BIOMASS Above Ground Biomass Product ATBD

Ref: BIO-BPS-AGB-ATBD-ARE-024912
Issue: 3.1.4
Date: 02/04/2026

Document information

Title	BPS AGB ATBD - Algorithms Theoretical Baseline Document
Internal Ref.	
External Ref.	BIO-BPS-AGB-ATBD-ARE-024912
Date	02/04/2026
Issue	3.1.4
Pages	54

Recipients	
Cristiano Lopes	ESA – ESRIN
Michele Caccia	ESA – ESRIN
Muriel Pinheiro	ESA - ESRIN

Prepared by	
Maciej Soja	Wageningen University & Research
Lars Ulander	Chalmers University of Technology
Shaun Quegan	University of Sheffield
Paolo Mazzucchelli	Aresys
Francesco Banda	Aresys

Checked by	
Davide D’Aria	Aresys

Approved by	
Davide D’Aria	Aresys

Change history

Date	Issue	Change	Author
16/06/2023	3.0	First issue of the document	Maciej Soja Lars Ulander Shaun Quegan Paolo Mazzucchelli Francesco Banda
04/08/2023	3.0.1	Post-PDR3 issue of the document • BPSPDR3-61: added "no-selection" criterion in section 3.2 • BPSPDR3-29: updated section 4.1.4 • BPSPDR3-24, BPSPDR3-25: clarified section 3.4.4.2.2 • BPSPDR3-27: updated section 4 • BPSPDR3-32: added insights on second iteration • BPSPDR3-22: editorials corrected	Paolo Mazzucchelli
17/11/23	3.1	CDR3 issue of the document • BPSPDR3-22: remaining editorial corrected • BPSPDR3-23: updated section 3.2.1 • BPSPDR3-27: updated section 4 • BPSPDR3-28: updated section 4.1.4.3 • BPSPDR3-31, BPSPDR3-32: updated section 4.1.4.5	Paolo Mazzucchelli
06/09/2024	3.1.1	FAT3 issue of the document: • BPSCDR2-10: updated description of "real elevation" • BPSPDR3-63: updated input table section 4.1.2 • Added selection for combination of reference data at second iteration	Paolo Mazzucchelli
13/12/2024	3.1.2	Delta_FAT3 issue of the document: • Updated eq. 3.14, 3.15 • Updated section 3.5.5 • Updated section 4.1.4.3 (power law model) • Updated section 4.1.4.6 • Renamed "decimationFactor" to "upsamplingFactor"	Paolo Mazzucchelli
30/04/2025	3.1.3	BPS-FTO-BATCH4: • Separated calibration screen table in skp and flattening (section 3.3.2); expanded description in how to use them (section 3.3.4)	Paolo Mazzucchelli Emanuele Giorgi
02/04/2026	3.1.4	Post-launch issue of the document (associated to BPS V4.4.1)	Francesco Banda Emanuele Giorgi

Date	Issue	Change	Author
		• Updated Eq. 3.13 (Sigma naught normalization): removed the spatial resolution normalization factor to preserve the dynamic range of the estimation	

Table of contents

1. Introduction

1.1. Context

BIOMASS is ESA's 7th Earth Explorer mission. The purpose of the BIOMASS is to reduce the uncertainty in the worldwide spatial distribution and dynamics of forest biomass in order to improve current assessments and future projections of the global carbon cycle. This objective will be achieved by the implementation of a P-band SAR mission, providing global maps of forest biomass stocks, forest disturbance and growth.
The BIOMASS mission will also provide 3D views of forests through the "tomographic" phase and will also support additional science needs arising from the opportunity to explore the Earth for the first time with a P-band SAR system from space offering the possibility to "see" below the vegetated areas and in beneath bare soil or icy regions.

To achieve its challenging objectives, the BIOMASS mission requires a specific data processing strategy that combines different technologies at different processing level in order to generate the final Above Ground Biomass global map.

1.2. Scope of the document

The purpose of this document is to provide the "Above Ground Biomass Algorithms Theoretical Baseline Document (ATBD)" for the BIOMASS Processing Suite (BPS) project.
This document is organized as follows:

- Section 1: Overview of the document;
- Section 2: Brief introduction about BPS and Above Ground Biomass processing;
- Section 3: Description of L2a Above Ground Biomass algorithms;
- Section 4: Description of L2b/L3 Above Ground Biomass algorithms.

1.3. Acronyms

Acronym	Description
AD	Applicable Document
ADS	Annotation datasets
ACM	Average Covariance Matrix
AGB	Above-ground biomass
ARESYS	Advanced Remote Sensing and Systems
ATBD	Algorithm Theoretical Baseline Document
AUX	Auxiliary

Acronym	Description
BIODEMPP	Biomass DEM Prototype Processor
BPS	BIOMASS Processing Suite
CFI	Customer Furnished Items
CFM	Computed Forest Mask
CM	Covariance Matrix
CoSCS	Coregistered SCS
CP	co-polarization
DEM	Digital Elevation Model
DGG	Discrete Global Grid
DTM	Digital Terrain Model
ECSS	European Cooperation for Space Standardization
EO-CFI	Earth Observation Custom Furnished Items
ESA	European Space Agency
GPP	Ground Processor Prototype
GPS	Global Positioning System
FD	Forest Disturbance
FH	Forest Height
FNF	Forest/Non-Forest Mask
GN	ground cancellation (Ground Notching)
IERS	International Earth Rotation Service
InSAR	Interferometric Synthetic Aperture Radar
INT	Interferometric Phase
L0	Level-0 product
L1c	Level-1c product
L2a	Level-2a product
L2b	Level-2b product
L3	Level-3 product
LOS	Line Of Sight
LUT	Look-Up Table
PF	Processing Facility
PFD	Product Format Specification Document
RD	Reference Document
SAR	Synthetic Aperture Radar

Acronym	Description
SCS	Single-look Complex Slant-range
SKP	Sum of Kronecker Products
SNR	Signal-to-Noise Ratio
std	Standard deviation
TBD	To Be Defined
TBC	To Be Confirmed
TOM	Tomographic Phase
XML	eXtended Markup Language
XP	cross-polarization

1.4. Applicable documents

[AD1] BIOMASS Production Model, BIO-ESA-EOPG-EEGS-TN-0046, issue 3.5
[AD2] BIO-BPS-TN-PRODMOD, BIOMASS Production Model Technical Note, I/R 1/2/1
[AD3] BIO-BPS-L1-PFD, BIOMASS L1a,b,c products format specification, I/R 1/2/2

1.5. Reference documents

[RD1] BIO-BPS-L1-STACK-ATBD, BIOMASS BPS L1c Stack ATBD, I/R 1/1/2
[RD2] BIO-BPS-L1-SAR-ATBD, BIOMASS BPS L1 SAR ATBD, I/R 1/1/2
[RD3] BIOMASS DEM Product Processor Prototype, DEM Processor Algorithm Theoretical Baseline Document, BIODEMPP-CoSCS-DEM-01, Issue 4.0, June 07th 2021
[RD4] M. Mariotti d'Alessandro, S. Tebaldini, S. Quegan, M. J. Soja, L. M. H. Ulander and K. Scipal, "Interferometric Ground Cancellation for Above Ground Biomass Estimation," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 9, pp. 6410-6419, Sept. 2020
[RD5] Snowdon, Peter. "A ratio estimator for bias correction in logarithmic regressions." Canadian Journal of Forest Research 21.5 (1991): 720-724
[RD6] Díaz-Francés, Eloísa, and Francisco J. Rubio. "On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables." Statistical Papers 54 (2013): 309-323
[RD7] BIO-BPS- AUX-FMT, BPS Auxiliary Product Format, I/R 3/2
[RD8] BIO-BPS-PPD, BIOMASS BPS Product Performance Description, I/R 3/0
[RD9] L2-ATBD-ARE-011567, BIOMASS L2 Algorithm Theoretical Baseline Document, I/R 1/5
[RD10] BIO-BPS-IODD, BIOMASS BPS IODD, I/R 3/0/2

[RD11] Soja, Maciej J., et al. "Mapping above-ground biomass in tropical forests with ground-cancelled P-band SAR and limited reference data." Remote Sensing of Environment 253 (2021): 112153

2. BPS and AGB Processing Overview

2.1. BPS Overview

The BIOMASS Processing Suite (BPS) is in charge of processing BIOMASS Level-0 data up to Level-3, generating a wide set of products. Product generation scheme is depicted in the following two figures respectively for the TOM phase and INT phase. The processing foresees the following steps:

- **L1 processing**: the BIOMASS image formation step, taking as input 1 L0 slice and generating as output N>=1 L1 frames;
- **Stack Processing**: the coregistration and stacking processing step that will define the required 7-images Tomographic stacks or the 3-images Interferometric stacks. It includes stack phase calibration;
- **L2a Processing**: processing step for each stack that provides the inputs to the further products Tile Generation performed by the L2b processors. For one global cycle the number of stacks varies mainly with latitude, approximately from a minimum of two at the Equator (ascending/descending) to about six at higher latitudes (three ascending/descending pairs);
- **L2b Processing**: processing that aggregates all the stack-based L2a products insisting on the same location on ground (Tile) in the same time interval and generates a single BIOMASS L2b product. A different number of stacks may be available for each DGG tile [RD8], based on tile dimensions and swath coverage (which is not related to DGG tiling pattern);
- **L3 Processing**: consolidation processing removing discontinuities in L2b products.

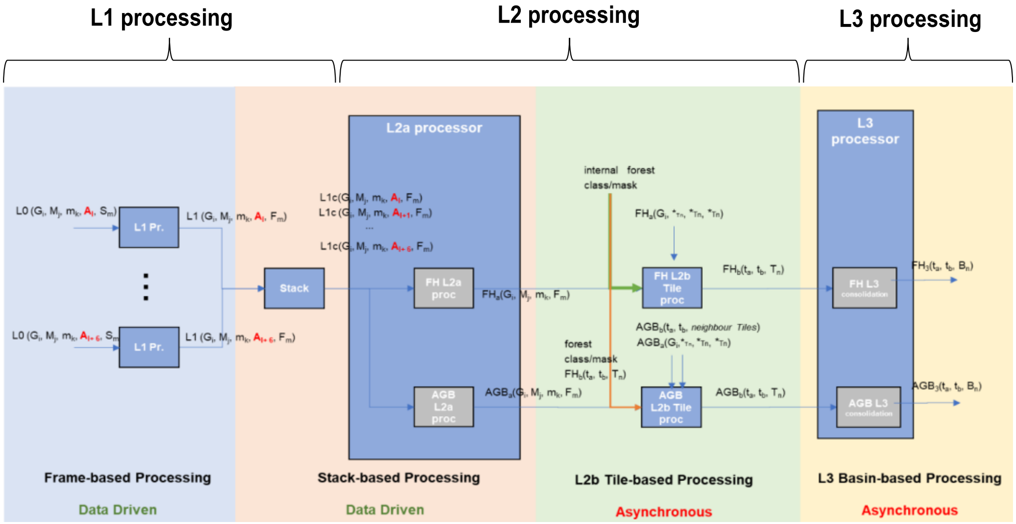

Fig.1 Biomass high-level processing scheme from L0 data to L3 products, TOM phase.

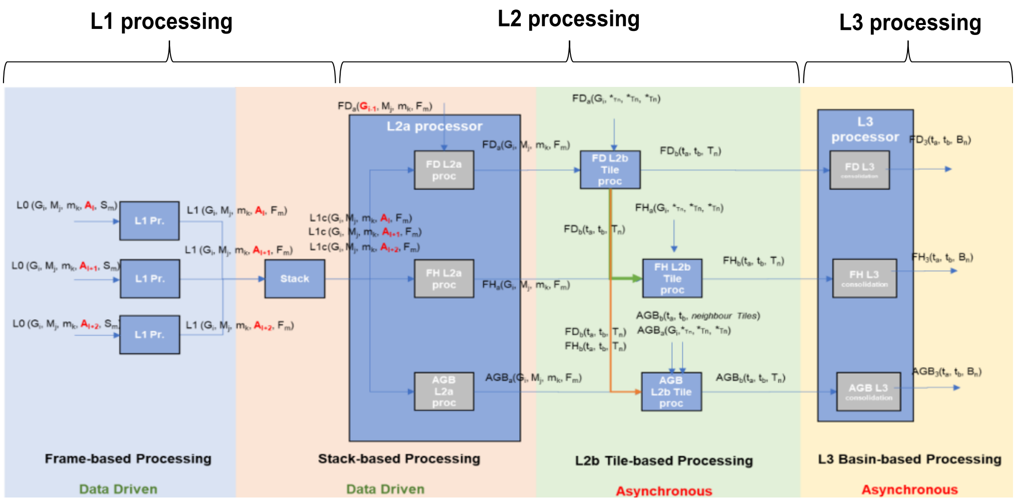

Fig.2 Biomass high-level processing scheme from L0 data to L3 products, INT phase.

