Scout-2 HydroGNSS

Algorithm Theoretical Baseline Document (ATBD) – L2/L3 PP/OP FB processor

ESA Contract No:	4000135686/21/NL/AD
Subcontract No:	MS000434
Issue / Revision:	1.6.0
Date:	10/05/2025
IFAC Doc Ref:	IFAC-ATBD-L2/L3-PP/OP-FB_v160
SSTL Doc Ref	#00655639

Prepared by Emanuele Santi

Revised by: Simone Pettinato, Fabrizio Baroni

IFAC STAFF

Export Control Document Status

UK Export Control	This document and any hardware referred to within has been UK Export Control Rated as:		
Hardware UK Export Rating Number:	N/A	Martin Unwin	
Information UK Export Rating Number:	N/A		
I confirm I am an approved SSTL Export Rater and this rating is correct as of the date above:		07/07/2025	
US Export Control	No US Content		
ITAR EAR	This document contains information controlled by the U.S. government and authorized for export only to the country of ultimate destination for use by the ultimate consignee or end-user(s) herein identified. It may not be resold, transferred, or otherwise disposed of, to any other country or to any person other than the authorized ultimate consignee or end-user(s), either in its original form or after being incorporated into other items, without first obtaining approval from the U.S. government or as otherwise authorized by U.S. law and regulations.		
ITAR	The information contained herein is controlled under US ITAR and is authorised for export under the provision of TAA/DSP:		
EAR	The information contained herein is controlled under US EAR and has been classified as ECCN:		

TABLE OF CONTENTS

LIST OF FIGURES

DOCUMENT REVISION STATUS

Last Edited Date	Revision/ Release Number	Status	Edited By	Pages / Paragraphs Affected	Change Ref
15/12/2021	0.1		Emanuele Santi	First draft version	
18/01/2022	0.2		Emanuele Santi	• Reference document list revised. • Improved description of the physical basis according to ESA's MPC16 RID to SRR. • Improved description of the algorithm implementation according to ESA's MPC16 RID to SRR. • Section from PP to OP added	
22/04/2022	0.3		Emanuele Santi	• File name changed to ATBD_L2PP_OP_FB • Input file description improved, and mask data description added. • Computational cost section added. • Revised according to RIDs MJU17-22.	
02/11/2022	0.4		Emanuele Santi	• File name changed to ATBD_L2PP_OP_FB vXX to account for progressive releases. • General revision of contents, sections reorganized. • Differences in architecture, inputs and outputs of PP, OP Alpha and OP Beta releases clarified. • inputs and outputs for each release summarized in three new tables for readability. • improved description of quality/anomalies flags.	
16/12/2022	0.5		Emanuele Santi	• added AD6. • Added RD13 and 14. • Document structure revised to match the format requested by ESA. • Changes in sections and subsections. • Changes in section titles • Modifications to reply to ESA's RIDs	
21/02/2023	0.6		Emanuele Santi	• Added RD16 • Modifications to reply to ESA's RIDs • Missing reference error fixed.	
23/11/2023	1.0		Emanuele Santi	• Updated for describing the changes in beta FB processor in v1.0 release	
18/01/2024	1.1		Emanuele Santi	• Added connection to cal/val plan (AD06) throughout the document. • Minor changes	

26/06/2024	1.5.0		Emanuele Santi	• Changes made to align with the v1.5.0 of the FB processor	
15/11/2024	1.5.1		Emanuele Santi	• Tables in section 4 completely revised to make better evidence of inputs used in the various implementations. • Other modifications to address the comments in ATBD_L2_PP_OP_FB v150_MPC.docx	
10/05/2025	1.6.0		Emanuele Santi	• Results of AGB global mapping from TDS-1 with OP beta processor added • Updated description for the recalibrating feature added to OP beta	

1 INTRODUCTION

1.1 SCOPE

The scope of this document is to describe the theoretical baseline and top-level functioning principles of the Forest Biomass processor, which aims at processing L1b data into a Level 2 (track-wise) and Level 3 global gridded 6-months product of Forest Above Ground Biomass. Three different releases are foreseen, namely Placeholder (PP), Operational Alpha (OP Alpha), and Operational Beta (OP Beta). All of them are based on Artificial Neural Networks, but they differ in inputs, outputs, and minor aspects in the implementation. The main differences between these releases are summarized here.

1.2 APPLICABLE DOCUMENTS

AD#	Title	Doc #	Revision	Date
1	HydroGNSS PDGS Algorithm ICD	00637207	v17	03/03/2025
2	Scout-2 HydroGNSS SRD Science Team Annex	00635466	v3.0	01/10/2021
3	Scout-2 HydroGNSS MRD	00632438	v2.0	08/09/2021
4	L2OP-FB User Manual	n/a	V1.5.0	26/06/2024
5	HydroGNSS ATBD RIDS	00663807	v0.3	21/03/2022
6	HydroGNSS Cal-Val plan	00644845	v6	26/09/2023

1.3 REFERENCE DOCUMENTS

1) Ferrazzoli P., L.Guerriero, N. Pierdicca, and R. Rahmoune, 2011. "Forest biomass monitoring with GNSS-R: theoretical simulations"'. DOI: 10.1016/j.asr.2010.04.025

2) Camps A., Park, H., Pablos, M., Foti, G., Gommenginger, C., Liu, P.-W. and Judge, J.,2016. "Sensitivity of GNSS-R Spaceborne Observations to Soil Moisture and Vegetation", IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., 9, 10, 2016, p. 4730-4741

3) Chew C. C., and E. E. Small, 2018. "Soil Moisture Sensing Using Spaceborne GNSS Reflections: Comparison of CYGNSS Reflectivity to SMAP Soil Moisture", Geophysical Research Letter, Vol. 45, 9, pp. 4049-4057, May 2018.

4) Carreno-Luengo H., Luzi, G. and Crosetto, M., 2019. "Sensitivity of CyGNSS Bistatic Reflectivity and SMAP Microwave Radiometry Brightness Temperature to Geophysical Parameters over land surfaces", IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., 12, 1, pp. 107-122, 2019

5) Clarizia M. P., N. Pierdicca, F. Costantini, and N. Floury, 2019. "Analysis of CYGNSS Data for Soil Moisture Retrieval", IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., 12, 7, pp. 1-9. DOI: 10.1109/JSTARS.2019.2895510

6) Egido A., S. Paloscia, L. Guerriero, N. Pierdicca, E. Motte, E. Santi, M. Caparrini, and N. Floury, 2014. "Airborne GNSS-R Soil Moisture and Above Ground Biomass Observations" IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., 7, 5, pp. 1522-1532.

7) Zribi M., D. Guyon, E. Motte, S. Dayau, J. P. Wigneron, N. Baghdadi, and N. Pierdicca, 2019. "Performance of GNSS-R GLORI data for biomass estimation over the Landes Forest", Int. J. Appl. Earth Obs. Geoinformation, 74, 150–158.

8) Dobson M. C., F.T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke, and N. Christensen, 1992. "Dependence of radar backscatter on coniferous forest biomass", IEEE Trans. Geosci. and Remote Sens., Vol. 30, Issue 2, Mar 1992.

9) Santi E., S. Paloscia, S. Pettinato, G. Fontanelli, M.P. Clarizia, D. Comite, L. Dente, L. Guerriero, N. Pierdicca, N. Floury, 2020. "Remote Sensing of Forest Biomass Using GNSS Reflectometry," in *IEEE*

Journal of Selected Topics in Applied Earth Observations and Remote Sensing, DOI: 10.1109/JSTARS.2020.2982993.

10) Carreno-Luengo H., G. Luzi, M. Crosetto, 2020. “Above-Ground Biomass Retrieval over Tropical Forests: A Novel GNSS-R Approach with CyGNSS”, Remote Sensing, 12, 2020. DOI:10.3390/rs12091368.
11) Avitabile V., M. Herold, G. Heuvelink, S. L. Lewis, O. L. Phillips, G. P. Asner et al., 2016. “An integrated pan-tropical biomass maps using multiple reference datasets”, Global Change Biology, 22, 1406–1420, DOI: 10.1111/gcb.13139.
12) Santi E., 2016. Neural Networks applications for the retrieval of hydrological parameters from microwave satellite sensors. Artificial Neural Networks - Models and Applications book, InTech Open, ISBN 978-953-51-2705-5. DOI: 10.5772/63165.
13) Technical Note N. 4: development of retrieval approaches. ESA/ESTEC CONTRACT n. 4000120299/17/NL/AF/hh/: Potential of Spaceborne GNSS-R for Land Applications (GNSS Overland). GLA-TN4/01.2019.
14) Technical Note N. 3: Neural Network retrieval of land parameters. ESA/ESTEC CONTRACT n. 4000130560/20/NL/FF/gp/: Exploiting CYGNSS GNSS-R over Land and Ocean for GeophYsics (ECOLOGY). CRAS-TN3/01.2021.
15) Technical Note on CAL/VAL PLAN for the FB PROCESSOR. N. IFAC-CAL-VAL-FB_v01, version v0.1, date 16/12/2022.
16) Technical note on the assessment of auxiliary data contribution to the AGB retrieval. N. IFAC-retrieval_contribution-FB_v01, version v01, date 21/02/2023.

