Climate Change Initiative+ (CCI+) Phase 2

Sea Surface Salinity

System Specification Document (SSD)

Customer: ESA

Ref.: ESA-EOP-SC-AMT-2021-26

Version: v4r1

Ref. internal: 4000123663/18/I-NB

Revision Date: 31/05/2023

Filename: SSS_cci_PHASE#02_D3.1-SSD-v4.1.docx

Deliverable code: D3.1

Signatures

Author	Frederic Rouffi ACRI-ST		
Reviewed by	Odile Fanton D'Andon ACRI-ST		
	Giovani Corato AdwaisEO		
	Jean-Luc Vergely ACRI-ST		
Approved by	Jacqueline Boutin (Science Leader) LOCEAN		
	Nicolas Reul (Science Leader) IFREMER		
	Rafael Catany (Project Manager) ARGANS		
Accepted by	Susanne Mecklenburg (Technical Officer) ESA		

Diffusion List
Sea Surface Salinity Team Members
ESA (Susanne Mecklenburg, Roberto Sabia)

Amendment Record Sheet

	Document Change Record	
Date / Issue	Description	Section / Page
JAN19/DRAFT0.1	Initial draft for internal discussion	New document
MAR19/i1r0	Document evolution for internal review	
NOV19/v1r1	AR update	
JUL20/v2r0	AR 2 update	Accounting for ESA comments
JAN21/v2r2	MS#4	Accounting for ESA comments: - List of acronyms - Include bibliography - Reference figures Correct definition of pct_var
FEB21/v2r3	MS#4	Clarification on Y1 and Y2 processing chain
SEP21/v3r0	Phase 1 final version	Updated for Year 3 production
MAY23/v4r0	Phase 2 intermediate version	Updated for first phase 2 version

Table of Contents

List of figures

List of tables

1 Introduction

1.1 Executive Summary

The System Specification Document (SSD) serves to define the design of the system utilised to generate the Sea Surface Salinity (SSS) Essential Climate Variable (ECV) time series. The system aims at evolving through the successive phases of the project targeting to improve the quality and accuracy of its production, coping with ESA requirements, and implementing the algorithms evolutions proposed by the science teams.

The first phase of the project ended in 2021, the SSD was initially structured as an answer to the System Requirement Document (SRD) aiming at specifying the system as a roadmap to help the system Engineers to maintain, operate and enhance the system, to develop the software and perform their testing and installation.

For this second phase, the document takes its source from the previous phase accounting for the same technical constraints.

Unlike the SSD produced in Phase 1, the Round Robin exercise will be described in the Product Validation and Algorithm Selection Report (PVASR) document.

1.2 Purpose and Scope

This document is the System Specification Document (SSD) dedicated to the definition of the End-to-End processing architecture aiming at producing the Essential Climate Variable (ECV) Sea Surface Salinity (SSS) dataset within phase 2 of the ESA Climate Change Initiative Extension (CCI+).

Th system utilises many data sources, such as satellite, in situ, and other relevant data gathered during phase 1, with the ultimate objective of achieving the most precise SSS ECV time series calculation possible.

It aims at supporting the scientist involved in the climate change assessment by providing the best quality long-term sea surface salinity monitoring dataset with the corresponding uncertainties. The algorithms used to produce the dataset are described in ATBDs and have been tuned along the way by the project Science Team during phase 1 to improve their reliability and adequacy with the CCI expectation, particularly for what concerns product format [AD-1].

The products generated by the system are customised according to the User's feedback, as exposed in the User Requirements Document [URD].

This document is the first release of phase 2 proposed in Year 1. It is also based on the experience acquired during the three Years of the project phase 1, and the feedback received by the validation team and the users of the products.

1.3 Structure of the document

The document contains the following major sections:

Section 1: Introduction to the document (present section)

Section 2: Overview of the processing system

Section 3: Justification of the Solution

Section 4: Architecture of the system

Section 5: GitHub repository

1.4 References

1.4.1 Applicable Documents

ID	Document	Reference
DSTD	CCI Data Standards, CCI-PRGM-EOPS-TN-13-0009	V2.3, 26/07/2021
SRD	System Requirement Document	SSS_cci-D3.1-SRD-i1r5
PUG	Product User Guide	SSS_cci-D4.3-PUG-i3r0
URD	User Requirement Document	SSS_cci-D1.1-URD-i3r0
DARD	Data Access Requirement Document	SSS_cci-D1.3-DARD-v1r3
PSD	Product Specification Document	SSS_cci-D1.2-PSD-v3r0
SoW	CCI+ Statement of Work	
ATBD	Algorithm Theoretical Baseline Document	SSS_cci-D2.3-ATBD-v3.1
PVASR	Product Validation and Algorithm Selection Report	SSS_cci-D2.1-PVASR-v3.1
ALGO_L2_SMOS	CATDS (2017). CATDS-PDC L3OS 2P Algorithm Theoretical Basis Document. Available at: https://www.catds.fr/content/download/78841/file/ATBD_L3OS_v3.0.pdf	ATBD_L30S_v3.0

ALGO_L2_SMAP	RSS SMAP Level 2 Sea Surface Salinity V4.0 40km Validated Dataset. Available at: https://podaac.jpl.nasa.gov/dataset/SMAP_RSS_L2_SSS_V4	RSS Technical Report 082219
ALGO_L2_AQUA	Aquarius Official Release Level 2 Sea Surface Salinity v5.0 ATBD. Available at: ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/docs/v5/	RSS Technical Report 120117
ALGO_L3_AQUA	Aquarius Official Release Level 3 Sea Surface Salinity v5.0. Aquarius L2 to L3 Processing Document. ATBD. Available at: ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/docs/v5/	AQ-014-PS-0017_Aquarius_L2toL3ATBD_DatasetVersion5.0

1.4.2 Reference Documents/Bibliography

ID	Document
RD-1	Systematic Observation Requirements for Satellite-based Products for Climate: Supplemental Details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in support of the UNFCCC (GCOS-92)", GCOS-107, September 2006 (WMO/TD No.1338). Available online at http://www.wmo.int/pages/prog/gcos/Publications/gcos-107.pdf
RD-2	2015 Update of Actions in The Response of the Committee on Earth Observation Satellites (CEOS) to the Global Climate Observing System Implementation Plan 2010 (GCOS IP-10). Available online at http://ceos.org/document_management/Working_Groups/WGClimate/WGClimate_The-CEOS-CGMS-Response-to-the-GCOS-2010-IP_Jun2015.pdf
RD-3	The Second Report on the Adequacy of the Global Observing Systems for Climate in Support of the UNFCCC, GCOS – 82, April 2003 (WMO/TD No.1143). Available online at http://www.wmo.int/pages/prog/gcos/Publications/gcos-82_2AR.pdf
RD-4	IPCC, 2022 Sixth Assessment Report https://www.ipcc.ch/assessment-report/ar6/

RD-5	The ESA Climate Change Initiative – Description, issue 1 revision 0 - 30/09/09 EOP-SEP/TN/0030-09/SP Available online at: http://cci.esa.int/sites/default/files/ESA_CCI_Description.pdf
RD-6	Climate Change Initiative website: http://cci.esa.int
RD-7	The 2022 GCOS Implementation Plan. Available online at: https://gcos.wmo.int/en/publications/gcos-implementation-plan2022
RD-8	Guideline for the Generation of Satellite-based Datasets and Products meeting GCOS Requirements, GCOS Secretariat, GCOS-128, March 2009 (WMO/TD No. 1488). Available online at: http://www.wmo.int/pages/prog/gcos/Publications/gcos-128.pdf
RD-9	Quality assurance framework for earth observation (QA4EO): http://qa4eo.org
RD-10	The ESA Data User Element: http://due.esrin.esa.int/
RD-12	EU Research Programmes on Space and Climate: H2020 (http://ec.europa.eu/programmes/horizon2020/en/h2020-section/space, https://ec.europa.eu/programmes/horizon2020/en/h2020-section/climateaction-environment-resource-efficiency-and-raw-materials) and Copernicus (http://www.copernicus.eu/).
RD-13	Implementation Plan for the Global Observing System for Climate in support to UNFCCC (2010 Update), GCOS-138, August 2010. Available online at: http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf.
RD-14	Systematic Observation Requirements for Satellite-Based Data Products for Climate - 2011 Update, GCOS-154, December 2011. Available online at: http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf.
RD-15	The Global Observing System for Climate: Implementation Needs, GCOS-200, October 2016. Available online at:

	https://library.wmo.int/opac/doc_num.php?explnum_id=3417, http://www.wmo.int/pages/prog/gcos/index.php?name=News
RD-16	Status of the Global Observing System for Climate - Full Report, GCOS-195, October 2015. Available online at: http://www.wmo.int/pages/prog/gcos/Publications/GCOS-195_en.pdf
RD-17	ESA CCI: CCI Project Guidelines EOP-DTEX-EOPS-SW-10-0002, 2010. Available at: http://cci.esa.int/sites/default/files/ESA_CCI_Project_Guidlines_V1.pdf
RD-18	ESA CCI Status 2012 v1.1, CCI-MNGT-EOPS-TN-12-0045, September 2012. Available at: http://cci.esa.int/sites/default/files/CCI_StatusReport_2012_for_web_complete.pdf
RD-19	M. Dowell, P. Lecomte, R. Husband, J. Schulz, T. Mohr, Y. Tahara, R. Eckman, E. Lindstrom, C. Wooldridge, S. Hilding, J. Bates, B. Ryan, J. Lafeuille and S. Bojinski, 2013: Strategy Towards an Architecture for Climate Monitoring from Space. Pp. 39. This report is available from: http://ceos.org/document_management/Working_Groups/WGClimate/WGClimate_Strategy-Towards-An-%20Architecture-For-Climate-Monitoring-From-space_2013.pdf
RD-20	S. Bojinski, J-L. Fellous, June 2013: Response by ESA to GCOS - Results of the Climate Change Initiative Requirement Analysis, GCOS Secretariat, CCI-PRGMEOPS-TN-13-0008. Available online at: http://cci.esa.int/sites/default/files/ESA_Response_to_GCOS_v3_2a.pdf
RD-21	Hollmann, R.; Merchant, C.J.; Saunders, R.; Downy, C.; Buchwitz, M.; Cazenave, A.; Chuvieco, E.; Defourny, P.; De Leeuw, G.; Forsberg, René; Holzer-Popp, T.; Paul, F.; Sandven, S.; Sathyendranath, S.; Van Roozendael, M.; Wagner, W. The ESA climate change initiative: Satellite data records for essential climate variables. American Meteorological Society. Bulletin, Vol. 94, No. 10, 2013, p. 1541-1552. Available online at: http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-11-00254.1

RD-22	(Joint Committee for Guides in Metrology, 2008, Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM), JGCM 100: 2008. Available online at: http://www.bipm.org/en/publications/guides/gum.html.
RD-23	Merchant, C., et al., 2017, Uncertainty information in climate data records from Earth observation, Earth Syst. Sci. Data Discuss., doi: 10.5194/essd-2017-16, 2017.
RD-24	Ohring, G., 2007: Achieving Satellite Instrument Calibration for Climate Change. National Oceanographic and Atmospheric Administration, 144 pp.
RD-25	Ohring, G., Tansock, J., Emery, W., Butler, J., Flynn, L., Weng, F., St. Germain, K., Wielicki, B., Cao, C., Goldberg, M., Xiong, J., Fraser, G., Kunkee, D., Winker, D., Miller, L., Ungar, S., Tobin, D., Anderson, J.G., Pollock, D., Shipley, S., Thurgood, A., Kopp, G., Ardanuy, P. And Stone, T., 2007, Achieving satellite instrument calibration for climate change. Eos, Transactions American Geophysical Union, 88, p. 136
RD-26	ESA Third Party Missions: www.esa.int/Our_Activities/Observing_the_Earth/Third_Party_Missions_overview
RD-27	Copernicus Space Component: www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Space_Component
RD-28	European Cooperation for Space Standardization: ecss.nl
RD-29	Data Standards Requirements for CCI Data Producers (v1.2, March 2015) cci.esa.int/sites/default/files/CCI_Data_Requirements_Iss1.2_Mar2015.pdf
RD-30	R. Somaraju and J. Trumpf, "Frequency, Temperature and Salinity Variation of the Permittivity of Seawater," IEEE Transactions on Antennas and Propagation, vol. 54, no. 11, pp. 3441-3448, 2006.
RD-31	Aquarius Level 2 Version 5.0 (final release) salinity retrieval algorithm and configured for SMAP (Meissner et al. 2017, 2018)

RD-32	Boutin, J., J.-L. Vergely, E. P. Dinnat, P. Waldteufel, F. D'Amico, N. Reul, A. Supply, and C. Thouvenin-Masson (2021). Correcting Sea Surface Temperature Spurious Effects in Salinity Retrieved From Spaceborne L-Band Radiometer Measurements, IEEE Transactions on Geoscience and Remote Sensing, 59(9), 7256-7269, doi:10.1109/tgrs.2020.3030488.
RD-33	Klein, L., and C. Swift (1977). An improved model for the dielectric constant of sea water at microwave frequencies, IEEE Transactions on Antennas and Propagation, 25(1), 104-111.
RD-34	Gaillard, F., T. Reynaud, V. Thierry, N. Kolodziejczyk, and K. v. Schuckmann (2016). In Situ–Based Reanalysis of the Global Ocean Temperature and Salinity with ISAS: Variability of the Heat Content and Steric Height, *Journal of Climate*, *29*(4), 1305-1323, doi:10.1175/jcli-d-15-0028.1
RD-35	Boutin, J., N. Reul, J. Koehler, A. Martin, R. Catany, S. Guimbard, F. Rouffi, et al. (2021), Satellite-Based Sea Surface Salinity Designed for Ocean and Climate Studies, Journal of Geophysical Research: Oceans, 126(11), e2021JC017676, doi:https://doi.org/10.1029/2021JC017676.

