# CCI

# BIOMASS

# END TO END ECV UNCERTAINTY BUDGET YEAR 1

# VERSION 1.0

| | |
|---|---|
| DOCUMENT REF: | CCI_BIOMASS_E3UB_V1 |
| DELIVERABLE REF: | D2.3-E3UB |
| VERSION: | 1.0 |
| CREATION DATE: | 2019-02-28 |
| LAST MODIFIED | 2019-02-28 |

## Document Authorship

| | NAME | FUNCTION | ORGANISATION | SIGNATURE | DATE |
|---|---|---|---|---|---|
| PREPARED | M. Santoro | | GAMMA | | |
| PREPARED | O. Cartus | | GAMMA | | |
| PREPARED | | | | | |
| PREPARED | | | | | |
| 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 |
|---|---|---|---|
| 0.1 | 2019-02-28 | First draft version | |
| 1.0 | 2019-02-28 | Finalised version | |

## Document Change Record (from Year 1 to Year 2)

| VERSION | DATE | DESCRIPTION | APPROVED |
|---|---|---|---|
| | | | |
| | | | |

## Table of Contents

# LIST OF FIGURES

# SYMBOLS AND ACRONYMS

| | |
|---|---|
| ADP | Algorithm Development Plan |
| AGB | Above-ground Biomass |
| ALS | Airborne Laser Scanner |
| ASAR | Advanced Synthetic Aperture Radar |
| ATBD | Algorithm Theoretical Basis Document |
| BCEF | Biomass Conversion & Expansion Factor |
| BEF | Biomass Expansion Factor |
| CCI | Climate Change Initiative |
| CCI-Biomass | Climate Change Initiative – Biomass |
| 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 |
| FAO | Food and Agriculture Organization |
| FBD | Fine Beam Dual |
| GCOS | Global Climate Observing System |
| GEZ | Global Ecological Zones |
| GLAS | Geoscience Laser Altimeter System |
| ICESAT GLAS | Ice, Cloud, and land Elevation Satellite Geoscience Laser Altimeter System |
| JAXA | Japan Aerospace Exploration Agency |
| MPI-BGC | Max Planck Institute for Biogeochemistry |
| PSD | Product Specification Document |
| PVASR | Product Validation and Algorithm Selection Report |
| SAR | Synthetic Aperture Radar |
| URD | User Requirements Document |
| VCF | Vegetation Continuous Fields |

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 | Algorithm Development Plan | | |
| 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 using 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-year 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 produced in the framework of the CCI Programme (e.g., Fire, Land Cover, Snow etc.).

Algorithms to estimate AGB and its changes are described in the Algorithm Theoretical Basis Document (ATBD) [RD-5]. The scope of this document is to define and quantify the uncertainties associated with the biomass estimates. This End to End ECV Uncertainty Budget document (E3UB) relies on indications in the User Requirements Document (URD) [RD-1], the Product Specifications Document (PSD) [RD-2] and the Data Access Requirements Document (DARD) [RD-3]. Advances, such as those described in the Product Validation and Algorithm Selection Report (PVASR) [RD-4], that may potentially be implemented in future revisions of the ATBD and in this document are described in the Algorithm Development Plan (ADP) [RD-6].

During Year 1 of the project, methods will be developed that lead to the generation of the global AGB product for the epoch 2017-2018. This document reports on the uncertainties of the datasets and models used to generate the AGB product for this epoch.

Section 2 provides the background of this E3UB, describing the strategy that underpins the algorithms implemented in CCI Biomass to estimate AGB. Section 3 describes the datasets (EO and auxiliary) used

to estimate AGB. The AGB retrieval methods used to generate a global map of AGB for the epoch 2017-2018 are presented in Section 4. An assessment of the retrieval uncertainty is presented in Section 5.

# 2 Background

Accuracy describes how well the estimate of a certain quantity (e.g., AGB) matches its true value. For an ensemble of data, two gross statistical measures of the accuracy of the estimator are commonly used: bias, which is the expected value of the difference between the estimated and true value, and a quantity indicating the variability of the estimate (standard error). More complete descriptors could include, for example, confidence Intervals on the estimates or the full error distribution.

In the case of AGB retrieval, the accuracy of the retrieved value depends on the accuracy of the input data and the accuracy of the estimation procedure. Figure 2-1 shows the flowchart of the CORE AGB estimation procedure implemented in year 1 of the CCI Biomass project to generate a global dataset of AGB estimates for the epoch 2017-2018 [RD-5].

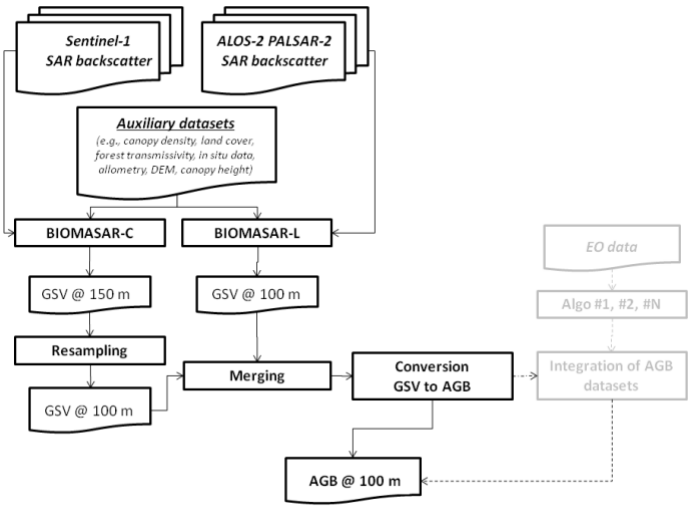


Figure 2-1: Functional dependencies of datasets and approaches forming the CCI Biomass CORE global biomass retrieval algorithm in year 1. The shaded part of the flowchart represents potential improvements following the implementation of additional retrieval techniques [RD-5].

