Climate Change Initiative+ (CCI+) Phase 1

Sea Surface Salinity

System Specification Document (SSD)

Customer: ESA

Ref.: ESA-CCI-PRGM-EOPS-SW-17-0032

Version: v3.1

Ref. internal: AO/1-9041/17/I-NB

Revision Date: 15/09/2021

Filename: SSS_cci-D3.2-SSD-v3.1.docx

Signatures

Author	Frederic Rouffi ACRI-ST		15/09/2021
Reviewed by	Odile Fanton D'Andon ACRI-ST		
	Giovani Coratto AdwaisEO		
	Stephen Emsley ARGANS		
	Manuel Arias ARGANS		
Approved by	Jacqueline Boutin (Science Leader) LOCEAN		15/09/2021
	Nicolas Reul (Science Leader) IFREMER		
	Rafael Catany (Project Manager) ARGANS		15/09/2021
Accepted by	Susanne Mecklenburg (Technical Officer) ESA		

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

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/v3r1 (this document)	Phase 1 final version	Updated for Year 3 production

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 utilized to generate the Sea Surface Salinity (SSS) Essential Climate Variable (ECV) time series. It is structured as an answer to the System Requirement Document (SRD) aiming to specify 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. The content of this document follows the requirements as stated in the Statement of Work (SOW Task 3 SOW ref. ESA-CCI-PRGM-EOPS-SW-17-0032).

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 1 of the ESA Climate Change Initiative Extension (CCI+).

The system described in this document capitalizes on the multiple sources of data available so far and that have also been acquired during the 3 Years of phase 1 exercise: this encompasses satellite, in situ and all other relevant data that may confer the best value to the computation of the SSS ECV time series.

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 tuned along the way by the project Science Team during phase 1 in order to improve their reliability and adequacy with the CCI expectation; in particular in term of product format [AD-1].

The products generated by the system are customized according to the Users feedback as exposed in the project [URD]. It results from the requirements established in the [SRD].

This document is the final version of phase 1 proposed in Year 3. It accounts for experience acquired during the three Years of the project 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

Section 6: Round Robin description

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-i1r4
DARD	Data Access Requirement Document	SSS_cci-D1.3-DARD-v1r3
PSD	Product Specification Document	SSS_cci-D1.2-PSD-v2r0
SoW	CCI+ Statement of Work	
ATBD	Algorithm Theoretical Baseline Document	SSS_cci-D2.3-ATBD-v3.0
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, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva,

	Switzerland, 104 pp. All four documents contributing to the Fourth Assessment Report are available online at http://www.ipcc.ch/publications_and_data/publications_and_data_reports.htm
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 web site: http://cci.esa.int
RD-7	GCOS Climate Monitoring Implementation Principles, November 1999. Available online at: http://www.wmo.int/pages/prog/gcos/documents/GCOS_Climate_Monitoring_Principles.pdf
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-11	IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp. Available online at: http://www.ipcc.ch/report/ar5/wg1/
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

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 formally known as 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 - element of the European Earth Watch programme.
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, 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 for Year 3 (from Jan-2010 to Sep-2020) within the frame of the CCI+ Sea Surface Salinity ECV project. The main objective of such production is to provide the best support to the users and more globally to the science community in better understanding the 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.

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 design of the system is primarily based on the following specification:

- ✓ 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 production is global, i.e. covering all Earth water surfaces and the main users are Climate Research groups and modelers. Nevertheless, some specific geographic areas are also subject to targeted studies (e.g. C-Band sensors analysis encompassing main river plumes that will be implemented during phase 2 of the project).
- ✓ The production facility implements improved algorithms as defined and proposed by the Science Team in the [ATBD].
- ✓ The production facility includes the processing of data of all available sensors and in-situ measurements at 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 modelling community (i.e. including research, academia 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 the CCI data produced during the project lifetime are freely distributed by ESA.

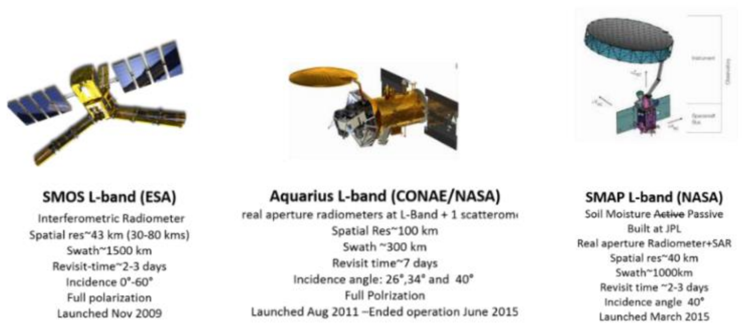
The needs expressed by the different project Teams (e.g. Science, Validation and Users) drive and have impacts on the architecture of the system. Indeed, during the project, many aspects either scientific or more widely technical may evolve. Therefore, in order to maintain the system up to date and operational, a capacity of adaptation and maturation are expected; in particular to cope with:

- The experience acquired from the system prototyping,
- The Round-Robin exercises or reprocessing demand aside from the generation of the CRDP,
- The CCI+ Salinity validation activity [WP250 and WP420],
- The feedbacks received from the Climate Research Group and Climate Modelling User Group and Data Engineering Working Group
- Improvement in technology,
- The evolution of the specifications (data format, algorithms, …)

Although the [SoW] clearly states that CCI+ Salinity project does not aim at building an operational processing system, the amount of data to be processed in the course of the project is considerable and it is increasing in time (i.e. yearly during the 3 Years life span of the project). 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 620) with a complete CCI reprocessing to level 2 SMOS SSS
- ✓ SMAP Level 2 from RSS (version 4.0)

  https://podaac.jpl.nasa.gov/dataset/SMAP_RSS_L2_SSS_V4

- ✓ AQUARIUS Level 3 from NASA (version 5)

  https://podaac.jpl.nasa.gov/dataset/AQUARIUS_L3_SSS_SMID_ANNUAL_V5

Main updates in version 3 of the dataset with respect to version 2 are as follows:

- SSS from SMOS have been generated from a complete L2 reprocessing with the following updates: OTT correction computed from ISAS-Argo instead of WOA climatology. Specific RFI filtering. ERA5 auxiliary data instead of ECMWF forecasts dataset. Dielectric constant model of [RD-33] (instead of [RD-34] in v2.3).

- Ice mask has been computed from SMOS retrieved pseudo dielectric constant (Acard parameter) estimated over the whole period

- SMOS seasonal latitudinal biases have been computed by using Pacific + Atlantic region (only Atlantic used in v2.3)

- Instantaneous rain effect has been corrected, relating surface salinity freshening to IMERG rain rate following [RD-32], before estimating bias correction and before L4 merging for SMOS and SMAP SSS. Bulk SSS is now available in the L4 product.

- SSS random uncertainty computation has been updated.

- Aquarius SSS have been resampled on the EASE 2 grid using an interpolation with a distance weighting (instead of the closest neighbour algorithm in the v2.3)

- SSS is now provided much closer to coast, but additional pixels are flagged with the same land-sea mask as in v2.3. So, users who wish to ensure using same pixels as in v2.3 should use this flag; users interested in S variability very close to coast, should not apply this flag but should use data close to coast with care.

- Representativity uncertainties considering the various spatio-temporal scales covered by the various sensors are taken into account for all sensors (only for Aquarius in v2.3) when estimating L4 fields.

This processing chain is further detailed in section 4.

The system is constituted with:

- SW inherited from existing chains. The processors are re-used as is or adapted to fit the specific needs of the project (resolution, grid, product format, interfaces …).

- Specific pieces of software developed for data adaptation and preparation (pre-processors) to prepare the required auxiliary files from different sources and to match the spatial and time acquisitions of the various EO instruments employed for the retrievals of SSS, or like post-processors to homogenize the datasets and comply with the requirements of the project providing datasets meeting some standards (e.g. netCDF/CF convention standards/CCI data standards see [DSTD]). The modularity of such system is then eased and ensured by construction through the use of harmonized dataset and duly interfaced processing items.
- CCI+SSS specific processors e.g. Level 4 SSS programs processing the data produced by the abovementioned elements.

This is applicable to the SMOS, SMAP and Aquarius processing chains 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 outcome of the Science Team work.

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 to get a pre-operational system at the end of each exercise.

The final chain is based on inputs from SMOS Level 1C products, L2 SMAP and L3 AQUARIUS.

2.4 Data access

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

3 Solution justification

3.1 Overview

The selection of the means to perform the computation of SSS ECV within CCI+ project has been studied; 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

Among the available systems that may be used or setup within the frame of this project, one can cite:

- Cloud computing
- Cluster computing

In some cases, a trade-off analysis may be performed to decide the alternative concepts offering the best solution applicable to a project.

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, 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.
- ✓ Moreover, the Level 2 SMOS Ocean Salinity processing, which is the most demanding processor in terms of computational resources, requires to be carefully scheduled because some dynamic ADFs are recomputed at the end of each acquisition day. Associated to the usage of openMP library, the processor is quite well fitted to be operated in a distributed memory environment, on machines with similar OS, with multi-cores available.
- ✓ The system is also setup 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 Architecture of the system

4.1 Project workflow

The SSS ECV production system is triggered upon updates/changes on 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 testing 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

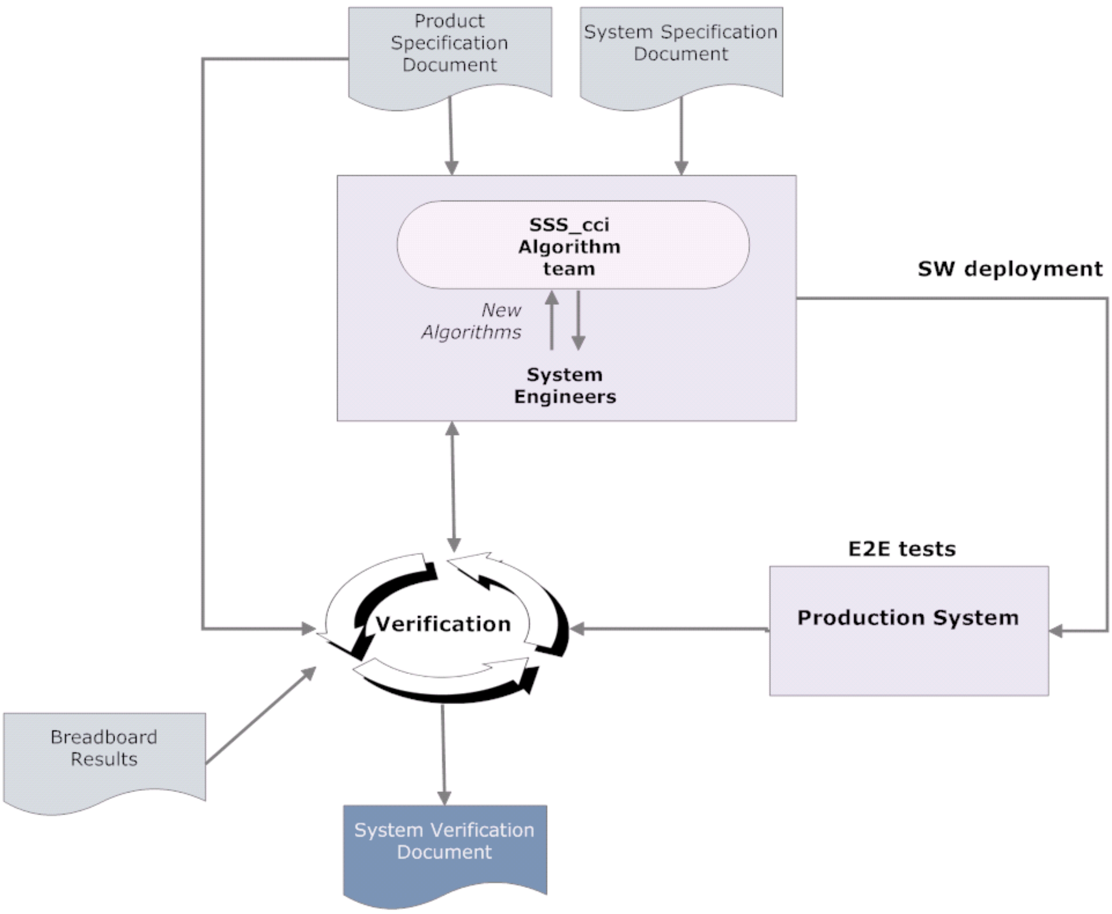


Figure 1: CCI+ Salinity system verification

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 dataflow between the above components and with the main external interfaces.

