

CCI
BIOMASS

ALGORITHM DEVELOPMENT PLAN
YEAR 1
VERSION 1.0

DOCUMENT REF:	CCI_BIOMASS_ADP_V1
DELIVERABLE REF:	D2.4-ADP
VERSION:	1.0
CREATION DATE:	2019-02-28
LAST MODIFIED	2019-02-28

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	2	28.02.2019	

Document Authorship

	NAME	FUNCTION	ORGANISATION	SIGNATURE	DATE
PREPARED	M. Santoro				
PREPARED	O. Cartus				
PREPARED					
VERIFIED	S. Quegan	Science Leader	Sheffield University		
APPROVED					

Document Distribution

ORGANISATION	NAME	QUANTITY
ESA	Frank Seifert	

Document History

VERSION	DATE	DESCRIPTION	APPROVED
0.1	2018-10-01	First draft version	
1.0	2018-11-15	Finalised version	

Document Change Record (from Year 1 to Year 2)

VERSION	DATE	DESCRIPTION	APPROVED

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	3	28.02.2019	

Table of Contents

List of Figures.....	4
Symbols and acronyms.....	5
1. Introduction.....	6
2. CCI Biomass CORE algorithm	7
3. Caveats of the CORE algorithm.....	9
4. Proposed development of CORE algorithm	10
4.1. Use of LiDAR observations	10
4.2. Revisiting the WCM to include allometry.....	12
4.2.1. Relating height to biomass	14
4.2.2. Relating canopy density to height	14
4.2.3. Estimation of tree attenuation	15
4.2.4. Validation of the WCM	16
4.3. Model training.....	17
4.4. Use of vegetation structural information.....	18
4.5. Use of Vegetation Optical Depth (VOD) observations	19
5. Conclusions.....	19
6. References	19

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	4	28.02.2019	

LIST OF FIGURES

Figure 2-1: Functional dependencies of datasets and approaches forming the CCI Biomass CORE global AGB retrieval algorithm in year 1. The shaded part of the flowchart represents potential improvements following the implementation of additional retrieval techniques [RD-5]..... 7

Figure 4-1: Observation geometry of the ATLAS laser instrument on-board the ICESAT-2 platform (taken from Neuenschwander and Pitts, 2019) 11

Figure 4-2: Observations of ICESAT GLAS canopy density and tree height RH100 for Sweden; vertical bars represent the interquartile range of canopy density for a given tree height range. The solid curve represents the least squares regression of Equation (4-9) to the observations. The curve has been extended to cover the range [0,40] m of potential heights in Sweden. The coefficient q in this example is equal to 0.0611. 15

Figure 4-3: Modelled backscatter using the WCM in Equation (4-10) and in Equation (2-1) (solid and dashed curves, respectively) for an ALOS PALSAR image acquired over a 30 x 30 km² area in southern Sweden on 2010/08/09. The backscatter observations have been grouped in 10 m³ ha⁻¹ wide stem volume (i.e., GSV, classes and illustrated in the form of average values (crosses) and 5th to 95th percentile; vertical bars). 16

Figure 4-4: Estimates of σ_{gr}^0 for an Envisat ASAR image obtained with self-calibration implemented in BIOMASAR-C (red circle) and with regression to observations of MODIS VCF and ASAR backscatter using Equation (4-3) (dark asterisk). The blue line represents the..... 18

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	5	28.02.2019	

SYMBOLS AND ACRONYMS

ADP	Algorithm Development Plan
AGB	Above Ground Biomass
ATBD	Algorithm Theoretical Basis Document
BCEF	Biomass Conversion & Expansion Factor
CCI	Climate Change Initiative
CCI-Biomass	Climate Change Initiative – Biomass
CD	Canopy Density
DARD	Data Access Requirements Document
E3UB	End to End ECV Uncertainty Budget
ECV	Essential Climate Variables
EO	Earth Observation
ESA	European Space Agency
FAO	Food and Agriculture Organization
GCOS	Global Climate Observing System
GEDI	Global Ecosystem Dynamics Investigation
GSV	Growing Stock Volume
ICESAT GLAS	Ice, Cloud, and land Elevation Satellite Geoscience Laser Altimeter System
PSD	Product Specification Document
PVASR	Product Validation and Algorithm Selection Report
PVP	Product Validation Plan
SAR	Synthetic Aperture Radar
SMOS	Soil Moisture & Ocean Salinity
SRTM	Shuttle Radar Topography Mission
URD	User Requirement Document
WCM	Water Cloud Model

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	6	28.02.2019	

Table 1-1: Reference Documents

ID	TITLE	ISSUE	DATE
RD-1	Users Requirements Document		
RD-2	Product Specification Document		
RD-3	Data Access Requirements Document		
RD-4	Product Validation and Algorithm Selection		
RD-5	Algorithm Theoretical Basis Document		
RD-6	End to End ECV Uncertainty Budget		
RD-7	Product Validation Plan		
RD-8	Algorithm Theoretical Basis Document of GlobBiomass project		

1. Introduction

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

One of the objectives of the CCI Biomass project is to generate global maps of AGB using a variety of Earth Observation (EO) datasets and state-of-the-art models for three epochs (2017-2018, 2018-2019 and 2010) and assess biomass changes on a 1-year difference and on a 10-years difference. The maps should be spatially and temporally consistent; in addition, they need to be consistent with other data layers thematically similar to the AGB dataset that are produced in the framework of the CCI Programme (e.g., Fire, Land Cover, Snow etc.).

