The underestimation occurred in the form of a saturated value shown in the PVASR [RD-4] and the PVIR [RD-8] documents, an indication that the retrieval had been cut off at an AGB level lower than the real maximum AGB. This occurred in regions where the characterization of the maximum AGB was poor, e.g., because extrapolated from national inventory averages, other maps, or scarce inventory samples. Since the datasets underpinning our spatially explicit layer of dense forest and maximum AGB have not improved in the last years, this approach is not supported any longer. It is indeed preferred to rely on direct observations relating LiDAR and inventory-based data through functions that may guarantee stronger reliability. However, the datasets based on the collation of inventory and map data streams are still used in CCI Biomass to benchmark this new form of estimating maximum biomass.

3.13.3.2 LiDAR-based maximum biomass

To overcome the problem related to a layer of maximum biomass primarily based on values collated from different sources, an alternative solution was sought that would be more data driven. The availability of global coverages of canopy height metrics from three sensors indeed suggested they could be used to improve values in regions where the original maximum AGB was poorly characterized. For this, we used the function relating canopy height and AGB and defined the maximum AGB as the value predicted by Equation (3-5) corresponding to the maximum value of canopy height.

Below we describe the procedure that was originally implemented to create the first global layer of maximum biomass from spaceborne LiDAR followed by the improvement that is part of the current version of the ATBD (v6).

V2 and v3 of the CRDP preceding this ATBD share the same approach but different sets of LiDAR observations, namely ICESat GLAS only and a combination of data from the three spaceborne LiDAR sensors, respectively. The combination was undertaken to overcome the issue of sparse sampling in each of the three datasets. For each laser sensor, a map of maximum canopy height with a pixel size of 0.25° was generated at first. The maximum canopy height in a pixel was defined as the 95th percentile of the canopy height histogram from the footprint-level data. The coarse resolution reduced the impact of incorrect canopy height estimates on the percentile value because this was based on many values even in regions with sparse footprints. In v2, this map represented the maximum biomass. In v3, the three maps of canopy height were combined with a weighted average. The weights were proportional to the area covered by all footprints/segments within a given area (0.25° in our case). GEDI was the main contributor in the tropics whereas ICESat GLAS and ICESat-2 data contributed in an equal manner in extra-tropical regions. To generate the maximum AGB from the maximum canopy height, we applied the function described in Equation (3-5) (Section 3.14.2). Figure 3-54 shows the layer of maximum AGB based on the LiDAR observations.

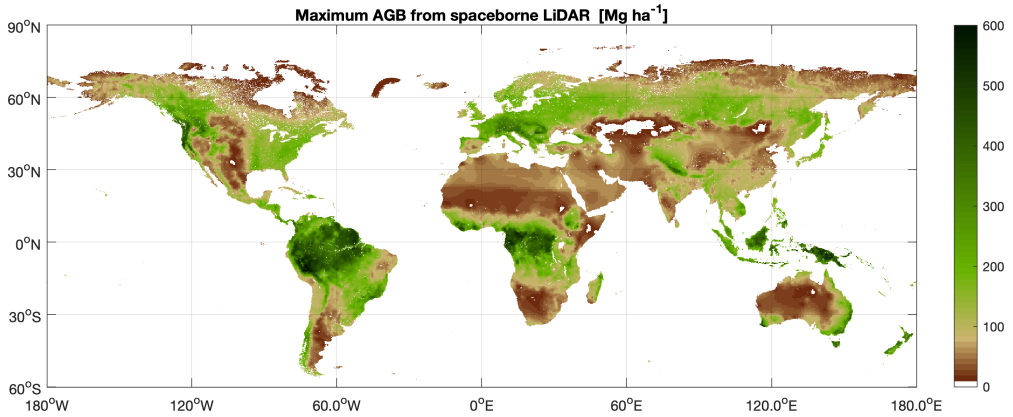


Figure 3-54: Map of maximum AGB derived from the LiDAR-based dataset of maximum canopy height. This layer was used in v3 release of the CCI Biomass CRDP.

With v3, for boreal and temperate forests, the AGB predicted from the LiDAR measurements was slightly higher, which is in line with investigations that demonstrated light saturation of the earlier CCI maps when the retrieval was constrained with the inventory/map-based maximum AGB (v1) and with the ICESat GLAS data only (v2) [RD-4] and [RD-8]. Smaller values were instead predicted across Alaska and most of the western north American continent. These results are plausible since the assessment of the earlier CCI maps (versions 1 and 2) indicated somewhat

higher values from the map than those reported by forest inventory. Across the wet tropics, higher AGB values were again predicted using the LiDAR-based measurements. This result is consistent with assessments of earlier CCI maps [RD-4]. In the dry tropics and the subtropics, instead, the maximum AGB predicted from the LiDAR data is often smaller than the inventory/map-based values. These were based primarily on other AGB maps or upscaled country statistics and therefore potentially biased. In regions, where inventory data were available (e.g., Madagascar and Australia), the lower values are explained because of the small number of LiDAR measurements in our database. The densest forest was hardly observed, which lead to a lower maximum AGB.

The availability of LiDAR observations from GEDI and ICESat-2 from several years, starting with v4, implied that a layer of maximum AGB could be based on data acquired by one sensor. Since v5 of the CRDP, we selected ICESat-2 data as the only contributor to the layer of maximum AGB because the ICESat-2 canopy heights appear to be more consistent with the ICESat GLAS values than GEDI and because ICESat-2 is global.

For v4 and v5, we first estimated the maximum canopy height for each year of the ICESat-2 dataset (i.e., 2019-2020 in v4 and 2019-2021 in v5) and then combined the results to give a final value. For v4, the maximum value of canopy height in a given year was defined as the 99th percentile of all height estimates corresponding to vegetation within an 0.1° grid cell. Maximum height was estimated in grid cells with dimensions of 0.1° × 0.1°. This was a compromise between stabilizing the value of the percentile expressing the maximum canopy height and retaining spatial detail. This definition tried to truly detect the tallest forest in a certain area and account for spatial variability. However, it likely captured the tallest tree within a segment, so was not properly a forest canopy height and was furthermore susceptible to noise. The average of the yearly height values weighted by the number of segments in each grid cell helped to reduce noise but could not remove potential biases. For v5, we first averaged the segment-wise data to 1 km grid cells to create yearly maps of mean canopy height and averaged them in time using the number of segments per grid cell as weights. Having accounted for noisy measurements in the original data, we defined the maximum canopy height as the 98th percentile of the 1 km values within 0.1° large areas. This was found to be sufficiently stable to avoid point-wise noise in regions of poor coverage. A 2-D median filter was then applied to reduce unrealistically high values. For v6, we created a single map of mean canopy height with a grid cell of 1 km from the data acquired between 2019 and 2022. The maximum canopy height was still defined as the 98th percentile of all canopy heights within 0.1° grid cells. Estimates based on less than 3 observations and with an estimate larger than 35 m corresponding to a tree cover of less than 30% were discarded. These thresholds were obtained after a detailed analysis of the data aiming at understanding the causes of biased height estimates. Figure 3-55 shows the map of maximum canopy height; values in correspondence of unvegetated surfaces (e.g., desert, and ice-covered regions) were obtained with inpainting from neighbouring regions where valid canopy height measurements were available.

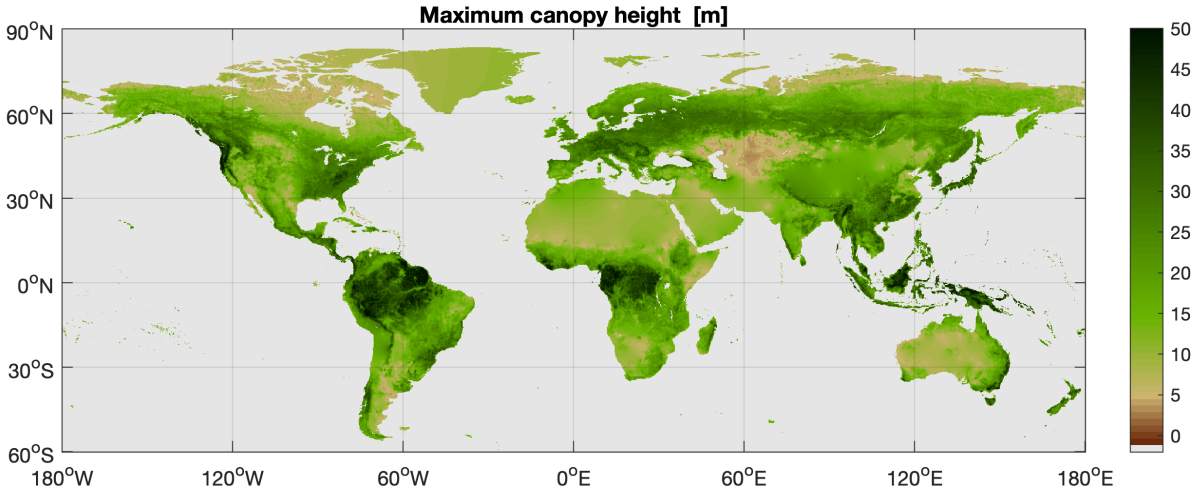


Figure 3-55: Map of maximum canopy height derived from ICESat-2 data acquired in 2019-2021.

Figure 3-56 shows the corresponding map of maximum AGB while Figure 3-57 indicates the difference between this map and the previous version.

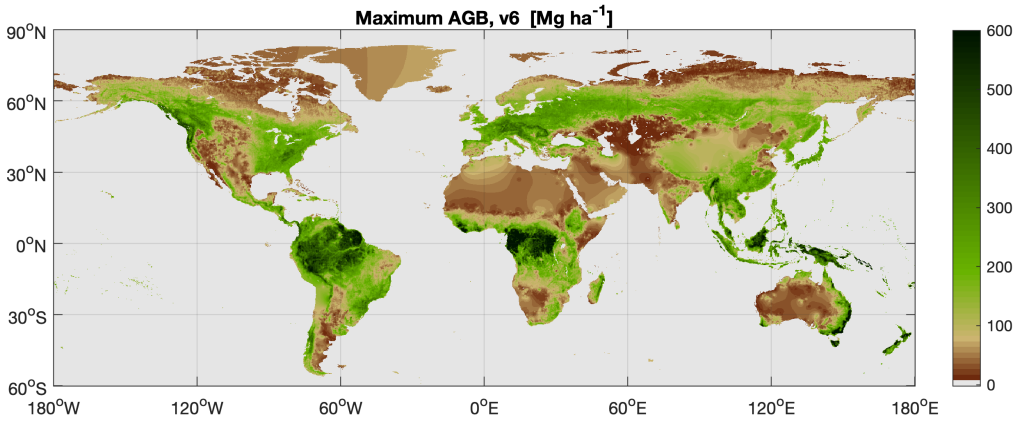


Figure 3-56: Map of maximum AGB derived from the ICESat-2-based dataset of maximum canopy height. This layer is used in the current release of the CCI Biomass CRDP.

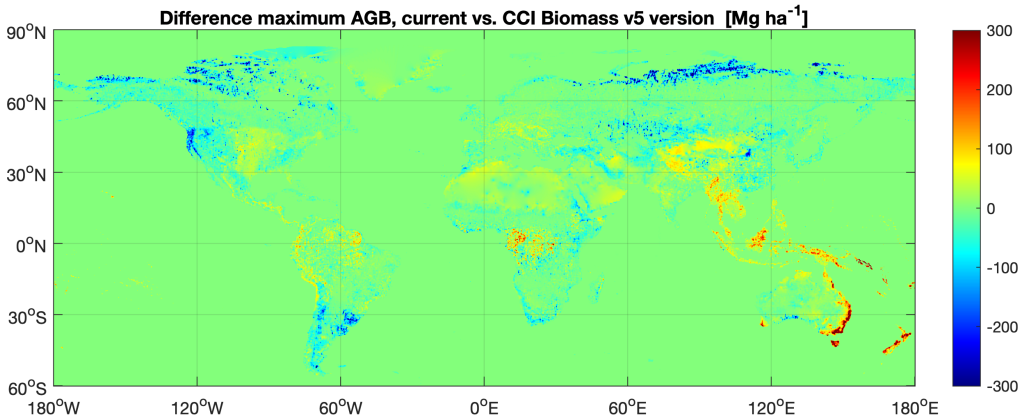


Figure 3-57: Difference of maximum AGB estimates from the current version (v5) and the previous version (v4).

The spatial patterns of maximum AGB correspond to those identified in the previous versions of the same layer but there are discrepancies in several regions. The maximum AGB in polar regions and some sparsely vegetated regions (e.g., Pampas and Kazakhstan) decreased to low values as a consequence of the filtering to remove unnatural high values of canopy height in sparsely vegetated regions. Improved filtering of the ICESat-2 data also led to a reduction of the maximum AGB in regions with steep topography (Pacific Northwest coast, Andes region and Himalaya). Improved structural functions in Southeast Asia and Australia led to a marked increase of the maximum AGB. The strong increase of the maximum AGB in the wet tropics in Africa are a consequence of having lowered the threshold on the number of laser photons per segment.

Although the layer of maximum AGB seems to have improved in each version of the CORE retrieval algorithm, it is not possible to quantify the accuracy of the current version. The accuracy of the current version can only be quantified indirectly through validation of the AGB maps against inventory data. Biases revealed by this exercise are usually explained because of an incorrect value of the maximum AGB.

The continual improvements to the layer of maximum of AGB imply variability of its values from version to version, which is illustrated by the temporal standard deviation of its values from v1 to v6 in Figure 3-58. The largest variability in time occurs in eastern Oceania, the Andes region, the western Congo Basin, peninsular Southeast Asia and the eastern US. In Asia and Oceania, the uncertainty is a consequence of continuous changes in the AGB data support and the stratification applied to characterized the structural function. The large difference of values in the Andes region and the eastern US is due to differences in the AGB data support (GlobBiomass dataset and NFI averages). Changes applied to the filtering of the ICESat-2 data explains the variability of the maximum AGB in the Congo Basin. Other regions of the world (e.g., Eurasian temperate and boreal forests, western and eastern Amazon) also showed remarkable fluctuations of the maximum AGB, which were explained by changes in the AGB data support. Changes in maximum AGB across versions of the CCI Biomass CORE retrieval algorithm imply variability of the corresponding AGB maps in the CRDP. The map in Figure 3-58 therefore indicates regions where the structural function relating height to AGB needs better characterization, i.e., where knowledge of forest structure needs to be improved.

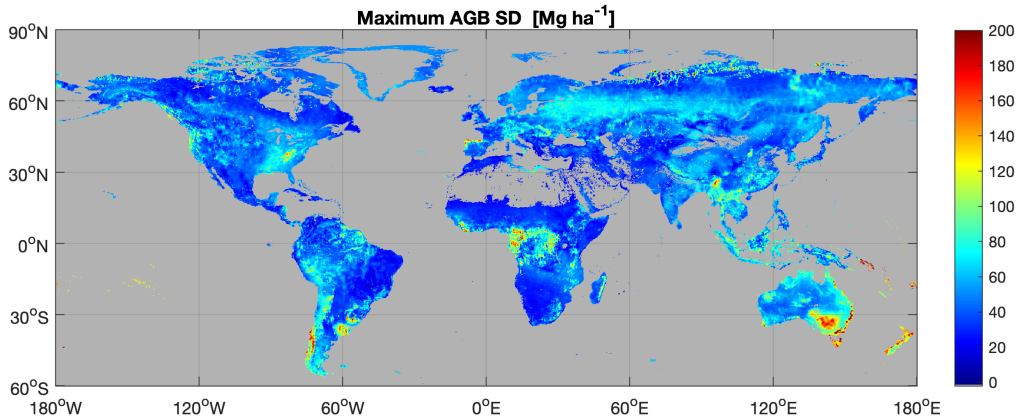


Figure 3-58: Map of the temporal standard deviation (SD) of the maximum AGB between v1 and v6 of the CCI Biomass CORE retrieval algorithm.

4 AGB estimation 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 ($m^3\ ha^{-1}$), 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 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 EO 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.

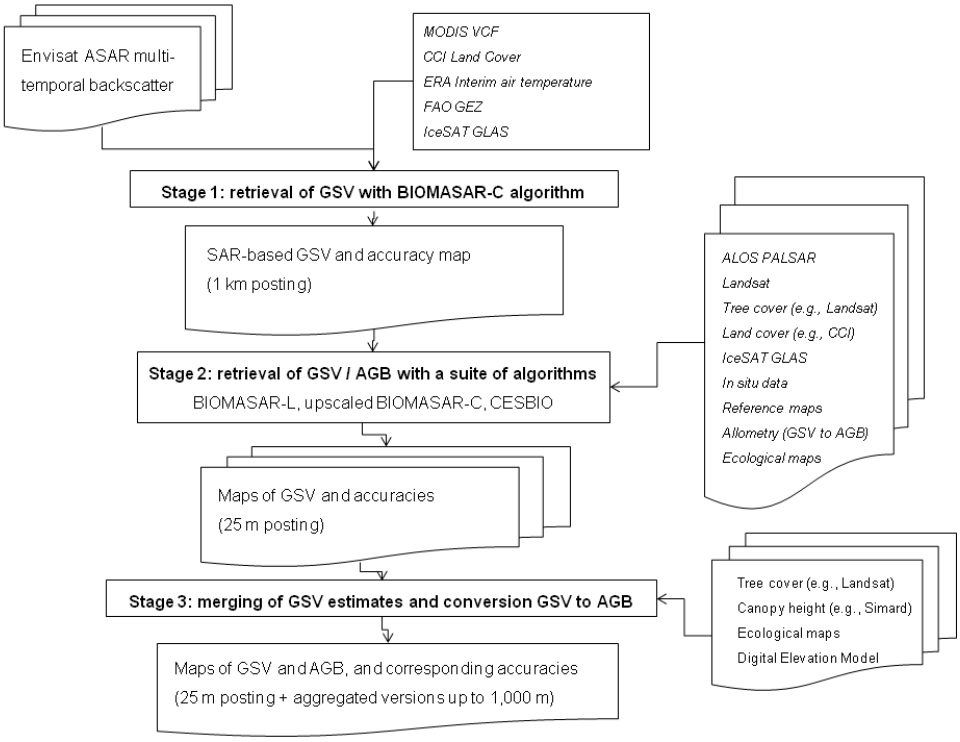


Figure 4-1: Flowchart of the GlobBiomass global biomass retrieval algorithm.

The three stages of the GlobBiomass algorithm were structured as follows (Figure 4-1**Error! Reference source not found.**).

- 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., 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 (Santoro et al., 2021), 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). Like 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 estimate of GSV and AGB has a corresponding estimate of its precision.
- 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-stages 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 to be used in this project.

The spatial resolution of the hyper-temporal dataset of Sentinel-1 C-band observations is substantially higher than in the GlobBiomass project (150 m vs. 1000 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.

We also consider that the L-band datasets available to this project consist of multi-temporal observations primarily from the ScanSAR mode, which comes with a spatial resolution of 50m and 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.

Since the start of the CCI Biomass project, the CORE algorithm evolved chronologically as follows:

- The GlobBiomass algorithm was adapted to Sentinel-1 and ALOS-2 datasets.

- Implement canopy height in the retrieval models by integrating a function relating canopy height-and-AGB.
- Estimate directly AGB rather than through an estimate of GSV. This step was also a consequence of reduced activity on the characterization between vegetation volume and organic mass in recent years.
- Identify ways to guarantee temporal consistency of AGB estimates from different satellite data sources (bias modelling, rescaling)

The biomass estimation procedure described in Figure 4-1 was detailed in Santoro et al. (2021) 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 functional flowchart that focuses on the interdependencies of datasets and algorithms in the current version of the CCI Biomass CORE global biomass retrieval algorithm (v6). For past versions of the CORE algorithm, we refer to older versions of the ATBD available at https://climate.esa.int/en/projects/biomass/key-documents/ (last access on 8 October 2024)

The CORE algorithm consists of the following processing steps.

- Multi-temporal C-band backscatter data are used to generate global estimates of AGB at 150 m pixel size using the BIOMASAR-C algorithm; these are then resampled to 100 m.
- Multi-temporal L-band backscatter data are used to generate global estimates of AGB at 100 m pixel size using the BIOMASAR-L algorithm.
- A set of auxiliary datasets is used to calibrate the parametric models embedded in the BIOMASAR algorithms since our strategy avoids *in situ* measurements for training.
- The AGB estimates from BIOMASAR-L and BIOMASAR-C are merged to reduce systematic errors in either of them.
- The precision of the estimates is characterized at each step shown in Figure 4-2**Error! Reference source not found.**, and an estimate of the AGB precision is attached to each 100 m pixel.
- The CORE algorithm can be expanded by linking it with additional datasets produced with algorithms that perform better than those proposed here.

Spatial averaging can 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.

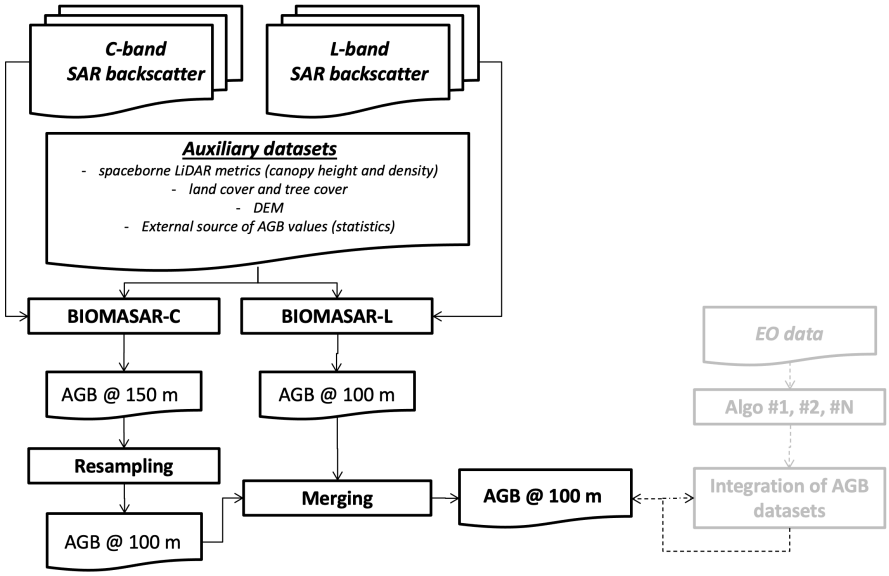


Figure 4-2: Functional dependencies of datasets and approaches forming the CCI Biomass CORE global biomass retrieval algorithm. The shaded part of the flowchart represents potential improvements following the implementation of additional retrieval techniques.

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

- BIOMASAR-C
- BIOMASAR-L
- Merging

4.3 The BIOMASAR-C algorithm

The theoretical basis of the BIOMASAR-C algorithm has been presented in Santoro et al. (2011, 2015a and 2021) and was extensively presented in Santoro et al. (2021). 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.

Figure 4-3 shows the flowchart of the current implementation of BIOMASAR-C, which is applied separately to Sentinel-1 and Envisat ASAR data (sigma0 SAR images). For each sigma0 SAR image, the forest backscatter model is trained, i.e., the model parameters are estimated, resulting in corresponding images for each of the two model parameters. The model training does not require reference measurements of biomass. The model parameters are estimated by relating canopy density (e.g., from a VCF type of product) and the backscatter observations in the sigma0 image. Once the model parameters are estimated, the forest backscatter model is inverted to express the biomass variable of interest (e.g., AGB) as a function of the measured backscatter. To allow for the estimation of AGB, it is necessary to reformulate the forest backscatter model by replacing the original predictors, namely canopy height and density, with AGB. This is achieved with the two functions relating forest variables, which require LiDAR metrics of canopy height and canopy density, and AGB reference data, here represented by statistics (mean value) derived from forest field inventory measurements. Note, that it is not mandatory to have access to the inventory measurements as the calibration of the function relating to AGB can equally be run on statistics published by NFIs.