1.4 ACRONYMS AND ABBREVIATIONS

The following abbreviations are used within this document:

E2ES	End to end simulator
ECMWF	The European Centre for Medium-Range Weather Forecasts
GNSS	Global Navigation Satellite System
L2OP	Level 2 Operational Processor
L2PP	Level 2 Placeholder Processor
LHCP	Left Hand Circular Polarization
RHCP	Right Hand Circular Polarization
PDGS	Payload Data Ground Segment
MRD	Mission Requirements Document
ICD	Interface Control Document
SSTL	Surrey Satellite Technology Limited
TBD	To be determined
TDS	TechDemoSat
ANN	Artificial Neural Networks
IFAC	Institute of Applied Physics “Carrara”
AGB	AboveGround Biomass

2 BACKGROUND AND SCIENTIFIC JUSTIFICATION

The retrieval concept is based on the well-known sensitivity of the GNSS operating frequency, namely L-band, to the water content of the observed targets. In particular, the L- band signal reflected by the soil surface is expected to be attenuated by the overlying vegetation in proportion to the vegetation optical depth (VOD) in turn related to the AGB (RD1).

In this respect, there is a growing interest on GNSS-R for land applications, including the retrieval of soil moisture (SM) and AGB (e.g., RD 2-5). With respect to the microwave radiometers operating in the same frequency band, e.g., SMAP and SMOS, which are routinely adopted for estimating VOD, the GNSS-R has the further advantages of a relative independence of the variation in thermal background, which conversely affect the radiometric measurements. Moreover, GNSS-R, due to the bistatic configuration, shown some saturation for AGB values of about 250 – 300 t/ha (RD 6-7) which are significantly higher than the 150 t/ha found for the L- band SAR (RD 8).

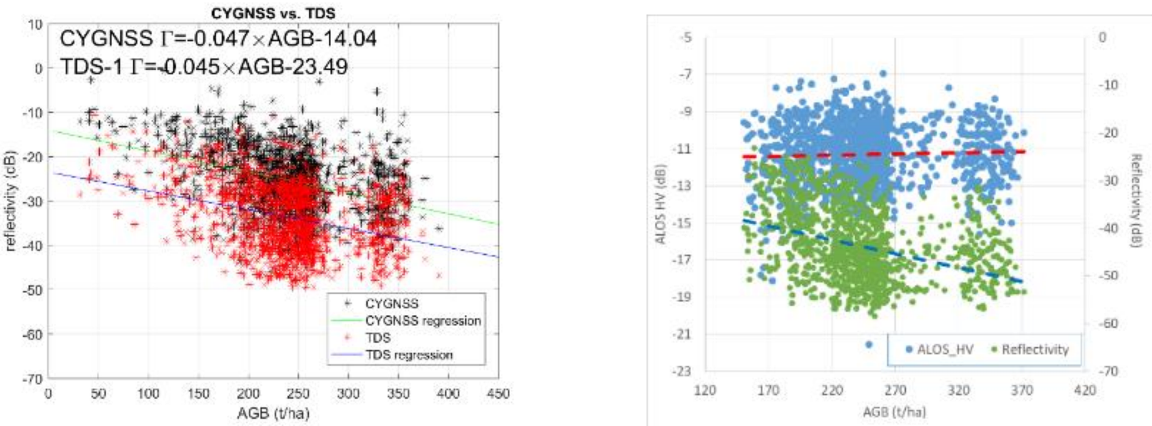
Among the GNSS-R observables, the feasibility of using the so-called Trailing edge for estimating AGB was demonstrated by Carreno Luengo et al. (RD 10), while the sensitivity of the equivalent reflectivity (Γ) to AGB was investigated by Santi et al. (RD 9). Γ is a function of the incidence angle θ at the SP which embeds the effects of surface moisture content, small scale roughness and eventual attenuation introduced by the vegetation cover. Γ can be derived from the following:

$$P_r^S = \Gamma(\theta)\frac{\lambda^2 G^t G^r P_t}{(4\pi)^2 (R_r + R_t)^2}$$

Where R_r is the distance from receiver to the specular point (SP), R_t from transmitter to SP, λ is the wavelength, G^t and G^r are the gain of transmitting and receiving (i.e., down-looking) antennas.

The outcomes of GNSS-Overland study (RD 9), which was carried out at both local and global scales by comparing the TDS-1 and CyGNSS observations to the improved pan-tropical biomass map from Avitabile et al. (RD 11), pointed out a decreasing trend when biomass increases. The behaviour of TDS-1 and CyGNSS Γ as a function of AGB is shown in Figure 1.4-1a). Data refer to three test areas of about selected as representative of different forest types: equatorial forest in Manaus, dense coniferous in Uruguay and mixed forest in Argentina (see RD 9). Despite the offset between CyGNSS and TDS-1, which depends on the different approach followed to compute Γ, given the parameters available in the data files, both sensors show almost the same sensitivity to AGB: 0.047·dB(t/ha)$^{-1}$ for CyGNSS and 0.045 dB(t/ha)^{1} for TDS-1.

The better sensitivity of GNSS-R reflectivity to the highest biomass range with respect to SAR backscatter σ° at the same frequency was also confirmed. The scatterplot of ALOS σ° in HV polarization and TDS-1 Γ, plotted as a function of AGB in the range 150 – 400 t/ha, is shown in Figure 1.4-1b).



a) b)

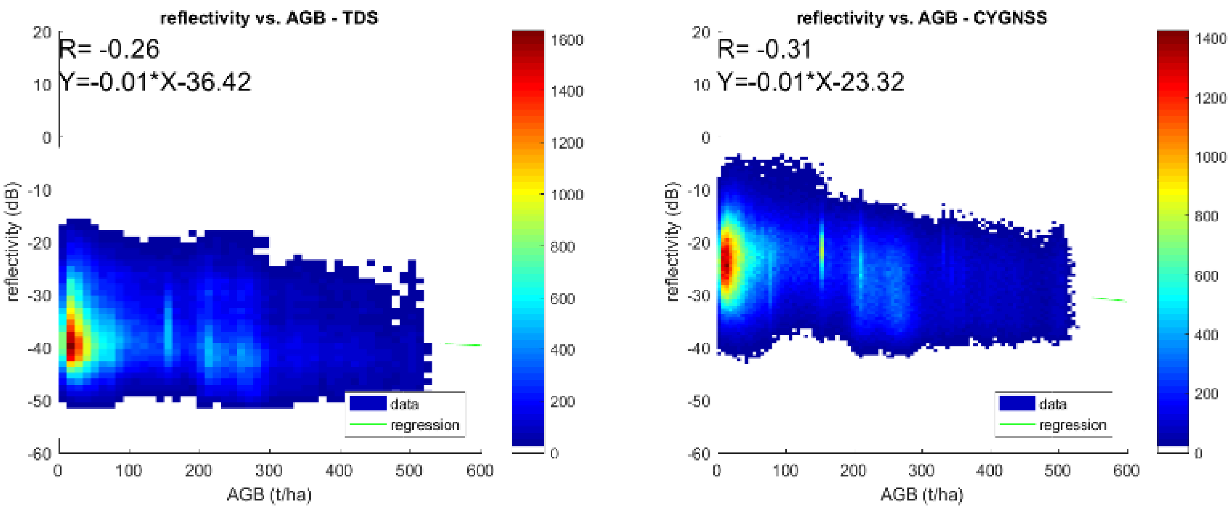
Figure 1.4-1. TDS-1 and CyGNSS reflectivity vs. AGB from pantropical map for the Manaus, Argentina and Uruguay test areas, b) ALOS backscatter and TDS-1 reflectivity as a function of AGB for biomass values higher than 150 t/ha (after Santi et al. 2020).