Algorithms to estimate AGB from Earth Observation (EO) data and its changes are described in the Algorithm Theoretical Basis Document (ATBD) [RD-5] whereas the End to End Uncertainty Budget (E3UB) document [RD-6] describes the accuracy associated with the estimates of AGB. The ATBD and the E3UB documents are live documents, updated once yearly to provide a thorough description of the algorithms implemented to generate the AGB and AGB change maps. The ATBD and the E3UB documents describe the CORE algorithm used in Year 1 of the CCI Biomass project to generate a global dataset of AGB estimates for the 2017-2018 epoch. The CORE algorithm is based on the GlobBiomass global retrieval algorithm [RD-8] (see <http://globbiomass.org/products/global-mapping/>) and inverts observations of EO data, globally available for the epoch of interest. These datasets are similar to those used to generate a global map of AGB for the 2010 epoch in the GlobBiomass project (i.e., C- and L-band SAR backscatter, supported by ICESAT GLAS waveforms and additional auxiliary data layers).

This document builds on the ATBD and E3UB documents to identify major elements that require development in Year 2 of the CCI Biomass project in order to overcome errors in the estimates of AGB produced in Year 1. In addition, we consider the review of the GlobBiomass dataset and algorithms reported in the Product Validation and Algorithm Selection Report (PVASR) [RD-4] and the Product

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	7	28.02.2019	

Validation Plan (PVP) [RD-7] (Note that the PVASR is not available in final form at time of writing). This Algorithm Development Plan (ADP) will guide the developments needed to improve the global AGB dataset produced in Year 1 (2017-2018 epoch) and will support the generation of the AGB dataset for Year 2 (2018-2019 epoch). As for the ATBD and the E3UB documents, the ADP relies on the Users Requirements Document (URD) [RD-1], the Product Specifications Document (PSD) [RD-2] and the Data Access Requirements Document (DARD) [RD-3].

Section 2 reviews the CCI Biomass CORE algorithm. Section 3 elaborates on the known major weaknesses of the CORE algorithm based on the initial assessment of AGB retrieval reported in the ATBD. The PVP and the analyses reported in the PVASR provide further information on these weaknesses. Section 4 lists potential solutions to the issues identified in Section 3.

2. CCI Biomass CORE algorithm

Figure 2-1 shows the flowchart of the CORE biomass estimation procedure implemented in Year 1 of the CCI Biomass project to generate a global dataset of AGB estimates for the 2017-2018 epoch [RD-5].

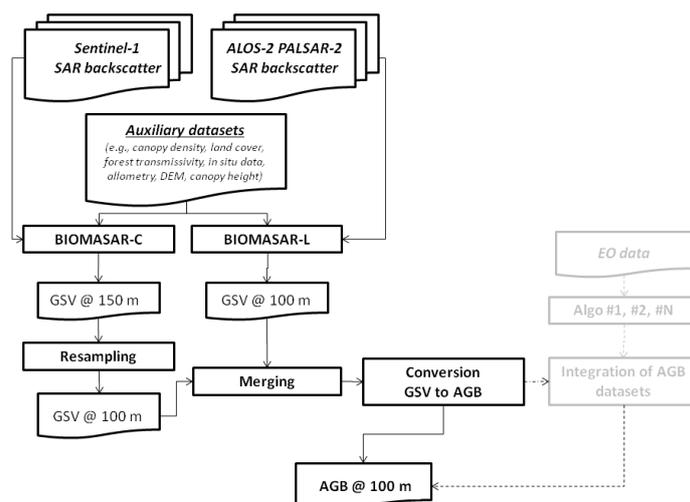


Figure 2-1: Functional dependencies of datasets and approaches forming the CCI Biomass CORE global AGB retrieval algorithm in year 1. The shaded part of the flowchart represents potential improvements following the implementation of additional retrieval techniques [RD-5].

Two independent estimates of growing stock volume (GSV, unit: m^3/ha) are obtained from the BIOMASAR algorithm adapted to ingest C-band SAR backscatter data from Sentinel-1 (BIOMASAR-C) and L-band SAR backscatter data from ALOS-2 PALSAR-2 (BIOMASAR-L). These estimates are combined using linear weighting to obtain a final estimate that should have higher precision than the original values. The C- and L-band datasets have different pixel spacing, so the GSV estimates from the BIOMASAR-C algorithm (with slightly lower resolution) are resampled to the geometry of the BIOMASAR-L estimates. Finally, GSV is converted to AGB using a Biomass Conversion and Expansion Factor (BCEF).

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	8	28.02.2019	

Both BIOMASAR-C and BIOMASAR-L invert a Water Cloud Model (WCM) to estimate GSV from a measurement of SAR backscatter.

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

In Equation (2-1), σ_{gr}^0 and σ_{veg}^0 represent the backscattering coefficients of the ground and vegetation layer, and β is an empirically defined coefficient expressed in ha m^{-3} relating to the two-way forest transmissivity. The three parameters are unknown *a priori* and must be estimated. Usual approaches achieve this by using a training set of *in situ* GSV measurements and corresponding SAR backscatter measurements. This is not viable when mapping large areas because *in situ* observations are only available at local sites and even national forest inventories do not have a sufficient density of measurements to accurately estimate the model parameters. An alternative would be to use existing maps (i.e., wall-to-wall estimates of GSV) but these carry a serious risk of introducing systematic errors that would propagate to the final estimate. Note that the same considerations apply to the WCM expressed in terms of AGB instead of GSV. Hence the CCI Biomass CORE algorithm is couched primarily in terms of estimating GSV. A fuller explanation of the reasoning behind this choice is given in Section 4.1 of the ATBD [RD-5].

