# ESA Climate Change Initiative
# Aerosol_cci+

## Deliverable D3.1b

## System Specification Document

## Version 3.0

**Document reference: Aerosol_cci+_SSD_v3.0.pdf**

# DOCUMENT STATUS SHEET

| | FUNCTION | NAME | DATE | SIGNATURE |
|---|---|---|---|---|
| LEAD AUTHOR | editor | Thomas Popp | 03.03.2020<br>21.04.2020 | |
| CONTRIBUTING AUTHORS | | Peter North<br><br>Yves Govaerts<br>Marta Luffarelli | 10.03.2020<br>24.04.2020<br>13.11.2020<br>31.03.2021<br>06.12.2021 | |
| REVIEWED BY | Science leader | Thomas Popp | 10.03.2020<br>27.04.2020<br>17.01.2021<br>31.03.2021<br>07.12.2021 | |
| APPROVED BY | Technical officer (ESA) | Michael Eisinger | | |
| ISSUED BY | Project manager | Thomas Popp | | |

# EXECUTIVE SUMMARY

Within this document a system specification of the scientific Aerosol_cci+ system used to process the planned limited test datasets is presented. This document does not describe an over-all system for the creation and repeated (re-)processing of satellite data based aerosol Essential Climate Variables (time series), which are accepted by the scientific community and used for IPCC assessments, since such an operational system has been transferred to and is now under the responsibility of the Copernicus Climate Change Service (C3S). A decentralized operational Aerosol retrieval processing system is now operated in the scope of the Copernicus contract C3S_312b_Lot2 (Atmospheric composition, including aerosol properties).

The system concerned here comprises a technical implementation to allow and support limited test dataset creation with new algorithm versions for the purpose of their evaluation. In order to ensure the highest possible quality and acceptance in the community the overall principle for the whole system is the leadership by scientific expertise.

After the introduction (Section 1) this document starts with a definition of the purpose of the system (Section 2) to specify subsequently its context (Section 3). Tasks of the Science Team are specified in chapter 4. In chapter 5, the operational scenarios for the ECV production system are discussed. Chapter 6 describes the “Environmental Characteristics of the System” including concrete implementation aspects. A matching matrix with the system requirements is made in Section 7 to assure traceability.

# DOCUMENT CHANGE RECORD

| Issue | Date | Modified Items / Reason for Change |
|---|---|---|
| 0.9 | 03.03.2020 | Document structure and first draft |
| 1.0 | 10.03.2020 | Update by responsible for mature algorithm<br>Review by science leader |
| 1.1 | 27.04.2020 | Add description of SU processing system in section 6 (page 25) as requested by ESA RID |
| 2.0 | 13.11.2020 | Update by responsibles for second algorithm |
| 2.1 | 17.01.2021 | Science leader review, minor formal updates |
| 2.2 | 31.03.2021 | Minor revisions as requested by ESA |
| 3.0 | 07.12.2021 | Minor revisions for the final version of the ACCI+ system (section 6) |

# LIST OF FIGURES

# TABLE OF CONTENTS

# 1 INTRODUCTION

## 1.1 Scope

The *System Specification Document* incorporates the requirements described in the System Requirements Document and specifies the characteristics of a scientific ECV production system from a developer's point of view. Only those requirements of the System Requirements Document are tackled in the appropriate part of this document to show the system design answer to the requirement, which are relevant for the processing and evaluating of test datasets. Considerations for the design of an operational system as the one implemented by the C3S have been discussed in detail in [RD3].

## 1.2 References

### 1.2.1 Applicable Documents

[AD1] The Statement of Work, reference ESA-CCI-EOPS-PRGM-SOW-18-018, issue 1, revision 6, dated May 31st, 2018, and its specific annex C.

[AD2] The Contractor's Proposal reference 3022091 revision 1.1 , dated 10 December 2018

### 1.2.2 Reference Documents

[RD1] Aerosol_cci+ User Requirements Document, Issue 5.0, 23.09.2021

[RD2] Aerosol_cci System Requirements Document, Issue 1.6, 15.09.2011

[RD3] Aerosol_cci2 System Specification Document, Issue 3, version 1.2, 30.04.2017

### 1.2.3 Acronyms

The following list contains only system relevant acronyms - no satellite sensor or algorithm acronyms.

| | |
|---|---|
| ADP | Algorithm Development Plan |
| AEROCOM | AERosol mOdel inter-COMparison initiative |
| ATBD | Algorithm Theoretical Base Document |
| CCI | Climate Change Initiative |
| DLR | Deutsches Zentrum für Luft- und Raumfahrt e.V. |
| ECMWF | European Center for Medium Range Weather Forecast |
| ECSS | European Cooperation for Space Standardization |
| ECV | Essential Climate Variable |
| FMI | Finnish Meteorological Institute |
| GCOS | Global Climate Observing System |
| I/O | Input / Output |
| ICARE | Thematic centre of CNES, CNRS, Univ. of Lille, holds satellite aerosol data |
| Level1 | calibrated satellite measurements in orbit projection |
| Level2 | derived satellite geophysical parameters in orbit projection |
| Level3 | geophysical parameters in gridded product |
| MPI | Max-Planck-Institut für Meteorologie |
| PVIR | Product Validation and Inter-comparison report |
| RF | Rayference |
| SRD | System Requirements Document |
| SSD | System Specification Document |
| SU | Swansea University |
| URD | User Requirements Document |
| WMO | World Meteorological Organisation |

# 2 PURPOSE OF THE SYSTEM

The purpose of this operational ECV System is to produce and repeatedly reprocess limited test datasets of satellite-based Essential Climate Variable time series of highest quality and credibility. The system specified in this document is based on a System Requirements analysis [RD2], which was derived from the User Requirements [RD1] of the Aerosol_cci project with its horizon encompassing the next 15 years (SR-OP-0320). The part of the system concerned here is the one which applies to improving algorithms to create and evaluate new updated versions in order to better meet the underlying user requirements. At this stage, data retrieved from dual view radiometers of ESA serve as a source for aerosol ECV products: (A)ATSR(-2), Sentinel-3A/-3b SLSTR.

The test Aerosol_cci ECV production system intends to provide an infrastructure for the creation of Aerosol_cci+ products of multi-spectral Aerosol Optical (AOD) and Fine Mode AOD. The test Aerosol_cci+ production system consists of two major components: a technical implementation to allow the required production and reprocessing of the datasets, and a coordination mechanism to ensure that the Aerosol_cci system is under permanent leadership of scientific experts. The close interaction of these two essential components assures the production of highest quality and credible products, which are accepted by the science community (SR-GE-0050).

The decentralized Aerosol_cci retrieval und production system has been transferred to the Copernicus Climate Change Service (C3S) project 312b_ Lot2 (Atmospheric Composition ECV production, including aerosol properties).

