CCI
BIOMASS

System Specification Document

PHASE 2 / YEAR 1
VERSION 4.0

Document ref:	CCI_BIOMASS_SSD_V4
Deliverable ref:	D3.2-SSD
Version:	4.0
Creation date:	2023-03-13
Last modified	2023-03-13

Document Authorship

	NAME	FUNCTION	ORGANISATION	SIGNATURE	DATE
PREPARED	Wiesmann	SysEng	GAMMA		13.03.23
PREPARED	Cartus		GAMMA		
PREPARED					
PREPARED					
PREPARED					
PREPARED					
PREPARED					
PREPARED					
REVIEWED	H. Kay	Project Coordinator	Aberystwyth University		
REVIEWED	R. Lucas	Project Manager	Aberystwyth University		
VERIFIED	S. Quegan	Science Leader	Sheffield University		
APPROVED					

Document Distribution

ORGANISATION	NAME	QUANTITY
ESA	Frank Martin Seifert	

Document History

VERSION	DATE	DESCRIPTION	APPROVED
0.1	2018-10-01	First draft version	
1.0	2019-05-31	Finalised version	
2.0	2020-02-25	Finalised version	
3.0	2021-03-31	Finalised version	
4.0	2023-03-13	Finalised version	

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

VERSION	DATE	DESCRIPTION	APPROVED
4.0	2023-03-13	Update reference document versions Corrected minor text errors Updated processing specifications for year 31 phase 2	

TABLE OF CONTENTS

TABLE OF FIGURES

SYMBOLS AND ACRONYMS

ADP	Algorithm Development Plan
AGB	Aboveground Biomass
AGBC	Aboveground Biomass Change
ALOS	Advanced Land Observing Satellite
ATBD	Algorithm Theoretical Basis Document
AWS	Amazon Web Services
CAL/VAL	Calibration Validation
CCI	Climate Change Initiative
CEOS	Committee on Earth Observation Satellites
CMS	Content Management System
COTS	Commercial Off The Shelf
DARD	Data Access Requirements Document
DEM	Digital Elevation Model
ECV	Essential Climate Variables
EO	Earth Observation
ESA	European Space Agency
FBD	Fine Beam Dual
FOSS	Free and Open Source Software
GCOS	Global Climate Observing System
GSV	Growing Stock Volume
IGOS	Integrated Global Observing Strategy
IPCC	Intergovernmental Panel on Climate Change
JAXA	Japan Aerospace Exploration Agency
OSCAR	Observing Systems Capability Analysis and Review Tool
PS	Processing System
PSD	Product Specification Document
PVASR	Product Validation and Algorithm Selection Report

PVP	Product Validation Plan
SAR	Synthetic Aperture Radar
SoW	Statement of Work
SRD	System Requirements Document
SSD	System Specification Document
URD	User Requirement Document

1 Introduction

The European Space Agency (ESA) Climate Change Initiative (CCI) aims to generate high quality Essential Climate Variables (ECVs) derived from long-term satellite data records to meet the needs of climate research and monitoring activities, including the detection of variability and trends, climate modelling, and aspects of hydrology and meteorology.

1.1 Purpose

This document is deliverable D3.2 System Specification Document (SSD) of the ESA CCI Biomass project (Phases 1 and 2) requested in the Statement of Work (SoW) [R-1]. The SSD incorporates the requirements described in the System Requirements Document (SRD) [R-6] and specifies the characteristics of an operational ECV production system from a System Engineering point of view.

The system design is based on the experience developed and applied in previous work, including the ESA GlobBiomass project. The specifications foresee a modular approach covering the sequential modules from data uptake to product verification, resulting in the workflow specifications. Operational scenarios specify roles relevant for the Processing System (PS): the most relevant roles fall to the development team for the development and implementation of the CCI Biomass PS; external experts for the validation of products; and ESA for overall control. The functional design outlines the foreseen services and processors to be implemented. Most relevant are the processors for the retrieval of the AGB products and product uncertainty. Services at this stage cover mainly support services for development, the software repository and the issue tracker. Important aspects in the functional design specifications are the development and documentation. The aim is to have a well-documented, easy to maintain processing system which can be further developed. At this stage the processing time specifications are not critical and are reachable with standard commercial off the shelf (COTS) servers. The processing system will be further advanced in the years to come. Consequently, this SSD is a living document and will be complemented when necessary.

1.2 Document Overview

- Section 2 gives an overview of the CCI Biomass PS. It describes its purpose and intended use as well as the main requirements, functions and components.
- Section 3 shows the main operational scenarios.
- Section 4 discusses the necessary infrastructure.
- Section 5 highlights the functional design from different perspectives, the users, system operators and developers view.
- Section 6 summarises information about the system life cycle design, implementation and maintenance costs and performance.
- Section 7 gives the system requirements of [R-6] traceability and evolution.

1.3 Document Status

This document is based on issue 3.0 of the System Requirements Document (SRD), issue 2.0 of the Data Access Requirements Document (DARD), Issue 3.0 of the Product Specification Document (PSD), and Issue 2.0 of the User Requirements Document (URD); refinement of this document will be necessary with future issues of these documents.

1.4 Reference Documents

Table 1 below lists the reference documents of the system requirements.

Table 1: Reference Documents

Acronym	*Title*	*Issue*
R-1	CCI+ SoW Issue 1.4 r2	EOP-SEP/SOW/0031-1.4 r2
R-2	User Requirements Document (URD)	3.0
R-3	Product Specification Document (PSD)	
		3.0
R-4	Data Access Requirements Document (DARD)	
		2.0
R-5	Data Standards Guidelines (DSWG)	
	System Requirements Document (SRD)	3.0
R-6	Algorithm Theoretical Basis Document (ATBD)	2.0
R-7	Algorithm Development Plan (ADP)	2.0

2 CCI+ Biomass Processing System Overview

This section gives an overview of the CCI+ Biomass PS with its main modules, functions and components. It also summarises its designated use and the system requirements.