In the BIOMASAR algorithm, model training relies on a self-calibration approach, using statistical measures from the histogram of the SAR backscatter for areas characterized by a specific canopy density range provided as an auxiliary dataset. Note that BIOMASAR does not use the actual values in the canopy density dataset but uses it as a mask. The estimate of σ_{gr}^0 corresponds to the mean value of the SAR backscatter for pixels characterized by low canopy cover (so-called “ground” pixels) around the pixel for which an estimate is sought. The estimate of σ_{veg}^0 is obtained from the mean value of the SAR backscatter for pixels characterized by high canopy cover (so-called “dense forest” pixels) around the pixel for which an estimate is sought. This value, referred to as σ_{df}^0 , is compensated for the residual ground contribution due to the fact that a canopy is not completely opaque. The estimation of σ_{veg}^0 from σ_{df}^0 and σ_{gr}^0 is done in slightly different ways depending on whether GSV is estimated from C-band (BIOMASAR-C) or L-band SAR backscatter data (BIOMASAR-L) [RD-5]. In addition, an estimate of the coefficient of the two-way forest transmissivity, β , is needed, together with an estimate of the GSV of dense forest (V_{df} , BIOMASAR-C) or the canopy height and density of dense forests (h_{df} and η_{df} , BIOMASAR-L) [RD-5]. The parameters characterising dense forest are obtained from LiDAR observations (h_{df} and η_{df}) or inventory-based measurements (V_{df}) downscaled with LiDAR observations. The forest transmissivity relies again on observations of canopy density and canopy height. In this sense, LiDAR observations are central to obtaining estimates of parameters of the WCM that are not directly obtainable from the SAR backscatter observations. To adapt the estimates of the transmissivity term to vegetation characteristics, stratification based on the FAO ecological zones is applied. This rather coarse delineation of ecological zones, however, does not allow a finer estimation of forest transmissivity to reproduce forest structure (e.g., all rainforest are represented by a single forest transmissivity coefficient).

This procedure takes into consideration that the observables available globally to estimate AGB are sub-optimal (C- and L-band backscatter) and are characterized by distortions due to environmental conditions, which can vary spatially even over small scales. BIOMASAR was built on a simple modelling framework that has to deal with signals with low sensitivity to biomass. To reduce the impact of spatially variable environmental conditions on the forest backscatter, BIOMASAR adapts the estimation of the model parameters to a pixel-by-pixel basis (sliding window estimation). In addition,

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	9	28.02.2019	

where available, BIOMASAR combines estimates of AGB from multiple observations so to reduce the contribution of the temporally variable environmental conditions on the retrieved AGB.

To improve the precision of the estimate of GSV, the independent estimates of GSV from C- and L-band data are combined using weighted averaging with rules related to the major sources of errors identified in the maps produced from BIOMASAR-C and BIOMASAR-L. Again, the modelling framework to obtain a merged estimate of GSV has been kept simple because the full characterization of systematic errors in each dataset is not possible (too few in situ observations, non-systematic SAR pre-processing errors etc.) [RD-5].

In a last step, GSV is converted to AGB, basically by implementing the FAO approach to estimating forest resources. More than 80 % of the countries reporting to the FAO Forest Resources Assessments declared their stocks in terms of volume and applied a conversion factor to derive estimates of total biomass and total carbon stocks. The major issue when attempting to estimate the conversion factor from GSV and AGB is the representativeness of databases of wood density and total-to-stem biomass proportions used to derive wall-to-wall rasters of the BCEF. Wood density and, even more so the fraction of biomass in a tree's components (branch, stem, roots), has strong spatial variability [RD-5] so that any interpolation to derive spatially explicit maps of the BCEF is likely to miss small-scale spatial patterns, although over large area the mean value is unbiased [RD-5].

3. Caveats of the CORE algorithm

The above brief summary of the CCI Biomass CORE algorithm highlights the major elements of the retrieval approach and acknowledges that this may not be the best possible algorithm but rather is a global approach constrained by the available EO data and on ground observations. The GlobBiomass and CCI Biomass CORE algorithms rely on a number of assumptions that appear viable when comparing large-scale averages of estimated AGB with corresponding values based on inventory information [RD-5] and [RD-7]. Nevertheless, these assumptions, which were made in order to allow the CORE algorithm to perform globally, also introduce systematic errors into the retrieved biomass, which may become apparent when focusing on particular areas [RD-4], [RD-5] and [RD-7]. In the ATBD, we provided a list of potential areas of improvement of the CORE algorithm. These are reported below and then expanded in the next Section with a proposed development of the CORE algorithm

- The retrieval of AGB misses the high values of AGB, because of the loss of sensitivity of the major biomass predictors in high biomass forest. For this, we see the use of tree height information as extremely relevant. New missions carrying laser instruments have the potential to make a substantial contribution to estimating AGB in high-density forest.
- The retrieval of AGB is based on simplifying assumptions that cause the retrieval models to be too general to capture the spatial variability of the relationship between observables and vegetation properties. Vegetation structural information, as developed in the DARD [RD-3], should provide the backbone for more targeted estimation of model parameters. Also, knowledge gathered by investigating the relationship between EO observables and AGB in specific forest classes should be exploited.
- Alternative solutions to estimating AGB which can overcome the systematic over- and underestimation issues of the CORE algorithm need to be explored. We are particularly interested in the potential offered by L-band VOD observations [RD-4] and the links between different forest variables (canopy density, canopy height, volume and biomass) [RD-4]. In

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	10	28.02.2019	

addition, the relationship between biomass-related parameters and wood density needs to be further addressed with specific attention on capturing the local scale spatial variability of wood density.

- Almost 20 % of the world's forests grow on sloping terrain, which causes significant distortions in optical and radar datasets that translate into significant biomass estimation errors. Pre-processing of the EO data has been demonstrated to help but, due to data policy and processing capability issues, it is not realistic for the CORE algorithm to include full pre-processing in the near future. Interaction with space agencies is being pursued to grant access to data relevant to improving the estimation of AGB, but this is likely to be a lengthy process. It is therefore important to put some effort into techniques that minimize the impact of slopes on the retrieval.

4. Proposed development of CORE algorithm

4.1. Use of LiDAR observations

Observations that sense forest structure are of major benefit to estimation of biomass. Unfortunately, the majority of EO data available globally is in the form of energy reflected to the sensor, so that biomass can only be inferred with parametric or non-parametric approaches (Santoro and Cartus, 2018). SAR interferometry and laser scanning instead generate observations that contain information on the vertical and the horizontal distribution of vegetation, thus providing a more direct measure of parameters involved in the computation of biomass (canopy height, density of canopy).