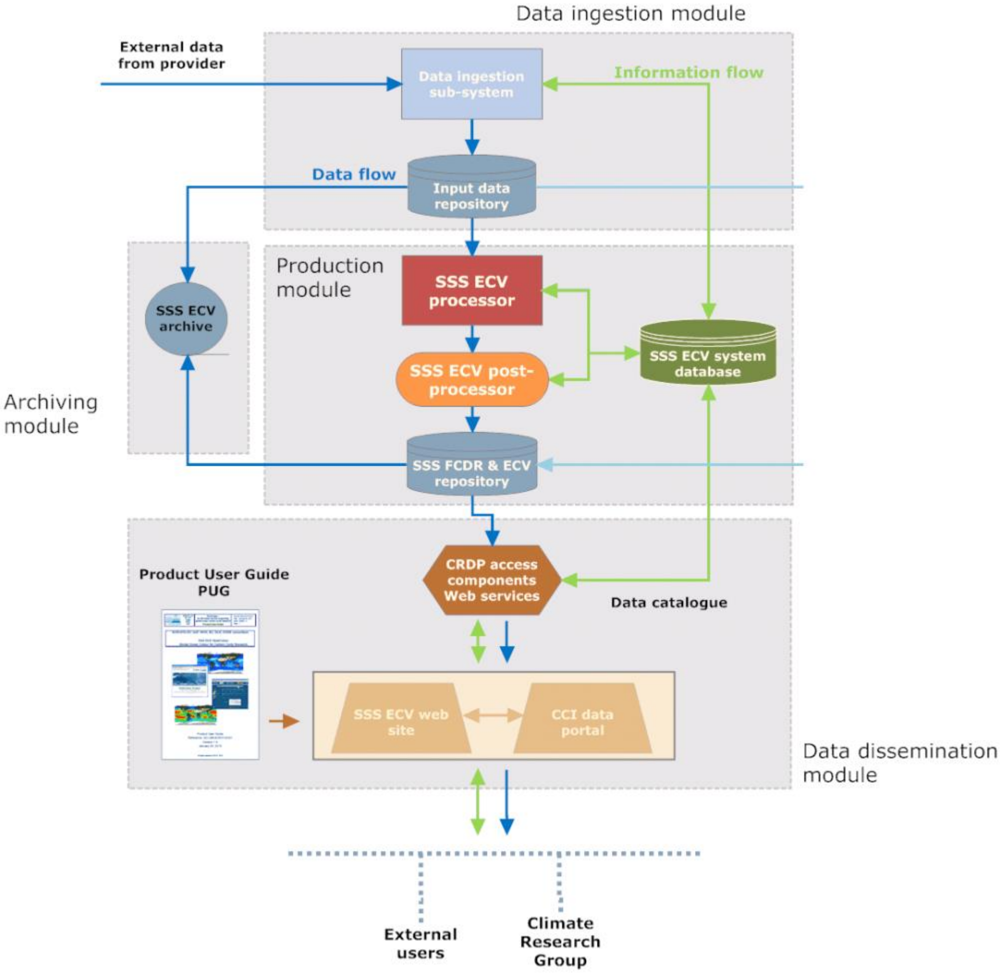


Figure 2: CCI+ Salinity production dataflow

The system is designed to answer the following matters:

- Being a help for the CCI+ Salinity Science Team to perform regular and performing computation in view to support them 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 purpose.
- Fully addressing the CCI+ expected production volume by running a complete end-to-end ECV processing system.
- The specification of the production system is a 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 modules data and information flows logic 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 retrieved input products,
- Ingesting the products in the CCI+ Salinity file repository
- Deriving metadata (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 inputs products, auxiliary data files, outputs directory. 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 to 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 customize 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, 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 on 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 dataflow 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 realized 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:
    - It supports the round robin exercises to help the science team in the tuning of their algorithms,
    - The full reprocessing activity doesn't last more than 4 months,
    - The EO missions to be considered includes both L-band (implemented in Year 1 to 3) and C-band radiometers (in phase 2)
- 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)	3 TB/yr	11 years (2009 – 2021)	33 TB
SMAP (L1)	511 GB/yr	6 years (2015 – 2021)	3 TB
Aquarius (L1)	27.4 GB/yr	4 years (2011 – 2015)	0.2 TB
SMOS L2 (UDP)	85 GB/yr	11 years	0.9 TB
SMOS L2 (DAP)	2.7 TB/yr	11 years	29.1 TB

SMAP (L2)	85-585 GB/yr	6 years	0.5-3.5 TB
Aquarius (L2)	6.6-34 GB/yr	4 years	0.03-0.2 TB
SMOS (L3)	0.4-9 GB/yr	11 years	0.04-0.1 TB
SMAP (L3)	0.01-8 GB/yr	6 years	Negligible-0.05 TB
Aquarius (L3)	2 GB/yr	4 years	0.01 TB
SMOS (L4)	0.25-77 GB/yr	11 years	Negligible-0.85 TB
Aquarius (L4)	0.06 GB/yr	4 years	Negligible
Total (rough estimations)			**67 – 72 TB**

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

SMOS AUX files (L1 to L2)	1 TB/yr	11 years	11 TB
SMAP AUX files (L1 to L2)	100 GB/yr (est)	6 years	0.6 TB
Aquarius AUX files (L1 to L2)	10 GB/yr (est)	4 years	0.1 TB
ISAS	27 GB/yr	11 years	0.3 TB
ARGO (Individual profiles)	7 GB/yr	18 years	0.2 TB
SST (OSTIA as proxy)	100 GB/yr	18 years	1.8 TB
WS (MetOp)	16 GB/yr	9 years	0.2 TB
Altimetry (Along-track)	1.5 GB/yr	18 years	Negligible
Total (rough estimations)			**15 TB**

Table 3: Initial estimation of storage requirements based on C-band data products.

AMSR-E (NSIDC-0301)	20 GB/yr	19 years	0.4 TB
AMSR-II (L1-R)	360 GB/yr	9 years	3.2 TB
Total (rough estimations)			**3.6 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...
- Due to the size of data, a "waiting for data" process queue is implemented in the system.
- If tapes are used, a message is sent to the operator with all information required for the data retrieval.
- The system is running in data-driven mode
- The load balancing of CPU and I/O is also very important when processing huge amount of data: pools of nodes is 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 send 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: contain 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 (level 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 of the product type. The storage area is based on file servers linked to the computation nodes though 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, of 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, web site...) 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).

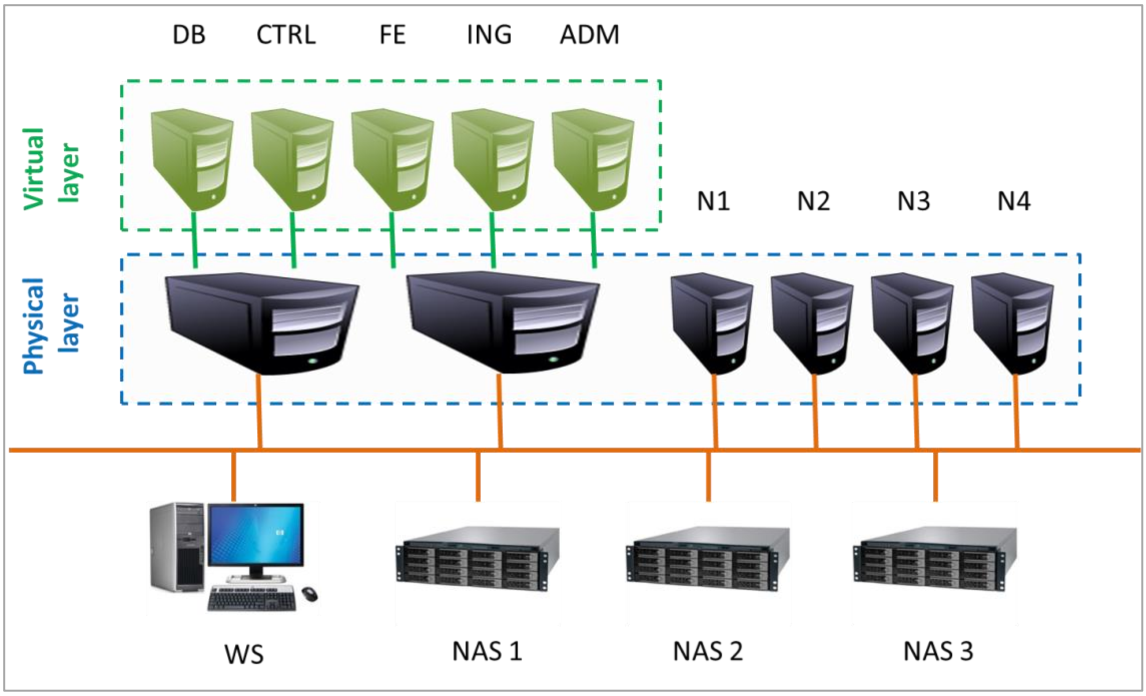


Figure 3: CCI+ Salinity processing system implementation

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 into dedicated database tables and visible from the computation nodes, the volumes can be used to store information. These NAS can also be connected to a ftp server (or other protocols such as WebDAV) to enable a 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:

- ✓ Direct models and retrieval implemented in SMOS L2OS processor version 6.71 (see ATBD available on https://smos.argans.co.uk/docs/deliverables/delivered/ATBD/SO-TN-ARG-GS-0007_L2OS-ATBD_v3.13_160429.pdf
- ✓ SMAP RSS products are available on

  https://podaac-tools.jpl.nasa.gov/drive/files/allData/smap/L2/RSS/V4/SCI
- ✓ AQUARIUS products are available on

  https://opendap.jpl.nasa.gov/opendap/SalinityDensity/aquarius/L3/mapped/V5/daily/SCI

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

The Processing L4 products (merge of several satellite SSS) is performed using a debiasing method similar to the one described in Kolodziejczyk et al. (2016) on which some adjustments are performed to take into account the various random and systematic errors of the various data sets. The algorithms implemented in the Level 4 processing chain are described in CCI+SSS document [ATBD].

4.5 Testing of new Algorithms

4.5.1 Round Robin testing

The CCI+ project aims at improving the existing algorithms to produce the best ECV time series during the round robin exercises.

Round robin (RR) are systematic tests that give a quick overview of the behavior of satellites products with respect to a set of in-situ data used as a reference. These tests have been used for now to help in the selection of CCI+SSS version 1 to version 3. For these tests, only data in the Atlantic Ocean have been used.

RR tests consist of computing metrics between different satellite products and in-situ data. These metrics are then compared to determine which product is closer to in-situ measurements. It is possible to compare one product to several references, for example to compare one CCI product to original processing (SMOS, SMAP and Aquarius).

A pdf file written using LaTeX is automatically generated resuming the results of the tests (PVASR).

The RR are based on comparisons with four kinds of in-situ data. Below is a short description of in-situ data, that can be useful to understand input files (section 2). The Product Validation and Algorithm Selection Report (PVASR) gives a more complete description of these data and of the methodology of the tests.

4.5.1.1 Round Robin in situ data

Repetitive merchant ship tracks

Repetitive ship measurements are used as a reference in RR. These data allow to estimate spatial and temporal statistics of satellite products.

To distinguish between different ship tracks, areas have been defined as in Figure 4. If the majority of the measurements of a ship transect is included in one of these areas, it is considered as following the corresponding ship track. For RR CCI v1 and v2 only AX20 and AX11 have been used.

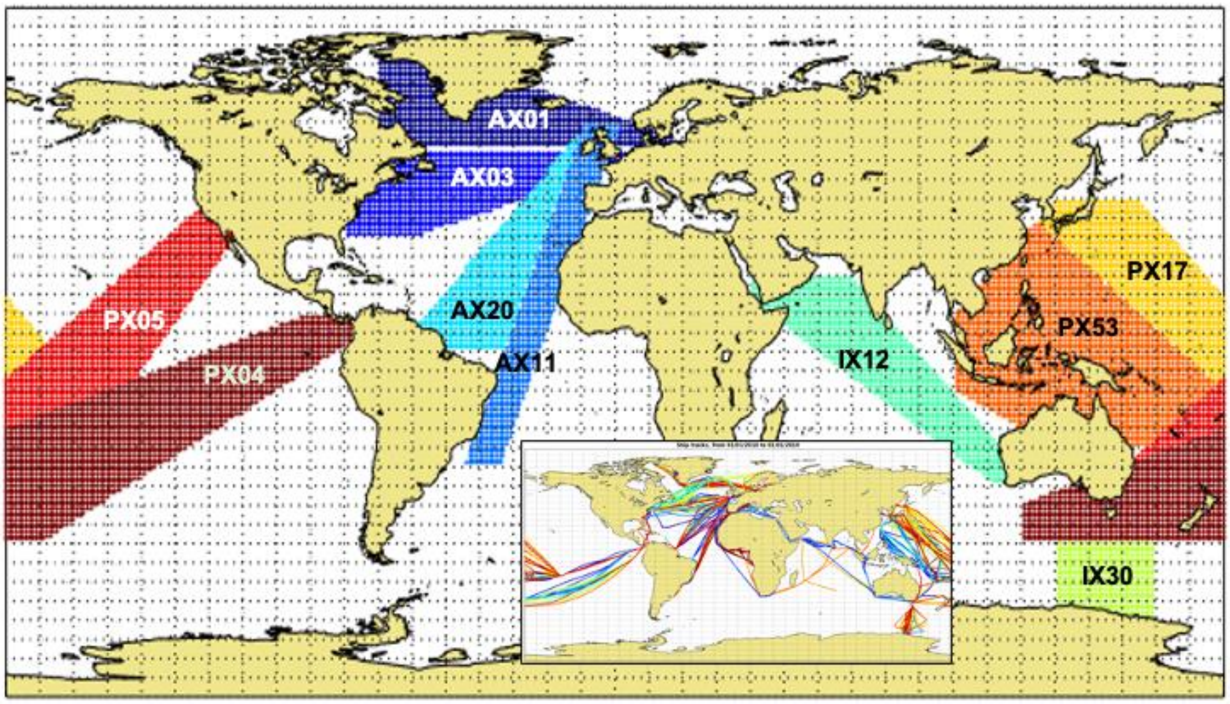


Figure 4: Areas definition for treatment of merchant ships tracks

All measurements of the same ship track are concatenated, and it is possible to apply a flag on insitu data at this stage with the field “flgis” in input.txt (see section III.6.1.1).