Two independent estimates of growing stock volume (GSV, unit: $m^3$ $ha^{-1}$) are obtained from the BIOMASAR algorithm adapted to ingest C-band data (BIOMASAR-C) and L-band data (BIOMASAR-L). The estimates are combined to obtain a final estimate that should be characterized by smaller errors than the original values. Since the C- and L-band datasets have different pixel spacing, the GSV estimates from the BIOMASAR-C algorithm have slightly lower resolution so are resampled to the geometry of the BIOMASAR-L estimates. Finally, GSV is converted to AGB with a Biomass Conversion and Expansion Factor (BCEF).

# 3 Methods to assign accuracies

In this Section, we detail methods to quantify the standard error for each of the individual global inversion methods and for the final product. First, we summarize the CCI Biomass CORE algorithm

(Section 3.1). Then, the accuracy of the GSV estimates from BIOMASAR-C and BIOMASAR-L are presented (Sections 3.2 and 3.3, respectively). The methods presented reflect our current understanding of the standard errors embedded in the retrieval algorithms.

## 3.1 The GlobBiomass CORE algorithm

The BIOMASAR algorithm inverts a Water Cloud Model (WCM) to estimate GSV from a measurement of SAR backscatter.

$$\boldsymbol{\sigma^0_{for} = \sigma^0_{gr} e^{-\beta V} + \sigma^0_{veg}\left(1 - e^{-\beta V}\right)} \tag{3-1}$$

In Equation (3-1), $\sigma^0_{gr}$ and $\sigma^0_{veg}$ represent the backscattering coefficients of the ground and vegetation layer, and β is an empirically defined coefficient expressed in ha $m^{-3}$. The three parameters are unknown a priori and need to be estimated. The estimate of $\sigma^0_{gr}$ corresponds to the mean value of the SAR backscatter for pixels characterized by low canopy cover (so-called "ground" pixels) around the pixel of interest. The estimate of $\sigma^0_{veg}$ 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 of interest. This value, referred to as $\sigma^0_{df}$, is compensated for the residual ground contribution due to the canopy not being completely opaque. The estimation of $\sigma^0_{veg}$ from $\sigma^0_{df}$ and $\sigma^0_{gr}$ is done in slightly different ways depending whether GSV is estimated from C-band SAR (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 canopy height and canopy density of dense forests ($h_{df}$ and $\eta_{df}$, BIOMASAR-L) [RD-5].

Once the model parameters have been estimated, the inversion of the WCM in Equation (3-1) is straightforward.

$$\boldsymbol{\hat{V} = -\frac{1}{\beta} ln\left(\frac{\sigma^0_{meas} - \sigma^0_{veg}}{\sigma^0_{gr} - \sigma^0_{veg}}\right)} \tag{3-2}$$

Given N observations of the SAR backscatter acquired at different times, the corresponding N estimates of GSV can be combined with a weighted average to form a new estimate of GSV. The resulting estimate will have higher precision than any of the individual GSV estimates but may not be closer to the true GSV if the estimates are biased.

$$\boldsymbol{V_{mt} = \frac{\sum_{i=1}^{N} w_i \hat{V}_i}{\sum_{i=1}^{N} w_i}} \tag{3-3}$$

The weights, $w_i$, in Equation (3-3) are defined as the vegetation-to-ground backscatter difference in dB, $\sigma^0_{veg}$ - $\sigma^0_{gr}$, normalized by the maximum backscatter difference:

$$\boldsymbol{w_i = \frac{\sigma^0_{veg,i} - \sigma^0_{gr,i}}{max\left(\sigma^0_{veg,i} - \sigma^0_{gr,i}\right)}} \tag{3-4}$$

Merging of the BIOMASAR-C and BIOMASAR-L estimates of GSV is implemented in the form of a weighted average, where the weights account for three different calculated weights combined into one [RD-5].

$$GSV = w(L)V_{mt,L} + w(C)V_{mt,C} \tag{3-5}$$

Estimation of AGB is a simple scaling of GSV with the BCEF.

$$AGB = BCEF \cdot GSV \tag{3-6}$$

where BCEF is the product of wood density (g/cm$^3$) and total-to-stem biomass ratio.

## 3.2 Quantifying the accuracy of the BIOMASAR-C GSV estimates

In CCI Biomass, the BIOMASAR-C algorithm is applied to multi-temporal observations of the SAR backscatter from Sentinel-1 imagery acquired in 2017 [RD-5].

The standard error (i.e., the precision, of a GSV estimate obtained from a single observation of the SAR backscatter with the BIOMASAR-C approach) is quantified by propagating the standard error of (i) the measured SAR backscatter, $\sigma^0_{meas}$, and (ii) the estimates of the forest backscatter model parameters: $\sigma^0_{gr}$, $\sigma^0_{df}$, β and $V_{df}$. Since the standard errors of the five random variables listed above are uncorrelated and small (see below), the standard error of the estimate of GSV obtained from a single backscatter observation, $\delta V$, is given by:

$$\delta V = \sqrt{(\delta\sigma^0_{meas})^2 \cdot \left(\frac{\partial V}{\partial \sigma^0_{meas}}\right)^2 + (\delta\sigma^0_{gr})^2 \cdot \left(\frac{\partial V}{\partial \sigma^0_{gr}}\right)^2 + (\delta\sigma^0_{df})^2 \cdot \left(\frac{\partial V}{\partial \sigma^0_{df}}\right)^2 + (\delta\beta)^2 \cdot \left(\frac{\partial V}{\partial \beta}\right)^2 + (\delta V_{df})^2 \cdot \left(\frac{\partial V}{\partial V_{df}}\right)^2} \tag{3-7}$$

Each of the partial derivatives in Equation (3-7) is derived from Equation (3-1) (Annex A).