The TanDEM-X and SRTM missions were conceived to acquire interferometric datasets that would allow the generation of surface elevation models (Farr et al., 2007; Krieger et al., 2007). Over forested terrain, an estimate of vegetation height can be inferred from the surface elevation assuming that the terrain elevation is known. To obtain the true vegetation height, an additional step that compensates the InSAR-based height of the vegetation for the penetration of microwaves into the canopy is required (Walker et al., 2007). Although high resolution and accurate (surface) elevation models based on interferometric data exist, there is no global dataset of terrain elevation, which hinders the use of interferometry for a “direct” measure of the vegetation vertical structure.

Laser instruments also measure the elevation of the Earth surface and, in the case of vegetation, return a profile of reflection intensity along the vertical direction. The GLAS instrument on-board the ICESAT satellite operated between 2003 and 2009 and recorded millions of waveforms along its orbital path. Unlike interferometric datasets, the signal recorded by a laser instrument contains also a ground return, so that an external dataset of terrain elevation is not required to estimate the height of vegetation. Waveform information in the GLA14 product was processed globally in the GlobBiomass project [RD-8] from which canopy density and several height percentiles were computed. A GLAS footprint has an approximately 70 m diameter and footprints were acquired sequentially along an orbit; however, the distance between orbits was around 60 km, leading to a somewhat sparse sampling of the Earth's vegetation. For this reason, it was preferred to use the ICESAT GLAS datasets of canopy height and canopy density to support modelling of backscatter as a function of biomass rather than as surrogate reference data for model training. This is indeed a widely used approach in global mapping efforts but was not applied in the GlobBiomass project because of the sparse sampling and the evidence that, at the time of the project, allometric functions relating height and AGB were poorly characterized globally.

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	11	28.02.2019	

To stress the importance of ICESAT GLAS height and canopy density in the context of global biomass modelling, we summarize the multiple contributions of the data below:

- Estimate the forest transmissivity coefficient in the WCM
- Model the maximum GSV to constrain the retrieval from SAR backscatter
- Define the merging rules to combine L-band and C-band GSV estimates

LiDAR senses vegetation structure and therefore is a primary candidate to estimate biomass. From a global point of view, however, the coverage until recently has been insufficient to achieve wall-to-wall estimates of AGB. The recent start of operations of two spaceborne LiDAR missions (GEDI on-board the International Space Station and ICESAT-2) potentially allows datasets similar to those obtained with airborne laser scanning systems to be obtained and these are likely to provide major support to estimation of biomass in the CORE algorithm.

The Global Ecosystem Dynamics Investigation (GEDI, <https://gedi.umd.edu>) instrument has operated since December 2018 and covers latitudes between 55°N and 55°S. GEDI is a full waveform LiDAR, specifically designed to observe vegetation. Sampling and spatial resolution are higher than ICESAT GLAS (1 km vs. 60 km cross-track, 60 m vs. 170 m along-track, 25 m vs. 70 m footprint diameter) so that a LiDAR-only biomass product (1 km spatial resolution) is an explicit target. The GEDI team is investing significant efforts in collecting reference data and characterizing allometric functions to establish the relationship between waveform parameters and AGB to achieve this target.

ICESAT-2 has operated since October 2018 (first official data release: spring 2019) and has global coverage. The sampling is slightly poorer than that of GEDI (Figure 4-1), whereas the size of a footprint is slightly smaller (20 m diameter) (Neuenschwander and Pitts, 2019). Unlike GEDI, the ATLAS instrument on-board ICESAT-2 is a photon counting laser, meaning that the observations are not in the form of a waveform and are therefore not optimized for vegetation studies. Although the focus of the mission is on cryosphere, vegetation studies are one of the thematic targets of the mission. Extraction of terrain and canopy heights from the ATLAS point clouds will produce the ATL08 geophysical data product (Neuenschwander & Pitts, 2019).

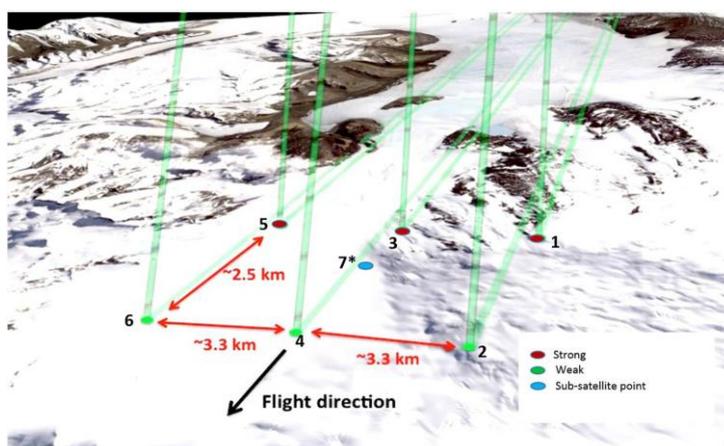


Figure 4-1: Observation geometry of the ATLAS laser instrument on-board the ICESAT-2 platform (taken from Neuenschwander and Pitts, 2019)

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	12	28.02.2019	

From the perspective of generating a global AGB data product, exploring the complementarity of data from both laser instruments is mandatory. However, this would involve a potentially huge amount of work to understand optimal trade-offs between datasets, filtering spurious observations, etc. If the current approach to estimate AGB in the CORE algorithm is retained, the higher density of samples compared to ICESAT GLAS would allow a finer definition of the parameters currently in use in the WCM and for merging. For example, transmissivity and maximum biomass could be adapted more locally. We do not foresee a direct use of the GEDI and ICESAT-2 LiDAR metrics for estimating biomass on a footprint basis as this is the focus of the GEDI and ICESAT-2 teams. Exploring the relationship between ALOS-2, Sentinel-1 and LiDAR observations is instead aimed at to overcome the limitations to retrieve biomass in forests where the sensitivity of the SAR backscatter to biomass is weak.