1.5 Acronyms

AD	Applicable Document
ADF	Auxiliary Data File
API	Application Program Interface
ATBD	Algorithm Theoretical Basis Document
CAR	Climate Assessment Report
CCI	The ESA Climate Change Initiative (CCI) is the Global Monitoring for Essential Climate Variables (GMECV) element of the European Earth Watch Programme.
CCI+	Climate Change Initiative Extension (CCI+) is an extension of the CCI over the period 2017–2024
CDR	Climate Data Record
CEOS	Committee on Earth Observation Satellites
CFOSAT	Chinese French Oceanography Satellite

CGMS	Coordination Group for Meteorological Satellites
CRDP	Climate Research Data Package
DARD	Data Access Requirements Document
DOI	Digital Object Identifier
DPM	Detailed Processing Model
DUE	Data User Element
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
ECSS	European Cooperation for Space Standardization
ECV	Essential Climate Variable
EO	Earth Observation
EOV	Essential Ocean Variable (of the OOPC)
ESA	European Space Agency
GEO	Group on Earth Observations
GCOS	Global Climate Observing System
GMECV	Global Monitoring of Essential Climate Variables - European Earth Watch programme element.
GUI	Graphical User Interface
H2020	Horizon 2020 programme
Hs	Significant Wave Height (see also SWH)
IPCC	Intergovernmental Panel on Climate Change
L1 / L2 / L3 / L4	Level 1, 2, 3, and 4 Products
L2OS	Level 2 Ocean Salinity
LUT	Look Up Table
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
Obs4MIPs	Observations for Model Intercomparison Projects
ODP	Open Data Portal
OOPC	Ocean Observation Panel for Climate
OPeNDAP	Open-source Project for a Network Data Access Protocol
OS	Ocean Salinity / Operating System
OTT	Ocean Target Transfer
Pi-MEP	SMOS Pilot Mission Exploitation Platform
PSD	Product Specification Document
PUG	Product User Guide
PVASR	Product Validation and Algorithm Selection Report

PVIR	Product Validation and Intercomparison Report
QA4EO	Quality Assurance Framework for Earth Observation
QC	Quality Control
RAM	Random Access Memory
R&D	Research and Development
RD	Reference Document
SAF	Satellite Applications Facility
SMAP	Soil Moisture Active Passive [mission of NASA)
SMOS	Soil Moisture and Ocean Salinity [satellite of ESA]
SoW	Statement of Work
SRD	System Requirements Document
SSD	System Specification Document
SSS	Sea Surface Salinity
SVR	System Verification Report
SWH	Significant Wave Height (see also Hs)
UCR/CECR	Uncertainty Characterisation Report (formerly known as the Comprehensive Error Characterisation Report)
UNFCCC	United Nations Framework Convention on Climate Change
URD	User Requirements Document
UUID	Universal Unique Identifier
VM	Virtual Machine
WGClimate	Joint CEOS/CGMS Working Group on Climate
WMO	World Meteorological Programme

2 Overview of the processing system

2.1 Purpose of the system

The system described in the present document addresses the generation of satellite-based Sea Surface Salinity time series (from Jan-2010 to Oct-2022) within the first part of the CCI+ Sea Surface Salinity ECV phase 2 project. The main objective of such product is to provide the best support to the users and, more globally, to the science community in better understanding climate change and, in particular, its effect on the salinity of the Oceans and further the impact of such a change on the other geophysics' indicators.

The processing system deployed in the frame of the project is driven by a dedicated system herein called DPMC that includes all requested functionalities, among others: a processing orchestrator compatible with the Satellite data processor and their interfaces, handling of multi-cores processing (cluster of processors), management of multiple processing baseline configurations (processor version and associated ADF), on-the-fly resources (re)-allocation, error management, processing historical records management.

In addition, the proposed processing system implements a configuration control system allowing simultaneous executions of several configurations of the processor and ADFs and strict traceability of the products generated with various configurations.

The main difference between the DPMC used in CCI+SSS phase 2 and in CCI+SSS phase 1 lies in the usage of Dockers to execute the computation. This change brings much flexibility to the whole system, particularly by suppressing the constraints linked to the operating system running the machines with the execution of heterogenous processors.

The design of the system follows a particular set of specifications:

- ✓ The overall objective of the system development is to setup and implement a production facility addressing the SSS ECV accounting for the requirements defined in the [URD], [SRD] and [PSD], capitalizing and improving upon users' feedback.
- ✓ The data generated comprehensively covers all sea surfaces on Earth. The main users are Climate Research groups and modelers. Nevertheless, some specific geographic areas are also subject to targeted studies (e.g. C-Band products on main tropical river plumes that will be implemented during this new phase of the project).
- ✓ The production facility implements improved algorithms as defined and proposed by the Science Team in the [ATBD].
- ✓ The production facility includes processing data of all available sensors and in-situ measurements to date, as well as newly acquired ones [DARD].

2.2 The system architecture context

The aim of producing an SSS ECV is to help the Climate community (i.e., academic climate modelers, data users and climate services) to better understand and characterize the changes in SSS and how these relate to other variables (e.g. SST, Humidity, etc.).

The time series produced by the system may also be used by third parties as ESA freely distributes the CCI data produced during the project lifetime.

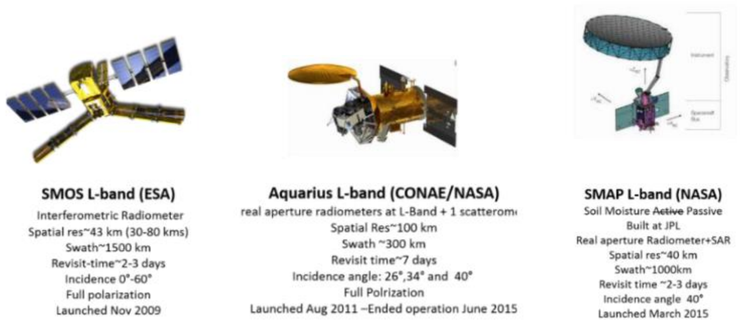
The needs expressed by the different project Teams (e.g. Science, Validation and Users) drive and impact the system's architecture. Indeed, many aspects, either scientific or more widely technical, may evolve during the project. Therefore, to maintain the system up to date and operational, a capacity for adaptation and maturation is expected, in particular, to cope with the following:

- The experience acquired from the system prototyping,
- The CCI+ Salinity validation activity,
- The feedbacks received from the Climate Research Group and Climate Modelling User Group and Data Engineering Working Group
- The Round-Robin exercises,
- Improvement in technology,
- The evolution of the specifications (i.e. data format, algorithms)

Although the [SoW] clearly states that the CCI+ Salinity project does not aim at building an operational processing system, the amount of data to be processed throughout the project is considerable. Moreover, this amount of data will continue to grow over time, with yearly and 1.5-year increments during the 3 Years life span of the project for phases 1 and 2, respectively phase 1 and phase 2). Hence the performance of such system must be realistic and scalable in time in particular in the fields of:

- CPU – multi-threading
- RAM Memory
- Disk space for archiving
- Distribution capacity to end users
- New upcoming data
- New algorithms
- Update of input data (e.g. reprocessed by provider) implying new TDS download

The system architecture mainly addresses the production of Level 4 products based on inputs from the following official production chains:



- ✓ SMOS level 1 data set (from ESA version 724) with a complete CCI reprocessing to level 2 SMOS SSS (for version 5 of phase 2)
- ✓ SMAP Level 2 from RSS (version 5.0) https://podaac.jpl.nasa.gov/dataset/SMAP_RSS_L2_SSS_V5
- ✓ Aquarius Level 3 from NASA (version 5) https://podaac.jpl.nasa.gov/dataset/AQUARIUS_L3_SSS_SMI_ANNUAL_V5

For the first part of CCI+SSS phase 2, the main updates in version 4 of the dataset with respect to previous version 3 are as follows:

- Use SSS SMOS L2 CATDS v700 over the period [01/2010-05/2010].
- Next, use SSS SMOS L2 DPGS v700 over the period [06/2010-10/2022].
- SMAP v5. SMAP 40 km resolution product filtering as suggested by the RSS ATBD. Use of the errors provided by the L2C products.
- Use Aquarius L3 NASA v5
- All the products are projected over the regular grid (0.25° resolution)
- Empirical SST correction for Aquarius and SMAP
- SSS latitudinal profile used as a reference for the seasonal latitudinal correction of SMOS, SMAP and Aquarius SSS: estimated from a seasonal SSS climatology derived from ISAS SSS (from a SMOS reference dwell line in previous versions).
- Empirical corrections at SMOS L2: dielectric constant model correction, roughness/foam correction, new rain rate correction, including wind speed information.

This processing chain is further detailed in section 4.

The system is constituted with:

- Software inherited from existing chains and from materials from CCI+SSS phase 1. The processors are reused as they are or adapted to fit the specific needs of the project (e.g. resolution, grid, product format, interfaces).

- Specific pieces of software developed for data adaptation, preparation (pre-processors) of the required auxiliary files from different sources, for matching the spatial and time acquisitions of the various EO instruments employed for the retrievals of SSS, or for homogenizing the datasets and comply with the project requirements of meeting some standards (e.g. netCDF/CF convention standards/CCI data standards see [DSTD]). The modularity of such a system is then eased and ensured by construction through the use of harmonized datasets and duly interfaced processing items.

- CCI+SSS specific processors, e.g. Level 4 SSS processing chain which merges the data produced by the abovementioned elements.

This is applicable to the SMOS, SMAP and Aquarius products and available products, which have been modified in turn in order to align part of the chains to fit with the best-selected algorithms as per the outcome of the Science Teamwork.

It is worth noting that the SMOS L2 processing chain is well-known by the Science and Engineering Teams through their work on L2 OS maintenance.

2.3 Production

The complete processing chain and tools are developed with the objective of getting a pre-operational system at the end of each exercise.

The chain for the final version of CCI+SSS phase 2 products will be based on inputs from SMOS Level 1C products (with a reprocessing to Level 2 OS), L2 SMAP and L3 Aquarius.

2.4 Data access

The access to the data archive and the product structure is detailed in [PUG, PSD].

3 Solution justification

3.1 Overview

The selection of the means to perform the computation of SSS ECV within the CCI+ project has been studied during phase 1 of the project, considering all aspects of pros and cons that bring one solution or the other. In particular in terms of:

- Cost-effectiveness
- System performance/Scalability
- Availability of the system
- Technical issues
- Data accessibility/archiving
- Security

3.2 Application to Salinity computation

The selection of the processing system for the Salinity ECV is naturally guided by the next considerations:

- ✓ For SSS ECV matters, the scalability of the processing system is not very challenging since the number of past and current missions providing salinity and addressing full Earth coverage is quite restricted (<= 3 over the duration of the project for the L-Band instruments). Therefore, for the course of the project, the processing resources needed to perform the full reprocessing of the data will not evolve significantly.
- ✓ The Level 2 SMOS Ocean Salinity processing is the most demanding processor in terms of computational resources. It requires to be carefully scheduled because some dynamic ADFs are recomputed at the end of each acquisition day. The processor is designed to work well with the openMP library for parallel computation and is especially well-suited for use in a distributed memory environment on machines with similar operating systems and multi-core processors.
- ✓ The system is set up to support the Science Team during the round-robin exercises that help to test and select the best algorithms to generate the time series.