Statistics are computed in boxes which dimensions are defined in the input file. To do so, a “mean transect” is computed over each area used in RR, and dimensions of boxes given in km or in degrees are calculated along this mean transect (see Figure 5)

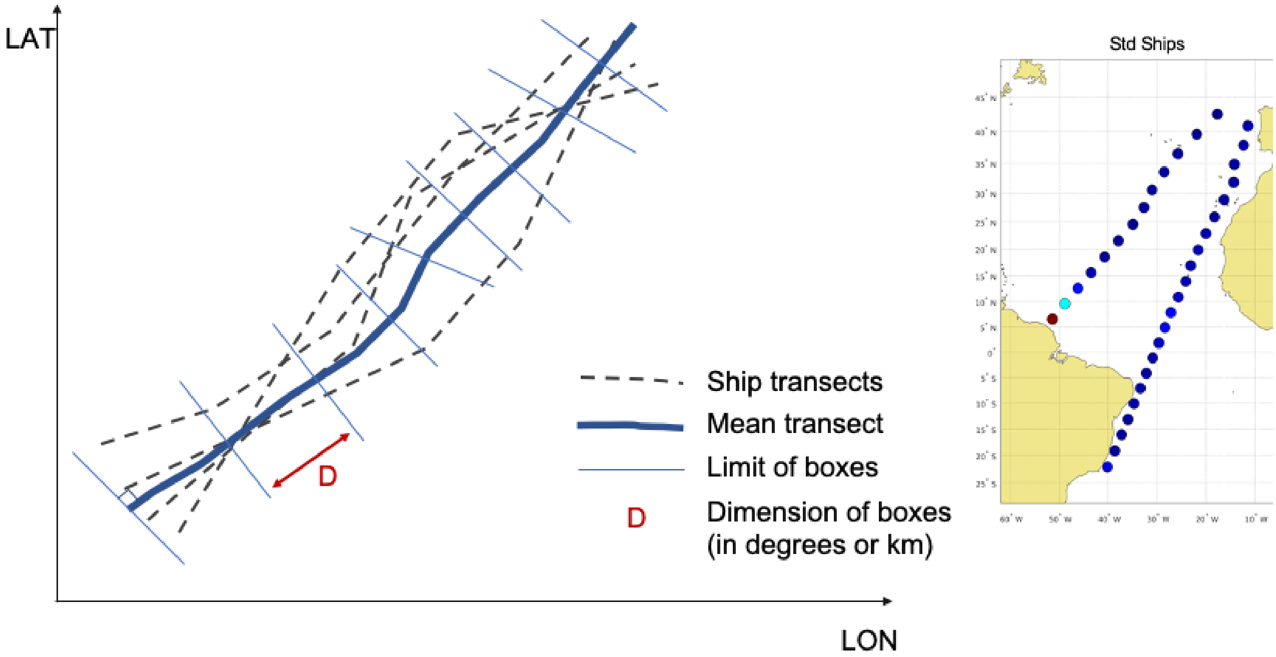


Figure 5: Methodology for the determination of boxes for the calculation of metrics over merchant repetitive ship tracks

Moorings

Three datasets of moorings are available for Round Robin tests: Prediction and Research Moored Array in the Tropical Atlantic (**PIRATA**), Tropical Atmosphere Ocean (**TAO**), Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (**RAMA**).

Binned ship tracks in northern latitudes

Ship transects in northern latitudes have been binned over boxes as described in PVASR. 4 transects are available: BAX01, BAX02, NAX01 and GAX01. Only BAX01 and BAX02 are used up in RR up to CCI v2 as other transects are too close to coast (see Figure 6).

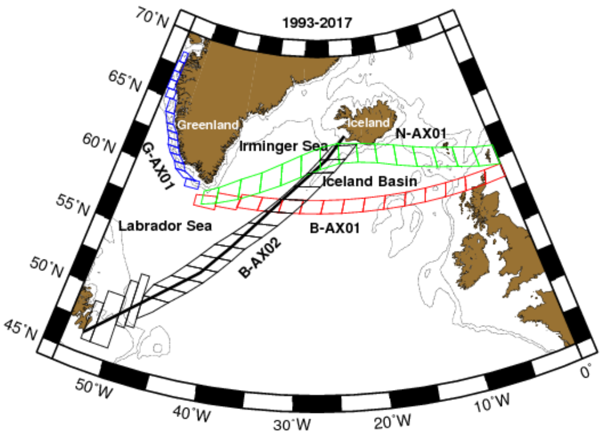


Figure 6: Definition of binned ship tracks in northern Atlantic

Science ships measurements

Science ship available for RR have been downloaded from GOSUD database 2019[1].

It includes for now the following tracks: Tara, Beautemps_Beaupre, Fleur_Australe, Pourquoi_Pas, Atalante, Thalassa, Le_Suroit, La_Louise, Mariposa, RARA_Avis, STAMM, BarkEuropa, Boogaloo.

4.5.1.2 Round Robin program architecture

On Figure 7; we present a description of the architecture of Round Robin program: squares correspond to input files, ovals to matlab programs and parallelograms to output files. Program description is given in annex.

There are two input files to fill before launching Round Robin tests, described in the annex. Input files must be in the same directory as the executable. The processor has been optimized using multi-threading.

Round Robin produce two kinds of output files, reports and figures that can be used in presentations and save files (format .mat or netCDF) (see sections 5 and 6 for further details).

[1] Kolodziejczyk Nicolas, Diverres Denis, Jacquin Stéphane, Gouriou Yves, Grelet Jacques, Le Menn Marc, Tassel Joelle, Reverdin Gilles, Maes Christophe, Gaillard Fabienne (2019). Sea Surface Salinity from French RESearcH Vessels : Delayed mode dataset, 2001-2018. SEANOE. https://doi.org/10.17882/39475#65555

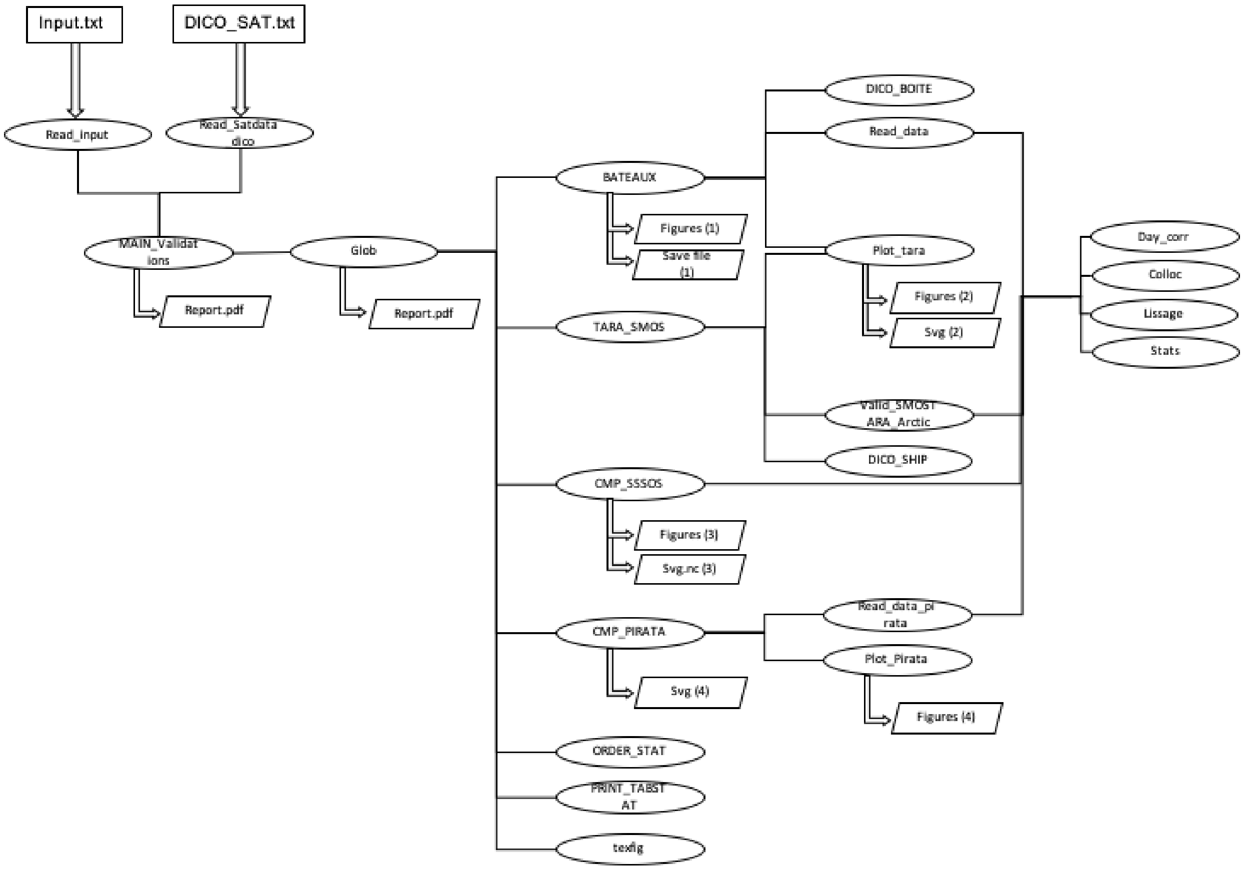


Figure 7: Architecture of Round Robin program

4.5.2 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 are SSS from SMOS or SMAP. SSS is corrected from various systematic errors, as land-sea contamination systematic errors, SSS systematic errors dependent on SST and, in case of SMOS from seasonal systematic errors 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 the EASE2 grid at 25 km resolution.

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].

4.5.3 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 been already corrected for some biases (e.g. land sea contamination or spatio-temporal drifts 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 V2 (in V1, the 7 days running product has been used).

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].

4.5.4 L4 algorithm evolutions

In a first stage, the generation of L4 products has consisted to setup and tune the Level 4 algorithm by using existing L2/L3 input products from different sensors producing salinity.

In a second stage, some new algorithms have been developed:

- to improve the current processors
- and to align the inputs.

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

4.5.5 Processing chain evolutions

The SMOS L2 processor has been tested using a new dielectric constant model, written following a single Debye relaxation law, derived according to the model of Somaraju and Trumpf (2006) [RD-30] and Boutin et al. (2021) [RD 32].

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 on Figure 8. Three entries are used; one for each main satellite data inputs. The CCI+ SMOS L2 processing chain was not used for the two first CRDPs (Y1 and Y2); instead, the data was retrieved from the CATDS production chain (IFREMER/CNES/ESA). This link is now discarded. In Year 3, it was replaced with the full L2 processing chain that has then been setup and activated (see section 4.6.1.1)

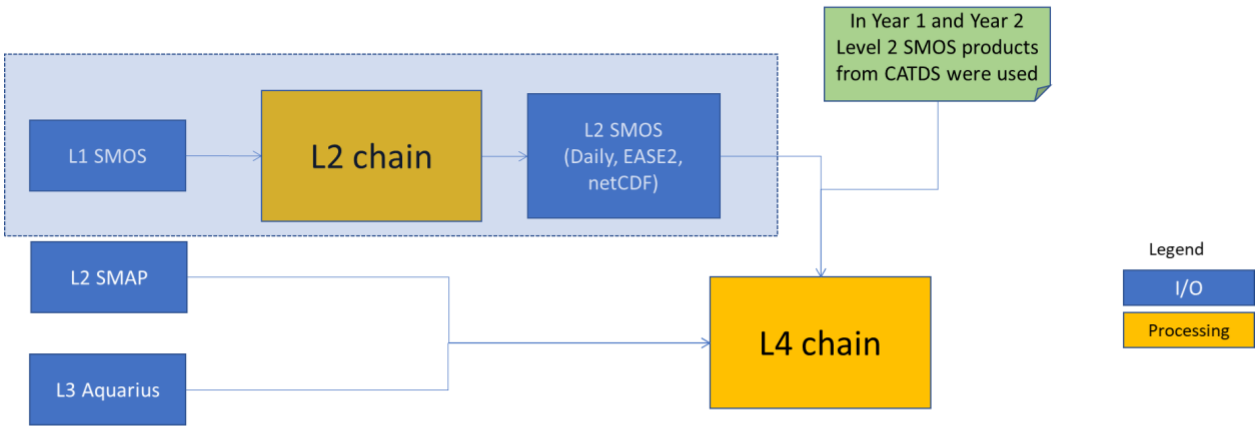


Figure 8: CCI+ Salinity production chain

4.6.1.1 Level 2 processing chain

SMOS

The L2 product generation chain for the CCI+ SSS ECV project is detailed on Figure 9. This production system has been deployed in Year 3 of the project.