The plot confirms the almost complete saturation of $\sigma°$ for AGB higher than 150 t/ha, which is reported as the saturation limit for L- band SAR. Γ is decreasing almost linearly down to the maximum AGB available in the dataset.

The global scale comparison between Γ and AGB confirmed the decreasing trend when biomass increases, although with expectably lower correlation (R $\simeq$ -0.3). The TDS-1 and CyGNSS Γ behaviours as a function of AGB for the global dataset are shown in the density plot of Figure 1.4-2. The variability of surface parameters other than AGB and the difference in the acquisition geometry caused the high dispersion of data especially in sparse forests, where the variable soil moisture and roughness mostly affected the signal.

Further analysis carried out in the framework of the ESA Ecology project by using CyGNSS data confirmed the suitability of Γ for the global scale monitoring of AGB, by also pointing out that the contribution of CyGNSS observables other than Γ to the retrieval is negligible.

However, both GNSS Overland and Ecology activities pointed out that the direct correlations between GNSS-R observables and target AGB are not sufficient for accurate retrievals and advanced algorithms are needed. Both studies exploited the ANN potential for implementing the AGB retrieval algorithms. ANN have a proven ability to cope with noisy data; moreover, they offer the possibility of easily merging different information sources into the retrieval, thus making possible to combine different GNSS-R observables, their observation angles, and ancillary data, as the land cover and the local slopes.



a) b)

Figure 1.4-2 a) TDS-1 and b) CyGNSS reflectivity plotted as function of AGB on a global scale.

These capabilities were proven to be very effective in reducing the uncertainties and improving the retrievals while, the poor robustness to outliers represented the main drawback of ANN: to overcome this problem, the training shall be representative of a variety of surface conditions as wide as possible (RD 12).

For this scope, the retrieval concept proposed in Overland (RD 9, 13) and Ecology (RD 14) is based on a data driven concept that needs a training set composed of GNSS-R observables and ancillary parameters with spatially and temporally co-located ground truth data to be used as target.

As an example of the possibility to estimate AGB from GNSS-R data using ANN, the results obtained in (RD 9 and 13) at global scale using CyGNSS data are shown in Figure 1.4-3 a), where the AGB estimated by the ANN by using CyGNSS data is represented as a function of the corresponding target AGB. The ANN inputs are Γ + SNR + incidence + geographic position (lat/lon) of the point. In absence of other auxiliary information, the latter input was intended to provide to the ANN a proxy of the climatological area and forest type (see RD9 and 13). The correlation coefficient is R$\simeq$0.82 and RMSE $\simeq$ 76 t/ha.

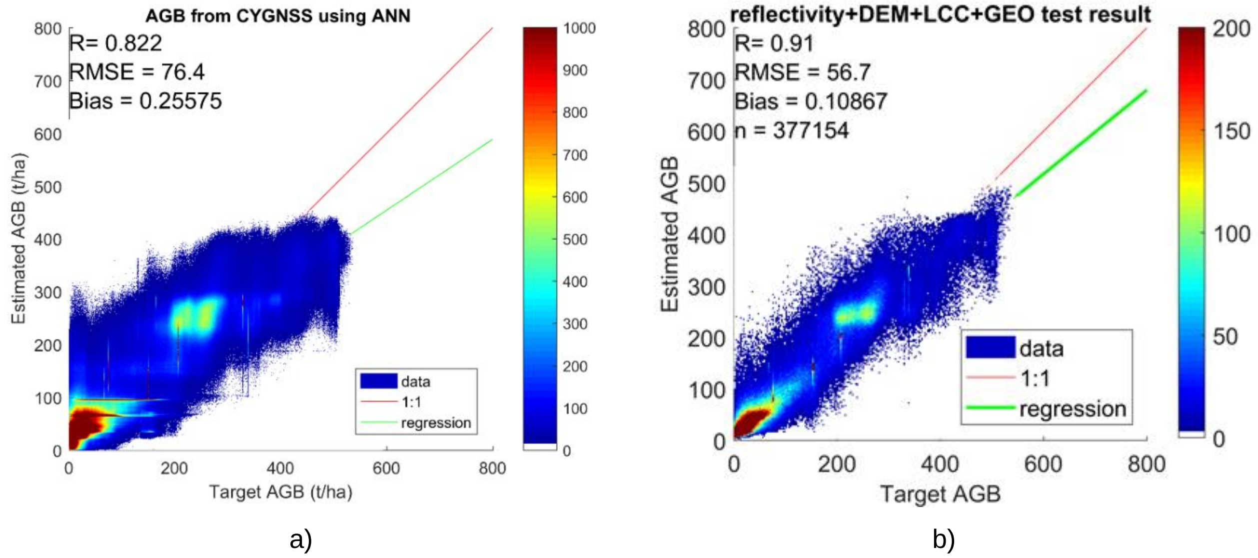


Figure 1.4-3. AGB validation on a global scale using CyGNSS data: a) outcome of GNSS Overland (RD 9), b) outcome of Ecology

Although obtained on a different dataset, the result improved with the inclusion of Land Cover Classification (LCC) and DEM + Local Slope carried out in the Ecology project (RD 14), as shown in Figure 1.4-3b). In this case, the correlation coefficient is R≃0.91 and RMSE ≃ 57 t/ha. The analysis carried out in Ecology confirmed as the most effective for the retrieval the combination Reflectivity, SNR, and incidence angle, while the inclusion of observables other than SNR provided no or negligible accuracy improvements, independent of their direct correlation to the target parameter (in any case low - see RD14). the ANN scheme implemented in Overland (RD 9 and 13) and Ecology (RD 14) is shown as example in Figure 1.4-4.

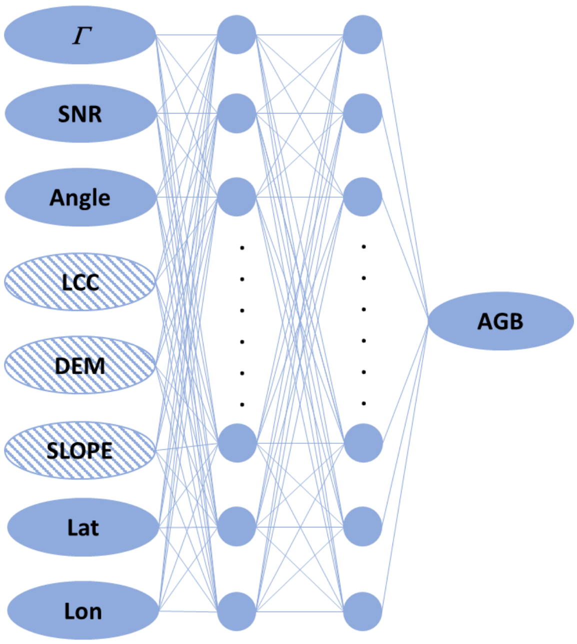
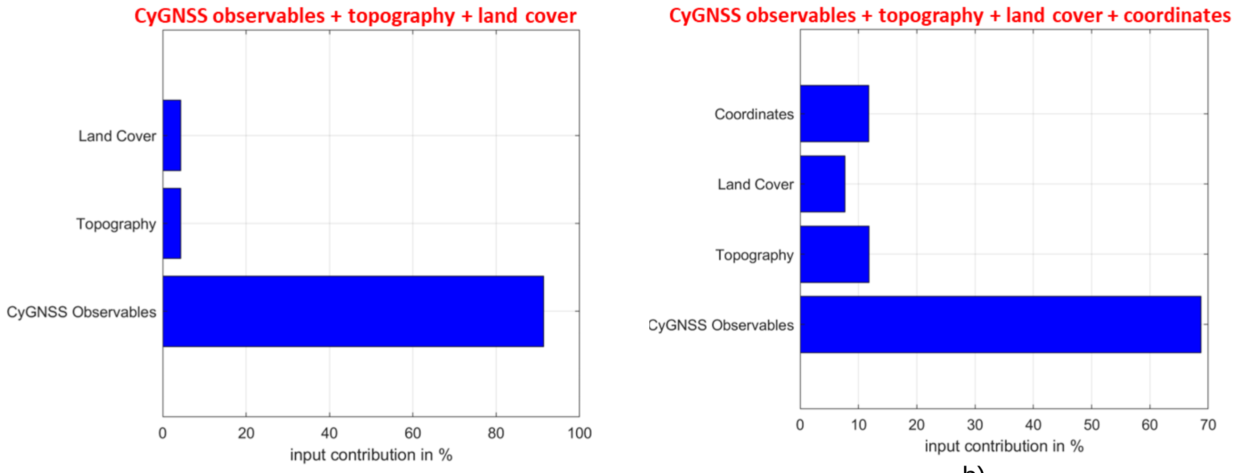


Figure 1.4-4. Scheme of the ANN from Overland and Ecology (same of Overland plus dotted inputs).