2.1 Context

As depicted in the general overview of Figure 1, the CCI Biomass PS generates products and supports the process of algorithm improvement, reprocessing and validation. It provides products and services to the biomass community, supporting their climate change impact assessment over a wide range of scales. The PS will be used by the CCI Biomass consortium but can also be applied by others as the overall workflow is largely generic. The key difference to data production in other science projects is their often-missing dissemination, i.e., the work ends with a publication and generated data products are not shared. The PS is specified to provide biomass products such as Above Ground Biomass (AGB; Mg ha^{-1}) and Above Ground Biomass Change (AGBC) based on state-of-the-art technology using the best suited and available Earth Observation (EO) data and algorithms. The products are produced in a transparent and documented way, with accompanying meta-data, documentation, uncertainty and validation reports (project outreach).

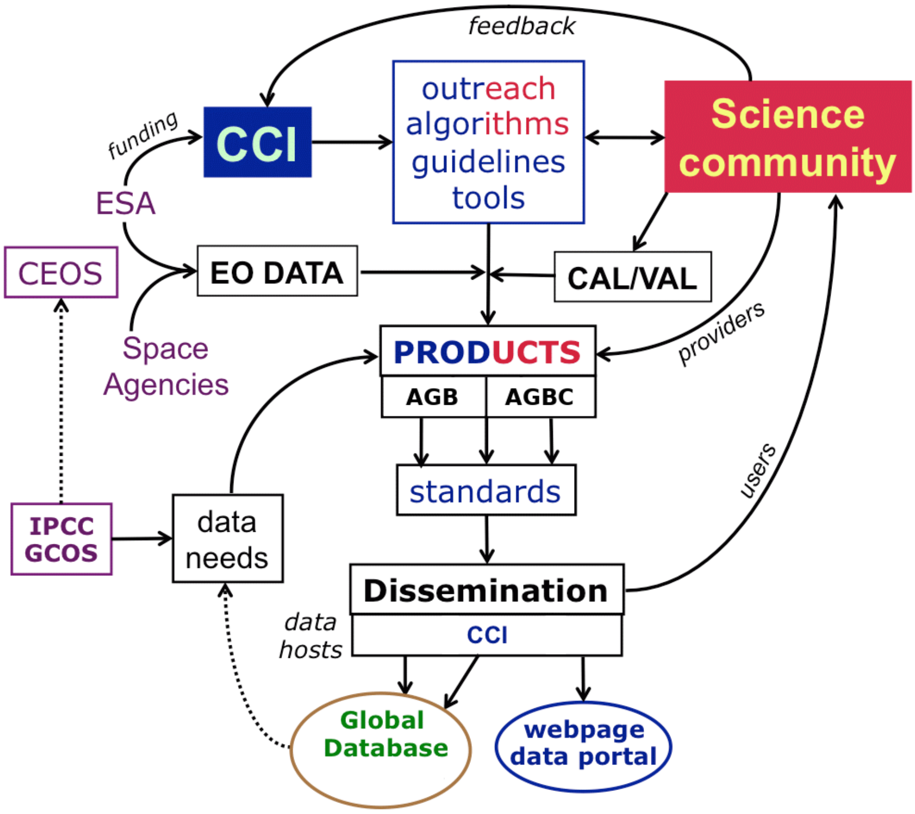

Figure 1: Biomass_cci Processing System Environment

The PS can be understood as a value-adding layer between the data provider and the users. A high-level relation diagram of the PS is given in Figure 2. There are interfaces to the different user communities, which receive products and can provide feedback. Another interface is with the EO data providers. Depending on the module, EO data are obtained from the providers at the Committee for Earth Observation Satellites (CEOS) level 3 or 4 and are ingested into the PS. Feedback is given to the providers about issues found with the data, processing improvements and requirements for the continuity of the service. Another interface is with third-party sources to receive ancillary and validation datasets.

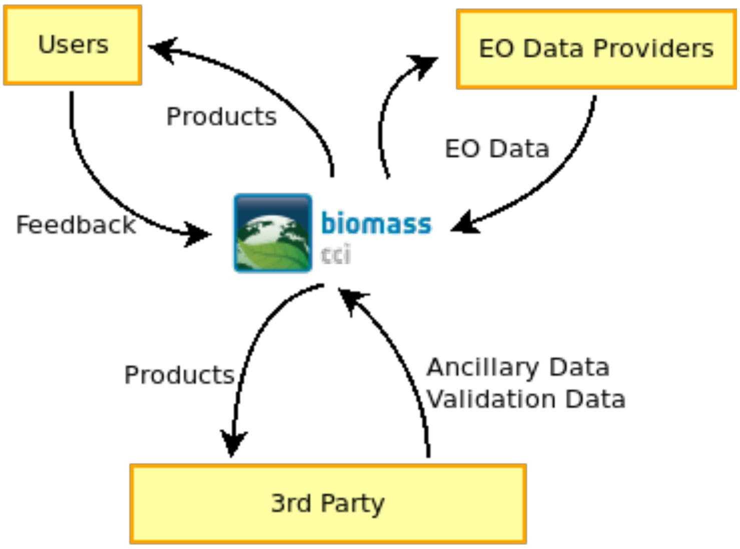

Figure 2: High-level relation diagram of the Biomass_cci PS.