# 3 CONTEXT OF THE SYSTEM

## 3.1 System Overview

In Figure 3-1 the Aerosol_cci system is depicted with all internal and external elements (within the blue box). Within the Aerosol_cci system box technical entities are depicted in blue, and coordination elements are depicted in black. The Figure 3-1 is taken from [RD3].

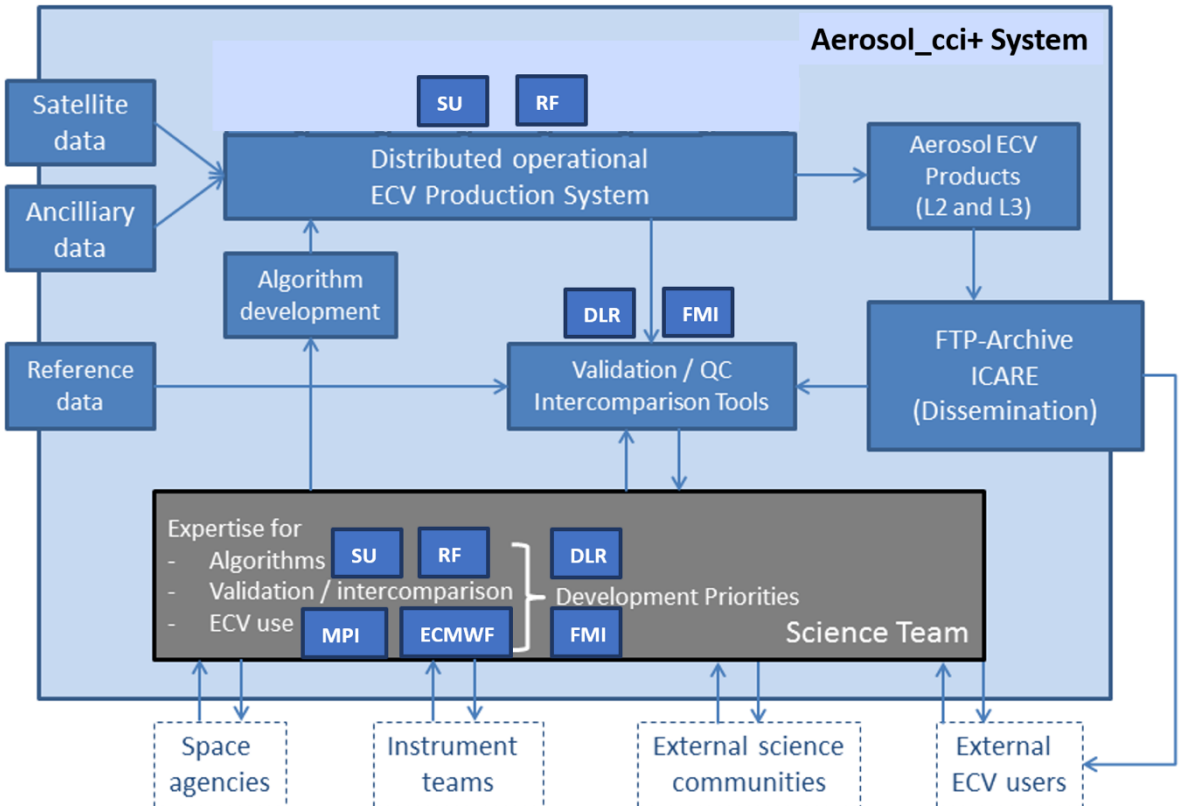


**Figure 3-1: Distributed Aerosol_cci system and its context (from [RD3])**

The **operational distributed ECV production system** consists firstly of hardware for the fast, complete and repeated reprocessing of large datasets. Secondly, it includes **archiving** functionality for the input data (**satellite and ancillary datasets**), for the **Aerosol ECV products** and for **reference datasets** used for validation and inter-comparison (ground-based datasets and external aerosol ECV datasets). Furthermore, it includes **standard tools** to support repeated testing and validation of new algorithms (SR-GE-0040) and capabilities for the **dissemination** of products and **documentation** to the users (SR-GE-0130). The distributed system is able to handle all relevant input data together with required external reference data. A **Science Team** assures scientific fitness for purpose of the aerosol ECV products. Based on the joint expertise of **climate users**, **validation and inter-comparison experts** and **algorithm developers** its key responsibility is the definition of **algorithm development priorities** to steer the operational product evolution (SR-GE-0041). To fulfil this responsibility the science team coordinates external links with **novel science**, **climate users**, **instrument teams** and **space agencies**. A continuous **algorithm development** task implements the priorities defined by the science team in order to ensure

product evolution to better meet end user needs and responding to their requests for improvements (SR-GE-0080). **Validation and quality control** for each new ECV dataset determines the error characteristics by comparison to external reference datasets (ground-based). In order to document the competitiveness or complementarity of the new ECV products, their statistical error characteristics must be related to other state of the art aerosol satellite datasets by inter-comparison with them (SR-DO-0160). In order to assure credibility, validation of the products and the uncertainty estimates provided as part of the products is conducted by experts who are independent from the algorithm developers (in line with SR-GE-0042). The requirements for the different actors and interactions within this system are described in more detail in [RD2].

Besides the technical elements purely dedicated to the automated creation of ECV datasets strong organisational interfaces and interactions are implemented to guarantee quality and credibility of the created products (SR-GE-0041) as well as give the user community the opportunity to influence development processes and processing priorities (SR-MA-230).

The project team is convinced that the best coordination mechanism to ensure the scientific quality of the ECV products is guaranteed through a science team with dedicated expertise for aerosol retrieval algorithms, validation and use of aerosol ECV products. This therefore is an answer to the requirements for both, internal coordination (SR-GE-0100) and external coordination (SR-GE-0110). As the science team organises the coordination mechanisms themselves these mechanisms can be flexible (SR-GE-0120). The science team also coordinates communication with other projects to maximise consistency of ancillary data used by other projects (SR-FU-0150). This is also the most efficient way to assure swift and continuous implementation of cutting edge research results such as new algorithms and upgrades (SR-QU-0180).

Due to this nature of the system we have to define the system interfaces in three layers:

- The systems **external interfaces**, where system means the whole structure depicted within the blue box in Figure 3-1.

- The **internal interfaces** within the system between the operational ECV production system and the other (coordination) elements within the blue box in Figure 3-1

- The **interface with the operational ECV production system**

In the scope of the **Copernicus Climate Change Service (C3S)** project C3S_312b_Lot2, the decentralized Aerosol retrieval and processing system is further developed. DLR hosts the machine to machine interface towards the Copernicus Climate Data Store for aerosol (and also ozone and greenhouse gas) datasets while ICARE provides the ftp server for collecting all aerosol datasets and FMI operates a backup service for DLR's interface. The Climate Data Store provides data access to users to the full mission ECV products (operational external interface). In order to maintain one integrated data collection of all historic Aerosol_cci datasets, the new test datasets under Aerosol_cci+ and the full mission data records of C3S, ICARE ftp continues to archive all aerosol data under those three projects.