The effective contribution of the auxiliary information to the AGB retrieval is assessed in RD14 and RD15. In particular, both reports proved that auxiliary data help in constraining the inverse problem, which is in general ill posed, since multiple combinations of surface and observation parameters have the same electromagnetic response. However, both RD14 and RD16 well demonstrated that the retrieval is driven by the GNSS-R observables, as it is shown in Figure 1.4-5 that represent the relative contribution of CyGNSS and auxiliary data for two different input configurations investigated in RD16, namely CyGNSS observables + topography + land cover and CyGNSS observables + topography + land cover + coordinates of the specular point: the latter is the same to which the result shown in Figure 1.4-3b) refer.



a) b)

Figure 1.4-5. total contribution to the AGB retrieval of CyGNSS and auxiliary data for the two configurations CyGNSS observables + topography + land cover and CyGNSS observables + topography + land cover + coordinates of the specular point.

3 PROCESSOR DESCRIPTION

The IFAC-CNR Above Ground Forest Biomass processor converts L1b data into a Level 2/Level 3 product of Forest Above Ground Biomass at global scale. The AGB retrieval from HydroGNSS L1b data is carried out by an Artificial Neural Networks (ANN) algorithm based on the outcomes of the research carried out during GNSS Overland (ESA n.4000120299/17/NL/AF/hh), Ecology (ESA 4000130560/20/NL/FF/gp) and SCOUT (ESA AO/1-9901/19/NL/CT).

3.1 ANN RETRIEVAL CONCEPT

The AGB retrieval from GNSS-R observation is based on Feed-Forward Multi-Layer Perceptron ANN (FF-MLP), because of their intrinsic capability of make synergic use of multiple inputs. The processor development is based on the FF-MLP available in the Matlab ® neural network toolbox.

FF-MLP are composed of one or more hidden layers with a certain number of neurons (also called perceptrons), that are fully interconnected. The inputs are transferred from a layer to another by the transfer function: The most common transfer functions are linear, hyperbolic tangent sigmoid (tansig), and log-sigmoid (logsig). The tansig is expressed by equation (2) and returns an output value between -1 and 1 when the input a value goes from –∞ to +∞:

$$\tanh(a) = \frac{\left(e^{a} - e^{-a}\right)}{\left(e^{a} + e^{-a}\right)} \tag{2}$$

Logsig is expressed by equation (3) and its output goes from 0 to 1 when a varies between –∞ and +∞.

$$g(a) = \frac{1}{\left(1 + e^{a}\right)} \tag{3}$$

In the training module, the training is obtained by feeding the ANN with GNSS-R observables, ancillary data, and corresponding target AGB values.

The ANN training is based on the so-called Back Propagation learning rule (BP), which iteratively attempts to minimize the mean square error (MSE) between the target value and the training output at each iteration. Such minimization is obtained by adjusting the weights and offsets to inhibit or excite the connections.

Following the data driven approach outlined in (RD 9, 13) and further exploited in the Ecology project (RD 14), a subset of the entire dataset is considered for training the algorithm and the remaining data for its validation, by predicting AGB from a set of satellite data not considered in the training phase. The sampling strategy depends on the distribution of target AGB. A random sampling is sufficient if the AGB is uniformly distributed in the entire range, while more sophisticated sampling (e.g., pseudo random) can be considered in other cases, to keep all the forest types and biomasses represented in the training set. As example, the percentage of the entire data set used as *training* and *validation subsets* in the implementations for Overland and Ecology were varying from 50% and 50% for applications to small areas and limited datasets, to 1% for training and 99% for test, in case of global retrievals (RD 9, 13 and 14). In the latter case, the 1% data for training allowed to reserve the most of dataset for test, by keeping a still satisfactory accuracy.

The *training* set is further subsampled randomly in 60%, 20% and 20% subsets. The first subset serves for iteratively adjusting the ANN weights and connection strengths using BP; the second and third subsets serve having a posteriori test at each training iteration. Such strategy can be considered already sufficient for validating the algorithm, however, in an attempt of keeping test and validation as independent as possible, we preferred to validate the ANN on the *validation subset,* which is not involved in the training.

Based on the so-called "early stopping" rule, the training stops as soon as the errors on the three training subsets are diverging. Training also includes a peculiar algorithm developed at IFAC for iteratively defining the best ANN architecture in terms of neurons, hidden layers, and transfer function, with the aim of minimizing underfitting and overfitting problems. Such iterative optimization is summarized in the flowchart of Figure 3.1-1.

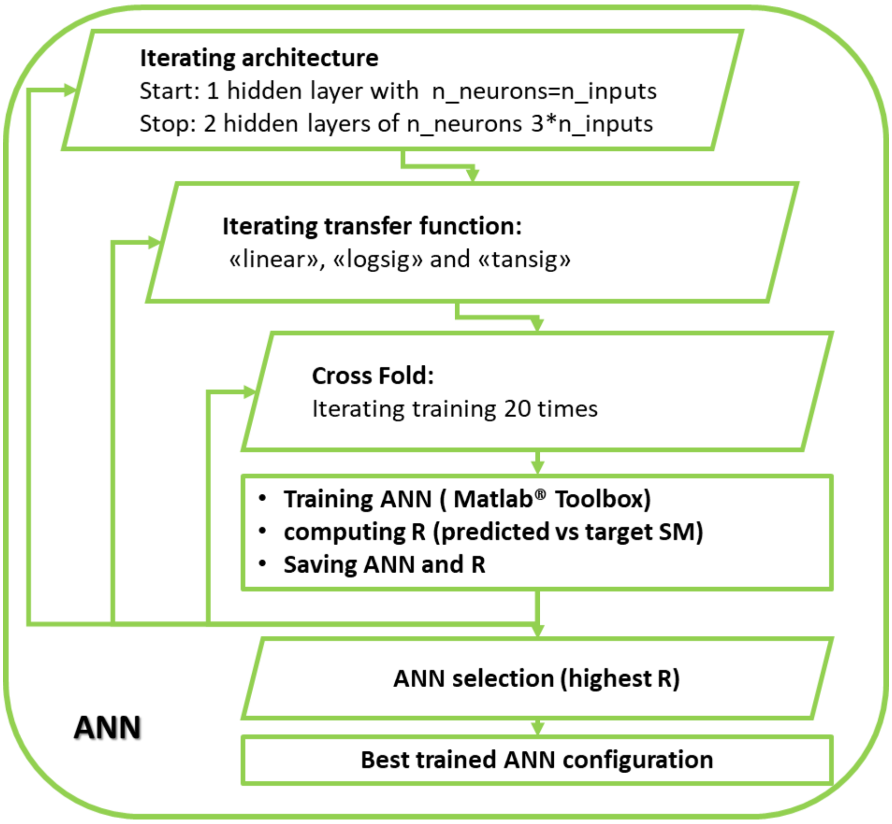


Figure 3.1-1. Scheme of the ANN iterative architecture definition developed at IFAC.

3.2 CURRENT AND FUTURE IMPLEMENTATIONS

Three different releases have been implemented, namely Placeholder (PP), Operational Alpha (OP Alpha), and Operational Beta (OP Beta). All of them are based on Artificial Neural Networks, but they differ in inputs, outputs, and minor aspects in the implementation. The main conceptual difference is that placeholder, and alpha processor are based on E2ES simulations, while the final delivery of OP beta will be adapted to the real HydroGNSS acquisitions. PP is based on the standard combination of Reflectivity, SNR, and elevation (or incidence) angle, which was demonstrated in Ecology as the most effective input combination for the retrieval (see RD14). OP alpha exploits the dual polarized data (from E2ES since no real data suitable for this scope are available so far), while the dual frequencies are exploited in OP beta, firstly using E2ES data and then HydroGNSS acquisitions. Based on the results obtained with simulated and CyGNSS data (see section 3.1), OP beta does include the auxiliary data listed in Table 4-I. We refer to RD16 for the effective contribution of these data to the retrieval accuracy. Future plans will also include the possibility to consider the coherent channel information as provided by the IEEC processor implementation, but this activity is still TBD depending on data availability.