2.2 User Requirements

The User Requirements are documented in the URD [R-2]. This document outlines the requirements for the CCI Biomass EECV obtained through engagement with users from across climate applications, including the detection of variability and trends, climate modelling, and other aspects relevant to AGB and biomass in general. The primary parameters to be evaluated are the requirements for GCOS.

A synthesis of input from the Global Carbon Observing System (GCOS), the Integrated Global Observing Strategy (IGOS) and the Observing Systems Capability Analysis and Review Tool (OSCAR) illustrates the full range of user needs, within which CCI BIomass requirements and product specification occupy a specific niche.

The general baseline requirements for AGB are global maps depicting globally the distribution of aboveground carbon stocks at a spatial resolution better than 500 x 500 m^2. The update frequency, latency and repeat interval vary between applications.

2.3 Main System Requirements

System requirements are compiled in the SRD [R-6]. The document lists system requirements grouped into functional, operational and performance requirements, affecting the system design. Section 7 provides the complete matrix of forward tracing from requirements to sections and indicates evolution of the requirements with time.

The high-level requirement is to generate the CCI Biomass products (BM-FUN-0030). The processing line will be well-defined and flexible for future updates and adaptations (better algorithms, new input data) (BM-SIZ-0080). The available data will be frequently reported and properly disseminated to the interested user communities (BM-INT-0090).

The main scenario is that the functional scope of the system is not restricted to the processing, reprocessing, validation and improvement cycle, although this is its main purpose. Also functions to make output products and documentation available to users are included in the scope of the system.

2.4 Main Functions

Requirements in this section are:

→ BM-FUN-1010 Develop and validate algorithms to approach the GCOS ECV

→ BM-FUN-0040 Generate and fully document a production system capable of processing and reprocessing the data

→ BM-OPE-0060 capitalise on existing assets

→ BM-FUN-1050 Reprocess with new or so far unprocessed data

The PS hosts input data, performs pre-processing, classification processing, supports validation, and serves the output to users ('Dissemination' in Figure 1). It supports the interaction between the development team and users by information services. Processor interfaces, configurable data management and version control, with easy transfer to operations, supports testing and development of new algorithms and continuous improvement.

To fulfil its purpose in such a context, the PS provides three high level functions:

- Production
- Dissemination
- Life Cycle Management

In the following, we will discuss the fundamental operations of the PS regarding these three functions. For production the focus is on processing and repeated reprocessing of complete products. Necessary functions are:

1. Storage to gather and store inputs, intermediate products, output products and auxiliary data;
2. Processors to produce output products from the input data;
3. Processing Control;
4. Quality Control of the intermediate and output products;
5. Comparison with reference data;
6. Documentation of the processing using meta-data;
7. Ingestion of new input data and auxiliary data.

In general, we distinguish between the pre-processing, the main processing and the post-processing functionality covering the preparation of the input data, the processor itself, and the product generation steps, respectively.

For dissemination, the focus is on the service for the biomass and climate community. The products are distributed through existing and in development platforms of the ESA CCI Data Portal and the CCI Biomass Website. Functions are:

1. Project Information/Introduction
2. Processing Information
3. Product Description (incl. meta-data)
4. Online Product Access
5. Validation Support
6. Feedback Handling
7. Long-term Preservation

Good Life Cycle Management helps improve service quality and reliability, crucial elements for the attractiveness of the provided service. A small effort should be necessary to implement an improved processor handling, improved algorithms and data from new sensors, given that basic characteristics of the data and the processing do not change. Consequently, fundamental necessary operations are:

- Test environment (for new processors)
- Access to test or benchmark input data (for tests and comparison)
- Version Management (this is linked to the point "Documentation of the Processing in Meta-data")

2.5 The modules of the PS

At this stage, AGB map production within the CCI Biomass project has one processing chain, and therefore a single processing module. However, as with some of the other CCI projects, there might be future products that rely on completely different processing chains in terms of input and ancillary data and the processor. The processing chains for each product are organised in modules that are part of the PS. In [R-6] the following modules are part of the PS:

1. Above Ground Biomass (AGB)
2. Above Ground Biomass Change (AGBC)

Furthermore, these modules will potentially be developed for different input data, resulting in separate modules.

Within the framework of the CCI Biomass project, we aim at developing the most efficient, consistent and sustainable system addressing the needs of the relevant communities. We also aim at investigating synergies among the modules especially for the interfaces, but potentially also with other CCI projects, if the system benefits. Hence, in the following, we address both levels: the module level and the system level. The distributed approach of the PS requires a community of producers and consequently a coordination mechanism to ensure continuity and management of multiple outputs.

2.6 High Level Decomposition

The functions and modules listed in the previous section are implemented as functional components. In this section, we outline the high-level architecture of ESA CCI at this subsystem level. Data uptake is either through the mostly web-based services of satellite data providers (e.g.,EOLI-SA, GLOVIS/Earthexplorer, S1/2 Science Hub) and those listed in [R-4]. A key requirement is that all EO data are open and freely available. Data are then processed in the product generation modules and distributed through the CCI Common Basic Services (CCI data portal). Within this document, we focus on the product generation (see Section 3.2). Furthermore, we provide an overview of the components of the PS as a starting point for the operational scenarios in Section 3 and design in Section 4. The specification with all components, functions and interfaces follows in Section 5.

On a high-level, we can distinguish three subsystems: production, user services, and data stewardship based on the way EO data are encapsulated. This can also be viewed from a functional perspective as production, dissemination and improvement. Processing storage of the production system is accessed by the module processors and need not be openly accessible (i.e., with write access to the storage medium). The user services and archive are taken care of by the CCI data portal and are not discussed here.