4.2. Revisiting the WCM to include allometry

The WCM with gaps (Attema & Ulaby, 1978; Askne et al., 1997) originally expressed the total forest backscatter as the sum of direct scattering from the ground through gaps in the canopy, ground scattering attenuated by the canopy and direct scattering from the vegetation.

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

In Equation (4-1), η represents the area-fill or tree cover factor (i.e., the fraction of the area covered by vegetation.) For short wavelengths, it can reasonably be assumed that this parameter corresponds to canopy density. The coefficients σ_{gr}^0 and σ_{veg}^0 represent the backscattering coefficients of the ground and vegetation layer, respectively. T_{tree} represents the two-way tree transmissivity. This can be expressed as

$$T_{tree} = e^{-\alpha h} \quad (4-2)$$

where α represents the two-way attenuation per meter through the tree canopy and h is the depth of the attenuating layer. The terms in Equation (4-1) can be grouped differently in order to highlight the contribution from the forest floor and the contribution from the vegetation

$$\sigma_{for}^0 = 1 - \eta(1 - e^{-\alpha h})\sigma_{gr}^0 + \eta(1 - e^{-\alpha h})\sigma_{veg}^0 \quad (4-3)$$

or likewise

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

where

$$T_{for} = 1 - \eta(1 - e^{-\alpha h}) \quad (4-5)$$

expresses the two-way forest transmissivity, combining gap fraction and two-way tree transmissivity.

Equation (4-5) relates the forest transmissivity to two forest structural parameters, height and canopy cover, as well as a parameter related to the ability of microwaves to penetrate through the vegetation, α . For the purpose of estimating a forest variable from an observation of SAR backscatter, Equation (4-1) is not useful since the area-fill factor is not a parameter of interest to foresters. For retrieval

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	13	28.02.2019	

purposes, the form of the WCM has been kept but the backscatter has been expressed as a function of a forest attribute related to biomass (i.e., AGB, Mg ha⁻¹) or stem volume / growing stock volume (GSV, unit: m³ ha⁻¹) introducing (semi-)empirical coefficients (e.g. the coefficient β of Equation (2-1)). This was found to be a source of error when retrieving biomass because empirical coefficients are difficult to relate to vegetation properties and environmental conditions.

Equation (4-1) can be written in terms of AGB if functional dependencies between canopy density, canopy height and AGB are identified. Canopy density and canopy height can be estimated from laser waveforms to establish a functional dependency where height would be the independent variable and canopy density the dependent variable. Having expressed the forest transmissivity as a function of height only, the next step would be to relate height to biomass by mean of allometric equations. In summary, Equation (4-5) can be rewritten in a general form as:

$$T_{for} = f(\eta, h, \alpha) = f(\eta(h(B)), h(B), \alpha) \quad (4-6)$$

where we assume that B represents biomass, which could be either GSV or AGB.

When comparing the definitions of forest transmissivity in Equations (2-1) and (4-1), the following equality should hold:

$$e^{-\beta V} = 1 - \eta(1 - e^{-\alpha h}) \quad (4-7)$$

This is obviously not the case, meaning that the approximation of describing the forest transmissivity as an exponential with a linear term (βV) can lead to biased estimates of AGB in some AGB ranges, which is a feature of the GlobBiomass dataset.

- To validate Equation (4-6), we looked at the three elements that are new to the modelling of SAR backscatter with the WCM compared to previous work.
- 1. Characterization of the allometry linking the biomass variable of interest (B, either AGB or V) and height

$$h = f(B)$$

2. Characterization of the relationship between area-fill (canopy density) and height

$$\eta = f(h) = g(f(B))$$

3. Simulation of forest backscatter with Equation (4) and benchmarking with observations of SAR backscatter and stem volume to estimate the two-way tree attenuation, α .

$$T_{for} = f(\alpha, B)$$

This work has been undertaken with ALOS-2 PALSAR-2 data in Swedish forests, and made possible by the close cooperation with the National Forest Inventory in providing countrywide coverage of data from field-measured inventory plots. Nonetheless, we try to put the results and related discussion in the context of global mapping to understand whether a potential development of this approach is feasible.

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	14	28.02.2019	

4.2.1. Relating height to biomass

The relationship between tree height and biomass in boreal and temperate forest can be expressed by a power-law function of the type (Askne et al., 1997; Santoro et al., 2002; Mette et al., 2004; Santoro et al., 2007; Cartus et al., 2010; Santoro et al., in preparation):

$$h = (aV)^b \quad (4-8)$$

It is remarked that in boreal forests, the term “biomass” can be used interchangeably between GSV, as in Equation (4-8), and AGB because of the few species and the similar branching patterns, leading to a small variability of BCEF values across the biome. The coefficients change depending on the way AGB is accumulated by the forest due to environmental conditions, availability of light, self-thinning and forest management, etc. While this appears to be a promising approach, it is clear that there is a strong requirement to accurately characterize the functional dependence between AGB and height. Saatchi et al. (2011) introduced pan-tropical allometric functions relating height and AGB, thus indicating that such an approach would also be viable in extra-boreal environments. An intriguing question at this stage is how detailed the characterization of the $h = f(B)$ function needs to be so that it does not introduce biases when retrieving AGB. This will be assessed by investigating allometric functions available (close interaction with WP1000), their generalization and their performance when inserted in the modelling framework of the CORE algorithm. It should also be noted that expressing height as a function of AGB would overcome the issue of retrieving GSV first and then scaling it with poorly characterized BCEF estimates. There is still a need to characterize the spatial distribution of wood density but the current CCI Biomass team lacks the relevant expertise to pursue this.

4.2.2. Relating canopy density to height

The relationship between canopy density and tree height has been investigated with ICESAT GLAS data. Canopy density was defined as the percentage of energy returned from heights of more than 2 m above the ground, whereas canopy height was defined as the RH100 height, (i.e., the distance between the signal beginning and the ground peak.) Unfortunately, we did not have *in situ* data to confirm the validity of such definitions as ICESAT GLAS footprints did not cover the area sampled by the Swedish NFI for their measurements.