3.2.1 Placeholder Processor

The algorithm used for the placeholder processor is a simplified implementation of the one developed for CyGNSS in the previous Ecology project. With respect to the one shown in Figure 1.4-4, the ANN developed for PP does not use auxiliary information (LCC, DEM, Slope and lat/lon), which was found not useful considering the small areas covered by the E2ES simulations, and the Monte Carlo approach used for simulating the GNSS-R observables that unlink the surface input parameters from the real surface condition. The AGB outputs are L2 track wise and L3 gridded at 25 km as specified in [AD2]. In advance of E2ES simulated data, the ANN training for this implementation has been based on the same dataset exploited in Ecology (RD14).

3.2.2 Operational Processor alpha

The OP alpha processor attempts to exploit the dual polarized data: its development has been based on the E2ES simulated data, depending on the unavailability of RHCP data from past missions for sensitivity investigation, training, and test. In detail, a dataset generated by two simulator runs on Gabon plus one on Finland, has been considered, for a total of 7000 vectors of simulated observables corresponding to AGB in a range from 0 to 350 t/ha for Gabon and from 0 to 250 t/ha for Finland.

Training has been carried out by randomly picking up 50% of the overall dataset for creating the training set and the remaining 50% for creating the test set, which was not involved in the training The training/test was repeated for two different configurations of inputs. The first configuration (herein referred as 1-pol) has as inputs the LHCP Γ and SNR plus the incidence angle. The second configuration adds the RHCP Γ and SNR. It should be noticed that, also OP alpha does not consider the LCC, DEM + Slope and lat/lon auxiliary information, which was found useful for global mapping but not for local tests in the relatively small areas covered by the E2ES simulations. The data organization for training/test is summarized in the flowchart of Figure 3.2-1.

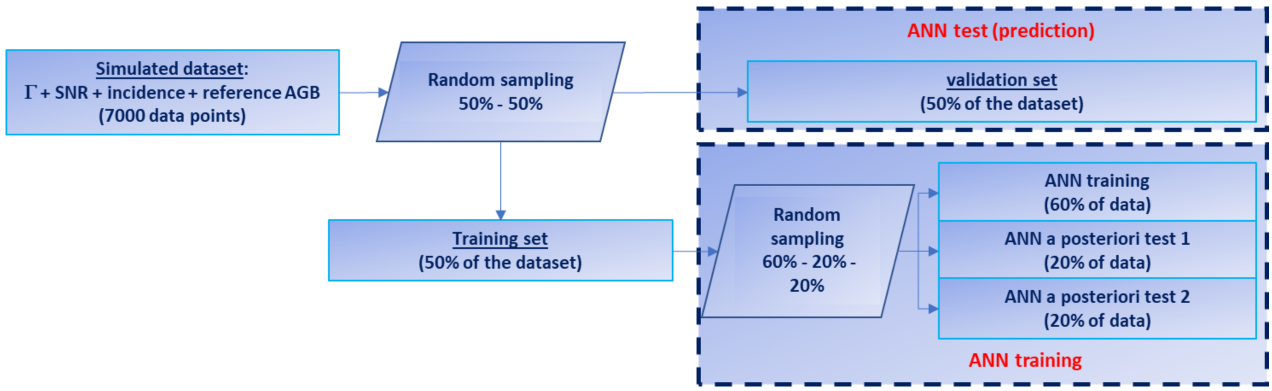


Figure 3.2-1. Data organization for training/testing the ANN.

Although Γ, especially in RHCP polarization, was found poorly correlated to AGB, with correlation coefficients R=-0.39 and R= -0.09 for LHCP and RHCP, respectively, the results obtained by applying the trained ANN to the test set (50% of data not involved in training) were quite encouraging and they proved the effectiveness of the added RHCP. As shown in Figure 3.2-2 indeed the 1-pol configuration obtained indeed R=0.61 and RMSE= 85 t/ha (Figure 3.2-2 a), while the dual polarization obtained R=0.74 and RMSE≃72 t/ha (Figure 3.2-2 b).

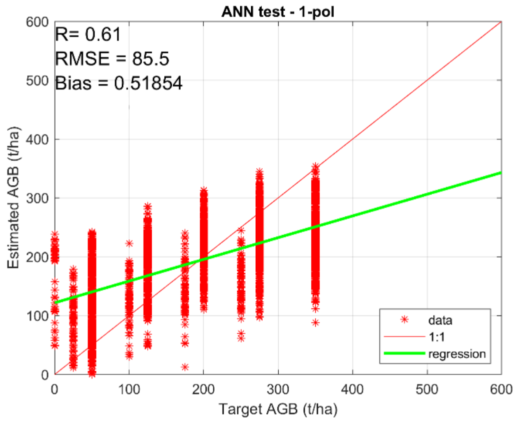


Figure 3.2-2. test of the ANN in configuration: a) single LHCP polarization, b) LHCP + RHCP polarizations.

The result, although not excellent, is really encouraging since it has been based on GNSS-R data only, without auxiliary information and it has been obtained using the same ANN for the two completely different test areas: significant improvements are therefore expected when moving to more sophisticated configurations.

A further attempt to evaluate the contribution of L5/E5 has been also carried out based on another E2ES simulated dataset including the second frequency: the result is shown in Figure 3.2-3.

