Climate Change Initiative+ (CCI+) Phase 1

Sea Surface Salinity
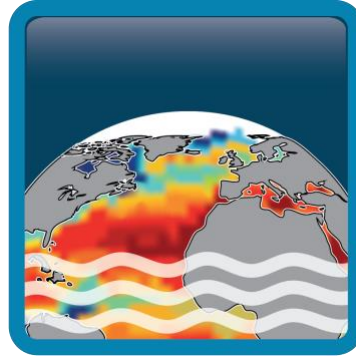
End-to-End ECV Uncertainty Budget (E3UB)

Signatures

Author	Jean Luc Vergely	*Vergely.*	19-12-2019
Reviewed by	Jacqueline Boutin (Science Leader)		
Approved By	Jacqueline Boutin (Science Leader)	*JBoutin*	19-12-2019
	Nicolas Reul (Science Leader)	<signature>	<date>
	Rafael Catany (Project Manager)	*Rafael Catany*	19-12-2019
Accepted by	Craig Donlon (Technical Officer)		

Diffusion List
Sea Surface Salinity Team Members
ESA (Craig Donlon, Paolo Cipollini)

Amendment Record Sheet

DATE / ISSUE	DESCRIPTION	SECTION / PAGE
15-07-2019 / v1.0	Delivery to ESA	New document
21-10-2019 / v1.1	Update reference document documents section	Section 1.3.2 / page 14
	Implemented description of the sensors mounted in each satellite	Section 2 / page 21 - 22
	Added overview about PiMEP	Section 4.4.2 / page 31
	Added definition PCTVAR	Section 5.3.1 / page 63
	Improved Figure 33. Now all markers are visible as described in figure caption.	Section 5.3.2 / page 65
19/12/2019 / v1.2	Edited description about PIMEP	Section 4.4.2 / page 31
19/12/2019 / v1.2	Added band width of Aquarius satellite is 25 MHz	Section 2.4 / page 22
19/12/2019 / v1.2	Added clarification text about Figure 32, stating that further analyses is needed and it will be provided in next version of the document.	Section 5.2.2/ page 60
19/12/2019 / v1.2	Edited caption Figure 34. Now the caption read "2016-2018" as seen in the main text too.	Section 5.3.2 / page 63

Table of Contents

List of figures

green, Bayesian method; in black, average over a 10-day rectangular time window. The blue dots correspond to the SSS corrected for the relative inter-sensor bias, the red dots to the uncorrected SSS. At the bottom, the associated errors.

List of tables

1   Introduction

1.1   Scope of this document

This document holds the End-to-End ECV Uncertainty Budget (E3UB) prepared by CCI+ Salinity team, as part of the activities included in the [WP240] of the Proposal (Task 2 from SoW ref. ESA-CCI-PRGM-EOPS-SW-17-0032).

The climate users of CCI+ Sea Surface Salinity (SSS) products need to know precisely their uncertainty from statistical indicators as realistic as possible. The simplest way to qualify SSS is to associate it with two statistical indicators that are the random error and systematic error. The purpose of this document is to provide an estimation of the systematic uncertainties in order to correct measured SSS before L3/L4 SSS merging and the way to estimate the L4 SSS random error.

Uncertainties arise due to many aspects that can be generally grouped into the following primary categories:

•   instrument measurement uncertainty: those relating to instrument hardware,

•   retrieval/algorithm uncertainty: those related to model uncertainties and auxiliary data a priori error (SST, WS, TEC, galactic correction, sun correction, …),

•   unknown: those uncertainties that are "unknown".

For each category standard practice [RD-47] requires an uncertainty budget to be derived including all aspects leading to a quantification of a root-sum-square (RSS) estimate of uncertainty.

The traditional way to estimate errors is to consider a statistical approach for the data processing. Least square methods allow, under the assumption of Gaussian error of the input data, to calculate errors on the model parameters. In theory, at each processing step, errors associated with the corresponding parameters are available at the input of the next calculation step.

However, in practice, there are processing steps in SMOS, SMAP or Aquarius data that do not propagate errors. L0-L1 processing does not propagate all the errors related to calibration problems or uncertainty on certain parameters (for example, for SMOS there is an uncertainty on antenna gains that is not propagated during reconstruction). In addition, the reprocessing of the L0→L1 data is not within the scope of this project. However, in some cases, it is possible to estimate errors empirically at the end of a processing.

1.2   Structure of the document

The E3UB is structured as follows:

This document is composed of 4 major sections:

- ✓ Section 2: Sensor main characteristics
- ✓ Section 3: L1 uncertainty characterization
- ✓ Section 4: L2OS uncertainty characterization
- ✓ Section 5: L3 and L4 error budget

L2/L3/L4 data sets will be provided each year and are described in the SRD.

This is the first version of the E3UB document addressing Year 1 activity. More accomplished versions will be proposed in Year 2 and 3 accounting for acquired experience and feedback received by the users of the products.

1.3   References

1.3.1   Applicable Documents

ID	Document	Reference
AD01	CCI+ Statement of Work	SoW
AD02	Product User Guide (PUG)	PUG
AD03	User Requirement Document (URD)	SSS_cci-D1.1-URD-i1r0
AD04	Product Specification Document (PSD)	SSS_cci-D1.2-PSD-v1r4
AD05	Algorithm Theoretical Baseline Document	SSS_cci-D2.3-ATBD_L3_L4-i1r0_v1.1

1.3.2   Reference Documents

ID	Document	Reference
RD01	Boutin, J., N. Martin, N. Kolodziejczyk, and G. Reverdin (2016a), Interannual anomalies of SMOS sea surface salinity, Remote Sensing of Environment, doi:http://dx.doi.org/10.1016/j.rse.2016.02.053	
RD02	Kolodziejczyk, N., J. Boutin, J.-L. Vergely, S. Marchand, N. Martin, and G. Reverdin (2016), Mitigation of systematic errors in SMOS sea surface salinity, Remote Sensing of Environment, doi:http://dx.doi.org/10.1016/j.rse.2016.02.061.	
RD03	Evaluation of measurement data – Guide to the expression of uncertainty in measurement, JCGM 100:2008	
RD04	SMOS ATBD L2OS v3.13, 29 April 2016	SO-TN-ARG-GS-0007

ID	Document	Reference
RD05	AQ-014-PS-0017_Aquarius_L2toL3ATBD_DatasetVersion5.0 Liang Hong, Normal Kuring, Joel Gales and Fred Patt	
RD06	AQ-014-PS-0018_AquariusLevel2specification_DatasetVersion5.0 Fred Patt, Liang Hong	
RD07	SMAP_RemSSS_Release_V2.0	
RD08	Meissner, T. and F. J. Wentz, 2016: Remote Sensing Systems SMAP Ocean Surface Salinities [Level 2C, Level 3 Running 8-day, Level 3 Monthly], Version 2.0 validated release. Remote Sensing Systems, Santa Rosa, CA, USA. Available online at www.remss.com/missions/smap, doi: 10.5067/SMP20-2SOCS (L2C files).	
RD09	Boutin J., J.-L. Vergely, S. Marchand, F. D'Amico, A. Hasson, N. Kolodziejczyk, N. Reul, G. Reverdin, J. Vialard (2018), New SMOS Sea Surface Salinity with reduced systematic errors and improved variability, *Remote Sensing Of Environment,* doi:http://dx.doi.org/10.1016/j.rse.2018.05.022	
RD10	Thomas Meissner + Frank Wentz Remote Sensing Systems, Santa Rosa, CA, RSS SMAP Salinity: Version 2 Validated Release. Algorithm Theoretical Basis Document (ATBD), September 13, 2016	RSS Technical Report 091316

1.4 Acronyms

AD	Applicable Document
ADP	Algorithm Development Plan
AOPC	Atmospheric Observation Panel for Climate
AR	Assessment Report (of the IPCC)
AR6	IPCC Scientific Assessment Report 6
ATBD	Algorithm Theoretical Basis Document
Aquarius	NASA mission
C3S	Copernicus Climate Change Service
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
CliC	World Climate Research Programme - Climate and Cryosphere Project
CLIVAR	WCRP Climate Variability and Predictability project
CMEMS	Copernicus Marine Environmental Monitoring Service
CMIP	Coupled Model Intercomparison Project
CMUG	Climate Modelling User Group
COP	Conference of the Parties
COWCLIP	Coordinated Ocean Wave Climate Project (of JCOMM)
CR	Cardinal Requirement
CRDP	Climate Research Data Package
CRG	Climate Research Group
CSCDA	Copernicus Space Component Data Access System
CSWG	Climate Science Working Group
DARD	Data Access Requirements Document
DEWG	Data Engineering Working Group
DOI	Digital Object Identifier
DPM	Detailed Processing Model
DTBT3	Database for Task 3
DUE	Data User Element
E3UB	End-to-End ECV Uncertainty Budget
EC	European Commission
ECMWF	European Centre for Medium Range Weather Forecasts
ECSAT	European Centre for Space Applications and Telecommunications

ECSS	European Cooperation for Space Standardization
ECV	Essential Climate Variable
EO	Earth Observation
EOV	Essential Ocean Variable (of the OOPC)
ESGF	Earth System Grid Federation
ESM	Earth System Model
EU	European Union
FCDR	Fundamental Climate Data Record
FIDUCEO	Fidelity and uncertainty in climate data records from Earth Observations
FOV	Field Of View
FP7	EU Framework Programme 7
FRM	Fiducial Reference Measurements
GAIA-CLIM	Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring
GEO	Group on Earth Observations
GCOS	Global Climate Observing System
GCW	Global Cryosphere Watch
GMECV	Global Monitoring of Essential Climate Variables - element of the European Earth Watch programme.
GNSS	Global Navigation Satellite System
GOOS	Global Ocean Observing System
H2020	Horizon 2020 programme
Hs	Significant Wave Height (see also SWH)
H-SAF	EUMETSAT's Hydrology Satellite Applications Facility
HDD	Hard disk
IOC	Intergovernmental Oceanographic commission (of UNESCO)
IODD	Input Output Data Definition

IP	Implementation Plan
IPCC	Intergovernmental Panel on Climate Change
ISAS	In Situ Analysis System (LOPS)
ISDB	in situ database (of Fiducial Reference Measurements and satellite measurements)
JAXA	Japan Aerospace Exploration Agency
JCOMM	Joint Commission on Oceanography and Marine Meteorology
KO	Kick-off
MOOC	Massive Open Online Course
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NOP	Numerical Ocean Prediction
NWP	Numerical Weather Prediction
Obs4MIPs	Observations for Model Intercomparison Projects
ODP	Open Data Portal
OOPC	Ocean Observation Panel for Climate
OTT	Ocean Target Transform
Pi-MEP	Pilot Mission Exploitation Platform
PMP	Project Management Plan
PSD	Product Specification Document
PUG	Product User Guide
PVASR	Product Validation and Algorithm Selection Report
PVIR	Product Validation and Intercomparison Report
PVP	Product Validation Plan
QA4EO	Quality Assurance Framework for Earth Observation

QSR	Quarterly Status Report
R&D	Research and Development
RTM	Radiative Transfer Model
RCP	Representative Concentration Pathways
RD	Reference Document
SAF	Satellite Applications Facility
SAR	Synthetic aperture Radar
SISS	Satellite and In situ [Working Group]
SLP	Sea Level Pressure
SMAP	Soil Moisture Active Passive [mission of NASA)
SMOS	Soil Moisture and Ocean Salinity [satellite of ESA]
SoW	Statement of Work
SRAL	SAR Radar Altimeter (of Sentinel-3)
SRD	System Requirements Document
SSD	System Specification Document
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SVR	System Verification Report
SWIM	Surface Waves Investigation and Monitoring (instrument of CFOSAT)
SWH	Significant Wave Height (see also Hs)
TB	Brightness Temperature
TBC	To Be Completed
TOPC	Terrestrial Observation Panel for Climate
TR	Technical Requirement

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
USB	Universal Serial Bus
USGS	United States Geological Survey
VOS	Volunteer Observing ships
WCRP	World Climate Research Programme
WGClimate	Joint CEOS/CGMS Working Group on Climate
WMO	World Meteorological Programme
WS	Wind Speed
WWA	World Wave Atlas (of FUGRO)

2 Sensor main characteristics

2.1 Introduction

This section presents the main characteristics of SMOS, SMAP and Aquarius sensors. It provides information about revisit time and mean footprint resolution.

2.2 SMOS

The main SMOS characteristics are:

- ✓ Interferometric radiometer with center frequency of 1.41 GHz and bandwidth of 27 MHz

- ✓ Data time coverage: 2010-now

- ✓ sub-cycle of 18 days

- ✓ Exact repetitive cycle : 149 days

- ✓ Earth Incidence Angle: 0-60°.

- ✓ Local ascending/descending time: 6 AM/PM.