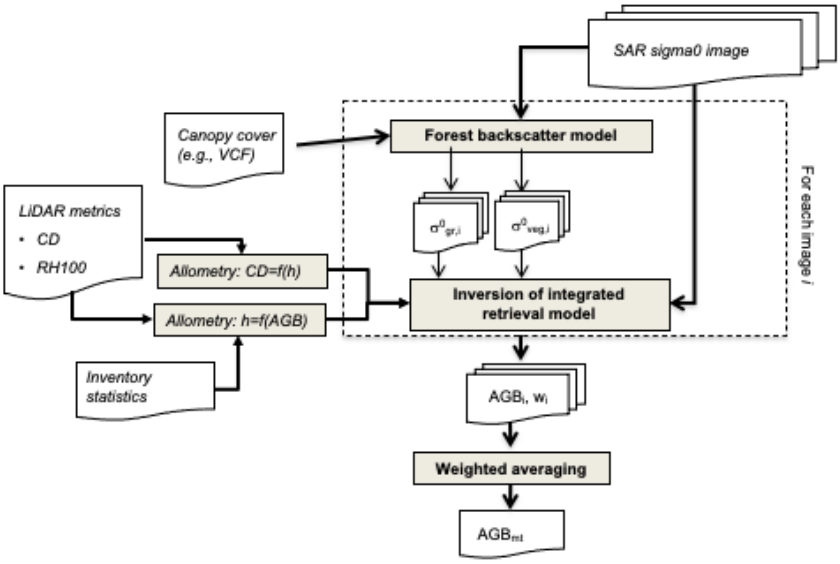


Figure 4-3: Flowchart of the BIOMASAR-C algorithm.

4.3.1 The forest backscatter model

At C-band, spatial and temporal variability of the backscatter make empirical modelling of biomass (whether as 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 this type of model in the retrieval of forest biomass (Santoro and Cartus, 2018) and the extensive knowledge gathered by the team developing the retrieval algorithm 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 expresses 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:

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

Here, η is the area-fill or canopy density factor, representing the fraction of the area covered by vegetation, σ^0_{gr} and σ^0_{veg} 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 the model relates two forest variables (canopy 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), Pulliainen et al. (1994) rewrote the original WCM in a form similar to Equation (4-1) that relates the SAR backscatter to stem volume, V, which can be considered equivalent to GSV:

$$\boldsymbol{\sigma^0_{for} = \sigma^0_{gr} e^{-\beta V} + \sigma^0_{veg}\left(1 - e^{-\beta V}\right)} \tag{4-2}$$

In Equation (4-2), β is an empirically defined coefficient expressed in ha m^{-3}. However, this coefficient has some physical meaning since, by comparing Equations (4-1) and (4-2), β, η and α are linked by the relation (Santoro et al., 2002):

$$\boldsymbol{e^{-\beta V} = 1 - \eta\left(1 - e^{-\alpha h}\right)} \tag{4-3}$$

Each of these terms represents the forest two-way transmissivity. 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 SAR backscatter may introduce systematic biases in the retrieval, regardless of the forest structure.

Equation (4-3) was used to generate GSV estimates part of the first CRDP of the CCI Biomass project. We refer to Santoro et al. (2021) for the description of BIOMASAR-C for GSV retrieval. Later versions of the CRDP only included AGB and were based on the original WCM with gaps in Equation (4-1) integrated with the models relating canopy density and canopy height in Equation (3-4), and the model relating canopy height and AGB in Equation (3-5). Here, it was assumed that the area-fill factor equals the optical canopy density in Equation (3-4).

As we want to express the backscatter as a function of AGB, we first need to invert Equation (3-7), thus obtaining $h = (b_1 \cdot AGB)^{b_2}$ with $b_1 = 1/p_1$ and $b_2 = 1/p_2$. The coefficients p_1 and p_2 were introduced in Equation (3-5). Equation (4-5) shows the forest backscatter (observable, σ^0_{for}) expressed as a function of AGB only.

$$\begin{aligned}\sigma^0_{for} = &\left[1 - \left(1 - e^{-q\cdot(b_1\cdot AGB)^{b_2}}\right)\left(1 - e^{-\alpha\cdot(b_1\cdot AGB)^{b_2}}\right)\right]\sigma^0_{gr} \\ &+ \left(1 - e^{-q\cdot(b_1\cdot AGB)^{b_2}}\right)\left(1 - e^{-\alpha\cdot(b_1\cdot AGB)^{b_2}}\right)\sigma^0_{veg}\end{aligned} \tag{4-4}$$

4.3.2 Model training

Regardless of whether the biomass variable of interest is GSV or AGB, the retrieval model has three unknowns, namely σ^0_{gr}, σ^0_{veg} and α. In addition, the model expressed as a function of GSV has an additional unknown: the coefficient of the forest transmissivity β.

If a set of reference GSV or AGB values is available, the model parameters can be estimated with a least squares regression to the reference measurements and the measurements of the SAR backscatter corresponding to the GSV or AGB observations. This approach is 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^0_{gr} and σ^0_{veg} by means of statistics of the backscatter for certain types of forest cover. For α, we currently assume 2 dB/m following a 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 yet been attempted. In case the response variable is GSV, the coefficient β is estimated from metrics of the ICESat GLAS waveforms. Refer to previous versions of the ATBD for the estimation of this model parameter.

Associating statistics of the backscatter to the model parameters σ^0_{gr} and σ^0_{veg} is referred to as self-calibration. Unlike previous versions of BIOMASAR-C when both model parameters were based on self-calibrated values, here we use the self-calibrated value of σ^0_{gr} as its estimate and the self-calibrated value of σ^0_{veg} as an initial estimate to be used in a least squares regression.

The self-calibrated value for σ^0_{gr} is here set as the 25th percentile of the backscatter distribution for pixels belonging to a class of low vegetation cover density ("ground" pixels). This percentile was found to lead to more accurate estimates of AGB than the median value or lower percentiles. The self-calibrated value for σ^0_{veg} is defined as the median of the backscatter for high vegetation cover density ("dense forest" pixels). The “ground” and “dense forest” pixels are extracted within an estimation window of finite size to allow the estimation of the two model parameters to adapt to the local conditions of the forest and of the backscatter. We apply a single canopy density threshold, i.e., pixels are labelled either as "ground" or as "dense forest" depending on whether the canopy density is below or above 30%. The estimation window corresponds to the size of a tile, i.e., 1° × 1°. It is assumed that sufficient samples are captured in each class with such a window size; a drawback can be the inclusion of pixels characterized by different properties of the backscatter because of different environmental conditions within the area covered by the tile.

The final estimate of σ^0_{veg} is then obtained by means of least squares regression between SAR backscatter observations and corresponding canopy density and height values with Equation (4-1) within the same estimation window. Here, σ^0_{gr} is known and equal to its self-calibrated value. In Equation (4-1), we replace canopy height with the model relating it to canopy density in Equation (3-4) to express the SAR backscatter as a function of canopy density only. The reason for this step is the availability of global datasets of canopy density. Even though global datasets of canopy height also exist, they are mostly based on inferences from optical data. Based on the physics behind optical data, it is assumed that canopy density derived from such data is more reliable than canopy height derived from the same type of data.

The tree cover density dataset used in the self-calibration and the regression steps is 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, areas not corresponding to natural vegetation are masked out because of their different features in terms of C-band backscatter, to avoid distorting estimates of σ^0_{gr}. For this, the CCI Land Cover dataset of the year closest in terms of availability to the year of the C-band data is used and pixels labelled as cropland, urban areas, bare soil, permanent snow and ice and water bodies are excluded from the training phase.

To account for the different levels of SAR backscatter depending on local incidence angle, the estimation of σ^0_{gr} and σ^0_{veg} is applied separately to each set of backscatter measurements and percent tree cover values characterized by a specific range of incidence angles. As a trade-off between precision, speed of computation and representativeness of the estimates, we divide 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 regress 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 is filtered for outliers, due, for example, to forest cover changes between 2010 and the year of acquisition of the SAR image.

The estimates of σ^0_{gr} and σ^0_{veg} obtained for a given image in a tile and for a given range of incidence angles are retained if based on at least 3 valid median values of the SAR backscatter. This avoids cases when only a small range of canopy densities is represented (e.g., very high or very low tree cover) thus causing erroneous estimates of one or the other parameter. When this requirement is not met, no estimate is associated with σ^0_{gr} and σ^0_{veg} for the specific image and range of local incidence angle.

To show the performance of the combined approach (i.e., self-calibration and regression) with respect to self-calibration only, we have selected four tiles along a north-south transect in Europe and Africa and therein a Sentinel-1 image.
Figure 4-4 to
Figure 4-7 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 represent well the levels of the backscatter of an unvegetated surface and the densest canopies. In contrast, 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 canopy density modelled with Equation (4-1) to confirm the suitability of this model relating canopy density and SAR backscatter. Note that these curves should not be interpreted as having a predictive meaning, since the model behind them was introduced only to estimate model parameters.

Figure 4-4 to
Figure 4-7 also show examples of σ^0_{gr} and σ^0_{veg} for each of the five ranges of local incidence angles. 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.

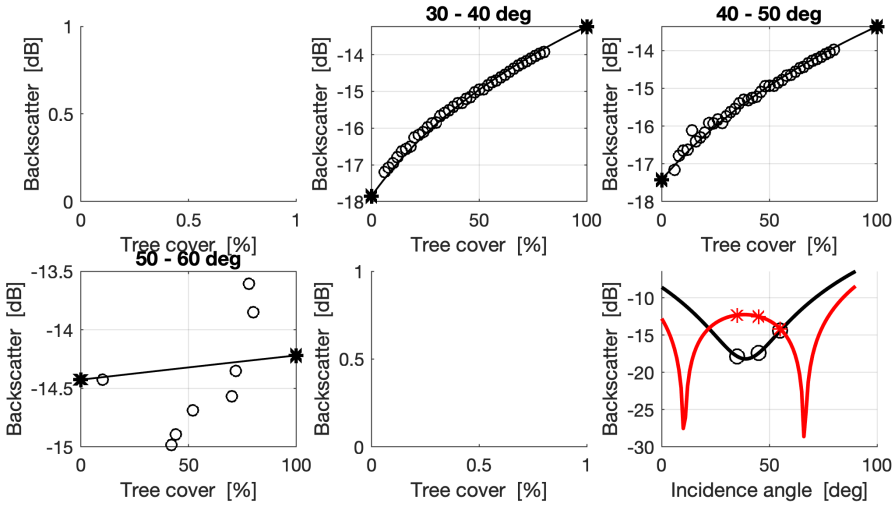


Figure 4-4: Panels with incidence angle range as title show (i) estimates of σ^0_{gr} and σ^0_{veg} obtained with the combined approach (black asterisks) and Equation (4-1) fitted to the median backscatter for a given canopy density (circles). The panel relating backscatter to incidence angle shows the estimates of σ^0_{gr} (black circles) and σ^0_{veg} (red asterisks) obtained with the combined approach and their quadratic fits (black curve for σ^0_{gr} and red curve for σ^0_{veg}) spanning the range of incidence angles between 0° and 90°. Dataset: Sentinel-1, VH-polarization., acquired during May 2017 along a descending orbit. Tile (top left corner coordinate): 64°N, 30°E (boreal forest). Note that the incidence angle ranges 20-30 deg and 60-70 deg were not represented in the dataset.

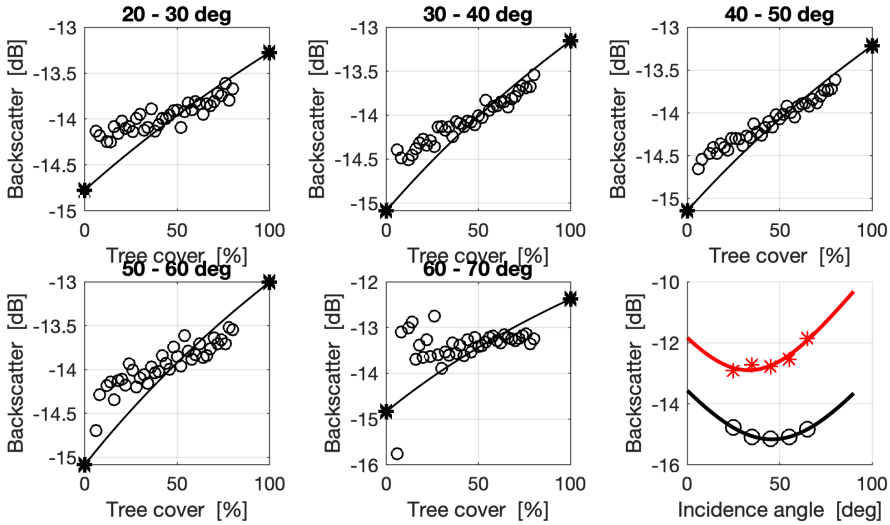


Figure 4-5: Same as in

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

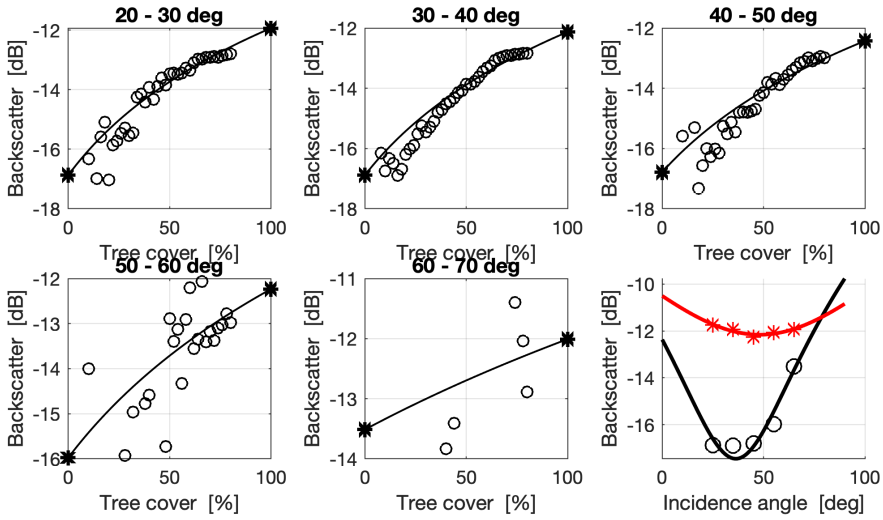


Figure 4-6: Same as in

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

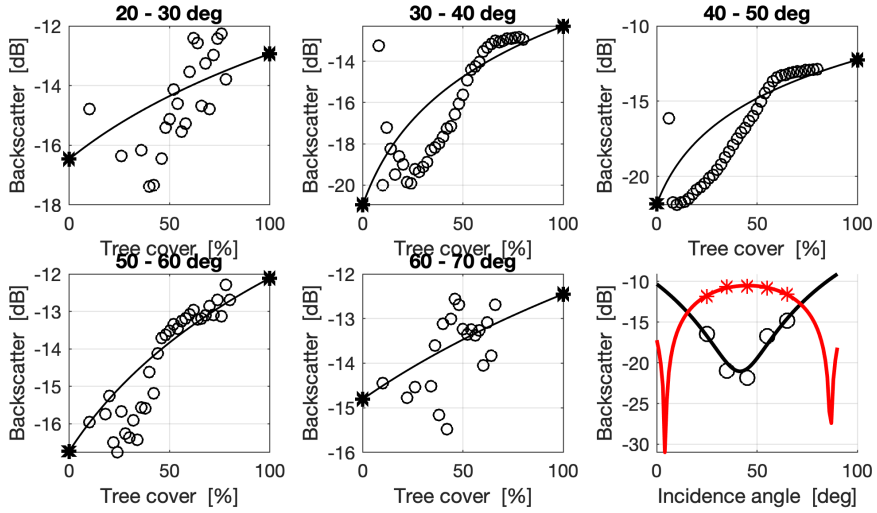


Figure 4-7: Same as in

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

Because of the multiple requirements imposed to obtain an estimate of σ^0_{gr} and σ^0_{veg}, it is likely that the 1° × 1° grid of estimates for a given SAR image may be incomplete (see e.g., Figure 4-4**Error! Reference source not found.**). In a post-processing step, we therefore interpolate spatially by infilling over valid estimates to fill gaps. This is done separately for each of the five intervals of incidence angle into 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 corresponding to the raster of local incidence angle for a given SAR backscatter image. Figure 4-8

shows an example of σ^0_{gr} and σ^0_{veg} rasters for the Sentinel-1 image used in Figure 4-6. The moderate topography in the 1° × 1° 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^0_{gr} shows a decrease of the backscatter for increasing incidence angle, as shown by the panel relating incidence angle and σ^0_{gr} estimates in Figure 4-6. For σ^0_{veg} we can see hardly any variability, as shown by the same panel in Figure 4-6. Both images however show the model parameter estimates follow the patterns of the local incidence angle. Figure 4-8 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^0_{gr} and σ^0_{veg} images because of the interpolation used to generate these images. The consequence is a seam in the AGB map obtained from the Sentinel-1 image (not shown here).

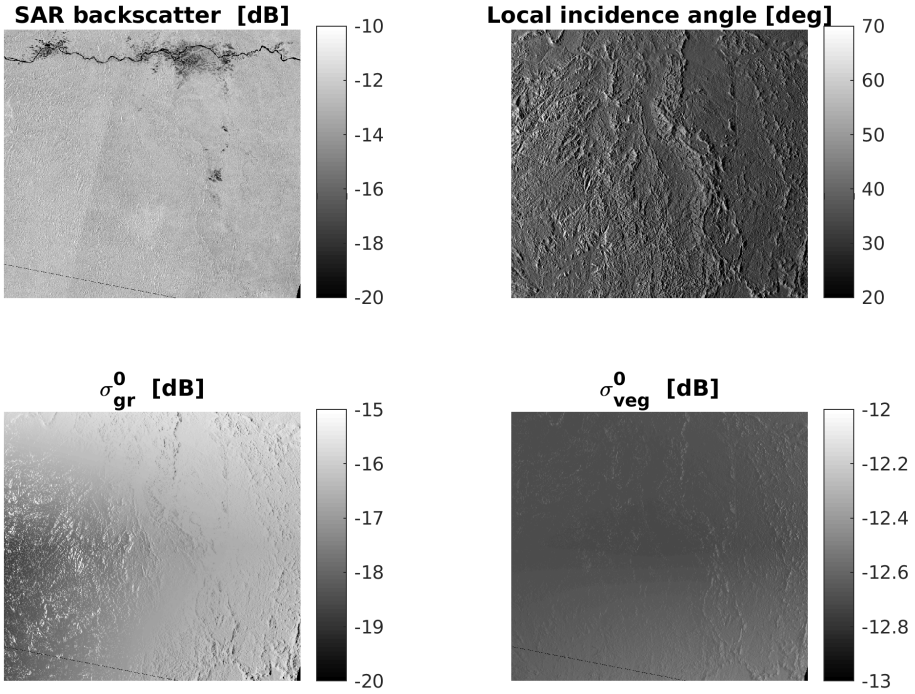


Figure 4-8: Illustrating the raster images of the estimates of σ^0_{gr} and σ^0_{veg} (bottom row) for the tile used in Figure 4-6. The top row shows the image of the SAR backscatter and the image of the local incidence angle.

4.3.3 Single image retrieval

The estimation of AGB from the backscatter model in Equation (4-4) requires numerical minimization 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.

Figure 4-9 shows the simulation of the Water Cloud Model in Equation (4-4) assuming that the maximum AGB is 362 Mg/ha. The backscatter first increases rapidly for 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. When a measurement of the backscatter falls outside this range, the inversion rule overrides the estimate because either it is negative or unrealistic. We define two intervals of backscatter to which the inversion either associates the maximum AGB or 0

Mg/ha depending on whether the measurement is above or below the range of modelled backscatter. These intervals account for the potential outliers of the backscatter. A backscatter value lying within an interval between the minimum or maximum modelled backscatter and this value plus 3 times the standard deviation of the measurements at low and high canopy density is associated with 0 and the maximum AGB, respectively. Otherwise, it is assumed to be unrealistic and discarded from the retrieval.

To visualize the difference between this approach and the approach that retrieves GSV, we also illustrate the model fit obtained with Equation (4-2), i.e., not exploiting functions that relate forest variables, in Figure 4-9. For a measurement of the backscatter, the integrated Water Cloud Model allows a higher AGB to be estimated for high AGB. This is an improvement considering the observation that BIOMASAR-C for GSV retrieval can be locally underestimated for high biomass forests.

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

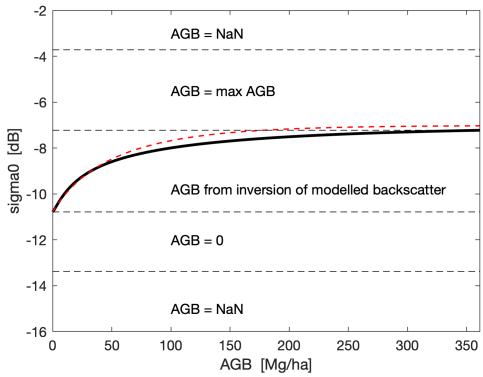


Figure 4-9: AGB retrieval rules depending on the level of the backscatter measurement.

4.3.4 Multi-temporal retrieval

Given N individual estimates of AGB from inversion of Equation (4-4), a weighted linear combination of the estimates is used to obtain the final estimate of AGB, AGB_{mt}, with Equation (4-5). This reduces the retrieval error with respect to each of the individual estimates (Kurvonen et al., 1999; Santoro et al., 2002; Santoro et al., 2011).

$$\boldsymbol{AGB_{mt}} = \frac{\sum_{i=1}^{N} \boldsymbol{w_i} AGB_i}{\sum_{i=1}^{N} \boldsymbol{w_i}} \tag{4-5}$$

The weights, w_i, in Equation (4-5) are defined as the vegetation-to-ground backscatter difference in dB, σ^0_{veg} - σ^0_{gr}, normalized by the maximum backscatter difference:

$$\boldsymbol{w_i} = \frac{\boldsymbol{\sigma^0_{veg,i} - \sigma^0_{gr,i}}}{\boldsymbol{max\left(\sigma^0_{veg,i} - \sigma^0_{gr,i}\right)}} \tag{4-6}$$

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 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.