The accuracy of a backscatter measurement is affected by the radiometric and calibration accuracies, thermal noise and speckle. The SAR pre-processing also introduces additional uncertainty related to: (i) the accuracy of the geocoding transformation and resampling between radar and map geometries; (ii) the horizontal and vertical accuracy of the Digital Elevation Model (DEM) used as reference for the map geometry, and (iii) the accuracy of the pixel area and local incidence angle used to normalize the backscatter for slope-induced effects on the backscatter. Since the pixel-level uncertainties in the DEMs used in this study are unavailable, we cannot estimate the variance of a backscatter measurement from the individual variances of the terms listed above. We therefore estimate it empirically by equating it to the Equivalent Number of Looks (ENL). In the ATBD [RD-5], our preliminary estimate of the ENL was 162 (median value) with a span of [90, 375] but most values being between 100 and 250. Assuming a constant ENL of 162, we obtain a standard deviation of 0.32 dB. This value is here used to characterize $\delta\sigma^0_{meas}$.

The standard error of the model parameters expressing the backscatter from the ground, $\delta\sigma^0_{gr}$, and dense forest, $\delta\sigma^0_{df}$, combines the standard error in the observations of the pixels labelled as unvegetated or dense forest, respectively, and the standard error of the backscatter for the tree cover values representing the "ground" and "dense forest" classes. Here $\sigma^0_{gr}$ and $\sigma^0_{df}$ are estimated from the histograms of the SAR backscatter measurements for unvegetated and dense forest pixels (see

Figure 9 in Santoro et al., 2011). The uncertainties can be set equal to the standard deviation of each histogram minus the standard error associated with each backscatter measurement, i.e., $\delta\sigma^0_{meas}$.

The estimates of the coefficient β and the corresponding standard error $\delta\beta$ are defined in the ATBD of the GlobBiomass project (available at http://globbiomass.org/products/global-mapping/) [RD-8]. For each ecological zone of the FAO Global Ecological Zones (GEZ) dataset (see [RD-5] for details), the coefficient β was set equal to the value of the exponential model relating MODIS Vegetation Continuous Fields (VCF) estimates of canopy cover and GSV from map datasets. Accordingly, the standard error was defined as the mean standard deviation of the observations. It is important to remark that this is an educated guess due to the lack of observations of forest transmissivity from different biomes.

The standard error of the $V_{df}$ parameter, $\delta V_{df}$, was also discussed in the ATBD of the GlobBiomass project [RD-8]. A global raster was obtained with RandomForest (Breiman, 2001) models for each ecoregion of the FAO GEZ dataset using a set of initial estimates of $V_{df}$ from various sources [RD-5] as response and the WorldClim [RD-5] and ICESAT GLAS layers of forest height and density [RD-5] as predictors. Figure 3-1 illustrates the predictive performance of the models for each FAO ecoregion with the comparison of Out-Of-Bag model predictions (i.e., bootstrap aggregation of predictions from 500 regression tree models) versus the initial estimates for $V_{df}$ from the reference database. These values suggest that the standard error associated with RandomForest predictions for $V_{df}$ is of the order of 15 to 60% with the largest error of 40 to 60% for sub-tropical and tropical dry forests. This approximation might be too coarse but in our opinion is currently the best achievable estimate.

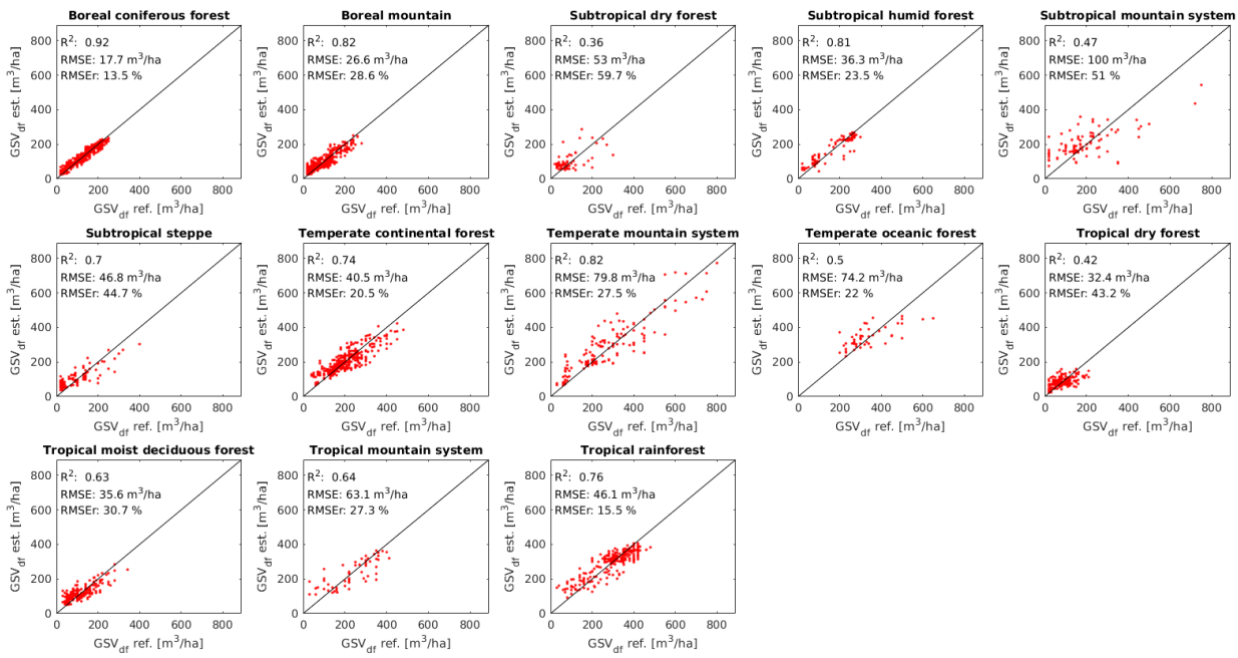


Figure 3-1: RandomForest predictions for the GSV of dense forests per 2°x2° grid cell derived from WorldClim and ICESAT GLAS versus estimates derived from inventory, provincial/state reports or existing maps.

The multi-temporal GSV estimate is obtained as a linear combination of GSV estimates from images acquired at different times. The scope of the multi-temporal combination is to reduce noise affecting each single estimate of GSV. Accordingly, the standard error of the multi-temporal estimate of GSV is

the result of a weighted average of standard errors, thus being smaller than the individual standard errors.