Observations of canopy density (CD) and canopy height (h) from the ICESAT GLAS dataset of Sweden show a clear relationship, which was found to be modelled with the following function (Santoro et al., in preparation):

$$CD = 1 - e^{-qh} \quad (4-9)$$

In Equation (4-9), q is an empirical parameter. Observations of CD and tree height from the ICESAT dataset for Sweden and the corresponding model fit with Equation (4-9) are shown in Figure 4-2.

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	15	28.02.2019	

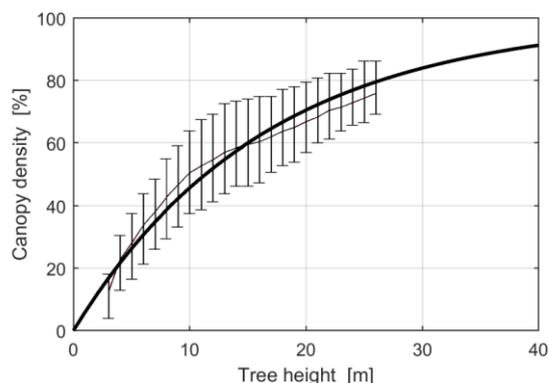


Figure 4-2: Observations of ICESAT GLAS canopy density and tree height RH100 for Sweden; vertical bars represent the interquartile range of canopy density for a given tree height range. The solid curve represents the least squares regression of Equation (4-9) to the observations. The curve has been extended to cover the range [0,40] m of potential heights in Sweden. The coefficient q in this example is equal to 0.0611.

When plotting the observations of canopy density and height in other regions of the world, we identified the same functional dependencies, albeit for different values of the q parameter. In rainforest, canopy density increased for increasing height very rapidly, then reached a plateau. In contrast, sparse forests at the northernmost latitudes and in the dry tropics showed a more linear type of relationship. These results suggest that a LiDAR-based characterization of the functional dependency between canopy density and height would effectively reduce the number of forest variables in the WCM of Equation (4-1) to height only. This would be feasible globally provided that the spatial variability of the q coefficient is well captured.

4.2.3. Estimation of tree attenuation

Having fixed the functional dependencies between height and AGB on the one hand, and canopy density and height on the other, the WCM becomes invertible once the coefficients σ_{gr}^0 and σ_{veg}^0 , and the two-way tree attenuation coefficient α have been estimated. Using the BIOMASAR approach, one can estimate σ_{gr}^0 and a measure of the backscatter from dense forest σ_{df}^0 , from which one can obtain σ_{veg}^0 . By replacing Equation (4-9) in the WCM, the total forest backscatter can be rewritten as:

$$\sigma_{for}^0 = \sigma_{gr}^0 (e^{-qh} + e^{-\alpha h} - e^{-(q+\alpha)h}) + \sigma_{veg}^0 (1 - e^{-qh} - e^{-\alpha h} + e^{-(q+\alpha)h}) \quad (4-10)$$

with

$$\sigma_{veg}^0 = \frac{\sigma_{df}^0 - \sigma_{gr}^0 (e^{-qh_{df}} + e^{-\alpha h_{df}} - e^{-(q+\alpha)h_{df}})}{1 - (e^{-qh_{df}} + e^{-\alpha h_{df}} - e^{-(q+\alpha)h_{df}})} \quad (4-11)$$

where the height terms can be expressed in terms of AGB using the allometry for the specific region of interest. Equation (4-10) can be used to estimate α by fitting the model to observations of the forest backscatter and biomass. For Sweden, the forest backscatter was simulated with the WCM in Equation (4-10) using α values between 0.1 dB m⁻¹ and 2 dB m⁻¹ in steps of 0.1 dB m⁻¹ and using a LiDAR-based map of biomass for Sweden as reference (Nilsson et al., 2017). Although α was found to increase from 0.3 dB m⁻¹ to 0.7 dB m⁻¹ with decreasing latitude (Santoro et al., in preparation), the non-negligible uncertainty of the estimates and the marginal effect of the value of α on the retrieval (for values below

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	16	28.02.2019	

1 dB m⁻¹) led us to conclude that the value of 0.5 dB m⁻¹ used in the past [RD-8] is acceptable for retrieval purposes, at least in the boreal zone. If the WCM in Equation (4-1) is to be pursued in a global context, investigations should focus on the spatial characterization of α in extra-boreal forests and at C-band. This requires LiDAR maps of GSV and AGB to act as reference for assessing the sensitivity of the WCM to α .

4.2.4. Validation of the WCM

The validity of the WCM in Equation (4-1) is confirmed by an example in Figure 4-3 where the modelled backscatter (solid line) is plotted, together with observations of ALOS PALSAR backscatter and GSV from the dataset of Nilsson et al., (2017). We also plotted the modelled backscatter obtained with Equation (2-1). The WCM in Equation (4-10) follows the trend in the measurements whereas the WCM in Equation (2-1) predicts somewhat lower SAR backscatter and a slower increase for increasing GSV (Santoro et al., in preparation). To quantify the impact of the new modelling scenario, consider a pixel with HV-polarized backscatter value of -14 dB. The inversion of Equation (4-1) (i.e., 4-10), produces an estimate of GSV of 30 m³ ha⁻¹. Inverting Equation (2-1) instead produces an estimate of 75 m³ ha⁻¹ (i.e., a value 2.5 times larger).