As a conclusion the cluster option is the one selected for the generation of the SSS ECV dataset within the CCI+ project.

4 The architecture of the system

4.1 Project workflow

The SSS ECV production system is triggered upon updates/changes of the algorithms and, if required, input data (e.g. reprocessed data):

- ✓ The Science Team provides new or updated algorithms to the development Team
- ✓ The Engineer Team, in charge of the development, implements the change and performs test exercises verified by the Science Team.
- ✓ The software is deployed for larger-scale testing or production
- ✓ Data are retrieved
- ✓ Production is triggered
- ✓ The documentation is updated (ATBD, SVR, …)
- ✓ The dataset is validated.

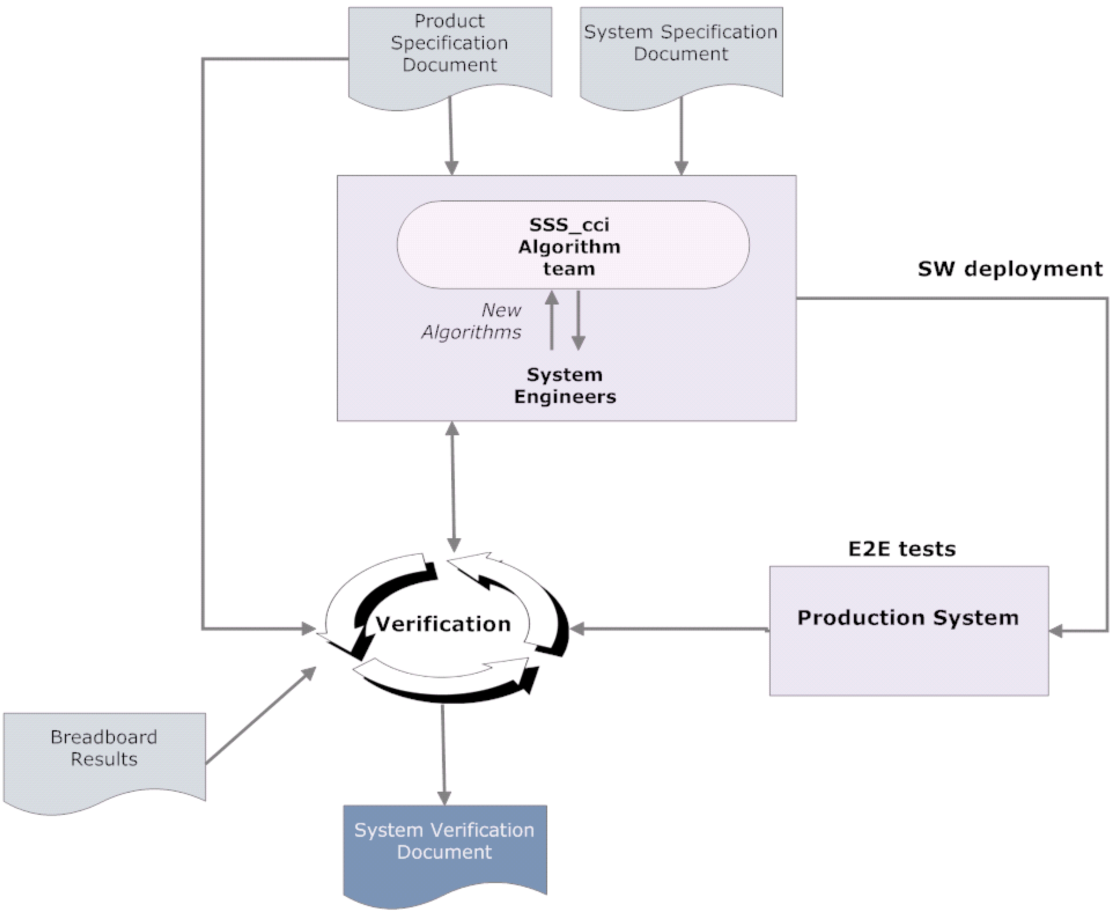


Figure 1: CCI+ Salinity system verification

4.2 Dataflow and main tasks

The production system is based on a dataflow that gives the utmost priority to the automation of the processing, thus, complying with the large processing resources requirements.

The system is composed of the following main components:

- ✓ The data ingestion module
- ✓ The production module
- ✓ The archiving module
- ✓ The data dissemination module

The following diagram on Figure 2 states the data flow between the above components and the main external interfaces.

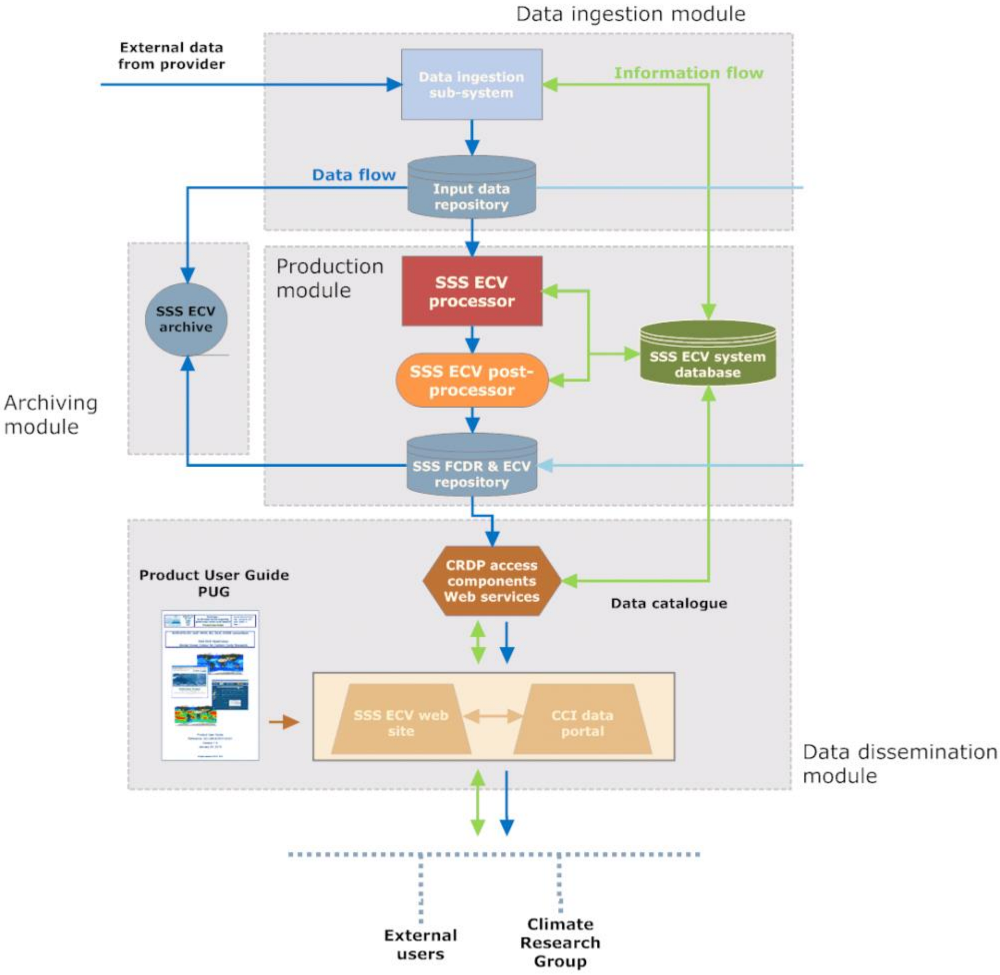


Figure 2: CCI+ Salinity production dataflow

The system is designed to answer the following matters:

- Being a help for the CCI+ Salinity Science Team allowing them to perform regular computations in view of supporting the Team during the different steps of the project. In that respect, in addition to the main production system, computing capacity through virtual machines (VM) on which the processors are installed and configured are made available to the Science researchers for algorithm testing purposes.
- Fully addressing the CCI+ expected production volume by running a complete end-to-end ECV processing system.
- The specification of the production system is living documentation. Based on the lesson learnt on the production each year through internal reviews and on users' feedback, the system will be improved and consolidated.

The different data modules and logic information flows constituting the production system, as well as the external interfaces, are detailed hereafter.

4.2.1 Data ingestion module

The external data is first sent to the data ingestion system, which aims at:

- Reading the input products,
- Ingesting the products in the CCI+ Salinity file repository
- Deriving metadata (e.g. name, validity periods) and
- Informing the CCI+ Salinity processing system database about the availability of the product. The database is the core of the processing system; it is further detailed throughout the technical document.

The metadata attached to each product are used in particular to sort and locate the data, and to properly configure the job orders which are further sent to the processing system.

The job order contains the path to all the input products, auxiliary data files, and output directories. It also sets the time period to be processed during a run.

4.2.2 Production module

Once the data is ingested and referenced in the database, the generation of products may start based on the list of orders associated with each data type to be sent to the processing.

The data generation is performed through successive steps in the production module:

- The CCI+ Salinity processor box is the computation element of the processing system. It aims at transforming EO products into higher-level products (e.g. SMOS Level 2 OS processor).
- The CCI+ Salinity post-processor box includes specific tools that will be used, for instance, to customise the main processor outputs according to some specific needs and to comply with the expected product format (CCI data standards requirements).

The processing system also creates diagnostic data files.

The temporal aggregation covers weekly and monthly productions.

Another important aspect covered by this module is the ability to handle several versions of the input data (products and ADF) – through the data ingestion module - as well as processors versions. This is managed through a configuration table that is prepared prior to the processing.

An orchestrator is available in the processing environment, based on generic and specific scripts using information extracted from the database as shown below.

In this environment, the following generic objects can be found:

- A "request". It is a set of database information that provides all characteristics related to a processing chain configuration known as Processing Baseline, which addresses both processor version and ADF configuration
- , a "batch", which is one instance of a request launched on one input product,
- a “job” which is a running batch.

At the upper level, the processing orchestrator is mainly composed of two generic scripts:

- **Scheduler.sh**, which is an infinite loop around a processing queue management service. This script is running on the controller server (see the virtual infrastructure of the system). The processing queue works in First-In-First-Out mode.
- **run_job.sh**, which is the script launched by the controller server of the selected computation node.

When the computation node takes control of a batch and starts to run the specific scripts dedicated to a processor, it sends information to the database such as start time, process Id, etc. This information is used to monitor what is currently running on the system.

At the end of a batch, the node sends information to the database, such as processing end time and batch status, and fills the process history tables. These history tables can be used to follow the evolution of the processing of the dataset and can be used to compute production statistics, including processing performance.

Both scripts are fully generic and rely on information stored in the database during the preparation of the batches. The batch Id is the only information required by these scripts to retrieve all required information from the database as shown in the above figure: input product identification, processing chain (scripts), processor (version), list of auxiliary product types, and version to be used, etc.

When an updated processor is integrated into the system, the database is updated to identify this new version for future use by the orchestrator.

Complex workflows and dataflows can be handled. Both on-request and data-driven processing can be easily enabled: this may be useful to quickly assess some changes by triggering on-demand computation on a limited number of inputs.

4.2.3 Archiving module

The CCI+ Salinity system implements a scalable archiving facility that is dimensioned to host the available and evolving input data and auxiliary data files and also to store the full data production of the system throughout the project lifetime.

The archiving module is fully linked to the system database.

4.2.4 Data dissemination module

Dissemination of the output products to the user community is achieved via the CCI open data portal; the unit is connected to the archiving system, thus allowing retrieving data through an FTP server.

4.2.5 System requirements

The system is based on the following statements:

- The system facility is based on a computation cluster using powerful hardware and a cluster management system. The number of computing nodes is sized to face the following computation constraints:
    - The full reprocessing activity does not last more than four months,
    - The EO missions to be considered cover L-band
- The production system is able to handle (archive, disseminate) and secure the following data volumes dimensioned to cope with the various missions used as input to the processing (see following tables providing the volume estimation per mission and per product level:
    - 50 TB of input level 1 products,
    - 10 TB of updated input level 1 products,
    - 50 TB of output level 2 products,
    - 1 TB of output level 3 products.

Table 1: Estimation of storage requirements based on generated products.

SMOS (L1c)	12 years (2009 – 2022)	36 TB
SMOS L2 (UDP)	12 years	1 TB
SMOS L2 (DAP)	12 years	32 TB
SMAP (L2C)	7 years	2.2 TB
SMOS (L3)	12 years	0.2 TB
Aquarius (L3)	4 years	0.01 TB
SMOS (L4)	12 years	1 TB

Total (rough estimations)	**74 TB**

Table 2: Estimation of storage requirements based on auxiliary/ancillary data.