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, 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 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 L-band backscatter. Like BIOMASAR-C, multi-temporal stacks of SAR backscatter observations are modelled individually, and biomass is obtained by integrating the estimates from the individual data takes. Major improvements in retrieval performance have been reported at C-band (Santoro et al., 2011, 2019; Cartus et al., 2019a, 2019b), but improvements could also be achieved at L-band (Santoro et al., 2006, 2015b; Cartus et al., 2012; Cartus et al., 2019a, 2019b), 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. One of the underlying assumptions of this model is that higher order scattering can be neglected. Although under typical conditions (rough forest floor, substantial attenuation in the canopy) stem-ground interactions can be neglected at L-band (Dobson et al., 1992; Pulliainen et al., 1999), a significant contribution from higher order scattering may arise from the canopy. Models indicate that higher order scattering effects are negligible for co-polarization, but not necessarily for cross-polarization. In Wang et al. (1998), higher order scattering increased the modelled L-HV backscatter from pine forest by 1.5 to 2 dB (at ~ 35° incidence angle). Karam et al. (1992) noted that higher order scattering from walnut orchards had a significant effect at HV polarization only for X-band, not L-band, regardless of the incidence angle. For a forested site in France, Picard et al. (2004) observed underestimation of L-HV backscatter when only first order scattering in a model was considered. However, when multiple scattering effects were included, the backscatter was overestimated compared to SIR-C L-HV measurements at 26° and 54° incidence angle. Picard et al. concluded that the improvements in 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 (4-1) combined with Equation (4-3) was used to support v1 of the CCI Biomass CRDP. Since v2, we considered 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, i.e., Equation (4-4).

As for 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 for BIOMASAR-C. 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.

The retrieval approach comprises three main steps:

1) calibration of the model in Equation (4-1) with the aid of a Landsat canopy density map, considering differences in the relationship between backscatter and canopy density dependent on incidence angle,
2) retrieval of AGB for each single acquisition in a multi-temporal stack of L-band observations by inverting Equation (4-4),
3) weighted multi-temporal combination of single image AGB estimates.

4.4.1 Model calibration

The model relating L-band backscatter to AGB is based on Equation (4-1). This 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 signal attenuation while passing through the canopy. In contrast to C-band, the two-way transmissivity term in the model ($\exp(-\alpha h)$) significantly affects the backscatter to η relationship (Cartus et al., 2018) as illustrated in Figure 4-10. Assuming a two-way attenuation of 0.5 dB/m, i.e., a value considered realistic for boreal forests (Praks et al., 2012; Askne et al., 2003), simulations based on Equation (4-1) demonstrate that the two-way transmissivity term is not negligible for L-band since the response of backscatter to changes in η depends strongly on the canopy height, in particular in low canopy height ranges (<15m).

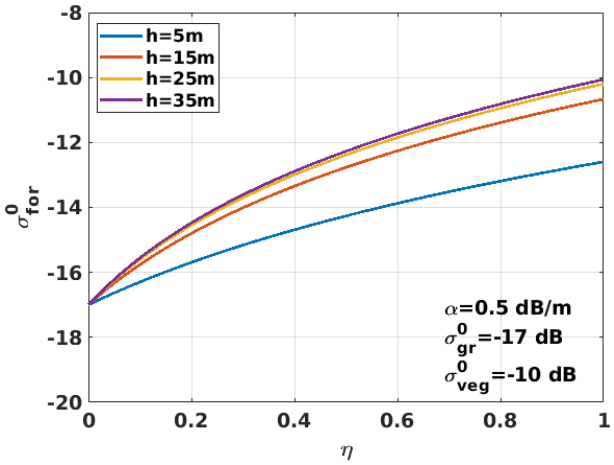


Figure 4-10: Simulated differences in the L-band HV backscatter response to changes in η for forests with canopy heights between 5 and 35 m.

Given knowledge of the local relationships between canopy density, η, and height h, the unknown model parameters σ^0_{gr}, σ^0_{veg}, and α can be estimated for each individual L-band backscatter observation with the aid of existing global maps of canopy density only (Hansen et al., 2013) when reformulating the original model in Equation (4-1) with height expressed as function of canopy density:

$$\sigma^0_{for} = (1-\eta)\sigma^0_{gr} + \eta\sigma^0_{gr}e^{-\alpha h(\eta)} + \eta\sigma^0_{veg}\left(1-e^{-\alpha h(\eta)}\right) \quad (4\text{-}8)$$

where

$$h = -\frac{\log(1-\eta)}{q} \quad (4\text{-}9)$$

Equation (4-9) is derived from Equation (3-4), where q characterizes the relationship between η and height.

The model in Equation (4-8) is calibrated for each orbit contained in a 1x1˚ tile of the ALOS-1/2 PALSAR-1/2 backscatter data provided by JAXA. To capture the dependence of the parameters on the local incidence angle, models are fitted separately for 10˚ wide ranges of the local incidence angle (i.e., 20-30˚, 30-40˚, 40-50˚, 50-60°, 60-70°, 70-80°). In previous versions of the BIOMASAR-L algorithm (up to version 5 of the CRDP), the model in Equation (4-8) was calibrated by fitting it to observed relationships between L-band backscatter and Landsat canopy density using least-squares regression. In this way, estimates for σ^0_{gr}, σ^0_{veg}, and α could be obtained. To produce version 6 of the CCI Biomass CRDP, we adopted a modified calibration approach like that originally implemented for the ESA GlobBiomass product (Santoro et al., 2021).. In this approach, estimates for σ^0_{gr} and σ^0_{veg} are obtained by identifying areas of sparse and dense forest cover and calculating the mean observed backscatter in such regions, respectively. The thresholds for identifying areas of low and high canopy density are defined adaptively for each 1° x 1° tile and backscatter image so that the average backscatter can be estimated across a minimum of 200 pixels. Before calculating the average backscatter in areas of low and high canopy density, areas are masked for which:

- land cover maps report cropland, bare ground, wetland or built-up areas,
- layover/shadow has been detected,

- annual Landsat-based forest cover change maps (Hansen et al., 2013) report gains or losses of forest cover between 2010, i.e., the year of the Landsat canopy density map (Section 3.8), and the year of the L-band backscatter image acquisition.

After masking out the aforementioned areas, the average backscatter in areas with sparse vegetation cover according to the Landsat canopy density is considered a direct estimate of σ^0_{gr}. The model parameter σ^0_{veg}, instead represents the backscatter from an ideal opaque forest canopy whereas the average backscatter observed over dense forest canopies, here referred to as σ^0_{df}, may still include significant ground contributions due to a forest layer with finite depth, remaining gaps in the canopy, and hence a two-way transmissivity larger than zero. To derive σ^0_{veg}, σ^0_{df} needs to be compensated for residual ground contributions with:

$$\sigma^0_{veg} = \frac{\sigma^0_{df} - \left(1 - \eta_{df} + \eta_{df} e^{-\alpha h_{df}}\right)\sigma^0_{gr}}{\eta_{df}\left(1 - e^{-\alpha h_{df}}\right)} \quad (4\text{-}10)$$

In Equation (4-10), η_{df} and h_{df} denote the canopy density and height of forests for which the average backscatter σ^0_{df} has been calculated. η_{df} is derived with the average Landsat canopy density of pixels considered for the estimation of σ^0_{df}. The average height of dense forests, h_{df}, is inferred from the gridded dataset of IceSAT-2 LiDAR heights representative of the densest forests.

In previous implementations of the BIOMASAR-L algorithm, the two-way tree attenuation coefficient was either defined with a global constant value of 0.5 dB/m or estimated when fitting the model in Equation (4-8) to observations of canopy density and L-band backscatter. In the latest version of the algorithm, used to produce version 6 of the CCI Biomass CRDP, we relied on spatially explicit estimates of the one-way tree transmissivity at nadir, α_{1w}, produced from SMOS and SMAP L-band radiometer data (Cartus et al., submitted). The two-way attenuation α in Equation (4-8) can be derived from the one-way attenuation α_{1w} coefficient with $2\alpha_{1w}/\cos(\theta_{loc})$. The radiometer-based estimates for α_{1w} are generally in the range of 0.2 to 0.3 dB/m, which translates to values for α of 0.5 to 0.75 dB/m.

The decision to estimate σ^0_{gr} and σ^0_{veg} via a self-calibration approach was motivated by the validation of previous releases of the CCI Biomass CRDP indicating that in semi-arid forest regions the CCI Biomass maps tended to underestimate the AGB in the lower AGB ranges. In most forest areas, the two different model calibration approaches, i.e., the regression-based approach deployed until version 5 of the CRDP, and the approach described above, lead to very similar results. This is exemplified in Figure 4-12a for a 1° x 1° large area in the Amazon Basin and all PALSAR-2 ScanSAR observations in 2018. The two model calibration approaches lead to almost identical results (red and blue curves). In contrast, in semi-arid forest regions such as Southern California (4-12b) and Mozambique (4-12c), the regression-based calibration approach tends to result in higher estimates for the parameter σ^0_{gr} and thus underestimation of the AGB particularly in low AGB ranges. The main advantage of a regression-based model calibration is the possibility to obtain estimates for σ^0_{gr} and σ^0_{veg} even when in the region of interest there are no forests with very sparse or dense canopies. However, there are major drawbacks to a regression-based parameter estimation:

- The latest Landsat canopy density map is available for the year 2010. Even when masking out areas where Landsat change products report gains or losses in canopy density, the uncertainty of the 2010 canopy density map, particularly in intermediate density ranges, is high. For model calibration via regression, such uncertainties in intermediate density ranges were found to affect the estimation of the model parameters. The requirement of

having a timely density map is clearly relaxed when focussing on sparse and dense forest only.

- Even when up to date, the reliability of optical density products has been reported to be lower in intermediate canopy density ranges and the estimates obtained in some ecoregions, e.g., semi-arid regions, tend to be biased, i.e., too low.

Once σ^0_{gr} and σ^0_{veg} 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 4-13). These polynomials are subsequently used to obtain spatially explicit estimates of the parameters using the local incidence angle maps (Figure 4-14). The initial model calibration is performed for each orbit on a 1x1˚ tile-by-tile basis. Since, within a given tile, only part of the incidence angle range in an ALOS-2 image may be represented, we here also consider derived estimates for the model parameters obtained for the backscatter images acquired from same orbit in adjacent tiles (in a 5x5˚ window). 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 error of the fit of the model in Equations (4-8) and (4-9) to the observed relationship between L-band backscatter and Landsat canopy density.

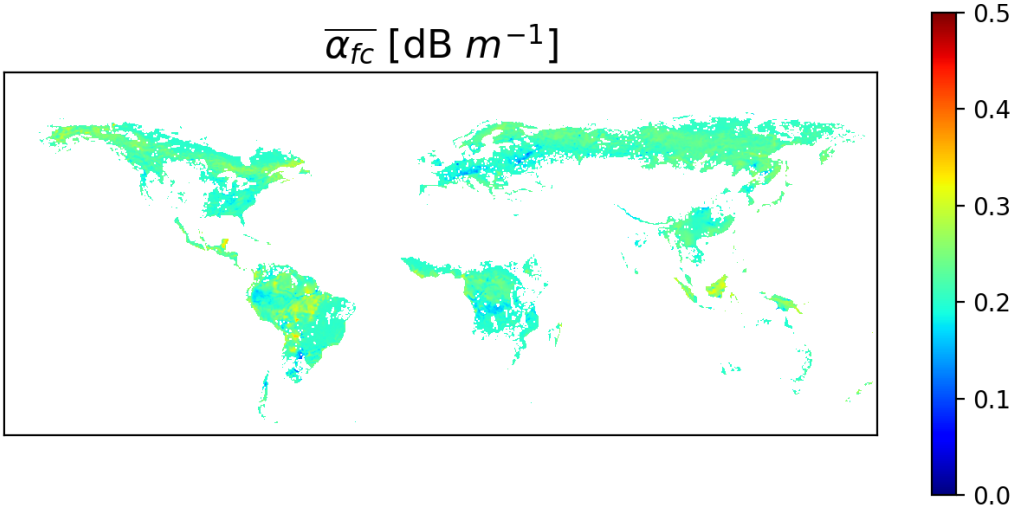


Figure 4-11. The average annual one-way attenuation at L-band derived from SOMS and SMAP L-band radiometer data (Cartus et al., submitted).

(a)

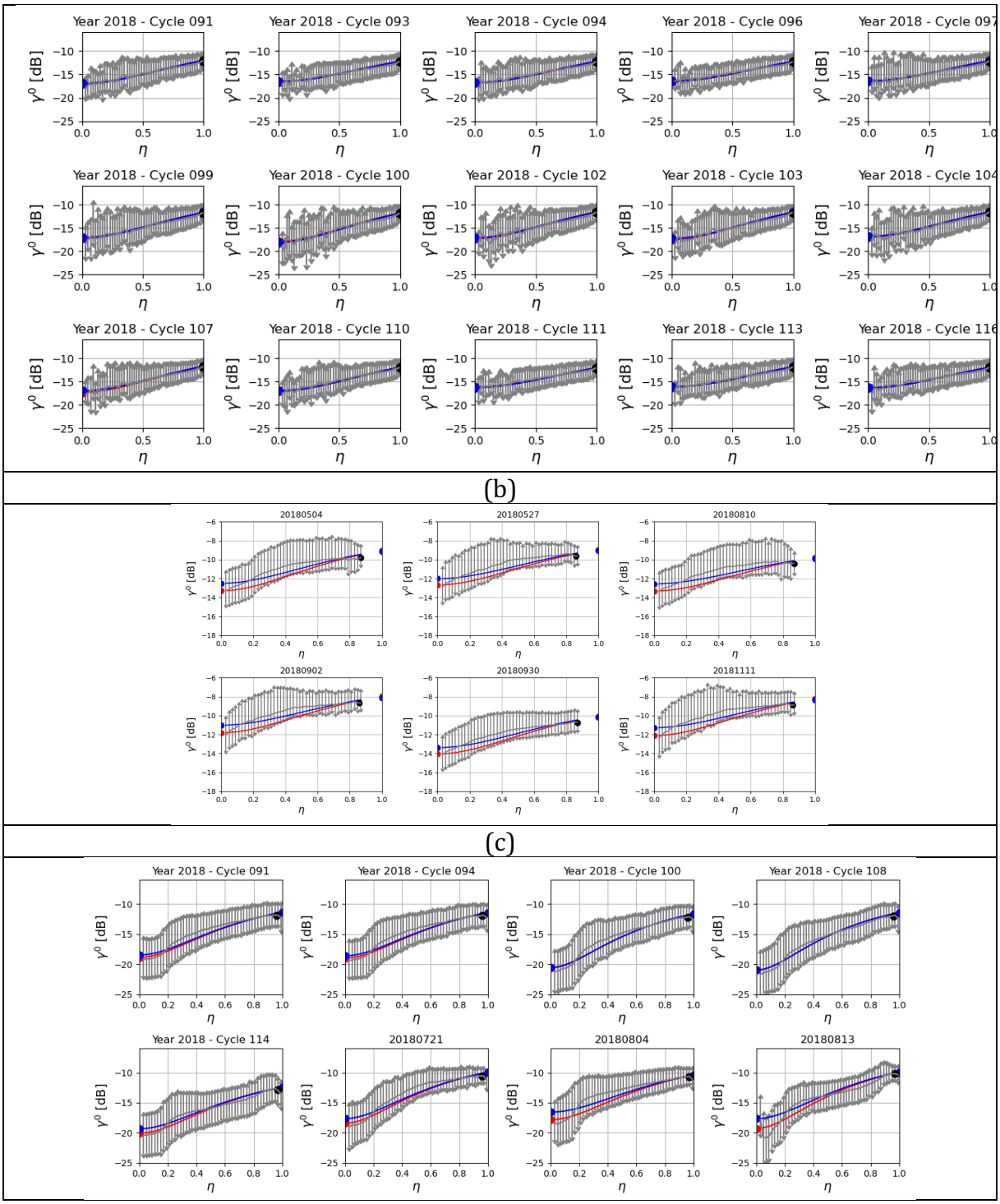


Figure 4-12. Observations of Landsat canopy density (Hansen et al., 2013) and ALOS-2 PALSAR-2 Fine-Beam and ScanSAR L-band HV polarization backscatter observations for three different 1° x 1° tiles covering the Amazon Basin (a), Southern California, USA (b), and Eastern Mozambique (c). The curves represent the fit of the model in Equation 4-8 when estimating σ^0_{gr} and σ^0_{veg} via least-squares regression to the observed relationship between backscatter and Landsat canopy density (blue curves) or based on thresholds of canopy density to calculate the average backscatter in correspondence of sparse (σ^0_{gr}) and dense (σ^0_{veg}) forest vegetation (red curves). In the case of Fine-Beam imagery, the figure reports the acquisition date. In the case of ALOS-2 ScanSAR data the observation cycle is indicated.

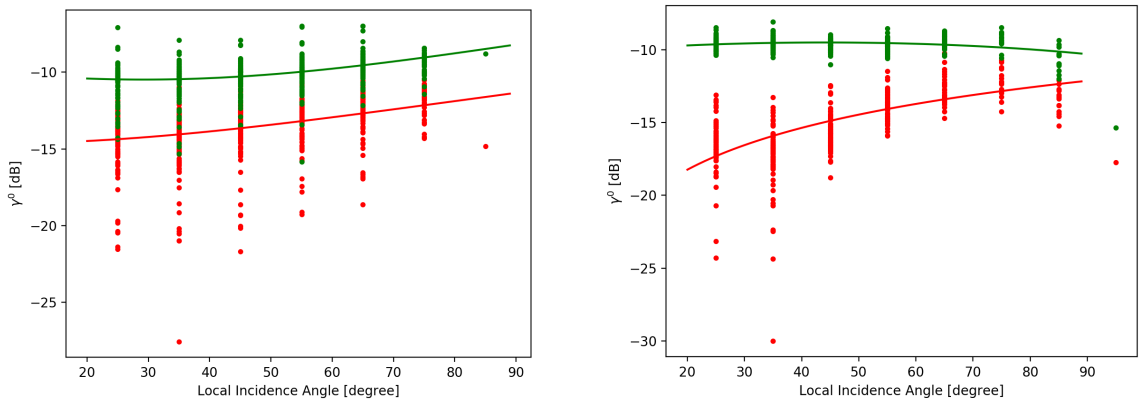


Figure 4-13: Estimates for the model parameters σ^0_{gr} (red dots) and σ^0_{veg} (green dots) as function of the local incidence angle for ALOS-2 orbits covering the Swiss Alps (left) and the Pacific Northwest of the United States (right). The curves represent the corresponding fit of a second order polynomial.

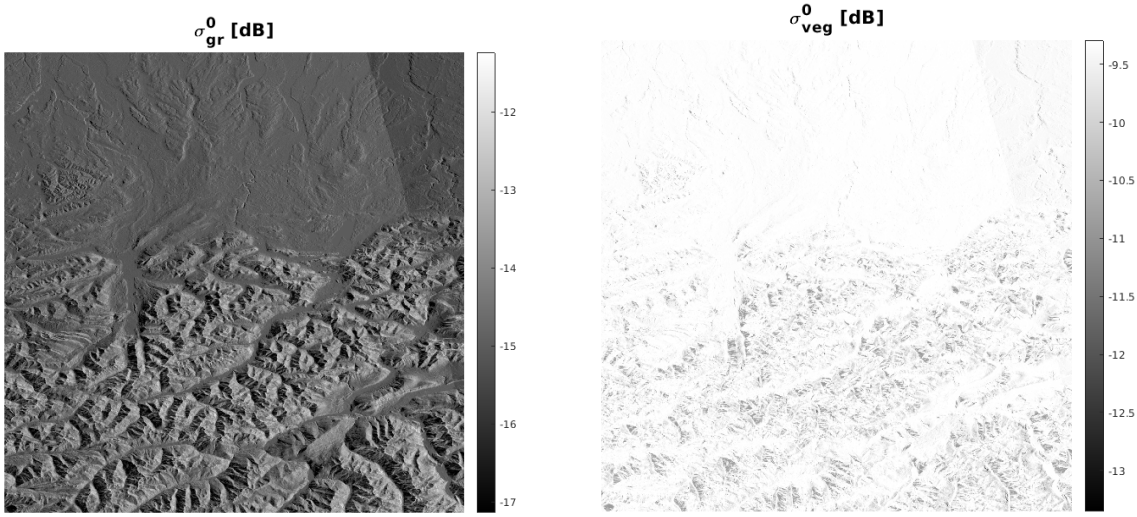


Figure 4-14: Spatially explicit estimates for the model parameters σ^0_{gr} and σ^0_{veg} for a 1x1˚ ALOS-2 tile (year 2016) covering the Swiss Alps.

4.4.2 Single image retrieval

Once the unknown parameters σ^0_{gr} and σ^0_{veg} are known, estimates for the parameter q, p_1 and p_2 relating canopy density, height and AGB may be used to retrieve AGB for each L-band backscatter image in the multi-temporal stack of observations using Equation (4-7). However, direct inversion of the model is not possible, and a numerical minimization approach is required. As in the case of the retrieval of BIOMASAR-C, the retrieval is constrained to a maximum AGB.

4.4.3 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 for BIOMASAR-C.

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 available for multi-temporal combination, or the quality of the pre-processing (calibration, topographic correction, geocoding). It is therefore advisable to combine the estimates that were obtained independently considering the respective strengths and weaknesses of each dataset.

In general, the purpose of integrating datasets is to 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 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

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. The weights are therefore produced at a scale of 0.01° 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 change 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 canopy density, η, canopy height, h, and two-way attenuation coefficient, α:

$$T_{for} = 1 - \eta\left(1 - e^{-\alpha h}\right) = 1 - \eta(h(AGB)\left(1 - e^{-\alpha h(AGB)}\right) \quad (4\text{-}10)$$

and may be formulated as function of AGB when relating η and h to AGB using Equations (3-4) and (3-5). As a consequence of the 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 attenuation terms, α, are generally required to describe the transmissivity as a function of biomass at C-band than at L-band. This is exemplified in Figure 4-15, which shows the transmissivity as a function of AGB for two different values of α to represent attenuation at C- and L-band. The derivative of Equation (4-10), which may be considered an indicator of the sensitivity of the signals to biomass, is therefore more negative for higher values of α (i.e. for C-band) in the lower ranges of AGB. In the higher ranges of AGB, instead, the derivative is higher for the lower values of α (i.e., for L-band).

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 (4-10), which in the case of AGB is written as:

$$\boldsymbol{w_M} = \left|\frac{\partial T_L}{\partial AGB}\right| - \left|\frac{\partial T_C}{\partial AGB}\right| \qquad (4\text{-}11)$$

The transmissivities at L- and C-band, T_L and T_C, are determined using values for α of 2 and 0.5 dB/m for C- and L-band, respectively, the local models relating forest variables, and a reference AGB map produced by converting a gridded dataset of the of the average canopy height according to IceSAT-2 LiDAR data at 0.01° x 0.01° resolution to AGB using Eq. 3-5.

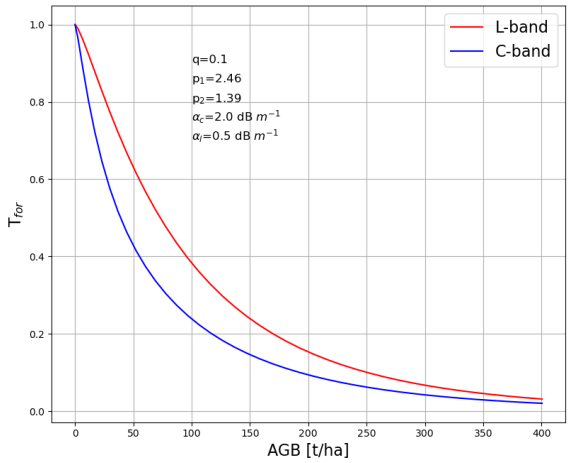


Figure 4-15: Forest transmissivity at C-band and L-band modelled as a function of AGB using forest structural parameters (q, p_1, p_2) appropriate to tropical forests of the Amazon Basin.