2.2. AGB Processing Overview

The AGB processing has as main goal the generation of maps of above ground biomass density at a spatial resolution roughly equivalent to ~200 m x 200 m (i.e. 4 ha) at the Equator [AD2]. It receives as input a stack of BIOMASS Level-1c images (SCS co-registered and calibrated stack), including its relevant annotated data [AD3]. The AGB processing is divided into three levels:

- L2a, generates a ground cancelled GN product, corresponding to L1 acquisitions filtered in order to attenuate ground scattering and enhance canopy scattering; interferometric diversity is used to perform filtering, so that only polarimetric diversity remains after filtering;

- L2b, generates above ground biomass density map from all the L2a products on a connected group of DGG tiles, as well as a quality map, training on external reference AGB [RD10];
- L3, consolidates the result removing discontinuities in L2b products.

NOTE: Provided that FD algorithm proper validation will only be possible with BIOMASS in orbit, the strategy proposed to deal with FH/AGB dependency on FD is to initially input to FH/AGB a forest/non-forest mask generated outside of BPS [AD1][AD2]. After testing FD performance with BIOMASS in orbit, FD dependency can be switched on again, provided it has a satisfactory performance. The diagram of Fig.2 reports nonetheless FD input for completeness.

NOTE: L2 processor receives from Stack processor L1c products with two co-polarization (CP) channels and one cross-polarization (XP) channel. Details on the aggregation of XP channels into one XP are reported in [RD1].

NOTE: Regarding stack images cardinality, their number can vary (not always 3 or 7) due to contingency reasons and it can go up to 8 in TOM phase.

2.3. AGB Processing Workflow and Algorithmic Overview

The main steps of AGB processing workflow are depicted in the following figure and described in detail in the following sub-sections. This algorithm is a tailoring of the one described in [RD11].

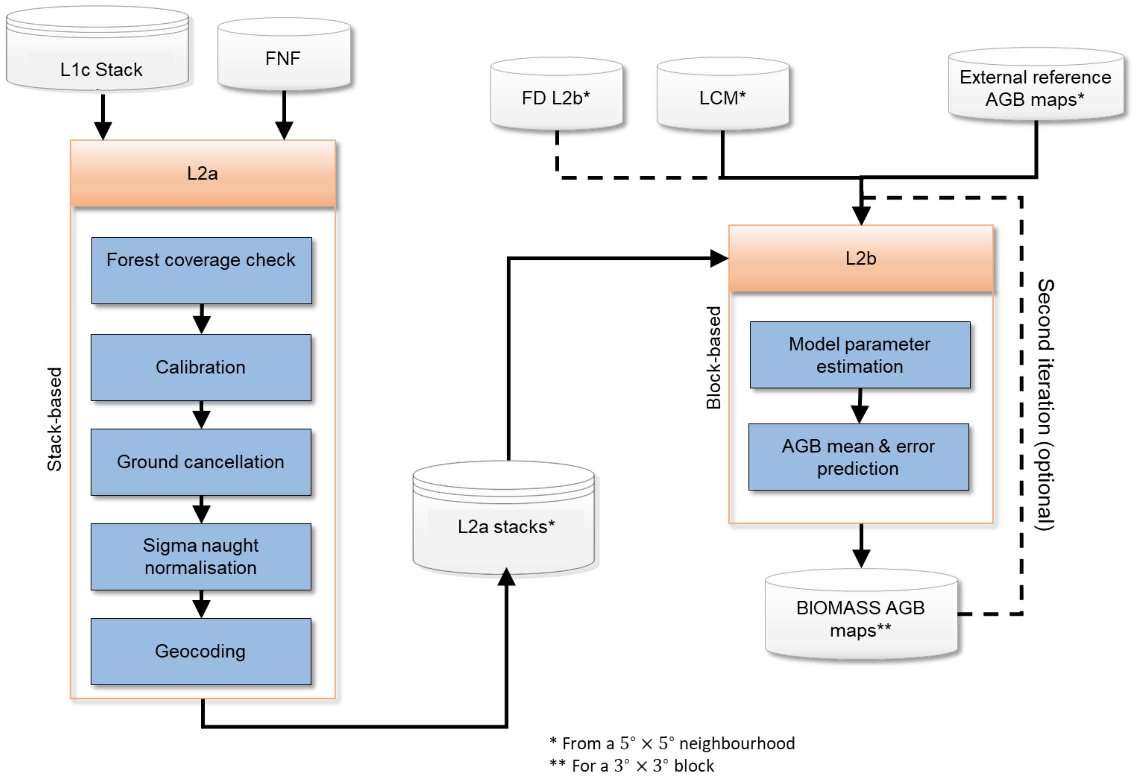

Fig.3 AGB processing detailed workflow. 5°x5° neighbourhood and 3°x3° processing block are shown in Fig.12.

2.3.1. Ground cancellation (L2A_P)

The main steps of the stack-based L2a block are:

- forest coverage check: predetermine which input data (as a whole image) should need to be processed;
- calibration: application of phase calibration and ground steering (terrain at zero elevation) screens to the stack;
- ground cancellation: removes the contribution of the ground scattering from the stack;
- sigma naught calibration: removes the first-order dependence on incidence angle and normalizes to resolution cell area;
- geocoding: projects the processed data on a geographic map, accounting for the different forest point locations with respect to the reference DEM.

2.3.2. Above-ground biomass density estimation (L2B_AGB_P)

The main steps of the L2b block are:

- selection of training data: checking for availability of training AGB data within the L2a stacks and selecting the appropriate source (either external reference AGB data or previous version of the BIOMASS AGB map);
- training data extraction: extracting data from the extent of a tile and its neighborhood, matching pixels with reference AGB and forest type information with pixels from L2a stacks;
- parameters estimation: estimating AGB model parameters;
- AGB mean and standard error estimation: extracting L2a stack data within the current tile and creating maps of AGB mean and standard error.

2.3.3. Note on interpretation of inputs and outputs tables

Inputs and outputs for each processing sub-step are summarized in specific subsections (two subsections for each substep). A specific tabular format is defined to improve the access to the document. The meaning of each column is exemplified in the following table.

Symbol	Name	Origin/ Destination	Size	Variability	Default	Description
Mathematical symbol of the quantity appearing in formulas.	Name of the quantity as related to other BPS documentation (e.g., AUX_FMT, L1 ATBDs, …)	BPS document name describing the origin of the quantity or processing step of origin/destination or BPS product of destination.	Dimensions of the quantity.	Range of values or update rate of an internal resource.	Default value for configuration parameters.	Verbal description of the quantity and its characteristics (e.g. measure units, …)

3. Ground cancellation

The overall scheme of the L2a processing is reported in Fig.3.
The key steps is ground cancellation. These steps are described in detail in the following subsections.

3.1. Forest Coverage Check

3.1.1. Overview and rationale

The availability of a general FNF map allows to predetermine which input data (as a whole image) should need to be processed. A threshold (*mF*) must be set as input parameter to select the minimum forest coverage (as a percentage of the whole area covered by the images) to start or skip computation.

The step is optional. If unset, change detection computation is always performed for all the data elements.

3.1.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
FNF	FNF	Internal resource (IODD)	$[N_{lat,FNF} \times N_{lon,FNF}]$	Yearly	-	A global forest-non-forest mask given as input
mF	forestCoverageThreshold	AUX_PP2_2A (AUX_FMT)	float value [scalar]	0.0-100.0	5.0	The minimum [%] relative forest coverage that triggers the computations
BB	Footprint	L1c main product header (L1_PFD)	lat-lon vertices [4 couples]	-	-	The extent of the L1c input data in geographical coordinates

3.1.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
TR	trigger	*Next steps*	Boolean [scalar]	-	True	Allows to trigger (True) the computation of

Symbol	Name	Destination	Size	Variability	Default	Description
						the change detection

3.1.4. Algorithm mathematical background

The values of the FNF mask that lie within the area covered by the image, defined by the bounding box *BB*, are extracted from the global FNF mask. The relative number of forested pixels is compared to the required minimum threshold *mF* to trigger any subsequent computation.

$$TR = \left(100 * \frac{ForestedArea|_{BB}}{TotalArea|_{BB}} > mF\right) \quad (3.1)$$

3.1.5. Notes on implementation

The substep should be as fast as possible, as its unique role is to avoid any un-useful computation. Thus, comparison is carried on at the original FNF mask spatial sampling (i.e. no reprojection needed), as any resampling does not have any impact on the relative percentage of forested area with respect to the total covered area.
It must be noted that any noValue pixel in FNF is treated as ForestedArea sample to avoid skipping computation because of the presence of noValues in FNF.
Details on the derivation of external FNF are reported in [RD10].

3.2. INT subsetting (TOM only)

3.2.1. Overview and rationale

The algorithm applies to TOM stacks and subsets the stack in order to select (or not) the acquisitions. The subsetting rule applicable at the moment follows a geometric criterion. Other rules could be defined post-launch. A **no-subsetting** rule is also available (“maintain all”, allowing to perform L2A_GN_P core algorithm on the full tomographic acquisition stack. **Note that this rule is only available for L2A_GN processing, and an error is generated if the rule is applied to other L2a processing submodules**.

This step is mandatory for TOM only.

3.2.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
SCS	L1c	Stack processor (L1_PFD)	M, each acquisition $[N_{rg} \times N_{az}]$	M=7/8 acquisitions (TOM phase), with two CP and one XP	-	L1c stack

Symbol	Name	Origin	Size	Variability	Default	Description
-	subsettingRule	AUX_PP2_2A (AUX_FMT)	String	-	"geometry"	Select 3 acquisitions from the 7/8 of TOM phase, choosing, with a geometrical rule, the baselines corresponding to the ones of INT phase. A no selection rule ("maintain all") is available for GN processing only

3.2.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
SCS	L1c	*Calibration step*	M, each acquisition $[N_{rg} \times N_{az}]$	M=3…8 acquisitions, with two CP and one XP	-	L1c stack

3.2.4. Algorithm mathematical background

3.2.4.1. Geometry criterion

The "geometry" criterion for subsetting selects among the seven/eight TOM stack the three acquisitions that give nominally consistent baselines with INT phase. As different combinations are possible, see Fig.4, the default rule is to select the central combination, which is corresponding to exactly the same nominal situation as INT. In case this combination presents a missing acquisition, the rule is to then evaluate remaining combinations in sequential order until a three-acquisitions combination is found. If no combination of three exists, the minimum required is two and the maximum baseline is equivalent to double TOM spacing. The contingency case with eight TOM acquisitions is reported in Fig.5.

```
                  0    1    2    3    4    5    6       TOM stack
Evaluation order  *    *    *    *    *    *    *
       2          *         *         *
       1               *         *         *           <-default
       3                    *         *         *
```

Fig.4 Selection for seven TOM.

```
                  0    1    2    3    4    5    6    7     TOM stack
Evaluation order  *    *    *    *    *    *    *    *
       2          *         *         *
       1               *         *         *               <-default
       3                    *         *         *
       4                         *         *         *
```

Fig.5 Selection for eight TOM (contingency).

3.3. Calibration

3.3.1. Overview and rationale

The calibration algorithm makes use of the L1c annotation [AD3] to apply phase calibration to L1c necessary for the L2a processing.

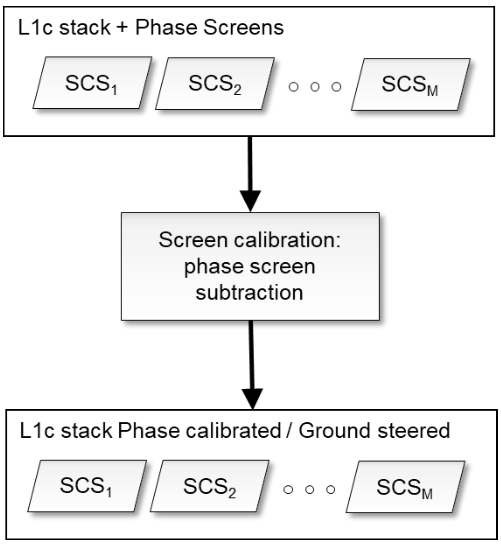

Fig.6 Calibration.

The step is mandatory. If disabled for debug, default rule is to copy inputs as step output.

3.3.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
SCS	L1c	Stack processor, or *INT subsetting step*	M, each acquisition $[N_{rg} \times N_{az}]$	M=3…8 acquisitions, with two CP and one XP	-	L1c stack
φ_{skp}	skpCalibrationPhasesScreen	L1c reference acquisition ADS (L1_PFD)	M, each $[N_{rg,\varphi} \times N_{az,\varphi}]$ (zero for reference acquisition)	-	-	SKP calibration phase screens [rad]
$\varphi_{flattening}$	flatteningPhasesScreen	L1c reference acquisition ADS (L1_PFD)	M, each $[N_{rg,\varphi} \times N_{az,\varphi}]$ (zero for reference acquisition)	-	-	Ground steering (flattening) geometry phase screens [rad]
-	applyCalibrationScreen	AUX_PP2_2A (AUX_FMT)	String	'skp', 'geometry', 'none'	skp	Choose the phase calibration to be performed: skp or geometry if phase screens calibration should be applied, none otherwise.