- ✓ four polarizations

- ✓ 3-dB (half power) footprint size: between 40 and 100 km (according to the incidence angle)

- ✓ Global coverage : 3 days

2.3 SMAP

The main SMAP characteristics are:

- ✓ Radiometer (6-meter mesh antenna) with center frequency of 1.41 GHz and bandwidth of 24 MHz

- ✓ Exact repetitive cycle of 8 days

- ✓ aft and fore acquisition

- ✓ Data time coverage: 04/2015 to now

- ✓ Conical scanning at 14.6 rpm. Scan time: 4.1 sec

- ✓ Earth Incidence Angle: 40°.

- ✓ Local ascending/descending time: 6 PM/AM.

- ✓ four polarizations

- ✓ 1000 km wide swath.

✓ 3-dB (half power) footprint size: 40 km.

✓ Global coverage : 3 days

2.4 Aquarius

The main Aquarius characteristics are:

✓ Radiometer (3 beams) with center frequency of 1.413 GHz and bandwidth of 25 MHz.

✓ Exact repeat cycle of 8 days

✓ Almost global coverage : 7 days

✓ Data time coverage: 08/2011 to 06/2015.

✓ Earth incidence angles: 28.7, 37.8, and 45.6°.

✓ Footprints for the beams are: 74 km along track x 94 km cross track, 84x120 km and 96x156 km yielding a total cross track of 390 km.

✓ Measurement every 1.44s (about every 10 km).

✓ Distance between beam swaths of about 100 and 150 km (across track).

✓ Local ascending/descending time: 6 PM/AM.

✓ TH and TV

✓ Aligned with a scatterometer (1.26 GHz); both instruments polarimetric.

2.5 AMSR

Not considered in year 1.

3 L1 uncertainty characterization

3.1 Introduction

L1 processing is very complex and requires specific expertise. At our level, it is not possible to review all L0-L1 processing methods. The purpose of this project is to replay L2 from existing L1 products. It is therefore not our objective to re-estimate TB errors but to take what comes out of the L0-L1 and work with them, as long as we have sufficient information to properly propagate TB errors to level 2. TBs at the L1 processing output may have systematic errors that are not corrected at L0-L1.

Remaining differences between forward model predictions and TB data from SMOS, Aquarius or SMAP are still found after instrumental calibration. These differences may exhibit seasonal patterns varying from ascending to descending passes, due mainly to errors in the thermal model/monitoring of the instrument, Radiative Transfer Model (RTM) inaccuracies, input auxiliary EO data , …etc.

3.2 SMOS sensor

SMOS is an L-band interferometer that measures the Fourier transform of the scene. Level 1 processing is the passage of visibilities (which integrate antenna gains) to the Fourier transform of TB then the passage of data TB in the frequency space domain to the physical space domain. These different processing require knowledge of antenna gains, with, as an additional difficulty, a spatial sampling of the observed frequencies lower than Shannon's sampling. Since the scene has infinite frequencies, this poses specific difficulties for the passage into physical space domain. In the following, this operation is called reconstruction

Complex calibrations for thermal drifts based on Noise Injection Radiometer data and several on-board thermistor measurements are used to calibrate the visibilities. Short-term calibration is regularly performed at raw level to compensate for high variability drifts. In addition, cold-sky calibration is performed at several occasions in a year when the satellite sensor is rotated upward sky during dedicated manoeuvres (used for the so-called Flat Target Transformation). Yet, systematic and seasonal image reconstruction errors are still found in the reconstructed level 1 data despite raw data calibration. This can occur, for example, due to the instrument response to a very strong L-band source in the field of view, such as the sun image and its tails corrupting the quality of the reconstructed brightness but also because of image reconstruction systematic errors (noise floor, aliasing, instrument impulse response function, antenna pattern uncertainties).

To compensate for these distortions in the image, a vicarious calibration is performed at level 2 by evaluating a mean spatial difference in the antenna coordinate frame between SMOS antenna TBs and a radiative transfer forward model of the brightness obtained along specific orbits in the middle of the Pacific. The forward model is derived using climatology of SSS or analysed in situ

data (ISAS fields) interpolated along the half orbits used for calibration. The Tb adjustment is named the Ocean Target Transformation (OTT).

Following this correction, it is possible to empirically validate the errors on TBs against the expected radiometric noise. These errors are in accordance with the expected radiometric noise, as soon as the various unmodeled contributions have been filtered (RFI, sun effect, etc.). On the other hand, there are reconstruction biases that are not corrected by the OTT, which then generates biases in the estimated parameters. For the time being, L1 processing do not make it possible to avoid such biases. However, the Gibbs 2 treatment currently being implemented should reduce coastal and ice edge biases.

In SMOS Level 2 processing, an empirical adjustment based on evaluated mean biases (between forward model and objectively analysed observations) found after averaging long periods of acquisition (4-5 years) is made on the input brightness depending on the target location, polarization, imaging position within the instrument field of view and passes type. Here again, these corrections depend on a reference SSS field (ISAS, HYCOM, climatology). This empirical correction improves the signal biases along the coast line but cannot correct for moving edges such as sea-ice border. New promising corrections are underway at Level 1 (so-called Gibbs-2).

3.3 SMAP and Aquarius sensors

To correct for residual drifts after raw data calibration, NASA algorithms thus use the median difference between Aquarius (or SMAP) data and forward radiative transfer model simulations of the brightness temperature obtained by using HYCOM model SSS or Argo SSS (depending on the release) as a forcing parameter. The difference is then averaged globally and the mean difference evaluated daily is used for post-calibration adjustments.

In addition, for SMAP (it is more complicated than for Aquarius: the SMAP antenna has some non-negligible emissivity), the antenna temperature predicted from thermal model has probably some errors, and a latitudinal correction has to be applied in addition to the globally average one. This makes the SMAP calibration mode similar to the calibration for SMOS based on the OTT.

Sharp brightness transition in the satellite pixels mixing open ocean and land masses, or sea ice, generate Gibbs like phenomena produced by the sharp truncation of the Fourier Transform of the signal. In addition, uncertainties in the modelled side-lobes of the radiometer antenna patterns used for image reconstruction/antenna temperature provide some signal leakage of the brighter sources (land, sea ice) into the lighter source (pure ocean). These so-called "land contamination" or "ice-contamination" need to be corrected for the input TB to retrieve an unbiased SSS as close as possible from the coast lines or ice-edges. A method of contrasting half-space is currently used in NASA algorithms to adjust antenna pattern corrections when Aquarius or SMAP pass through two sharply contrasted (in the TB sense) semi-infinite surface (from sea to land, for instance).

In the CCI+SSS project, we will develop algorithms using the same SSS reference field (in situ fields such as ISAS) used to vicariously calibrate the Level 1 data used as input for the three sensor Level 2 algorithms. To ensure satellite and in situ SSS consistency for these calibration steps, we

will ensure that the satellite data used to evaluate the corrections can be safely compared to *in situ* observations (i.e., removal of data acquired in presence of rain, or, in zones with high SSS horizontal/vertical/temporal variability). We will thus analyse the consistency of the SSS fields from the 3 sensors in the pixels as a function of distances from coasts and sea ice edges, as function of latitude and time and propose corrections to adjust for large-scale inter-sensor differences and on a reference in situ fields (ISAS field).

3.4   RFI filtering

SMOS, Aquarius and SMAP missions operate in the L-band protected spectrum (1400-1427 MHz) that is nevertheless now known to be vulnerable to radio-frequency interference (RFI). Areas affected by RFI might experience data loss or result in inaccurate soil moisture and ocean salinity retrieved values. To alleviate this situation, several strategies were put into place to filter the data from RFI perturbed measurements. As SMOS, launched in 2009, was the first satellite to operate in L-band, it does not have any on-board hardware/software to filter RFI, so that RFI filtering/mitigation only rely on data post-acquisition processing. This issue is significantly less important for SMAP (and to a least extent for Aquarius), as they are (were) equipped with on-board frequency/time-domain-based RFI filters.

Over the ocean, SMOS data are contaminated by RFI emitted principally from land. The impact on the reconstructed brightness temperature can be positive or negative and is not limited to the location of the on-ground antenna causing the interference but affects measurements as soon as there is the line of sight between the instrument and the RFI source (Corbella, Martin-Neira, Oliva, Torres, & Duffo, 2012). Due to the interferometer principle from a Y-shape antenna, the contamination is not circular symmetric in SMOS images, but presents six main tails spreading from the RFI source. In the case of SMAP and Aquarius, the RFI contamination is different as they operate real-aperture radiometer and on-board data filtering with enhanced detection capabilities. To protect against RFI, Aquarius employs rapid sampling (10 ms, milliseconds) and a "glitch" detection algorithm that looks for outliers among the samples. Samples identified as RFI are removed, and the remainder is averaged to produce an RFI-free signal for the salinity retrieval algorithm. The RFI detection algorithm appears to work well over the ocean with modest rates for false alarms (5%) and missed detection but RFI are still detected in Aquarius (Le Vine and De Matthaeis, 2014).

SMAP takes a multidomain approach to RFI mitigation by utilizing an innovative onboard digital detector back end with digital signal processing algorithms to characterize the time, frequency, polarization, and statistical properties of the received signals. Almost 1000 times more measurements than what is conventionally necessary are collected to enable the ground processing algorithm to detect and remove interferences. A strategy to provide probability maps and or flags of the RFI contamination for each CCI Level 2 product from SMOS, SMAP and Aquarius will be proposed in order to ease a further estimate of each dataset uncertainty and allow an RFI tracking during the multi-sensor Level 3 merging process.

4 L2OS uncertainty characterization

4.1 Introduction

SSS random uncertainties are provided in the L2OS products. These errors can be assessed by comparing estimated SSS and reference SSS (in-situ). The random and systematic uncertainties can be obtained in a relative way by comparing average products from different sensors and orbit type (ascending or descending).

4.2 Methods

4.2.1 Random error estimation

4.2.1.1 Introduction

The uncertainty on the data can be obtained in different ways. The redundancy of the error estimation is essential since the different calculation methods have their advantages and their limits. In addition, redundancy allows detecting possible problems relating to specific operating points.

4.2.1.2 Error propagation

The basic uncertainties on salinity correspond to that provided in salinity level 2 products.

Level 2 algorithms are used to propagate the TB noise characterized by the radiometric accuracy, the error of all geophysical parameters (wind speed, surface temperature, etc.) on the salinity. The propagation methods generally assume a Gaussian statistic, a linearization of the direct model in the vicinity of the solution and least square type retrieval. We will see to what extent this approach is common to all sensors. The final error obtained depends on the a priori errors on the parameters. In order to homogenize L2 errors from the different sensors (SMOS, SMAP, Aquarius), a review of error propagation methods has been carried out. We then propose a strategy to standardize the error calculation and the a priori errors to be assigned to the geophysical parameters. This strategy of standardizing the error calculation must be in phase with the standardization of the auxiliary data and the uniformization of the direct and inverse models on the set of sensors. Note that for the time being, model errors are not propagated in L2 SSS. After that, when computing L3 products by combining different SSS, it is possible to use the L2 error in order to weight properly the SSS before average.

Basically, the theoretical SSS a posteriori error depends on the radiometric accuracy (σ_{TB}) and the error on the auxiliary data. If these two error sources are given, the SSS error also depends mainly on the sensitivity of TB according to the SSS. This sensitivity increases with the SST. This means that the SSS error, σ_{SSS}, increases at high latitudes. The relation is as follows:

$$\sigma_{SSS} = \sigma_{TB}.\left(\frac{1}{\frac{\partial \text{TB}}{\partial \text{SSS}}(\text{SST})}\right) \qquad \text{Eqn 4-1}$$

This yields, according to a $\frac{\partial \text{TB}}{\partial \text{SSS}}(\text{SST})$ estimation:

$$\sigma_{SSS} = \frac{\sigma_{TB}}{0.015.SST + 0.25}$$

Where $0.015.SST + 0.25$ is an approximation of $\frac{\partial \text{TB}}{\partial \text{SSS}}(\text{SST})$. SSS errors due to WS uncertainties and SST uncertainties could be added quadratically to this relation after propagation.

4.2.1.3 Error from external data comparison

The error balance obtained by propagation does not generally include the errors on the models themselves (model of galactic noise, roughness, solar contamination, etc.). In order to estimate the true errors and to validate the errors obtained by propagation, it is necessary to compare the satellite SSS data with in-situ measurements. This can be done directly with SSS L2. It is however preferable to average the L2 data of the sensors before comparison in order to reduce the noise level and to have an estimation of the bias. The comparison with the external data is carried out on L2 and L3 products. It is then possible to map the error in order to highlight any problem areas. This approach has been performed sensor by sensor (L2) and after aggregation of different sensors (L3 and L4).

4.2.1.4 Error from self-consistency analysis

We have 3 sensors that provide independent measurements. Two comparisons can be made, depending on the period: SMOS-Aquarius over the period 2012-2015, SMOS-SMAP over the period 2015-2020 and SMOS-SMAP-Aquarius over the period April to June 2015. Note that if we standardize the direct and auxiliary data, the random and systematic errors on the SSS data will not be completely independent. On the other hand, this could make possible to qualify errors related to the unexpected behaviour of the instruments (drift, problem of reconstruction, contamination by RFI ...).