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, a limited number of observations per year and within a few seasons (primarily summer) acquired in the FB modes of ALOS-1 PALSAR-1 and ALOS-2 PALSAR-2 were available for most of the northern hemisphere, which is a 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, many Sentinel-1 and Envisat ASAR 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 tended to be 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. Although the Sentinel-1 dataset was reduced to monthly averages, the number of observations per pixel was still higher than the L-band datasets. Previous experience with 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 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 added to the weighting by scaling the original weight defined in Equation (4-12) by the square root of the number of observations available at C- and L-band at a given pixel location:

$$\boldsymbol{w_M} = \left|\frac{\partial T_L}{\partial AGB}\sqrt{N_L}\right| - \left|\frac{\partial T_C}{\partial AGB}\sqrt{N_C}\right| \quad (4\text{-}12)$$

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

$$\boldsymbol{w_{M,norm}} = \frac{(w_M - min(w_M))}{max(w_M - min(w_M))} \quad (4\text{-}13)$$

The rescaling is undertaken separately for each FAO ecoregion. Given the way Equations (4-12) and (4-13) are defined, w_M is greater in higher ranges of biomass and low in lower ranges of biomass (i.e., $w_{M,\ norm}$ 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 $1\text{-}w_{M,norm}$.

4.5.3 Final weight for merging

The individual AGB maps produced with BIOMASAR-C and BIOMASAR-L do contain artefacts because of processing related issues in the calibration and topographic correction of the C- and L-band SAR data. Furthermore, an incomplete multi-seasonal coverage, particularly in the case of the L-band imagery, may lead to local differences in the annual AGB maps produced with either of the two datasets that exceed the possible range of biomass changes on the ground due to growth, mortality, degradation, etc. To maximize the inter-annual agreement between the maps, the weights for each year, calculated as described above, were refined locally considering the stack of C- and L-band derived biomass maps for the years 2007, 2010, and 2015 to 2022. A refined weight for each of the ten years was obtained by minimizing a cost function (CF):

$$CF = \sum_{y1=2007}^{2021} \sum_{y2=y1+1}^{y1+2} \left|\left(B_{C,y1} + w_{L,y1}\left(B_{L,y1} - B_{C,y1}\right)\right) - \left(B_{C,y2} + w_{L,y2}\left(B_{L,y2} - B_{C,y2}\right)\right)\right|^2 \quad (4\text{-}14)$$

based on the differences between the merged C- and L-band derived biomass maps for any given year and the merged biomass estimates for the following (one or two years). In the minimization, the weights determined as described in Section 4.5.2 serve as starting point for the optimization. The search space for the new weights is constrained to a range of +/- 10% of the initial value. Since this is computationally demanding, the weight refinement is carried out at an aggregated scale of 0.01°, i.e., after aggregating all C- and L-band derived maps to 0.01° resolution. In the weight refinement, however, tropical evergreen forests, as classified in the CCI Land Cover maps for the given year, were not considered because of potential biases in the BIOMASAR-C maps in this type of forest. Instead, the weights were locally adjusted so that all weight was assigned to the BIOMASAR-L maps. In addition, weights were over-ruled in areas where land cover maps

report cropland, since the BIOMASAR-L maps tend to exhibit less erroneous biomass assignments over crops.

The approach followed for weight refinement may be seen as a temporary, rather conservative, solution with respect to the quantification of biomass changes throughout the time series of biomass maps (2007, 2010, 2015 to 2022) since the resulting series of maps present biomass changes only when unequivocally detected by both the C- and L-band derived products.

An example of weights that have been determined for the BIOMASAR-L map produced for the year 2020 is illustrated in Figure 4-16. Although the illustration was derived for the previous version of the CORE retrieval algorithm (v5), the distribution of the weights was not altered in the current version of the CORE algorithm. The merging weighted the C- and L-band datasets almost equally except for: (i) tropical rainforests, where it relied exclusively on the BIOMASAR-L AGB estimates due to under-prediction by BIOMASAR-C and (ii) short forests and vegetation, where the proportion of BIOMASAR-L AGB estimates dominated. In other parts of the world the merged map represents a mix of BIOMASAR-C and BIOMASAR-L with generally increasing weights assigned to BIOMASAR-L in low and high biomass forests.

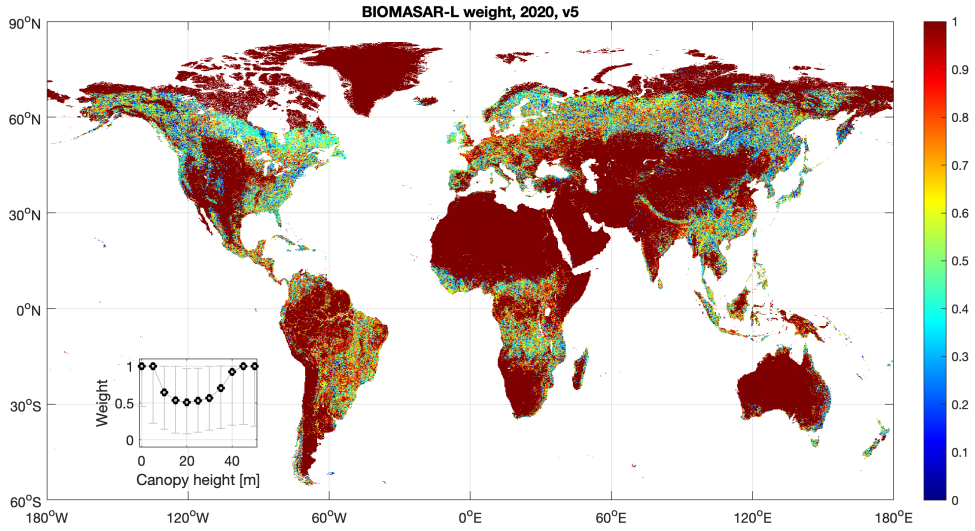


Figure 4-16: Map of the weights used to merge the BIOMASAR-C and BIOMASAR-L maps for the year 2020. The map shown represents the weights applied for the L-band products. The inset shows the distribution of the weights as a function of canopy height from the ICESat-2 dataset. For each bin of canopy height, the circle and the vertical bar represent the median and the interval between the 5th and the 95th percentile of the weights.

5 AGB change estimation methods

AGB changes between two epochs can be quantified either by differencing signals assumed to be sensitive to AGB or by differencing estimates of AGB. Differencing signals is viable if a change in the signals can be attributed to a change in biomass. The specifications of the data products to be delivered by the CCI Biomass project can only be achieved through a combination of high-resolution satellite data. In addition, for each epoch to be mapped, the satellite dataset consists of a different set of observations. A biomass change product based on signal differences is therefore not feasible.

The AGB change estimates obtained in CCI Biomass are defined as the difference between AGB estimates at two points in time (AGB_{y2} and AGB_{y1}). Since each value is an estimate of the true AGB, an additional term, representing the AGB bias ($Bias_{y2}$ and $Bias_{y1}$) needs to be accounted for. The bias term is relevant when the set of predictors used to estimate AGB at each point in time is not the same, as in the case of the three epochs targeted in CCI Biomass

$$\Delta AGB = \left(AGB_{y2} - Bias_{y2}\right) - \left(AGB_{y1} - Bias_{y1}\right) \quad (5\text{-}1)$$

In principle, the bias may vary in space and with AGB value, so it is hard to characterise and remove.

6 AGB bias estimation

The evaluation of the AGB maps against plot-based values of AGB [RD-8] revealed that the CCI Biomass maps are affected by biases. The cause of these biases is often an interplay of multiple factors, including the weak sensitivity of the EO data to biomass, the models relating biomass to the EO data and the uncertainty of the auxiliary layers used to calibrate the retrieval model.

The AGB bias is estimated with the non-parametric Random Forest (RF) regression approach (Breiman, 2001), which is an ensemble model of decision trees from bootstrapped samples of training data that produces averaged predictions (Araza et al., 2022).

Several RF models were initially tested with a set of many covariates that could possibly influence the bias (Chave et al., 2004; Rejou-Mechain et al., 2014), including the AGB estimates themselves and their standard deviation [RD-5], terrain elevation, slope and aspect angles, tree cover, precipitation, temperature, biome, longitude and latitude. Using all and partial combinations of the covariates, multiple RF models using the default RF hyper-parameters were tested. The models were evaluated using a randomly held out 30% of the 0.1° data to assess the proportion of the variance of residuals explained by the model. We then visually inspected the bias for indications of geographic correlation among covariates, as suggested in Meyer et al. (2019). After this initial investigation, the covariates were limited to AGB, AGB uncertainty, tree cover and slope.

The predictive power of the covariates in the RF model is assessed by the Variable Importance Measure (VIM) and Partial Dependence Plots (PDP). VIM is the mean decrease in accuracy of an RF model after data permutation of a covariate. When evaluated against the CCI Biomass dataset of 2017 produced in year 1, the VIM ranked the AGB estimates as the most important predictor, followed by tree cover density, AGB standard deviation, slope and aspect angle, with a proportion that decreased from 26% to 17%(Figure 6-1).

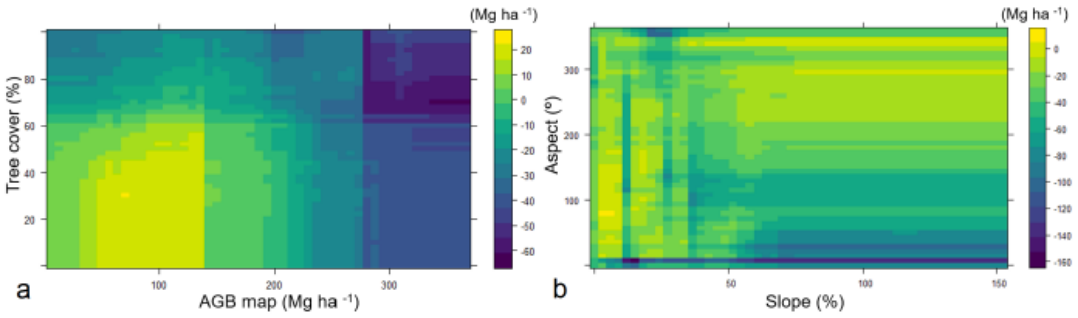


Figure 6-1: PDP of predicted bias as a function of CCI Biomass 2017 AGB and tree cover at 0.1° (a) and slope and aspect angle (b). A PDP shows the marginal effect of covariates on bias prediction. The PDPs are displayed as matrices, color-coded with bias and with the axes labelled by the values of a covariate pair, e.g., bias plotted against AGB map and tree cover.

In CCI Biomass, the bias is modelled at 0.1° to form weighted bootstrap samples. Refer to [RD-7] for the rationale underpinning the choice of this spatial scale in the process of bias evaluation and bias modelling.

7 Results

This Section presents mapping results based on the current version (v6) of the CORE algorithm (from C- and L-band data separately and merged estimates) and compares them with results obtained with previous versions of the CORE algorithm.

7.1 AGB estimation

7.1.1 BIOMASAR-C

In previous versions of this ATBD, we compared the performance of the CCI Biomass CORE algorithm with the GlobBiomass algorithm (https://climate.esa.int/en/projects/biomass/key-documents/). For AGB maps based on Sentinel-1, we demonstrated the improved performance of BIOMASAR-C, thanks to the higher spatial resolution of the C-band data (20 m vs. 1,000 m) and the availability of cross-polarized backscatter. We then compared the performance of BIOMASAR-C with the semi-empirical terms in the WCM and with the models relating forest variables. The accuracy of the AGB estimates improved when these models were integrated into the forest backscatter model, confirming the plausibility of the integrated modelling framework.

The performance of BIOMASAR-C using Sentinel-1 data is shown in the examples in Figure 7-1**Error! Reference source not found.**. The area in the panel on the left-hand side is intensively managed boreal forest with frequent clear-cuts (rectangular shapes) and regenerating forests (light green areas). The level of detail is high, as demonstrated by the linear features (deforested corridors), despite the moderate resolution of the Sentinel-1 data used in CCI Biomass (150 m). This is a consequence of the multi-temporal approach followed in BIOMASAR-C, which filters out temporally uncorrelated noise and maximizes the part of the backscatter signal related to forest structure. The area in the panel on the right-hand side includes undisturbed rainforests in the

Brazilian Amazon to the north and disturbed forests to the south (fishbone pattern). The multi-temporal features of the Sentinel-1 dataset allowed clear distinction of the two levels of biomass.

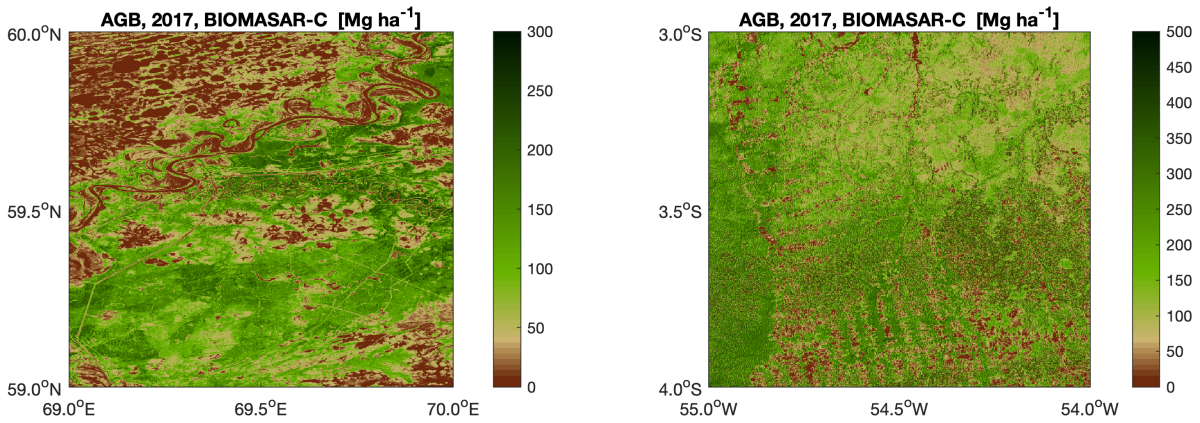


Figure 7-1: Examples of AGB estimates obtained with BIOMASAR-C applied to 1° × 1° tiles of Sentinel-1 data acquired in 2017.

The performance of BIOMASAR-C for an extended region is shown in Figure 7-2, with an AGB map of the Alpine range. Artefacts are infrequent and the map is in line with the impression given by the Sentinel-1 colour composite in Figure 3-2 that suggested a spatially consistent representation of SAR backscatter globally. Topography disturbs the estimates of AGB only locally, whereas the spatial distribution appears to be well captured in such a patchy landscape consisting of forests, cropland, lakes and urban areas. The latter are a systematic issue because they are associated with the highest possible AGB because of the very high backscatter. The same occurs over moraines in the Alps where the soil roughness occasionally causes high backscatter, depending on the local incidence angle. This experimental evidence highlights the necessity of masking non-forest areas or accounting for the different levels of reliability of BIOMASAR-C and BIOMASAR-L in the merging process.

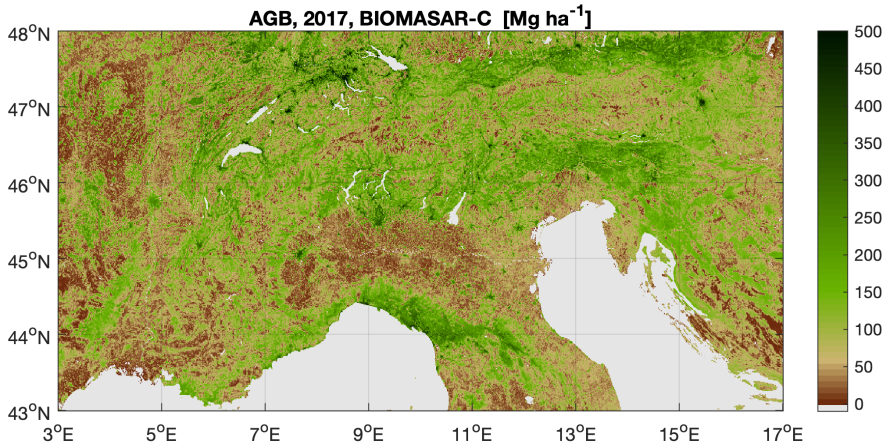


Figure 7-2: Estimates of AGB for the region of the European Alps from Sentinel-1 imagery acquired in 2017 using BIOMASAR-C.

The global AGB map for 2017 obtained with BIOMASAR-C is shown in Figure 7-3**Error! Reference source not found.**. The coverage is almost global with a few minor gaps in north Canada and Russia. Extended unvegetated regions (Greenland, Canadian Isles, Sahara) were not mapped because the computing efforts required to process the Sentinel-1 images would have been disproportionate with respect to the real benefit to the global map. Note that in such regions BIOMASAR-L produced reliable estimates of AGB from ALOS-2 data using however only a minor fraction of the computing resources that would have been needed to process the Sentinel-1 dataset.

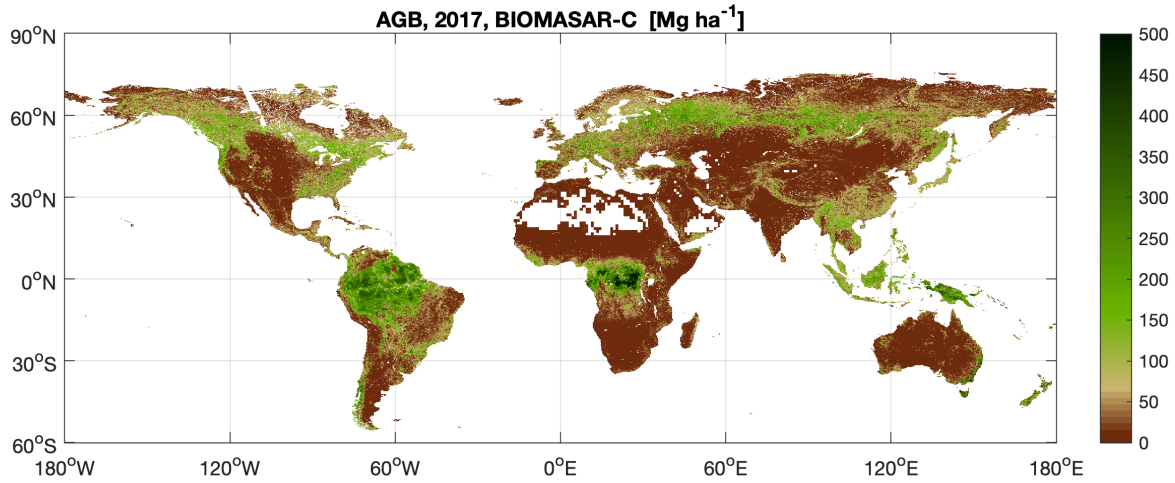


Figure 7-3: Map of AGB obtained with the BIOMASAR-C algorithm applied to the multi-temporal dataset of Sentinel-1 backscatter observations of 2017. Pixel size: 150 m.

The spatial distribution of AGB is consistent with known patterns. High biomass was estimated in the wet tropics as well as along the Pacific Northwest coast, the Andes, and Southwest Australia. Low biomass was estimated in dry regions of the world (savannas and tundra). The gradient of decreasing AGB from the Equator to the North Pole is also well represented. Abrupt variations of AGB occurred occasionally along the edge of two adjacent Sentinel-1 orbits or swaths. Figure 7-4 shows two examples of AGB offsets. The left panel of Figure 7-4 covers an area in the Southeast U.S. The AGB offset is due to an incorrect estimate of the noise floor, which appears as banding in the Sentinel-1 IWS images. This banding was also observed over Southeast China and cannot be undone. The right panel of Figure 7-4 instead includes an area in Gabon characterized by dense tropical forest where slight differences in the estimates of the model parameter in adjacent scenes translated to clear offsets. Even though we tried to adapt the WCM to the local incidence angle, such artefacts appeared because the sensitivity of the SAR backscatter to AGB was a fraction of a dB so that even a slightly incorrect estimate of one of the WCM parameters in images from adjacent orbits resulted in an AGB offset.

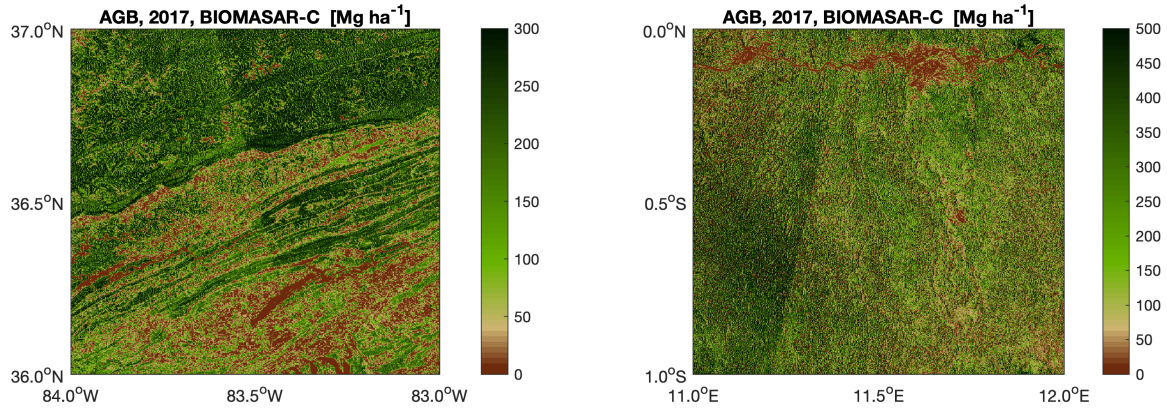


Figure 7-4: Examples of AGB estimates obtained with BIOMASAR-C for 1° × 1° tiles and characterized by AGB offsets across seams in Sentinel-1 imagery.

Figure 7-5 shows a quantitative assessment of the AGB estimates obtained with BIOMASAR-C. The reference consists of NFI-based averages of AGB per administrative or ecological unit, which were used to set up the model between canopy height and AGB. Accordingly, we computed the AGB average for each unit included in the reference dataset. The comparison is not per se a validation of the BIOMASAR-C estimates but helps to assess the quality of the output of BIOMASAR-C as regards the occurrence of systematic biases. The spatial patterns of AGB appear to have been reproduced with overpredictions in regions with AGB < 200 Mg ha^{-1}, primarily in Europe. The wide range of backscatter values for unvegetated terrain affected the estimation of σ^0_{gr} leading to smaller values than in reality and a modelled backscatter lower than the measured values in the low-to-moderate range of AGBs. The regions most distant from the identity line mainly correspond to small administrative units in Asia and Central Europe, with a small forest fraction

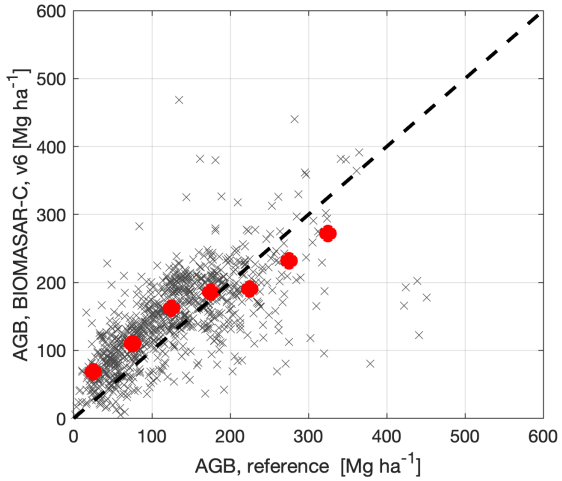


Figure 7-5: Scatterplot comparing averages of AGB from NFIs at administrative or ecological region with corresponding averages from the BIOMASAR-C AGB map of 2017 (v6). The red markers report the median AGB from the BIOMASAR-C AGB map for 50 Mg ha^{-1} wide bins of reference AGB.