SMOS AUX files (L1 to L2)	1 TB/yr	12 years	12 TB
ISAS	27 GB/yr	12 years	0.4 TB
Total (rough estimations)			**13 TB**

- A database management system is implemented as the heart of the cluster.
- All data and tasks are managed by this database: software, auxiliary data, scripts, input products, output products, computation nodes, nodes system configuration and status, input and output disks, computation pools, backups, requests, running tasks, waiting tasks...
- The scripts and software are loaded into a docker and executed on bare-metal machines.
- Due to the size of the data, a "waiting for data" process queue is implemented in the system.
- The system is running in data-driven mode.
- The load balancing of CPU and I/O is very important when processing huge amounts of data: pools of nodes are defined, with overlaps and priority rules between requests.
- Several output disks are used, and the system automatically distributes the output flux on the available disks in order to minimise or even eliminate any possible I/O bottleneck.
- The system automatically detects processing anomalies and applies corrective tasks: for example, if a job takes too much time to complete, the job is cancelled and queued again, the corresponding node is disabled, and a message is sent to the operator.
- Computation nodes: the computation nodes are automatically controlled by a cluster management system installed on a cluster controller and a database management sub-system.
- A cluster controller and a database management sub-system contains all information needed to process the data (filenames, physical location, working directories, output directories, processor configurations, ADF configurations, auxiliary files, computation module sequences, active nodes, computation pools, processing queues, automatic error management, ...).
- This system is fully duplicated on a redundant computer that can be used in case of severe failure.
- A storage area: contains all input/output products (levels 1, 2 and 3) and all the auxiliary files. The auxiliary files are duplicated on two different servers to minimise the impact of a storage server failure. The lifetime of the data in this system depends on the product type. The storage area is based on file servers linked to the computation nodes through a Gb interface.
- A nearline area: is used to store nearline data, i.e. data not frequently accessed. This could be the case, for example, for intermediate products used for the generation of final disseminated products.

- A duplication area: is used for the generation of archives, dissemination tapes (LTO2/LTO4) and DVD. The duplication area is also used to ingest external input data, delivered by FTP or tapes.
- The tape format for archiving purpose and data dissemination is LTO4 or LTO2 for small quantities.
- This system is connected to the system management network through a dedicated router.
- Operator workstations are connected to this network.
- The data distribution means of the CCI+ system (CCI data portal, website...) is hosted on computers connected to the system management network and communicates to the production server and, in particular, to the database and disk servers through a router in order to protect the system from any attempt of intrusion.
- When published, the data will be available through the server hosted by CEDA which will mint a DOI to the dataset.

4.3 Hardware

The CCI+ Salinity processing system is intended to run on both physical and virtual infrastructures (see Figure 3).

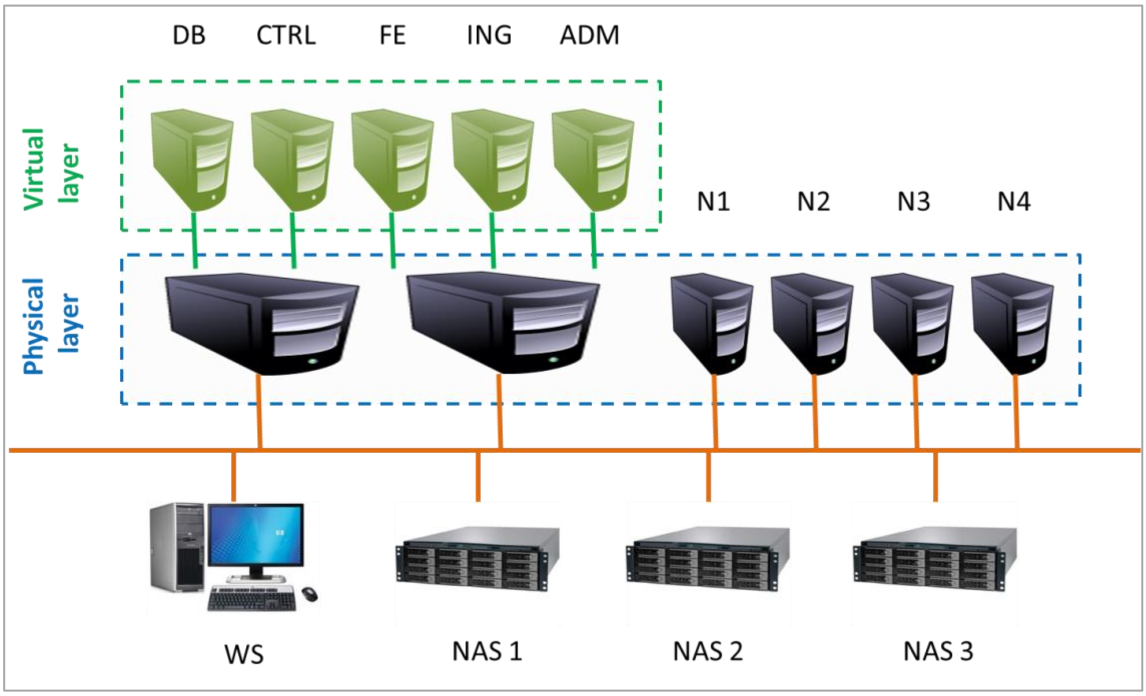


Figure 3: CCI+ Salinity processing system implementation

The components mentioned in the above schema are:

- **DB:** database management server. Hosts the database.

- **CTRL:** processing system controller. Hosts the orchestrator scripts (scheduler.sh). Send batches to the computation nodes (run_job.sh).
- **FE**: front-end server.
- **ING:** ingestion server. This server is dedicated to the ingestion of the input products.
- **ADM:** administration server. This server is used to perform all manual operations on the system, including on-request orders such as reprocessing launch.
- **N1…Ni:** computation nodes. The number of nodes is a function of the targeted system performance.
- **WS:** operator workstation. Allows remote access (ssh) to any server.
- **NAS1…NAS3:** storage areas. All files identified in the database are stored on these volumes. Once identified in dedicated database tables and visible from the computation nodes, the volumes can be used to store information. These NAS can also be connected to an FTP server (or other protocols such as WebDAV) to enable direct access to the products as soon as they are generated.

4.4 Algorithm development

The basis for the level 2 CCI algorithms is taken from:

- ✓ SMOS L2 products which are available after registration on:

  https://earth.esa.int/eogateway/catalog/smos-science-products:

  Direct models and retrieval implemented in SMOS L2OS processor version 700 (see ATBD available on https://earth.esa.int/eogateway/documents/20142/37627/SMOS-L2OS-ATBD.pdf.
  RE07 CATDS L2OS products have been used over the period [01/2010 – 05/2010] (commissioning period) and DPGS L2OS v7 products over the period [06/2010 – 10/2022].

- ✓ SMAP RSS products which are available after registration on
  https://podaac.jpl.nasa.gov/dataset/SMAP_RSS_L2_SSS_V5
- ✓ Aquarius products which are available after registration on
  https://podaac.jpl.nasa.gov/dataset/AQUARIUS_L3_SSS_SMI_ANNUAL_V5

The basis for building SMOS level 3 products is taken from the annex of Yin et al. (2012) (taking into account the error in retrieved SSS estimated in the course of the SSS retrieval and currently in use for binned products at CATDS and CP34 centres).

The Processing of L4 products (merge of several satellite SSS) is performed using a debiasing method described in Boutin et al. (2021), in which some adjustments are performed to take into account the various random and systematic errors of the various data sets. Detailed algorithms implemented in the Level 4 processing chain are described in CCI+SSS document [ATBD].

4.5 Testing of new Algorithms

4.5.1 Level 2P production

CCI L2 products are Level 2 Pre-Processed (L2P) products as defined in the CCI data standards document (see AD.9).

L2P files are daily files with ascending/descending orbit separation and are available for both SMOS and SMAP sensors.

The main content of L2P products is SSS from SMOS or SMAP. SSS is corrected from various systematic errors, such as land-sea contamination systematic errors, SSS systematic errors dependent on SST, seasonal and latitudinal systematic errors, as outputs of the Level 4 CCI SSS data version 2 chain.

SMAP SSS is harmonized with SMOS SSS by using a bilinear interpolation on a 0.25° regular grid.

The Level 2P dataset is not made available to the Users external to the CCI group. Further details about the equations are provided in [ATBD].

4.5.2 Level 3C production

The level 3 (L3) products are time and space-averaged products obtained sensor by sensor, without mixing inter-sensor information. Here, we consider simple averages of swath Level 2 SSS products, which may have already been corrected for some biases (e.g. land-sea contamination or spatiotemporal drift corrections).

SMOS and SMAP L3 products are computed from Level 2P CCI+SSS output. These products are the swaths L2P SMOS and SMAP data generated by the algorithms described in the previous section and further in [ATBD].

For Aquarius, the official end-of-mission public data release from the Aquarius/SAC-D mission is used. Aquarius Level 3 sea surface salinity standard mapped image data contains gridded 1-degree spatial resolution SSS averaged over daily, 7-day, monthly, and seasonal time scales. We use the daily non-averaged dataset for generating the CCI+SSS L4 dataset V4 (as for L4 dataset V3.2).

There are three types of L3C products:

- L3C products with data averaged weekly on a daily sliding window, cumulating ascending and descending orbits;
- L3C products with data averaged monthly on a 15-day sliding window, cumulating ascending and descending orbits;
- L3C products with data averaged monthly on a 15-day sliding window for ascending and descending orbits separately;

The Level 3C dataset is not made available to the Users. Further details about the equations are provided in [ATBD].

4.5.3 L4 algorithm evolutions

In the first stage, the generation of L4 products consisted of the setup and tuning of the Level 4 algorithm by using existing L2/L3 input products from different sensors producing salinity. All the L4 LUTs have been revised due to the change of grid and the extended period.

L4 v3.2 products (the last products from phase 1) show some limitations:

-low coverage at high latitude

-too strong ice filtering

-varying seasonal biases at high latitudes

Specific diagnostic has been made in order to correct these features. The problem of coverage comes from some LUT, which do not provide information at high latitudes. Extrapolation has been performed in order to fill in the missing information. Concerning ice filtering, a simplified algorithm using Acard has been performed. Changes in dielectric constant and wind speed corrections have been implemented to improve the modelling of the L-Band radiative signal at high latitudes.

Some minor changes have been implemented:

-management of different SSS estimators for the SMOS sensor (four different SSS estimators: SSSuncorr, SSScorr, SSSacard, SSSrfi).

-job trace content.

The quality of the results has been assessed by exploiting the round-robin results.

4.6 Generation of the sea surface salinity ECV

4.6.1 Processing chain

The processing chain used to generate the CRDP for SSS ECV is drawn in Figure 4. Three entries are used, one for each main satellite data input.

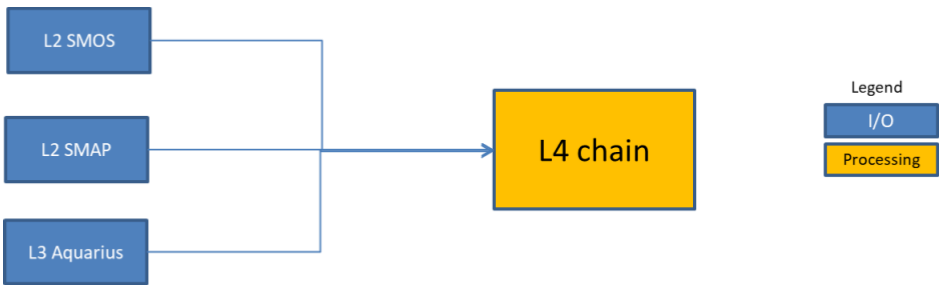


Figure 4: CCI+ Salinity production chain

4.6.1.1 Level 2 processing chain

SMOS

The full L2 product generation chain for the CCI+ SSS ECV project, used in v3.2, is detailed in Figure 5. **This production system is not used for the first release of phase 2, as instead existing SMOS L2 products are used.** This production system will be used for the second release replacing the EASE2 global grid in a regular rectangular 0.25° grid.