3.3.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
SCS	L1c	*Ground cancellation processing step*	M, each acquisition $[N_{rg} \times N_{az}]$	M=3…8 acquisitions, with two CP and one XP	-	Phase calibrated and ground steered (terrain at zero level) L1c stack

3.3.4. Algorithm mathematical background

This step aims to remove all the phase disturbances from the L1c stack and set terrain phase at zero height (i.e., subtract DTM phase). Each polarization channel is processed independently. The SCS L1c stack provided as input contains a set of calibration and ground steering phase screens (SKP calibration screens and flattening phases screens described in [RD1]). Depending on *applyCalibrationScreen* configuration, the used screen is:

$\varphi = \varphi_{skp} + \varphi_{flattening}$ if *applyCalibrationScreen= "skp"*

$\varphi = \varphi_{flattening}$ if *applyCalibrationScreen= "geometry"*

The operation implemented in the calibration step of the L2 processor is simply the subtraction of the selected calibration phase screens from the input image stack phase. The operation can be expressed as:

$$SCS(r, a, n) = SCS(r, a, n)e^{-j\varphi(r,a,n)} \quad (3.2)$$

where $\varphi(r,a,n)$ is the phase screen accounting for all the disturbance terms and terrain topography relative to the baseline n for each pixel in azimuth a, slant-range r. In the case of the reference acquisition, no compensation is done (zero baseline). If the phase screens have a lower resolution than the SCSs then a preliminary interpolation step is needed. The operation can be expressed as:

$$\varphi(r, a, n) = f\left[\varphi(r, a, n)|_{(rg,\varphi,az,\varphi)}\right]\Big|_{(rg,az)} \quad (3.3)$$

where *f* is a bilinear interpolant function.

3.4. Ground cancellation

3.4.1. Overview and rationale

Ground cancellation [RD4] is the interferometric combination of two or more images so that the backscattered echo coming from the ground level is cancelled. This processing step can be carried out starting from the calibrated and ground steered stack of SAR images (so that the phase of the ground contribution is aligned in the images). Consequently, the radar signal coming from a particular elevation gets emphasized. This specific elevation depends on the spatial baseline characterizing the interferometric system, depending on the geometry of acquisition, and cannot usually be chosen when processing the data. However, the availability of more than two interferometric images allows this specific elevation to be selected. This is referred to as interpolated ground cancellation. In case the quantity of interest is ground cancelled backscatter (intensity), filtering can be computed for all possible reference selection and intensity is finally averaged. This is called multi-reference ground cancellation, which is the default setting. The processing steps are described in the following, for two images and with several images.

The step is mandatory. If disabled for debug, default rule is to set I_{GN} (see section 3.4.3) to primary image acquisition.

3.4.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
SCS	L1c	*Calibration processing step*	M, each acquisition $[N_{rg} \times N_{az}]$	M=3…8 acquisitions, with two CP and one XP	-	Phase calibrated and ground steered (terrain at zero level) L1c stack
k_z	Wavenumber	L1c Stack ADS (L1_PFD)	M, each $[N_{rg,k} \times N_{az,k}]$ (zero for reference acquisition)	-	-	Vertical wavenumbers [1/m] computed with respect to incidence on inflated ellipsoid
z_0	emphasized ForestHeight	AUX_PP2_2A (AUX_FMT)	float value [scalar]	Global single value* for all polarizations and all pixels	30.0	Desired elevation emphasized from ground cancellation, meters [m]
-	disableGroundCancellationFlag	AUX_PP2_2A (AUX_FMT)	Boolean [scalar]	-	False	True to disable ground cancellation: for debug.
-	operationalMode	AUX_PP2_2A (AUX_FMT)	String	-	“multi reference”	Choose the Ground Cancellation method to use
-	imagesPair Selection	AUX_PP2_2A (AUX_FMT)	List of integers	-	-	Optional, needed only when operationMode=”insarPair” and more than 2 images are available

**NOTE: analysis can be carried out post-launch to assess whether this value can be made forest-class dependent.*

The default operation mode of ground cancellation is operationalMode="multi reference" with full INT stack [RD7]. To operate with a single pair configuration must be set to “insar pair” and the pair can be selected through imagesPairSelection. In case of only two images available, the operationalMode is automatically set to “insar pair”.

3.4.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
$\|I_{GN}\|^2$	-	*Sigma naught normalisation*	$[N_{rg} \times N_{az}]$	Two CP and one XP (independent from M)	-	Ground cancelled intensity image

3.4.4. Algorithm mathematical background

The flowcharts of the ground cancellation algorithm are shown in Fig.7, Fig.8. Each polarization channel is processed independently. A preliminary interpolation step is needed to bring wavenumbers onto the same L1c grid, in similar fashion as for the calibration screens (Section 3.3.4).

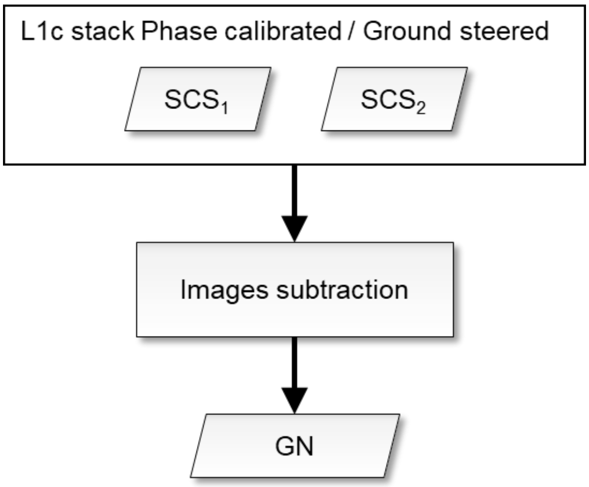

Fig.7 Ground cancellation algorithm with two images.

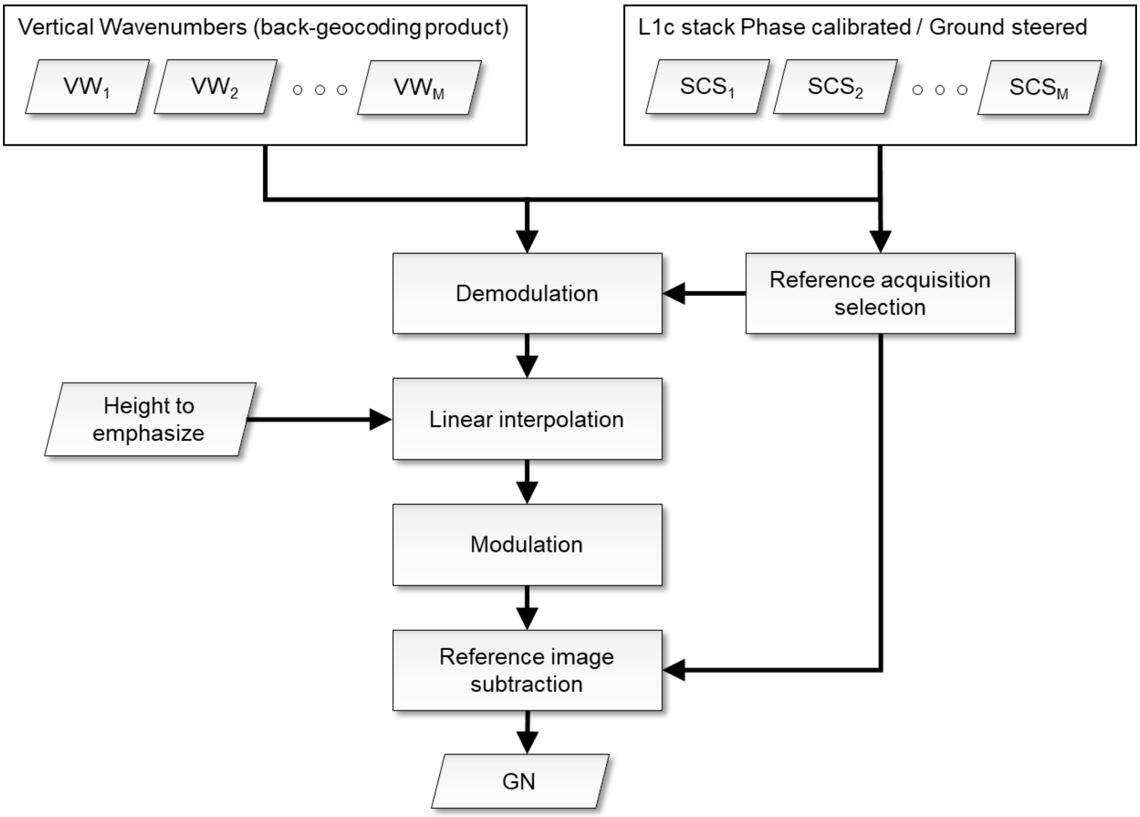

Fig.8 Ground cancellation algorithm with more than two images.

3.4.4.1. Case of two images

The processing for just two images is a simple subtraction of the second image from the first, as represented by

$$I_{GN}(r,a) = SCS(r,a,1) - SCS(r,a,2) \tag{3.4}$$

which holds for each polarization channel, for pixel in azimuth a, slant-range r. The first image is not required to be the reference acquisition, as symmetrical quantities, such as the covariance matrix or impulse response function of the ground cancellation (absolute value) are not affected by the order chosen for subtraction [RD4].

3.4.4.2. Interpolated ground cancellation single reference

For a stack of coherent ground steered SCS images (M>2) a further degree of freedom is available: a given elevation above the ground level can be chosen and emphasized together with removal of the ground echo. The desired elevation z_0 must be converted into a wavenumber before feeding it to the routine, using the relation:

$$k_{z_0} = \frac{\pi}{z_0} \tag{3.5}$$

A synthetic SCS associated with the desired wavenumber can be produced by linearly interpolating the stack. To ensure that the backscattering contribution associated with the desired elevation is not distorted by the interpolation process the stack is demodulated so that the desired elevation is steered around 0 m; the elevation to be demodulated coincides with half of the desired elevation. The demodulation is expressed by

$$I_{k_z}(r,a,n) = SCS(r,a,n)e^{-jk_z(r,a,n)\frac{z_0}{2}} \tag{3.6}$$

The linear interpolation algorithm must recognize the previous $k_z^-(r,a) = max\{k_z(r,a,n)\epsilon[k_z(r,a,n) < k_{z_0}]\}$ and the next $k_z^+(r,a) = min\{k_z(r,a,n)\epsilon[k_z(r,a,n) > k_{z_0}]\}$ sampled wavenumbers of the stack and then implement

$$I_{k_{z_0}}(r,a) = I_{k_z}^-(r,a) + [I_{k_z}^+(r,a) - I_{k_z}^-(r,a)]\frac{k_{z_0} - k_z^-(r,a)}{{k_z}^+(r,a) - k_z^-(r,a)} \tag{3.7}$$

where $I_{k_z}^-$ and $I_{k_z}^+$ are the SCS images associated with the wavenumbers k_Z^- and k_Z^+ respectively. After interpolation the spectrum is shifted back.

$$I_{k_{z_0}}(r,a) = I_{k_{z_0}}(r,a)e^{jk_{z_0}\frac{z_0}{2}} \tag{3.8}$$

The synthetic image (independent from M) is then subtracted from the reference image

$$I_{GN}(r,a) = I_{ref}(r,a) - I_{k_{z_0}}(r,a) \tag{3.9}$$

3.4.4.2.1. Reference acquisition selection

In this case the reference acquisition I_{ref} should be chosen as the one guaranteeing the widest scene coverage where interpolation can be performed, around the desired value k_{z_0}

$$I_{ref} = max_{SCS_n} \sum_{r,a} (k_{z_{min}}^{\ n}(r,a) < k_{z_0})(k_{z_0} < k_{z_{max}}^{\ n}(r,a)) \quad (3.10)$$

where $k_{z_{min}}^{\ n}$ and $k_{z_{max}}^{\ n}$ are the maximum and minimum wavenumbers of the set $k_z(r,a,n) - k_z(r,a,ref)$ assuming SCS_n as I_{ref}.

The reference acquisition selection is performed in a fully automatic way by the algorithm and requires no configuration inputs or others. The aim of this step is to set the reference acquisition to the one guaranteeing correct scene coverage and avoid gaps. The reference selected by this step is not necessarily corresponding to the original L1c reference of the stack. Anyway, the change of reference computed by this step is valid only for the purpose of ground cancellation computation and does not imply any change of reference outside the scope of this step. This step applies only to the interpolated single reference case, since in the two images case there is no need to specify a reference acquisition as described in sec. 3.4.4.1).

3.4.4.2.2. Multi-reference selection

In case the quantity of interest is ground cancelled backscatter (intensity), filtering is be computed for all possible reference selection and intensity is finally averaged.
Differently from what described in sec.3.4.4.2.1 this means the algorithm computes automatically as many ground cancellation results as the acquisitions in the stack, assuming each acquisition as a reference (denoted by *n* in following equations), following the same equations detailed in section 3.4.4.2. Results are finally averaged:

$$|I_{GN}(r,a)|^2 = I_{valid}(r,a) \sum_{n} \left| I_{ref}^{\ n}(r,a) - I_{k_{z_0}}^{\ n}(r,a) \right|^2 \quad (3.11)$$

where $I_{ref}^{\ n}$ and $I_{k_{z_0}}^{\ n}$ are reference and interpolated secondary assuming SCS_n as I_{ref} (which implies wavenumber rescaling as described in sec.3.4.4.2.1).
A weighting mask is computed, to account for each choice of reference for points that would need extrapolation, which are to be discarded and effective valid occurrences to be averaged

$$I_{valid}(r,a) = \frac{1}{\sum_n 1 - \{(k_{z_{min}}^{\ n}(r,a) > k_{z_0}) \vee (k_{z_0} > k_{z_{max}}^{\ n}(r,a))\}} \quad (3.12)$$

where $k_{z_{min}}^{\ n}$ and $k_{z_{max}}^{\ n}$ are the maximum and minimum wavenumbers of the set $k_z(r,a,n) - k_z(r,a,ref)$ assuming SCS_n as I_{ref}. Thus, only valid (i.e., non-extrapolated values) ground cancelled pixels are averaged (summed and scaled by the effective number of contributions as implied in eq. 3.11). If no valid ground cancelled results are available at a certain location, the corresponding pixel is set as invalid.