7.1.2 Comparing BIOMASAR-C versions in CCI CORE algorithms

Similar to what was reported for the BIOMASAR-C v6 AGB map, the scatterplot in Figure 7-6 compares the NFI-based averages of AGB with the corresponding average values from the last set of estimates of the BIOMASAR-C map. The overall trend remains unchanged with overall lower AGB estimates. These results relate to the fact that no major structural change was introduced except for a new definition of the starting value for the regression of the WCM to estimate σ^0_{veg}. Because of the weak sensitivity of the C-band backscatter to AGB in many regions of the world, this slight change caused a substantial increase of the AGB estimates. Modifications introduced in v6 concerned the function relating height to AGB in Southeast Asia and Australia and, thereof, maximum AGB, which likely caused some changes in data points characterized by high biomass.

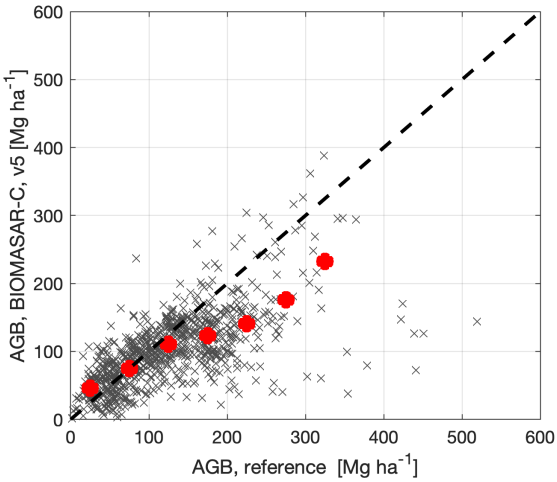


Figure 7-6: Same as in Figure 7-5 but based on v5 of the BIOMASAR-C algorithm.

7.1.3 BIOMASAR-L

An analysis of previous BIOMASAR-L estimates compared to the GlobBiomass dataset was reported in earlier versions of this document. As for BIOMASAR-C, we first show two full resolution maps (Figure 7-7) of the AGB estimates for the same 1° × 1° areas used in Figure 7-1. The L-and C-band estimates present similar spatial patterns but the former exhibits stronger contrast between unvegetated regions and dense forests. With BIOMASAR-L, the estimated AGBs were closer to zero in unvegetated regions and closer to the maximum biomass in dense forests. The comparison of the C- and L-band estimates indicates that the two datasets (i.e., Sentinel-1 and ALOS-2) do not produce the same levels of AGB, which corroborates our view that merging the datasets is necessary to overcome systematic biases in one or other AGB dataset.

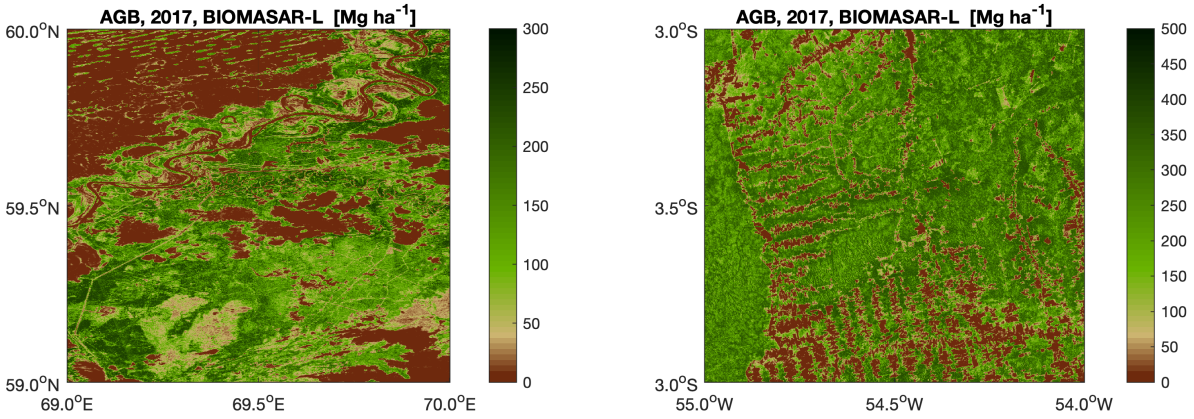


Figure 7-7: Example of AGB estimates obtained with BIOMASAR-L applied to the same 1° × 1° tile shown in Figure 7-1.

Figure 7-8 shows the AGB map from BIOMASAR-L for the Alpine region and the year 2018, which can be compared with the output from BIOMASAR-C in Figure 7-2. This extended AGB map has more contrast than the map obtained with BIOMASAR-C and confirms the indications reported for Figure 7-7.

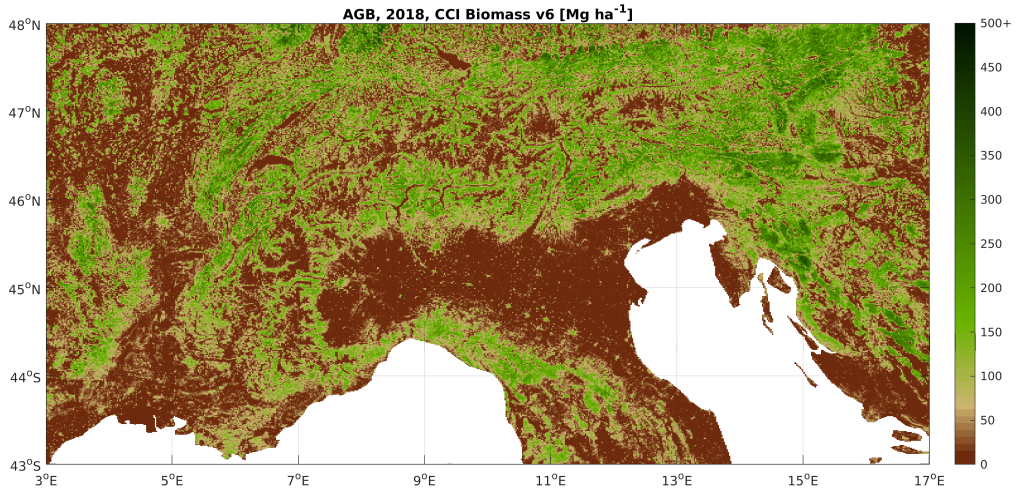


Figure 7-8: AGB map of the Alpine region obtained with BIOMASAR-L using ALOS-2 data acquired in 2018.

An example of a global map from BIOMASAR-L is displayed in Figure 7-9. The spatial distribution of AGB is similar to that obtained with BIOMASAR-C, but the contrast is stronger, which means that BIOMASAR-L can resolve the highest and lowest biomasses better than BIOMASAR-C. The map in Figure 7-9 shows that the spatial distribution of AGB estimated from ALOS-2 PALSAR-2 data is more homogeneous than from C-band thanks to the filtering during the pre-processing of the data. Nonetheless occasional striping occurs, corresponding to the overlap between adjacent SAR images in regions where the sensitivity of the backscatter to AGB is weak and even small calibration errors of few tenths of a dB lead to significant differences in the estimated AGB (Figure 7-10).

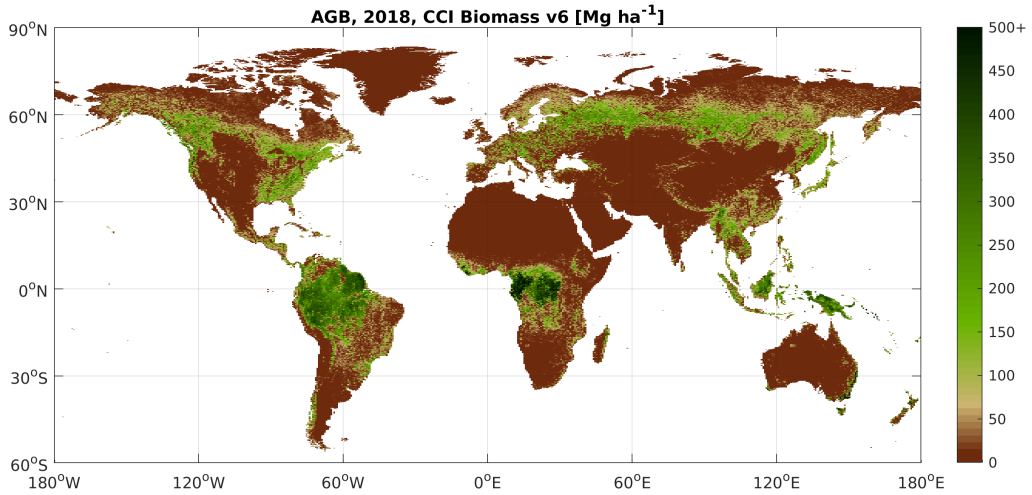


Figure 7-9: AGB map obtained with BIOMASAR-L using ALOS-2 data acquired in 2017.

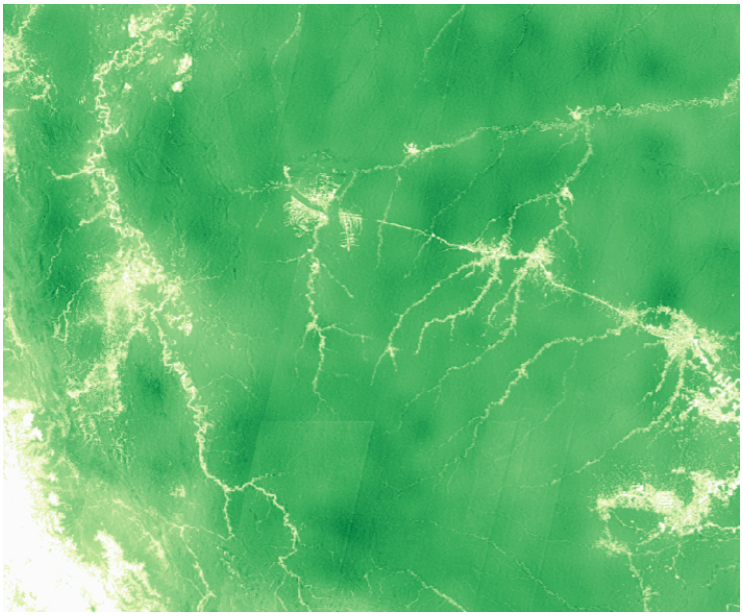


Figure 7-10: AGB estimates from ALOS-2 data for the southwestern Amazonian region.

The scatterplot comparing NFI-based averages of AGB at the level of administrative or ecological units with AGB averages from the BIOMASAR-L map of 2017 shows strong agreement and no apparent loss of sensitivity of the retrieved AGB. The red circles indicate that, on average, BIOMASAR-L tends to slightly overestimate AGB below 200 Mg ha^{-1} and underestimate in the high biomass range (> 300 Mg ha^{-1}). This tendency was also identified in BIOMASAR-C, although the L-band estimates appear to be more in line with the NFI-based averages, except for several outliers that did not occur in the BIOMASAR-C map. These results suggest merging of the two maps to mitigate systematic biases in each single map.

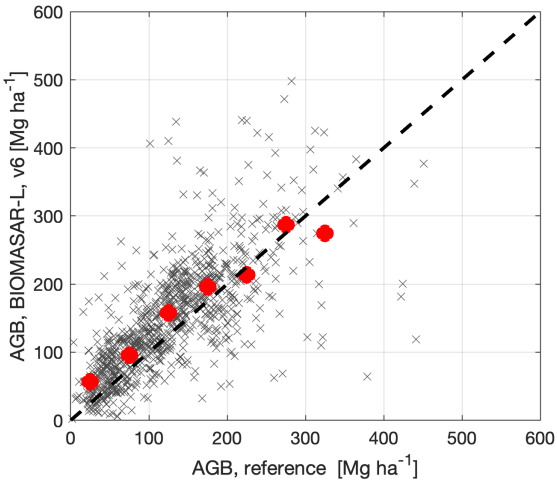


Figure 7-11: Scatterplot comparing averages of AGB from NFIs at administrative or ecological region with corresponding averages from the BIOMASAR-L AGB map of 2017. The red markers report the median AGB from the BIOMASAR-C AGB map for 50 Mg ha^{-1} wide bins of reference AGB.

7.1.4 Comparing BIOMASAR-L versions in CCI CORE algorithms

The BIOMASAR-L algorithm has undergone several modifications to produce v6 of the CCI Biomass maps, so the maps produced with v5 and v6 of the algorithm exhibit differences. Figure 7-12 illustrates these differences for the maps produced for 2020. They occur for three main reasons:

1. Changes in the height-to-biomass model (locally).
2. Changes in the maximum retrievable biomass (globally).
3. Changes in the estimation of the model parameters (globally).

The current AGB maps produced from ALOS-1 PALSAR-1 or ALOS-2 PALSAR-2 data tend to have a higher AGB across dense forest areas in the humid tropical, temperate, and boreal forest regions. AGB is instead lower than in previous versions across forests in the Pacific Northwest of the United States and British Columbia, Canada, as well as for forest regions in Central Chile dominated by pine and eucalyptus plantations. These differences are primarily associated with differences in the IceSAT-2 LiDAR-based map of maximum forest heights and, consequently, the maximum retrievable AGB. Also, across semi-arid forest regions in Southern Africa and South America, the new version presents a moderately higher AGB in low AGB ranges. These differences are associated with the modified approach for estimating the model parameter σ^0_{gr} in Equation (4-8) (Section 4.4.1). The new estimates for σ^0_{gr} tend to be lower and hence the AGB estimates higher in low AGB ranges.

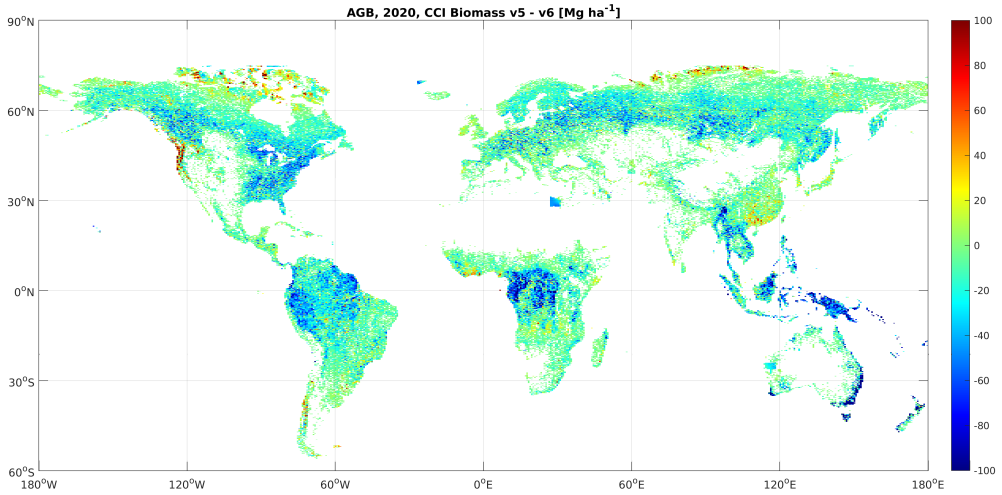


Figure 7-12: Differences between the AGB maps for the year 2020 produced with v5 and v6 of BIOMASAR-L, respectively.

Finally, we show an assessment of the BIOMASAR-L estimates from v5 with respect to NFI-based averages (Figure 7-13). Figure 7-13). The distribution of the AGB averages is almost unaltered compared to v6 in Figure 7-11.

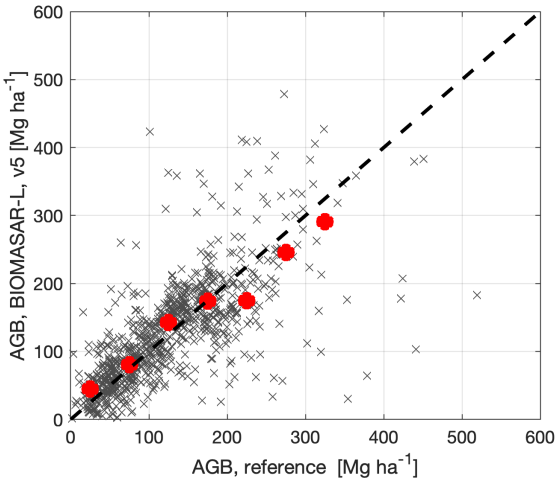


Figure 7-13: Same as Figure 7-11 but based on v5 of the BIOMASAR-L algorithm.

7.1.5 Merging AGB estimates

The merging approach, which had already been implemented for generating the GlobBiomass product, was also found to improve the estimates of biomass derived from the C- and L-band datasets in v1 to v5 of the CCI Biomass AGB maps. It was therefore applied also to produce v6 of the CRDP. Overall, the merging approach favours the L-band derived AGB over values from C-band for forests with low or high AGB. In forests with moderate AGB, instead, L-band derived estimates are given more weight. However, deviations from this simple rule (Equation (4-11)) were required to account for the varying number of annual backscatter observations available, i.e., in particular the variable number of Fine-Beam observations from ALOS-1 PALSAR-1 and ALOS-2 PALSAR-2 (Equation (4-12)). With the focus of the CCI Biomass project moving towards

quantification of AGB changes, the merging of maps from C- and L-band data acquired in multiple years since v5 of the CRDP also considers the interannual consistency of the resulting merged AGB maps (Equation (4-14)), i.e., the weights for combining C- and L-band derived biomass maps from a given year are adjusted so that the interannual variability between merged AGB maps for consecutive years is minimized. As can be seen in Figure 7-10, the individual C- and L-band based AGB maps do present local artifacts due to calibration errors as well as an inconsistent annual coverage with observations from different seasons locally. Because of these imperfections in the individual C- and L-band products, a harmonization across multiple years is performed to produce a temporal series of maps which allow for assessing annual and decadal changes in AGB. As illustrated in Figure 7-14, the harmonization leads to a better interannual consistency (i.e., higher correlation, lower RMSD) between the AGB maps (compared to v4 of the CRDP), a fundamental requirement for an analysis of biomass trends.

Figure 7-15 presents the difference in AGB between the v6 maps for the years 2010 and 2020 for Central Canada, the Northern United States and Central Brazil. The map clearly shows the areas of deforestation and forest fires (red). However, the AGB differences also indicate areas of increasing biomass (green to blue). These increases are not verified and at this stage suggest changes between the maps produced for 2007, 2010, and 2015-2022 need to be interpreted with care (see also Section 7.2).

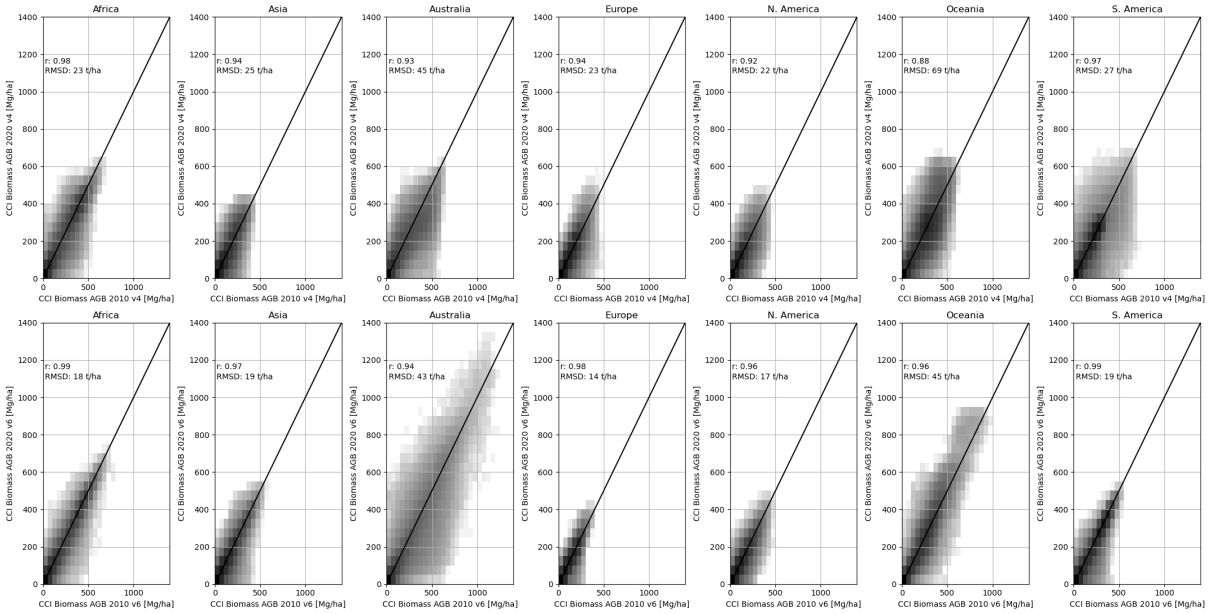


Figure 7-14: Comparison of merged AGB maps for 2010, 2017 and 2020 per continent and at kilometric scale for the CRDP v4 and v6, produced without and with interannual harmonization of weights for merging C- and L-band AGB maps, respectively.

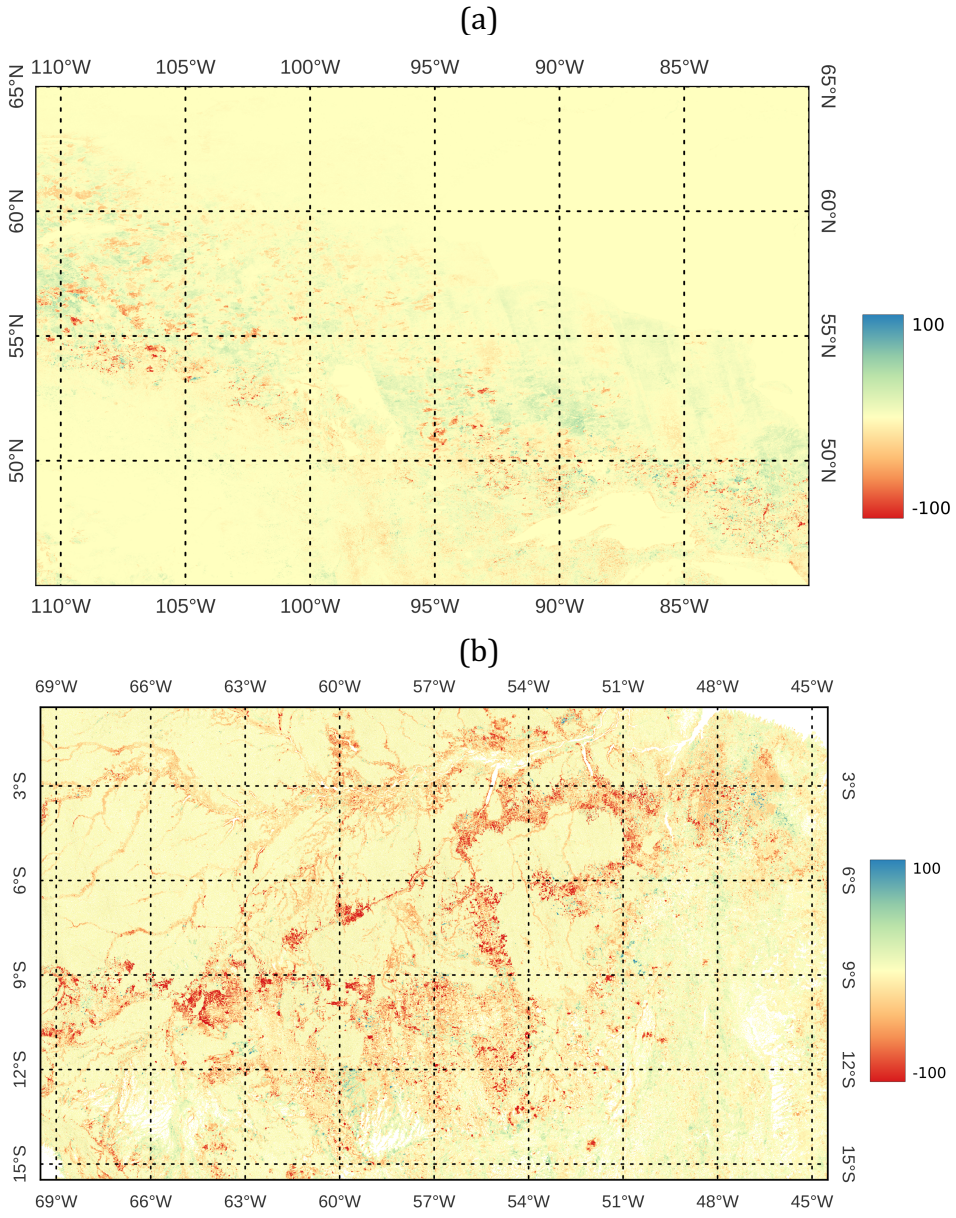


Figure 7-15: Difference between CCI Biomass v6 AGB maps for 2020 and 2010 in Central Canada/Northern USA (a) and Brazil (b). Forests subject to biomass loss show up in red, potential biomass increase in blue.