In the scope of the **CCI+ program**, the limited test datasets processed within Aerosol_cci+ are archived and provided to users (test external interface). Internally, we will continue using the ICARE ftp archive (which is now one component of the C3S_312b_Lot2).

Within the Aerosol_cci+ system the science team is responsible for the algorithm development. This means it converts the validation and user feedback into evaluated development priorities (SR-OP-0480) regarding algorithm development. To achieve this goal and thus to fulfil SR-GE-0042 (scientific credibility) and SR-GE-0050 (science team composition) the science team has to be composed of algorithm experts, independent validation experts (SR-OP-0470) and ECV use experts.

Another internal interface of the science team is the external reference input needed for repeated quality checking, validation and inter-comparison of ECV products and thus fulfil requirement SR-GE-0040. All documents issued by the science team describing the input, the products, their validation and their utilization are stored and distributed.

One key interface is the one with the operational processing system of C3S. This is achieved as far as possible by joint leadership (coordinated requirements, algorithm evolution, evaluation) and joint partnership (using new algorithm developments in Aerosol_cci+ for the next reprocessing in C3S). In cases of disjunct partnership, new Aerosol_cci+ algorithm developments can only be transferred to C3S within the next ITT.

**Algorithm development**

The algorithm development function of the Aerosol_cci system has to interface with the operational ECV production system to integrate new algorithm implementations into the operational ECV production system. The major benefit is that algorithm development can focus on the core retrieval issues since it can fully rely on the operational production system for handling of all input and output datasets, formats, metadata, etc.

The algorithm development task receives its priorities from the science team which has iterated these between user needs, validation results and algorithm possibilities. Algorithm development is then conducted by EO experts within the science team.

**Validation / QA**

Supervision and interpretation of the statistical analysis performed for validation is conducted by the validation experts in the science team and is supported by standard tools. The output of validation and inter-comparison exercises is delivered to the entire science team.

The validation task has interfaces to the archive (access the ECV products and reference datasets), to the operational ECV production system (standard tools for statistical validation), and to the science team (delivery of validation reports and recommendations).

# 4 TASKS OF THE SCIENCE TEAM

In Aerosol_cci+, a science team was established with the main aim to

> **Assure scientific leadership of ECV development and processing in order to achieve highest quality and credibility of aerosol ECVs in the climate scientific community.**

## 4.1 Fulfilment of the tasks

In order to fulfil the functions described in [RD3] the science team uses an efficient but comprehensive means of work. We consider as best suited an organisation with

- a science team leader,
- a flexible team of scientific experts to conduct scientific-technical work,
- space agency representatives as observers

The science team leader (part time to assure personal involvement with related scientific work and rotated among the leading members of the science community) manages all practical tasks conducted by the science team members (coordination between scientific experts and the operational system, coordination and quality assurance of documentation, preparation of decisions, representation at relevant user communities, interaction with space agencies, cal/val teams, international retrieval teams and other ECVs).

The science team assembles several times a year. For the time of their participation science team members concentrate on the relevant tasks (e.g. for the life cycle of their algorithm or validation source). The entire science team works as one team and has to endorse major decisions (algorithm development priorities, ECV validation acceptance). The science team must also include international experts.

To conduct scientific work, the science team must comprise of experts from three sub domains focusing on algorithm development, product validation and user interaction.

- Algorithm experts contribute on case-by-case basis (for dedicated limited time and specific task) regarding the priority development plan endorsed by the science team.
- Validation experts conduct validation and inter-comparison tasks on a long-term basis to ensure stability of quality control and reference datasets (possibly partly replaced every few years).
- User interaction includes updating user requirements every two years and test utilization and endorsement of new ECV products in conjunction with climate models and / or in collaboration with other ECV teams.

The science team bases its work on the following terms of reference:

1. **To define algorithm development priorities** the science team shall prepare an annual algorithm development (priority) plan based on newly arising user requirements (summarized into an updated user requirements document by the user experts), latest ECV product validation and assessment results (summarized in validation and assessment reports by the validation and user experts) and proposed new or improved algorithms (new or updated algorithm theoretical

baseline documents prepared by the algorithm experts or external retrieval scientists) after consultation with the entire science team and of recent scientific literature. This priority plan is discussed, altered where deemed necessary and endorsed by the whole science team at an annual review meeting and implemented afterwards by the science team. It is the overarching goal of this task to base algorithm development priorities on consensus within the international aerosol retrieval community – however focusing on European instruments. In order to meet SR-OP-0170, a schedule on the availability of ECVs must be published (SR-OP-0170). This activity intends to foster the ECV's use.

**To formally approve the quality level of new ECV datasets** the science team will conduct validation of the new datasets by the related experts, who can rely on automatic tools for extraction of validation statistics to derive and judge the respective validation (validation report). Additionally test utilization may be organized. The science team will review at its annual review meetings these validation and assessment reports for any new ECV dataset and prepare a release note for this new dataset which summarises in condensed manner (typically 1 page) the overall validation results, the major use demonstrated, the added value over earlier datasets and recommendations and limitations for the use of the dataset. Experts for product validation and inter-comparison shall be consulted to validate ECV product quality (SR-OP-0470).

2. **To assure the use of state-of-the-art algorithms, input data and auxiliary data** the science team leader and the scientific experts shall rely on several means

    i. be involved in aerosol satellite retrieval activities which will allow them to regularly screen novel related science

    ii. actively participate to relevant conferences and organize specific workshops to discuss potentially beneficial improvements of algorithms and ECV products

    iii. actively contribute to international coordination (retrieval and climate science communities, space agencies, international scientific programs and bodies such as AEROSAT, AEROCOM, ICAP, GCOS, WMO, etc.) via bi-lateral exchange and conference participation

    iv. to assure the contribution of Aerosol_cci to the international scientific dialogue by maintaining the communication mechanisms (web portal, newsletter, training courses and materials, etc.)

    v. pro-actively seek continuous flow of knowledge with instrument cal/val teams to assure proper understanding of calibration uncertainties and their improvements

    vi. collaborate with validation data originators (in situ, airborne) in order to understand their limitations and assure that all necessary external reference data are available to allow proper validation and improvement of algorithm parameterisations

    vii. collaborate with the aerosol climate model community to promote the use and assure proper use of aerosol ECV datasets and obtain feedback on their added value, limitations and required improvements

viii. use the relevant information and recommendations for updating the underlying specification documents (data access requirements, product specifications) which will then feed into the algorithm development priority plan with one particular focus on identifying gaps and possible solutions (e.g. new parameters, regional, …)

ix. To contribute to definition of mission requirements in order to assure the long-term continuity of (European) aerosol ECV measurements

x. To promote open data access for any input and auxiliary datasets needed and negotiate data access where necessary

xi. To act as an interface to the outside world for open and transparent communication of issues and potential solutions in order to gradually reduce uncertainties and errors of the ECV datasets

3. **To lead the data processing system by scientific needs** the science team leader consulted by the experts will define any necessary upgrades of hardware, technical software for ECV processing, tools for validation, dissemination, and visualisation and with regard to new algorithms or input data.