If the backscatter observations were independent of each other, the standard error of the final GSV estimate would correspond to the weighted sum of the standard errors of each individual GSV estimate. In reality, observations are correlated, so the standard error of the multi-temporal GSV estimates is the sum of a variance component and a covariance component that accounts for the correlation between errors.

$$\delta(V_{mt})^2 = \sum_{i=1}^{N} w_i^2\, \delta(V_i)^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w_i\, w_j\, Cov(V_i, V_j) \tag{3-8}$$

where

$$Cov(V_i, V_j) = \delta V_i \delta V_j r_{ij} \tag{3-9}$$

The variance component of the standard error of the multi-temporal GSV estimate is modelled as a linear combination of the single-image GSV variances given by Equation (3-3). Here, it is assumed that the multi-temporal weights are the best estimates of the individual variances of the individual estimates of GSV.

The covariance component is then expressed in a similar manner where individual error co-variances are weighted. The symbol $r_{ij}$ represents the correlation of errors between the estimates of GSV from image i and image j. Such correlation has been neglected in previous studies due to the weak correlation between observations at coarse spatial resolution. Envisat ASAR observations with a pixel size of 1,000 m repeat from the same orbit every 35 days; observations adjacent in time were taken from adjacent orbits and/or different nodes. From this weak correlation, it was assumed that also the correlation of the errors was small. In the case of Sentinel-1, this assumption has been revisited for two reasons. The higher spatial resolution compared to the Envisat ASAR dataset (100 m vs. 1,000 m) allows a more detailed spatial characterization of the biomass. The constellation, formed by the Sentinel-1A and -1B units, observes any point on the ground with the same viewing geometry every 6 days; this increases to 12 days for a single satellite. This increases the probability of correlation between retrieval errors.

Computing the correlation of errors requires a reference dataset. The only viable solution is to use maps of AGB, since plot measurements are typically too sparse to allow a spatially explicit characterization of the error covariance. Laser-based maps of AGB are probably the most suitable reference dataset for characterizing the error covariance. Although such maps are not free from errors, these are neglected here for simplicity. Obviously, maps lacking a complete characterization of errors and with low accuracy should be discarded, which poses a serious issue when attempting to generate wall-to-wall values of the error covariance. This is discussed later in this Section.

Figure 3-2 shows a matrix reporting at each bin the correlation of the errors between Sentinel-1 images i and j using as a reference the GSV predicted from country-wide laser scanning of Sweden (Nilsson et al., 2017). The index on each axis represents the date of acquisition in 2017 (i.e., 1, 2, 3 etc. mean first, second, third, etc.) image acquired in 2017. The correlations have been computed for all pixels within a 1° x 1° grid cell in North Sweden. The impact of the area selected to compute the error correlation is negligible. In Figure 3-2, we observe higher correlations towards the top-left and bottom-right corners, which represent data acquired during the first 3 months of the year (image index between 1 and 180) and the last 2 months of the year (image index 700 to 814; i.e., under winter/frozen

conditions). Similarly, somewhat high correlations can be seen at the top-right and bottom-left corner of the matrix, corresponding to combinations involving an image acquired at the beginning and end of 2017, with both characterized by winter/frozen conditions. Occasionally, high correlation is observed for some of the winter-summer pairs, whereas pairs formed by images acquired under unfrozen conditions were mostly characterized by lower correlation.

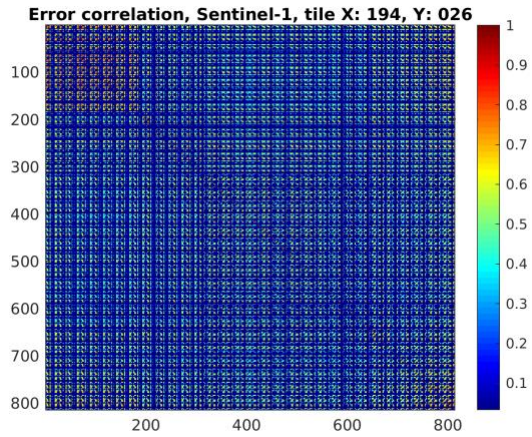


Figure 3-2: Matrix of error correlations for Sentinel-1 GSV estimates for a 1° x 1° area in northern Sweden (Lat: 63°-64°N; Lon: 14°-15°E, corresponding to tile X: 194 and Y: 026)

The histogram of correlations shows the distribution of the error correlation for all combinations of images acquired over the area visualized in Figure 3-3. The mean correlation coefficient was 0.53 with a span of 0.16 to 0.82.

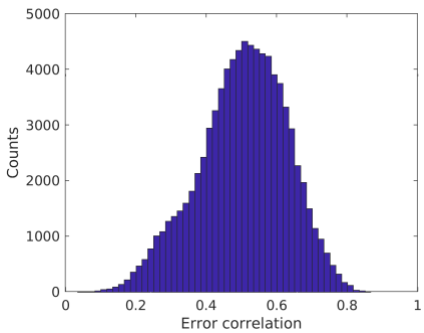


Figure 3-3: Histogram of error correlation for the dataset illustrated in Figure 3-2.

Since the computation of the error correlation requires a reference dataset, the question is whether a model or some general rules can be identified to characterize the correlation, given that no such dataset is available globally. Figure 3-4 and Figure 3-5 show two examples of correlation of errors between a given image and all other images for the 1° x 1° tile considered in Figure 3-2. Figure 3-4 is based on an image acquired under winter/frozen conditions and shows a slight seasonal trend of higher correlation when the other image was also acquired during winter. Figure 3-5 is instead based on an image acquired under summer/unfrozen conditions and shows fairly constant behaviour. These results indicate that a model expressing correlation as a function of time (e.g., exponential decay) does not apply. A model that replicates seasonality would be more suitable but difficult to implement as it

changes in space depending on local climatic conditions. Hence, in the first instance, one could consider using a simple generic value, such as the median of all correlations.