Finally, it is possible to compare SSS by triple collocation by adding the in-situ data products. The difficulty of this inter-comparison lies in the fact that the spatial resolution of SSS in situ and satellite are not identical (error of representativity to be taken into account) and that the satellite salinities are affected by correlated errors (due to the use of common auxiliary data). However, it is possible to estimate a minimum error level on the sensors.

4.2.1.5 Qualitative estimation of error and identification of outliers

In some cases, we know that the estimation of the error at L2 is uncertain (especially if one identifies problems of convergence in retrieval algorithms). In this case, it is important to identify the outlier SSS and flag it accordingly. In addition, some statistical indicators for TB residues may show that the error obtained by propagation is underestimated. It is then possible to empirically re-evaluate the SSS errors upwards. Finally, it is possible to use data history to identify areas where SSS is most likely erroneous (RFI maps, seasonal maps of solar or galactic contamination, etc.).

4.2.2 Systematic error estimation

4.2.2.1 Introduction

Estimating systematic errors is much more difficult than estimating errors. In fact, in most Level 2 products, the systematic error is not directly estimated. There are two types of systematic errors: relative systematic differences (inter-sensor or intra-sensor) and absolute systematic errors (in comparison with 'truth'). The systematic errors should be corrected with identical techniques for the 3 sensors, this being an essential prelude before combining the data of the various sensors.

4.2.2.2 Estimation of relative systematic differences

The solution to this problem is not to estimate absolute SSS, but rather to analyse salinity anomalies. This approach has been applied to SMOS and has yielded very good results (Boutin et al, 2016). In particular, specific algorithms allow correcting the relative across track systematic errors (Kolodziejczyk et al. 2016) and ascending-descending latitudinal biases. This type of algorithm will be extended to the other sensors and it will thus be possible to estimate the relative biases for all the sensors.

Moreover, the bias also depends on the operating point and the sensitivity of SSS to TB. Some biases related to TB bias can be corrected a posteriori, for example in relation to SST.

In order to characterize the bias, we can distinguish 2 types of bias:

- a coastal bias independent of time which is related to the instrument function (reconstruction problem for SMOS and pollution by the existence of side lobes for SMAP and Aquarius). Even though the land emissivity is expected to vary seasonally, it is so large compared to ocean emissivity (~a factor 2) that at first order it can be considered as constant.

- a seasonal latitudinal bias that depends on sun and galactic noise contamination and possibly other instrumental drifts that are considered here periodic over a period of one year.

We consider that latitudinal bias is independent of the basin (Atlantic, Pacific or Indian Ocean) and that it applies in addition to coastal bias in an additive way.

The general formulation of the bias, for a given grid node at the position (lat,lon) is as follows:

$$\text{SSS}_{\text{obs}}(X, t, \text{orb}, \text{lat}, \text{lon}) = \text{SSS}(t) - \text{bc}(X, \text{orb}, \text{lat}, \text{lon}) - \text{bl}(X, \text{orb}, \text{t\_month}, \text{lat}) \qquad \text{Eqn 4-2}$$

with bc, coastal bias and bl latitudinal bias. SSS_obs is the observed salinity, SSS(t) corresponds to the unbiased SSS sought. X corresponds to a subset of data that is assigned in the same way through the bias. In the case of Aquarius, this may be the considered antenna beam or, in the case of SMOS, the position of the measurement in the swath, in the case of SMAP the aft and fore views.

It is possible to calculate bc and bl independently starting with the calculation of bl on open sea areas taken far from the coast. Then, a latitudinal correction is applied to the coastal pixels. From these latitudinal bias corrected data, we can estimate bc. Since the number of independent subsets of data is relatively large, the different biases can be estimated in a self-referenced way, i.e. there is no need for an external reference when considering anomalies and not absolute salinities. Note that the Eqn 4-2 requires a simultaneous estimation of SSS(t) (or anomalies with respect to a reference salinity given by the measurements themselves) and biases bc and bl since we do not use an external reference that gives us SSS(t). This is a very important point because, in this situation, we estimate the L4 products, represented by SSS(t) at the same time as we characterize the biases. The error propagation occurs at the time of this estimate. In view of this remark, an estimation method should be proposed. We have chosen to perform a Bayesian least square method that includes a time correlation length. We can process each grid node independently of each other and thus maintain the native spatial resolution of the sensors.

The resolution of this equation will follow the method described in the paper by Kolodziejczyk et al. 2016 which presents an application on SMOS.

An improvement of this correction has been proposed, in particular as regards the inclusion of SST. This approach remains valid for all L-band sensors.

The correction of the inter-dwell or latitudinal instrument bias does not depend in principle on geophysical conditions. However, if the brightness temperature bias (ΔTB) is generally independent of geophysical conditions and therefore, in a particular way, of the sea surface temperature (SST), this is not the case for the SSS bias (ΔSSS) ([RD09]). Indeed, the sensitivity of the inversion (transition from brightness temperatures to SSS) depends strongly on SST. The sensitivity of TB to SSS decreases with a decrease in SST. Therefore, a given ΔTB bias will not have the same impact on SSS at low or high temperatures. More precisely, we have:

$$\Delta\text{SSS} = \Delta\text{TB}.\left(\frac{1}{\frac{\partial \text{TB}}{\partial \text{SSS}}(\text{SST})}\right)$$

The lower the sensitivity $\frac{\partial \text{TB}}{\partial \text{SSS}}(\text{SST})$, the greater the bias on the SSS, for a constant TB bias. This behaviour obviously does not simplify the management of bias correction in SSS since, at a given point, SST can vary greatly from one season to another.

If we measure an SSS bias at SST=SST0, it is like measuring a different SSS bias at SST=SST1:

$$\Delta\text{SSS}(\text{SST} = \text{SST1}) = \Delta\text{SSS}(\text{SST} = \text{SST0}) \left(\frac{0.015\ \text{SST0} + 0.25}{0.015\ \text{SST1} + 0.25} \right)$$

$$\equiv \Delta\text{SSS}(\text{SST} = \text{SST0}).\,\text{coeff}_{\text{SST0}}(\text{SST1})$$

$\text{coeff}_{\text{SST0}}(\text{SST1})$ represents the multiplicative coefficient to be applied to the calculation of the bias when it is observed at a different SST (otherwise, this coefficient is equal to 1). So the idea is to compute the bias for a given SST, that is, to reduce, for each measure, to an average SST that it does not necessarily correspond to the SST observed at the time of the acquisition.

4.2.2.3   Estimation of absolute systematic differences.

The estimation of SSS(t) from Eqn 4-2 is affected by a global bias and the SSS(t) estimate contains essentially the SSS anomalies. The relative correction described in the previous section does not allow reaching absolute SSS field.

The absolute systematic error calculation requires correction based on climatology or in-situ data. This type of systematic error is currently corrected independently on each of the sensors at the L3-L4 level. For SMOS, the daily correction element is the OTT, then there are different correction algorithms for the TBs and the SSS at L2 or L3.

At SSS level, the resolution of Eqn 4-2 gives L4 SSS anomalies. At this level, it is possible to add at these SSS anomalies a constant shift in order to reach an absolute SSS.

4.3   Spatial sampling

The definition of the grids on which the SSS are projected can be done at several levels:

1/ from user considerations that wish to work on regular rectangular grids in (lat,lon), oversampled.

2/ from pragmatic considerations related to information content. In this case, it is a question of working on a grid that allows to switch to any other grid with a minimum loss of information. Given the resolution of the sensors, it is a question of sampling at a frequency twice as high as the resolution. The average resolutions of SMOS, SMAP and Aquarius are respectively 50 km, 40 km and 150 km. A 0.25° longitudinal sampling should then allow interpolation on any grid. The EASE grid is an area conservative grid (Figure 1), regular in longitude but not in latitude with an average sample length of 25 km. The level 2 SSS will be represented on this grid before analysis and averaging.
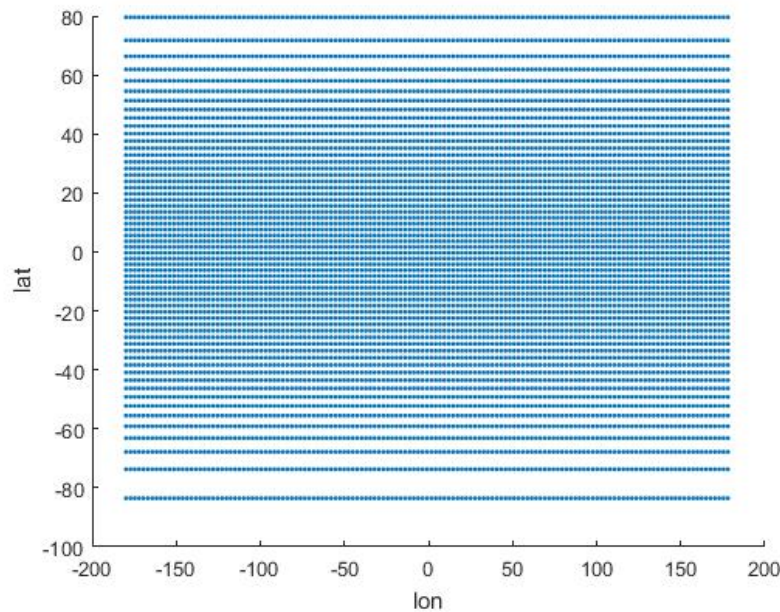
Figure 1: EASE grid (undesampling of a factor of 10).

4.4   Uncertainty estimation methods applied to SMOS, SMAP and Aquarius.

4.4.1   Introduction

We will use L2 and L3 products to characterize the error on the SSS estimator and get a first idea of the bias behaviour for each sensor.

By definition, L2 and L3 products are single sensors. L2 are given orbit by orbit while L3 are averages obtained at different spatial and temporal scales. The purpose of the project is not to produce SSS L3 but rather SSS L4. However, the production of L3 allows for sensor-by-sensor analysis, which can simplify the diagnosis because the information from the different sensors has not yet been mixed. In addition, L3 products (not produced in year 1 by CCI project) contain not only SSS fields but also monthly standard deviations that are close, in most cases, to the expected errors on L2. L3 products are also easier to handle than L2 products for diagnostic purposes because they consist of less noisy SSS average fields and have a reduced volume.

We will try to apply common methods to the 3 sensors knowing that our analysis is at level 2.

4.4.2   Pilot Mission Exploitation Platform PI-MEP

(from https://www.smos-pimep.org/overview.html)

The Pilot Mission Exploitation Platform (Pi-MEP) is a ESA funded project, which aims at supporting enhanced SSS validation and scientific process studies over ocean.

Pi-MEP project objectives are:

- Focus 1 - Enhanced validation of satellite SSS and products assessment
- Focus 2 - Oceanographic exploitation and case-studies monitoring

The **Pi-MEP** is designed to allow **systematic comparisons between available datasets** by providing **comparable QC metrics** for all these SMOS data derived SSS products, as well as for the two other NASA missions. This will enable:

1. the user to choose which satellite SSS product is best adapted for their own specific application,
2. to improve the Level 2 to Level 4 SSS retrieval algorithms by better systematically identifying the conditions for which a given SMOS, or other satellite, SSS products are of good or degraded quality.
3. to in fine converge towards best approaches and generate less but better satellite SSS products.

A large ensemble **of in situ SSS data distributed by different data centers** can be used to infer SMOS, Aquarius or SMAP SSS data product quality. This includes in situ data from the following sources:

- **ARGO float data** (CORIOLIS)
- **Moored buoy** data (TAO, PIRATTA, RAMA, STRATUS, NTAS, SPURS1-2, WHOTS)
- **Thermo-Salinograph** data installed on Voluntary Observing Ships (LEGOS, SAMOS)
- **Thermo-Salinograph** data installed on Research Vessels (GOSUD, Polarstern, NCEI-0170743)
- **Thermo-Salinograph** data installed on Sailing Ships (GOSUD)
- **Surface Drifters** (LOCEAN)
- **Equipped marine mammals** (MEOP)
- **Analysed in situ data fields** (IFREMER/LOPS)
- **Dedicated Campaign data** (e.g. SPURS)

So the PI-MEP could be very useful in order to estimate the quality of the L2OS and L3OS products.

4.4.3   External data

In order to compare the SSS of the different sensors with external data, we will use the ISAS data and the different comparisons made at PI-MEP. It should be noted that ISAS data are very spatially smoothed (600 km) which can lead to interpretation difficulties when comparing better spatially defined satellite fields over areas where spatial variability is high. Figure 2 shows such a problem: the comparison of SMOS SSS and ISAS shows a larger std of differences than the SMOS-SMAP difference, this being due to the fact that SMOS and SMAP share a relatively close spatial resolution. The errors of representativity are therefore significant in the comparison mechanism.