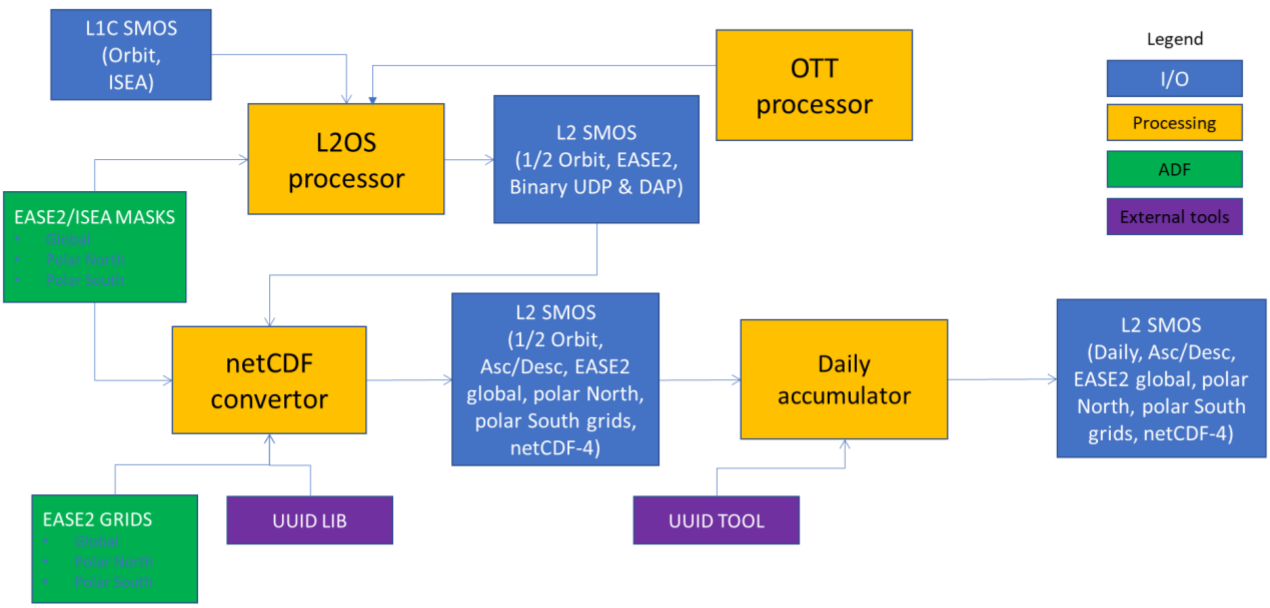


Figure 9: CCI+ Salinity processing Level 2 chain that has been deployed in Year 3

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 3 grids: an EASE2 rectangular 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. Rectangular grid is the EASE2 global cylindrical equal-area projection (see NSIDC page https://nsidc.org/data/ease/ease_grid2.html). 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. They have been respectively downloaded from

https://podaac-tools.jpl.nasa.gov/drive/files/allData/smap/L2/RSS/V4/SCI

and

https://opendap.jpl.nasa.gov/opendap/SalinityDensity/aquarius/L3/mapped/V5/daily/SCI

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:

- Weekly L4 and
- Monthly L4

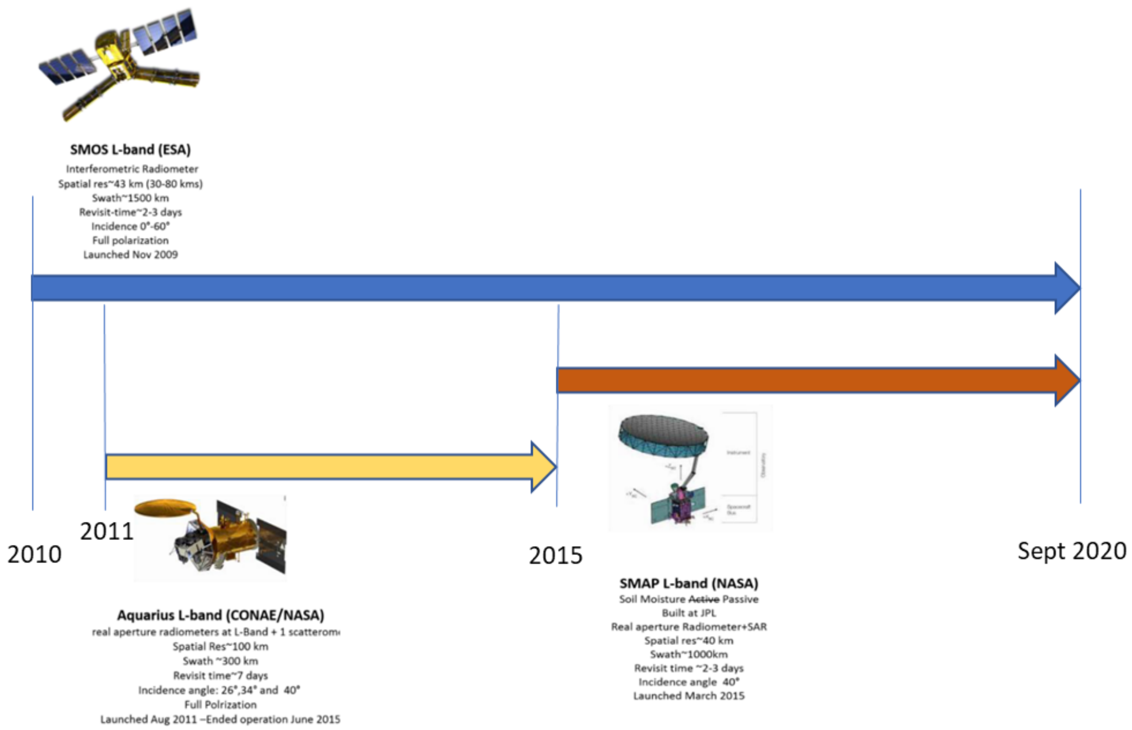
Both products are formatted in NetCDF and are in conformance with the data format convention applied on the CCI projects [DSTD]. The product structures are described in [PSD].

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)
- Other sensor data such AMSR (C-BAND) will be included during phase 2 of the project.



Satellite	Y1	Y2	Y3
SMOS	10	10	11
AQUARIUS	5	5	5
SMAP	4	5	6

Figure 10: 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 10 provides the acquisition coverage for each satellite.

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

  Version of the Level 2 OS processor is 6.71. Data are projected on an EASE2 cylindrical equal area grid (It is a global coverage grid at 25 km resolution.), separated from ascending and descending acquisition, aggregated on daily map, and formatted in netCDF.

  Dataset name example and quick-look image (Figure 11):

  SM_TEST_MIR_OSUDP2_20160826T111930_20160826T121249_662_001_8

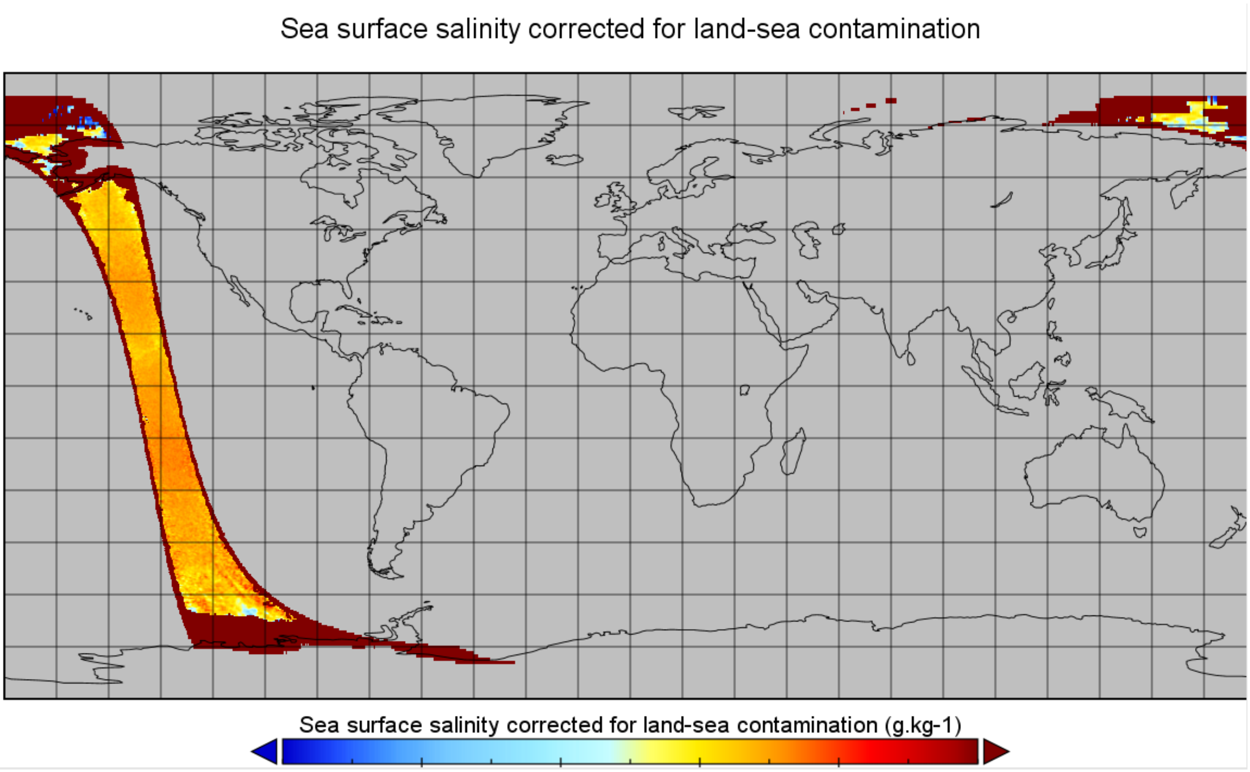


Figure 11: SMOS Salinity daily map (one time-step is represented)

- ✓ AQUARIUS Level 3 data are used in version v5.0 with a latitudinal correction within AQUARIUS processing using ARGO data. No extra correction added. We use the daily non averaged dataset for generating the CCI+SSS L4 dataset V2 (in V1, the 7 days running product were used).

  The original L3 mapped products are given on a Plate-Carrée equidistant cylindrical grid. They consist in 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 12):

Q2012002.L3m_DAY_SCIA_V5.0_SSS_1deg.bz2

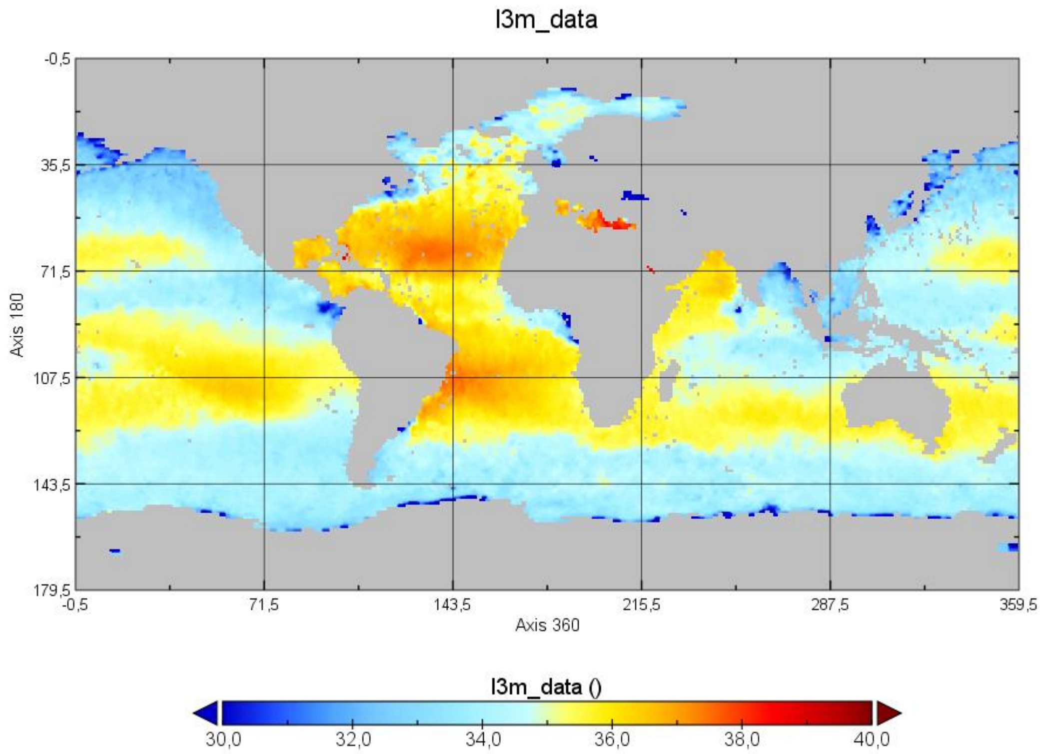


Figure 12: Aquarius Salinity 7-days map

✓ SMAP Level 2 data in version v4.0 are used with a latitudinal correction within SMAP RSS processing using ARGO data. No extra correction added.