7.2 AGB change estimation

The CCI Biomass project targets estimation of AGB changes for annual intervals between 2015 and 2022 and for a decadal interval, i.e., between 2010 and 2020. Verification of the global AGB maps produced for v5 of the CRDP reveals a consistent spatial distribution of AGB but different levels, particularly in the dense tropics. This is due to the different EO datasets available in 2010 (multi-temporal moderate resolution Envisat ASAR observation and a few ALOS PALSAR observations) compared to years starting with 2015 (multi-temporal, high resolution observations from Sentinel-1 and ALOS-2 PALSAR-2).

Since global AGB change maps are not ready at the time of finalizing this document, the discussion is centred around problems that were evident in older releases of the CRDP and that may still affect the biomass change maps included in the CRDP following this report.

Figure 7-16 shows an example of the AGB difference between AGB maps of 2020 and 2010 from v5 of the CRDP and a latitudinal profile of the AGB difference. In this example, we compare maps averaged to 0.1° to identify major patterns of change and to judge the overall quality of a change product derived from the difference of two maps.

The latitudinal pattern indicates constant AGB or slight decrease. In the absence of a truly global reference dataset of AGB differences, these trends cannot be confirmed. Compared to v4, where the magnitude of the changes was larger and some of the detected changes were dubious, here we see a more constrained range of changes between 2010 and 2020. Northern latitudes are characterized by increase of AGB and local strong losses. Widespread losses in tropical regions were detected in the Amazon and the Congo Basin. However, we also detected some increases in both regions, as well as in China and in temperate South America. The result is an improvement of the AGB merging step where temporal consistency of the AGB estimates was introduced in v5.

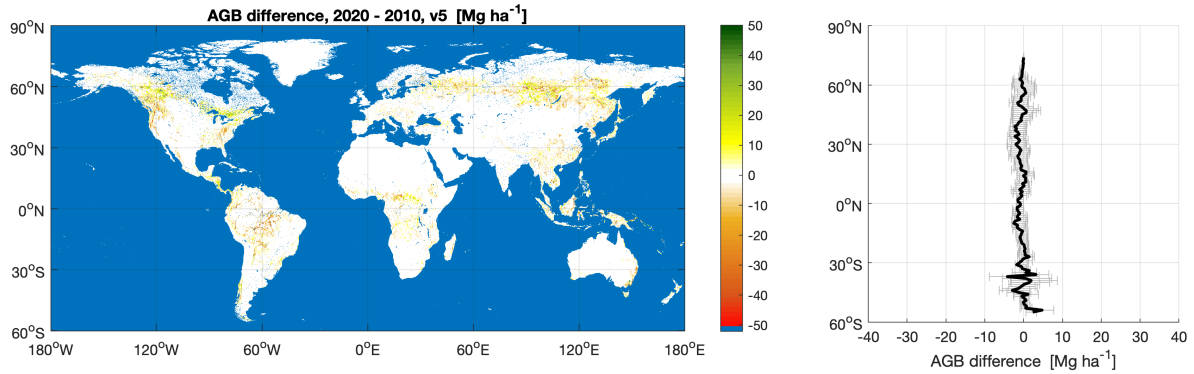


Figure 7-16: Difference between the 2020 and the 2010 AGB datasets (left) and latitudinal profile of the AGB difference (right). The AGB maps were averaged to 0.1° before taking the difference. The colour ramp is constrained between +/- 50 Mg ha^{-1} to enhance contrast. The latitudinal profile shows the average AGB difference as a function of latitude (thick line) and the interquartile range of AGB difference at a given latitude (horizontal bars). The AGB maps used in this example are part of the CRDP v5 of the CCI Biomass project.

The variability of the AGB changes is, however, much stronger when considering the full resolution of the maps. An example is shown by Figure 7-17 with the AGB difference map obtained at full resolution from the 2020 and 2010 maps of v5. The 10° x 10° region in Figure 7-17 covers tropical rainforests along the Amazon River, with significant deforestation occurring south of the river (fishbone pattern). The AGB difference map shows both positive and negative values, some of them substantially larger than would be expected. Further evidence for the low reliability of these AGB change estimates is the strong variability of the AGB difference between adjacent pixels. We interpret this to mean that the AGB difference is dominated by errors and uncertainties in the individual maps. In summary, the interpretation of CCI Biomass change maps at full resolution must be very cautious. There is, however, one exception, which concerns areas affected by deforestation. Figure 7-17 shows several red-coloured spots at the edges of previous deforestation. The strong loss of biomass between 2010 and 2020 can be explained with the expansion of deforestation into intact rainforests.

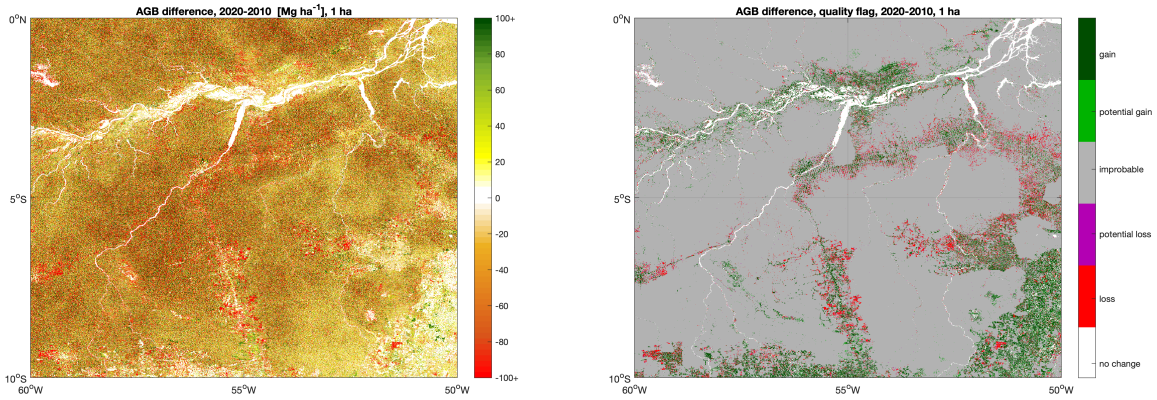


Figure 7-17: Difference between the AGB estimates obtained for 2020 and 2010 (left) and index of reliability of the AGB difference estimate (right).

To quantify the reliability of an estimate of AGB difference, we followed the approach developed in the GlobBiomass project for mapping AGB changes at regional scale. Each of the two estimates contributing to an AGB difference may have a substantial standard deviation, typically 40%-60% of the estimated value [RD-5]. The standard error of the AGB difference is therefore even larger because the variances of the individual estimates are summed. As a result, even a large AGB difference obtained from the maps may not be significant. In the GlobBiomass project, two scenarios were depicted. Let us assume that AGB_1 is the estimate at the first point in time and AGB_2 is the estimate for the same pixel at the second point in time. Let us then assume that SD_1 and SD_2 are the standard deviations of the two estimates, respectively. The probability that the AGB difference corresponds to a high or low probability of true change depends on whether the intervals $AGB_i \pm 1\ SD_i$ are disjoint (Figure 7-18, upper graphic) or overlap (Figure 7-18, lower graphic). Because of the rather large intervals of AGB, we also introduce a moderate level of reliability referred to as potential change. Potential change occurs when the intervals overlap but the estimate at one point in time is outside the interval $AGB_j \pm 1\ SD_j$ for the second interval in time (Figure 7-19).

Though this approach oversimplifies the issue, it is a valuable way to generate a simple auxiliary layer that can inform map users about the reliability of the difference estimate.

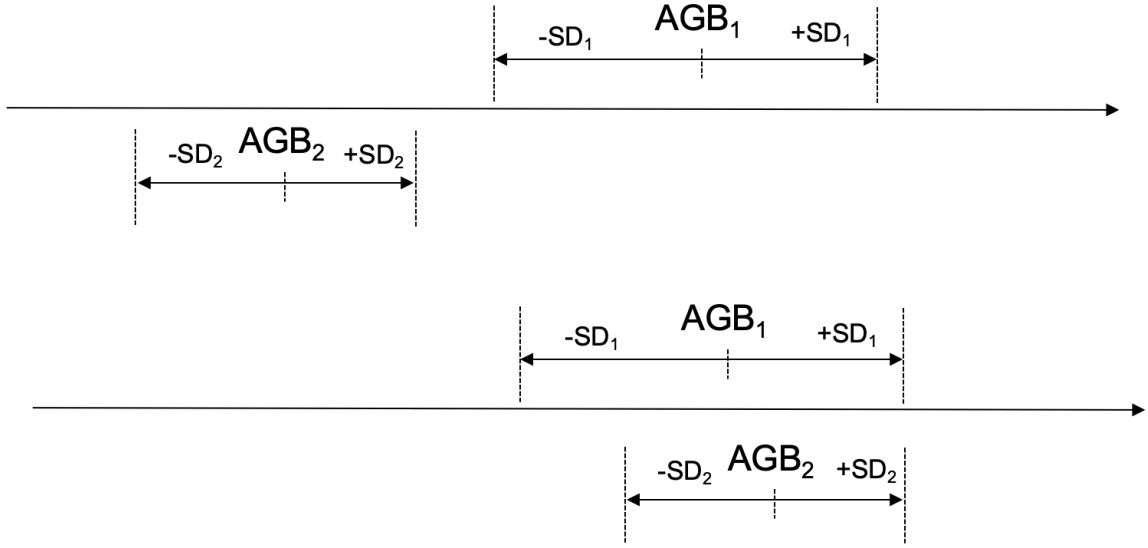


Figure 7-18: Upper graphic: disjoint intervals of AGB estimates at points in time 1 and 2 indicating high reliability of an AGB change defined as AGB difference. Lower graphic: overlapping intervals of AGB estimates at points in time 1 and 2 indicating low reliability of AGB change defined as AGB difference.

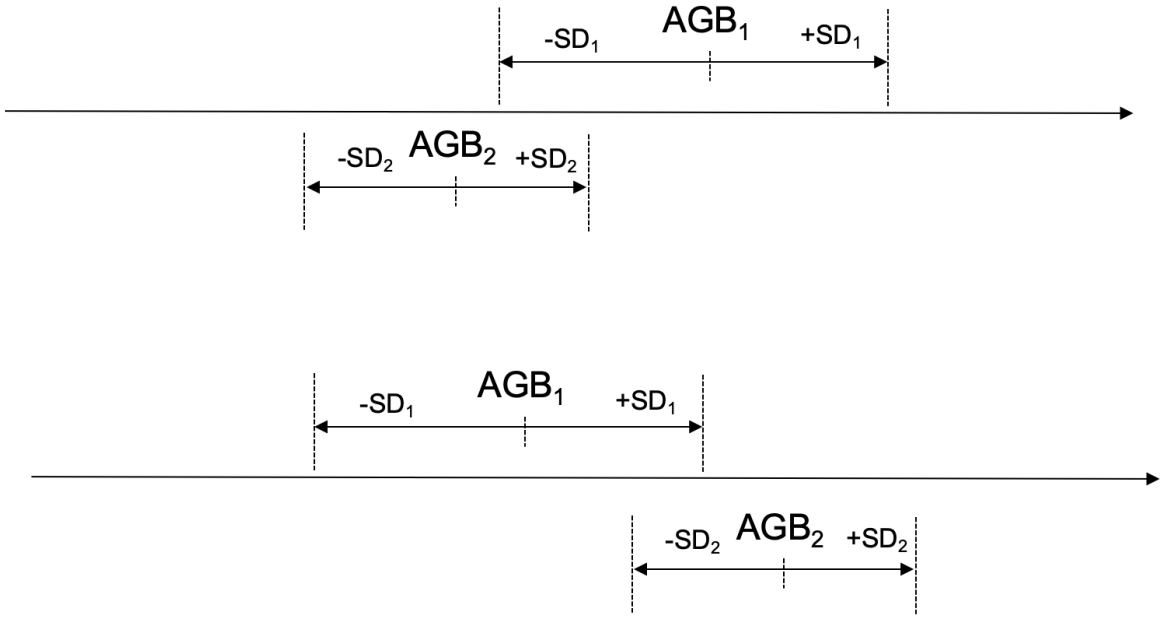


Figure 7-19: Partial overlap of intervals AGBi+/-SDi corresponding to a definition of potential AGB loss ($AGB_2 < AGB_1-SD_1$, upper graphic) or potential AGB gain (lower graphic, $AGB_2 > AGB_1+SD_1$).

Figure 7-20 shows an example of a reliability map corresponding to the AGB difference between 2020 and 2010. Low reliability corresponds to having overlapping AGB distributions and characterizes most of the area, even if the AGB change is non-negligible. Very few areas are characterized by an intermediate reliability, corresponding to a definition of potential AGB change. Areas with a high reliability, i.e., with disjoint AGB distributions, occur seldom but always

correspond to areas where AGB dropped from the level of a mature forest to a level close to 0 Mg ha^{-1}.

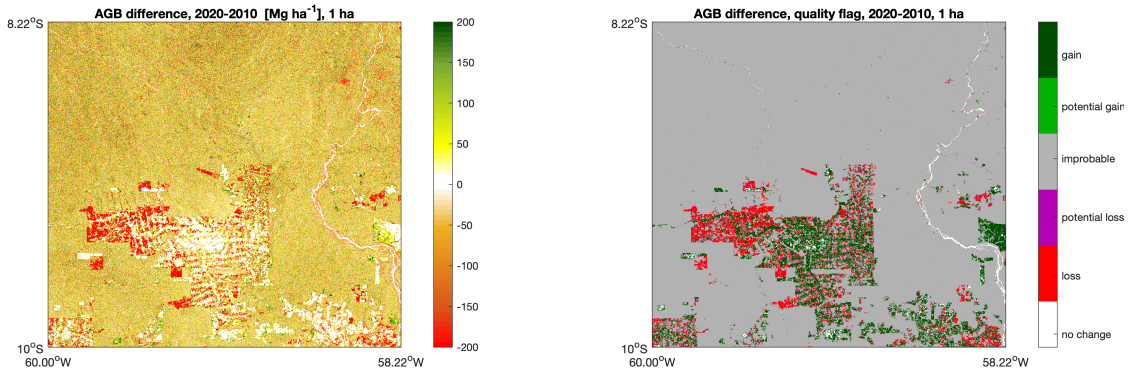


Figure 7-20: Zoom of Figure 7-17 in an area characterized by expanding deforestation into intact forests.

We conclude that an AGB difference map should be handled with care and a layer indicating its reliability should accompany the AGB differences to decide whether such differences are meaningful. In this context, we do not consider the effect of spatial averaging to reduce the variance of the error in an AGB estimate. Nonetheless, the concept developed here for the maps at full resolution applies to maps obtained after spatial averaging to coarser resolution.

It should finally be noted that, regardless of the spatial scale at which the AGB difference is obtained, this approach does not account for biases affecting each AGB map. Should the two maps be affected by different biases, the AGB change obtained by differencing AGBs at the two points in time will have an additional bias term (difference of biases) which will affect the accuracy of the estimate of AGB change.

8 References

Araza, A., de Bruin, S., Herold, M., Quegan, S., Labriere, N., Rodriguez-Veiga, P., Avitabile, V., Santoro, M., Mitchard, E.T.A., Ryan, C.M., Phillips, O.L., Willcock, S., Verbeeck, H., Carreiras, J., Hein, L., Schelhaas, M.-J., Pacheco-Pascagaza, A.M., da Conceição Bispo, P., Laurin, G.V., Vieilledent, G., Slik, F., Wijaya, A., Lewis, S.L., Morel, A., Liang, J., Sukhdeo, H., Schepaschenko, D., Cavlovic, J., Gilani, H., Lucas, R., (2022). A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps. Remote Sensing of Environment 272, 112917. https://doi.org/10.1016/j.rse.2022.112917

Askne, J., Dammert, P.B.G., Ulander, L.M.H., Smith, G. (1997). C-band repeat-pass interferometric SAR observations of the forest. IEEE Transactions on Geoscience and Remote Sensing 35, 25–35. https://doi.org/10.1109/36.551931

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.

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

Attema, E.P.W., Ulaby, F.T. (1978). Vegetation modeled as a water cloud. Radio Science 13, 357–364.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., 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.

Balzter, H., Cole, B., Thiel, C. andSchmullius, C. (2015). Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. Remote Sensing 7, 14876–14898. https://doi.org/10.3390/rs71114876

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.

Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N.E., Herold, M. and Fritz, S. (2020). Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe. Zenodo. https://doi.org/10.5281/senodo.3939049

Cartus, O., Santoro, M. and Kellndorfer, 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

Cartus, O., Santoro, M., Wegmüller, U., &Rommen, B. (2019a). Benchmarking the retrieval of biomass in boreal forest using P-band SAR backscatter with multi-temporal C- and L-band observations. *Remote Sensing*, 11,1695.

Cartus, O., & Santoro, M. (2019b). Exploring combinations of multi-temporal and multi-frequency radar backscatter observations to estimate above-ground biomass of tropical forest. *Remote Sensing of Environment*, 232, 111313.

Cartus, O., Siqueira, P., Kellndorfer, J., 2018. An Error Model for Mapping Forest Cover and Forest Cover Change Using L-Band SAR. IEEE Geosci. Remote Sensing Lett. 15, 107–111. https://doi.org/10.1109/LGRS.2017.2775659

Cartus, O., Santoro, M., Jimenez, C., Prigent, C., Schwank, M., Araza, A., Wegmüller, U. (submitted). A Parametric Approach for Global Estimation of Forest Aboveground Biomass with SMOS and SMAP L-band Radiometer Data. Remote Sensing of Environment,

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.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and scaling for tropical forest biomass estimates. Philosophical Transactions of the Royal Society B: Biological Sciences 359, 409–420.

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. andYamakura, 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. andQie, 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.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Science of Remote Sensing 100002. https://doi.org/10.1016/j.srs.2020.100002

Fahrland, E., 2022. Copernicus DEM Product Handbook (No. Date: 3 June 2022, Version 4.0).

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. andPeñuelas, J. (2019). Satellite-observed pantropical carbon dynamics. Nat. Plants 5, 944–951.

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.

Frison, P.-L., Fruneau, B., Kmiha, S., Soudani, K., Dufrêne, E., Toan, T.L., Koleck, T., Villard, L., Mougin, E. andRudant, J.-P. (2018). Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. Remote Sensing 10, 2049. https://doi.org/10.3390/rs10122049

Fritz, S., See, L., Perger, C., McCallum, I., Schill, C., Schepaschenko, D., Duerauer, M., Karner, M., Dresel, C., Laso-Bayas, J.-C., Lesiv, M., Moorthy, I., Salk, C.F., Danylo, O., Sturn, T., Albrecht, F., You, L., Kraxner, F., Obersteiner, M., 2017. A global dataset of crowdsourced land cover and land use reference data. Sci Data 4, 170075. https://doi.org/10.1038/sdata.2017.75

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C., Sohlberg, R.A., 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Earth Interactions 7, 1–15. https://doi.org/10.1175/1087-3562(2003)007<0001:GPTCAA>2.0.CO;2

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.

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.

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.

Kay, H., Santoro, M., Cartus, O., Bunting, P., Lucas, R., (2021). Exploring the Relationship between Forest Canopy Height and Canopy Density from Spaceborne LiDAR Observations. Remote Sensing 13, 4961. https://doi.org/10.3390/rs13244961.

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

Kindermann, G.E., McCallum, I., Fritz, S., Obersteiner, M., 2008. A global forest growing stock, biomass and carbon map based on FAO statistics. Silva Fennica 42, 387–396. https://doi.org/10.14214/sf.244

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.

Labrière, N., and Chave, J., CCI Biomass project – In situ datasets, Processing and validation of CMS Kalimantan data, internal report, April 2020.

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.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., et al. The Ice, cloud, and land elevation satellite-2 (ICESat-2): Science requirements, concept, and implementation. Remote Sens. Environ. 2017, 190, 260–273.

Meyer, H., Reudenbach, C., Wo llauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications: Moving from data reproduction to spatial prediction. Ecological Modelling, 411:108815.

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.

Neuenschwander, A., Pitts, K., 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. Remote Sensing of Environment 221, 247–259.

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.

Praks, J., Antropov, O., Hallikainen, M., 2012. LIDAR-aided SAR interferometry studies in boreal forest: Scattering phase center and extinction coefficient at X- and L-band. Remote Sensing 50, 3831–3843.

Pulliainen, J. T., Heiska, K., Hyyppä, J. and Hallikainen, M. T. (1994). Backscattering properties of boreal forests at the C- and X-bands. IEEE Transactions on Geoscience and Remote Sensing, 32 (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.

Réjou-Méchain, M., Muller-Landau, H.C., Detto, M., Thomas, S.C., Le Toan, T., Saatchi, S.S., Barreto-Silva, J.S., Bourg, N.A., Bunyavejchewin, S., Butt, N., Brockelman, W.Y., Cao, M., Cárdenas, D., Chiang, J.-M., Chuyong, G.B., Clay, K., Condit, R., Dattaraja, H.S., Davies, S.J., Duque, A., Esufali, S., Ewango, C., Fernando, R.H.S., Fletcher, C.D., Gunatilleke, I.A.U.N., Hao, Z., Harms, K.E., Hart, T.B., Hérault, B., Howe, R.W., Hubbell, S.P., Johnson, D.J., Kenfack, D., Larson, A.J., Lin, L., Lin, Y., Lutz, J.A., Makana, J.-R., Malhi, Y., Marthews, T.R., McEwan, R.W., McMahon, S.M., McShea, W.J., Muscarella, R., Nathalang, A., Noor, N.S.M., Nytch, C.J., Oliveira, A.A., Phillips, R.P., Pongpattananurak, N., Punchi-Manage, R., Salim, R., Schurman, J., Sukumar, R., Suresh, H.S., Suwanvecho, U., Thomas, D.W., Thompson, J., Uríarte, M., Valencia, R., Vicentini, A., Wolf, A.T., Yap, S., Yuan, Z., Zartman, C.E., Zimmerman, J.K., Chave, J., 2014. Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. Biogeosciences 11, 6827–6840. https://doi.org/10.5194/bg-11-6827-2014

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.