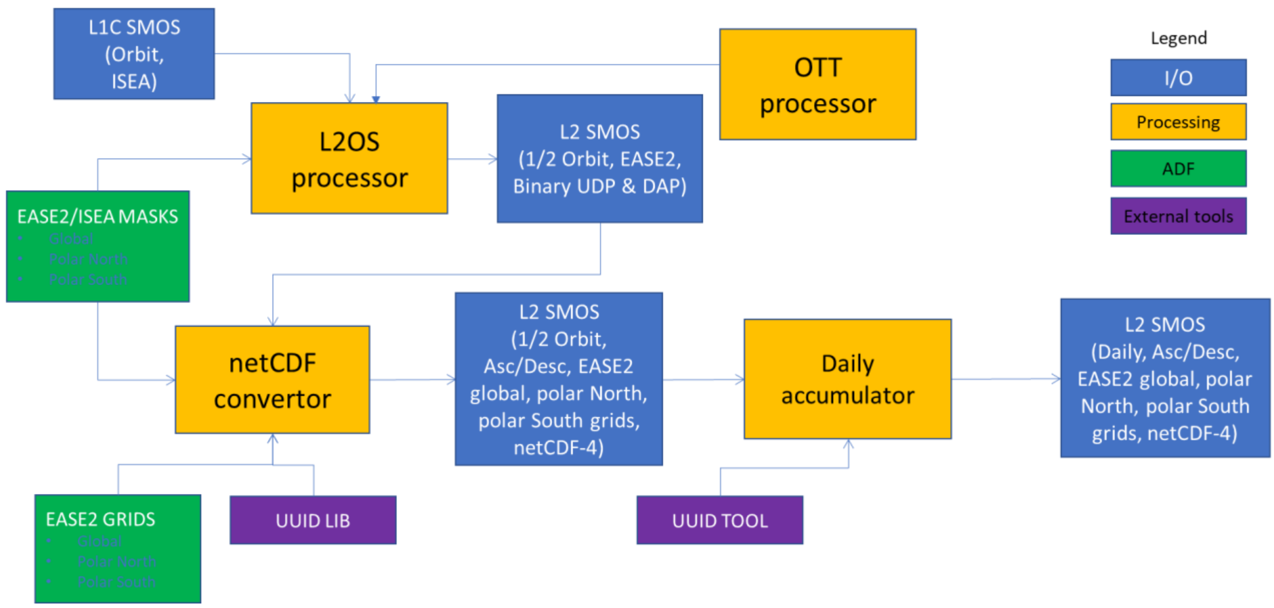


Figure 5: CCI+ Salinity processing Level 2 chain that has been deployed in Year 3

The products generated by the above chain cover a full day split into ascending and descending half-orbits. SMOS level 2 data have been produced over three grids: an EASE2 global grid (common to level 3 and level 4 datasets) and two polar EASE2 grids (North and South) that will be used for product generation during phase 2. During phase 2, the EASE2 global cylindrical equal-area projection (see NSIDC page https://nsidc.org/data/ease/ease_grid2.html) will be replaced by a regular rectangular 0.25° grid. The netCDF internal compression is applied (compression factor of 4). Accumulation of traces is performed using NCO tools (http://nco.sourceforge.net/).

The UUID library is used to generate the tracking_id chain.

A detailed view of the Level 2 processing chain is provided in section 4.9.1.

SMAP & Aquarius

Existing L2 SMAP and L3 Aquarius products have been used as input to the L4 processing chain.

The full sketch of L4 computation is illustrated in section 4.9.

4.6.1.2 Level 4 processing chain

The L4 processing chain intends to produce the following:

- Weekly L4 and
- Monthly L4

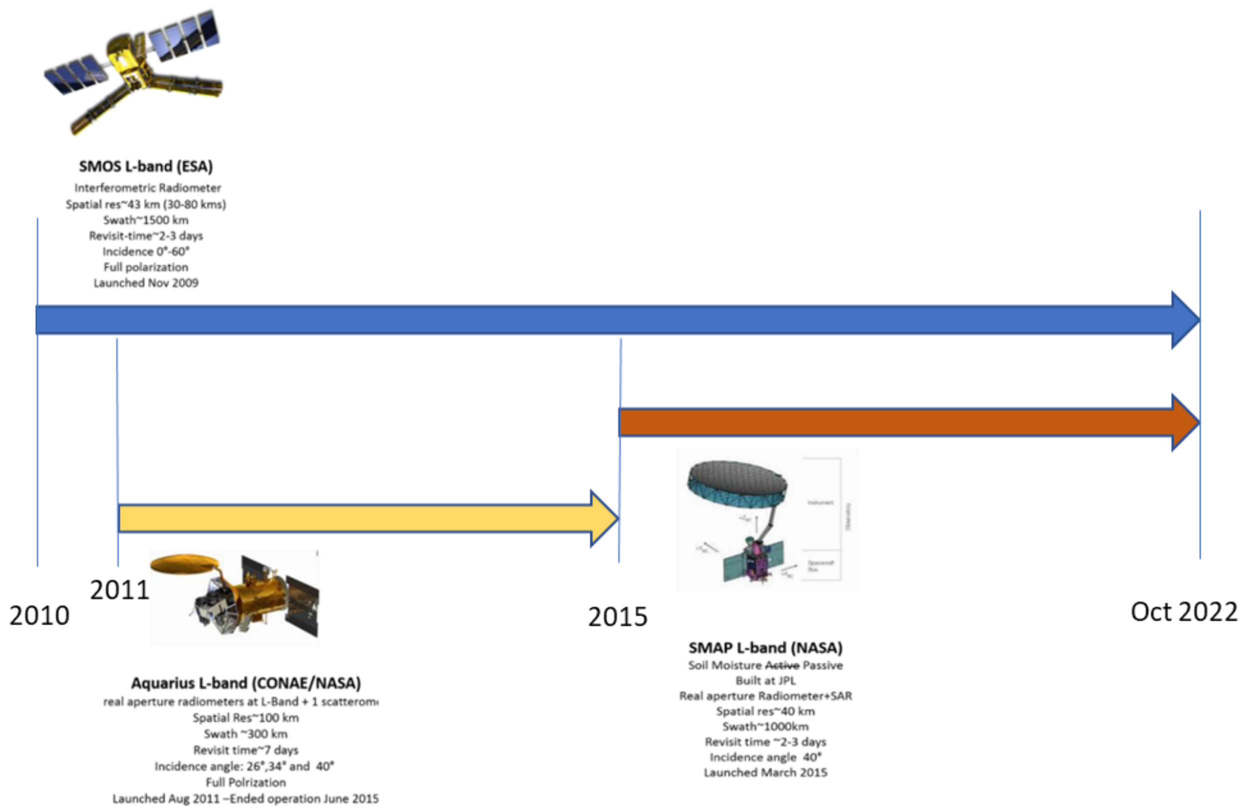
Both products are formatted in NetCDF and are in conformance with the data format convention applied to the CCI projects [DSTD]. The product structures are described in [PSD].

4.7 The input data

4.7.1 Satellite data

The main sources of satellite-based SSS data are:

- SMOS (Soil Moisture and Ocean Salinity)
- Aquarius
- SMAP (Soil Moisture Active Passive)



	Phase 1			Phase 2
Satellite	Y1	Y2	Y3	Y1
SMOS	10	10	11	12
AQUARIUS	4	4	4	4
SMAP	4	5	6	7

Figure 6: Satellite acquisition time coverage – cumulative number of Years of data during the project

The above inputs are originally generated on different grids, so a homogenisation of the data has been performed prior to the L4 processing. Figure 6 provides the acquisition coverage for each satellite.

- ✓ SMOS is computed on an Icosahedron Snyder Equal Area (ISEA) Aperture 4 Hexagonal (ISEA4H) global grid.

  A version of the Level 2 OS processor is v700. Data are projected on a regular grid (It is a global coverage grid at 0.25° resolution.), separated from ascending and descending acquisition, aggregated on daily map, and formatted in netCDF. Rain rate correction is performed from collocated IMERG instantaneous rainfall.

Dataset name example and quick-look image (Figure 7):

SM_TEST_MIR_OSUDP2_20160826T111930_20160826T121249_700_001_8

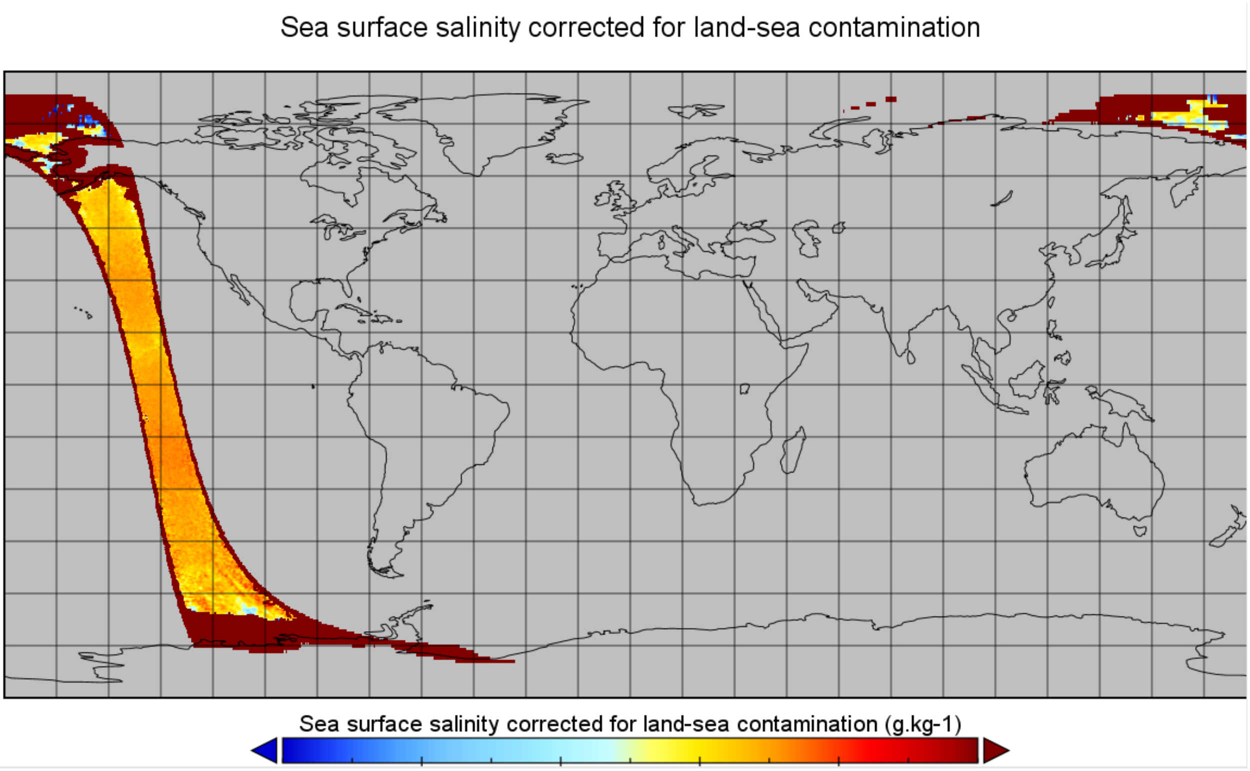


Figure 7: SMOS Salinity daily map (one time-step is represented)

- ✓ Aquarius Level 3 data used are from version v5.0, including a latitudinal correction within Aquarius processing using ARGO data. We use the daily non-averaged dataset for generating the CCI+SSS L4 dataset V2 (in V1, the 7 day running product was used). Due to latitudinal seasonal residual biases, two additional corrections have been applied at L3 (before L4 Optimal Interpolation): the first one is a SST dependent correction, and the second one is a seasonal latitudinal correction using seasonal latitudinal profiles from ISAS climatology as reference.

The original L3 mapped products are given on a regular rectangular grid. They consist of binned data accumulated for all Level 2 products over a period of 7 days (Aquarius Level-3 Standard Mapped Image). The data are formatted in netCDF at a resolution of 1 degree.