The Original L2C products have 40 km and 70 km (39 km x 47 km elliptical footprint) spatial resolutions. They are based on L1B SMAP RFI filtered antenna temperatures version 4 [SMAP_L2C]. Data are formatted in netCDF.

Dataset name example and quick-look image (Figure 13):

RSS_SMAP_SSS_L2C_r00870_20150401T004312_2015091_FNL_V04.0.nc

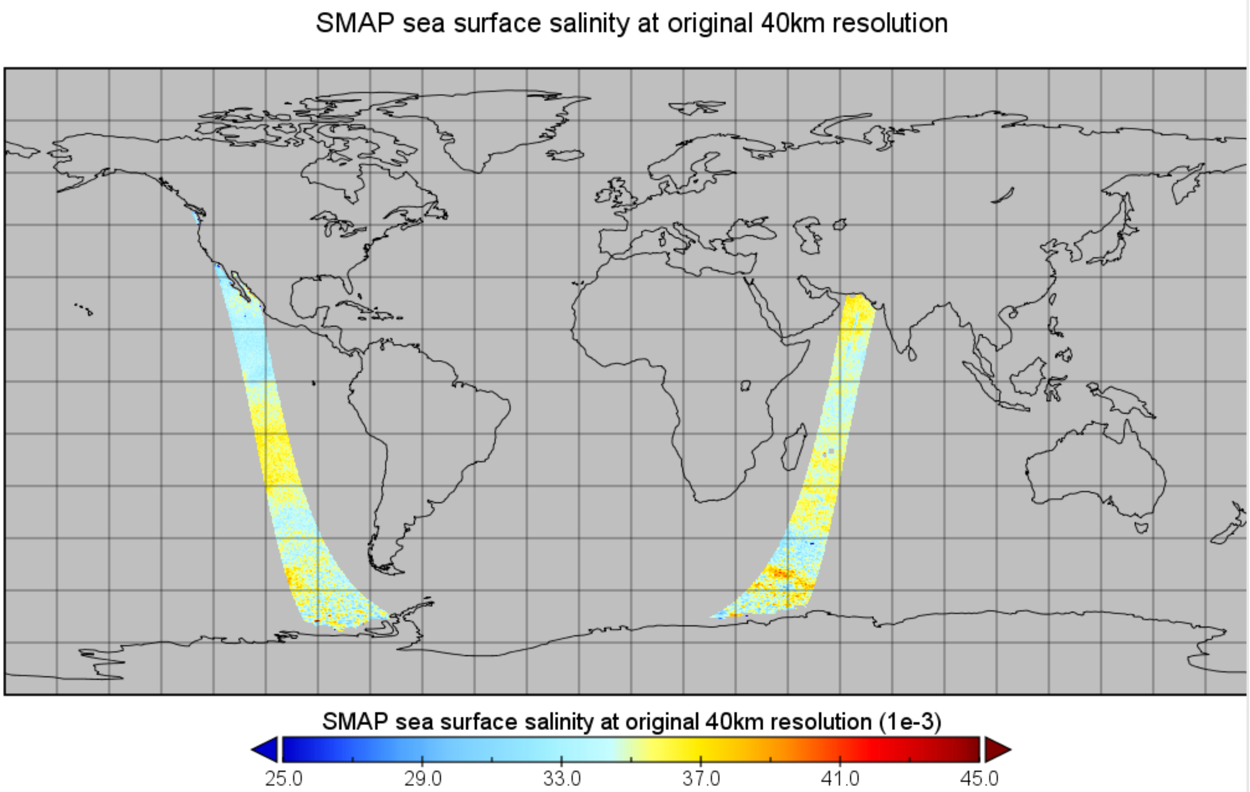


Figure 13: SMAP Salinity map

A projection of the SSS on the EASE2 grid has been applied on 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) are also used (see a description in [RD-34] document). 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. The monthly fields are also used to compute the OTT (see below).

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

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

Ex: ESACCI-SEASURFACESALINITY-L4-SSS-MERGED_OI_7DAY_RUNNINGMEAN_DAILY_25km-20100111-fv3.21.nc

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

Ex: ESACCI-SEASURFACESALINITY-L4-SSS-MERGED_OI_Monthly_CENTRED_15Day_25km-20100201-fv3.21.nc

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

- ✓ monthly and weekly SSS fields : obtained from 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 V2) : SMOS ice detection (Dg_ice descriptor from SMOS L2OS product) is integrated over a period of +/-30 days for monthly data or +/-10 days for weekly data. If the integrated value is greater than 0, the ice flag is raised.
- ✓ coast flag (CCI V2): 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 14 shows the SMOS L2OS processing data flow logic.

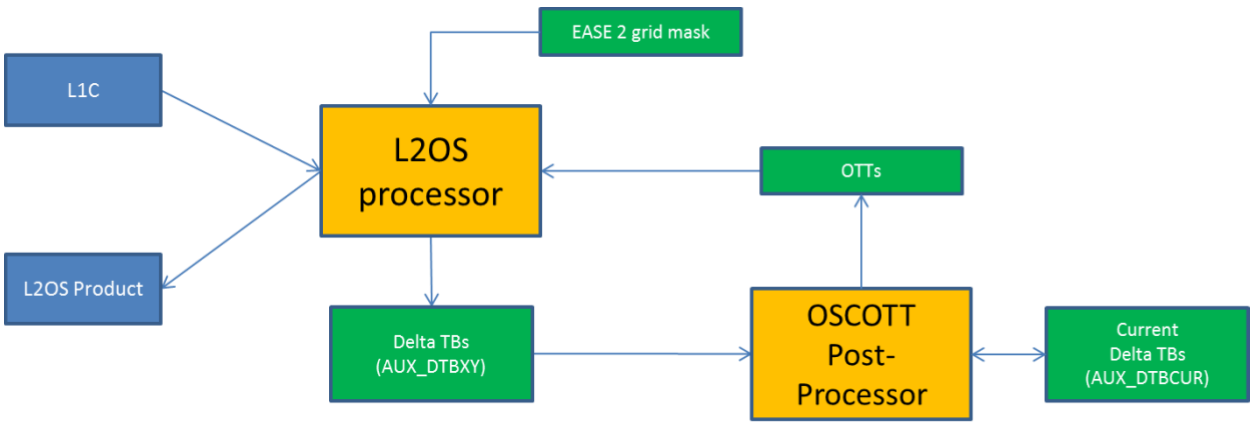


Figure 14: Overview of the SMOS L2OS/OSCOTT processor data flow

4.9.1.1 Performance testing

The following table summarizes the performances obtained with the L2OS processor with a variable number of cores. 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 1 core is not the best solution and doesn't 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 4 : L2OS Processing time as a function of number of cores

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

4.9.1.2 Processing logic based on interlaced L2OS and Ocean Target Transform

As depicted on the previous figure, the SMOS L2OS processor is triggered on L1C products. It is important to note that the run requires specific dynamic ADF named OTT for Ocean Target Transform. These ADF are generated by a side processor named OSCOTT that uses the intermediate Delta TB ADF created by the L2OS processor.

The computation flow is the following:

- ✓ The L2OS processor is launched on a full day of acquisitions (o = 29 : 15 ascending nodes and 14 descending nodes)

- ✓ The Delta TB issued by the L2OS are used as input to the OSCOTT processor that in turn generates updated OTT and an ADF called Current Delta TB that contains a 10 days sliding window information on delta TB; nominally centred on the day processed by the L2OS (5 days before and after the date).
- ✓ In order to reach the nominal operating point, the processing of 10 days of SMOS data is necessary to initialize the process (see Figure 15).

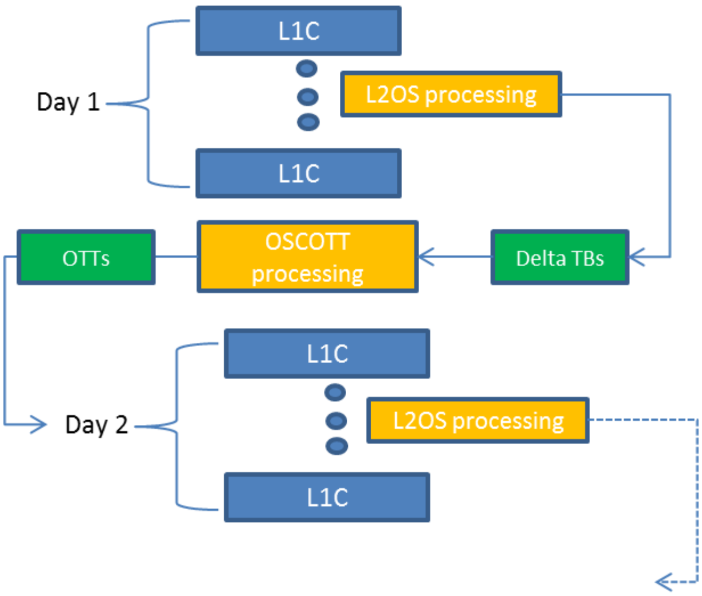


Figure 15: SMOS L2OS/OSCOTT initialization flow

The above processing flow also evidences that the parallelisation of the SMOS processing is limited to one full day of half-orbits since the result of the OSCOTT run is expected before starting the next day.

Nevertheless, there is a possibility to split the production in separate chains of N-Days of data.

- ✓ The counterpart is to initialize each chain with 10 days of processing in order to reach the nominal behaviour (processing overhead to initialize the sliding window for the OTT computation).
- ✓ The number N of chains fully relies on the available hardware (number of cores, blades, RAM, I/O capacity)
- ✓ The multithreading capability is deactivated.; so 1 core is used at a time.
- ✓ The machines considered in the evaluation below are 32 cores with 256 Gb of RAM. It is likely that a mix of 28 and 32 cores will be used during the reprocessing.
- ✓ One L2OS instance consumes about 6 Gb of RAM

Example:

Parameter	Value
Number of orbits per day	14.5
Number of half-orbits per day	29
Number of days of SMOS data	4015 (11 Years)
Half-orbit processing time	1h30 on 1 core
A day of data processed in	90 minutes (29 instances in parallel on 1 machine which leaves enough resource for the system)
OSCOTT processing time	15 minutes per day

If the SMOS dataset is split in 1 Year of data, 11 machines are necessary to process it.

Accounting for the 10 days overhead, the 11 Years of data are processed in parallel on the 11 machines in:

(4015 + 110) / 11 * 105' ~ 28 days

On 15 machines, the processing time goes down to:

(4015 + 150)/15 * 105' ~ 21 days

4.9.1.1 Processing logic based on separate computation of L2OS and Ocean Target Transform

To overcome the restriction in parallelisation, all the inputs (delta TBs) to OSCOTT can be generated by the L2OS processor in OTT mode before triggering the computation of the OTT. Indeed, 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 16).

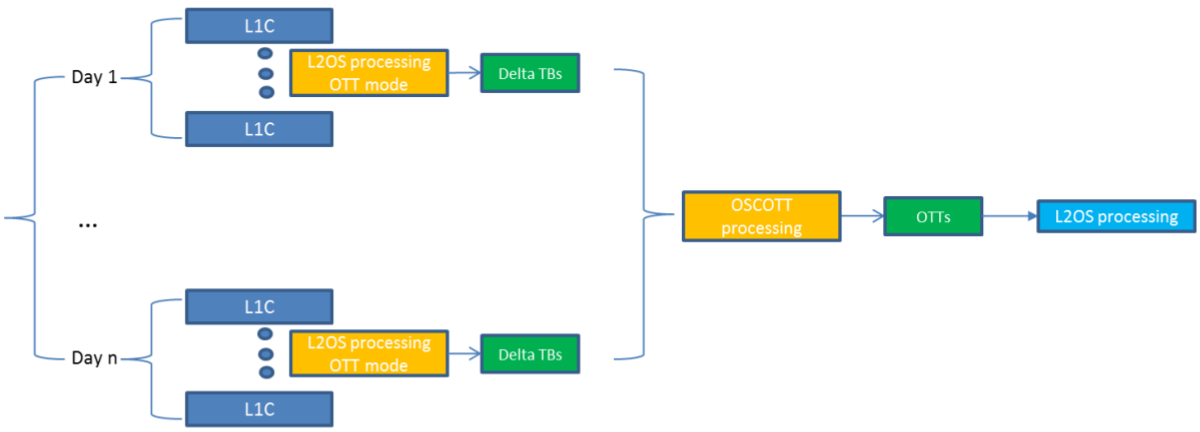


Figure 16: SMOS L2OS/OSCOTT scenario for an optimal parallelisation

With the above schema, the order of the computation doesn't 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 nominal mode in order to generate the Level 2 products.