3.5. Sigma naught normalisation

3.5.1. Overview and rationale

The sigma naught normalization block aims to estimate the second order statistics of the ground cancelled intensity data and to normalize for some first-order incidence angle variability.

The step is mandatory. If disabled for debug, default rule is to skip correction.

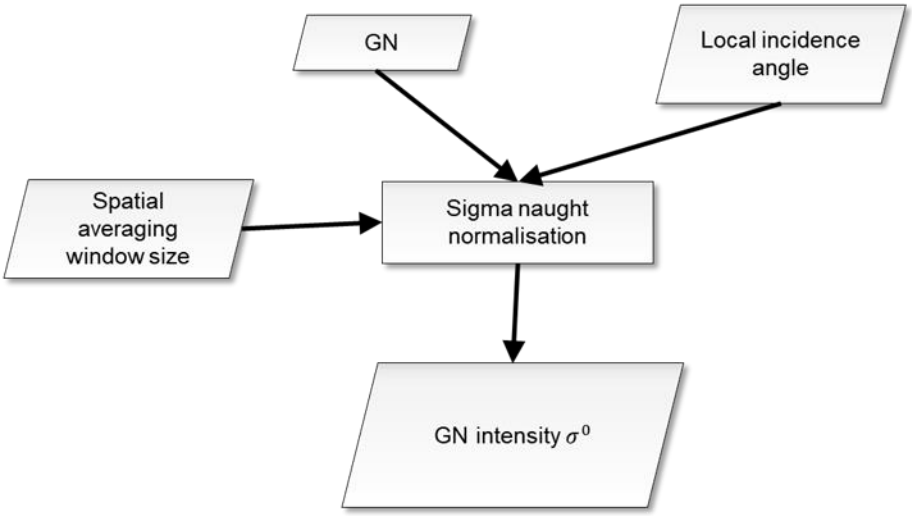

Fig.9 Sigma naught normalisation.

3.5.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
$\lvert I_{GN}\rvert^2$	-	*Ground cancellation processing step*	$[N_{rg} \times N_{az}]$	Two CP and one XP (independent from M acquisitions)	-	Ground cancelled intensity image
pRes	productResolution	AUX_PP2_2A (AUX_FMT)	float value [scalar]	-	200.0	Value in [m] to be used as the resolution on ground range map and also to perform the intensity averaging in radar coordinates
-	upsamplingFactor	AUX_PP2_2A (AUX_FMT)	int value [scalar]	-	2	upsamping factor, that modules the

Symbol	Name	Origin	Size	Variability	Default	Description
						overall decimation in multilooking
θ	incidenceAngle	L1c reference acquisition ADS (L1_PFD)	$[N_{rg,\theta} \times N_{az,\theta}]$	-	-	L1c stack reference acquisition incidence angle (on inflated ellipsoid) annotation [deg]
α	terrainSlope	L1c reference acquisition ADS (L1_PFD)	$[N_{rg,\alpha} \times N_{az,\alpha}]$	-	-	Terrain slope [deg]
B_r	L1c range-bandwidth	L1c reference acquisition ADS (L1_PFD)	-	-	-	L1c range bandwidth [Hz]
B_a	L1c azimuth bandwidth	L1c reference acquisition ADS (L1_PFD)	-	-	-	L1c range bandwidth [Hz]
v_a	L1c velocity annotation	L1c reference acquisition ADS (L1_PFD)	-	-	-	Average azimuth velocity [m/s]
rcFlag	radiometricCalibrationFlag	AUX_PP2_2A (AUX_FMT)	Boolean [scalar]	-	True	Flag to apply radiometric calibration

3.5.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
σ_{GN}	-	*Geocoding*	$[N_{pol} \times N_{rg,subs} \times N_{az,subs}]$	-	-	Ground-notched backscatter coefficient of current stack
ϑ	-	*Geocoding*	$[N_{rg,subs} \times N_{az,subs}]$	-	-	local incidence angle (wrt. ground surface normal) [rad]

3.5.4. Algorithm mathematical background

The sigma naught normalization block aims to estimate the intensity of the ground cancelled intensity data. The size of the spatial averaging window used to estimate the second order statistics depends on the final resolution of the products to be provided as an input. The output is a 3D array with dimensions: $[N_{pol} \times N_{rg,subs} \times N_{az,subs}]$ where N_{pol} the number of polarizations, and $N_{rg,subs}$ and $N_{az,subs}$ are the numbers of slant-range and azimuth samples after subsampling (half window size, well complying with the Nyquist-Shannon sampling theorem). For each (subsampled) range-azimuth position a N_{pol} vector proportional to sigma naught is returned. Mathematically:

$$\sigma_{GN_p}(r,a) = \frac{1}{N_{\Omega_r} N_{\Omega_a}} \sum_{r\in\Omega_r} \sum_{a\in\Omega_a} \sin[\theta(r,a) - \alpha(r,a)] \left|I_{GN_p}(r,a)\right|^2 \quad (3.13)$$

where p denote the p-th polarimetric channel; r and a denote the slant-range and azimuth indices; Ω_r and Ω_a denote the neighborhood in the slant-range and azimuth directions of the desired position; $I_{GN_p}(r,a)$ denotes the pixel value of ground cancelled complex image in polarimetric channel p; α indicate terrain slopes in range direction, θ the incidence angle, $\vartheta = \theta - \alpha$ the local incidence angle (wrt. ground surface normal). The correction term $\sin[\theta(r,a) - \alpha(r,a)]$ is applied only if *rcFlag* is set to True. A preliminary interpolation step is needed to bring angles onto the same L1c grid, in similar fashion as for the calibration screens (Section 3.3.4).

The averaging window size is computed as:

$$|\Omega_r| = \left\lceil \frac{pRes}{\rho_r} \sin\theta_m \, (\Delta_r B_r)^{-1} \right\rceil \quad (3.14)$$

$$|\Omega_a| = \left\lceil \frac{pRes}{\rho_a} (\Delta_a B_a)^{-1} \right\rceil$$

where $\rho_r = \frac{c}{2B_r}$ and $\rho_a = \frac{v_a}{B_a}$ denote the L1c resolution computed from system bandwidth, Lightspeed *c*, average azimuth velocity v_a, *pRes* is the L2 resolution, $\theta_m = \frac{1}{N_r N_a} \sum_{r,a} \theta(r,a)$ is the average incidence angle across the swath, while the terms $(\Delta_i B_i)^{-1}, i = r, a$ represent the oversampling factor. SCS invalid pixels will not be used during processing. The presence of an invalid SCS pixel causes the generation of an invalid covariance sample. The number of looks *L* [RD6] is computed as:

$$L = \left\lceil \frac{pRes^2}{\rho_r \rho_a} \sin\theta_m \right\rceil \quad (3.15)$$

3.5.5. Notes on implementation

It should be noted that, as implied by the equations above, multilooking is implicitly performed while computing vector proportional to sigma naught. The ratio of subsampling (i.e., the step where σ_{GN} full matrix is sampled) of the final output matrix is ruled by the “upsamplingFactor” parameter: when the assumed value is 1, full subsampling is applied ($N_{rg|az,subs} = N_{rg|az} / N_{\Omega r|az}$)

while when its value is higher, the final output matrix will be bigger ($N_{rg|az,subs} = N_{rg|az} / N_{\Omega r|az}$ * upsamplingFactor)

3.6. Geocoding

3.6.1. Overview and rationale

The purpose of this step is to project the processed data onto a geographic map, accounting for the different forest points location with respect to reference DEM.

NOTE: It should be noted that the use of the BIOMASS DTM [RD3] is to be preferred to other DEM since it represents the true terrain level at P-band, necessary for correct geocoding.

NOTE: this functionality is used by all three AGB, FD, FH processors with different I/O. mathematical formulation is described highlighting the mapping of the different processors I/O to geocoding formulas in each processor document. The functionality relies on L1c ADS [RD1] and not on the L1 geocoding tool [RD2].

The step is mandatory.

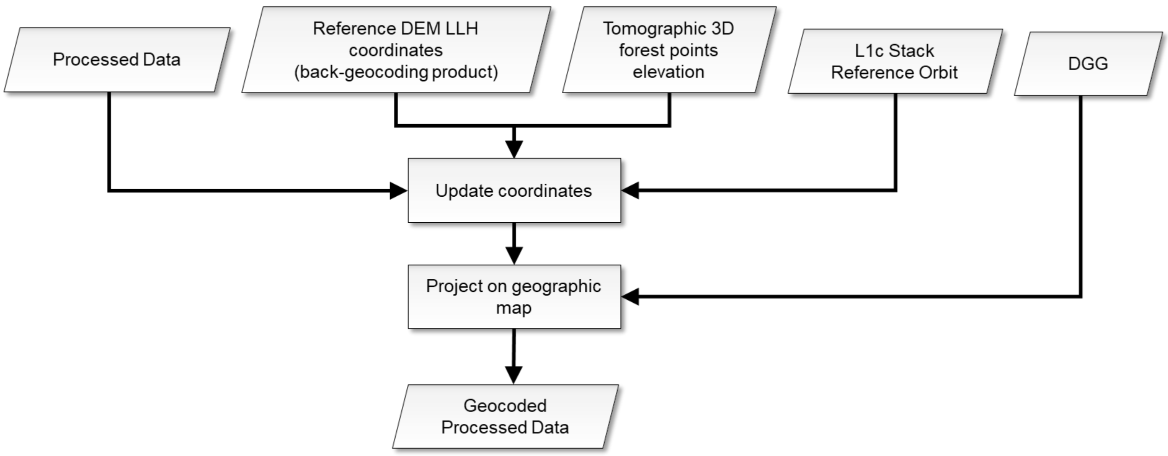

Fig.10 Geocoding.

3.6.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
$\sigma_{GN}, \vartheta \rightarrow Rd$	-	*Sigma naught normalisation processing step*	$[N_{pol} \times N_{rg,subs} \times N_{az,subs}]$, $[N_{rg,subs} \times N_{az,subs}]$	-	-	Processed data, in this case the ground-notched backscatter coefficient and local incidence angle of current stack

Symbol	Name	Origin	Size	Variability	Default	Description
(lat, lon, z_{P_0})	latitude longitude height	L1c reference acquisition ADS (L1_PFD)	3x $[N_{rg,\mathrm{llh}} \times N_{az,\mathrm{llh}}]$	-	-	Reference DEM latitude-longitude-height coordinates
$z_0 \rightarrow \delta z_{tomo}$	emphasizedForestHeight	AUX_PP2_2A (AUX_FMT)	float value [scalar]	Global single value (see previous note) for all polarizations and all pixels	30.0	Tomographic 3D forest points elevation [m], in this case the desired elevation emphasized from ground cancellation
DGG	*DGG*	Internal lookup table	-	-	-	DGG sampling
$\hat{\imath}_{az}$	Orbit ADS	L1c reference acquisition ADS	-	-	-	L1c stack reference acquisition orbit annotation

3.6.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
$Rd \rightarrow \boldsymbol{X}$	-	*FNF annotation processing step*	$[N_{pol} \times N_{lat} \times N_{lon}]$, $[N_{lat} \times N_{lon}]$	-	-	ground-notched backscatter coefficient and local incidence angle of current stack projected onto a geographic map

3.6.4. Algorithm mathematical background

3.6.4.1.1. Update reference DEM coordinates

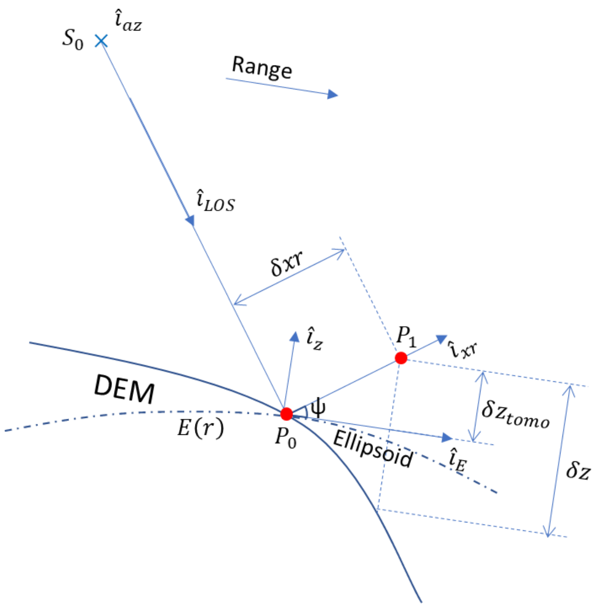

Fig.11 Determining the elevation of one forest point. The picture represents a lateral section for one sensor azimuth position. Details of the represented quantities are given in the text description.