Dataset name example (ascending product) and quick-look image (Figure 8):

Q2012002.L3m_DAY_SCIA_V5.0_SSS_1deg.bz2

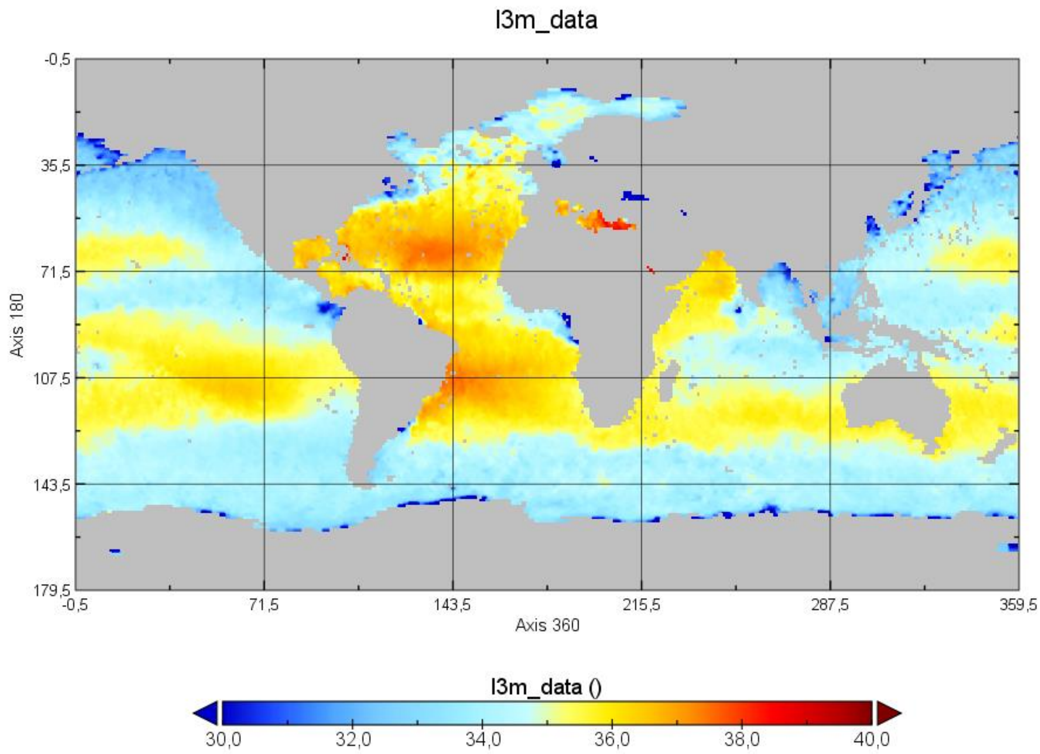


Figure 8: Aquarius Salinity 7-days map

- ✓ SMAP Level 2 data in version v5.0 are used with a latitudinal correction within SMAP RSS processing using ARGO data. As for Aquarius, two additional corrections have been applied. The first one is an SST dependent correction and the second one is a seasonal latitudinal correction using seasonal latitudinal profiles from ISAS climatology as reference.

  The Original L2C products have 40 km and 70 km (original 39 km x 47 km elliptical footprint) spatial resolutions. In comparison with v4, v5 SMAP products show improved SSS close to ice edge. Data are formatted in netCDF.

  Dataset name example and quick-look image (Figure 9):

  RSS_SMAP_SSS_L2C_r00870_20150401T004312_2015091_FNL_V05.0.nc

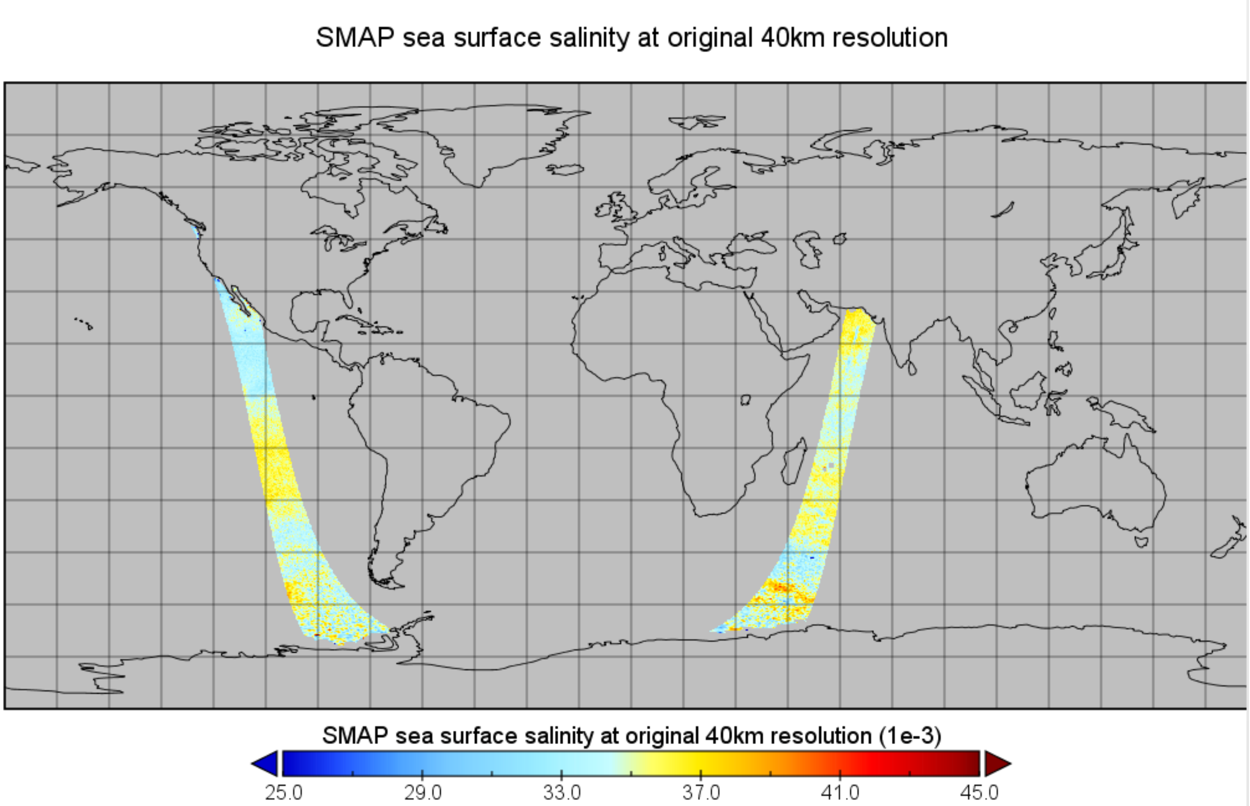


Figure 9: SMAP Salinity map

A projection of the SSS on the 0.25° regular grid has been applied to SMAP and Aquarius data to homogenise the data at the entrance of the L4 processor.

4.7.2 Other information

Optimally interpolated in-situ measurements from ARGO (ISAS fields, Gaillard et al. 2016) (see a description in [RD-34] document) are used for two purposes: 1/ a monthly climatology of latitudinal SSS profiles is built from ISAS fields and used as reference for estimating seasonal latitudinal corrections, and 2/ the statistical distribution of SSS in each grid point is used to calibrate the mean SSS value over the full time period. For the latter, in each spatial grid point, a quantile of the statistical distribution of the salinity in the time domain is used for final calibration as input to the Level 4 processor.

Representativity uncertainties and SSS variabilities derived from ocean simulations are also inputs to the OI (see more in the ATBD).

4.8 The output data

As exposed in the previous sections, the Level 4 products are computed over two time periods:

- ✓ 7-day running mean at the one-day time sampling

Ex:

-for the global product: ESACCI-SEASURFACESALINITY-L4-SSS-GLOBAL-MERGED_OI_7day_RUNNINGMEAN_Daily_0.25deg-20181101-fv4.21.nc

-for polar products: ESACCI-SEASURFACESALINITY-L4-SSS-POLAR-MERGED_OI_7day_RUNNINGMEAN_Daily_25kmEASE2-NH-20181101-fv4.21.nc

- ✓ One month at 15-day time sampling, centred.

Ex:

- for global product: ESACCI-SEASURFACESALINITY-L4-SSS-GLOBAL-MERGED_OI_Monthly_CENTRED_15Day_0.25deg-20181101-fv4.21.nc

- for polar products: ESACCI-SEASURFACESALINITY-L4-SSS-POLAR-MERGED_OI_Monthly_CENTRED_15Day_25kmEASE2-NH-20181101-fv4.21.nc

The L4 products are formatted in netCDF 4. They contain the following variables:

- ✓ Monthly and weekly SSS fields: obtained from the OI algorithm.
- ✓ SSS error: obtained from OI algorithm.
- ✓ Number of outliers over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).
- ✓ Number of data over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).
- ✓ pct_var : 100x(SSS a posteriori error)^2/variability (%).
- ✓ quality flag =1 if the fraction of outliers (n outlier/n data) present over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data) is larger than 0.1.
- ✓ flag ice (CCI V4): SMOS ice detection from Acard average. If Acard and the SST are lower than a threshold, the flag is raised.
- ✓ Coast flag (CCI V4): raised if the grid point is far from the coast at a distance less than 50 km

The products comply with the data standard of the CCI+ project [DSTD].

4.9 Overview of the processing chains

4.9.1 SMOS Level 2 processor

Figure 10 shows the SMOS L2OS processing data flow logic that has been used for processing version 3 CCI dataset during project phase 1 and that will be used in the generation of the second dataset of CCI phase 2.

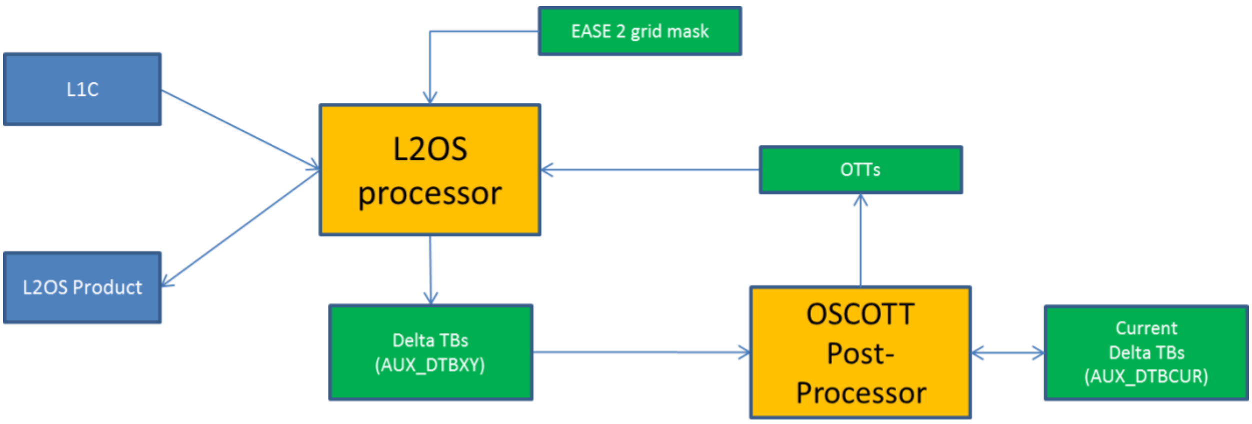


Figure 10: Overview of the SMOS L2OS/OSCOTT processor data flow

4.9.1.1 Performance testing

The following table summarises the performances obtained with the L2OS processor with a variable number of cores during the last phase 1 exercise. For a reprocessing activity like in the present CCI project, and since the amount of available RAM is rather big (256 Gb), the use of more than one core is not the best solution and does not lead to an overall improvement of the processing time. Therefore, the recommendation would be to use a non-multi-threaded processing configuration in that context.

Threads	In seconds	Gain wrt 1 core config
1	5427	-
2	3149	-42%
3	2545	-53.1%
4	2488	-54%
5	2421	-55%

Table 3: L2OS Processing time as a function of the number of cores

Based on this consideration, two different processing logics can be applied to compute the full SMOS data from L1C to Level 2.

1. based on interlaced L2OS and Ocean Target Transform, or
2. based on separate computation of L2OS and Ocean Target Transform

The second procedure has shown a better efficiency during the phase 1 exercise. It will be applied in phase 2 as well.

We provide hereafter some details on the method.

To address limitations in parallelisation, all the inputs (delta TBs) to OSCOTT can be generated by the L2OS processor in OTT mode before triggering the OTT computation. This is because the number of OTT zones is limited to 2 so this restricts the computation to the number of data acquired over those zones (see Figure 11).