The production subsystem contains the production and development. Production control, processing storage and the processors provide the basic infrastructure for processing. A test environment with read access on input data serves the development needs. Where applicable, the main functions of the processing system and processing storage will be available for development. Due to the distributed processing system, development of new versions will always be performed in parallel to running an older version.

2.7 Hardware Infrastructure

The processing within the CCI Biomass project is undertaken in a distributed way. Consequently, the different PS modules will be installed on different processing hardware infrastructure.

2.7.1 AGB Sentinel-1 Processor

Currently, the production is foreseen to happen partly in the cloud on Amazon Web Services (AWS) and on the GAMMA Linux cluster. For the pre-processing of the Sentinel-1 data, AWS is used. It allows parallelisation of the orthorectification of the 100k input files and benefits from the data archive being located on the Amazon servers. After retrieving the intermediate products, the processing is undertaken at GAMMA on a Dalco Linux cluster.

3 CCI+ Biomass PS Workflow and Operational Scenarios

3.1 Roles

The development team consists of scientists and operators that manage the production and continuous development. Actors in the operational scenarios are users with different roles depending on how they use the system:

- CCI Biomass Users (client)
  - Interested in best existing biomass products
  - Skilled in biomass applications
  - Provide feedback and proposals
  - Request a data format compatible with their communities
- Development team
  - Mandated to run CCI Biomass
  - Are in dialogue with users
  - Advances the PS further
  - Issues product versions
- Validation experts
  - Are part of the international community
  - Support the development team
  - Provide local expertise
  - Feedback on the products
- Auditor/Project Manager
  - Project supervision

3.2 User Information and Data Access

Users access the CCI Biomass data products using the ESA CCI Open Data Portal. Updates to the system are disseminated via the CCI Biomass website.

Typical functionality performed in the context of user information and data access:

- The development team submits new products to the ESA CCI Open Data Portal. New versions or major updates will be announced on the CCI Biomass website.
- A user accesses the CCI Biomass web site as an entry point and gets general information on CCI Biomass. It also provides links to software, services and resources.

- The user uses the CCI Search catalogue to find CCI Biomass products. The catalogue also provides metadata and quicklooks.
- The user downloads product documentation (e.g.,the Product User Guide or PUG).

3.3 Processing System Workflow

The basic processing system production workflow is given in Figure 3. It has the following main parts: pre-processing, retrieval, product generation and verification/validation. The input data are EO data and auxiliary data such as a Digital Elevation Model (DEM), land cover maps, and masks. The output of the production is the AGB products. In the following, the CCI Biomass modules will be presented in further detail.

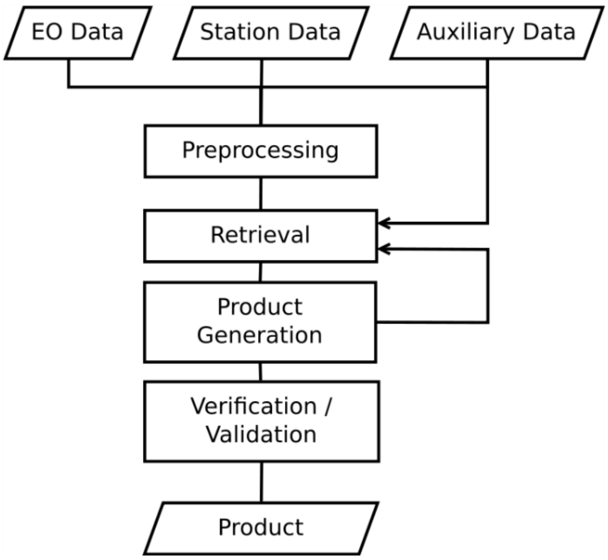

Figure 3: Biomass_cci module high level diagram.

3.3.1 Pre-processing

The following EO datasets are ingested in the biomass map generation and require pre-processing in the CCI Biomass project:

- Sentinel-1
- ALOS-1 PALSAR
- ALOS-2 PALSAR-2

ENVISAT ASAR C-band data have been used to produce the AGB map for 2010. The ASAR data have been pre-processed in the context of previous projects.

The radar data pre-processing covers all steps from Synthetic Aperture Radar (SAR) data intake to orthorectified calibrated backscatter images. This includes the preparation of auxiliary data needed for the processing, and sampling to a common reference grid etc.

Processing of the Sentinel-1 (S1) data is implemented to obtain calibrated backscatter images with sub-pixel co-registration accuracy, arranged in a structure that allows easy access and management.

To this end, a global grid tied to the geographic reference system and having its origin at 90° N, 180° W, with tiles of relatively small size is used. The S1 images (with an original spatial resolution of approximately 15 m) are geocoded to a pixel size of 1/720th of a degree in both latitude and longitude and tiled into 1° × 1° tiles (i.e., 720 × 720 pixels). The commercial software package by GAMMA Remote Sensing is used for the pre-processing of the S1 data. The flowchart of the S1 data pre-processing is shown in Figure 4. For the details, refer to the ATBD of the CCI Biomass project. All IWS (Interferometric Wide Swath) images in Ground Range Detected (GRD) format in the S1 archives for the years 2017 to 2020 are considered. The large number of images necessitates implementation of the S1 pre-processing on AWS, which is why the pre-processing is not considered to be part of the system developed for CCI Biomass.