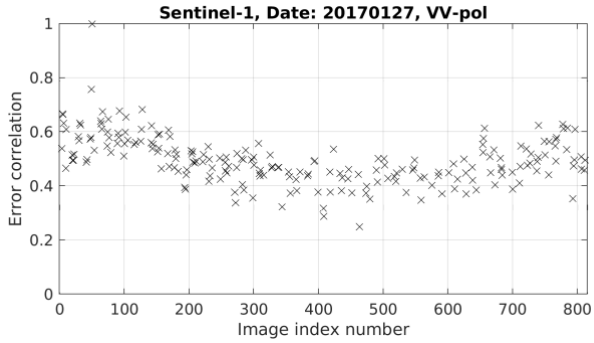


Figure 3-4: Error correlation between GSV estimated from a Sentinel-1 image acquired on 20170127 and all other Sentinel-1 images covering the 1° x 1° tile introduced in Figure 3-2.

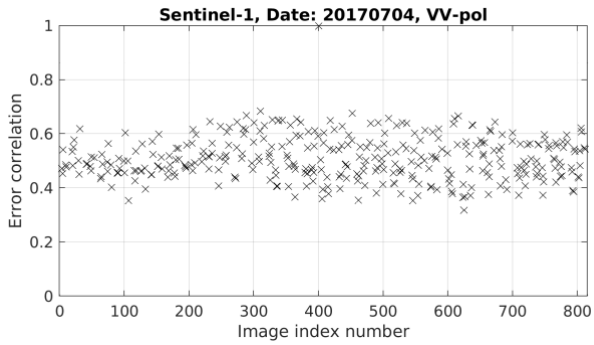


Figure 3-5: Error correlation between GSV estimated from a Sentinel-1 image acquired on 20170704 and all other Sentinel-1 images covering the 1° x 1° tile introduced in Figure 3-2.

The total standard error and its components (Equation 3-8) are illustrated in Figure 3-6 for the 1° x 1° tile in North Sweden (lower panels). We also include the map of GSV retrieved with the BIOMASAR-C algorithm and an image showing the number of Sentinel-1 backscatter observations used to retrieve GSV (upper panels). It is notable that, at this stage, GSV has been retrieved regardless of the land cover, (i.e., values of GSV have also been associated to water bodies and cropland). These areas need to be masked out for any further analysis of the data. Because of the very large number of observations used to retrieve GSV, the variance of the standard error is very small. The covariance term instead is large. In some areas the magnitude of this covariance term is comparable to the estimate of GSV, which seems to be related to the smaller number of observations used for the retrieval. Although this result is not entirely clear and deserves further attention, it is clear that the total standard error is represented basically by the covariance term.

The same analysis could have been repeated for other areas where laser-based maps of AGB are available so as to reinforce the findings obtained in a boreal environment. Nonetheless, since the computation of the error correlation matrix requires a reference dataset, a global data-driven characterization of this matrix is not possible. For this reason, we have computed the GSV standard error by assuming a constant error correlation, equal to the mean of the correlations shown in the histogram of Figure 3-3. The result is identical to the bottom-right panel in Figure 3-6 showing the covariance term of the standard error, suggesting that this simplifying assumption may be introduced to give a global characterization of the error covariance term.

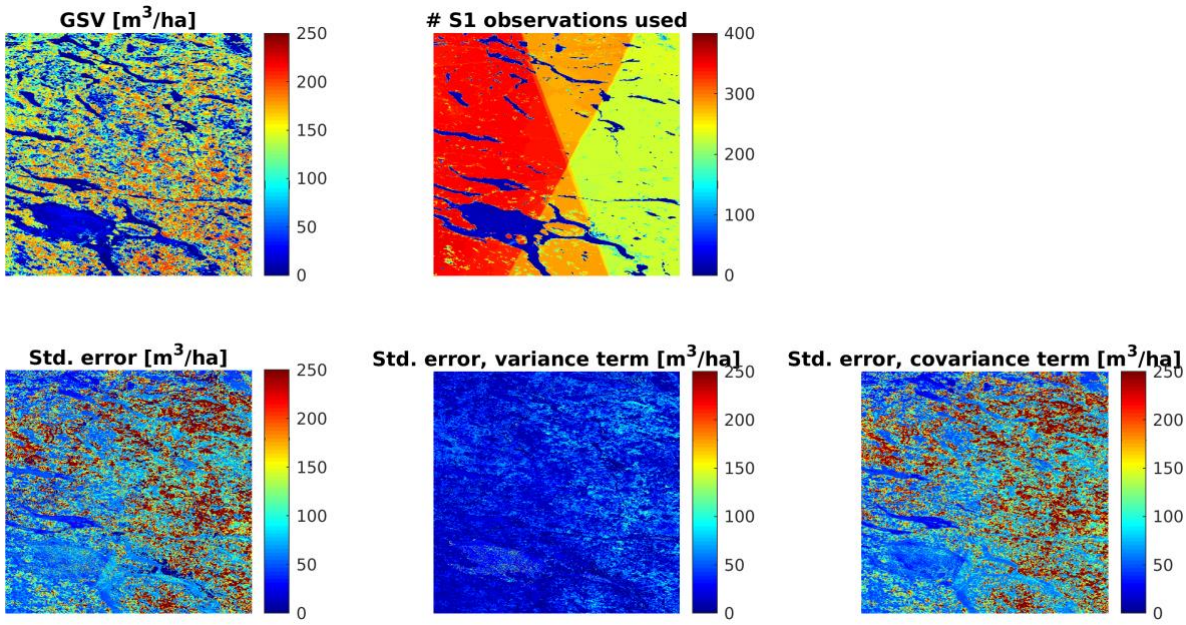


Figure 3-6: Comparing total standard errors using the correlation matrix shown in Figure 3-2 (left) and a constant correlation matrix with correlation equal to the mean of the error correlation in Figure 3-3.

## 3.3 Quantifying the accuracy of the BIOMASAR-L GSV estimates

In CCI Biomass, the BIOMASAR-L algorithm is applied to multi-temporal observations of the SAR backscatter from ALOS-2 PALSAR-2 imagery acquired in 2015, 2016 and 2017, and provided in the form of mosaics by the Japan Aerospace Exploration Agency (JAXA) [RD-5].