We assume that 15 machines with 32 cores can be dedicated to the task. Leaving one core to the system, 465 processes can be run in parallel.

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 are used as input to the full L2OS processing in:

4015 * 90' / 465 ~ 16 days

We see that the total processing time is slightly better for the second which is also simpler to setup (no waiting process);

This is the configuration retained for the Year 3 production. It has evidenced a very good efficiency.

4.9.1.2 Input data

The Level 1C SMOS product contains multi-angular brightness temperatures in 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 for 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 initialize ESA EARTH EXPLORER CFI functions
AUX_MISP__	Mispointing angles used to initialize 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	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 5: SMOS L2 static ADF list

Dynamic ADFs

Dynamic inputs	Origin	Comment
AUX_BULL_B	Trace included in BULL_B used to initialize ESA EARTH EXPLORER CFI functions	IERS Bulletin B Same BULL_B for whole day
MPL_ORBSCT	FOS	Orbit Scenario File
AUX_ECMWF_	ECMWF pre-processor	One ECMWF per half orbit
AUX_OTT1F_	OSCOTT pre-processor Each contains 10 days of DTBXY as generated by the L2OS (sliding window)	Ocean Target Transformation LUTs including asc/desc values derived for the 3 roughness models The same OTT is used for the whole day of SMOS data
AUX_OTT2F_		
AUX_OTT3F_		
AUX_DGGRFI	ESA DPGS	Current RFI probability on the DGG

Table 6: 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 7: SMOS L2 output files

4.9.1.5 OSCOTT post-processor

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

Table 8: OSCOTT input list

Outputs	Description
AUX_OTTxF_	Ocean Target Transformation LUTs including asc/desc values derived for the 3 roughness models
AUX_DTBCUR	Current delta TBs (to be further reused as input to the next OSCOTT run)

Table 9: 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 10: 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) one and configured for SMAP by Meissner et al. in 2017, 2018 – [RD-31]. The algorithm flow is represented on Figure 17.

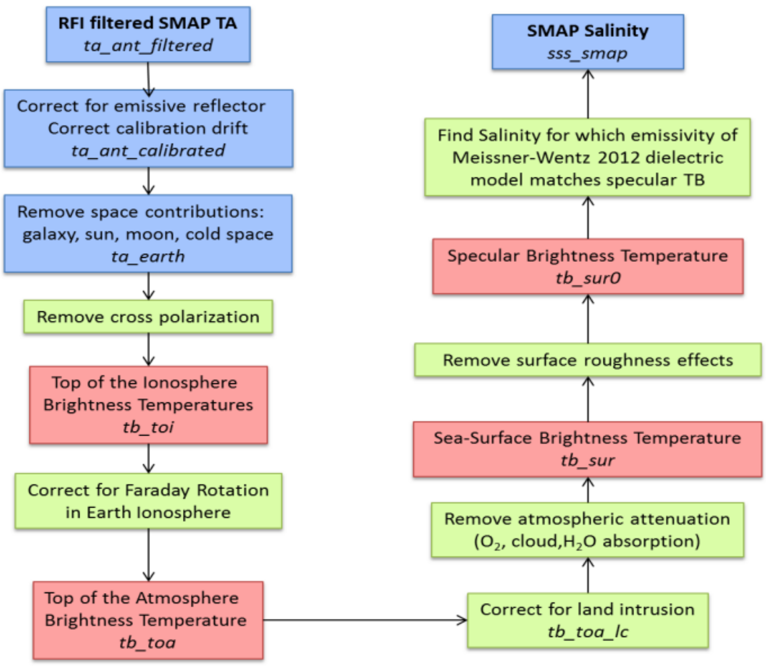


Figure 17: SMAP salinity retrieval algorithm flow diagram

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 18) can be retrieved here: ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/L1/SSS_1a/

4.9.3.2 Processing resources

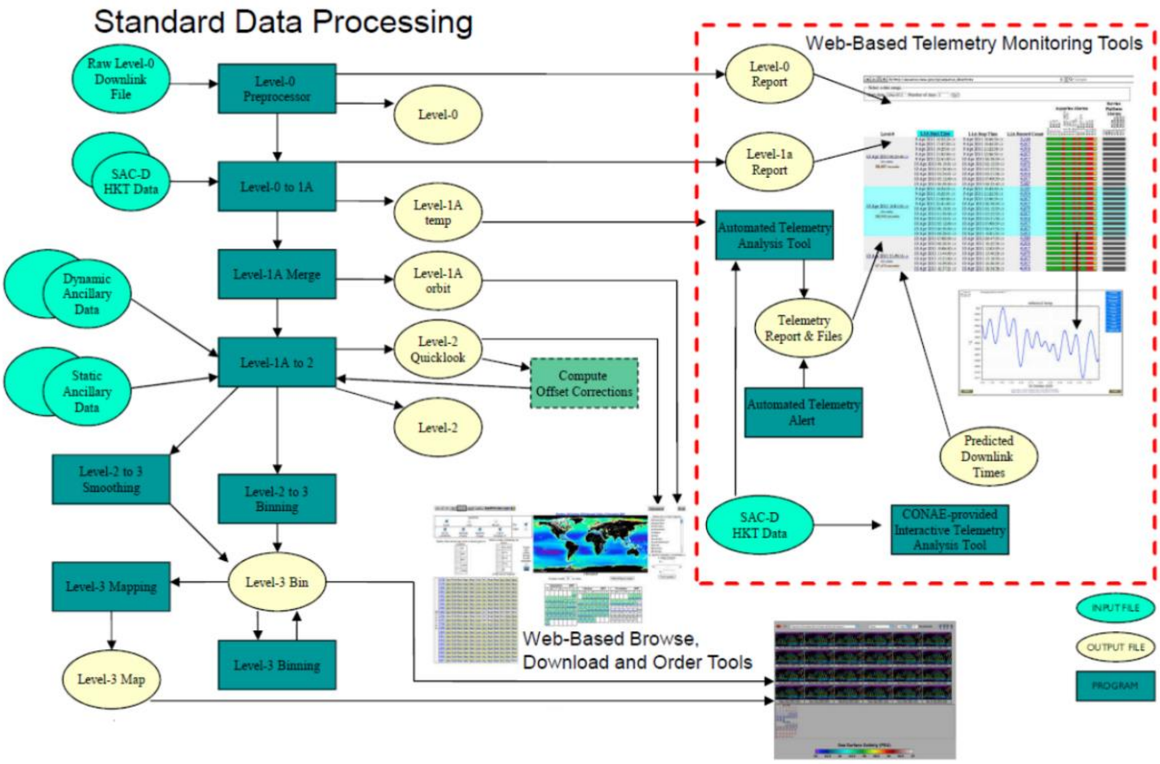


Figure 18: AQUARIUS algorithm flow diagram

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 1 week.

With

- ✓ a spatial sampling : 25 km EASE V2 grid;
- ✓ time sampling : 15 days (monthly products) and 1 day (weekly products).

The processing chain breakdown is detailed on Figure 19.

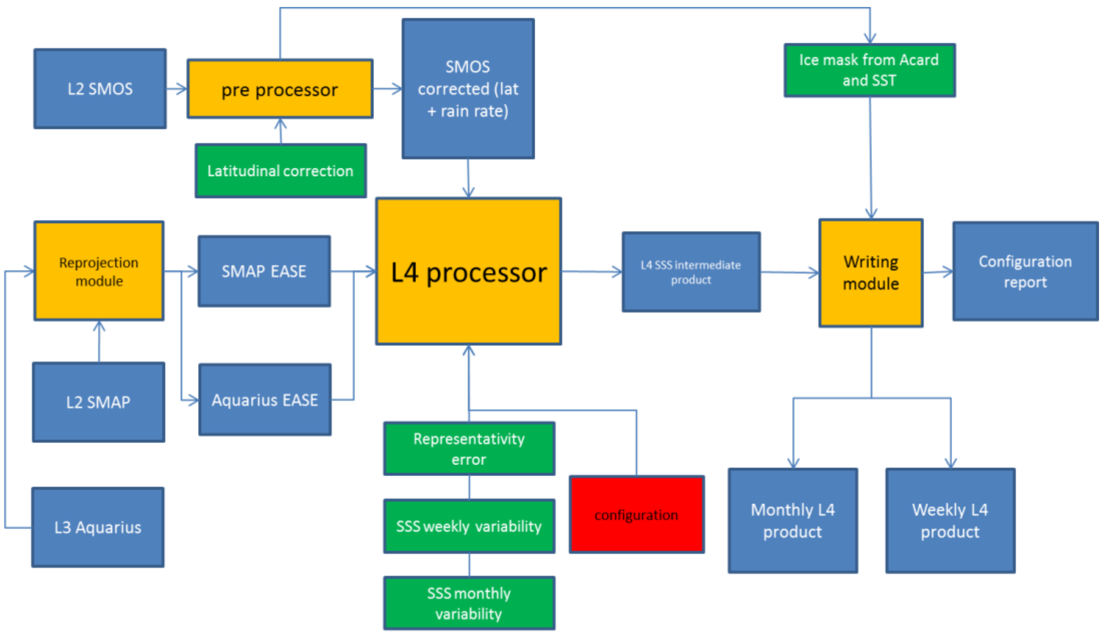


Figure 19: CCI+ Salinity processing Level 4 chain

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 EASE-2 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, 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 taking the averaged SSS from the SMOS central across swath location 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 Mercator model. This eliminates outliers that deviate too far from expected values.

5 CCI+SSS GitHub repository

To manage the source code, a GitHub repository was setup for the SSS ECV project at: https://github.com/CCI-SALINITY

The full Matlab code used to produce the SSS CRDP of Year 1 to 3 (Level 4 products) has been uploaded in the repository. Level 2P and Level 3C processing chains have also been uploaded on the server.

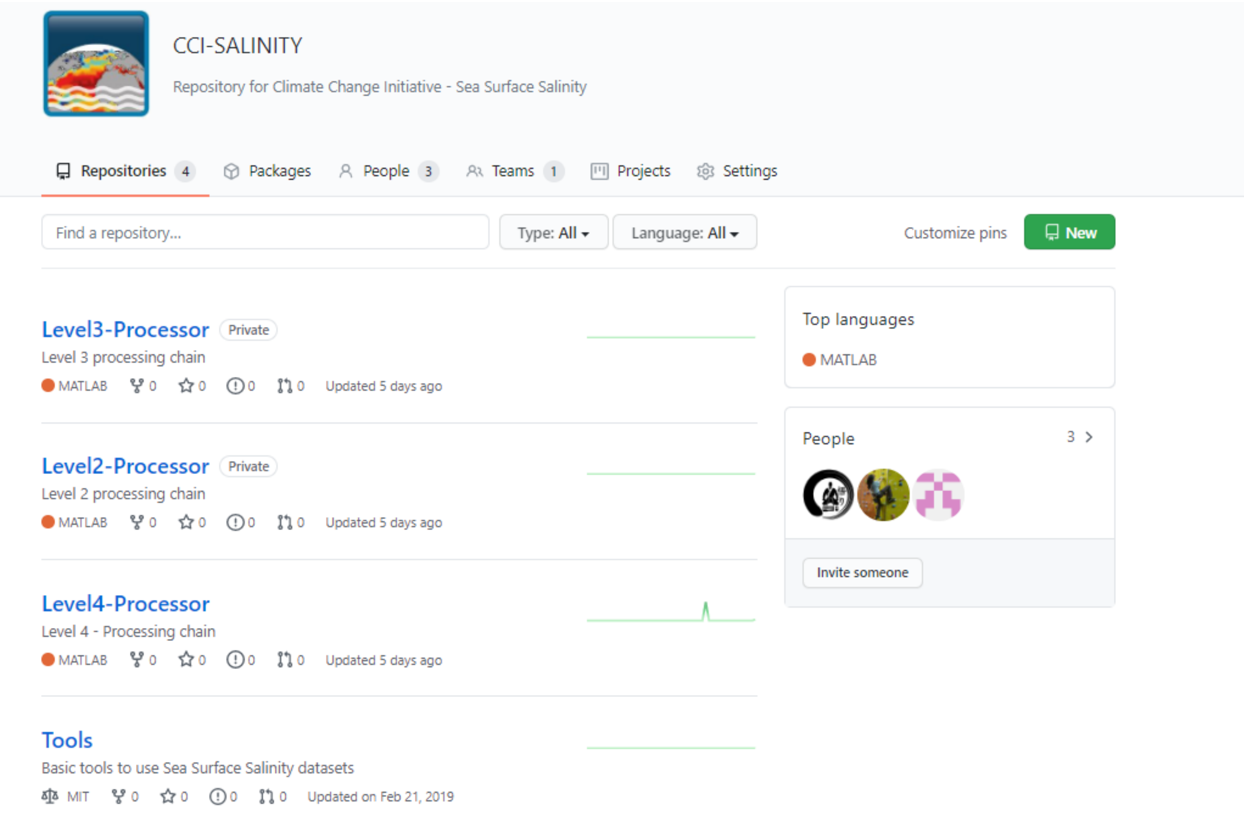


Figure 20 : Github repository screen shot

Description on each item is also included in the web pages (Readme files).