ALOS-1 PALSAR and ALOS-2 PALSAR-2 datasets are available through the Japanese Aerospace Exploration Agency (JAXA), which releases fully pre-processed (i.e., calibrated, geocoded, tiled) datasets in a ready-to-use format. Global mosaics of the L-band backscatter at two polarizations acquired in Fine Beam Dual-polarization mode (FBD) are released on an annual basis. So far, FBD mosaics with a spatial resolution of ~25 m x 25m for the years 2007 to 2010 and 2015 to 2020 have been released. In addition, per-observation-cycle mosaics of ALOS-2 L-band backscatter acquired in two polarizations in ScanSAR mode have been released. Most of the per-cycle mosaics with a spatial resolution of ~50 m x 50 m cover the wet tropics. Up to now, available ScanSAR mosaics cover the period between 2016 and mid 2022.

Post-processing of the ALOS-2 PALSAR-2 data was therefore limited to aggregation to the CCI Biomass target resolution of 1/1125° (i.e., 0.00088889°x0.00088889°).

Since v3 of the CCI Biomass products, JAXA has released all individual ALOS-1 PALSAR Fine-Beam images in Level 2.2 format, i.e., images processed to radiometrically terrain-corrected level. Post-processing of these images was also limited to aggregation to the CCI Biomass target resolution and creation of tiles.

Through a JAXA-ESA collaboration, JAXA has provided exclusive access to all individual ALOS-2 PALSAR-2 Fine-Beam observations acquired between 2015 and 2021 in a slant-range detected format, from now on referred to as KC. Pre-processing of the imagery comprised topographic corrections, geocoding to the target pixel grid with 0.00088889° pixel posting, and tiling. Because of the large data volume, pre-processing was again performed on AWS.

For further details on the ALOS-1 PALSAR and ALOS-2 PALSAR-2 post-processing, refer to the CCI Biomass ATBD.

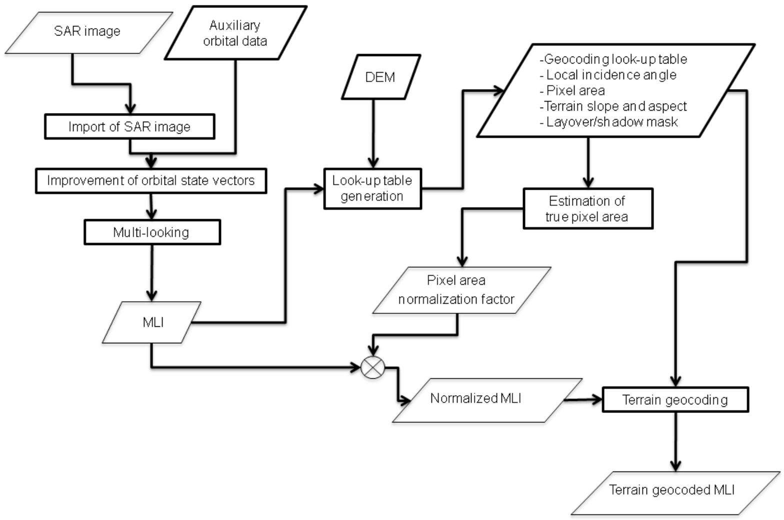

Figure 4: Flowchart of Sentinel-1 pre-processing.

3.3.2 Retrieval

In the retrieval process, the algorithms from the Algorithm Development Plan (ADP) (R-8) are applied. In the module, AGB and related standard errors are estimated. In Year 1 of the project, the structure of the AGB retrieval algorithm (**Error! Reference source not found.**) followed the general design of the retrieval algorithm implemented to produce the GlobBiomass map products for 2010, which is considered the baseline algorithm in CCI Biomass. In the following years of the project, the retrieval algorithms have been modified to better reflect structural differences of forests. Biome-specific adjustments have been considered to complement/improve upon what the baseline algorithm produces regionally. EO datasets and approaches with potential for large-scale mapping of biomass have been identified in the CCI Biomass Product Validation and Algorithm Selection (PVASR) document. Implementation, which must comply with the system design for the baseline approach, and validation of these approaches was pending at the time of writing this document and is therefore not addressed below.

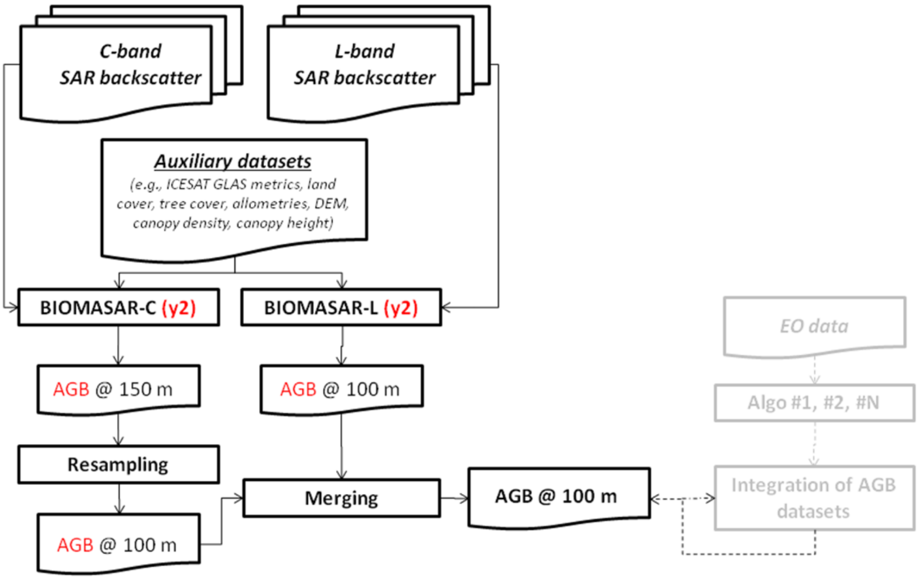

Figure 5: Flowchart of the CCI Biomass retrieval algorithm; note that the general workflow was not modified in year 3 of the project.