This step aims to account for the different forest points location with respect to reference DEM. A preliminary interpolation step is needed to bring DEM latitude-longitude-height onto the same L1c grid, in similar fashion as for the calibration screens (Section 3.3.4). Determining the elevation of one forest point is shown in Fig.11:

- (lat, lon, z_{P_0}) is the reference DEM for the point $\boldsymbol{P}_0$, which can be converted to Cartesian coordinates $(X_{P_0}, Y_{P_0}, Z_{P_0})$
- z is normal direction to ellipsoid reported to DEM at point $\boldsymbol{P}_0$, i.e., given the tangent plane to ellipsoid a X + b Y + c Z + d = 0 then $\hat{\imath}_z = [a, b, c\,]$
- $\hat{\imath}_{az}$ is the azimuth sensor direction, determined from reference orbit annotation
- $\hat{\imath}_{LOS} = \frac{\boldsymbol{P}_0 - \boldsymbol{S}_0}{|\boldsymbol{P}_0 - \boldsymbol{S}_0|}$ is line of sight vector, $\boldsymbol{S}_0$ is the sensor position for the considered azimuth
- $\hat{\imath}_{xr} = \hat{\imath}_{az} \times \hat{\imath}_{LOS}$ is the cross range vector
- the ellipsoid direction at reference DEM is $\hat{\imath}_E \cdot \hat{\imath}_z = 0$, i.e., $\hat{\imath}_E = \frac{d[\boldsymbol{E}(r)]}{dr}$ and $\boldsymbol{E}(r) = [X_E(r), Y_E(r), Z_E(r)]$ are ellipsoid coordinates as a function of range
- δz_{tomo} is one tomographic forest point elevation, in this case the desired elevation emphasized from ground cancellation

- δz is the elevation of point P1 versus DEM (with respect to WGS84 geodetic system). Note that while this is represented for completeness in **Errore. L'origine riferimento non è stata trovata.**, it is not explicitly needed to compute P1 [x,y,z] position in the following eq 3.17.

Then

$$\delta xr = \frac{\delta z_{tomo}}{|\hat{\imath}_{xr} \times \hat{\imath}_E|} = \frac{\delta z_{tomo}}{sin(\psi)} \quad (3.16)$$

The geographic position of the forest point is:

$$\boldsymbol{P}_1 = \boldsymbol{P}_0 + \hat{\imath}_{xr} \cdot \delta xr \quad (3.17)$$

3.6.4.1.2. Projection onto a geographic map

Once the updated coordinates of the points are computed, interpolation of each raster layer to regular map coordinates projects the processed data onto a geographic map. In this case the information layers ground-notched backscatter coefficient and local incidence angle $Rd = \sigma_{GN}, \vartheta$ are both geocoded using the desired elevation emphasized from ground cancellation. Geocoding is done in a way that the products are sampled the same way onto the DGG sampling [RD8]. The operation can be expressed as:

$$Rd(lat, lon, z) = f\left[Rd(lat, lon, z)|_{\boldsymbol{P}_1}\right]\Big|_{DGG} \quad (3.18)$$

where *f* is a bilinear interpolant a function, *Rd* is each raster layer to be interpolated.

3.7. FNF annotation

3.7.1. Overview and rationale

The purpose of this step is to annotate the FNF in the L2a product.

The step is mandatory.

3.7.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
FNF	FNF	Internal resource (IODD)	$[N_{lat,FNF} \times N_{lon,FNF}]$	Yearly	-	A global forest-non-forest mask given as input
fMIT	forestMaskInterpolationThreshold	AUX_PP2_2A (AUX_FMT)	float value [scalar]	0.0-1.0	0.5	Threshold to fix rounding of pixels with decimal values originated from binary FNF

Symbol	Name	Origin	Size	Variability	Default	Description
						interpolation onto L2a grid
Rd	-	*Geocoding step*	$[N_{pol} \times N_{lat} \times N_{lon}]$, $[N_{lat} \times N_{lon}]$	-	-	ground-notched backscatter coefficient and local incidence angle of current stack projected onto a geographic map

3.7.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
Rd	-	GN L2a product of current cycle (AGB_PFD)	$[N_{pol} \times N_{lat} \times N_{lon}]$, $[N_{lat} \times N_{lon}]$	-	-	ground-notched backscatter coefficient and local incidence angle of current stack projected onto a geographic map
FNF	FNF	GN L2a product of current cycle (AGB_PFD)	$[N_{lat} \times N_{lon}]$	-	-	Annotated forest-non-forest mask

3.7.4. Algorithm mathematical background

The operation can be expressed as:

$$FNF(lat, lon) = f[FNF(lat, lon)|_{FNF}]|_{Rd} \tag{3.19}$$

where *f* is a bilinear interpolant a function. Since interpolation generates border effects a threshold *fMIT* to fix rounding of pixels with decimal values is applied:

$$FNF = FNF > fMIT \tag{3.20}$$

4. AGB estimation

The overall scheme of the L2b processing is reported in the figure below. The key step is the conversion of multi-stack, ground-notched backscatter coefficients to estimates of AGB mean and standard error. The associated steps are described in detail in the following subsections.

The world is divided into processing blocks covering all forested areas and not overlapping with each other (Fig.12, top). This step is run separately for each 3°×3° processing block, where the data for parameter estimation are extracted from a 5°×5° neighbourhood while the estimated AGB mean and standard error maps are only saved for the 3°×3° processing block: the neighbourhood consists of the processing block and includes a 1° buffer zone as shown in Fig.12, bottom.

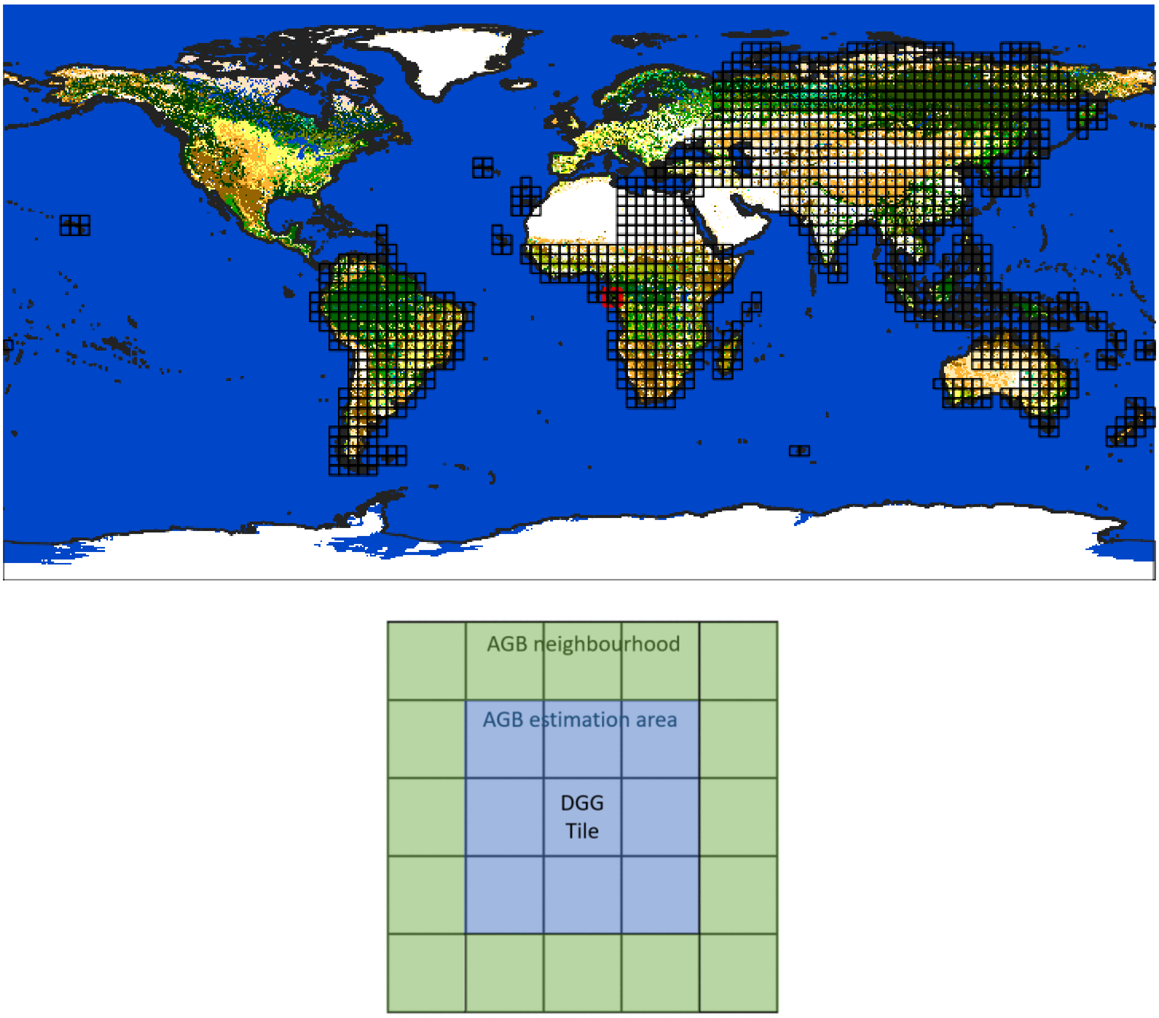

Fig.12 World covered by 3°x3° processing blocks (top) and 5°x5° neighbourhood for AGB estimation (bottom).

The algorithm performs up to two iterations for each processing block. The following approach to orchestration is followed:

1. perform first iteration for all blocks: during first iteration, only external reference AGB mean and standard error are present as input over all the 5°x5° neighbourhood.
    a. the invalid AGB estimates (where corresponding regression parameters cannot be estimated) are tagged in output at a pixel level **AND** a flag is raised.
2. perform second iteration only for all blocks where invalid AGB (but still recoverable) estimates are present (based on the flag).The second iteration uses the output from the first iteration as reference (as an addition to external AGB estimates). Only pixels tagged as invalid form first iteration are recomputed.

Fig.13 shows the flow chart for Level 2b processing for one processing block. Input reference AGB mean and std differs from first and second iteration:

- **first iteration:** external reference only
- **second iteration:** external reference + first iteration AGB output

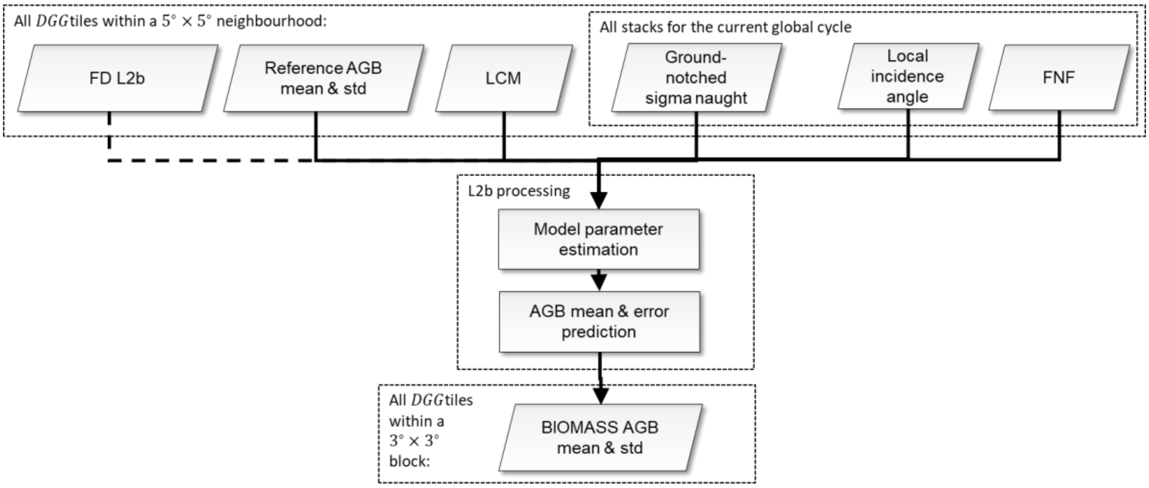

Fig.13 Complete flow-chart of the L2b processing.

While the tessellation of the world in 3°x3° processing blocks is proposed as a manageable way to assure the expected parameter variability, the algorithm formulation itself allows any possible definition of parameter estimation and AGB estimation manifolds. They both can be as small as 1°x1° tile only: thus, areas on continental borders / land sea interfaces are nicely and implicitly managed (as described in section 4.1.4.2).

4.1. Model parameter estimation

4.1.1. Overview and rationale

This step first checks the availability and values of reference AGB data, Level 2a stacks, and forest class map. Subsequently, model parameters are estimated using regression.

Fig.14 shows the flow chart for this module.

The step is mandatory.