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.

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., 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., Wegmuller, U., Kellndorfer, 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., Wegmüller, U. and Askne, J. (2018). 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., Rozendaal, D., Avitabilie, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, Á., Cavlovic, J., CazzollaGatti, R., da ConceiçãoBispo, P., Dewnath, N., Labrière, N., Liang, J., Lindsell, J., Mitchard, E.T.A., Morel, A., Pacheco Pascagaza, A.M., Ryan, C.M., Slik, F., Vaglio Laurin, G., Verbeeck, H., Wijaya, A. and Willcock, S. (2021). The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth System Science Data 13, 3927–3950. https://doi.org/10.5194/essd-2020-148

Santoro, M., Cartus, O., Fransson, J. E. S., &Wegmüller, U. (2019). Complementarity of X-, C- and L-band SAR backscatter observations to retrieve forest stem volume. *Remote Sensing*, 11, 1563.

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.

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.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. andSchmullius, 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.

Van De Kerchov, R., Zanaga, D., Iordache, M.-D., Wevers, J., Kirches, G., Cartus, O., Carter , S., Tsendbazar, N., Li, L., Georgieva, I. (2021). ESA WorldCover - Technical Specification. Report WorldCover_TS_v1.2, ESA_ESRIN.

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.

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

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment 144, 152–171.

9 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° × 1° large tile characterized by high biomass forest, pasture, agriculture, and mountainous terrain with steep slopes between Austria and Slovenia. This is representative for the behaviour 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 behaviour 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.

9.1 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° × 1° 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 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.

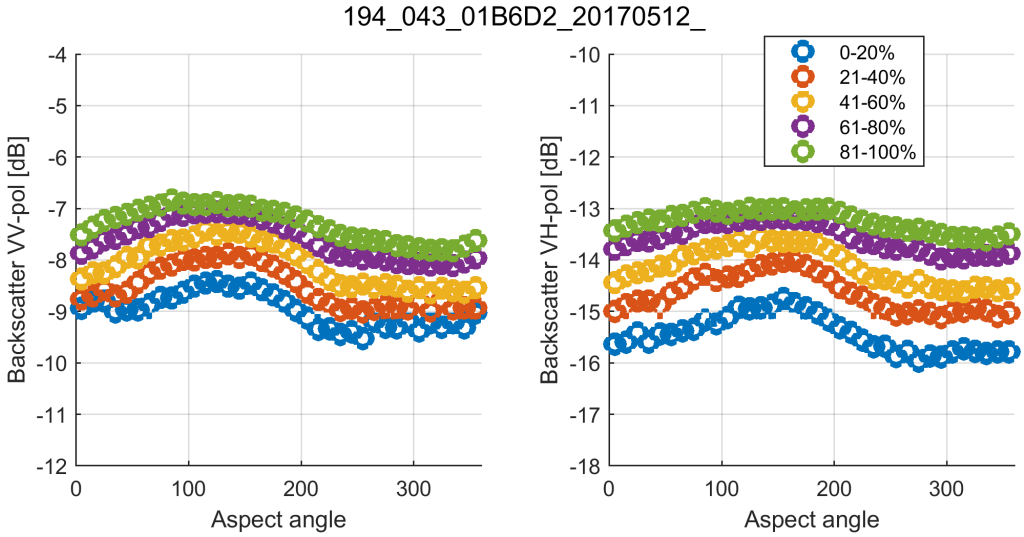


Figure A 1: Observations of Sentinel-1 backscatter as a function of terrain aspect angle for the 1° x 1° tile with top left coordinate: 14°E, 47°N. The tile includes temperate forests and alpine terrain between Austria and Slovenia. Aspect angle is defined as being zero when the normal to the surface 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).

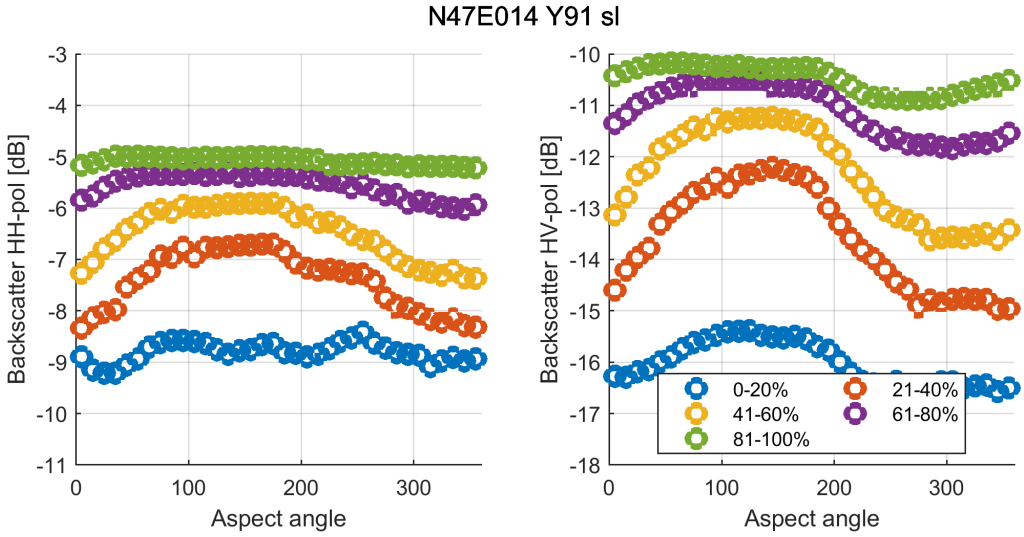


Figure A 2: Observations of ALOS PALSAR backscatter as a function of terrain aspect angle for the same 1° x 1° tile as in Figure A1. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and aspect angle interval (10° interval).

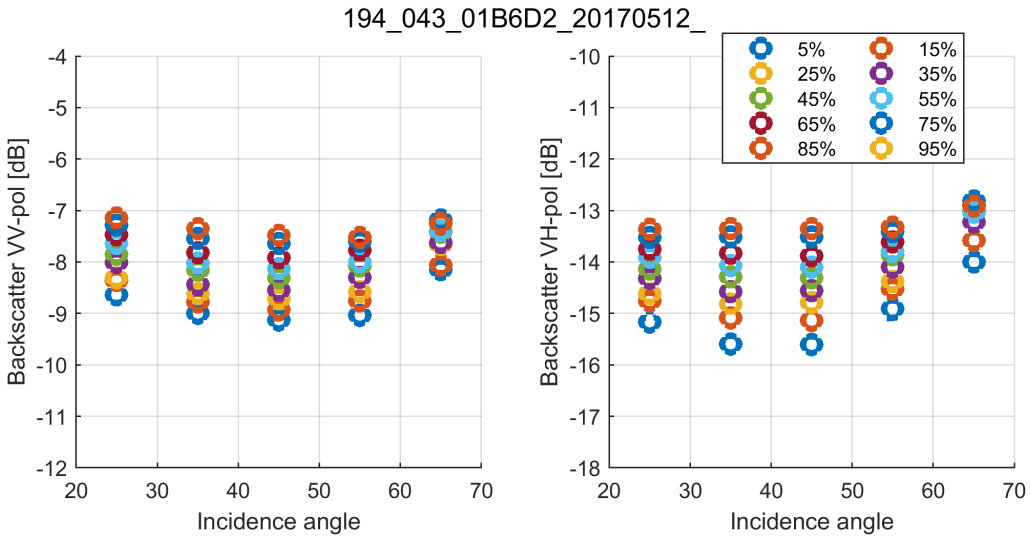


Figure A 3: Observations of Sentinel-1 backscatter as a function of local incidence angle for the same dataset and tile in Figure A1. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

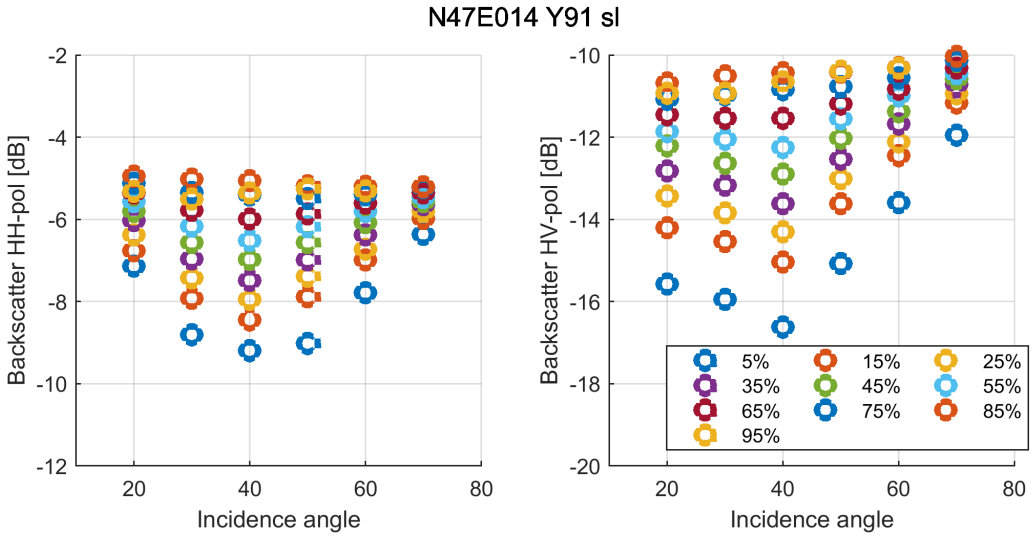


Figure A 4: Observations of ALOS PALSAR backscatter as a function of local incidence angle for the same dataset and tile in Figure A2. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

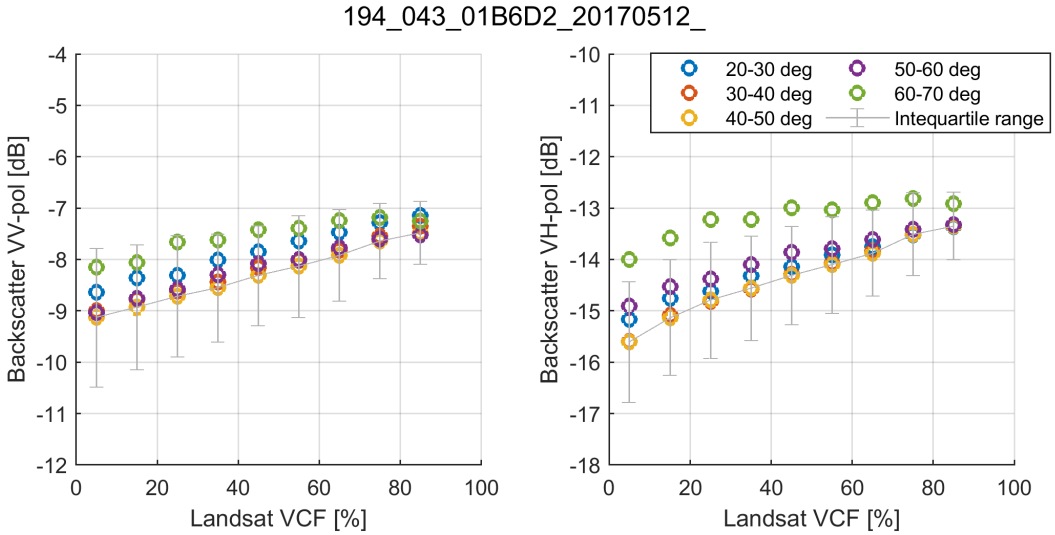


Figure A 5: Observations of Sentinel-1 backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A1. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval). The vertical bars represent the interquartile range of backscatter measurements in each VCF range and incidence angle interval of flat terrain.

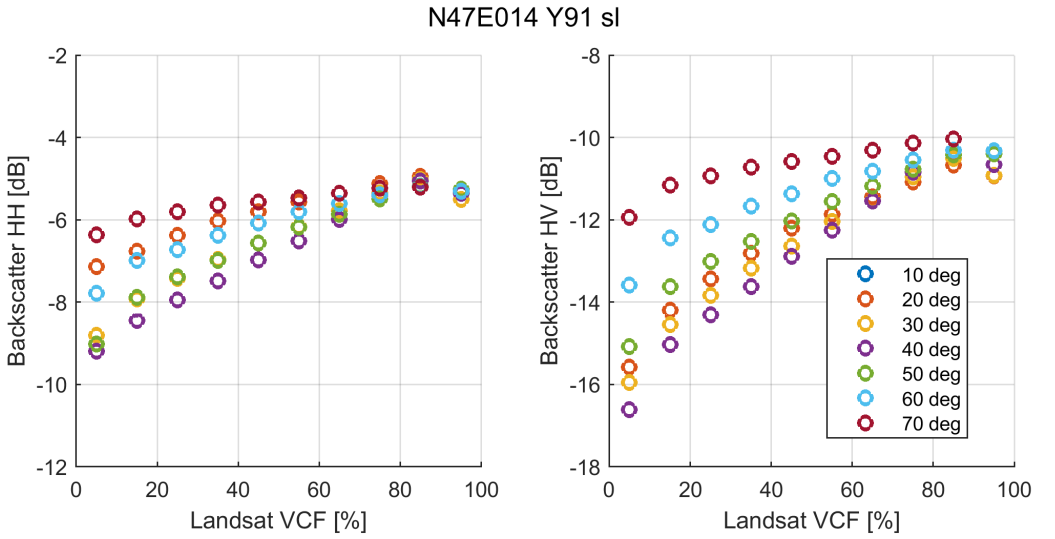


Figure A 6: Observations of ALOS PALSAR backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A2. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval).

9.2 Flat to hilly terrain, tropical forest

The analysis was repeated for a more moderate topography and a different forest type. The results are in line with previous observations, 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. For canopy density ranges where this is not as clear, the reason is the small number of pixels. Similar, 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° × 1° 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 backsscatter to canopy density was evident above 80%.

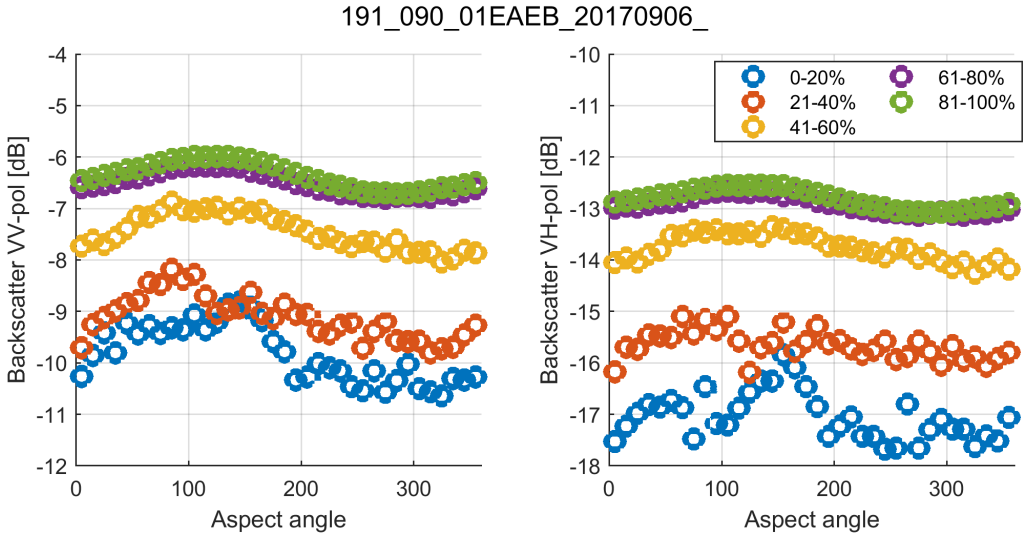


Figure A 7: Observations of Sentinel-1 backscatter as a function of terrain aspect angle for the 1° x 1° tile with top left coordinate: 11°E, 0°N. The tile includes the forest site of Lope, Gabon. Aspect angle is defined as being zero when the normal to the surface is oriented northwards. Observations 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).

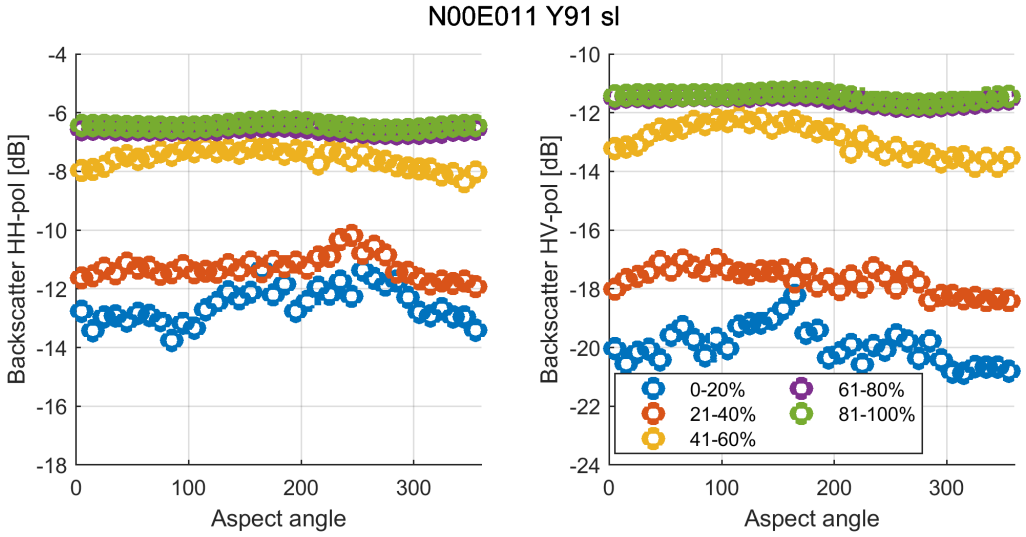


Figure A 8: Observations of ALOS PALSAR backscatter as a function of terrain aspect angle for the same 1° x 1° tile as in Figure A7. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and aspect angle interval (10° interval).

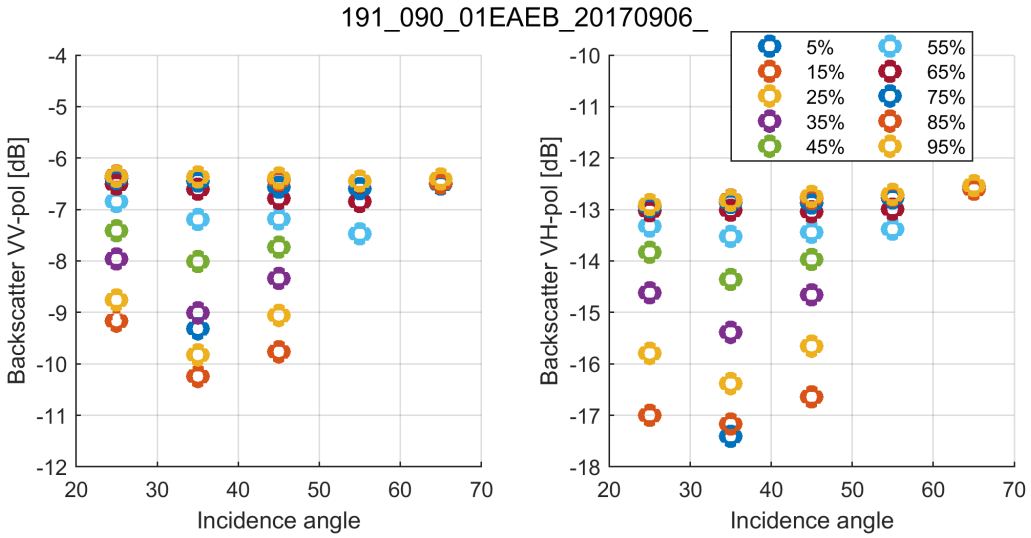


Figure A 9: Observations of Sentinel-1 backscatter as a function of local incidence angle for the same dataset and tile in Figure A7. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval (10° interval).

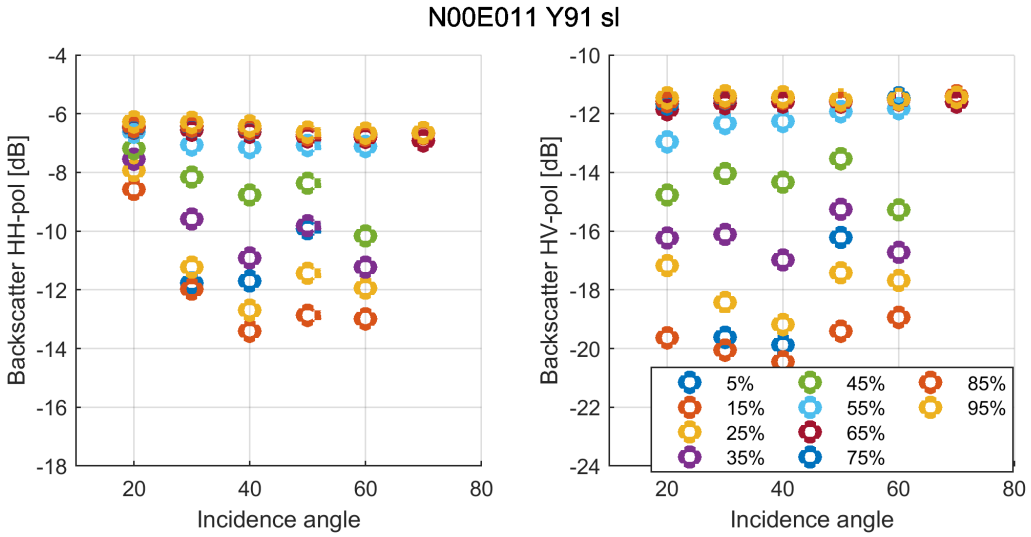


Figure A 10: Observations of ALOS PALSAR backscatter as a function of local incidence angle for the same dataset and tile in Figure A8. Observations are stratified per canopy density. Each circle represents the median backscatter for a given canopy density interval (see legend) and incidence angle interval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

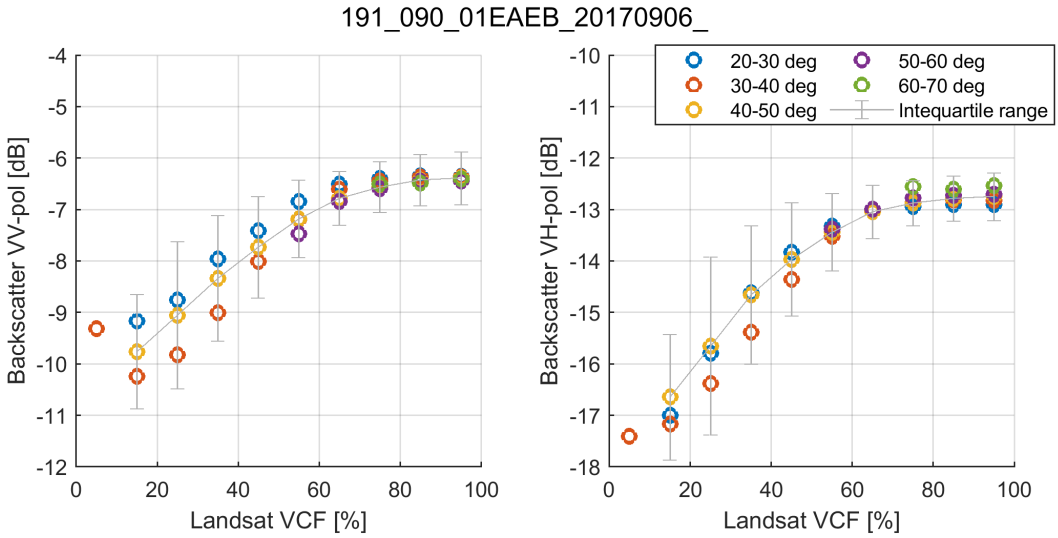


Figure A 11: Observations of Sentinel-1 backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A7. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval). The vertical bars represent the interquartile range of backscatter measurements in each VCF range and incidence angle interval of flat terrain.