To best exploit the global EO datasets currently available, the retrieval is arranged in three stages (**Error! Reference source not found.**).

In the first stage, a global dataset of AGB is derived from the multi-temporal dataset of S1/ASAR backscatter images with a pixel spacing of ~150x150 m^2 using the BIOMASAR-C algorithm (Santoro et al. 2011, Santoro et al. 2015). To avoid confusion with an implementation of BIOMASAR using L-band data in the next stage, we use the acronym BIOMASAR-C for Stage 1.

The second stage for producing AGB maps is based on the BIOMASAR-C model calibration and inversion approach, adapted for dual polarization backscatter from ALOS-2. It is therefore referred to as BIOMASAR-L.

The third stage is the product generation stage described in Section 3.3.3 below.

The global AGB maps produced with BIOMASAR-C and BIOMASAR-L product are accompanied by maps depicting the pixel-level standard deviation of the estimates. Error models have been defined to characterize uncertainties associated with the backscatter measurements, the estimation of the model parameters for each image in the multi-temporal stack of observations, and the error associated with supporting datasets, such as ICESAT GLAS.

3.3.3 Product Generation

In the product generation step, the resulting maps are put in the delivery product format as requested and specified in the PSD (R-3). The product generation process includes the conversion of the maps to the product format specifications. This is the third stage shown in **Error! Reference source not found.**, in which the set of AGB estimates are merged to form the final estimate. The rules to do so are described in the ATBD for CCI Biomass. The final maps are provided in an equiangular map projection with WGS-84 as datum in Geotiff and netCDF formats.

3.3.4 Verification/Validation

The products undergo procedures to ensure product integrity and quality. This step must not be confused with the product validation process.

The system operators oversee system (module) verification. The verification is undertaken separately for the different PS modules. It covers:

- System (module) validation (after an upgrade or new installation)

The PS provides tools that facilitate these tasks such as

- Benchmark test data (BM-OPE-4060)
- Test tools (BM-OPE-4060)
- Verification Report (BM-RAM-5)

3.4 Algorithm Improvement

The development team decides about features or processes to be improved in order to meet user requirements. The development team implements the improvements as new versions of processors and if necessary, as a modified workflow definition. The development team tests and validates the new products. The development team also decides about new versions to be released.

An improvement cycle is defined as:

1. New requirements are identified and analysed
2. Modified processor implementation
3. Test production
4. Validation
5. Decision to a) go to 1, and iterate again, b) implement go on, c) stop here
6. Release a new version (code freeze) while retaining older ones for cross check
7. Start production

After each validation, a decision is taken about whether the improved algorithm is accepted, further development is needed or the development is stopped. Only in the case of acceptance does the development lead to a new version of the processor and full reprocessing of the archive or implementation for new data. This cycle of innovation and improvement typically takes 5-10 years. With an effective control mechanism in place, the update cycle can be reduced to a few years.

4 Functional Design

Herewe discuss and present the major functional blocks of the CCI Biomass PS.

4.1 Services

Most of the necessary services are provided through the CCI environment, such as the data exchange storage, the document management system and user interaction. For the PS development, a software repository and an issue tracker are hosted at GAMMA.

4.1.1 Processor software repository

An important element of the modern software development process is source control (or version control). Cooperating developers commit their changes incrementally to a common source repository, which allows them to collaborate on code without resorting to crude file-sharing techniques (shared drives, email). Source control tools track all prior versions of all files, allowing developers to "time travel" backward and forward in their software to determine when and where bugs are introduced. These tools also identify conflicting simultaneous modifications made by two (poorly communicating) team members, forcing them to work out the correct solution (rather than blindly overwriting one or the other original submission). The software repository contains the actual processing code and all prior versions. The write access to the processor repository is restricted to the development team. As all software changes are updated directly in the repository, the software changes are published almost immediately and are made available for review.

4.1.2 Issue Tracker

During software development a Redmine issue tracker is used (http://www.redmine.org/). Redmine is a flexible project management web application written using Ruby on the Rails framework. Redmine integrates the version control system into its user interface and manages the access control to the version control system resulting in a state-of-the-art Free and Open Source Software (FOSS) software development environment.

4.2 Processors

The processors cover the necessary tools to produce the different CCI Biomass products. CCI Biomass is organised in modules covering the production of the different products. In general, a distributed processing approach is followed. Consequently, the modules are portable.

Requirements:

→ BM-SIZ-0080 Flexible production

→ BM-FUN-5020 Data overwrite

Component	**Purpose**	**Content**	**Implementation**
AGB processor	Generate L2/L3 Biomass_cci AGB Product	Input Data	Heritage from GlobBiomass
AGB change processor	Generate L2/L3 Biomass_cci AGB Change Product		Biomass_cci

4.3 Concept for Continuous Improvement

Continuous improvement is an important aspect in the CCI projects. To ensure a transparent process Software Modularity, Software Version Concept, Version Control, and Version Numbering are important issues. This section defines the structures and functions that extend the production environment for continuous improvement. Focus is on flexibility, rapid testing and prototyping. The concepts described are processors, versioning and a test environment. The concept of processors and versioning contribute to the modularity of the system.

Requirements:

→ BM-OPE-4020 Development under version control

→ BM-OPE-4030 Decoupled from own research

The software of the PS and the processing algorithms code is under version control. The software repository contains the actual processing code and all prior versions. All software changes are documented in the repository. Version numbering of the processor is reflected in the repository by revisions and tags. Revisions are usually linked to committals and indicate the sequential order of documented changes. Tags are set to indicate software releases of frozen software states. Subversion is a good candidate for version control. Together with Redmine, it is a complete FOSS version control and issue tracking system.