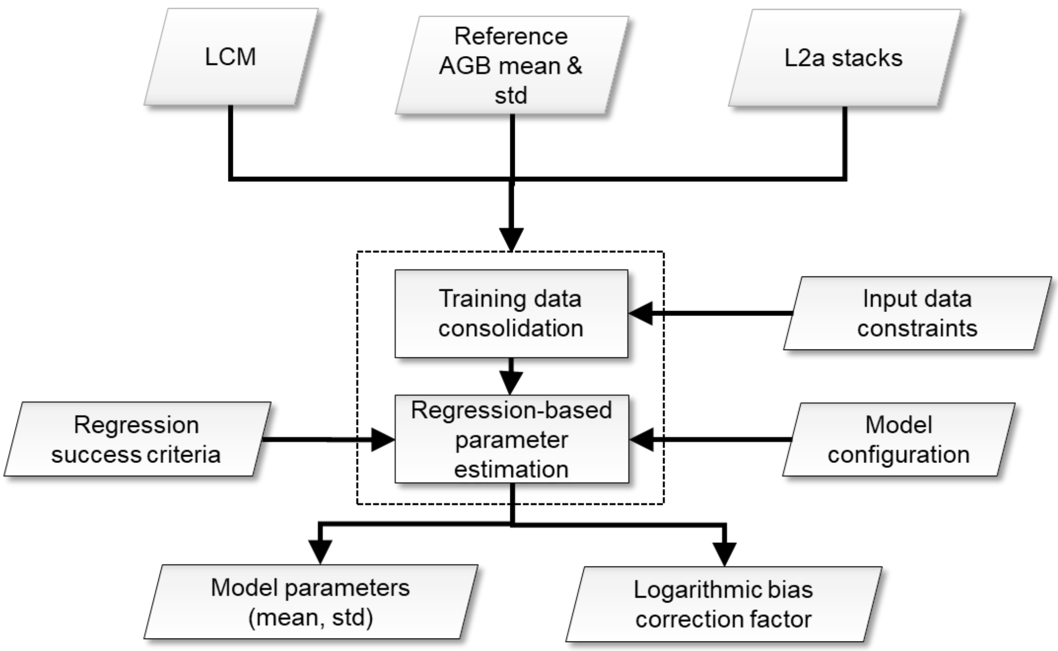

Fig.14 Complete flow-chart of the model parameter estimation module.

4.1.2. Inputs

Symbol	Name	Origin	Size	Variability	Default	Description
σ_{GN}	ground-notched backscatter coefficient	GN L2a products in a 5°×5° neighborhood in a global cycle (AGB_PFD)	$[N_{pol} \times N_{lat} \times N_{lon}]^{t}$	-	-	ground-notched backscatter coefficient, in linear units
ϑ	local incidence angle	GN L2a products in a 5°×5° neighborhood in a global cycle (AGB_PFD)	$[N_{lat} \times N_{lon}]^{t}$	-	-	local incidence angle [deg]
W_r	Reference AGB mean	Internal resource (IODD)	$[N_{lat,W} \times N_{lon,W}]^{t_W}$	-	-	A map of AGB mean used as reference [t/ha] in a 5°×5° neighborhood
U_r	Reference AGB std	Internal resource (IODD)	$[N_{lat,W} \times N_{lon,W}]^{t_W}$	-	-	A map of AGB std used as reference [t/ha] in a 5°×5° neighborhood
W_{fc}	"First cycle" AGB mean	AGB L2b product in a 5°×5° neighborhood in a global cycle (AGB_PFD)	$[N_{lat,W} \times N_{lon,W}]^{t_W}$	-	-	A map of AGB mean used as reference [t/ha] in a 5°×5° neighborhood
U_{fc}	"First cycle" AGB std	AGB L2b product in a 5°×5° neighborhood in a global cycle (AGB_PFD)	$[N_{lat,W} \times N_{lon,W}]^{t_W}$	-	-	A map of AGB std used as reference [t/ha] in a 5°×5° neighborhood
DGG	DGG	Internal lookup table	-	-	-	DGG tiles in a 5°×5°

Symbol	Name	Origin	Size	Variability	Default	Description
						neighborhood
FNF ^{t}	FNF	GN L2a products in a 5°×5° neighborhood in a global cycle (AGB_PFD)	$[N_{lat} \times N_{lon}]^{t}$	-	-	Annotated forest-non-forest mask
CFM	CFM	FD L2b product in a 5°×5° neighborhood in a global cycle (FD_PFD)	$[N_{lat,L2b} \times N_{lon,L2b}]^{t^{L2b}}$	optional	not provided	Computed Forest Mask
-	forestMaskingFlag	AUX_PP2_AGB (AUX_FMT)	bool	-	True	True if forest/non-forest map masking should be performed
BB ^{t}	Footprint	L2a main product header (AGB_PFD)	lat-lon vertices (4 couples)	-	-	The extent of the L2a input data in geographical coordinates in a 5°×5° neighborhood
mDGG	minimumL2aCoverage	AUX_PP2_AGB (AUX_FMT)	float value [scalar]	0.0-100.0	10.0	The minimum [%] tile coverage that triggers the computations
refSel	referenceSeelection	AUX_PP2_AGB (AUX_FMT)	String	‘refOnly’, ‘firstIterationOnly’, ‘weightedMean’	‘refOnly’	Combinatorial behaviour for multiple reference data at second iteration
LCM	Land Cover Map	Internal resource (IODD)	$[N_{lat,LCM} \times N_{lon,LCM}]$	Yearly	-	Landcover product classifying the type of land and forest

Symbol	Name	Origin	Size	Variability	Default	Description
Ω_{fcrej}	rejectedLandcover Classes	AUX_PP2_AGB (AUX_FMT)	List of integers	-	-	Set of landcover class indices to ignore. LCM description can be found in (IODD)
$\underline{\sigma}_p^0, \overline{\sigma}_p^0$	backscatterLimits	AUX_PP2_AGB (AUX_FMT)	N_{pol} ×2 float value [scalar]	$0 < \underline{\sigma}_p^0 < \overline{\sigma}_p^0 < \infty$	0.0001,100 for all polarisations	Lower and upper limits on backscatter in linear units
$\underline{\vartheta}, \overline{\vartheta}$	angleLimits	AUX_PP2_AGB (AUX_FMT)	2 float value [scalar]	$0 \le \underline{\vartheta} < \overline{\vartheta} \le \frac{\pi}{2}$	0,π/2	Lower and upper limits on the local incidence angle in radians
$\underline{W}, \overline{W}$	meanAGBLimits	AUX_PP2_AGB (AUX_FMT)	2 float value [scalar]	$0 \le \underline{W} < \overline{W} < \infty$	$10^{-3}, 10^3$	Lower and upper limits on the AGB mean in t/ha
$\underline{U}, \overline{U}$	stdAGBLimits	AUX_PP2_AGB (AUX_FMT)	2 float value [scalar]	$0 \le \underline{U} < \overline{U} < \infty$	$10^{-3}, 10^3$	Lower and upper limits on the AGB std in t/ha
$\underline{R}, \overline{R}$	relativeAGB Limits	AUX_PP2_AGB (AUX_FMT)	2 float value [scalar]	$0 \le \underline{R} < \overline{R} < \infty$	0,0.3	Lower and upper limits on AGB standard deviation relative AGB mean (coefficient of variability), in linear units
indexingL	indexingL	AUX_PP2_AGB (AUX_FMT)	String	'none', 'p', 'j', 'k', 'pj', 'pk', 'jk', 'pjk'	'pj'	l parameter variability with polarization 'p', date 'j', forest class 'k'. See section 4.1.4.3 for further details.

Symbol	Name	Origin	Size	Variability	Default	Description
indexingA	indexingA	AUX_PP2_AGB (AUX_FMT)	String	'none', 'p', 'j', 'k', 'pj', 'pk', 'jk', 'pjk'	'pk'	a parameter variability with polarization 'p', date 'j', forest class 'k'. See section 4.1.4.3 for further details.
indexingN	indexingN	AUX_PP2_AGB (AUX_FMT)	String	'none', 'p', 'j', 'k', 'pj', 'pk', 'jk', 'pjk'	'p'	n parameter variability with polarization 'p', date 'j', forest class 'k'. . See section 4.1.4.3 for further details.
useConstantN	useConstantN	AUX_PP2_AGB (AUX_FMT)	Boolean [scalar]	-	False	If True, indexingN=p is forced
n_p	valuesConstantN	AUX_PP2_AGB (AUX_FMT)	N_{pol} floats	(-∞,∞)	(1.0,1.0,1.0)	Values to use if useConstantN=True
mFV	minimumPercentageOfFillableVoids	AUX_PP2_AGB (AUX_FMT)	float value [scalar]	0.0-100.0	5.0	The minimum [%] of invalid pixels that triggers a new iteration
-	*regressionSolver*	AUX_PP2_AGB (AUX_FMT)	string	'double', 'float'	double	Numerical precision of the regression

4.1.3. Outputs

Symbol	Name	Destination	Size	Variability	Default	Description
$\widehat{W}$	Estimated AGB mean	AGB L2b product (AGB_PFD)	$[N_{lat,L2b} \times N_{lon,L2b}]^{t^{L2b}}$	-	-	AGB mean map [t/ha] for the 3°×3° processing block

Symbol	Name	Destination	Size	Variability	Default	Description
$\widehat{\sigma_W}$	Estimated AGB std	AGB L2b product (AGB_PFD)	$[N_{lat,L2b} \times N_{lon,L2b}]^{t^{L2b}}$	-	-	AGB std map [t/ha] for the 3°×3° processing block
rerun	*agbTileStatus*	AGB L2b product (AGB_PFD)	String	'Complete',, "NotComplete', 'Partial'	-	Flag raised when a new iteration is needed
HM	heat map	AGB L2b product (AGB_PFD)	$3[N_{lat,L2b} \times N_{lon,L2b}]^{t^{L2b}}$	0-	-	heat map (three values (1 float and 2 integers) per pixel) for the 3°×3° processing block
acquisitionIdImage	acquisitionIdImage	AGB L2b product (AGB_PFD)	$N_{L2a}[N_{lat,L2b} \times N_{lon,L2b}]^{t^{L2b}}$	0,1	-	Incidence matrix for the contributing data

4.1.4. Algorithm mathematical background

As a first step, the minimum required coverage over L2b DGG tiles in a 5°×5° neighbourhood *j* is checked to proceed with computations:

$$TR = \left(100 * \frac{A(\cup_i BB^i|_{DGG} \cap TT^j|_{DGG})}{A(TT^j|_{DGG})} > mDGG\right) \tag{4.1}$$

where BB^i are the L2a bounding boxes, TT^j is the footprint of the current L2b DGG tile collection and *A* operator computes the area of the corresponding polygon.

4.1.4.1. Definitions

- $j \in \{1, \dots, N_{dates}\}$ is an acquisition date index. Within the current 5°×5° neighbourhood, all stacks acquired from the same orbit and on the same date will have the same acquisition date index.
- $p \in \{1, \dots, N_{pol}\}$ is a polarisation index, with N_{pol} being the total number of polarisations
- $k = k(x, y) \in \{1, \dots, N_{fclass}\}$ is a forest class index. Note that the forest class is defined by the landcover class map LCM, so for each pixel (x, y), there is only one k
- x_0, y_0 denote longitude and latitude for the centre of the block

- $\Delta x_B, \Delta y_B,\ \Delta x_N, \Delta y_N$ denote block & neighbourhood width and height in degrees

4.1.4.2. Training data consolidation

This section describes the necessary steps to check the availability of input data and to define which pixels can be used for subsequent model parameter estimation.

The management and provision of 1° tiles and resources (stacks, auxiliary: AGB reference and **AGB estimates at previous iteration when available**) is dealt by the orchestrator and not described here [AD1]. A 3°×3° processing block will be formed by up to 9 tiles and a 5°×5° neighbourhood will be formed by up to 25 tiles. One acquisition is defined as one or more stacks from the same date and orbit. Throughout this part, pixels are identified via their (x, y) coordinates and SAR acquisitions are identified using index j, derived using the date-of-acquisition. In the following sections of the document, U and W are defined as the combination of W_r and W_{fc}, U_r and U_{fc}, following the selected behaviour described in the note at the end of this section.

Let Ω_B denote the set of pixel coordinates (x, y) within the current processing block:

$$\Omega_B = \left\{(x, y): |x - x_0| \leq \frac{\Delta x_B}{2} \wedge |y - y_0| \leq \frac{\Delta y_B}{2}\right\} \tag{4.2}$$

and Ω_N denote the set of pixel coordinates (x, y) within the current neighbourhood:

$$\Omega_N = \left\{(x, y): |x - x_0| \leq \frac{\Delta x_N}{2} \wedge |y - y_0| \leq \frac{\Delta y_N}{2}\right\} \tag{4.3}$$

The total SAR data coverage is the set of all pixels covered by at least one acquisition and fulfilling the requirements on backscatter and incidence angle:

$$\Omega_{SAR} = \Omega_1 \cup \ldots \cup \Omega_j \cup \ldots \cup \Omega_{N_{acq}} \tag{4.4}$$

where

$$\Omega_j = \left\{(x, y): \vartheta_j(x, y) \in \left[\underline{\vartheta}, \overline{\vartheta}\right] \wedge \left(\sigma^0_{pj}(x, y) \in \left[\underline{\sigma}^0_p, \overline{\sigma}^0_p\right] \forall p \in \left[1, \ldots, N_{pol}\right]\right)\right\} \tag{4.5}$$

The reference AGB data coverage is the set of pixels that contain reference AGB data and forest class fulfilling input data requirements:

$$\Omega_{AGB} = \left\{(x, y): W(x, y) \in \left[\underline{W}, \overline{W}\right] \wedge U(x, y) \in \left[\underline{U}, \overline{U}\right] \wedge R(x, y) \in \left[\underline{R}, \overline{R}\right] \wedge F(x, y) \notin \Omega_{fcrej}\right\} \tag{4.6}$$

where $R(x, y) = \frac{U(x,y)}{W(x,y)}$ is the AGB coefficient of variation (standard deviation relative mean) and Ω_{fcrej} is the set of landcover class indices to ignore. The set of landcover classes to ignore shall be made consistent with the input FNF information $\Omega_{fcrej} = \Omega_{fcrej} \cup \{1 - FNF^{\,t}\}$ for all t L2a stacks, or CFM in case this is provided $\Omega_{fcrej} = \Omega_{fcrej} \cup \{1 - CFM^{\,L2b}\}$ for all t^{L2b} L2b stacks.