## 4.2 Concrete implementation for Aerosol_cci+

During the ESA-funded Aerosol_cci+ project (2019-2022), the science team as defined in this section is realized by the project consortium partners. Annual review meetings backed up with the associated deliverable documents (user requirements, validation and assessment reports, annual algorithm development report and plan, …) provide the concrete implementation of the priority decisions. The project science leader acts as science team leader who organizes the internal and external coordination defined for the science team. The international embedding of the project-internal science team is achieved by AEROSAT, the open International Satellite Aerosol Science Network, which is co-chaired by the Aerosol_cci science leader and a NASA colleague (recently also a new deputy co-chair has been appointed from the Aerosol_cci+ team). AEROSAT holds annual workshops and disseminates its discussions and conclusions through its website (aero-sat.org). Through their leading activities in AEROSAT Aerosol_cci+ team members take part in the international discussions on retrieval improvements and new developments and they expose the Aerosol_cci+ results to feedback in this forum. Where appropriate, team members of Aerosol_cci+ act jointly with AEROSAT working groups (e.g. recently within 2 focused collaboration activities which led to 2 distinct peer-reviewed publications) or other CCI project groups (e.g. at CCI colocation or integration meetings, leading role in joint CCI publications). Through participation to science conferences and the organisation of user workshops Aerosol_cci+ strengthens up its feedback loops.

## 4.3 Prioritising and Handling Updates

The data processing system has to be able to implement new or revised algorithms (or modules) following the agreement on algorithm development priorities coordinated within the science team between end users, validation experts and algorithm experts.

In order to facilitate this implementation the algorithm experts will have access to (modular) algorithm code implemented into the operational system so that they can easily make their changes and the subsequent re-implementation can be facilitated easily. Before re-implementation the science team conducts sensitivity studies, algorithm experiments and validation studies with a limited amount of data outside the operational system. Validation and inter-comparison experts shall have the possibility to request reprocessing of a specific time series of an ECV product to validate results of an algorithmic change (SR-OP-0400).

Following the first implementation verification within the data processing system of a data amount agreed to be sufficient within the science team has to be made.

One of the main requirements to the Aerosol_cci system as well as to all other CCI systems is a strong link to the science community. Therefore, the operational scenarios include scenarios for changes in algorithms or addition of new algorithms, and the addition of new input data.

### 4.3.1 Scenario Agreement on priorities for algorithm development

The process of identification of priorities for algorithm development is a key necessity to ensure cutting edge research addressing the user needs (SR-MA-0240, SR-MA-0250). These needs are rapidly identified and implemented, to focus science community algorithm development expertise on ECV requirements, and to justify funding for algorithm development – this implements SR-QU-0180.

To support an easy exchangeability of processing algorithms the system has to be modular in a way, which allows the exchange of separate modules in a processing algorithm – this is also an answer to SR-GE-0080, SR-OP-0420 and SR-MA-0250.

Starting from the validation of existing ECV products disseminated to the user community user feedback is obtained on existing ECV products, their validation results, and new ECV product requirements from the wider user community by the Science Team User Experts. On the basis of this user feedback, the Science Team Algorithm Experts identify potentially feasible areas of algorithm development, based on *e.g.* a knowledge of the potential information content of the satellite sensor data, known algorithm weaknesses, *etc.* (*e.g.* opportunity to use sun glint for retrieval of absorption properties, use of IR channels for better characterisation of desert dust, improvement of AATSR dual view collocation for improved retrieval performance, improvements to uncertainty characterisation, improvements to cloud masking, *etc.*). The Science Team produces an annual report on algorithm development priorities, including requirements for improved space agency (and ancillary data) products. This annual report is then disseminated to Space Agencies (formal request and prioritisation of improvements to standard products) and the EO science community (help direct and justify proposals for algorithm development funding). The Science Team then reviews new algorithm developments appearing in peer reviewed publications, and selects those which address documented high priority user needs and performs a trade-off analysis to identify the most promising or urgent algorithm upgrades, taking into account their potential impact, technical maturity, feasibility and cost. Finally, the Science Team recommends these priorities for implementation – this implements the requirement SR-DO-374 for a priority development plan based on broad scientific consensus.

| | |
|---|---|
| **Scenario Name:** | **Agreement on priorities for algorithm development** |
| **Start-up Preconditions** | **New ECV requirement or new algorithm or module available** |
| **Actors** | **Space_agencies, EO_science_community**<br>**Science_team in all its capacities:**<br>**Climate_research_ECV_users,**<br>**Algorithm_experts,_**<br>**Validation_and_intercomparison_experts** |
| | Science_team Validation_and_intercomparison_experts perform validation of existing ECV products and provide these results to the user community<br>Deliverable: Product Validation and Intercomparison Report (PVIR) |
| | User feedback on existing ECV products, their validation results, and new ECV product requirements is collected from the wider user community by the Science_team Climate_research_ECV_users<br>Deliverable: User Requirements Document (URD) |
| | The Science_team Algorithm_experts identify potentially feasible areas of algorithm development.<br>Input: URD, knowledge of the potential information content of the satellite sensor data, known algorithm weaknesses |
| | Science_team produces an annual report on algorithm development priorities, including requirements for improved space agency (and ancillary data) products<br>Deliverable: Algorithm Development Plan (ADP) – chapter 1. |
| | Annual report is disseminated to:<br>(a) Space_agencies to formally request and prioritise improvements to standard products,<br>(b) EO_science_community to help direct and justify proposals for algorithm development funding. |
| | The Science_team annually reviews new algorithm developments appearing in peer reviewed publications, and selects those which address documented high priority user needs. |
| | The Science_team performs a trade-off analysis to identify the most promising or urgent algorithm upgrades.<br>Criteria applied to judge proposed algorithm upgrades in the trade-off analysis: potential impact, technical maturity, feasibility and cost.<br>Deliverable: Algorithm Development Plan (ADP) - chapter 2 |
| | In conclusion the Science_team recommends these priorities for implementation. For each identified change, upgrade or new development task the Science_team states a specified goal (e.g. improved accuracy, coverage, new output variable, …)<br>Deliverable: Algorithm Development Plan (ADP) - chapter 3. |
| **Result** | **Work plan for development and testing of new algorithm or module is agreed upon**<br>**A proposal for funding is submitted to respective funding source** |