Data processors help to organise the data processing in modules. Due to the differing input and output datasets/formats, the modules are normally not shared among products (even if the functionality is the same). A processor is a software component that can be parametrised and that generates a (higher level) output product of a certain type from one or several input products of certain types. A PS module consists of the sequential call of processors. Each processor has its own version information. Parameters are usually provided as command line arguments, environment variables or as information in a parameter file. Feedback is received by a return code, messages on STDOUT/STDERR and in log files.

Component	**Purpose**	**Content**	**Implementation**
Software Repository	Stores all versions of the processor code in a transparent way, with branches, authorship	Code	Subversion Redmine Tools

4.4 System Documentation

The system documentation contains the PS documentation consisting of manuals, specifications and reports, as well as the product documentation consisting of product specifications, manual and validation reports. At this stage, no advanced functionality such as collaborative editing etc. seems to be necessary so that the basic functionality of any FOSS content management system (CMS) might be sufficient for this task.

Requirements addressed by this section are:

→ BM-FUN-1070 Product Description

→ BM-INT-3010 Self-standing documentation

→ BM-RAM-5030 The system developed shall be detailed as a separate self-standing document

Component	Purpose	Content	Implementation
Documentation Management	Stores documentation in a structured and transparent way	Documentation	CCI Biomass Project Website at ESA

The PS documentation includes requirement documents, design and interface control documents, test documents, manuals, and maintenance information. CCI Biomass deliverables to name here are the ATBD and the PSD. The SRD and SSD define requirements and design of the system.

5 Development, Life Cycle, Cost and Performance

This section discusses the system development in the future, potential development strategies, efforts and costs. The development is driven by several factors such as the availability of new technology, faster algorithms, new scientific findings and improved product algorithms, new available EO data, and user needs.

5.1 Re-use and Development

Development is needed to bring the existing prototypes of the PS modules to a higher operational level satisfying the requirements listed in the previous sections and to add the missing components such as those for user services, data handling, life cycle management, archiving etc.

Requirements addressed by this section are:

→ BM-OPE-4040 User Requirements

→ BM-OPE-4030 Development decoupled from research

→ BM-OPE-4070 Freeze prototype

→ BM-FUN-5040 Verification of implementation

In the table below we summarise the tools that were used, adapted, configured and integrated during development of the PS within the CCI and beyond.

Name	Usage	Remarks
Subversion	Version control	FOSS

Redmine	Issue tracking	FOSS
Python3	Scripting, netcdf4 reader, pyresample, rasterio, gdal, numpy, scipy	FOSS
GAMMA ISP, DIFF&GEO	AWS S1 preprocessing	proprietary
Biomasar-C retrieval module	MATLAB	
Biomasar-L retrieval module	Python3	

5.2 System Life Cycle Drivers and Considerations

The PS needs to be incrementally adapted to integrate new functional extensions, improved algorithms and input datasets. New EO data make adaptations necessary and most likely also have an impact on the hardware infrastructure. The life cycle plan cannot be static as it is not foreseeable. Currently, the following driving factors are identified:

- Availability of the existing processor module prototypes
- Functional extension of the system
- New workflows
- Improved algorithms
- New Sensors
- Hardware improvements
- Dependencies on 3rd parties (other ECVs, data providers, new users)

To deal with the first two points, the system is initially based on the prototype. Incrementally, additional components and functions are added and interfaces to data providers and users are extended. The third and fourth point of workflow and algorithm development requires the addition of tools for validation and user feedback.

New sensors and increased data volume are a qualitative change, too. The existing methods need to be adapted to make use of new sensors. For the longer perspective, renewal of hardware and optional change of software layers must be considered. The PS design is prepared for this by the modularity of its functional components. The last item is not so relevant for CCI Biomass now as the dependence on other CCI projects is minor.

Requirements:

→ BM-FUN-1010 Long-term storage

→ BM-FUN-1020 Unique identifier

→ BM-FUN-1030 Structured storage

→ BM-FUN-1050 Reprocess Products

⟶ BM-FUN-1060 Partial Processing

⟶ BM-OPE-4020 PS shall be under version control

⟶ BM-RAM-4060 Test tools

⟶ BM-RAM-5040 Verification

5.3 Sizing and Performance Analysis

In the SRD, there are no specific requirements concerning the processing time performance. Now it is mainly labour hours that drive the processing rather than CPU core hours. Full reprocessing of historical data requires a variable amount of work, depending on the product (see tables below). The data storage budget for inputs and outputs for historical data and for the yearly increase of acquired data is in the low TB range initially but might grow to more substantial TB values with new sensors becoming available.

There exist requirements on disk space that are modest:

⟶ BM-SIZ-2060 Space for input data

⟶ BM-SIZ-2050 Space for auxiliary data

⟶ BM-SIZ-2040 Space for output data

⟶ BM-FUN-2030 Run on available hardware

⟶ BM-FUN-2010 Time series

Below the budgets for data storage and processing capabilities are estimated. The budget for data storage mainly depends on the amount of input data to be managed. This comprises historical data and data acquired continuously.

5.3.1 AGB module

We estimate the number of satellite scenes (Sentinel-1 and ALOS-2 PALSAR-2 coverage) required to map AGB.