This representativity error could be calculated on the basis of high spatial resolution models.
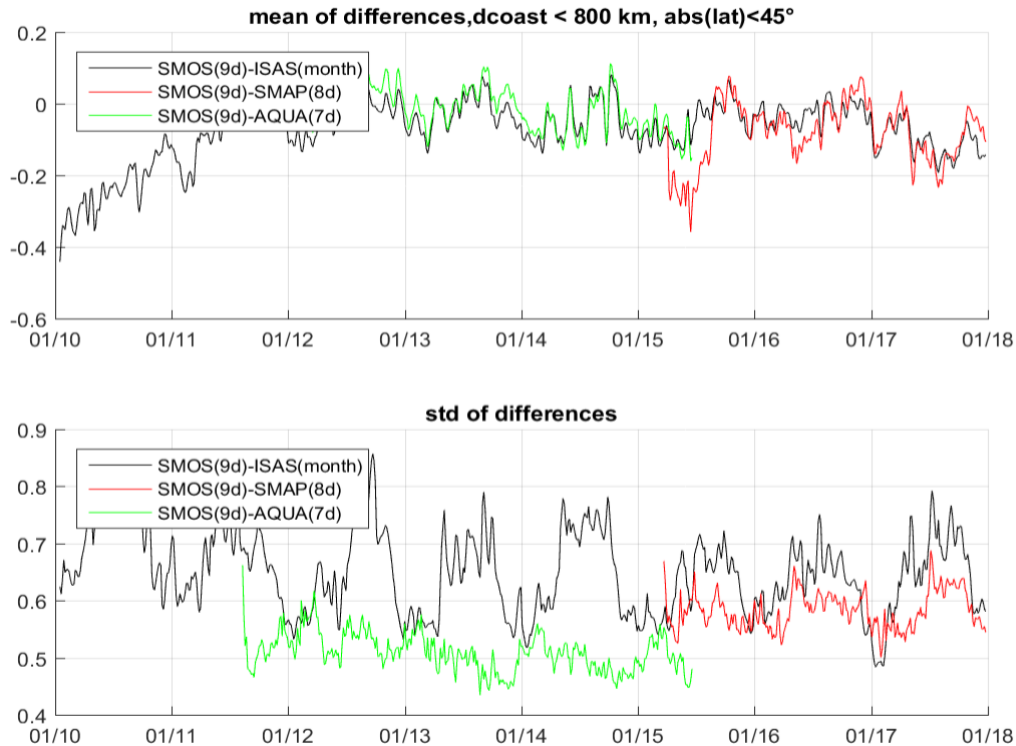

Figure 2: Average time series analysis. Comparison between SMOS-SMAP_Aquarius and ISAS. Spatial resolution effect. Top: mean SSS differences; bottom: standard deviation difference.

4.4.4 Data analysis

4.4.4.1 Introduction

Data analysis is carried out on the basis of self-consistency of SSS and comparison with external data. These comparisons will initially be based on objective analysis of ARGO in-situ data (ISASv6.2).

In year 1, the time interval considered covers early 2010 to mid-2018.

In year 1, we start from the existing L2 products (produced by the various agencies) knowing that the end of the implementation of L2 treatments is planned at the end of year 1.

The aim is to analyse the errors as they are currently managed in the different processors, starting from the ATBDs and validation reports of each sensor.

Random errors must be well known in order to be able to properly weight the different SSSs when developing L3 and L4 products. We propose to validate the theoretical errors provided in the products or to empirically estimate the errors that affect the SSS estimator (in the case of

SMAP L2C products, such errors are not provided). We know that the theoretical error (which assumes that the direct model and instrumental response are known) depends essentially on radiometric measurement errors and the sensitivity of TBs to SSS. This sensitivity depends essentially on the sea surface temperature SST. We will therefore attempt to characterize the theoretical error according to SST.

With regard to systematic errors, there are several causes that generate them:

-the instrument, which is known with a certain level of precision, undergoes poorly controlled and therefore poorly corrected solicitations (antenna temperature, antenna gains, etc.).

-the direct model used for inversion is not perfect (dielectric constant, sun, galactic, TEC...)

In all cases, systematic errors result from limited knowledge of the signal and sensor.

Currently, the various sensors have their own correction strategies for brightness temperatures.

However, as we will show, there are residual biases in salinity. Here, we are trying to build a salinity field from salinities from the different sensors. The aim is to mix SSS as homogeneous as possible and thus to correct inter-sensor bias.

Two possibilities for addressing these residual biases:

-to improve knowledge of the signal and sensors and to act on brightness temperatures and direct models.

-to compute empirically the bias and to correct it afterwards.

Here, we start from L2 (or L3) products that represent the best we can do at a given time in term of modelling and TB calibration. The empirical approach is therefore required at this level. It is clear that there is a feedback loop between the empirical calculations proposed here and improved calibration of instruments and forward models.

These biases affect the data differently depending on the across-swath position and the orbit (ascending or descending). More precisely, we know that glint effects depend on the season and the geometry of observation. There are a solar glint and a galactic glint which can, depending on geometry and latitude, have a greater or lesser impact on the signal. Similarly, the flux affecting the antenna back lobes depends on the orientation of the antennas and the position of the different sources and their intensity (for example, even if the back lobe gains are very low, a source like the sun can affect the total signal).

In year 1, we consider that the same bias affects the following subsets of measurements:

- for SMOS, the same bias is considered for each SSS from a given dwell line and for each position (lat,lon). The dwell lines are sampled across track every 25 km.

- for SMAP, the bias is considered independently for fore and aft measurements and for each (lat,lon) position.

SSS from ascending and descending orbits are also differentiated.

- for Aquarius, we will use the L3 data.

Therefore, we have several datasets from the different sensors and it is necessary to solve the Eqn 4-2 in order to estimate the biases bc and bl. As underlined in section 4.2.2.2, the accurate estimation of the biases is done at the same time as the SSS L4 are calculated. This is presented in section 5.3.

4.4.4.2   SMOS L2OS data

The SMOS data used in year 1 are as follows:

SMOS RE05 data set (v622) CATDS (Gibbs 1 processing).

4.4.4.2.1 Estimation of random error by propagation

A maximum-likelihood Bayesian approach is used in the L2 inversion algorithm, taking advantage of *the a priori* information available about geophysical parameters (SSS, SST, wind speed, TEC, etc.), hereafter denoted P_i. With this formalism, errors on TB and on the retrieved geophysical parameters are assumed to be Gaussian. The following cost function χ^2 is minimized:

$$\chi^2 = \sum_{i=1}^{N} \frac{[A_{meas\,i} - A_{model\,i}]^2}{\sigma_{Ai}^2} + \sum_{i=1}^{M} \frac{[P_{j0} - P_j]^2}{\sigma_{Pj0}^2}$$

This means that the error on the a priori P_{j0} (WS, SST …) parameters are propagated on the SSS estimator.

The theoretical a posteriori variance (error) $\sigma_{P_i}^2$ can be computed by the Levenberg-Marquardt algorithm as follows (Zine et al., 2008):

$$\begin{bmatrix} \sigma_{P1} \\ \cdots \\ \sigma_{PM} \end{bmatrix} = \sqrt{diag(\mathbf{M}^{-1})}$$

where $\mathbf{M}$ is the pseudo-Hessian, with $\mathbf{M} = \mathbf{F^T C_o^{-1} F}$, where $\mathbf{C_o}$ is the *a priori* covariance matrix and $\mathbf{F}$ the matrix of derivatives:

$$diag(\mathbf{C_o}) = \begin{bmatrix} \sigma_{A1}^2 \\ \dots \\ \sigma_{AN}^2 \\ \sigma_{P10}^2 \\ \dots \\ \sigma_{PM0}^2 \end{bmatrix}$$

The two components of this *a priori* covariance matrix $\mathbf{C_o}$ are:

- $\sigma_{An}^2 = \sigma_{A_{meas\_n}}^2 + \sigma_{A_{model\_n}}^2$ which includes $\sigma_{A_{meas\_n}}^2$, the estimated instrument brightness temperature uncertainty, and $\sigma_{A_{model\_n}}^2$ the estimated forward model error. Both are considered in the antenna reference frame. The radiometric error $\sigma_{A_{meas\_n}}^2$ is already given in the antenna reference frame. The model error $\sigma_{A_{model\_n}}^2$ is given in the ground reference frame and propagated to antenna frame before the retrieval (using the ground to antenna rotation matrix, see Appendix B of Zine et al., 2008).

- σ_{PMi}^2, the a priori variance of the geophysical parameter P_{Mi}

Providing the L2OS users with an improved uncertainty σ is key for a number of application, such as proper L2 SSS merging at Level3 and 4 (σ can be used to properly weight multiple L2 SSS observations in a specific space-time window), or for assimilation into Ocean General Circulation models, etc.

Typically, $\sigma_{A_{meas\_n}}^2$ ranges from 1.5 to 3.5 K depending on the distance to the sub satellite point. Radiometric accuracy is computed based in two main parameters: integration time of the snapshot and footprint size, or the equivalent area introduced into the computation of the measurement in the Fourier space. This means that it depends on incident angle and, therefore, there is a cross-track dependency/variation, but also there is dwell line dependency. In a first approach, we took the model error $\sigma_{A_{model\_n}}^2$ to be constant and equal to 0.5 K for H and V polarization and 0.1 K for Stokes-3 and Stokes-4. In addition, in the current version of the processor (v662), the geophysical parameter a priori uncertainties are constant as function of time and space and given as follows: $\sigma_{SSS} = 100$ psu; $\sigma_{SST} = 1°C$; $\sigma_{WS} = 2$ m/s and $\sigma_{TEC} = 10$ TECu.

A first improvement has been done by multiplying SSS a posteriori error by χ, which is the normalised square root of the cost function χ^2 after convergence.

4.4.4.2.2 Estimation of random error by using empirical approach

In order to estimate the actual errors empirically, we consider, for each grid node, time series of 40 days of data taken at different times of the year. The SMOS revisit time being about 4 days, we have about ten SSS over this period. Assuming that the SSS does not change over this period (which is true most of the time), the std of the time series gives an idea of the error. However, the error depends on the position of the measurement according to the center of the swath. In

order to normalize the data, we create a reduced dimensionless centered variable X (normalizing by the theoretical error a posteriori σ_{SSS} multiplied by the χ):

$$X=((SSS-<SSS>)/\widetilde{\sigma_{SSS}}) \text{ with } \widetilde{\sigma_{SSS}}=\sigma_{SSS} \times \chi$$

This new random variable should follow a Gaussian distribution of mean 0 and standard deviation 1 if the theoretical error is realistic. If the theoretical error is underestimated, the std of this new variable is greater than 1. In this way it is possible to homogenize SSS affected by different theoretical errors.
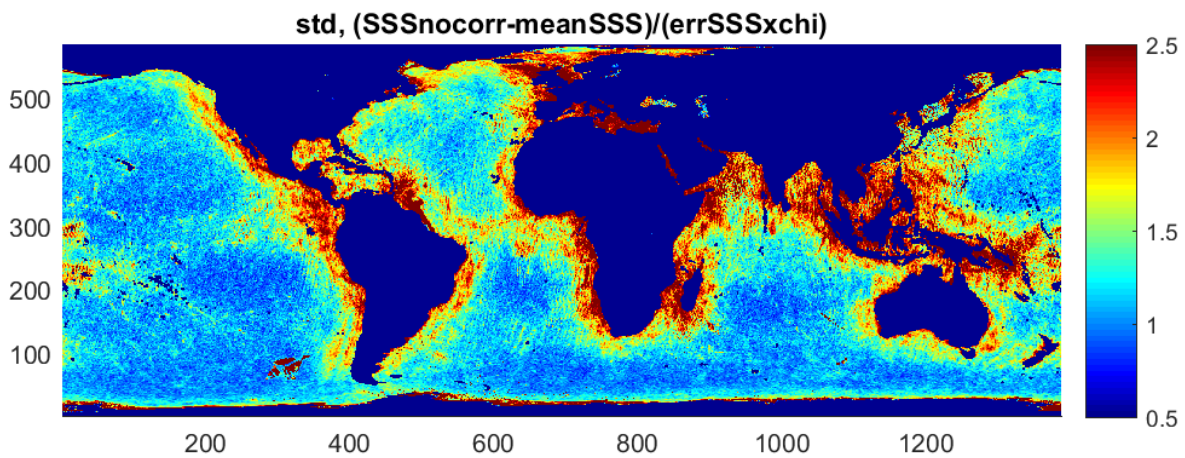
Figure 3: Standard deviation of the reduced centered variable. A value of 1 is expected if the SMOS error estimate is correct and a value greater than 1 if the SMOS theoretical error is too small compared to the true error. Example for the month of March 2012, 40 days of data.