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

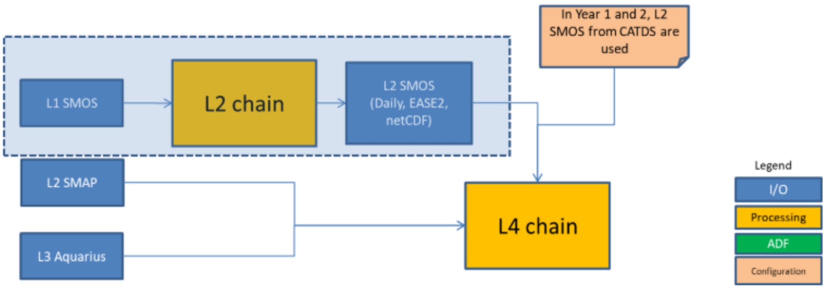
Year1	Add files via upload	4 months ago
Year2	Update image	5 days ago
README.md	Update README.md	5 days ago

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

6 Round Robin description

Sub-programs list and description: in this table, orange fields correspond to programs that include parallelization.

Matlab program	Description
MAIN_validations	Main program : reads inputs, launches RR for each set of products, creates a global report of all comparisons
Glob	Program to run RR tests, for one set of products
BATEAUX	Treatment of merchant ship tracks data
TARA_SMOS	Treatment of science ship tracks
CMP_SSSOS	Treatment of binned ship tracks in northern atlantic
Cmp_Pirata	Treatment of moorings data
Plot_Pirata	Generation of figures for moorings
Valid_SMOS_TARA_Arctic	Collocate science ships to satellites
plot_Tara	Generation of figures for science ships
read_data	Read satellite data for the RR test
readdata_pirata	Read satellite data for comparisons with moorings
read_satdata_dico	Read DICO_SAT.txt
stats	Compute statistics between two datasets.
Colloc	Main collocation program
day_corr	Detection of satellite files to read
read_input	Read input.txt
Lissage	Smoothing of ship tracks
r2test.m, varFtest.m	Tests of significance
texfig, printtabstats, orderstats	Programs to create reports

Table 11: Matlab programs description

6.1 Input files description

There are two input files to fill before launching Round Robin tests, described below. Input files must be in the same directory as the executable.

6.1.1 Input.txt

Input.txt is a list of parameters to run the Round Robin tests. Each field must be separated from its value with a tabulation.
In this file, "#" at the beginning of a line is a comment.
Lines beginning with "%" define the structure of the file and must remain unchanged.
The following table summarize the fields to fill in this file. Annex 1 is an example of input.txt file.

section	field	description	available
General	prods	Products to compare in RR, by pair (reference,product). Possibility to compare to several references but only one product in one set. Possibility to define several sets of products, separated with ";", RR compare each set separately.	Fields 'name' of DICO_SAT.txt file (see section 6.1.2)
	period	Period of study, limited by the product availability. There must be one period for one set of products.	SMOS SMAP Aquarius Other period : beginning,ending dates (format: ddmmyyyy)
	titreT	Title of generated reports	-
	log	Create log file or not	1/0
	Datasave	Directory to save stats	-
	Figsave	Directory to save figures	-
	Outputdir	Directory to save output reports	-
	Auxfigdir	Directory of auxiliary figures	-
	DCmask	Distance to coast file	-
	rc	Force the recalculation of the report even if stats are saved for titreT	1/0
	rt	Force the actualization of figures	1/0
	rs	Force the recalculation of stats	1/0
	maptype	Choose to plot products stats or differences between products stats and reference stats	prod/diff

	revserseRefProd	To inverse order of reference/product. This allows to compare several products to one reference.	1/0
TRAITEMENTS MOORING	treat	To treat moorings	1/0
	datavec	List of mooring data to include in tests	PIRATA,RAMA,TAO
	DATAdir	Directory of mooring data	-
	Periodicity	Choose periodicity of satellites to collocate to in-situ: “dtsat” read satellite files accordingly to satellite time resolution, to avoid redundancies; “one” use every satellite file available in the period of study	one,dtsat
	tstflg	To test flag or not	1/0
TRAITEMENTS SHIP_TRACKd	treat	To treat ship tracks data in tests	1/0
	datavec	List of ship tracks areas to include in tests	See section
	DATAdir	Directory of ship tracks data collocated	-
	DATAdir0	Directory of ship tracks , original files	-
	dlat	Difference in latitude between boxes to calculate statistics (see section 85734.32896.0.0)	In degrees
	dkm	Difference in km between boxes to calculate statistics (see section 85865.32896.0.0)	In km
	dcm	Minimal distance to coast of in-situ data in km (see section 85984.32896.0.0)	In km
	flgis	Flags on in-situ data (ship tracks) dc=distance to coast (km) Lat=latitude (deg) Lon=Longitude (deg) sss=salinity (pss) (see section 86177.32896.0.0)	Example: dc>40&Lat<45&Lon<120&sss>32
	tstflg	To test flags	1/0
	rmk	To force the production of collocalisations	1/0
TRAITEMENTS SCIENCE_SHIP	treat	To treat science ships data in tests	1/0
	datavec	List of ships to include. To select a specific year in the data, add the year after ship name (ex: TARA;2013)	TARA,Beautemps_Beaupre,Fleur_Australe,Pourquoi_Pas,Atalante,Thalassa,Le_Suroit,La_Louise,Mariposa,RARA_Avis,STAMM

			,BarkEuropa,Boogalo o
	DATAdir	Directory of science ships data	-
	dtm	Maximum gap in time to cut the track in figures	In day
	tstflg	To test flags	1/0
TRAITEMENTS HIGH_LAT	treat	To treat high latitude data in report	1/0
	datavec	List of tracks to include in tests	BAX01,BAX02,GAX01,NAX01
	DATAdir	Directory of in-situ data	-
	tstflg	To test satellite flags	1/0
METRICS	METRICS_FIGURES	Metrics to include as figures in the report	Number corresponding to metrics : 1: nb_points 2: std_diff 3: rmsd 4: bias 5: r2 6: std 7: err 12: robust std_diff
	METRICS_STATS	Metrics to include as statistics (tables) in reports	Number corresponding to metrics : 1: nb_points 2: std_diff 3: rmsd 4: bias 5: r2 6: std 7: err 12: robust std_diff
Interrupteursfiguresrapport Ship_tracks	MAPS	Include maps in ship tracks sections	1/0
	SCATTER	Include scatterplots in ship tracks sections	1/0
	TIMESERIES	Include time series in ship tracks sections	1/0
	EXAMPLES	Include examples in ship tracks sections	1/0
Interrupteursfiguresrapport	MAPS	Include maps in moorings sections	1/0
	SCATTER	Include scatterplots in moorings sections	1/0

Moorings	TIMESERIES	Include time series in moorings sections	1/0
Interrupteursfiguresrapport Science_ships	MAPS	Include maps in science ships section	1/0
	SCATTER	Include scatterplots in science ships section	1/0
	TIMESERIES	Include time series in science ships section	1/0
Interrupteursfiguresrapport High_latitude	MAPS	Include maps in high latitude section	1/0
	SCATTER	Include scatterplots in high latitude section	1/0
	TIMESERIES	Include time series in high latitude section	1/0
	STATS	Include stats figures in high latitude section	1/0
	HM	Include hovmöller in high latitude section	1/0
EXAMPLES	Examples_Mouillage	List of examples for moorings	
	Examples_Ship_Track	List of examples for ship tracks	

Table 12: input.txt fields

6.1.2 DICO_SAT.txt

DICO_SAT.txt is a dictionary that gives all information about satellite products useful for RR tests. It includes the following fields separated by a tabulation.
The first line of this dictionary is a DEFAULT line. When satellite fields remain empty, they are set to the default value described in this first line.
Each field "name" must be unique. This field is used in input.txt file, when selecting the products to compare in RR.
The following table describes fields included in DICO_SAT.txt file. An example of DICO_SAT.txt is in section 6.3

Name	Name of product
path	Path to the directory where corresponding satellite daata are stocked
var	Variable to read in satellite file
varerr	Error variable to read in satellite file (optional)
flg	Flag on satellite data (optional)
dt	Temporal resolution of satellite product
freq	Frequency of satellite products
res	Spatial resolution
moyt	Method of temporal averaging (g: gaussian, e: exponential, s: simple)
cax	Axis limits for figures (optional)
mis	Name of mission/project
str1	First part of name, before dates, in satellite files
str2	Second part of name, before dates, in satellite files (optional)
str3	Third part of name, before dates, in satellite files (optional)
permut	To apply a permutation in lat,lon (optional)
coef	Multiply this product by a constant in tests (optional)
bias	Add a constant bias to this product in tests (optional)
Printstat	1/0 : define if stats should appear in figures for this product (optional)
nc.d	Beginning of nc file to read (optional)
nc.f	End of nc file to read (optional)

Table 13: DICO_SAT fields

6.1.3 Output files

Round Robin produces two kinds of output files, reports and figures that can be used in presentations and save files (format .mat or netCDF). Below is the list of these output files. Files positions in program correspond to the description in Figure 7.

6.1.4 Figures and reports

Position in program	Name of file	directory	description
Report.pdf	TitreT.pdf (see input.txt)	Outputdir (see input.txt)	Round Robin report
Figures (1)	**Maps**: Map[d/diff]_[metric]_[product]_[area].[fig/png] **Scatterplot**: Scatterplot_[metric]_[products]_[area].[fig/png] **Stats**: Stats_[products]_[area].[fig/png]	FIGURES/ MShips/	Figures corresponding to ship tracks treatment
Figures (2)	**Map of salinity of satellite products along ship track**: sss[product]_[ship_name]_[area]_[date].[fig/png] **Map of differences between salinity of satellite products and ship salinity**: DIFF_[ship_name]_sss[product]_[area]_[date].[fig/png] **Time series of salinities along ship track:** CMP_[products]_COL_[ship_name]_[area]_[date].[fig/png]	FIGURES/ ScienceShips/	Figures corresponding to Scientific Ship tracks
Figures (3)	**Maps**: Map[d/diff]_[metric]_[product]_OSSSS_	FIGURES/	Figures corresponding to High latitudes binned ship tracks

	Comparisons_with_SSS-OS.[fig/png] **Scatterplots of products statistics wrt ship measurements:** Scatterplot_[metric]_[products]_[transect].[fig/png] **Time series of satellites minus in-situ:** TS_[product]_[transect]_SSSOS_JLV.[fig/png] **Time series of statistics:** Stats_[products]_[transect].[fig/png] **Hovmöllers:** HM_[product]_[transect]_SSSOS_JLV.[fig/png]	SSS_OS/	
Figures (4)	**Maps**: Map[diff]_[metric]_[product]_[period].[fig/png] **Scatterplots**: Scatterplot_[metric]_PIRATA_[products]_[period].[fig/png] **Time series**: TS_PIRATA1_[mooring]_[products]_[period].[fig/png]	FIGURES/ PIRATA_SSS/	Figures corresponding to moorings treatment

Table 14: Figures and reports produced by Round Robin

6.1.5 Save and intermediate files

Save file (1)	Colocs_[spatial_resolution]_[name_of_original_nc_file]_[product]_SSS-SST.nc	INSITU/COLOCS/	Colocations between ship tracks and satellites, in netCDF files. See full description of these files in section 6.4
Save file (1)	stats_bateauxd_[products]_[period].mat	VALIDATIONS/	Statistics for each box over a merchant ship transect + coordinates of boxes, global statistics over ship transect
Save file (2)	STATS_[ship_name][products]_[period]period.mat	VALIDATIONS/	Statistics of products wrt scientific ship track
Save file (2)	[ship_name]_globalColoc_[product]_[period].mat	VALIDATIONS/	Colocation of product with scientific ship track
Save file (3)	_STATS_HMSSSOS__[product]___[transect]___[period]_period.mat	VALIDATIONS/	Colocation of product wrt binned salinity in northern latitudes (in time and space)
Save file (3)	_STATS_SSSOS___[products]___[period]_period.mat	VALIDATIONS/	Global statistics of different products wrt binned salinity in northern latitudes
Save file (3)	SSSOS_Coloc_[product]_[period]period.mat	VALIDATIONS/	Extraction of satellite product over the area of binned tracks in northern latitudes
Save file (4)	MOORING_Colocs_[time_resolution]_[mooring_name]_[product]_[dates].mat	VALIDATIONS/	Colocations of products wrt moorings.