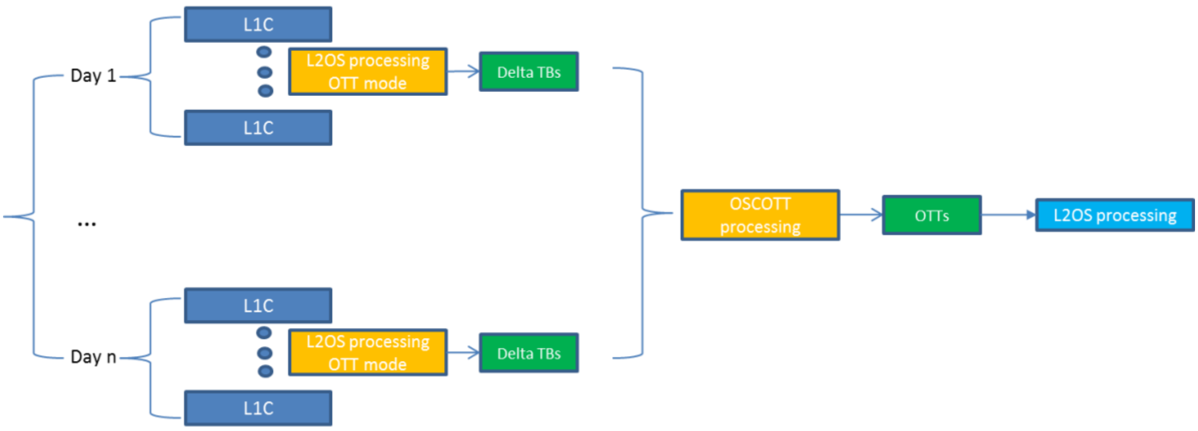


Figure 11: SMOS L2OS/OSCOTT scenario for an optimal parallelisation

With the above schema, the order of the computation does not matter so any computation can be run in parallel:

- ✓ The Delta TBs are first fully generated using the L2OS processor in a mode limiting the computation to the delta TB output. The processing time for each half-orbit is about 45 minutes.
- ✓ Then the OTT are computed for each day of the time period
- ✓ Finally, the L2OS is triggered in a nominal mode in order to generate the Level 2 products.

Based on our calculations, we have determined that 15 machines with 32 cores can be allocated for the task at hand. With one core reserved for the system, it is possible to run 465 processes in parrallell.

There are about 8100 half-orbits that cross the OTT zones; the corresponding processing time is assessed as:

8100 * 45’ / 465 ~ less than 1 day

Then the OSCOTT processor is run to generate the OTT ADF:

8100 * 15' / 465 ~ less than 1 day

Finally, the OTT ADF is used as input to the full L2OS processing in the following:

4015 * 90' / 465 ~ 16 days

The total processing time is slightly better for the second which is also simpler to set up (no waiting process).

4.9.1.2 Input data

The Level 1C SMOS product contains multi-angular brightness temperatures in the antenna frame (X-pol, Y-pol, T3 and T4) at the top of the atmosphere, geo-located in an equal-area grid system (ISEA 4H9 - Icosahedral Snyder Equal Area projection).

MIR_SCSF1C are available for salinity retrieval.

The pixels are consolidated in a pole-to-pole product file (50 minutes of sensing time), with a maximum size of about 550MB (for land and sea together) per half orbit (29 half orbits per day). Spatial resolution is in the range of 30-50 km.

4.9.1.3 Auxiliary data files

The auxiliary data files used in the Level 2 processing are listed in the following tables.

Static ADFs

Static ADF type	Description
AUX_DGG___	ISEA4-9 Discrete Global Grid used in geolocation
AUX_BFP___	Antenna best-fit plane used to initialise ESA EARTH EXPLORER CFI functions
AUX_MISP__	Mispointing angles used to initialise ESA EARTH EXPLORER CFI functions
AUX_FLTSEA	Physical Constants Needed by Flat Sea Model
AUX_RGHNS1	Look Up Tables needed by L2 Processor for the IPSL Ocean Roughness Model
AUX_RGHNS2	Look Up Tables needed by L2 Processor for the IFREMER Ocean Roughness Model.
AUX_RGHNS3	Look Up Tables needed by L2 Processor for the ICM-CSIC Ocean Roughness Model.

AUX_GAL_OS	AUX_GALAXY Map convolved with the Weighting Function AUX_ WEF___
AUX_GAL2OS	Galactic Map Product including asc/desc values
AUX_SGLINT	Bi-Static Scattering Coefficients Look-Up Table used in Sun glint correction.
AUX_ATMOS_	Physical Constants used by Atmospheric Model
AUX_DISTAN	Distance to the coast and monthly Sea/Ice Flag information over Discrete Global Grid
AUX_SSS___	Monthly Sea Surface Salinity over Discrete Global Grid
AUX_FOAM__	Physical Constants used by Foam Model
AUX_CNFOSF	The processor configuration file in full polarisation
AUX_MSOTT_	Mixed scene land-sea correction OTT Look-Up Tables
AUX_FARA_x	Faraday angle based on algorithm improvements and refined VTEC

Table 4: SMOS L2 static ADF list

Dynamic ADFs

Dynamic inputs	Origin	Comment
AUX_BULL_B	Trace included in BULL_B used to initialise ESA EARTH EXPLORER CFI functions.	IERS Bulletin B Same BULL_B for the whole day
MPL_ORBSCT	FOS	Orbit Scenario File
AUX_ECMWF_	ECMWF pre-processor	One ECMWF per half orbit
AUX_OTT1F_ AUX_OTT2F_ AUX_OTT3F_	OSCOTT pre-processor Each contains ten days of DTBXY as generated by the L2OS (sliding window)	Ocean Target Transformation LUTs, including asc/desc values derived for the three roughness models The same OTT is used for the whole day of SMOS data
AUX_DGGRFI	ESA DPGS	Current RFI probability on the DGG

Table 5: SMOS L2 dynamic ADF list

4.9.1.4 Output files

Outputs	Description	Comment
MIR_OSUDP2	Level 2 Ocean Salinity User Data Product	
MIR_OSDAP2	Level 2 Ocean Salinity Data Analysis Product	N/A as not used
AUX_DTBXY_	Delta TBs for the L2OS post-processor (input to OSCOTT post-processor	

Table 6: SMOS L2 output files

4.9.1.5 OSCOTT post-processor

Inputs	Description
AUX_CNFOSF	Processor configuration file in full polarisation (from L2OS)
AUX_DTBXY_	Delta TBs (from L2OS output)
AUX_DTBCUR	Current delta TBs (not mandatory/generated or updated by OSCOTT)

Table 7: OSCOTT input list

Outputs	Description
AUX_OTTxF_	Ocean Target Transformation LUTs, including asc/desc values derived for the three roughness models
AUX_DTBCUR	Current delta TBs (to be further reused as input to the next OSCOTT run)

Table 8: OSCOTT output list

4.9.2 SMAP Level 2 processor background

4.9.2.1 Input/output data

The SMAP salinity retrieval algorithm is run on Level 2B files and produces calibrated SMAP Level 2C surface ocean brightness temperatures (TB) and sea surface salinity (SSS) values.

4.9.2.2 Ancillary data files

Ancillary input	Description
sea surface temperature	Canadian Meteorological Center. 2016 GHRSST Level 4 CMC 0.1deg Global Foundation Sea Surface Temperature Analysis. Version. 3.0. doi: 10.5067/GHCMC-4FM03, https://podaac.jpl.nasa.gov/dataset/CMC0.1deg-CMC-L4-GLOB-v3.0
sea surface wind speed and direction	CCMP V2.0 near-real-time wind speed and direction. http://www.remss.com/measurements/ccmp/ (Wentz et al. 2015.).
atmospheric profiles for pressure, height, temperature, relative humidity, cloud water mixing ratio	NCEP GDAS 1-deg 6-hour. HGT, PRS, TMP, TMP, RH, CLWMR. Available from http://nomads.ncep.noaa.gov/
IMERG rain rate	Huffman, G. et al., 2018. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG) Version 5, LATE RUN, 30 minutes, NASA, https://pmm.nasa.gov/sites/default/files/document_files/IMERG_FinalRun_V05_release_notes-rev3.pdf
solar flux	Noon flux values from US Air Force Radio Solar Telescope sites 1415 MHz values. Available from NOAA Space Weather Prediction Center, www.swpc.noaa.gov
total electron content (TEC)	IGS IONEX TEC files. Available from ftp://cddis.gsfc.nasa.gov/pub/gps/products/ionex/
sea ice fraction	NCEP sea ice fraction. Available from http://nomads.ncep.noaa.gov/pub/data/nccf/com/omb/prod/
land mask	1 km land/water mask from OCEAN DISCIPLINE PROCESSING SYSTEM (ODPS). Based on World Vector Shoreline (WVS) database and World Data Bank. Courtesy of Fred Patt, Goddard Space Flight Center, frederick.s.patt@nasa.gov
galactic map	Dinnat, E.; Le Vine, D.; Abraham, S.; Floury, N. Map of Sky Background Brightness Temperature at L-Band. 2018. Available online at https://podaac-tools.jpl.nasa.gov/drive/files/allData/aquarius/L3/mapped/galaxy/2018
reference salinity (HYCOM)	Hybrid Coordinate Ocean Model, GLBa0.08/expt_90.9, Top layer salinity. Available at www.hycom.org

Scripps ARGO salinity (only included in rain-filtered L3 monthly files)	monthly 1-degree gridded interpolated ARGO SSS field provided by Scripps. Available at www.argo.ucsd.edu/Gridded_fields.html

Table 9: SMAP L2C ancillary file list

4.9.2.3 Salinity retrieval

The SMAP Level 2C salinity retrieval algorithm was adapted from the Aquarius Level 2 Version 5.0 (final release) and configured for SMAP by Meissner et al. in 2017, 2018 – [RD-31]. The algorithm flow is represented on Figure 12.

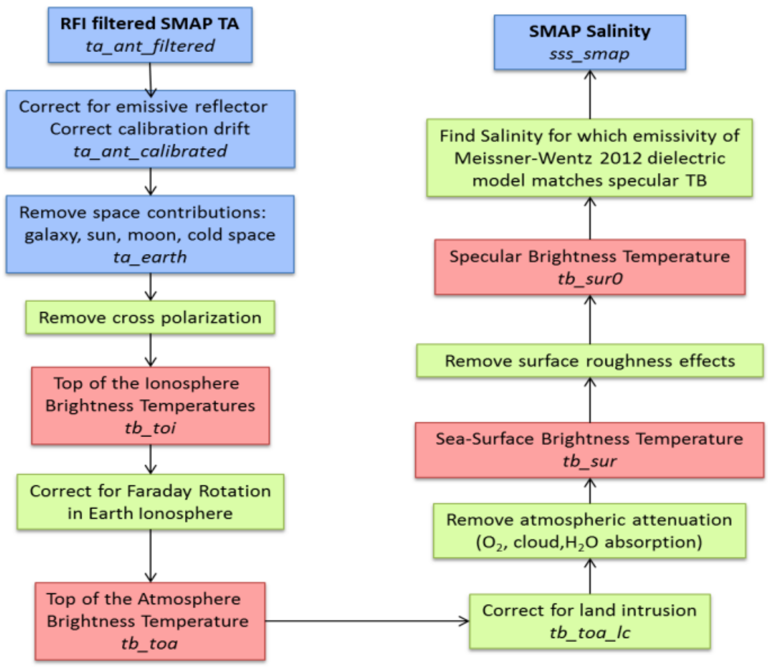


Figure 12: SMAP salinity retrieval algorithm flow diagram

4.9.3 Aquarius Level 3 processor background

4.9.3.1 Input data

Aquarius Level 1A used as input to Level 2 processing (see Figure 13) can be retrieved here: ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/L1/SSS_1a/

4.9.3.2 Processing resources

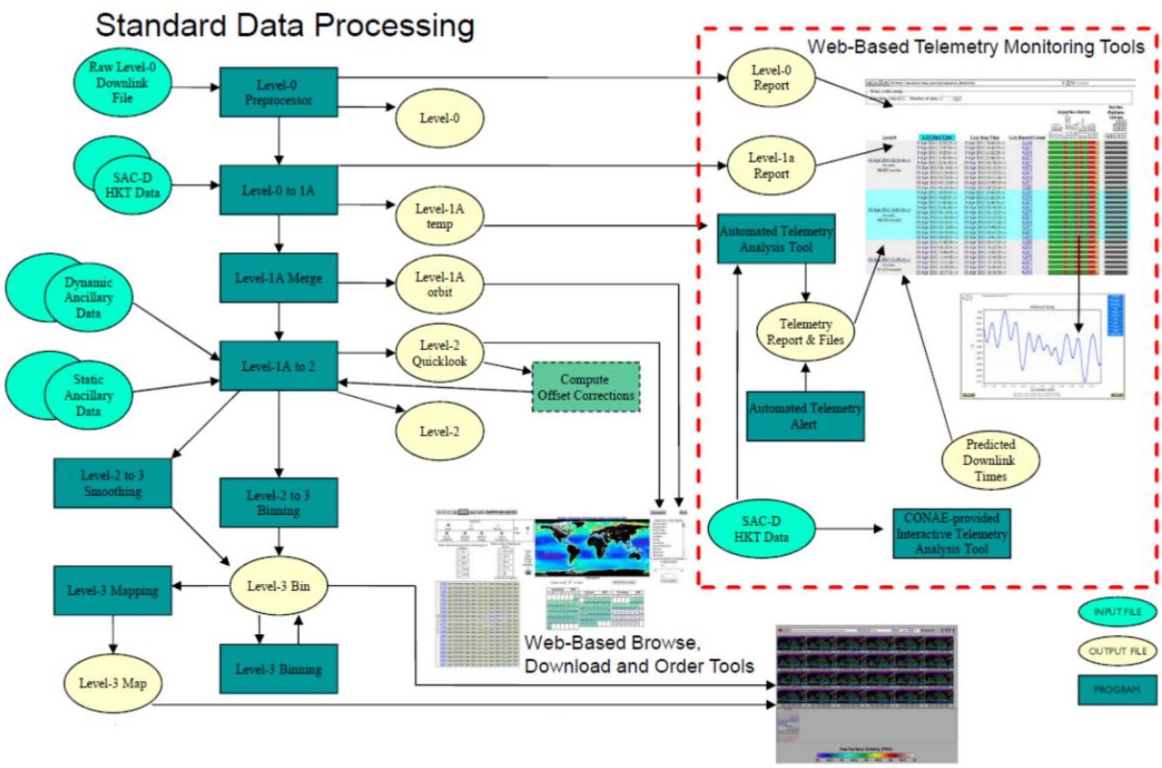


Figure 13: Aquarius algorithm flow diagram

4.9.4 Level 4 processor

The aim of the L4 processing is:

- ✓ To merge products from different satellite sensors
- ✓ To produce SSS at:
    - o a spatial resolution of about 50 km
    - o a time resolution of 1 month or one week.