### 4.3.2 Scenario Prototyping and Implementation of new algorithm or module

Once an algorithm development plan has been worked out in the previous scenario and the funding for the activity has been secured the science team members (possibly extended by new relevant experts needed in this work) conduct all necessary steps to prototype and implement the new algorithm(s) or module(s). The algorithm experts start with coding the new algorithm based on its underlying scientific

publication and scientifically testing it with a small amount of data. This step includes acquisition of any needed new ancillary data or validation reference datasets. Once this scientific case study code is regarded ready it is implemented into the ECV production system by the processor experts together with support by the algorithm experts as new prototype (separated from the existing operational algorithms) and used for production of a substantial data amount which allows statistically significant evaluation. Then the automatic standard tools for statistical analysis are invoked to prepare the statistical analysis for validation to ground-based reference datasets and for inter-comparison to other datasets (satellites, models). On the basis of this validation output the science team validation and inter-comparison experts review the results and the entire science team draws a final conclusion which takes into account statistical validation / inter-comparison quantities, coverage achieved and spatial and temporal correlation with observed aerosol features. By comparing the new products with the existing operational products the increment is assessed and compared to the underlying user requirements and the specified improvement goal which had been defined in the algorithm development plan - this is also an answer to SR-DO-0160. When the new product satisfies the specified goal its prototype implementation is switched free for operational production and the respective (re-)processing and subsequent validation of the new ECV dataset are initiated. The sequence of these steps has to be described within a manual (SR-DO-0440).

| | |
|---|---|
| **Scenario Name:** | **Prototyping and implementation of new algorithms or modules** |
| **Start-up Preconditions** | **Science_team has agreed on new algorithm or module to be developed and implemented**<br>**Funding for this activity is secured** |
| **Actors** | **Science_team: Algorithm_experts, Validation_and_intercomparison_ experts, Climate_research_ECV_users**<br>**Operational_ECV_production_system, Processor_experts** |
| | The Algorithm_experts code the new algorithm and test it scientifically with a small amount of data.<br>Acquisition of any needed new ancillary data or validation reference datasets is made.<br>Input: ADP, scientific publications describing new algorithm |
| | The new algorithm / module is implemented into the operational_ECV_production_system by the Processor_experts together with the algorithm experts as new prototype (separated from the existing operational algorithms). |
| | The prototype is used for production of a substantial data amount which allows statistically significant evaluation. |
| | The automatic standard tools for statistical analysis are invoked to prepare the statistical analysis for validation to ground-based reference datasets and for inter-comparison to other datasets (satellites, models). |
| | The Science_team Validation_and_intercomparison_experts review the results and the entire Science_team draws a final conclusion on the value of the new algorithm which takes into account statistical validation / inter-comparison quantities, coverage achieved and spatial and temporal correlation with observed aerosol features.<br>By comparing the new products with the existing operational products the increment is assessed and compared to the underlying user requirements and the specified improvement goal which had been defined in the algorithm development plan. |
| | When the new product satisfies the specified goal its prototype implementation is switched free for operational production and the respective (re-)processing and subsequent validation of the new ECV dataset can be initiated (see operational (re-)processing scenario which is then invoked). |
| **Result** | **New algorithm or module is included into the operational_ECV_production_system** |

### 4.3.3 Scenario Addition_Of_New_Input_Data

If a new satellite mission produces new relevant data for aerosol ECVs this data has to be included into the Aerosol_cci system (SR-OP-0430). To facilitate this step easily the algorithm implementation has to be modular to allow swift replacement of sensor specific pre-processing tasks (SR-FU-0141). Trigger of this scenario will be an input of the Climate research ECV users within the Science Team who require the usage of these new data (SR-MA-0240). The Algorithm experts will develop or adjust the affected modules of an appropriate algorithm which will be implemented by the system experts. The workflow rule sets of the Aerosol_cci system are then extended to include the new input data and the matching processor into the workflows. Also the new sensor input products have to be added to the Aerosol_cci system inputs as well as the ancillary data needed by the algorithm implementation.

# 5 SCENARIOS FOR TEST PRODUCTION SYSTEM

## 5.1 Processing

**Functions**

Production Control: Production Control handles the overall workflows for the production of all products and generates processing assignments for the processing workflow management.

Processing Workflow Management: This function receives assignments for product generation from the production control function. It then allocates processing resources, fetches needed input products and starts an appropriate processor. It includes an intelligent mechanism to determine daughter products dependant on a new input dataset and assure reprocessing only of these (SR-GE-0030) After processing it hands back control to the production control. Intermediate products will be used for the next processing steps if they have been stored earlier. Final products will be archived. For archiving of the intermediate products two alternatives exist. They can be stored in the product archive (then allowing also distribution of these to users for scientific analysis) but this would increase I/O operations between processor and archive and archive size. Or they could be stored in a separate storage under control of the processor thus increasing I/O efficiency. In the latter case the sequence of product processing may be affected. However, the current principle of level2 retrieval algorithms is operation on units of complete orbits, so that independence between the major processing units is given. For level3 products based on ECV level2 products the dependence on availability of all level2 products remains for both alternatives. Thus the choice made is for storing intermediate products under control of the processors (not a permanent archive) and archiving a limited amount of intermediate products defined by the science team in the product archive (see 4.1.4).

Processor: This component is the implementation of a scientific algorithm which creates ECV products. It also can be a tailoring or quality control processor.

The system must be able to perform a processing of the complete Aerosol ECV products (SR-FU-0360).

The repeated ingestion of new and improved input data into the aerosol_CCI production system raises the need for a frequent reprocessing of data which has to be organized (SR-OP-0390).

The production control has to contain information on inter-dependencies between products and processors in order to assure complete reprocessing (if ordered) of all dependant products when one processor or one input dataset is updated (SR-GE-0030).

To achieve a minimal need of processing during an update-reprocessing the system has to keep a selection of the intermediate products available (SR-OP-0380). Here a trade-off between archive costs and flexibility needs to be made, which means that not all intermediate products can be stored permanently. (The System shall keep intermediate products of the most recent ECV products: SR-OP-380).

Validation and inter-comparison experts shall have the possibility to request reprocessing of a specific time series of an ECV product to validate results of an algorithmic change (SR-OP-0400).

**Interfaces**

Options for processor integration depend on the processor's implementation:

Direct inclusion into workflow management can be achieved if the processor is implemented in the same language as the workflow management system. Then, parameter handover will be handled directly. Implementation of that option has to take special care for error handling as the processor implementation could directly interfere with the workflow management system. This would increase the efforts for re-coding of algorithms.

If a processor is delivered as an executable (preferably static linked) package the processor has to be executed by the workflow management system through a system call. Handover of input product locations and processing parameters can be handled through command line parameters (possible problem if a high number of parameters has to be handed over) or through an options file.

The last option is to be preferred, as it is much easier for an operator to track back any errors.

For handing back results the options file can define a "result file" where the processor can write back result parameters.