The BIOMASAR-C error model used to quantify the standard error of a GSV estimate, $\delta V$, from a backscatter observation is also applied in the case of BIOMASAR-L. However, some modifications are required because of the differences in the estimation of the parameter $\sigma^0_{veg}$.

In the model inverted for GSV, $\sigma^0_{veg}$ is expressed as a function of the average backscatter observed over dense forests, $\sigma^0_{df}$, the average canopy density of dense forests, $\eta_{df}$, the average height of dense forests, $h_{df}$, and the two-way attenuation coefficient, $\alpha$ [RD-5]. The error model therefore needs to be reformulated to consider the error associated with $\eta_{df}$, $\alpha$, and $h_{df}$:

$$\delta V = \sqrt{\begin{array}{c}(\delta\sigma^0_{meas})^2 \cdot \left(\frac{\partial V}{\partial \sigma^0_{meas}}\right)^2 + (\delta\sigma^0_{gr})^2 \cdot \left(\frac{\partial V}{\partial \sigma^0_{gr}}\right)^2 + (\delta\sigma^0_{df})^2 \cdot \left(\frac{\partial V}{\partial \sigma^0_{df}}\right)^2 + \\ (\delta\beta)^2 \cdot \left(\frac{\partial V}{\partial \beta}\right)^2 + +(\delta\eta_{df})^2 \cdot \left(\frac{\partial V}{\partial \eta_{df}}\right)^2 + (\delta\alpha)^2 \cdot \left(\frac{\partial V}{\partial \alpha}\right)^2 + (\delta h_{df})^2 \cdot \left(\frac{\partial V}{\partial h_{df}}\right)^2\end{array}} \quad (3\text{-}10)$$

As for BIOMASAR-C, the accuracy of the estimates of $\sigma^0_{df}$ and $\sigma^0_{gr}$ is estimated as the standard deviation of the histograms of backscatter observations in areas of low and high canopy density, since the histograms summarize the uncertainties associated with estimating the parameters due to spatially variable imaging conditions, uncompensated topographic effects, etc. (e.g., variable soil/canopy moisture).

Reported accuracies of height and canopy density estimates derived from ICESAT GLAS are used to determine the uncertainty associated with estimating $h_{df}$ and $\eta_{df}$. Following the results in Los et al. (2012) and Simard et al. (2011), who validated GLAS-based height estimates at boreal, temperate, sub-tropical, and tropical forest sites, we assumed standard errors for height estimates at the GLAS footprint level as between 4 m (boreal) and 10 m (tropics). While there are a large number of studies on the estimation of canopy cover and closely related variables, such as fractional cover, gap probability or transmittance from LiDAR, only a few have presented comprehensive validation. It is thus not possible to provide forest type-specific numbers for the error in the BIOMASAR-L retrieval associated with errors in the parameter $\eta_{df}$. As indicated by Garcia et al. (2012), the estimation of canopy cover from ICESAT GLAS with the ratio of energy returned from the canopy to the total energy returned may be of the order of 15 to 20%. We therefore assume a conservative global error of 20%. Note that the parameters $h_{df}$ and $\eta_{df}$ are estimated with the average GLAS height and density across all footprints covering an ALOS-2 1° × 1° tile that cover dense forest according to Landsat. It is therefore assumed that the standard error for the two parameters reduces as the square root of the number of GLAS footprints used in the estimation increases.

The accuracy of the forest transmissivity coefficient $\beta$ and the related two-way attenuation coefficient $\alpha$ are the most difficult to specify and it is only possible to provide a best guess, for BIOMASAR-C. In the case of $\beta$, the associated uncertainty may be inferred from the relationship between the forest transmissivity, simulated with the aid of GLAS height and optical canopy density estimates, and GSV (see Figure 3-7). The results presented in Figure 3-7 suggest that the uncertainty associated with the forest transmissivity parameter increases with increasing $\beta$. The 95% error bounds of the estimate for $\beta$ increased from +/-0.002 ha m$^{-3}$ in the case of low values of $\beta$ that are valid in boreal and subtropical dry forests to +/-0.007 ha m$^{-3}$ for the highest values of $\beta$ that are applied in the tropics. For the two-way attenuation coefficient $\alpha$, we assume a standard error of 0.25 dB m$^{-1}$, which is roughly consistent with the range of values reported in the literature (Ulaby et al., 1990; Chauhan et al., 1991; Shinohara et al., 1992; Sheen et al., 1994; Kurum et al., 2009; Praks et al., 2012).

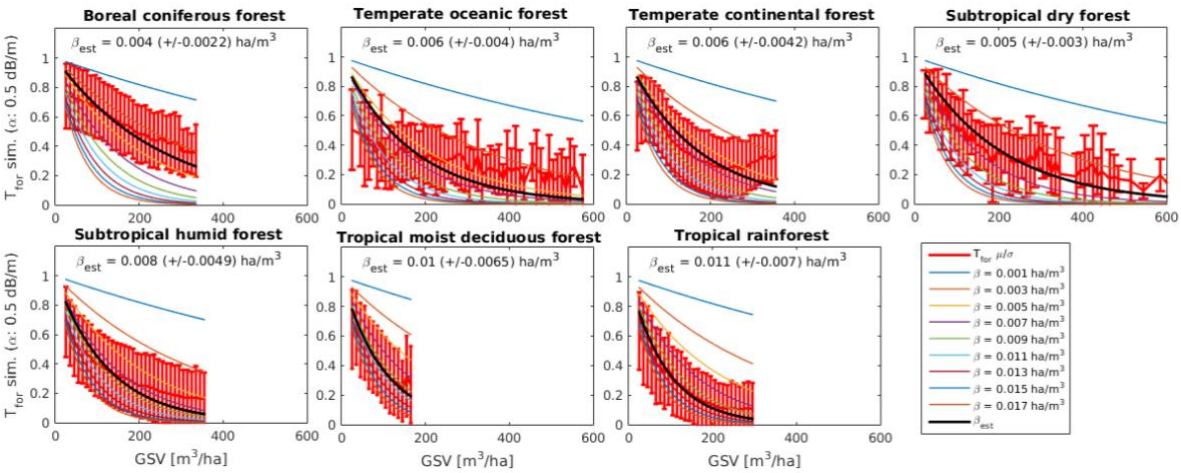


Figure 3-7: Estimates for β (incl. 95% confidence bounds) obtained for different FAO eco-regions. For details, refer to the ATBD of the GlobBiomass project [RD-8].