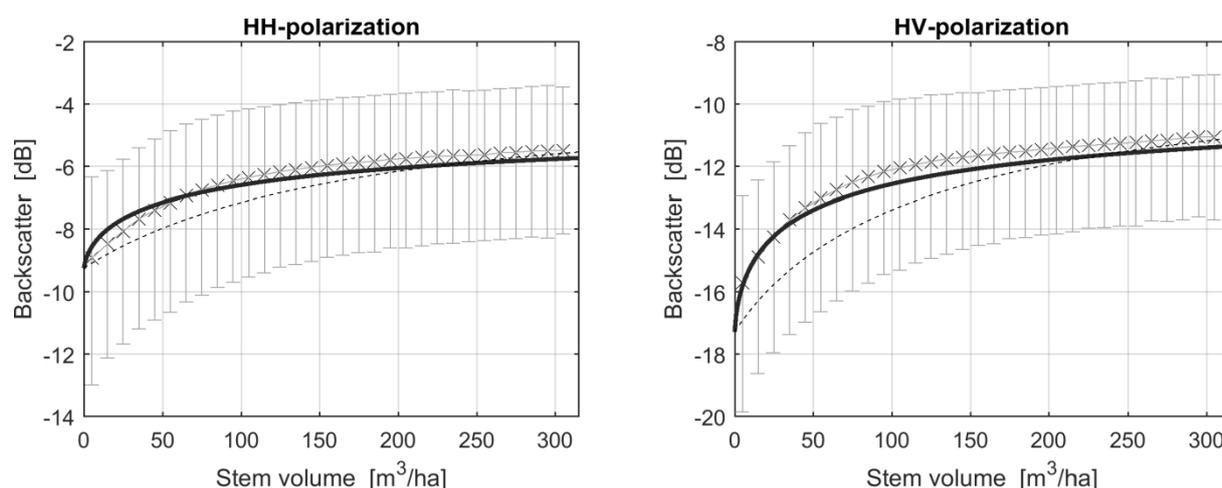


Figure 4-3: Modelled backscatter using the WCM in Equation (4-10) and in Equation (2-1) (solid and dashed curves, respectively) for an ALOS PALSAR image acquired over a 30 x 30 km² area in southern Sweden on 2010/08/09. The backscatter observations have been grouped in 10 m³ ha⁻¹ wide stem volume (i.e., GSV, classes and illustrated in the form of average values (crosses) and 5th to 95th percentile; vertical bars).

This preliminary suggests that overestimation in the low biomass range can be addressed in the boreal zone, but the question remains whether this approach will perform equally well in other biomes and forest conditions. We intend to test this in dry sparse forests. Note that the backscatter values modelled with Equation (4-10) are not substantially different from those obtained from Equation (2-1) so this solution is primarily for forests with low AGB values.

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	17	28.02.2019	

4.3. Model training

The WCM, regardless whether expressed as a function of GSV, height or AGB (see PVASR), contains parameters that need to be estimated in order to make the model invertible.

The traditional approach to train the WCM consists of regressing observations of the remote sensing observable (e.g., the SAR backscatter) to in situ observations of the forest variable of interest (e.g., AGB). Accurate estimation of the model parameters requires that the entire range of AGB is represented by the samples of in situ observations and the size of the inventory plots is comparable to the size of the remote sensing observation (the pixel). Once the model has been trained, the question remains to which extent the estimates of the model parameters can be considered to be representative in nearby landscapes. Environmental conditions at the time of image acquisition or forest structure might be different. Such differences would reflect in a different shape of the WCM, which, if unaccounted for, would translate to biases in the retrieved biomass.

While it is acknowledged that training the WCM with reference data is the most accurate way to describe the relationship between remote sensing observables and forest variables, the issue of representativeness of the model was the main driver to develop the BIOMASAR algorithm, i.e., an algorithm that would invert the WCM globally being robust to environmental conditions and vegetation structural characteristics under the evidence that reference datasets are not available for most regions of the world. The self-calibration approach of the backscatter model parameters, the assumption that forest transmissivity can be characterized at ecoregion level and the introduction of a maximum retrievable biomass respond to the requirements set above. In some cases, there are some crude assumptions behind the model training approaches implemented in BIOMASAR and they are not error-free. The PVASR has documented for dry tropical forests in southern Africa that the GlobBiomass dataset was affected by systematic underestimation in the low biomass range. Looking at the estimates of the model parameter representing the backscatter of the ground (i.e., the backscatter for AGB = 0 Mg/ha) estimated with BIOMASAR at C-band using self-calibration, we understand now that the value is often 0.5 to 1 dB higher than what would be obtained by fitting the WCM to “observations”. This is illustrated in Figure 4-4 for an Envisat ASAR image acquired during the dry period of the year over a 2° x 2° area in central Mozambique. The red circle represents the estimate of σ_{gr}^0 using self-calibration (MODIS VCF < 15% in a 100 km x 100 km area). The black asterisk represents the estimate obtained by training the WCM expressed in Equation (4-3) as a function of area-fill regressing ASAR backscatter observations to MODIS VCF observations. The modelled backscatter appears to follow the trend of the observations, which instead would have not been true if one had used the σ_{gr}^0 estimated with self-calibration. The reason for the higher value obtained with self-calibration is that the estimate was obtained as the mean of backscatter measurements for MODIS VCF between 0 and 15%, a range in which the backscatter presents a rapid increase (Figure 4-4). Although this result was unique to southern African landscapes, it indicates that it is necessary to reconsider how σ_{gr}^0 is estimated by using more robust approaches to avoid that, ultimately, the biomass is biased across an entire landscape. In this sense, the PVASR suggested possible approaches (inclusion of soil moisture estimates, regression to reference data) that would need to be benchmarked against self-calibration.