From the four sets: Ω_B, Ω_N, Ω_{SAR}, and Ω_{AGB} we compute the following:

- Valid pixels with reference AGB data:

$$\Omega_{REF} = \Omega_{SAR} \cap \Omega_N \cap \Omega_{AGB}$$

- Pixels with SAR data but without AGB data (i.e., pixels with potential to-be-mapped):

$$\Omega_{-} = (\Omega_{SAR} \cap \Omega_B) \setminus \Omega_{AGB}$$

The algorithm continues to the parameter estimation sub-step if:

$$\Omega_{REF} \neq \emptyset$$

i.e., there are some training data available, and:

$$\Omega_{-} \neq \emptyset$$

i.e., there are some pixels where mapping is needed.

If either of these conditions is not fulfilled, the algorithm returns an error message stating that the mapping cannot be done, either due to the lack of training data ($\Omega_{REF} = \emptyset$) or the lack of need to map ($\Omega_{-} = \emptyset$).

The process of finding Ω_{SAR} presented above requires the algorithm to read entire images so that the requirements can be checked (e.g., $\sigma^0_{pj}(x,y) \in [\underline{\sigma}^0_p, \overline{\sigma}^0_p]$). This check is dealt in two steps: first using only the corner coordinates for each acquisition to find its extents, and then reading the entire images only for the acquisitions that cover the training data.

Finally, we define the set of all pixels for a given acquisition-forest class combination jk as:

$$\Omega_{jk} = \{(x,y) \in \Omega_j \cap \Omega_N \cap \Omega_{AGB} : F(x,y) = k\} \quad (4.7)$$

Such sets can belong to an area that covers only a subset of the original 3°×3° processing block 5°×5° neighbourhood: continental borders and sea/land interfaces are implicitly dealt within this algorithmic substep.

NOTE: when combining AGB previous estimates and external reference AGB estimates in the second iteration, the behaviour can be selected by user (with parameter 'refSel') as follows:

- **refOnly:** *reference estimates are preferred when both data are available at the same pixel (default behaviour)*
- **firstIterationOnly:** *AGB previously estimated values are preferred when both data are available at the same pixel*
- **weightedMean:** *a weighted mean (where weights are the inverse of AGB standard deviation U and U_{fc}) is computed when both data are available at the same pixel*

4.1.4.3. Regression-based parameter estimation

The actual number of regression parameters is ruled by the combination of user-selected parameter variability (by input flags indexingL, indexingA, indexing) **AND** the number of different forest classes, different acquisition dates and valid polarizations in the neighbourhood. Thus, it can differ for each processing block.
All possible choices for indexingL, indexingA, indexing are valid. However, they may lead to invalid regression (thus generating invalid AGB estimates), because of lack of input reference data. The default parameter variability is chosen as:

- backscatter intensity shall depend on specific acquisition conditions, thus the same parameter shall be used for same acquisition date (and polarization);

- AGB reference shall depend on forest class only (and polarization), as different forest classes depend on different allometries;
- incidence angle shall depend on polarization only.

As an example, a different parametrization where all parameters depend only on polarization can be chosen to limit the number of invalid AGB estimates (due to the lack of input data), if in the currently processed area the different forest classes are known to share similar allometries, and the acquisition conditions (i.e. meteorological conditions) are known to be similar even if acquisition dates are different.

The Power Law Model describes the dependence of radar observation σ^0_{pj} on AGB W and the local incidence angle θ. It requires three parameters for each unknown group: a scaling factor and two exponents, one for AGB *α* and one for the incidence angle normalisation *n.*

$$\sigma^0_{pj}(x,y) = A_{pj} \cdot W(x,y)^{a_{pk}} \cdot \cos\vartheta_j(x,y)^{n_p} \tag{4.8}$$

More insights can be found in [RD11]. The power law model solution can be expressed as an equivalent linear regression problem in log domain.Thus, as a default, the following forward model is used:

$$s_{pj}(x,y) = l_{pj} + a_{pk}w(x,y) + n_p c_j(x,y) \tag{4.9}$$

Where

$$\begin{aligned} s_{pj}(x,y) &= 10\lg\left[\sigma^0_{pj}(x,y)\right] \\ w(x,y) &= 10\lg[W(x,y)] \\ c_j(x,y) &= 10\lg\left[\cos\vartheta_j(x,y)\right] \end{aligned} \tag{4.10}$$

Based on this, we define a measurement vector as a complete set of input information for one backscatter value, including coordinates, indices for polarisation, acquisition, forest class, AGB mean and incidence angle data. We formalise it as:

$$\mathbf{b}_{pj}(x,y) = [x \quad y \quad p \quad j \quad k(x,y) \quad w(x,y) \quad c_j(x,y) \quad s_{pj}(x,y)] \tag{4.11}$$

We define the complete set of measurement vectors available in the data as:

$$\mathcal{B} = \left\{\mathbf{b}_{pj}(x,y)\ \forall(x,y) \in \Omega_{jk}, \forall k \in \{1,\dots,N_{fclass}\}, \forall p \in \{1,\dots,N_{pol}\}, \forall j \in \{1,\dots,N_{date}\}\right\} \tag{4.12}$$

Where

$$N_{meas} = n(\mathcal{B}) \tag{4.13}$$

We order the elements of $\mathcal{B}$ in an arbitrary order and then we assign an individual index $r \in \{1,\dots,N_{meas}\}$ to each element. The rth element is defined as:

$$\mathbf{b}_r = [x_r \quad y_r \quad p_r \quad j_r \quad k_r \quad w_r \quad c_r \quad s_r] \tag{4.14}$$

We define the training data table as:

$$[\boldsymbol{T}] = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_r \\ \vdots \\ \mathbf{b}_{N_{meas}} \end{bmatrix} = [\mathbf{x} \quad \mathbf{y} \quad \mathbf{p} \quad \mathbf{j} \quad \mathbf{k} \quad \mathbf{w} \quad \boldsymbol{c} \quad \mathbf{s}] \tag{4.15}$$

Where **x,y,p,j,k,w,c** and **s** are vectors containing specific data for all measurements. These data are, in respective order: longitude, latitude, polarisation index, acquisition date index, forest class index, mean reference AGB (in dB), cosine of the local incidence angle (in dB), and the vegetation backscatter coefficient (in dB).

Fitting of the forward model is solved through ordinary least squares and the following equation:

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{s} \tag{4.16}$$

where $\mathbf{A}$ is the design matrix and $\boldsymbol{\beta}$ is the vector of unknown parameters.

The design matrix $\mathbf{A}$ consists of three sub-matrices:

$$\mathbf{A} = [\boldsymbol{A}_l \quad \boldsymbol{A}_a \quad \boldsymbol{A}_n] \tag{4.17}$$

and $\boldsymbol{\beta}$ consists of three sub-vectors:

$$\boldsymbol{\beta} = \begin{bmatrix} \mathbf{l} \\ \mathbf{a} \\ \mathbf{n} \end{bmatrix} \tag{4.18}$$

The following describes the generation of the three sub-matrices of $\mathbf{A}$, and the meaning of the three sub-vectors of $\boldsymbol{\beta}$.

- $\mathbf{A}_l$ is a $N_{meas} \times N_l$ design matrix for parameter l and $\mathbf{l} = [l_1 \quad \cdots \quad l_{N_l}]^T$ is an N_l-element vector of its instances, i.e., different values to-be-estimated with least squares and corresponding to each unique combination of p and j. Each row of $\boldsymbol{A}_l$ corresponds to a unique measurement and each column to a unique instance for parameter l. With N_{pol} polarisations and N_{date} acquisitions,
$$N_l = N_{pol} N_{date}$$
can be used (all acquisitions feature all polarisations).

  For each unique combination of j and p, the corresponding instance index c_l (i.e., column in $\mathbf{A}_l$) can be created using:
$$c_l(j, p) = p + N_{pol}(j - 1) \in [1, \ldots, N_l]$$
($\max(p) = N_{pol}$). The element of $\mathbf{A}_l$ found at row r, column c is consequently:
$$A_l^{(r,c)} = \begin{cases} 1 & c = c_l(j_r, p_r) \\ 0 & \text{otherwise} \end{cases}$$
where j_r, p_r are the acquisition and polarisation indices for measurement r.
- $\mathbf{A}_a$ is a $N_{meas} \times N_a$ design matrix for parameter a and $\mathbf{a} = [a_1 \quad \cdots \quad a_{N_l}]^T$ is an N_a-element vector of its instances, i.e., different values to-be-estimated with least squares and corresponding to each unique combination of p and k. Each row of $\boldsymbol{A}_a$

corresponds to a unique measurement and each column to a unique instance for parameter a. With N_{pol} polarisations and N_{fclass} forest classes, $N_a = N_{pol} N_{fclass}$

For each unique combination of k and p, the corresponding instance index c_a (i.e., column in $\mathbf{A}_a$) can be created using: $c_a(k,p) = p + N_{pol}(k-1) \in [1, \ldots, N_a]$ $(\max(p) = N_{pol})$. The element of $\mathbf{A}_a$ found at row r, column c is consequently: $A_a^{(r,c)} = \begin{cases} w_r & c = c_a(k_r, p_r) \\ 0 & \text{otherwise} \end{cases}$
where k_r, p_r are the forest class and polarisation indices for measurement r.

- $\mathbf{A}_n$ is a $N_{meas} \times N_n$ design matrix for parameter n and $\mathbf{n} = [n_1 \quad \cdots \quad n_{N_n}]^T$ is an N_n-element vector of its instances, i.e., different values to-be-estimated with least squares and corresponding to each unique p. Each row of $\boldsymbol{A}_n$ corresponds to a unique measurement and each column to a unique instance for parameter n. With N_{pol} polarisations,
$$N_n = N_{pol}$$

For each unique p, the corresponding instance index c_n is simply p. The element of $\mathbf{A}_n$ found at row r, column c is consequently: $A_n^{(r,c)} = \begin{cases} c_r & c = p_r \\ 0 & \text{otherwise} \end{cases}$
where p_r is the polarisation index for measurement r.

Using ordinary least squares, the model parameters can be estimated to:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{s} \tag{4.19}$$

And their standard deviations to:

$$\widehat{\boldsymbol{\delta}} = \sqrt{\frac{\|\mathbf{s} - \mathbf{A}\hat{\mathbf{x}}\|^2}{N_{meas} - N_l - N_a - N_n} \cdot \operatorname{diag}[(\mathbf{A}^T\mathbf{A})^{-1}]} \tag{4.20}$$

For measurement r, the three parameters can be extracted from vector $\widehat{\boldsymbol{\beta}}$ using:

$$\begin{aligned} \hat{l}_r &= \hat{\beta}_{c_l(j_r,p_r)} \\ \hat{a}_r &= \hat{\beta}_{N_l + c_a(k_r,p_r)} \\ \hat{n}_r &= \hat{\beta}_{N_l + N_a + p_r} \end{aligned} \tag{4.21}$$

where $\hat{\beta}_v$ is the vth element of vector $\widehat{\boldsymbol{\beta}}$.

It must be noted that design matrix A can be rank-deficient (i.e., a column has all zero entries). In this situation, the corresponding regression parameter(s) cannot be solved, and a (subset) of AGB estimates cannot be computed, and will be tagged as invalid data in output.

Now an estimate of w_r for measurement r is obtained using:

$$\hat{w}_r = \frac{s_r - \hat{n}_r c_r - \hat{l}_r}{\hat{a}_r} \tag{4.22}$$

Finally, we estimate the logarithmic bias correction as the ratio of the average reference AGB to the average estimated AGB, see [RD5]:

$$\rho = \frac{\sum_{x,y} 10^{0.1w(x,y)}}{\sum_{x,y} 10^{0.1\hat{w}(x,y)}} \tag{4.23}$$

Where the estimated AGB for pixel (x, y) is calculated from a weighted average of all log-AGB estimates at that location, with the estimated parameter $\hat{a}_r$ used as the weight:

$$\hat{w}(x,y) = \frac{\sum_{r\in\mathcal{R}(x,y)} w_r a_r^2}{\sum_{r\in\mathcal{R}(x,y)} a_r^2}$$
$$\mathcal{R}(x,y) = \{\forall r \in [1, N_{meas}]: (x_r, y_r) = (x,y)\} \tag{4.24}$$

NOTE: while the description is formally carried out for the default parameter variability, eq. 4.15, 4.16 and 4.17 are independent of the selected parametrization, and are of general validity. Moreover, the exact cardinality of the regression problems depends also on the specific population (number of acquisitions, number of forest classes) in the current neighbourhood.
Moreover, from the implementation side, it must be noted that the least squares model inversion can be split into three orthogonal problems, if and only if all parameters depend on polarization: in this case, the above formulation is completely separable in the polarization domain.
Note that the least square problem can be solved with double (64bit) or single (32bit) precision, based on regressionSolver flag.