The std map of this new random variable is given in Figure 3. The open sea areas give a rather encouraging result except in areas contaminated by RFIs (e. g. Fiji). Coastal areas have theoretical errors that are greatly underestimated. A coastal correction has been proposed by N. Kolodziejczyk & al. 2016. If we take, instead of the SSS L2, the corrected SSS, then we obtain the map on Figure 4 which shows a much better behaviour near the coast. The histograms of the reduced centred variable are presented in Figure 5, and we note that the distributions are very close to a Gaussian of zero mean and standard deviation 1. The relatively high values of the standard deviation (1.8 and 1.25) are mainly related to distribution tails and the sensitivity of the std to outliers. We can see that for the open sea, the theoretical error weighted by the χ gives results very close to the true error. For the coast, there are still high uncorrected contaminations. They are dominant in the vicinity of the Asian coast where RFIs are very strong and not temporally stable. In some regions, the high value of the std is due to the dynamics of the SSS over 40 days (rainy areas such as ITCZ, river plumes). In this case, the stability assumption of the SSS is not ensured.

Figure 6 shows std(X) according to the time (between 01/2012 and 01/2015). The dashed curves are obtained from a robust computation allowing discarding outliers. Robust std(X) is very close to 1 for open ocean. For coastal pixel, robust std reaches 1.3. This means that the proposed

approach (multiplication of the theoretical error by the χ) is not totally efficient and will need to be improved in the future in such areas.
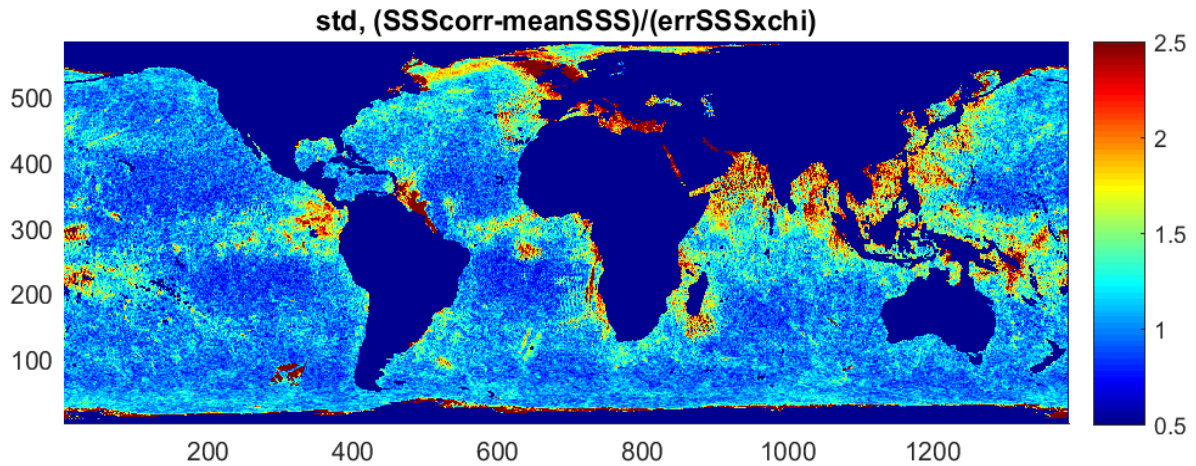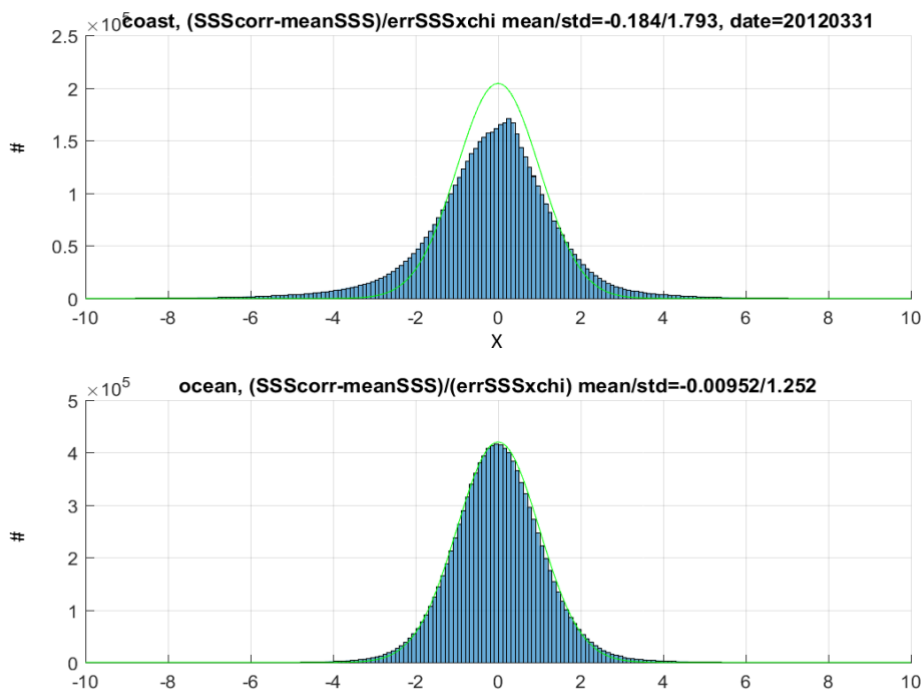

Figure 4: same as Figure 3 from corrected SSS.

Figure 5: histogram of the new random variable X (reduced centered SSS) after applying a coastal correction. At the top for pixels near the coast (dcoast<400 km); at the bottom for pixels in the open ocean. Example for the month of March 2012, 40 days of data.
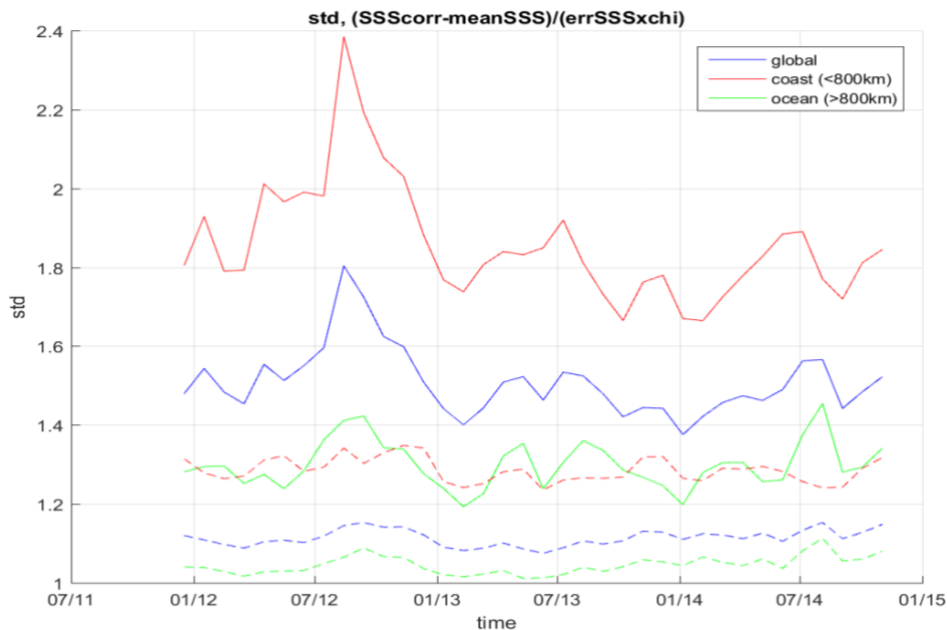
Figure 6: std(X) over a 2 year period. The continuous curves are obtained by using the classical std. Dashed curves come from robust std computation (allowing to remove outliers). Blue, for all the sea pixel; green for full ocean pixel; red for coastal area.

So, to conclude, we will use the SMOS theoretical error multiplied by the adjustment χ. This error, in some cases, is underestimated, especially on the Asian coast and places heavily contaminated by RFIs. This is a sign of the presence of outliers. The algorithm of Kolodziejczyk et al. 2016 allows us to get rid of some of these outliers by adding a 3 sigma filter from data intercomparison.

4.4.4.2.3 Estimation of systematic error

N. Kolodziejczyk et al. 2016 showed that TB biases affecting SMOS data are dependent on position on FOV and (lat,lon) so we find an SSS bias dependence on dwell-line and position (lat,lon), see Figure 8 and Figure 9.

Also, a bias depending on latitude and season was quantified (Figure 10). We are therefore in the situation of Eqn 4-2 with 2 biases to manage.

In addition, we found a bias related to the behaviour of the Klein and Swift dielectric constant model at low temperatures and to the SST noise?(Figure 7).
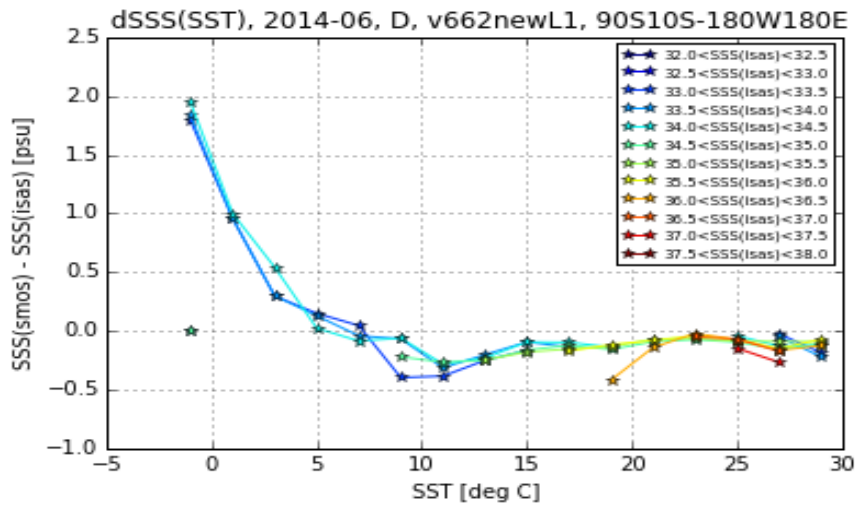
Figure 7: SSS bias as a function of SST from K&S dielectric constant (from Zhou et al. 2017).

This bias is corrected using an empirical fit of Zhou et al. 2017: bias(SSS)=0.0136.SST²-0.2553.SST+1.1874. It must be subtracted and applied over the interval[-2 8.5] °C. Note that this bias should be corrected in year 2 at the L2OS processor level.

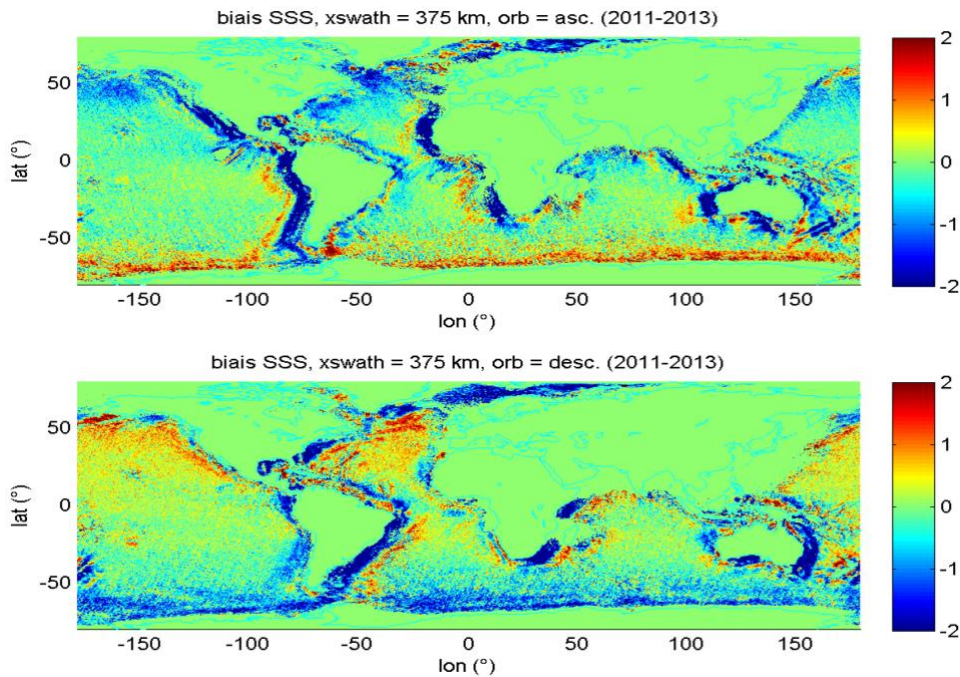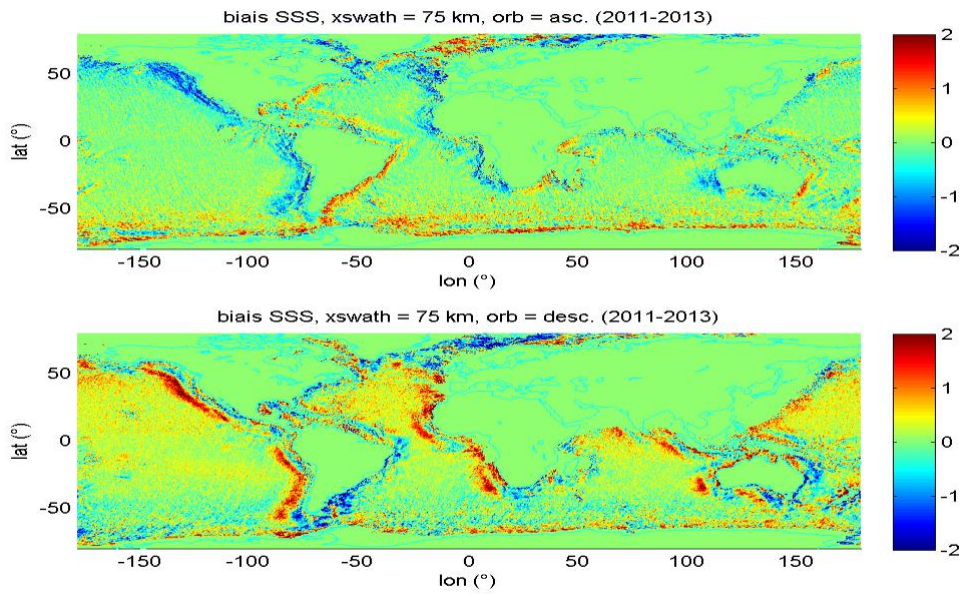*Figure 8: example of relative bias (with respect to the central dwell) calculated for xswath=375km. At the top, ascending orbits; at the bottom, descending orbits. (from CATDS, 2014 report)*