"The aerosol_cci-System shall perform an automatic quality check, inter-comparison and validation against available validation tools for each ECV product which is created", SR-QU-370: Validation of produced ECVs is performed within workpackages 3100 ("Level 2 validation"), 3200 ("Level 3 validation and scoring against AERONET") and 3300 ("Uncertainty validation").

## 5.2 Validation and Intercomparison

The validation component consists of independent lines, each using their specific validation module and applying a work sharing (e.g. within CCI+: DLR focusing on lv2, FMI focusing on comparisons to other satellite aerosol products; within C3S: MetNo focusing on lv3). All relevant reference data are available at the validation modules: mainly AERONET (MAN) sun photometer multi-spectral AOD, NDACC lidar stratospheric profiles and derived products, but also inter-comparison data from other satellites / algorithms such as MODIS, MISR, SEAWIFS, OSIRIS, and model data from AEROCOM median or CAMS. The Aerosol_cci ECV products are downloaded from the ICARE server, where all EO partners store their products. The validation output is integrated into the validation reports and disseminated via the CCI Open Data Portal.

## 5.3 Archiving

**Functions**

Archive Control: The archiving control interfaces with other entities of the operational ECV system to archive and deliver all products, to search products and to change metadata.

Ingestion: The ingestion function interfaces with external entities to ingest new products (satellite input, ancillary input, and reference datasets) into the operational ECV production system (SR-FU-0290, SR-FU-0300).

Archive: The archive stores mass data. This includes all input data of the operational ECV production system as well as all output data (SR-OP-0270, SR-OP-0280).

In the current implementation, intermediate products are not archived at all since their amount and diversity is large. However, if they were archived it would facilitate their validation (as is seen in the current experiments this can be highly relevant). It is thus concluded that the system design should allow archiving of a limited amount of intermediate data for a limited time period (where the science team puts a focus on assessing a certain intermediate module output). It is not planned to archive all intermediate data of all product versions in order to limit the archive size. For optimizing I/O efficiency a separate storage (not a permanent archive) is used for intermediate products during processing.

A further point to be assured is that the hard disks used during data processing (part of the operational ECV production system) should be separate from the archive used for serving the ECV products to the users. The user ftp archive needs to be a big disk without high performance requirements. The processing disks need to be fast, but only sized to store a fraction of the total input, intermediate and output data e.g. one orbit, one day or one month of data at a time. Using the same physical disks would enlarge costs (a very large set of hard disks of very high performance) and user access speed would be compromised whenever the products are being reprocessed.

Metadata: Metadata of input and output products are stored in a metadata database.

**Interfaces**

Archive Product: This interface of the archive control allows internal entities to hand over products including meta information to the archiving function.

Change Metadata: By using this interface meta information connected to a product can be changed.

Get Product: Internal entities can get products from the archive by using this interface.

Product: This archiving-internal interface hands over products between the archive and the archive control.

Product (Metadata): This archiving internal interface hands over metadata between archiving control and archive.

Furthermore, related documentation specified in [RD3] needs to be available and again to simplify work with the products needs to be linked to the products through their metadata. To facilitate and promote use of new ECV data sets, a short overview release note (SR-DO-0341) highlighting new features and added value together with limitations and recommendations for the product use shall be provided together with a more detailed product user guide (SR-DO-0342). Also the Algorithm Theoretical Baseline Document - ATBD shall be referenced within each product (SR-DO-0371, SR-QU-0450).

It is an absolute must that all level2 and all level3 data are constantly available via an online ftp access or from a web catalogue service. Also appropriate collections must be available “via one click” (e.g. the entire level3 daily products of one ECV for one year in one zip archive or all years of level3 monthly data in another zip archive).

In support of process studies a tailoring mechanism (to extract regional sub samples from global products) must be available (either for the operator or in the user interface – depending on the frequency of this utilization type). If this type of use becomes more frequent a shopping basket shall be offered for access to individual files, where a user can select a set of individual products.

For access to documentation an online content management system should be provided which offers transparent access to version controlled documents. Respective documents must be linked from each product (ATBD, release note, validation report, and metadata).

Both, documents and product catalogue / simple ftp access have to be directly available from a web portal, which offers also the possibilities to announce new product releases and disseminate user feedback and publications evaluating or exploiting the ECV datasets.

Another functionality of the user interface should be the online access to tools for e.g. manual visualisation, data inter-comparison, statistical analysis, difference plots, seasonal / annual averaging, or regional averaging as they are currently available from the AEROCOM and ICARE browser tools (SR-FU-0191). The required functionalities for these tools need to be defined in more detail and associated costs for their implementation need to be assessed.

# 6 ENVIRONMENTAL CHARACTERISTICS OF THE SYSTEM

## Concrete Implementation of Aerosol_cci+

This section gives an overview of the current status of the major system components operated and maintained by the partners. Since the envisaged data volumes are comparatively small (comprising at project end one year of global data per sensor), the hardware specification is regarded uncritical.

| | | |
|---|---|---|
| **ESA / DLR** | Operating Content Management System (https://climate.esa.int/en/projects/aerosol/) | - The ownership and hosting platform of domain name (https://climate.esa.int/) is by ESA<br>- DLR manages the content based on interim results of the project |
| **ICARE** | Data archive service | - ICARE operates the internal aerosol ftp server for C3S_312b_Lot2; within this context the (comparatively) small data archive for the Aerosol_cci+ test datasets is also held in order to keep Aerosol_cci heritage and current Aerosol_cci+ data together and in close link with the C3S contract.<br>- If needed DLR can also provide a “project-owned” ftp / https archive for the Aerosol_cci+ datasets |
| **ESA** | Data access service | - Selected limited test datasets can be uploaded to the CCI Open Data Portal as far as appropriate; since generally users are requesting operational datasets (consistent full mission records, rather recent time series extensions) the main interest of users goes to the C3S Climate Data Store (CDS)<br>- The CCI Open Data Portal could hold datasets which are |

| | | |
|---|---|---|
| | | not provided to users at the CDS (e.g. level2 datasets) |
| **SU / RF** | Test dataset processing component | - SU and RF provide computing facilities to enable repeated processing and quality control of test datasets from the 4 dual view sensors ATSR-2, AATSR, SLSTR (Sentinel-3A and -3B); they acquire the needed level1b input data and any further required auxiliary input data |
| **DLR / FMI** | Evaluation component (validation, inter-comparison) | - DLR and FMI provide computing facilities to enable the validation and inter-comparison of the Aerosol_cci+ test datasets; they also acquire the needed reference datasets (AERONET, other aerosol satellite datasets) |
| **MPI / ECWMF** | User test cases | - MPI and ECWMF provide computing facilities to conduct 2 specific user test cases with Aerosol_cci+ test datasets (and as appropriate possibly extended C3S datasets) and their own model capacities |