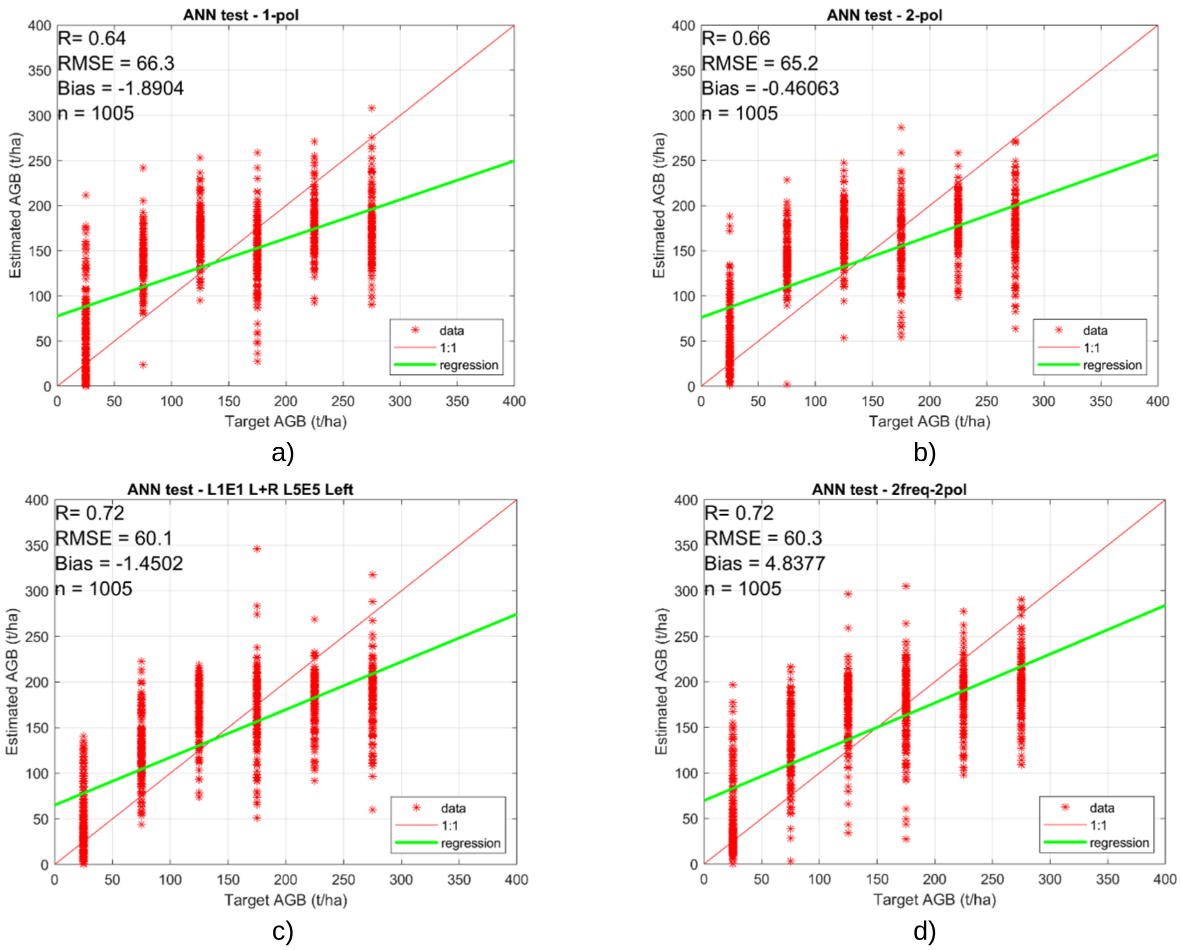


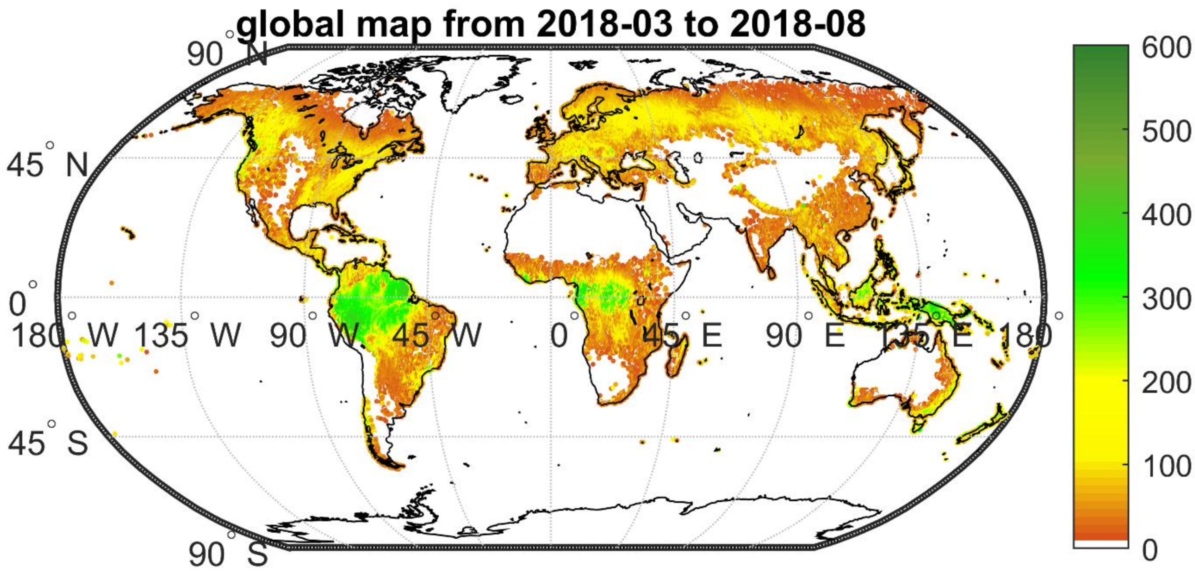
Figure 3.2-3. test of the ANN in configuration: a) L1/E1 LHCP polarization, b) L1/E1 LHCP + RHCP polarizations, c) L1/E1 LHCP + RHCP + L57E5 LHCP, d) all the four channels.

The overall metrics for L1 LHCP and RHCP are slightly worsened with respect to those shown in Figure 3.2-2 because of a different parametrization of the input “confounding” factors (SMC, HSTD and so on…), however the improvement in adding the third channel (L5/E5 LHCP) is still evident, while the addition of L5/E5 RHCP does not improve the retrievals. Although with some limitation, this result can be transferred to the operational algorithm, suggesting that the L5/E5 RHCP could be not included in the algorithm inputs.

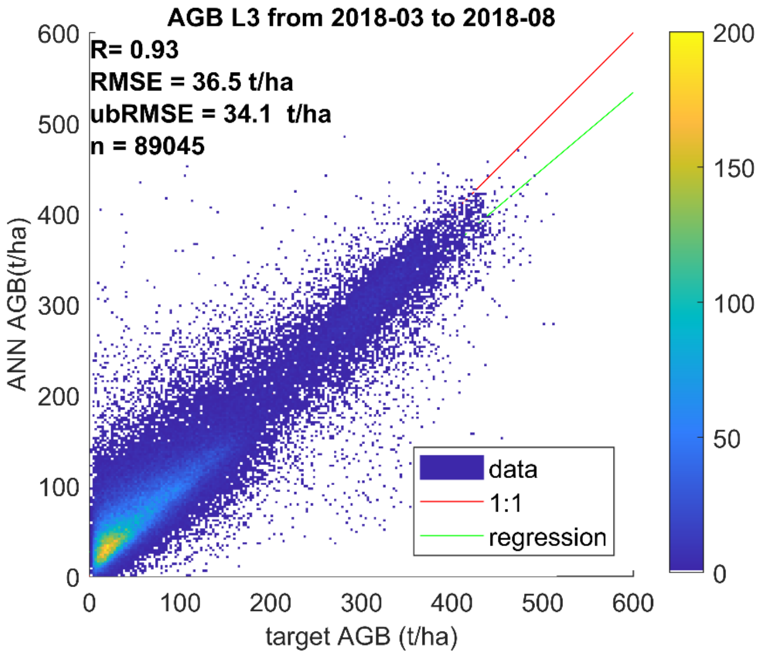
3.2.3 Beta Operational Processor (Beta OP)

The Beta OP processor aims at exploiting the dual polarization and dual frequency for increasing the retrieval accuracy and possibly improving the target resolution up to the 1 km goal (see AD02). The processor currently generates a L2 track-wise output, which is computed track by track starting from the L1b data, and an L3 gridded global product averaged over six months. The actual output resolution of the L3 product is 25 km, TBD the generation of output AGB at 1 km goal resolution (see AD02). While L2 AGB track-wise is generated track by track based on the data acquired by the satellite selected in the configuration file (HydroGNSS-1 or HydroGNSS-2), the L3 AGB gridded will make use of all the 6 months data acquired by both HydroGNSS-1 and HydroGNSS-2 satellites. Depending on the availability of different polarization and frequencies in each track, namely the L1/E1 and L5/E5 left and right polarized data, the FB processor selects the most appropriate ANN among the pre-trained configurations available for generating the L2 product. The ANN have been so far based on previous implementation using CyGNSS or trained using the data generated by the E2ES simulator but will be recalibrated and finally assessed for the operational use with payload commissioning data collection 2 (AD06).

An important problem to solve with this release is related to the AGB reference data to use for training the ANN to be used in the live system (i.e., operating with real HydroGNSS data). This does not affect the PP nor OP alpha processors, which, being based on E2ES, necessarily had as targets the same input AGB used for Monte Carlo simulations (see section 3.2.2). In the studies described in RD9, 13 and 14, the target AGB was derived from RD11, but this would imply limiting the coverage to pantropical latitudes, by excluding for instance the boreal forests that HydroGNSS is capable to cover. Other suitable sources of reference data have been identified in the Geocarbon and ESA CCI AGB. For instance, the results of an implementation based on Geocarbon and TDS-1 data collected between March and August 2018 are shown in Figure 3.2-4: Figure 3.2-4 a) shows the global AGB map generated by the ANN and Figure 3.2-4 b) the result of a validation against Geocarbon AGB represented as density plot colour coded for data occurrence.



a)



b)

Figure 3.2-4. a) Global AGB map at 25 km resolution generated by ANN from 6 months of TDS-1 data (March – August 2018), b) AGB map validation: estimated against reference AGB from Geocarbon.

These or other possibilities will be better identified and assessed when consolidating the cal/val strategy that has been now identified in AD06: in this respect, the situation, although quickly evolving, is so far challenging and the possibilities are quite limited, as pointed out in AD06 and RD15. It is also worthy to notice that the possibility of using as cal/val the airborne data and test sites in New Zealand (Rongowai) is currently under discussion with the team who is managing the experiment (see AD06, section 6.3.1).

⚠ **Independent of the input configuration and target AGB, adding, or removing inputs for exploiting the most effective combination of GNSS-R observables and auxiliary and compensating the differences between datasets is not an issue. The processor has been structured manage autonomously changes in inputs without requesting interactions from the user.**

3.3 PROCESSOR FLOWCHART

3.3.1 PP and OP alpha

PP and OP alpha processors are composed by a *training* and a *test* module: *training* is executed once after defining the training set. Training includes an iterative architecture definition to set the ANN hyperparameters, and it returns the optimal ANN for the given problem, including transfer function type, number of neurons and hidden layers, coefficients for weights and biases.