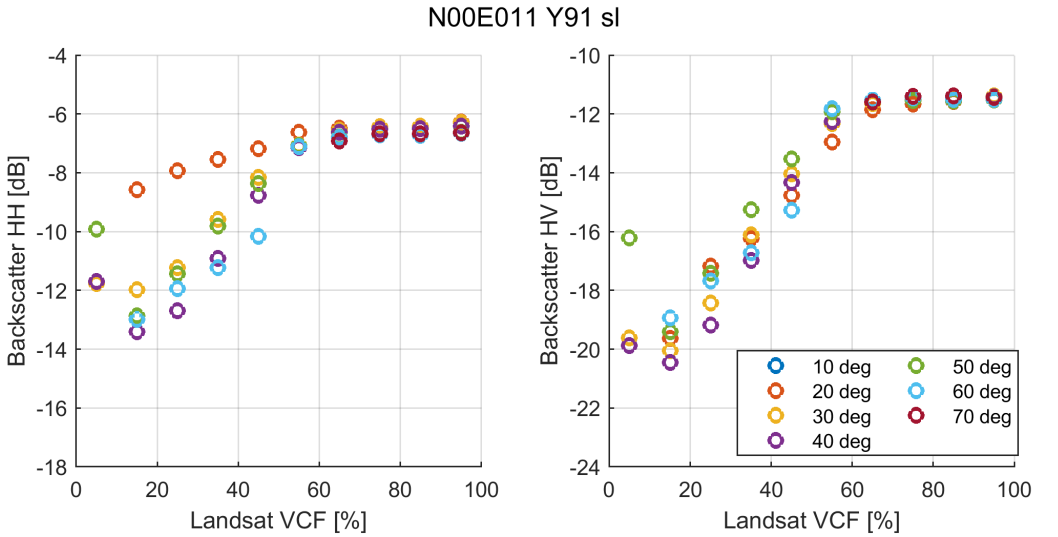


Figure A 12: Observations of ALOS PALSAR backscatter as a function of canopy density (Landsat VCF) for the same dataset and tile in Figure A8. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and canopy density interval (10° interval).

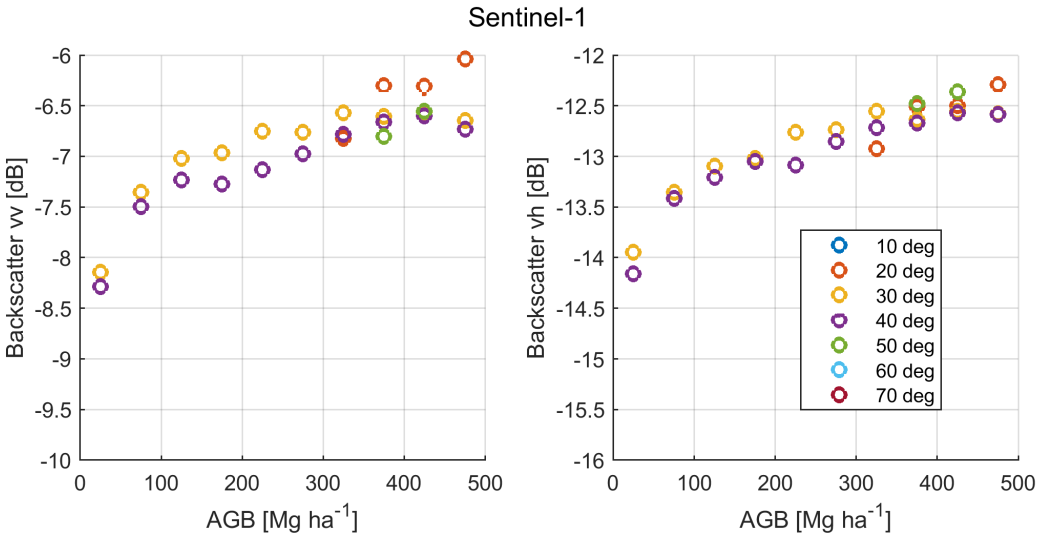


Figure A 13: Observations of Sentinel-1 backscatter as a function of AGB over Lope, Gabon, (Labriere et al., 2018) for the same dataset in Figure A7. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and AGB interval (50 Mg ha^{-1}).

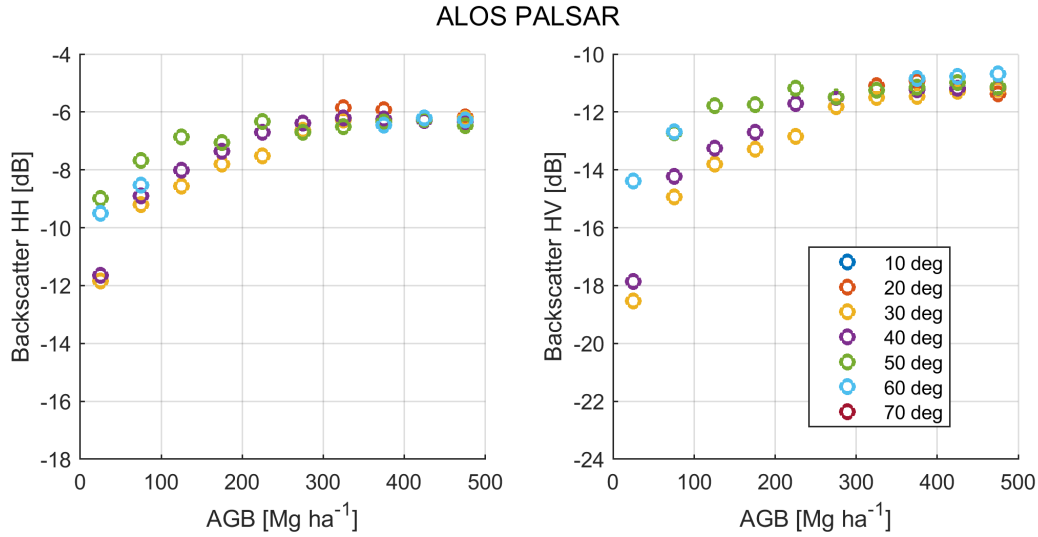


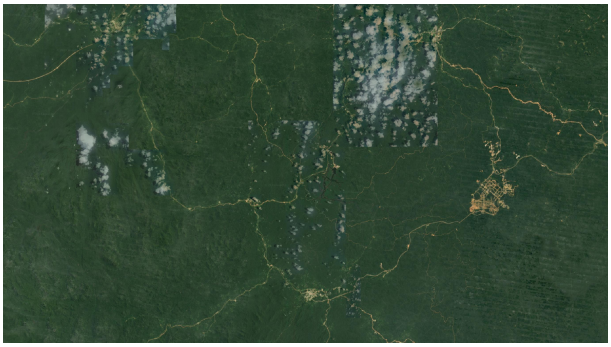
Figure A 14: Observations of ALOS PALSAR backscatter as a function of AGB over Lope, Gabon, (Labriere et al., 2018) for the same dataset in Figure A8. Observations are stratified per incidence angle. Each circle represents the median backscatter for a given incidence angle interval (see legend) and AGB interval (50 Mg ha^{-1}).

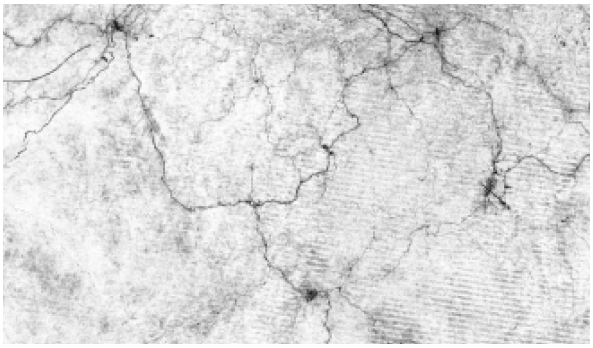
10Annex B

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 analysed; 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.

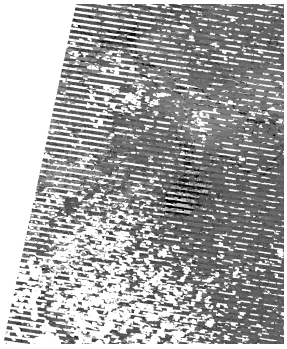
Figure B1 shows four percent tree cover datasets with spatial resolution between 30 m and 1,000 m of potential interest to support the model training. For reference, we included the optical image in Google Earth. This area (100 km × 60 km) is in Equatorial Guinea. The nearly horizontal, repeated segments 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 because of frequent cloud cover, which introduced several 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 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.

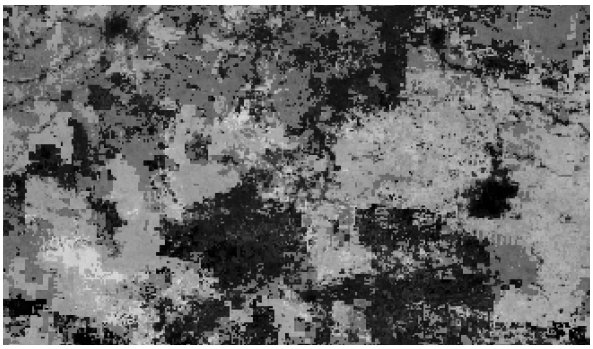




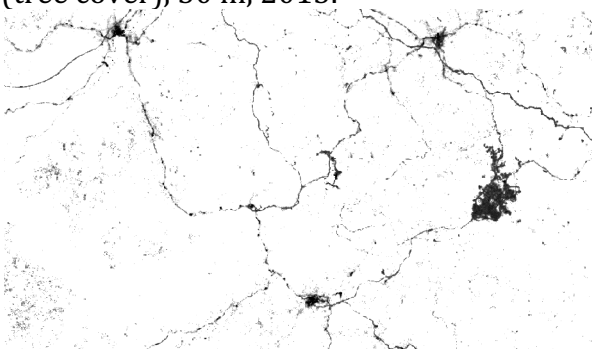
Landsat percent tree cover, 30 m, 2010.



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



MODIS Vegetation Continuous Fields v6, 250 m, 2017.



Proba-V tree cover fraction, 100 m, 2015.

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.

11 Annex C

From the GlobBiomass dataset, we derived averages at 0.25° and compared with the average AGB from plot inventory measurements described in the Product Validation Plan [RD-7]. The agreement was strong below 250 Mg ha^{-1} (Figure C1). Above this level, the map-based AGBs were underestimated, the discrepancy between map-based and plot-based AGB averages increasing with increasing AGB. This difference explains the negative bias and the rather high relative root mean squared difference between map-based and plot-based averages (Figure C1).

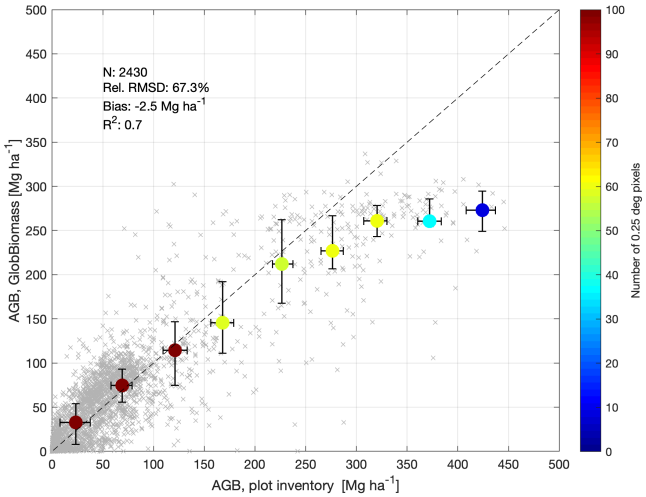


Figure C 1: Scatter plot comparing average AGB from plot inventory observations and map-based average AGB from the GlobBiomass dataset at 0.25°. Circles and bars represent the median and inter-quartile ranges of map-based averages for 50 Mg ha^{-1} wide bins of plot-based average AGB. Individual pairs of AGB values are represented by crosses. Retrieval statistics reported in this figure include the number of grid cells, i.e., pixels, the RMSD relative to the average plot-based AGB, the bias and the coefficient of determination R^2.

12 Annex D

Country	No. units	Type of unit	Reported biomass	Year	URL
Albania	1	Country	AGB	~ 2015	Avitabile et al., 2018
American Samoa	1	Country	AGB	2012	https://apps.fs.usda.gov/fiadb-api/evalidator
Argentina	6	Ecoregion	AGB	2017	https://www.argentina.gob.ar/ambiente/bosques/segundo-inventario-nacional-bosques-nativos
Australia	8	Territory	AGB	2016	https://www.agriculture.gov.au/abares/forestsaustralia/sofr/sofr-2018/
Bangladesh	5	Socioeconomic zone	AGB	2016-2019	https://forestecosyst.springeropen.com/track/pdf/10.1186/s40663-021-00284-1.pdf
Belarus	6	Province	GSV	2010	http://www.metla.fi/julkaisut/workingpapers/2010/mwp170.pdf
Belize	1	Country	AGB	2018	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Bhutan	20	Province	GSV	2012-2015	http://www.bhutantrustfund.bt/wp-content/uploads/2018/11/National-Forest-Inventory-Report-Vol.-I-DoFPS.pdf
Bosnia	1	Country	AGB	~ 2015	Avitabile et al., 2018
Brazil	6	Ecoregion	AGB	2018	Brazilian Forest at a glance 2019, Ministry of Agriculture, Livestock and Food Supply, Brazilian Forest Service
Bulgaria	1	Country	AGB	~ 2015	Avitabile et al., 2018
Burkina Faso	13	Province	AGC	2014	http://cns.bf/IMG/pdf/rapport_second_inventaire_forestier_national2.pdf
Cabo Verde	9	Island	AGB	2012	http://www.caboverdeifn.ifer.cz/?page_id=79
Cambodia	3	Ecoregion	AGB	2014	https://cambodia-redd.org/wp-content/uploads/2016/01/Forest-biomass-in-Cambodia-from-field-plots-to-national-estimates.pdf
Canada	12	Ecozone	AGB	2006-2017	https://nfi.nfis.org/en/standardreports
Chile	11	Region	GSV	2020	https://ifn.infor.cl/index.php/descargas-recursos/descargas/category/2-documentos-inventario-forestal
China	31	Province	AGB	2014-2018	China forest resources report, 9th forest inventory
Colombia	5	Biogeografic region	AGB	2015-2019	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Comores	3	Island	AGB	2010	FAO FRA country report 2020
Congo	5	Ecoregion	CO2 equivalent	2014	https://www.fao.org/3/cb2941fr/cb2941fr.pdf
Costarica	11	Life zone	CO2 equivalent	2014-2015	https://www.sirefor.go.cr/Sirefor/publicaciones_tabla?nombre=INF
Croatia	1	Country	AGB	~ 2015	Avitabile et al., 2018
Cyprus	1	Country	AGB	~ 2015	Avitabile et al., 2018
Democratic Republic of the Congo	26	Province	AGB	2011-2016	https://redd.gouv.cd/wp-content/uploads/2020/10/NERF-de-la-RDC.pdf (based on Xu et al., 2021)
Dominican Republic	4	Ecoregion	AGB	2018	https://fdocuments.ec/document/informe-final-inventario-nacional-forestal-de-repblica-ndice-elaboracin.html
Ecuador	9	Ecoregion	AGC	2009-2013	http://enf.ambiente.gob.ec/web_enf/?page_id=1239
El Salvador	4	Ecoregion	AGB	2018	https://cidoc.marn.gob.sv/documentos/inventario-nacional-de-bosques-de-el-salvador/
Estonia	1	Country	AGB	~ 2015	Avitabile et al., 2018
Ethiopia	4	Ecoregion	AGB	2010	https://redd.unfccc.int/files/ethiopia_frel_3.2_final_modified_submission.pdf
Fiji	1	Country	AGB	2006	http://fijireddplus.org/resources/publications/NFI2006DraftReport.pdf
Finland	19	NFI unit	AGB	2016-2020	https://statdb.luke.fi/PxWeb/pxweb/en/LUKE/LUKE__04%20Metsa__06%20Metsavarat/1.29_Puuston_biomassa_metsa_ja_kitumaalla.px/
French Guyana	1	Country	AGB	2015	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Guam	1	Country	AGB	2013	https://apps.fs.usda.gov/fiadb-api/evalidator
Guatemala	3	Forest type	AGB	2002-2003	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Honduras	4	Forest type	AGC	2020	https://icf.gob.hn/wp-content/uploads/2021/08/Anuario_Estadistico_Forestal_de_Honduras_2021.pdf
Iceland	1	Country	AGB	~ 2015	Avitabile et al., 2018
India	41	State	AGC	2016	https://fsi.nic.in/isfr-2021/chapter-9.pdf
Indonesia	7	Island	AGB	2014	http://dirjenppi.menlhk.go.id/kcpi/dokumen/national_frel_final%20revisi_10des.pdf
Ivory Coast	3	Ecoregion	AGB	2014	https://www.fao.org/3/i8019f/i8019f.pdf
Japan	47	Prefecture	GSV	2013-2017	https://www.rinya.maff.go.jp/j/keikaku/genkyou/h29/attach/pdf/3-13.pdf
Kosovo	1	Country	AGB	~ 2015	Avitabile et al., 2018
Laos	4	Forest type	AGB	2019	https://nfms.maf.gov.la
Latvia	1	Country	AGB	~ 2015	Avitabile et al., 2018
Liberia	15	Province	Total biomass	2018	https://www.forestcarbonpartnership.org/system/files/documents/Liberia%20National%20Forest%20Inventory.pdf
Liechtenstein	1	Country	AGB	~ 2015	Avitabile et al., 2018
Luxembourg	1	Country	AGB	~ 2015	Avitabile et al., 2018
Madagascar	4	Ecoregion	AGB	2017	https://redd.unfccc.int/files/2017_frel_mdg_modified_submission.pdf
Malawi	2	Region	Total biomass	2018	https://cepa.rmportal.net/Library/inbox/national-forest-inventory-2018-report
Marshall Islands	1	Country	AGB	2018	https://apps.fs.usda.gov/fiadb-api/evalidator
Mexico	32	State	AGB	2005-2009	de Jong et al., 2010
Micronesia	1	Country	AGB	2016	https://www.fs.usda.gov/pnw/projects/pnw-fia-pacific-islands-inventory
Mongolia	4	Ecoregion	AGB	2017	http://forest-atlas.gov.mn/DataSetResults.aspx
Montenegro	1	Country	AGB	~ 2015	Avitabile et al., 2018
Mozambique	10	Province	AGB	2018	https://www.biofund.org.mz/wp-content/uploads/2019/01/1548412245-Relatório%20do%20%20IV%20Inventário%20Florestal%20Nacional.pdf
Nepal	3	Physiographic region	AGB	2017	https://nepalindata.com/resource/STATE-OF-NEPAL%275-FORESTS/
Nicaragua	4	Forest type	AGB	2007-2008	https://cambioclimatico.ineter.gob.ni/bibliografia/Mitigacion%20del%20cambio%20climatico/Informe%20Final%20Inventario%20forestal.pdf
Nigeria	6	Ecozone	AGB	2019	https://www.fao.org/3/cb0037en/cb0037en.pdf
North Macedonia	1	Country	AGB	~ 2015	Avitabile et al., 2018
Northern Mariana Islands	1	Country	AGB	2015	https://www.fs.usda.gov/pnw/projects/pnw-fia-pacific-islands-inventory
Pakistan	12	Forest type	AGC	2008-2012	https://redd.unfccc.int/files/1_unfccc_frel_pakistan__final_with_proofread_-final.pdf
Palau	1	Country	AGB	2014	https://apps.fs.usda.gov/fiadb-api/evalidator
Panama	1	Country	AGB	2013-2015	https://chm.cbd.int/api/v2013/documents/05B386D2-5BCD-A52D-6097-F853803CC619/attachments/205145/inventario%20Nacional%20Forestal%20-%20Resultados%20Fase%20Piloto%202013-2015.pdf
Papua New Guinea	1	Country	AGB	> 2010	Proceedings_of_the_second_NFI_Research_Conference_compressed.pdf
Paraguay	6	Ecoregion	AGC	2014	http://www.infona.gov.py/index.php?cID=296
Peru	6	Ecozone	AGB	2013-2018	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Puerto Rica	1	Country	AGB	2019	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Republic of Korea	10	Province	GSV	2014	https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjRkbvRtZn6AhWYh_0HHeGSC7kQFnoECAUQAQ&url=https%3A%2F%2Fwww.forest.go.kr%2Fkfsweb%2Fcmm%2FFms%2FFileDown.do%3Bjsessionid%3DNLnLiHubhFNUOX5DhDrmaHEZurg8MExsUujmYxegMKFeFV2Ug2pXlg5gj
Russia	83	Province	GSV	2011-2020	Private data sharing, D. Schepaschenko (IIASA)
Sudan	3	State	GSV	2017	https://redd.unfccc.int/files/sudan_frl_submission_to_unfccc_january_2020.pdf
Suriname	2	Ecozone	AGC	2017	https://sbbsur.com/wp-content/uploads/2017/04/TechnischrapportEmissieFactors_CarbonStocks.pdf
Taiwan	8	Province	GSV	2021	https://www.forest.gov.tw/EN/0001465
Tanzania	25	Province	GSV	2015	https://www.tfs.go.tz/uploads/NAFORMA_REPORT.pdf
Thailand	3	Forest type	AGB	2013-2018	https://redd.unfccc.int/files/thailand_frel_frl_report.pdf
Togo	4	Ecoregion	AGB	2015-2016	https://redd.unfccc.int/files/nrf_togo_06_1_20_rev_18_08_20_finale.pdf
Turkey	1	Country	AGB	~ 2015	Avitabile et al., 2018
Uganda	80	Province (district)	AGB	2005	https://www.nfa.go.ug/images/reports/biomasstechnicalreport2009.pdf
Ukraine	25	Province	GSV	2010	Private data sharing, D. Schepaschenko (IIASA)
United Kingdom	1	Country	AGB	~ 2015	Avitabile et al., 2018
United States	50	State	AGB	2010-2021	https://apps.fs.usda.gov/fiadb-api/evalidator
United States Virgin Islands	1	Country	AGB	2014	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Uruguay	1	Country	AGB	2009-2016	Inventarios Forestales Nacionales de America Latina y el Caribe: Hacia la armonización de la información forestal
Vietnam	5	Forest type	Total carbon	2016	https://redd.unfccc.int/files/2016_submission_frel_viet_nam.pdf
Zambia	1	Country	AGB	2009-2016	https://redd.unfccc.int/files/zambia_frel-2020-technical_assessment.pdf