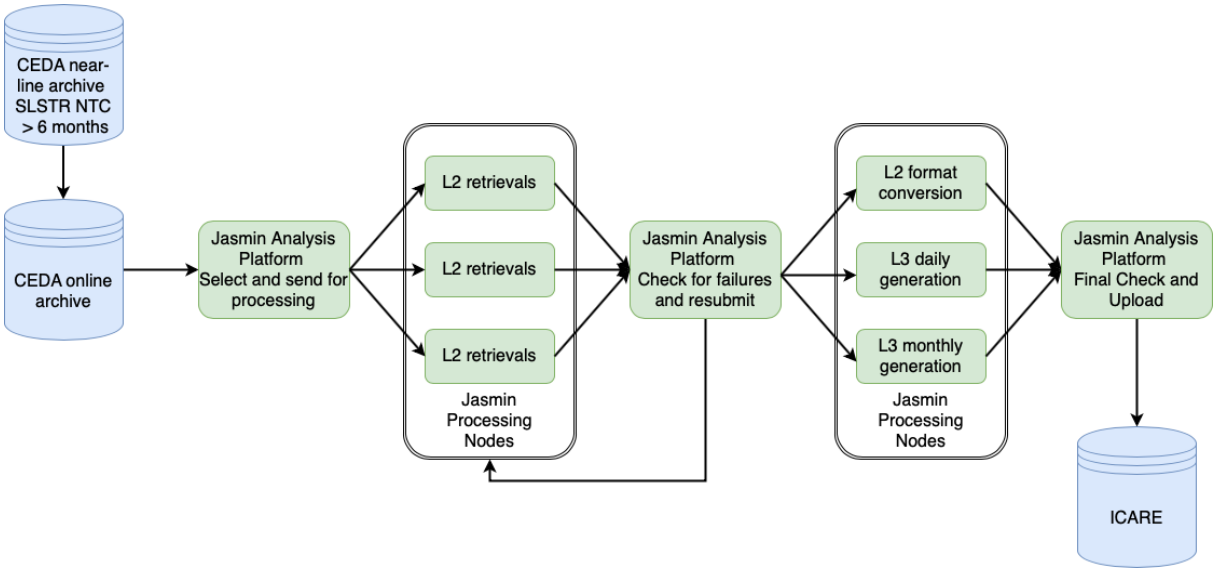


**Figure 6-1: Block diagram of SU Aerosol CCI processing**

The main processing node in year 1 of Aerosol_cci+ is at Swansea University (SU). We provide here a short summary of the SU processing system, including a specification of system performance, operating environment and hardware.

Processing developed by Swansea University is implemented on the JASMIN/CEMS data analysis environment of the UK Centre for Environmental Data Analysis (CEDA). This enables the direct processing of (A)ATSR and SLSTR from S3A and S3B L1b input files, available from the CEDA / NEODC archive. The JASMIN analysis platform provides a high performance distributed computing system under, with approximately 14,000 processing cores as of November 2021. It operates a CentOS Linux system with a SLURM scheduling mechanisms for parallel computation. The main retrieval code for ATSR-2 and ASTSR is written in C, while SLSTR is written in C++, requiring approximately 10,000 lines of code for each. Preprocessing and AERONET comparison is written in JAVA, and processing chain scripting in Bash. The retrieval system consists of a small number of command line tools providing preprocessing, cloud screening, AOD retrieval post processing and conversion to required output format. These tools are chained by shell scripts to perform the entire retrieval from L1b orbit files to L2 and L3 AOD. Version control is maintained using Git with a backup to GitHub, a publicly available, no-cost cloud service. Documentation is by comment within code, with major changes recorded at the level of ATBD updated versions. Under CCI, code documentation is in the process of being upgraded using the *Doxygen* markup system.

The current performance enables processing of one month of input data for SLSTR in about 6 hours using parallel processing available on JASMIN, and 1 hour for (A)ATSR. enabling the SU-ATSR retrieval to reprocess the 17 year archive in about 7 days, and one year of SLSTR in 3 days. While (A)ATSR is stored permanently, typically only the recent 6 months of NTC SLSTR are directly available, with temporary availability of remaining months transferred from tape archive on request. Currently 30TB space is reserved on CEDA for product outputs, with copy to the iCARE central repository for C3S_312b_Lot2 and Aerosol_cci+ once a due deliverable dataset is completed.

The processing of SLSTR data with the CISAR/GEDAP innovative algorithm is performed at Rayference premises. The S3A/SLSTR data are downloaded through the Python Sentinel API and stored in a dedicated 56 TB storage disk. The L1B data are then by the TileMaker module of the GEDAP algorithm, written in Python, and formatted into input tiles. The resulting input tiles are then processed by 6 dedicate linux servers, 3 of which with 36 cores, 2 with 28 cores and 1 with 20 cores. These servers run the TileMaker module and the TileProcessor module, written in Python (scheduler), C++ (pre-inversion operations) and Fortran 90 (inversion of the RTM). The scheduling is managed by a Dask scheduler, which distributed the tasks among the available workers.

The GEDAP algorithm is composed of ~28000 lines of code in C++, handling all the inversion pre- and post- processing operations, ~10000 lines in Fortran, which represents the core of the inversion algorithm and ~7000 lines of code in Python, which takes care of the scheduler. Version control is maintained using Git and the cost-free Git management software GitLab. Documentation is written within the code, using the Doxigen markup system. Major changes are documented in the ATBD updated versions.

The processing time of the CISAR/GEDAP algorithm is about 2s per pixel per accumulation period. Given the optimal parallelisation of the code at pixel level thanks to OPENMP API, given the current available resources at Rayference, i.e. 184 cores, 4 months of data over an area of about 20 milion pixels (e.g. Australia) can be processed in about 20 days.

The resulting architecture is described in Figure 2. The dask client is on a linux machine with 20 cores. The dask scheduler distributes the task among the available server on the cluster.