Data	Product	Time Span	Historical Data	New Data
Sentinel-1	GRD Level 1c	2017-2020	3.1 TB	
ALOS-2	Mosaics FBD and ScanSAR, KC data	2017-2020	4 TB	
ALOS-1	Mosaics FBD and L2.2	2010	0.7 TB	
ENVISAT ASAR	ScanSAR	2010	6 TB	

5.4 Cost Estimation

Requirements addressed by this section are:

→ BM-OPE-0100 Min maintenance and cost

The costs for the PS are composed of costs for storage, processing, network, development and integration, operations, dissemination and labour.

The cost estimates per year in Euros are summarised in the following table. Costs from operational production of on-demand services move from the products to the user services after initial setup.

Product	**Component**	**Year 1**	**Year 2**	**Year 3**
AGB	Hardware	10k	10k	10k
	Development	44k	44k	44k
	Operations	5k	5k	5k
	External data pre-processing (AWS)	23k	23k	
	Total	82k	82k	59k

6 Requirements Traceability

id	Title	Ref
BM-FUN-0010	Develop and validate algorithms to approach the GCOS ECV and meet the wider requirements of the Climate Community (i.e. long term, consistent, stable, uncertainty-characterized) global satellite data products from multi- sensor data archives. (CR-1)	§2.3
BM-FUN-0020	Produce, validate and deliver consistent time series of multi-sensor global satellite ECV data products for climate science. (CR-2)	§2.3
BM-FUN-0030	The CCI+ Biomass system shall generate the following product:	§2.3

	Aboveground biomass maps @ 100 m resolution Maps @ 100 m resolution AGB difference maps for the three epochs considered in the project (spatial resolution to be defined).	
BM-FUN-0040	Generate and fully document a production system capable of processing and reprocessing the data, with the aim of supporting transfer to operational activities outside CCI (such as C3S). [CR-4]	§2.3
BM-OPE-0050	All project documentation shall be made publicly available via the CCI Open Data Portal: http://cci.esa.int.	§2.3
BM-OPE-0060	The PS shall capitalise on existing European assets through their reuse, particularly Open Source scientific tools and prototype ECV processing systems from prior projects. (heritage)	§2.3
BM-INT-0070	The global Biomass community shall play an active role in its creation according to given guidelines and advice from a strategic operations team. They shall also give feedback from the implementation to the strategic team.	§2.3
BM-SIZ-0080	The system shall implement a data production line that is sufficiently flexible to continuously update and extend the database (e.g. with data from new sensors or better acquisitions).	§2.3
BM-INT-0090	The available data shall be frequently reported and properly disseminated to the interested user communities.	§2.3
BM-OPE-0100	Minimum maintenance and cost	§2.3
BM-FUN-1020	The Products shall be uniquely identified.	§5.2
BM-FUN-1030	The PS shall store data in a structured way using type, revision, date.	§2.3
BM-FUN-1040	If input data is retrieved directly from a third-party ground segment, the PS has to ensure that links are maintained, and functionality is regularly checked.	§4.3
BM_FUN-1050	The PS shall also be able to reprocess parts of the products.	§5.2
BM-FUN-1060	The PS shall be able to do partial processing.	§5.2

BM-FUN-1070	Produce products according to the Product Description in the PSD	§2.3
SW-SIZ-2010	The PS shall be able to do processing in due time.	§5.3
BM-SIZ-2020	The PS shall be able to do reprocessing in due time.	§5.3
BM-SIZ-2030	The PS shall be able to run on the available hardware infrastructure	§2.7
PF-SIZ-2040	The PS shall have sufficient space for output data	§5.3
PF-SIZ-2050	The PS shall have sufficient space for auxiliary data	§5.3
PF-SIZ-2060	The PS shall have sufficient space for input data	§5.3
BM-INT-3010	The PS shall have a self-standing documentation	§4.4
BM-INT-3020	The PS shall have the capability and interfaces to extend for future adaptations.	§5.3
BM-OPE-4020	Development of the PS shall be under version control.	§5.1
BM-OPE-4030	The system should be decoupled from the research	§5.1
SW-OPE-4040	Development of the system shall be based on the user requirements, the selected algorithms and the developed standardized validation protocols.	§2.2
BM-OPE-4050	The PS development shall be overseen by a science team that drives the development process.	§3.1
BM-OPE-4060	Each PS installation includes a set of test tools, data and benchmark data to test PS integrity (end-to-end, interfaces)	§5.1
BM-OPE-4070	If a module is based on a prototype, the prototype state has to be frozen until it is implemented.	§5.1
BM-OPE-4080	The verification is regarded as successful, when all tests are within TBD limits. Hashes are to be preferred where applicable.	§3.3
BM-FUN-5020	The operational processor shall not overwrite existing data. Versioning shall be used instead.	§5.2
BM-RAM-5030	The system developed shall be detailed in a separate self-standing document providing an	§5.1

	overview of the system and its components, functionality of the system and its subsystems, inputs, outputs, resource key interfaces, and resource requirements.	
BM-FUN-5040	Verification of the correct implementation of the prototype system against the algorithms developed is a fundamental part of the process.	§2.3
BM-RAM-5050	The verification shall be documented in a Verification Report. It shall contain the chosen approach and the justification, the selected verification data set and the verification results.	§2.3
BM-RAM-5060	The PS shall provide means against data loss of its input / output products.	§2.3
BM-FUN-5070	All data stored in the system shall be available for the long-term (at least 15 years).	§2.3

7 References

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U. and Wiesmann, A. (2011). Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sensing of Environment*,115 (2), 490-507.

Santoro, M., Eriksson, L. E. B. and Fransson, J. E. S. (2015). Reviewing ALOS PALSAR backscatter observations for stem volume retrieval in Swedish forest. *Remote Sensing*,7 (4), 4290-4317.