With

- ✓ a spatial sampling: 25 km EASE V2 grid;
- ✓ time sampling: 15 days (monthly products) and one day (weekly products).

The processing chain breakdown is detailed in Figure 14.

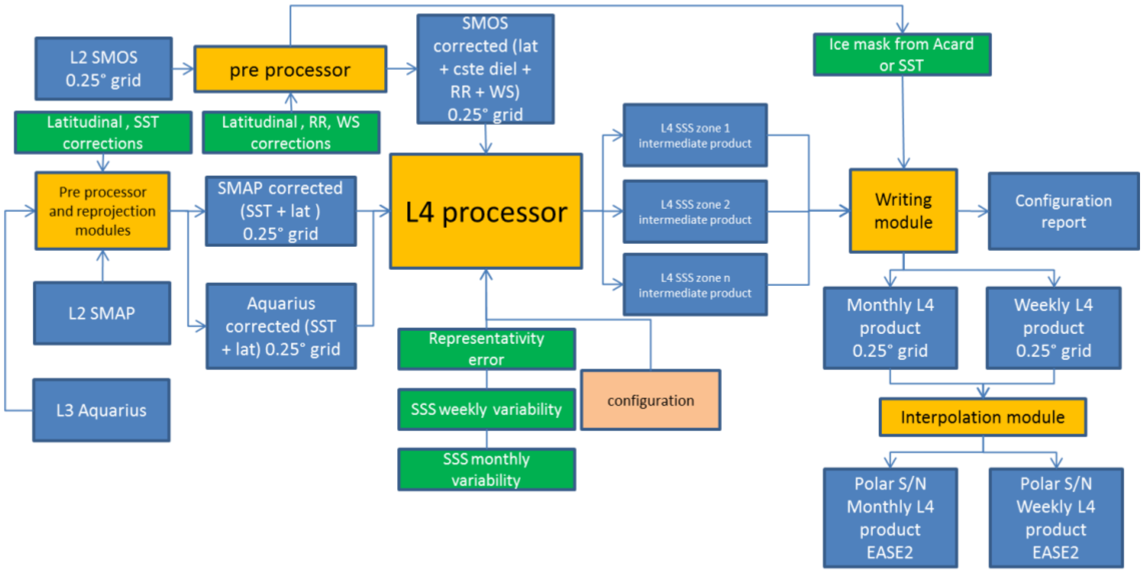


Figure 14: CCI+ Salinity processing Level 4 chain

The main processing steps are listed hereafter:

1. Pre-processing of the SSS L2/L3 products from the different sensors; Latitudinal correction and reprojection on the 0.25° regular grid
2. 3-sigma filtering and temporal Optimal Interpolation to generate monthly SSS without inter-sensor bias removal
3. 3-sigma filtering and temporal Optimal Interpolation to generate weekly SSS without inter sensor bias removal
4. Across-track and inter sensor bias removal
5. 3-sigma filtering and temporal Optimal Interpolation to generate monthly SSS. Error propagation
6. 3-sigma filtering and temporal OI to generate weekly SSS using monthly SSS as prior. Error propagation.

The content of the product is detailed in [PSD].

Compared with previous versions, several implementation errors have been identified and corrected (see [ATBD] section 4.2). Algorithmic changes have also been implemented concerning input data, computed variables, and period covered. They are all detailed in [ATBD] section 4.3.

4.9.4.1 Monthly products

The monthly SSS are evaluated in 3 steps:

1) A first estimation of the biases and time series of SSS, spatial grid node by spatial grid node is performed,

2) A 3-sigma filtering of the observed SSS in comparison with the estimated SSS is done.

3) A second estimate of SSS biases and time series after removing outliers.

The relative biases used to derive monthly SSS are estimated by taking the averaged SSS from ISAS as a priori.

4.9.4.2 Weekly products

To estimate the weekly SSS, the biases calculated on the monthly SSS are used. The weekly fluctuations are estimated around the monthly SSS as a priori. A 3-sigma filter is used where:

$$sigma = sqrt(error\_L2OS^2 + variability^2).$$

The variability is estimated from the Mercator model. This eliminates outliers that deviate too far from expected values.

5 CCI+SSS GitHub repository

The GitHub repository at https://github.com/CCI-SALINITY has been established to oversee the source code for the SSS ECV project. The complete Matlab code utilised to produce the SSS CRDP of Year 1 to 3 (Level 4 products) is available in the repository. Level 2P and Level 3C processing chains have also been uploaded on the server.

All materials developed during Phase 2 of the project will also be uploaded on the server.

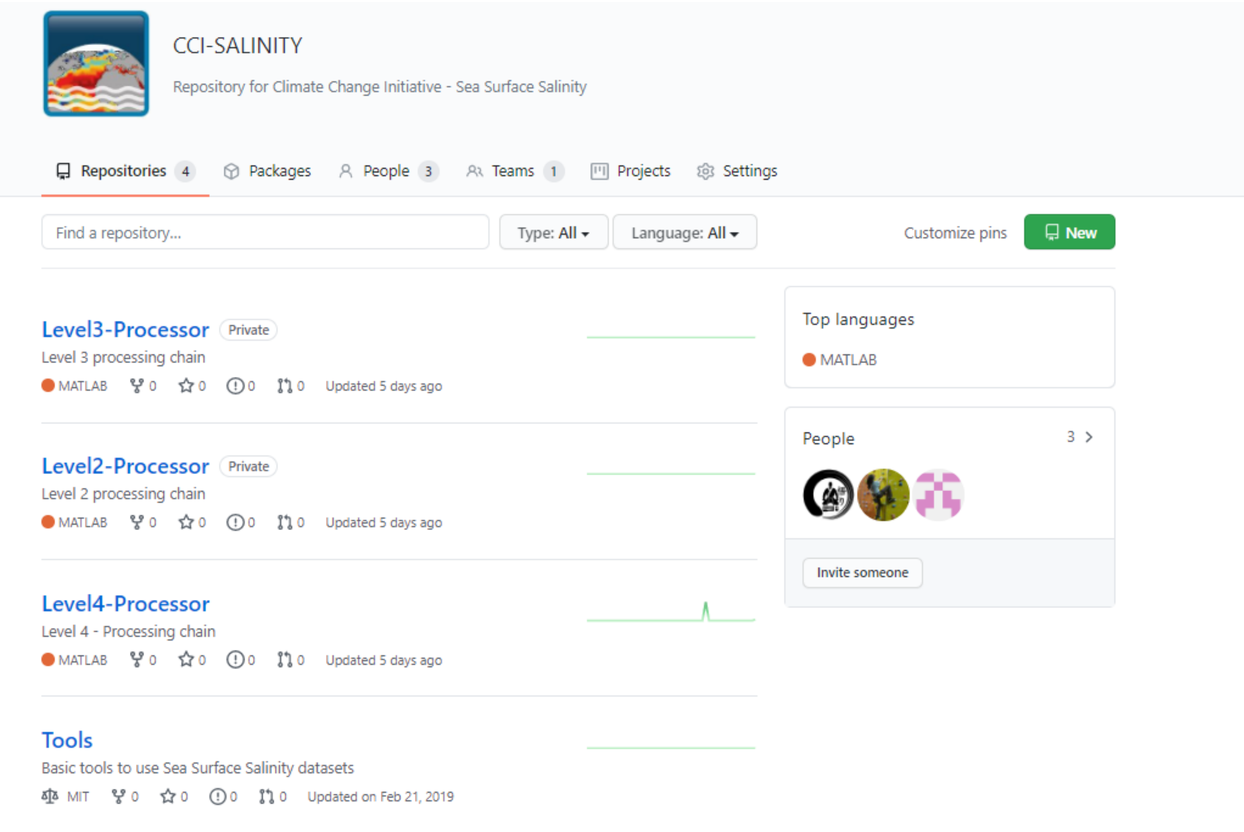


Figure 15: GitHub repository screenshot

The description of each item is also included in the web pages (Readme files).

FRouffi committed 9ee80f6 5 days ago | 37 commits | 1 branch | 0 tags

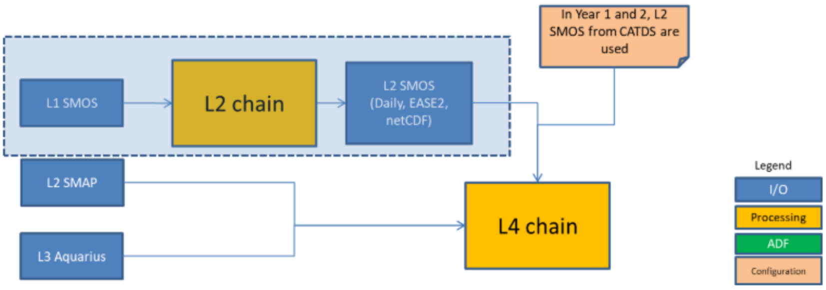
Year1	Add files via upload	4 months ago
Year2	Update image	5 days ago
README.md	Update README.md	5 days ago

README.md

Level4-Processor

Level 4 - Year 1 and 2 Processor

The processing chain used in Year 1 and 2 to generate the CRDP for SSS ECV is drawn on the following picture. 3 entries are used; one for each main satellite data inputs.



Level 4 - Processing chain

Readme

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

MATLAB 100.0%

FRouffi committed 29ea10d 5 days ago — History

..

extract_bias	Upload of year 2 version	5 days ago
lecL2_L3	Upload of year 2 version	5 days ago
merge_product_function	Upload of year 2 version	5 days ago
write_function	Upload of year 2 version	5 days ago
CCI salinity full production chain.png	Upload of year 2 version	5 days ago
CCI salinity production chain.png	Upload of year 2 version	5 days ago
CCI salinity satellite data 1.png	Upload of year 2 version	5 days ago
CCI salinity satellite data 2.png	Update image	5 days ago
Readme.md	Upload of year 2 version	5 days ago

Readme.md

Software for L4 product generation.CCI+SSS (year2)

aux_files

is the sub-directory which contains the auxiliary files required as input for other routines

(auxiliary files are not part of the source in github)

- latlon_ease.mat : EASE grid specification
- maskdmin_ease2.mat : EASE GP distance from coast
- SM_OPER_AUX_MINMAX_20050909T023037_20500101T000000_624_001_2.nc : SSS variability and SSS min/max values
- ERR_REP_50km1d_50km30d_smooth.mat : representativity error from Mercator (+ smoothing) between maps at 50km-1d resolution and maps at 50km-30d
- ERR_REP_150km7d_50km30d_smooth.mat : representativity error from Mercator (+ smoothing) between maps at 150km-7d resolution and maps at 50km-30d (Aquarius)
- isas_CATDS : SAS15 and NRT projected on EASE grid
- smos_isas_rmsd_ease_smooth.mat : rmsd of SSS from SMOS and ISAS data on EASE grid.
- smosA_20140101.mat : contains xswath specification

End of Document