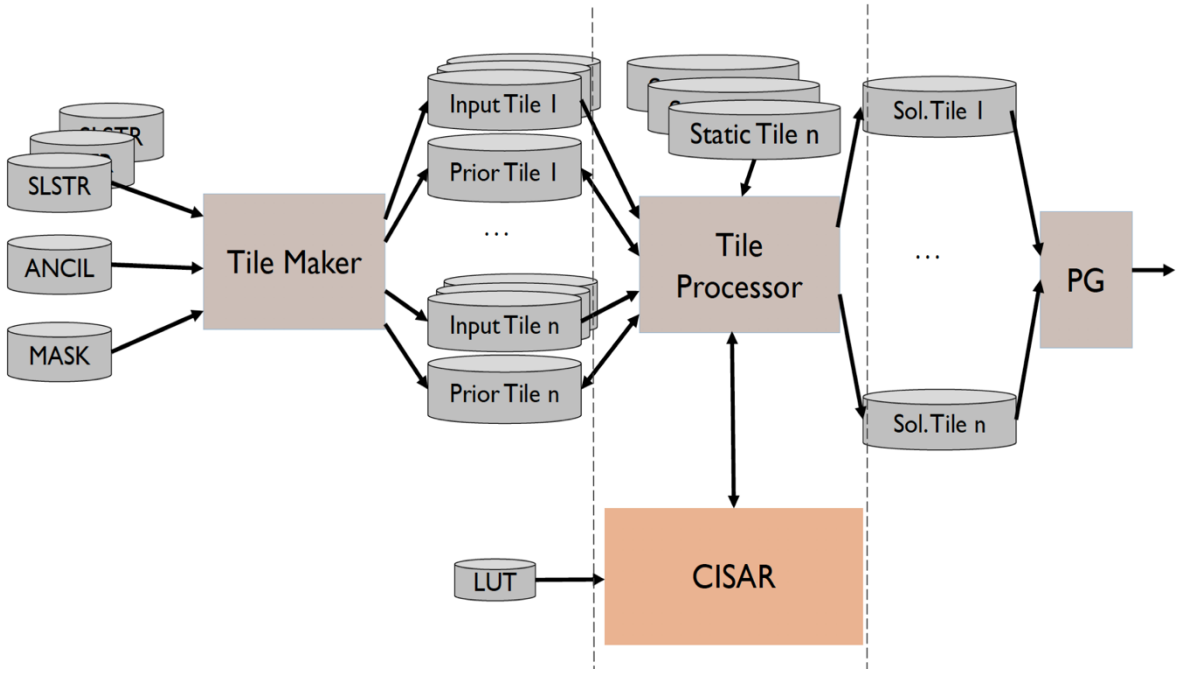


**Figure 6-2: Rayference data flow configuration. The dask scheduler running on the client is in charge of the scheduling. For each Tile *i* to be processed, the scheduler distributes the static and input tile generation, the inversion and the product generation tasks among the available workers.**

# 7 REQUIREMENT MATCHING MATRIX

The following table provides the matching of **those requirements** in the system requirements document [RD2**] relevant for the test aerosol datasets** to the sections of this system specification document.

| ID | Type | Description | SSD section |
|---|---|---|---|
| SR-GE-0040 | General | The system shall support the repeated validation and inter-comparison of the aerosol ECV products produced with external reference and comparison datasets. | 3.1, 3.2 |
| SR-GE-0041 | General | The ECV products produced by the aerosol_cci system must have scientific credibility from the users' point of view, e.g. by close interaction with end users, by performing regular inter-comparisons and validation against datasets that the users already trust and by submitting aerosol_cci results in peer reviewed journals that the users read and respect. | 3.1 |
| SR-GE-0042 | General | The team performing validation and inter-comparison tasks has to be independent from algorithm developers. | 3.1, 3.2 |
| SR-GE-0050 | General | In order to ensure highest quality and credibility of ECV products the system shall be developed, maintained and operated under the control of an Aerosol Science Team containing internationally recognised aerosol experts in the following domains:<br>- climate users<br>- algorithm developers<br>- validation experts | 2, 3.2 |
| SR-GE-0051 | General | The system must be open to assure free and open product availability and transparency of algorithm definitions. | 3.2 |
| SR-GE-0080 | General | The system shall be able to easily integrate new algorithm versions or completely new algorithms for limited test purposes (to qualify algorithm improvements) and for complete reprocessing. | 3.1, 4.3.1 |
| SR-GE-0100 | General | The system shall contain internal coordination mechanisms to assure the quality of algorithm development. This includes validation and inter-comparison expertise and agreement on algorithm development priorities between algorithm experts and validation experts / user representatives. | 3.1 |

| | | | |
|---|---|---|---|
| SR-GE-0110 | General | The system shall contain external coordination mechanisms to assure international coordination, effective interfaces with instrument calibration teams, and close links with other ECV teams (consistency of products). | 3.1 |
| SR-GE-0120 | General | The internal and external coordination mechanisms shall be flexible to allow adjustment based on experiences made throughout the continued work | 3.1 |
| SR-GE-0130 | General | The system shall contain an archive and dissemination facility including a repository for documentation on algorithms, validation and utilization of ECV products | 3.10, 3.2 |
| SR-FU-0150 | Functional | aerosol_cci system shall maximise the consistency of ancillary data used with other datasets used by the end users (e.g. from CCI Clouds and Fire projects, explicitly prioritized in the user requirements document). | 3.1 |
| SR-DO-0160 | Documentation | The system shall assess competitiveness or complementarity of any new aerosol earth observation product to the Aerosol CCI product and this should be documented and proven at the end of the product development and evaluation cycle. | 3.1, 4.3.2 |
| SR-QU-0180 | Quality | The system shall assure swift and continuous implementation of cutting edge research results (new algorithms and upgrades). | 3.1, 4.3.1 |
| SR-FU-0191 | Functional | The system shall support users to explore the ECV products (e.g. via AEROCOM, MPI, ICARE tools). | 5.3 |
| SR-QU-0260 | Quality | The science Community and users shall interface with the system for quality feedback | 3.2 |
| SR-DO-0342 | Documentation | Each product created by the aerosol CCI system and its recommended use shall be described in more detail within a product user guide highlighting limitations of its use. | 5.3 |
| SR-DO-371 | Documentation | Algorithms shall be described in ATBDs which are available for the end users. This document shall be endorsed by the science team on each major algorithm revision | 5.3 |
| SR-DO-372 | Documentation | A validation report shall be produced for every processing run. The report shall be created by validation experts and endorsed by the science team | 5.2 |

| | | | |
|---|---|---|---|
| SR-DO-373 | Documentation | An inter-comparison report shall be created by inter-comparison experts and endorsed by the science team | 5.2 |
| SR-DO-375 | Documentation | User requirements shall be described in a URD. This document shall be endorsed by the science team on bi-annual basis or release of major new user requirements (e.g. GCOS) to initiate the next algorithm development and production cycle. | 3.2 |
| SR-DO-378 | Documentation | An error characterization report (ECR) shall document the summary of findings on product uncertainties. It shall be endorsed by the science team for every completed processing run. | 5.2 |
| SR-DO-379 | Documentation | The science team shall publish a collection of utilizations made and reported feedback by end users of the ECV products on bi-annual basis. Wherever possible users shall be encouraged to publish these results in peer-reviewed journals. | 3.2 |
| SR-OP-0470 | Interface | Experts for product validation and inter-comparison shall be consulted to validate ECV product quality | 4.1, 5.2 |
| SR-OP-0480 | Interface | Feedback received from climate research ECV users and product validation and inter-comparison experts shall influence development priorities of the algorithm implementations | 3.2 |
| SR-QU-0450 | Quality | The characteristics of processing algorithms have to be annotated within the products meta data | 5.1 |

End of the document