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	18	28.02.2019	

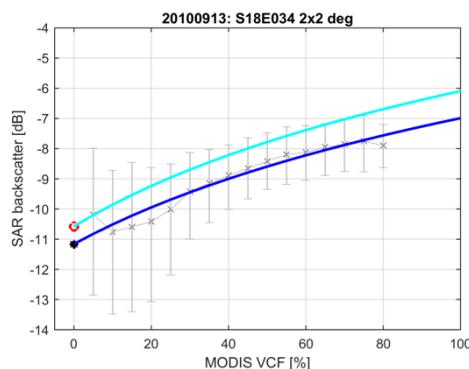


Figure 4-4: Estimates of σ_{gr}^0 for an Envisat ASAR image obtained with self-calibration implemented in BIOMASAR-C (red circle) and with regression to observations of MODIS VCF and ASAR backscatter using Equation (4-3) (dark asterisk). The blue line represents the fit of Equation (4-3) to the observations. Observations are represented in the form of mean backscatter and 5th-95th percentile (crosses and vertical bars) for 5% intervals of VCF. The cyan line represents the model fit of Equation (4-3) using the model parameters estimates obtained with self-calibration.

4.4. Use of vegetation structural information

One of the limitations of the currently implemented BIOMASAR algorithms is the coarse representation of vegetation structure. The world is divided into the FAO ecological zones, for each of which some of the parameters of the WCM are then estimated. Vegetation structural information as developed in the DARD [RD-3] should provide more targeted estimation of model parameters, in particular the transmissivity terms. A test that uses the segmentation and classifications proposed in [RD-3] and [RD-4] for various regions is planned in Year 2. We expect this to give a better adaptation of the shape of the WCM to the local vegetation conditions, but it will not contribute to systematic removal of biases as discussed in Section 4.2.

In the same vein, knowledge gathered by investigating the relationship between EO observables and AGB in specific forest classes will be exploited. When evaluating the GlobBiomass map in flooded forest and in mangrove forests, the specific scattering mechanisms occurring at C- and L-band are not being correctly accounted for in the WCM, which is the major source of error. The AGB of mangroves is often underestimated because the absorption of microwaves into the canopy leads to low backscatter, and the AGB of flooded forest is often overestimated because of double-bounce effects causing higher backscatter.

An alternative way to estimate the AGB of mangrove ecosystems has been presented in the PVASR [RD-4] where TanDEM-X elevations are used to estimate height, which are then converted with allometric functions to biomass (Simard et al., 2018). The important aspect of this approach is that models and related accuracies are documented and therefore transferable to the CCI Biomass retrieval framework. The first indications reported in the PVASR, however, indicate that only height and allometry do not seem to produce error-free estimates. In this respect integration of backscatter and heights should be explored to avoid that allometries ultimately impact the final estimate of AGB.

For flooded forests, the most practical solution is to remove images acquired during inundation periods and to exploit the temporal sequence of observations to identify possible outliers of high backscatter (e.g., characterized by high soil moisture). This was not feasible in the GlobBiomass project because

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	19	28.02.2019	

the AGB of flooded forests was based on the single estimate from the L-band ALOS PALSAR mosaic of 2010. However, it is now practicable because we have Sentinel-1 time series with a spatial resolution similar to ALOS-2 data (Figure 2-1) and a multi-temporal coverage of ALOS-2 observations from the ScanSAR mode in the tropics.

4.5. Use of Vegetation Optical Depth (VOD) observations

From the analyses reported in the PVASR [RD-4], it is clear that the AGB of high AGB forests needs to be better estimated. L-band VOD observations from 25 km resolution SMOS data may be able to address this issue, but it is unclear whether an estimate at such coarse resolution can be transferred to the moderate resolution of the AGB maps to be generated in the CCI Biomass project. An analysis of downscaling approaches and their accuracy is therefore needed, which should be linked to the work proposed in Section 4.1 to arrive at a unified approach.

5. Conclusions

Since it is still in its early stages, the development of the CORE retrieval algorithm of the CCI Biomass project has many aspects with a wide range of potential approaches. Some appear to be more mature (e.g., use of LIDAR data) whereas others require substantial efforts, in particular those aiming at ingesting new data sources and the associated need for an end to end characterization of uncertainties.

6. References

- Askne, J., Dammert, P. B. G., Ulander, L. M. H. and Smith, G. (1997). C-band repeat-pass interferometric SAR observations of the forest. *IEEE Transactions on Geoscience and Remote Sensing*, 35 (1), 25-35.
- Cartus, O., Santoro, M., Schmullius, C. and Li, Z. (2011). Large area forest stem volume mapping in the boreal zone using synergy of ERS-1/2 tandem coherence and MODIS vegetation continuous fields. *Remote Sensing of Environment*, 115 (3), 931-943.
- Farr, T., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. and Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Review of Geophysics*, 45 (RG2004), 1-33.
- Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M. and Zink, M. (2007). TANDEM-X: a satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45 (11), 3317-3341.
- Mette, T., Papathanassiou, K., Hajnsek, I., Pretzsch, H. and Biber, M. (2004). Applying a common allometric equation to convert forest height from Pol-InSAR data to forest biomass. In: *IGARSS'04*. City: IEEE Publications, Piscataway, NJ, 269-272.
- Neuenschwander, A. and Pitts, K. (2019). The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment*, 221, 247-259.
- Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J. and Olsson, H. (2017). A nationwide forest attribute map of Sweden predicted

	Ref	CCI Biomass Algorithm Development Plan v1		
	Issue	Page	Date	
	1.0	20	28.02.2019	

using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sensing of Environment*, 194, 447-454.

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

Santoro, M. and Cartus, O. (2018). Research pathways of forest above-ground biomass estimation based on SAR backscatter and interferometric SAR observations. *Remote Sensing*, 10 (4), 608, doi:10.3390/rs10040608.

Santoro, M., Fransson, J. and Cartus, O. (in preparation), Integrating SAR backscatter, ICESAT GLAS metrics and allometric functions towards an improved estimation of forest biomass.

Simard, M., Fatoyinbo, L., Smetanka, C., Rivera-Monroy, V. H., Castañeda-Moya, E., Thomas, N. and Van der Stocken, T. (2019). Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nature Geoscience*, 12, 40-45.

Walker, W. S., Kelldorfer, J. M., LaPoint, E., Hoppus, M. and Westfall, J. (2007). An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, 109 (4), 482-499.