Figure 9: example of relative bias (with respect to the central dwell) calculated for xswath=75km. At the top, ascending orbits; at the bottom, descending orbits. (from CATDS, 2014 report)
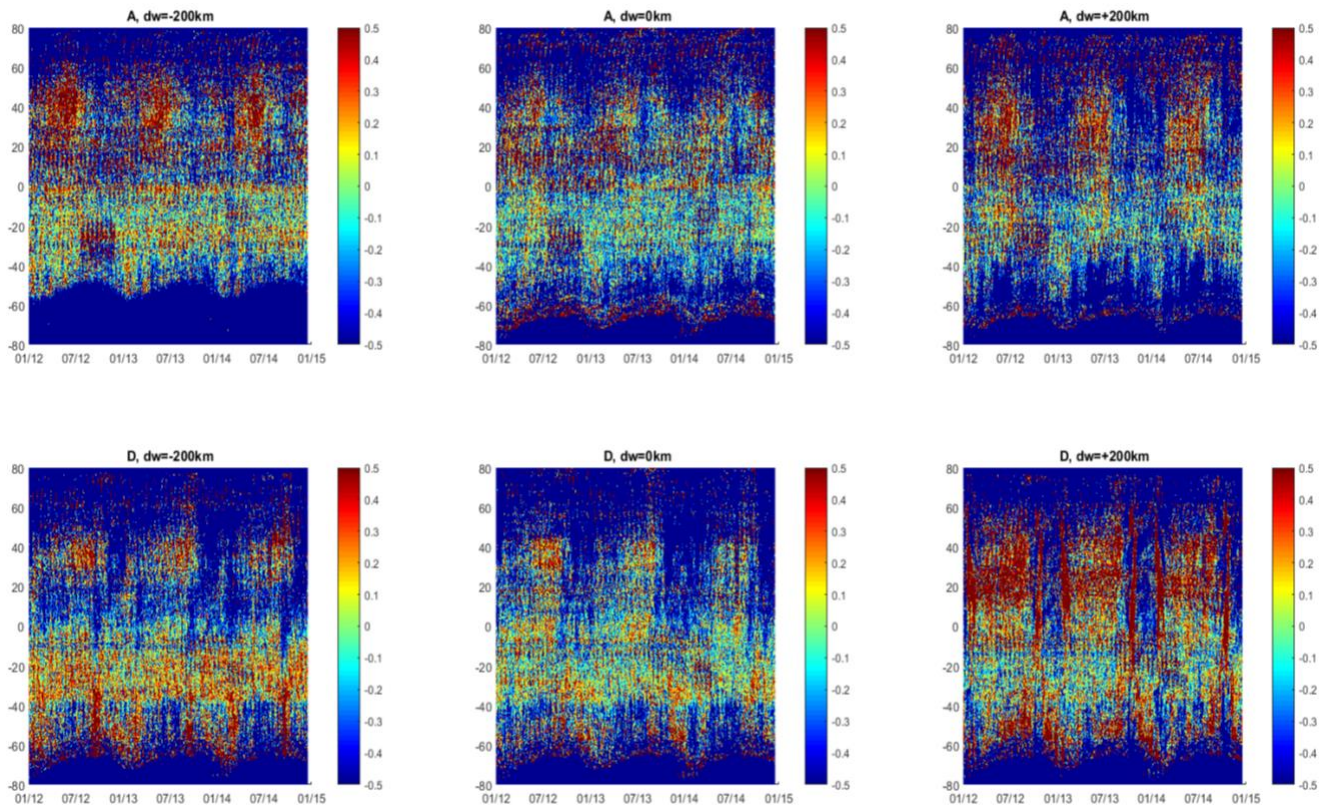
Figure 10: Hovmoller plot SMOS-ISAS before latitudinal correction. Top : ascending orbits, bottom : descending orbits. 3 dwell lines. dcoast>400km.

4.4.4.2.4 flagging the data

As we have seen, there are still SSSs for which the associated errors are underestimated. Here, we choose to exclude SSS that are associated with a χ adjustment greater than 3.

4.4.4.3   SMAP L2OS data

In year 1, we started using SMAP L2C v2.0 products (SMAP_RemSSS_Release_V2.0). During the study, a new version v3.0 has been made available with lower latitudinal bias than v2.0. That is why we preferred to use the latter. This section presents analysis elements on both versions.

4.4.4.3.1 Estimation of random error

Theoretical error is not available in L2C products.

We derived these errors empirically by using fore and after collocated SSS (Figure 11 and Figure 12) of v2.0 (the v3.0 gives very similar results to the v2.0 in term of random error). The std of the difference is very close to a modelled error with a 0.45K radiometric noise.

The afte-fore difference allows the error to be estimated more or less independently of the biases present in the direct model. Indeed, if we have a bias on wind speed, for example, this bias is found in the same way in the fore and afte targets. This is also true for the modelling of the dielectric constant, TEC, SST, etc. On the other hand, some TB contributions depend on geometry. For example, the reflection of galactic noise can be very different as well as the solar glint. However, we have applied flags that reduce its effects. The afte-fore comparison therefore probably gives a good idea of the expected theoretical noise, regardless of model bias/errors.

Comparison with external data shows that the theoretical error is less than the dispersion of the difference (Figure 13). Part of this extra-error is due to representativity error of ISAS data which are given with a 300 km resolution.
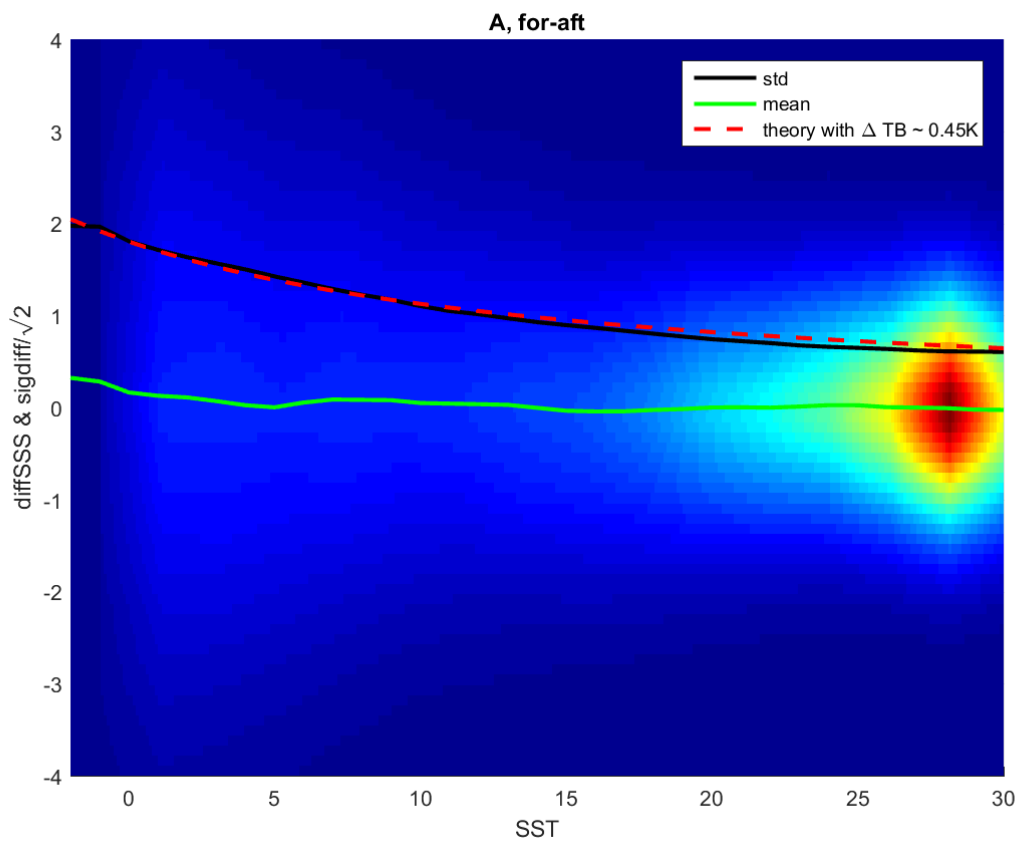