Table 15: Save and intermediate files

6.2 Example of input.txt file

```
prods
      SMOS_L3QA_317,CCI_v2019_1.8_7,CCI_v2020_2.3_7,CCI_v2019_1.8_7,CCI_v2020_2.4_7,CCI_v2019_1.8_7;SMOS_L3QA_317,CCI_v2019_1.8_7,CCI_
v2020_2.3_7,CCI_v2019_1.8_7,CCI_v2020_2.4_7,CCI_v2019_1.8_7;SMOS_L3QA_317,CCI_v2019_1.8_7,CCI_v2020_2.3_7,CCI_v2019_1.8_7,CCI_v2020_2.
4_7,CCI_v2019_1.8_7;SMOS_L3QA_317,CCI_v2019_1.8_7,CCI_v2020_2.3_7,CCI_v2019_1.8_7,CCI_v2020_2.4_7,CCI_v2019_1.8_7,SMAP_RSS4_40,CCI_v20
19_1.8_7;SMOS_L3QA_317,CCI_v2019_1.8_7,AQUARIUS_CAPv5,CCI_v2019_1.8_7,CCI_v2020_2.3_7,CCI_v2019_1.8_7,CCI_v2020_2.4_7,CCI_v2019_1.8_7

#Produits a comparer (reference, produit)

period 01012010,01012020;SMOS;SMAP;Aquarius
#periode d'etude (meme nb d elements que prods)

titreT Report_PVASR_wkly_noflag
#titre du document latex genere

log    0
#fichier log (0 ou nom du fichier)

Datasave     VALIDATIONS/
#dossier de sauvegarde stats

Figsave      FIGURES/
#dossier de sauvegarde figures

Outputdir    Reports/
#dossier de stockage des rapports

Auxfigdir    fig/
#dossier de figures auxilliaires pour les rapports (logos etc)

DCmask INSITU/maskLOCEAN.nc
#lien vers fichier de distances à la cote

rc     1
#rc=1 => force le traitement meme si il y a un fichier enregistré pour titreT
rt     1
#rt=1 => force le tracé des figures, meme si elles existent déjà
rs     1
```

```
#rs=0 =>force le calcul des stats même s'il y a un fichier de stats enregistré

maptype      diff
#diff=tracer les différences par rapport aux références, normal=tracer les produits (old)

revserseRefProd     1
#1 si on inverse les références et les produits dans "prods", 0 si ordre normal

################################################################################
%%TRAITEMENTS

%MOORING
#mooring: traitement de mouillages

treat  1
datavec      PIRATA
#,RAMA,TAO
DATAdir      INSITU/
Periodicity  dtsat
#one,dtsat
tstflg 1



%SHIP_TRACKd
#ship_track_d: traitement des ship tracks

treat  1

datavec      Ship_Atl_s
#all,All_ship,Ship_Atlantic,Ship_Atl_s,Ship_Atl_n,Ship_Pacific,Ship_Pac_o,Ship_Pac_e,Ship_Indian,
#AX01,AX03,AX20,AX11,PX04,PX05,PX17,PX53,IX12,IX30
#MarineMammals,MarineMammals_north,MarineMammals_south,
#MM_USA,MM_Canada,MM_Europe,MM_NZeland,MM_South_Am,MM_Indian

#dlat: ecart en latitude/longitude pour le calcul des stats
dlat   3
```

```
#dkm: ecart en km pour le calcul des stats
dkm    300

#dcm: distance max en km pour la classe 'proche de la cote'
dcm    400

#flgis: flag a appliquer; dc=distance cote min (km), Lat=latitude, Lon=longitude, sss=sss
flgis  dc>40&Lat<45
#&Lat<45&sss>32

DATAdir       INSITU/COLOCS/
DATAdir0      INSITU/ORESSS/orig/
tstflg 1
rmk    0



%SCIENCE_SHIP
treat  1
datavec       TARA;2013
#all_z,all,TARA;2013,Beautemps_Beaupre,Fleur_Australe,Pourquoi_Pas,Atalante,Thalassa,Le_Suroit,La_Louise,
#Mariposa,RARA_Avis,STAMM,BarkEuropa,Boogaloo
DATAdir       INSITU/GOSUD_DM/smooth/
dtm    0.5
tstflg 1




%HIGH_LAT
treat  1
datavec       BAX01,BAX02
#BAX01,BAX02,GAX01,NAX01
DATAdir       INSITU/sss_bin_nasg_netCDF_v3.0/
tstflg 1

################################################################################
%%METRICS
```

```
#nb_points,std_diff,rmsd,bias,r2,std,err
METRICS_FIGURES     1,2,3,4,5,12
#metriques que l'on veut intégrer dans le rapport (figures)
METRICS_STATS1,2,5,4,12
#metriques que l'on veut intégrer dans le rapport (stats)


################################################################################
%%Interrupteursfiguresrapport
#pour choisir les figures que l'on veut intégrer dans le rapport

%Ship_tracks
MAPS   1
SCATTER       1
TIMESERIES    1
EXAMPLES      1

%Moorings
MAPS   1
SCATTER       1
EXAMPLES      1

%Science_ships
MAPS   1
SCATTER       1
TIMESERIES    1

%High_latitude
MAPS   1
SCATTER       0
TIMESERIES    0
STATS  1
HM     1



################################################################################
```

```
%%EXAMPLES
#Pour tracer des exemples de mouillage ou de transect

Examples_Mouillage  15,16;5,1;5,1
#correspond a l'ordre de TRAITEMENTS.MOORINGS.datavec

Examples_Ship_Track 139,2;367,2;727,1;729,1;235,1;237,1;237,3;1225,3
```

6.3 Example of DICO_SAT.txt

Name	path	var	varerr	flg	dt	freq	res	moyt	cax	mis
DEFAULT				0					32,38	
SMOS_L3QA_317	path1	Mean_Sea_Surface_Salinity	SSS_error_mean		10	1	50	s	20,37	SMOS L3Q
CCI_v2019_1.8_7	path2	sss	sss_random_error	sss_qc==0	7	1	50	g	20,27	CCI
AQUARIUS_CAPv5	path3	sss_cap			7	1	150	g	20,37	Aquarius
SMAP_RSS4_40	path4	sss_smap_40km			8	1	40	g	20,37	SMAP
SMAP_RSS4_40_M	path5	sss_smap_40km			30	30	40	g		SMAP
CCI_v2019_1.8_30	path6	sss	sss_random_error	sss_qc==0	30	15	50	g		CCI
CCI_v2020_2.3_7	path7	sss	sss_random_error	sss_qc	7	1	50	g	20,27	CCI
SMOS_CECV4_18_q80	path8	SSS	eSSS		18	4	50	g		SMOS CEC
AQUARIUS_CAPv5_M	path9	sss_cap			30	30	150	g		Aquarius
SMOS_L3QA_317_M	path10	Mean_Sea_Surface_Salinity	SSS_error_mean		30	30	50	s		SMOS L3Q
CCI_v2020_2.3_30	path11	sss	sss_random_error	sss_qc==0 & lsc_qc==0 & isc_qc==0	30	15	50	g		CCI
CCI_v2020_2.3_7	path12	sss	sss_random_error	sss_qc==0 & lsc_qc==0 & isc_qc==0	7	1	50	g		CCI

str1	str2	str3	permut	coef	bias	printst	nc.d	nc.f
			1,2,3	1	0	1	1,1,1	Inf,Inf,Inf
SM_RE06_MIR_CSQ3A__	T000000_	T235959_317_001_7.DBL	2,3,1				1,1	Inf,Inf
ESACCI-SEASURFACESALINITY-L4-SSS-MERGED-OI-7DAY-RUNNINGMEAN-DAILY-25km-	-fv1.7.nc							
sss	.v5.0cap.nc						1,1	Inf,Inf
RSS_smap_SSS_L3_8day_running_	_FNL_v04.0.nc						1,1	Inf,Inf
RSS_smap_SSS_L3_monthly_	_FNL_v04.0.nc						1,1	Inf,Inf
ESACCI-SEASURFACESALINITY-L4-SSS-MERGED-OI-Monthly-CENTRED-15Day-25km-	-fv1.7.nc							
ESACCI-SEASURFACESALINITY-L4-SSS-MERGED_OI_7DAY_RUNNINGMEAN_DAILY_25km-	-fv02.3.nc							
SMOS_L3_DEBIAS_LOCEAN_AD_	_EASE_18d_25km_v04.nc							
sss	.v5.0cap.nc						1,1	Inf,Inf
SM_RE06_MIR_CSQ3A__	T000000_	T235959_317_001_7.DBL	2,3,1				1,1	Inf,Inf
ESACCI-SEASURFACESALINITY-L4-SSS-MERGED_OI_Monthly_CENTRED_15Day_25km-	-fv02.3.nc							
ESACCI-SEASURFACESALINITY-L4-SSS-MERGED_OI_7DAY-RUNNINGMEAN-DAILY-25km	-fv02.3.nc							

N.B.: Fields “path_X” must be replaced by paths to corresponding products.

6.4 Description of colocation files with respect to merchant ship tracks

Variable name	dimension	description
Lat_ship_orig	Time_ship_orig	Latitude of original ship track [degrees_north]
Lon_ship_orig	Time_ship_orig	Longitude of original ship track [degrees_east]
Time_ship_orig	Time_ship_orig	Time of original ship track [Julian day]
SSS_ship_orig	Time_ship_orig	SSS of original ship track [PSU]
SSS_qc_ship_orig	Time_ship_orig	SSS quality check of original ship track
SSS_error_ship_orig	Time_ship_orig	SSS error of original ship track
SSS_ship_orig_smooth	Time_ship_orig	SSS of original ship track after gaussian smoothing [PSU]
SST_ship_orig	Time_ship_orig	SST of original ship track [K]
SST_qc_ship_orig	Time_ship_orig	SST quality check of original ship track

SST_error_ship_orig	Time_ship_orig	SST error of original ship track
SST_ship_orig_smooth	Time_ship_orig	SST of original ship track after gaussian smoothing [K]
Lat_ship_pixel	Day_ship_pixel	Latitude of ship track averaged over satellite pixels [degrees_north]
sLon_ship_pixel	Day_ship_pixel	Longitude of ship track averaged over satellite pixels [degrees_east]
Day_ship_pixel	Day_ship_pixel	Time of ship track averaged over satellite pixels [Julian day]
Sec_ship_pixel	Day_ship_pixel	Seconds from 0h of ship track averaged over satellite pixels [seconds]
SSS_ship_pixel	Day_ship_pixel	SSS of ship track averaged over satellite pixels [PSU]
SSS_std_ship_pixel	Day_ship_pixel	SSS std of ship track averaged over satellite pixels [PSU]
SSS_error_ship_pixel	Day_ship_pixel	SSS error of ship track averaged over satellite pixels [PSU]

SST_ship_pixel	Day_ship_pixel	SST of ship track averaged over satellite pixels [K]
SST_error_ship_pixel	Day_ship_pixel	SST error of ship track averaged over satellite pixels [K]"
SSS_sat_pixel	Day_ship_pixel	SSS of colocated satellite pixels [PSU]
Lat_sat_pixel	Day_ship_pixel	Lat of colocated satellite pixels [degrees_north]
Lon_sat_pixel	Day_ship_pixel	Lon of colocated satellite pixels [degrees_east
Time_sat_pixel	Day_ship_pixel	Time of colocated satellite pixels [Julian day]
flg_sat_pixel	Day_ship_pixel	flag of colocated satellite pixels (1=good;0=bad)
Error_sat_pixel	Day_ship_pixel	Error of colocated satellite pixels [PSU]
SSS_sat_pixel_mean_lon	Day_ship_pixel	SSS of colocated satellite pixels averaged over 5pixels in longitude [PSU]

SSS_sat_pixel_lon	nlon, Day_ship_pixel	SSS of 7 nearest pixels in longitude to colocated satellite pixels [PSU]
Error_sat_pixel_lon	nlon, Day_ship_pixel	Error of 7 nearest pixels in longitude to colocated satellite pixels [PSU]
sigma	1	std of gaussian smoothing [km]

End of Document