Equation (3-10) quantifies the standard error of GSV estimates derived from a single L-band observation, as in the case of BIOMASAR-L applied to one channel of the yearly JAXA L-band SAR backscatter mosaics (see [RD-5]). In a multi-temporal scenario (i.e., when multiple mosaics and/or Fine Beam Dual (FBD) and ScanSAR data are used (see [RD-5]), the standard error of the final GSV is

obtained, similarly to the case of BIOMASAR-C, as a weighted multi-temporal combination of single image standard errors and error covariance:

$$\delta(V_{mt})^2 = \sum_{i=1}^{N} w_i^2\, \delta(V_i)^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w_i\, w_j\, Cov(V_i, V_j) \qquad (3\text{-}11)$$

It the retrieval errors between single image GSV estimates are uncorrelated, the second term in Equation (3-11) becomes zero. Figure 3-8 shows the standard error for the BIOMASAR-L data product of GSV from ALOS-2 PALSAR-2 data assuming uncorrelated errors. For the map of GSV obtained with the BIOMASAR-L algorithm, refer to the ATBD of this project [RD-5].

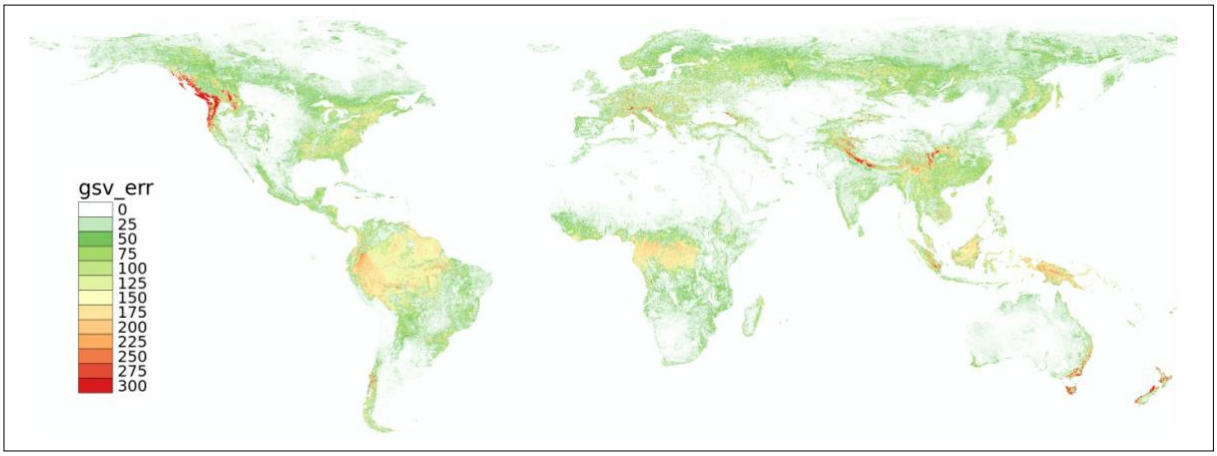


Figure 3-8: Standard error of BIOMASAR-L estimates of GSV assuming uncorrelated errors (i.e., first term of Equation (3-11)).

In practice, uncorrelated errors are unlikely, especially at low frequencies when the signal presents seasonal fluctuations and only minor date-to-date variability due to specific environmental conditions at the time of image acquisition.

In order to characterize the error correlation, we tested the use of airborne laser scanner (ALS) derived estimates of GSV/AGB as reference. For the AGB reported using laser-based maps, AGB was converted to GSV using the global database of BCEF estimates compiled in the frame of the GlobBiomass project [RD-8].

Figure 3-9 exemplifies the error correlation for a tropical forest site in Lope, Gabon, and a boreal forest site in Krycklan, Sweden. The laser-based maps of biomass used as reference were produced in the context of ESA's airborne radar campaigns AFRISAR and Biosar-2. In Lope, where FBD as well as ScanSAR data could be used to retrievie GSV, the error correlation was consistently at a level of 0.5 to 0.6. Similar results were observed for other forest sites in the tropics. In Krycklan, instead, where only the three ALOS-2 PALSAR-2 FBD mosaics from 2015, 2016, and 2017 were available and only the HV-polarized backscatter was used to retrieve GSV, the error correlation was close to zero. With a negligible covariance term and only three estimates to combine, the multi-temporal combination for areas where only FBD data are available hardly improves the standard error of the estimates.