Figure 11: SSS difference between fore and afte collocated SSS. Ascending orbits. In blue, the std of the difference; in green, the mean of the difference; in dashed red, the theoretical error expected for a radiometric uncertainty of 0.45K. The color background is indicative of the point density.
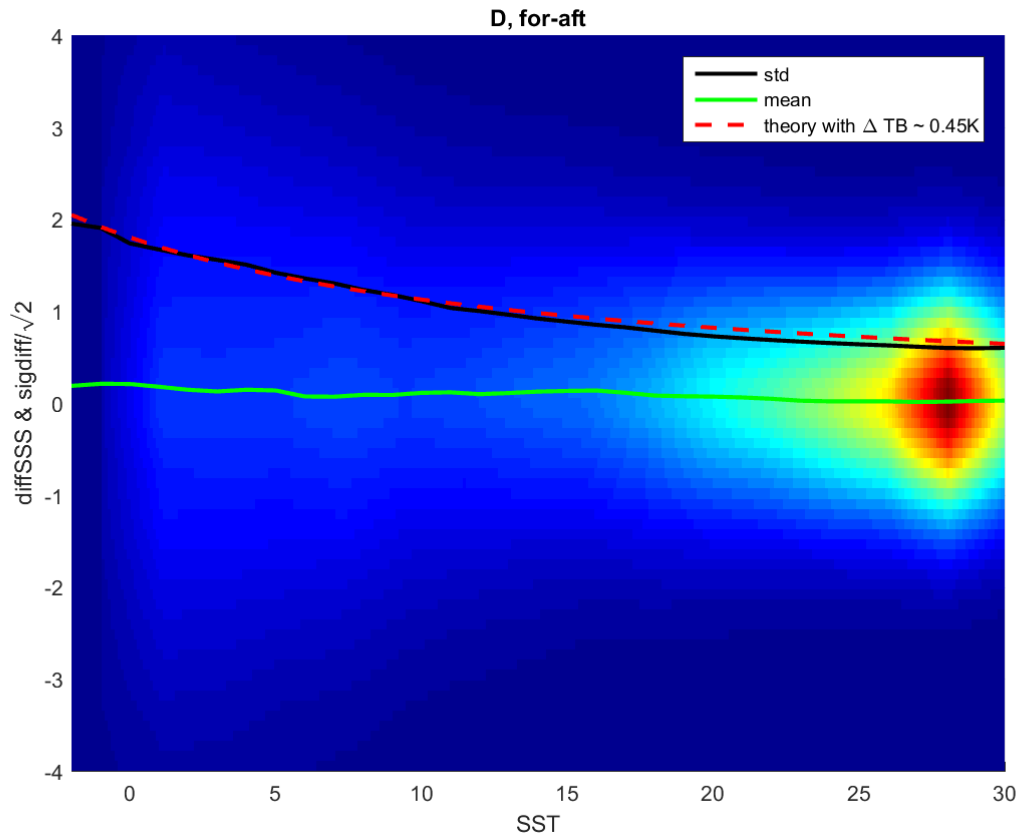
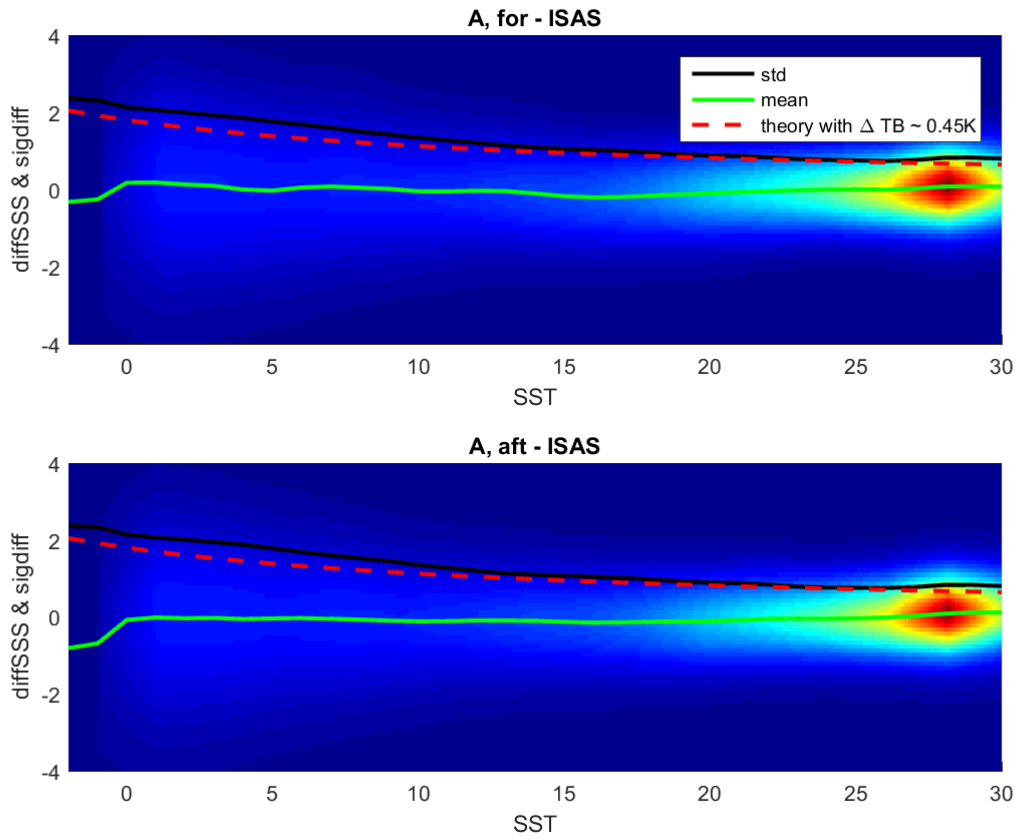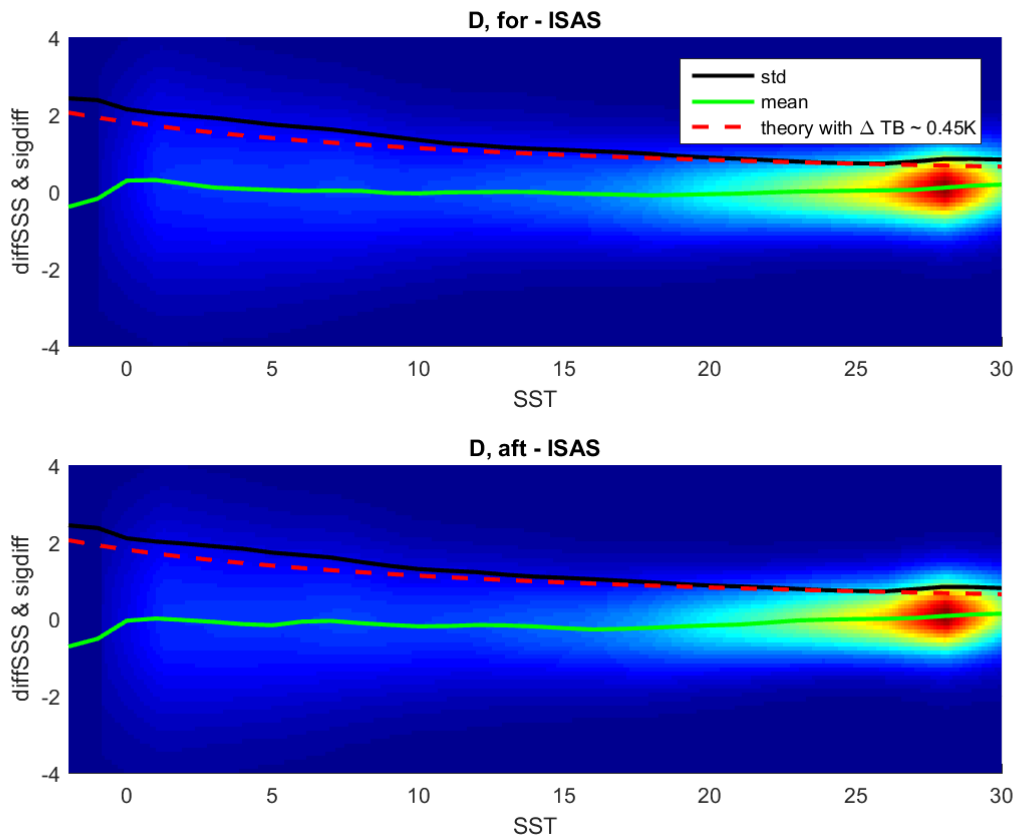*Figure 12: same as Figure 11 for descending orbits.*

Figure 13: difference between SMAP and ISAS SSS. Top : fore SSS; bottom : afte SSS. In black, the std of the difference; in green, the mean of difference; in dashed red, the theoretical error expected for a radiometric uncertainty of 0.45K. The color background is indicative of the grid points density.

Figure 14: same than Figure 13 for descending orbits

To summarize, we will consider the following theoretical error model:

sigSSS=ΔTB/(0.015.SST+0.25) with ΔTB=0.45K.

4.4.4.3.2 Estimation of systematic error (version RSS v2.0)

In order to estimate the systematic error, SSS from ascending and descending orbits have been compared. A way for doing this is to compute, over a large period, the mean difference between SSS ascending and SSS descending (Figure 15).
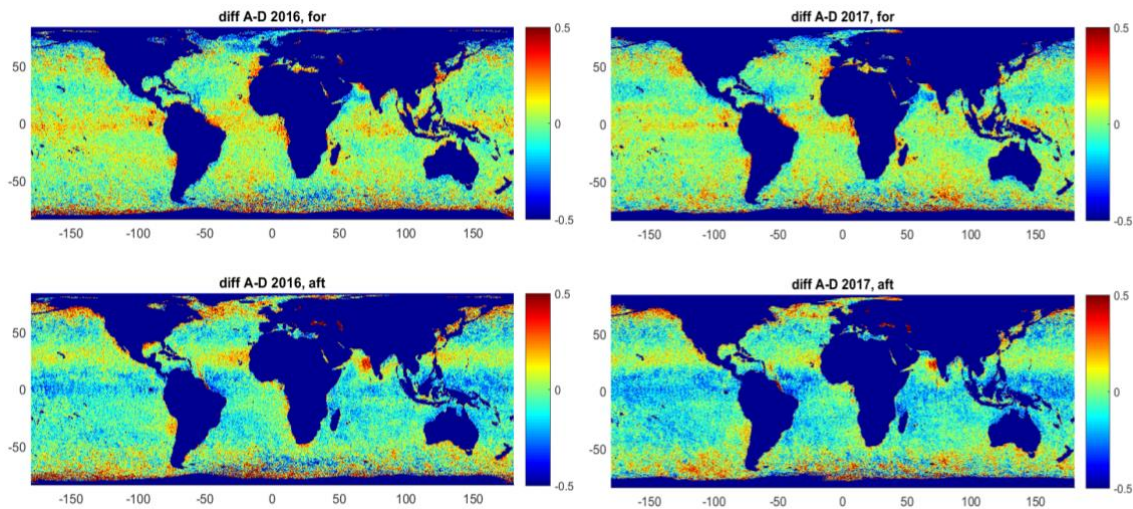
Figure 15: SSS self-consistency analysis. Mean difference between SSS ascending and SSS descending. Top : fore line of sight; bottom : afte line of sight. Left : year 2016; right : year 2017. L2C v2.

Comparison between fore and afte line of sight can be useful in order to make a self-consistency analysis of the data (Figure 16).
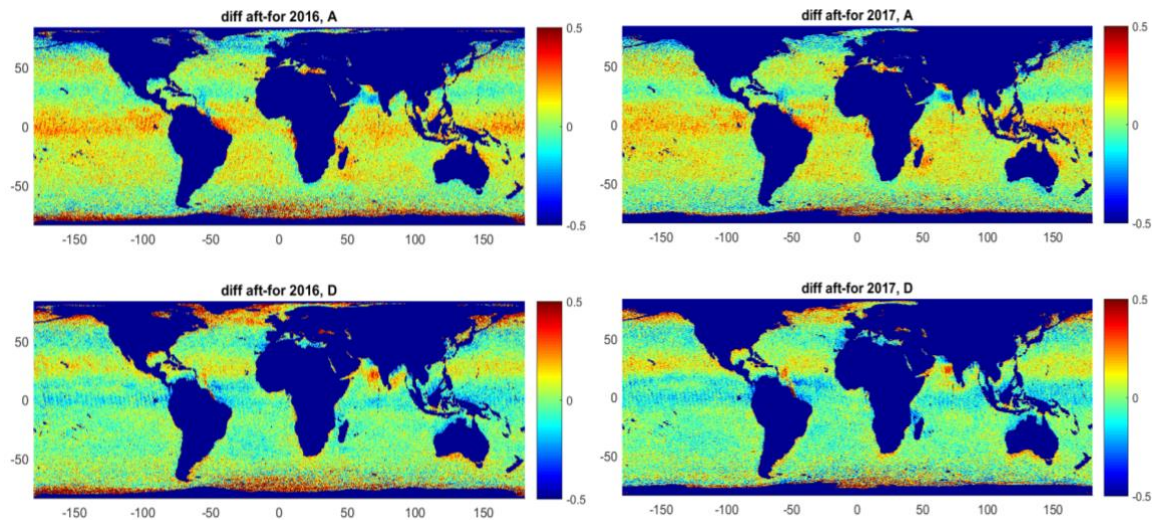

Figure 16: SSS self-consistency analysis. Mean difference between SSS fore and SSS afte.Top :ascending orbits; bottom : descending orbits. Left : year 2016; right : year 2017. L2C v2.

The Figure 15 and Figure 16 show that the SMAP data present differences according to the orbit type and the fore and after sight. The max amplitude of the differences is of the order of 0.3 psu, more pronounced close to equator. Specific differences could be found in coastal area, maybe due to RFI contamination (for instance in the Arabian Sea or in the Amazon plume). The signatures of the differences are mainly latitudinal (Figure 18, Figure 19 and Figure 20) and depend slightly on the month.

The signatures are very close from one year to the other (2016 and 2017). This means that the bias could be considered more or less stable in time at the first order.

The bias however present second order signatures (periodic patterns) which vary in time over the year (Figure 20). This second order term will be taken into account during the project 2nd year, if required.
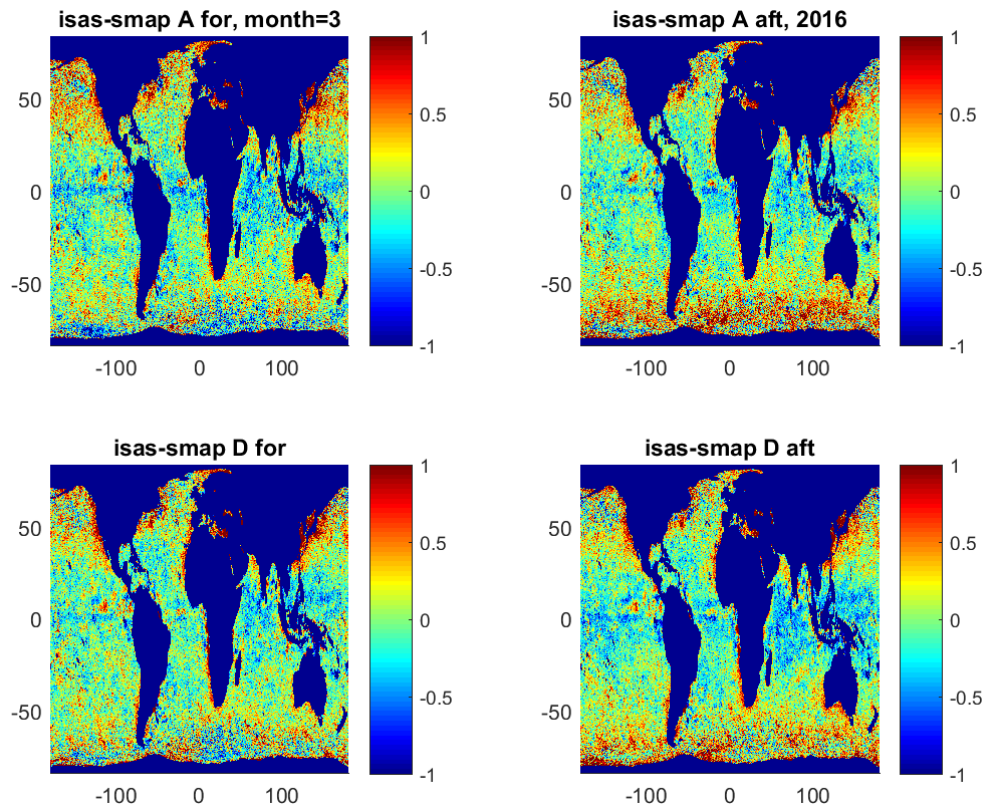