4.1.4.4. AGB mean and standard error estimation

This step first consolidates the prediction data (Level 2a stacks and forest class map) and extracts pixels that will be mapped. Next, model parameter values are assigned to the extracted pixels. Finally, AGB mean and standard error are estimated for these pixels and a map is saved. The output consists of maps of AGB mean and standard error.

Fig.15 shows the flow chart for this step.

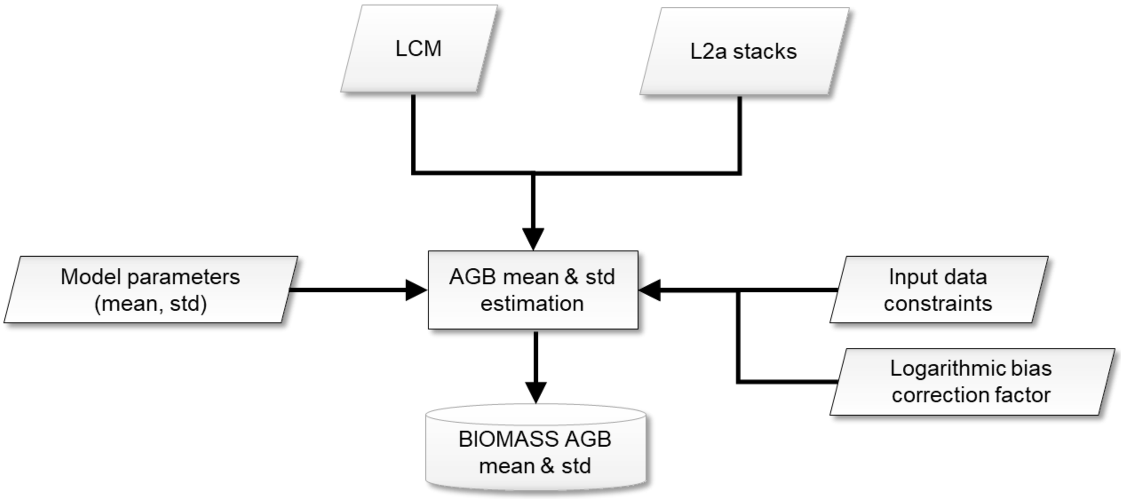

Fig.15 AGB mean and standard error estimation.

Mappable pixels (i.e., those that contain the required SAR data within the block) are defined as:

$$\Omega_{pred} = \Omega_{SAR} \cap \Omega_B \tag{4.25}$$

To estimate AGB, we solve for $w(x,y)$:

$$w(x,y) = \frac{1}{a_{pk}}\left(s_{pj}(x,y) - n_p c_j(x,y) - l_{pj}\right) \tag{4.26}$$

Based on this, we define a measurement vector as a complete set of input information for one backscatter value, including coordinates, indices for polarisation, acquisition, and forest class, and incidence angle information. We formalise it as:

$$\mathbf{d}_{pj}(x,y) = [x \quad y \quad p \quad j \quad k(x,y) \quad c_j(x,y) \quad s_{pj}(x,y)] \tag{4.27}$$

We define the complete set of measurement vectors available in the data as:

$$\mathcal{D} = \left\{\mathbf{d}_{pj}(x,y)\ \forall(x,y) \in \Omega_{pred}, \forall p \in \{1, \dots, N_{pol}\}, \forall j \in \{1, \dots, N_{date}\}\right\} \tag{4.28}$$

Where

$$N_{pred} = n(\mathcal{D}) \tag{4.29}$$

We order the elements of $\mathcal{D}$ in the same mapping order described in section 4.1.4.3 and then we assign an individual index $u \in \{1, \dots, N_{pred}\}$ to each element. The uth element is defined as:

$$\mathbf{d}_u = [x_u \quad y_u \quad p_u \quad j_u \quad k_u \quad w_u \quad c_u \quad s_u] \tag{4.30}$$

We now extract model parameter estimates $\hat{l}_u, \hat{a}_u, \hat{n}_u$ and standard deviations $\widehat{\sigma_{l_u}}, \widehat{\sigma_{a_u}}, \widehat{\sigma_{n_u}}$ for measurement u from vector $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ of equations 4.18/19, using the same mapping mechanism as described in section 4.1.4.3.

Now the mean log-AGB for measurement u can be estimated using:

$$\widehat{w}_u = \frac{s_u - \hat{n}_u c_u - \hat{l}_u}{\hat{a}_u} \tag{4.31}$$

Now, we want to estimate the standard deviation of the log-AGB estimate above. To do this, we will use the ratio distribution together with the mean and standard deviation estimates from $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$. We will assume that both the numerator and denominator are uncorrelated, non-central Gaussian distributed. Under these circumstances, a closed-form expression for the standard deviation can be used if the ratio distribution is approximated as Gaussian.

For the numerator, the mean and standard deviations are (dropping index u for simplicity of the derivation that follows):

$$\begin{aligned}\mu_{num} &= s - \hat{n}c - \hat{l} \\ \sigma_{num} &= \sqrt{c^2\widehat{\sigma_n}^2 + \widehat{\sigma_l}^2}\end{aligned} \tag{4.32}$$

And for the denominator:

$$\begin{aligned}\mu_{den} &= \hat{a} \\ \sigma_{den} &= \widehat{\sigma_a}\end{aligned} \tag{4.33}$$

[RD6] states that the ratio distribution can be approximated by a Gaussian distribution with standard deviation

$$\sigma = \frac{\mu_{num}}{\mu_{den}}\sqrt{\frac{\sigma_{num}^2}{\mu_{num}^2} + \frac{\sigma_{den}^2}{\mu_{den}^2}} \tag{4.34}$$

if the following requirements are satisfied:

$$\begin{aligned}\mu_{num} &> 0 \\ \frac{\sigma_{num}}{\mu_{num}} &\leq 1 \\ \mu_{den} &> 0 \\ \frac{\sigma_{den}}{\mu_{den}} &\ll 1\end{aligned} \tag{4.35}$$

The standard deviation for $\widehat{w}_u$ becomes (with re-introduced index u):

$$\widehat{\sigma_w}_u = \frac{1}{\hat{a}_u^2}\sqrt{\hat{a}_u^2\left(c_u^2\widehat{\sigma_n}_u^2 + \widehat{\sigma_l}_u^2\right) + \widehat{\sigma_a}_u^2\left(s_u - \hat{n}_u c_u - \hat{l}_u\right)^2} \tag{4.36}$$

Now, we average over all measurements with the same pixel to get the pixel-level estimates:

$$\begin{aligned}\widehat{w}(x,y) &= \frac{\sum_{u\in\mathcal{U}(x,y)} \hat{a}_u^2\,\widehat{w}_u}{\sum_{u\in\mathcal{U}(x,y)} \hat{a}_u^2} \\ \widehat{\sigma_w}(x,y) &= \sqrt{\frac{\sum_{u\in\mathcal{U}(x,y)} \hat{a}_u^2\,\widehat{\sigma_w}_u^2}{\sum_{u\in\mathcal{U}(x,y)} \hat{a}_u^2}} \\ \mathcal{U}(x,y) &= \left\{\forall u \in \left[1, N_{pred}\right]: (x_u, y_u) = (x,y)\right\}\end{aligned} \tag{4.37}$$

If we used the natural logarithm to transform AGB, then we could use the lognormal distribution to calculate the back-transformed mean and std from the above. However, we used decibels, so we first need to scale the estimates before we apply the lognormal distribution statistics to estimate the AGB mean and standard deviation estimates:

$$\widehat{w}'(x,y) = \frac{\widehat{w}(x,y)}{10 \lg e}$$
$$\widehat{\sigma_w}'(x,y) = \frac{\widehat{\sigma_w}(x,y)}{10 \lg e} \tag{4.38}$$

Subsequently we get the following using the properties of the lognormal distribution:

$$\widehat{W}(x,y) = \rho \exp[\widehat{w}'(x,y)] \exp\left[\frac{\widehat{\sigma_w}'^2(x,y)}{2}\right]$$
$$\widehat{\sigma_W}(x,y) = \rho \sqrt{\left[\exp\left(\widehat{\sigma_w}'^2(x,y)\right) - 1\right] \exp\left(2\widehat{w}'(x,y) + \widehat{\sigma_w}'^2(x,y)\right)} \tag{4.39}$$

4.1.4.5. Invalid pixel management and HeatMap definition

the AGB mean and std are computed over all the processing block (3°x3°), where:

- backscatter data are available;
- corresponding regression parameters have been correctly estimated.

While the first condition cannot be overcome, for the latter condition the availability of enriched reference data in the 5°x5° neighbourhood can help to build a full-rank regression matrix.

Thus, the number of invalid AGB estimates caused by missing regression parameters only shall be counted (i.e., meaning that only AGB estimates that cannot be computed due to lack of valid parameters while solving eq. 4.30), and a flag is raised asking to rerun the regression with enriched reference (i.e., second iteration taking advantage of the results of neighbouring processing blocks):

$$rerun = \left(100 \times \frac{invalidEstimateNumber}{totalEstimateNumber}\right) > \mathrm{mFV} \tag{4.40}$$

The orchestration takes advantage of the presence of the rerun flag in the output to trigger a subsequent iteration for AGB estimation step only for processing blocks that can take advantage of it.

In fact, three statuses are identified:

"Complete": all pixels are correctly processed

"NotComplete": some pixels are invalid and cannot be recovered by a subsequent iteration (or their number is under the selected threshold in eq. 4.39)

"Partial"; corresponding to rerun Boolean flag set to true (eq. 4.39): a subsequent iteration is required to fill-up estimation voids.

It must be noted that such approach, while limited to two iterations (that shall be adequate to compute AGB estimates for all processing blocks), can in principle iteratively run for any number of iterations.

The iterative run of L2B_AGB_P ends when:

- no rerun flag is set to true for any output tile (that means to string value: “Partial” in Output Product Header)

OR

- The maximum number of iterations (currently set to 2) has been reached.

The denominator of eq. 4.36 is saved as the first layer of the HeatMap HM:

$$HM(x,y) = \sum_{u \in \mathcal{U}(x,y)} \hat{a}_u^2 \tag{4.41}$$

Two additional layers will store in the HeatMap HM, at each pixel location:

- how many times reference data are used for the current regressions
- how many times additional reference data (i.e., belonging to a set previously solved by regression) are used for the current regressions

Moreover, the full traceability of the source data (i.e. L2a’s contributing for L2b data at the pixel level) is assured with the acquisitionIdImage, a volume having the following features:

- the number of layers corresponds number of contributing L2a data, where *i-th* refers to the referred to the *i-th* L2a contributing data
- for each pixel and layer, it assumes 0 value if the *i-th* L2a selected image did not contribute to the *j-th* pixel of the L2b product, 1 otherwise.

4.1.4.6. Notes on further developments

Default parametrization may be subject to updates, depending on analysis of preliminary results on real data. In particular, as (due to the limited extent of angles for satellite acquisition geometries) the regression terms ***l*** and ***n*** are highly correlated, it may be safe to assume a fixed value (***n***=2) for the third regression term (for all polarizations), and solve a simplified problem.

The current error model formulation prevents the analysis and use of covariances (i.e. off-diagonal terms in the covariance matrix). A possible alternative solution (that however presents some drawbacks) is to express the regression forward problem as a function of AGB *w(x,y)* known values.

With such a formulation, not only the error model can easily take into account the full covariance matrix, but also the variable error on AGB known values can be easily accounted for. However, ***l,a,n*** regression parameters always depend on polarization, date and forest class (that may lead to unsolvable unknowns because of lack of apriori knowledge). Biomes may be clustered in biome classes to reduce the unsolvable areas.

It must be noted that the regression problem is not generally separable (it is separable **only** for specific choices of parametrization, where the same dependency appear for all ***l,a,n*** regression parameters)

Some notable examples are:

- ***pj,pk,none*** → Regression **cannot** be be separated in smaller problems
- ***pj,pk,p*** → Regression can be separated in 3 smaller problems of dimension (Nj+Nk+1) x M/3 (where M is ther number of prior entries for regression, and priors are equally present for each polarization)
- ***pjk, pjk, pjk*** → Regression can be separated in Np*Nj*Nk problems of variable dimension 3 x M' (where M' depends on the distribution of priors)

The suggested way to lower memory requirements is taking advantage of the sparse (while unstructured) nature of incidence matrix **A**, and directly create the regression terms $\mathbf{A}^T\mathbf{A}$, $\mathbf{A}^T\mathbf{s}$, **Ax** (eq. 4.19, 4.20) without explicitly building incidence matrix **A**.

5. Appendix: effect of GN

The power associated with each pixel of I_{GN} can be expressed as [RD4]

$$|I_{GN}(r,a)|^2 = 2\int \sigma\,(r,a,z)^2[1-\cos(\,k_z(r,a)z)]\ dz \tag{5.1}$$

where $\sigma\,(r,a,z)^2$ is the uncorrelated reflectivity density of the distributed target along z, and $k_z(r,a) = k_z(r,a,1) - k_z(r,a,2)$ is the interferometric pair vertical wavenumber.