In the L2 product generation, L1b and auxiliary input data are loaded, for each specular point (SP) in the L1b track the auxiliary data are computed by interpolation and reprojection and the input are passed to the ANN. The QF (see AD2) are also screened. Then the ANN is loaded from the database of pre-trained ANNs created by the *training* module and the “per pixel” inversion is carried out to generate the AGB track-wise output which is then checked for anomalies and uncertainties. The ANN files for PP and OP alpha contain the full trained ANN according to Matlab® standard format. The main steps of the processing are summarised in the flowchart of Figure 3.3-1.

Figure 3.3-1 Top Level Functional Diagram of the PP and OP alpha FB processors.

In the L3 product generation, all the L2 AGB data track-wise are globally averaged over the 6 months period and gridded at the target resolution (25 km). Finally, the L2 or L3 AGB is outputted by the system as NetCDF + KMZ formats, according to AD2.

3.3.2 OP Beta

In the OP beta implementation, an ANN optimization has been added to the ONLINE module to account for signal fluctuations that affect the L1b data acquired over the same target and compensate their effect in the AGB retrievals. Such optimization has been implemented to address the request of having also a L2 track-wise AGB product, which was not initially foreseen. Optimization keeps the ANN architecture and transfer function output of the iterative hyperparameter definition, but it adjusts the values of weights and offsets based on a subsample of SP collected over forests in each track and corresponding targets derived by reprojection of the reference AGB (Geocarbon/CCI) stored in the auxiliary.mat file. Therefore, the pre-trained ANN files for OP Beta only contain the hyperparameters' specifications, as the number of neurons and hidden layers, the transfer function type and the sampling rate for extracting the calibration set from L1b files.

In OP beta, the L3 has been conceived as a "rolling" product that is generated every month based on the previous 6 months of data, instead of generating the product twice per year as originally foreseen. The accuracy assessment of such "rolling" approach has still to be finalized. Along with the AGB maps, quality check of results, flag anomalies and output uncertainty figures will be also produced. The assessment of the anomalies will be mainly based on outliers and quality flags in the input data. The uncertainties will be assessed mainly based on the AGB range for which the retrieval is feasible. Depending on the cal/val strategy definition (still TBD), the validation results can be included, or an accuracy flag provided. For all the details and specifications about resolution and data formats see AD1.

The possibility to generate L3 by restarting from L1b inputs averaged and gridded over six months has been also evaluated and found to have a significant impact on the computational cost (the ANN inversion to compute AGB has to be repeated) with negligible effects on accuracy.

The possibility to achieve the 1 km resolution is also under evaluation and it will be finally assessed based on the payload data collection 2, when available (see AD06).

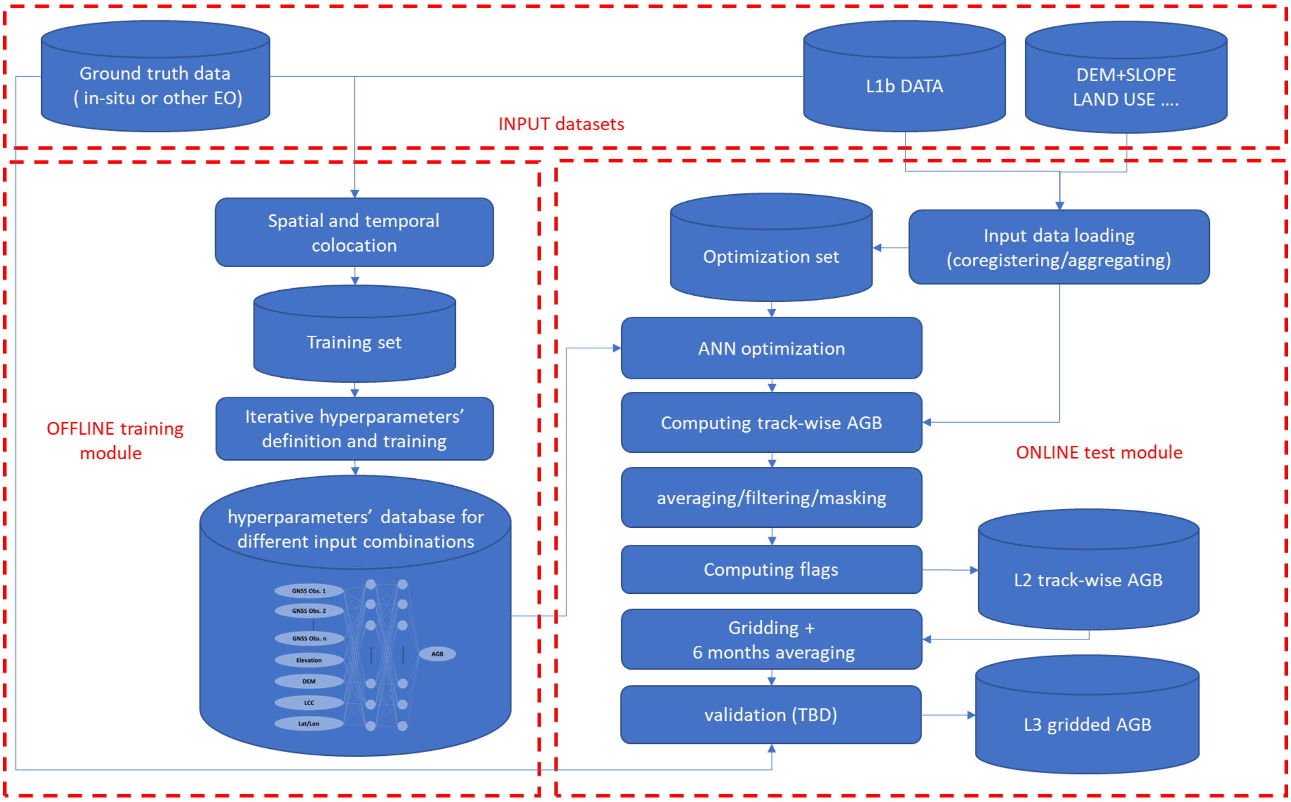


Figure 3.3-2 Top Level Functional Diagram of the OP beta FB processor.

⚠ **in all the PP and OP implementations, the training module has not been designed to be part of the online processor, but it is executed offline at IFAC. In beta OP only, the selection of the most suitable ANN configuration as well as the optimization of weights and biases is performed automatically by the processor based on the available data.**

4 INPUT DATA

4.1 PROCESSOR INPUTS

This section describes the inputs required by the FB processor to run, by evidencing the different input configurations for Placeholder (PP), Operational Alpha (OP Alpha), and Operational Beta (OP Beta) releases, respectively.

According to AD1, the L1b data needed for PP/OP FB processor either from E2ES or satellite observation, are as follows:

- **Equivalent reflectivity and SNR**, which are the main driving observables for the retrieval, computed from the Calibrated DDM peak power considering L1/E1 and L5/E5 left and right circular polarizations when available (from L1b data files).
- **Specular point location** (latitude and longitude - from L1b data files).
- **Incidence angle** (from L1b data files).
- **Very low SNR** flag (from L1b data files). This is not a direct input of the ANN but is used to filter out data that cannot be reliably used for the retrieval.
- **Snow flag** from L1b data files. In this case too, the flag is not used as ANN input, but it is used to filter out data that cannot be trusted for the AGB retrieval.

- **Auxiliary information**, including DEM, SLOPE and Land cover/land use map (LCC) for LIVE system implementation (auxiliary.mat file) or soil moisture and soil roughness at SP used in the simulations for E2ES implementation (InOutReference.mat file). We suggest using DEM from GTOPO30 (see [AD01]) and LCC from ESA CCI (see [AD01]) but other sources are also fine. These data should be resampled at the same resolution of the output AGB map (target 25 Km, goal 1 Km) and saved as matfiles. DEM is a static information, and updates are not strictly necessary. Land use is quasi static, and updates are only suggested every time a new product is released by ESA CCI. Resampling of LCC maps could not be straightforward, the suggested criterion is to aggregate at 25 km (target) or 1 km (goal) and retain only the dominant class in each pixel.
- **Configuration file** (.mat file) containing the ANN and input parametrization for each configuration.

⚠ **We refer to section 2, RD14 and RD16 for assessing the contribution of SP latitude and longitude and other auxiliary data to the retrieval accuracy.**

⚠ **The processor has been conceived to automatically adapt to E2ES or LIVE system with no supervised intervention by operators, provided that the required inputs from one or another configuration are available and correctly placed in their directories.**

The detailed input configuration for the PP, OP Alpha, and OP Beta releases is summarized in Table 4-I, with reference to [AD01] for naming and data formats.