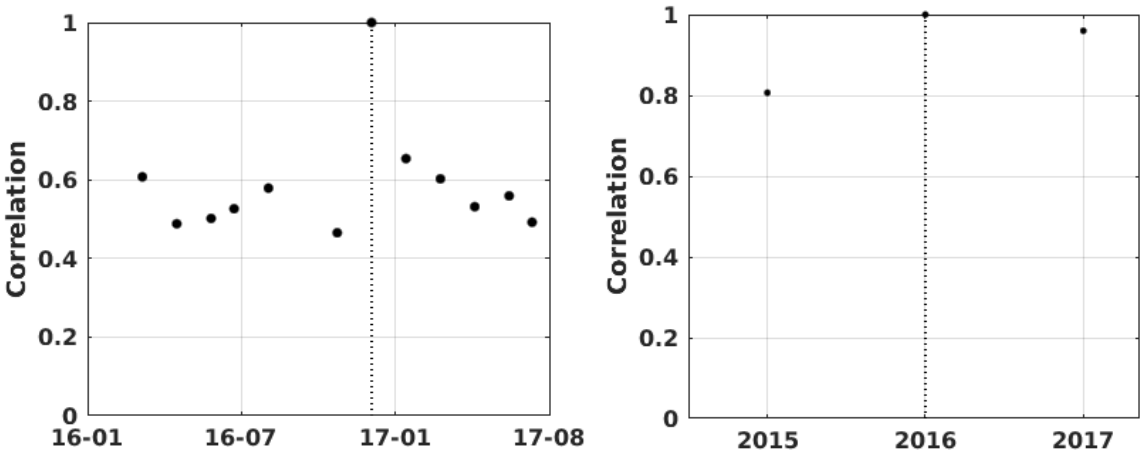


Figure 3-9: Correlation of GSV retrieval errors for a multi-temporal stack of L-band images acquired over Lope, Gabon (left) and Krycklan, Sweden (right). The vertical line denotes the reference image against which the correlation of retrieval errors with respect to the other available images was assessed.

The results obtained so far are considered to be a consequence of the fact that, over continuous tropical forest, the AGB sensitivity of L-band backscatter is low, so that different temporally uncorrelated noise sources in the measurements (thermal noise, speckle, quantization noise, etc.) have a more pronounced effect on the retrievals than in areas where the sensitivity to AGB is higher. In addition, it should be noted that the FBD mosaics were compiled with a preference for images acquired under dry summer conditions whereas the ScanSAR dataset comprises multi-seasonal observations (i.e., the measurements reflect a wider range of imaging conditions). As for BIOMASAR-C, further investigations are required to assess the representativeness of these results across different forest ecosystems. However, some simplifying assumptions will be needed to characterize the error correlation matrix because of the unavailability of a global reference dataset.

## 3.4 Quantifying the accuracy of the merged GSV estimates

When combining BIOMASAR-C and BIOMASAR-L derived GSV estimates, Equation (3-12) can be used to calculate the standard error of the merged product starting from the standard errors of the BIOMASAR-C and BIOMASAR-L estimates.

$$\delta(GSV)^2 = w^2(L)\delta(V_{mt,L})^2 + w^2(C)\delta(V_{mt,C})^2 \qquad (3\text{-}12)$$

At the time of writing, the standard error of C- and L-band derived estimates has not been assessed.

## 3.5 Quantifying the accuracy of the conversion from GSV to AGB

In the GlobBiomass project, a global raster of the BCEF was obtained by producing two independent raster datasets of wood density and total-to-stem biomass ratio. Each data layer was generated from extensive in situ databases in cooperation with the Max Planck Institute for Biogeochemistry (MPI-BGC). In the first year of the CCI Biomass project, the GlobBiomass datasets of wood density and biomass expansion are used; refer to the ATBD of the GlobBiomass project [RD-8] for details on how these and the corresponding accuracy were derived.

Since AGB is a product of GSV, wood density and biomass expansion factor (BEF), the accuracy of the AGB estimates is obtained as:

$$\delta AGB = \sqrt{\left(\frac{\partial AGB}{\partial WD}\right)^2 \cdot \delta WD^2 + \left(\frac{\partial AGB}{\partial BEF}\right)^2 \cdot \delta BEF^2 + \left(\frac{\partial AGB}{\partial GSV}\right)^2 \cdot \delta GSV^2} \qquad (3\text{-}13)$$

where $\delta WD$ and $\delta BEF$ represent the accuracy of the wood density and BEF terms respectively, and $\delta GSV$ represents the accuracy of the merged GSV dataset. The partial derivatives of Equation (3-13) are reported in Annex B. At time of writing, the standard error of AGB estimates has not been assessed.

# ANNEX A

Partial derivatives of Equation (3-7)

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

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

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

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

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

# ANNEX B

Partial derivatives of Equation (3-13)

$$\left(\frac{\partial AGB}{\partial WD}\right)_{BEF,GSV} = BEF \cdot GSV \tag{S6}$$

$$\left(\frac{\partial AGB}{\partial BEF}\right)_{WD,GSV} = WD \cdot GSV \tag{S7}$$

$$\left(\frac{\partial AGB}{\partial GSV}\right)_{WD,BEF} = WD \cdot BEF \tag{S8}$$

# 4 References

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

Chauhan, N. S., Lang, R. H. and Ranson, K. J. (1991). Radar modeling of a boreal forest. *IEEE Transactions on Geoscience and Remote Sensing,* 29 (4), 627-638.

Garcia, M., Popescu, S., Riano, D., Zhao, K., Neuenschwander, A., Agca, M. and Chuvieco, E. (2012). Characterization of canopy fuels using ICESAT/GLAS data. *Remote Sensing of Environment,* 123, 81-89.

Kurum, M., Lang, R. H., O'Neill, P. E., Joseph, A. T., Jackson, T. J. and Cosh, M. H. (2009). L-Band radar estimation of forest attenuation for active/passive soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing,* 47 (9), 3026-3040.

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.

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

Praks, J., Antropov, O. and Hallikainen, M. (2012). LIDAR-aided SAR interferometry studies in boreal forest: Scattering phase center and extinction coefficient at X- and L-band. *Remote Sensing,* 50 (10), 3831-3843.

Sheen, D. R., Malinas, N. P., Kletzli, D. W., Lewis, T. B. and Roman, J. F. (1994). Foliage transmission measurements using a ground-based ultrawide band (300-1300 MHz) SAR system. *IEEE Transactions on Geoscience and Remote Sensing,* 32 (1), 118-130.

Shinohara, H., Homma, T., Nohmi, H., Hirosawa, H. and Tagawa, T. (1992). Relation between L-band microwave penetration / backscattering characteristics and state of trees. In: *IGARSS'92*. City: IEEE Publications, Piscataway, NJ, 539-541.

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.

Ulaby, F. T., Whitt, M. W. and Dobson, M. C. (1990). Measuring the propagation properties of a forest canopy using a polarimetric scatterometer. *IEEE Transactions on Antennas and Propagation,* 38 (2), 251-258.