Table 4-I. list of inputs for FB processor in different releases

		Used By		
	Input data	**PP**	**OP Alpha**	**OP Beta**
From L1b files	**Incidence Angle at specular point**	Yes	Yes	Yes(1)
	Reflectivity at specular point L1/E1 LHCP	Yes	Yes	Yes(1)
	SNR L1/E1 LHCP	Yes	Yes	Yes(1)
	Reflectivity at specular point L1/E1 RHCP	No	Yes	Yes(1)
	SNR L1/E1 RHCP	No	Yes (coupled with the corresponding reflectivity)	Yes(1)
	Reflectivity at specular point L5/E5 LHCP	No	Functionality implemented and assessed on E2ES simulations	Yes(1)
	SNR L5/E5 LHCP	No	As above	Yes(1)
	Reflectivity at specular point L5/E5 RHCP	No	As above	Yes(1)
	SNR L5/E5 RHCP	No	As above	Yes(1)
	L1B Quality flags (see [AD01]): they will include at least low and very low SNR	No	No	Yes
	Snow Flag (see [AD01])	No	No	Yes(2)

Auxiliary data for Live	**Global DEM** at the same resolution of the output AGB map (target 25 Km, goal 1 Km): suggested GTOPO30 but other sources are also fine (NetCDF/tiff or other formats) - (see [AD01])	No	No	Yes[2]
	Land Use Flag (see [AD01]) **or Land cover/land use map** from ESA CCI (Net CDF or tiff) resampled at the same resolution of the output AGB map (target 25 Km, goal 1 Km)	No	Yes	Yes[2]
	Mask for Sea/open water, not forested areas, and all the areas in which the AGB cannot be retrieved. Provided as logical map with “1” for valid pixels and “NaN” elsewhere.	Yes	No (Replaced by land use flag)	No (Replaced by land use flag/map)
Auxiliary data for E2ES	**Soil Moisture at SP** used in the simulations (from InOutReference.mat file)	No	No	Yes[3]
	Soil roughness at SP used in the simulations (from InOutReference.mat file)	No	Yes	Yes[3]
Configuration files	**ANN architecture** (from net.mat file containing the configuration of the trained ANN - neurons, hidden layers, weights, transfer function and so on)	Yes[4]	Yes[4]	Yes[4]
	ANN input configuration (from net.mat file containing the list of input observables and auxiliary data for the given configuration)	Yes	Yes	Yes
	Color palette for the KML preview (optional - from Colorbar.mat)	Yes	Yes	Yes

[1] Extension to and assessment with HydroGNSS measurements will be based on the payload data collection 2, when available (see AD06).

[2] In LIVE system only, not used in E2ES. Extension to and assessment with HydroGNSS measurements will be based on the payload data collection 2, when available (see AD06)

[3] Used in E2ES only, not used in LIVE system.

[4] in PP and OP Beta the configuration file contains the full net file according to Matlab® standard format, in OP only the architecture parameters are stored (number of neurons, number of hidden layers, transfer function and so on), since weight and biases are recalibrated at each run.

⚠ **We refer to AD1 and AD4 for folder structure, file structure and contents, and naming convention for both L1b and ancillary data.**

4.2 ANN INPUTS

The OP Beta processor adapts automatically to the different availability and suitability of input data in different tracks (e.g. constellations that do not transmit L5/E5 and/or too low right signal) by selecting the most appropriate ANN configuration based to the available frequencies (namely E1/L1 and E5/L5), and polarizations (namely left and right) and L1b quality flag. The selection is carried out track by track before generating the L2 track-wise product. In detail, three input configurations are considered with different auxiliary data if working in E2ES or LIVE system. The inputs for each configuration (namely ANN1IN, ANN2IN and ANN4IN) are summarized in Table 4-II.

Table 4-II. List of the inputs for the current ANN configurations.

		ANN1IN	ANN2IN	ANN4IN
E2ES and LIVE	E1/L1 LHCP	YES	YES	YES
	E1/L1 RHCP	NO	YES	YES
	E5/L5 LHCP	NO	NO	YES
	E5/L5 RHCP	NO	NO	TBD depending on the results with payload data collection 2
LIVE system	LAT/LON	YES(1)	YES (2)	YES (2)
	DEM+SLOPE	YES(1)	YES (2)	YES (2)
	Land Use	YES(1)	YES (2)	YES (2)
E2ES	SMC	YES(3)	YES (3)	YES (3)
	HSTD (Soil Roughness)	YES(3)	YES (3)	YES (3)

(1) Implementation based on the concept developed for the CyGNSS AGB algorithm (see RD09).

(2) Used only in LIVE system. Extension to and assessment with HydroGNSS measurements will be based on the payload data collection 2, when available (see AD06).

(3) Used only in E2ES. It is derived from the InOutreference.mat file generated by the simulator and containing the input parametrization for each simulated dataset. The auxiliary inputs used in the LIVE system are useless for site specific cases as those covered by the E2ES simulations.

5 OUTPUT DATA

The algorithm outputs are (see AD1 and AD4):

- AGB map track-wise (L2) and 6-month globally gridded on EASE Grid (L3) at the target resolution (25 km) in NetCDF + KMZ formats (TBD the possibility to reach the goal resolution of 1 km). Open water/urban and all the not forested areas, for which the forest AGB cannot be estimated, are flagged as NaN/not valid in the output product, based on auxiliary data used in each release, as listed in Table 4-I.
- Quality check of results, Flag anomalies and output uncertainty figures. This does include quality flags of the L1B data and count of outliers or not valid values in the output AGB map. The latter is not included in the PP and has been implemented starting from the alpha OP processor. TBA the assessment using HydroGNSS data, since the accuracy is intended for the gridded (L3) product, and 6 months of data (currently not available) are needed for computing it.

⚠ **L2 output is generated track by track, and therefore cannot meet the accuracy requirements, but quality and anomalies flags are computed anyway.**

The list of the output divided by release is shown in Table 5-I.

Table 5-I. list of the outputs generated by each processor.

	Generated by processor		
output dataset	**PP**	**OP Alpha**	**OP Beta**
L2 AGB track-wise as netcdf + kmz file - see [AD01/AD04]. This product is generated track by track based on the data acquired by the satellite selected in the configuration file (HydroGNSS-1 or HydroGNSS-2)	Yes	Yes	Yes(1)

L3 AGB gridded as netcdf file - see [AD01/AD04]. Differently from the L2 product, the L3 generation will make use of 6 months data acquired by both HydroGNSS-1 and HydroGNSS-2 satellites.	Yes (is an L2 gridded since the simulator cannot generate global data)	Yes (As PP)	Yes(1)
quality flag Binary flag for good/bad retrieval at each SP location (L2) or pixel in the grid (L3)	Yes	Yes	Yes(1)
AGB Uncertainty for L2/L3 It is computed over the entire track and accounts for the presence of very low SNR, snow, not forested areas and outlier AGB values. values range is between 0% (maximum uncertainty – AGB value not reliable at all) and 100% (minimum uncertainty – AGB maximum reliability)	No	No	Yes(1)

(1) Implemented and tested with TDS-1 derived data. Extension to and assessment with HydroGNSS measurements will be based on the payload data collection 2, when available (see AD06).

⚠ **We refer to AD4 for details on the destination directory structure and file naming and contents.**

6 COMPUTATIONAL COST

The processor has been locally tested at IFAC with the following hardware/software setup:

- CPU: Intel® Core ™ 19-10900 CPU @2.80 GHz
- RAM 128 Gb
- HDD 4Tb
- OS: Windows 10
- Matlab Runtime to run the executable standalone.
- Standard Administrator rights for the execution.

Running the code in the machine listed above took about 30 seconds for generating L2 product from each 6-hours group of data. The ANN requires only a few seconds to estimate AGB, the remaining is equally divided between reading input data and writing output (NetCDF + KMZ). The generation of L3 map from one month of L2 data (both HydroGNSS-1 and HydroGNSS-2) takes about 180 sec.

7 FUTURE ACTIVITIES

Besides adapting and testing the beta OP processor with HydroGNSS data, when available, there is a plan to exploit the coherence information as delivered by the processor developed by IEEC and evaluate its possible contribution to the retrieval accuracy. This activity will require further research which is still TBD depending on resources availability.

8 INSTALLATION AND USE

For all the details about installation of the processor and use, please refer to the user manual [AD04].