Figure 17:SSS isas - SSS smap. Top left : ascending orbits, fore SSS; top right : ascending orbits afte SSS; bottom left : descending orbits, fore SSS; bottom right : descending orbits, afte SSS. Year 2016. L2C v2.
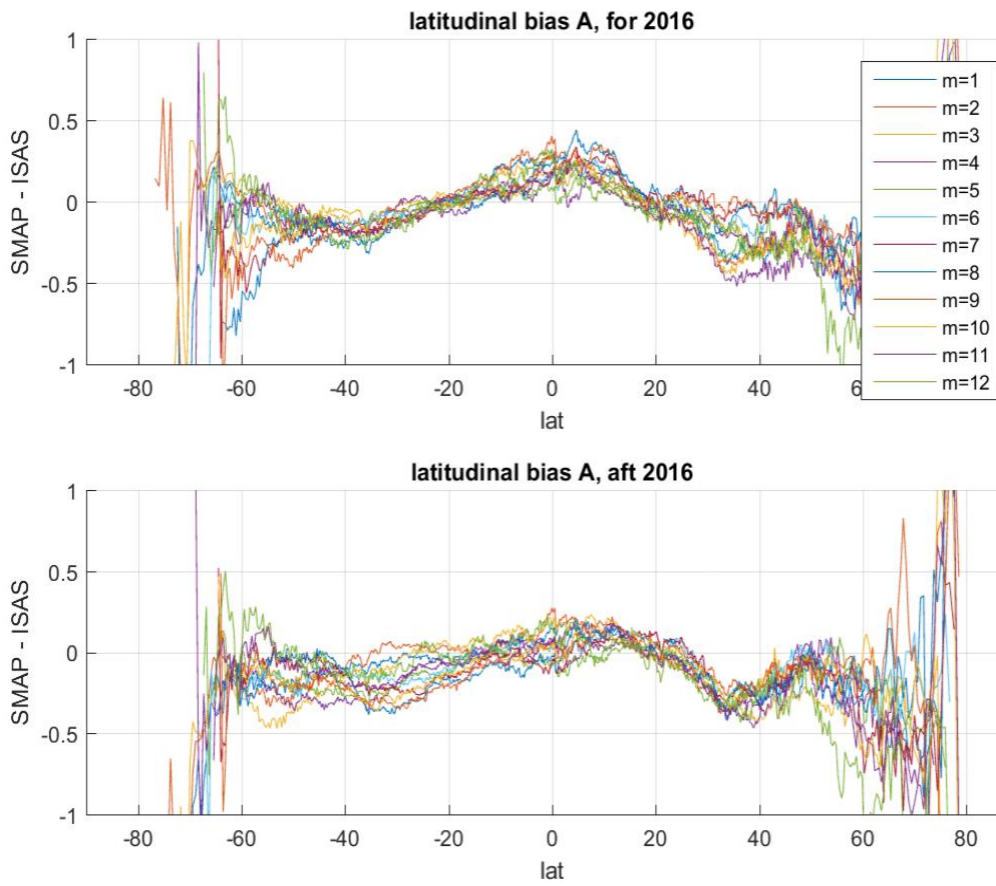
Figure 18: latitudinal difference between SMAP ascending and ISAS. Top: fore SSS; bottom: afte SSS. Each curve represents a month. Year 2016. L2C v2.
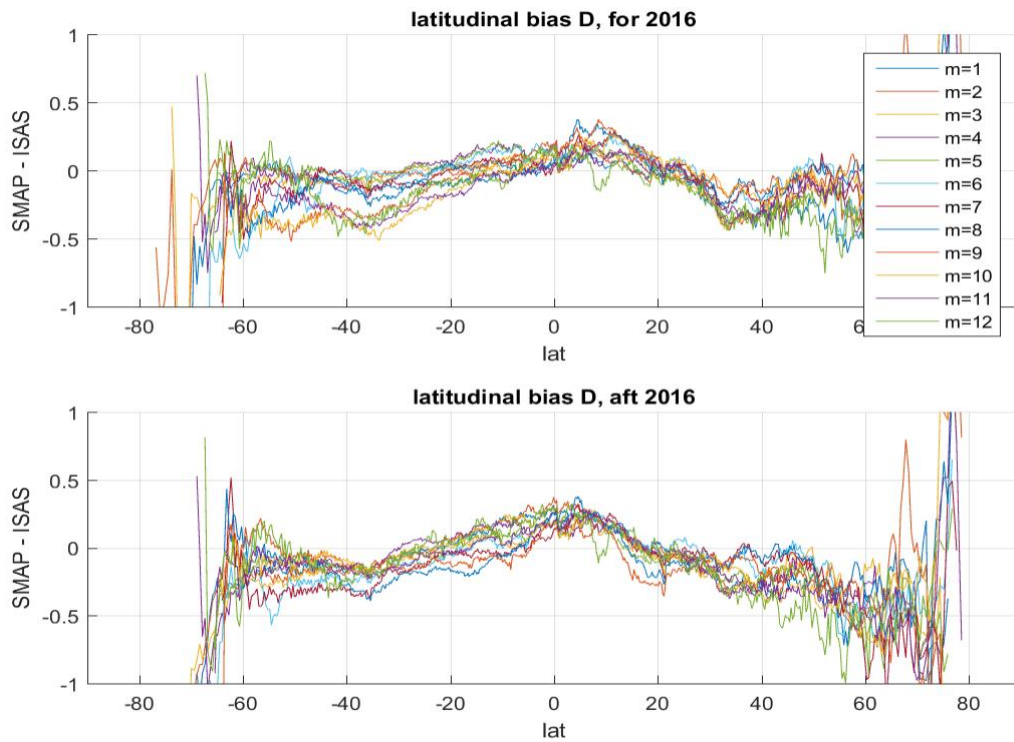
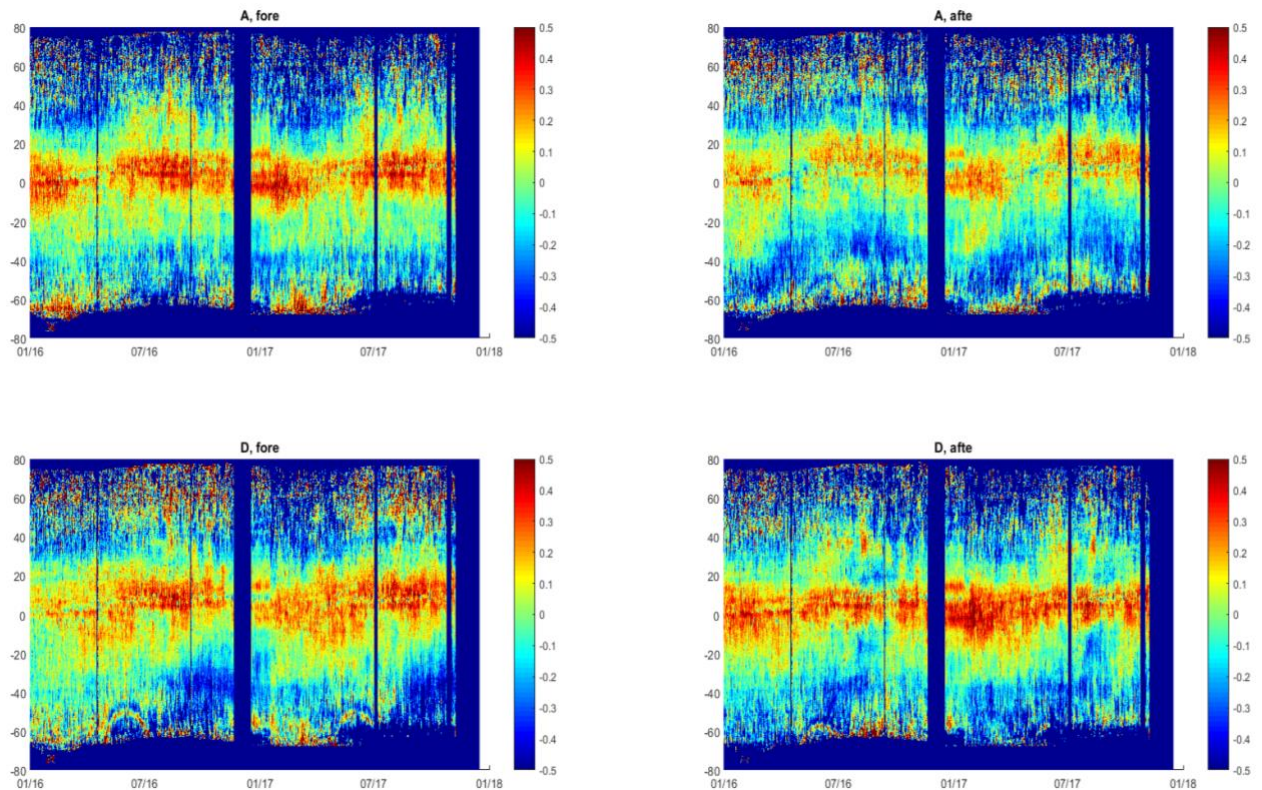*Figure 19: same as Figure 18 for descending orbits. Year 2016. L2C v2.*

Figure 20:Hovmoller plot SMAP-ISAS. Top left : ascending orbits, fore SSS; top right : ascending orbits afte SSS; bottom left : descending orbits, fore SSS; bottom right : descending orbits, afte SSS. dcoast>400km. L2C v2.

During year 1 activities, we will neglect time varying bias in SMAP data. We consider that we are dominated by a stable latitudinal bias. We will consider that this bias differs from ascending/descending passes and from fore/afte line of sights. In this situation, the bias is conform with Eqn 4-2. The latitudinal bias shall be not identified as bl but as bc (bias constant in time). It is (for year 1 only) like a coastal, time-invariant bias that extends to the open ocean and depends on latitude.

4.4.4.3.3 Estimation of systematic error (version RSS v3.0)

After September 2018, a new SMAP release has been delivered by RSS.

The main change concerns the latitudinal calibration: HYCOM has been replaced by ARGO.

This has a strong impact on the latitudinal bias signatures (Figure 21 and Figure 22).

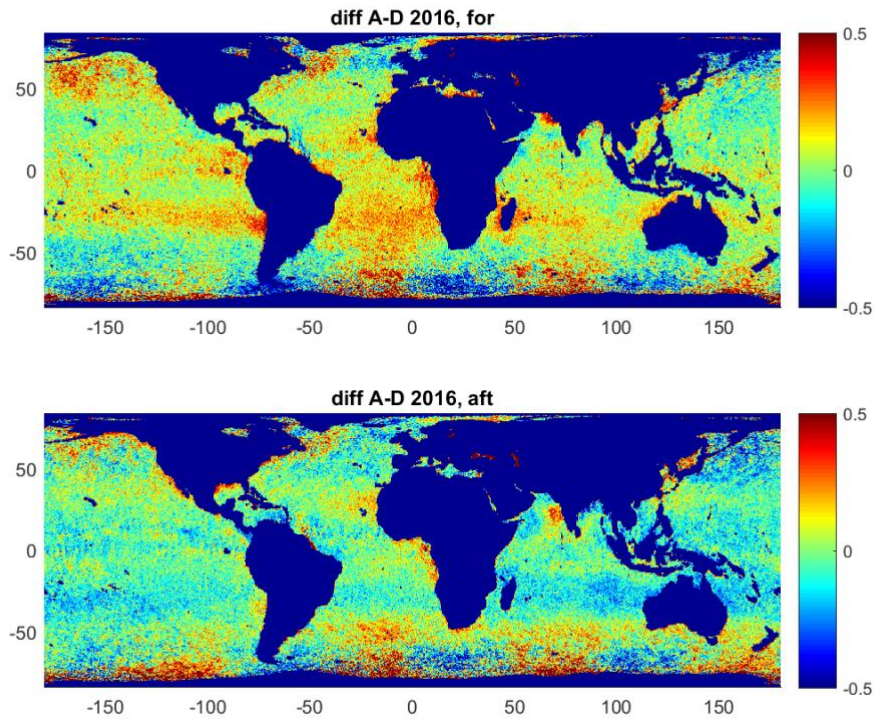The latitudinal bias vanishes in version v3 (Figure 23 and Figure 24).
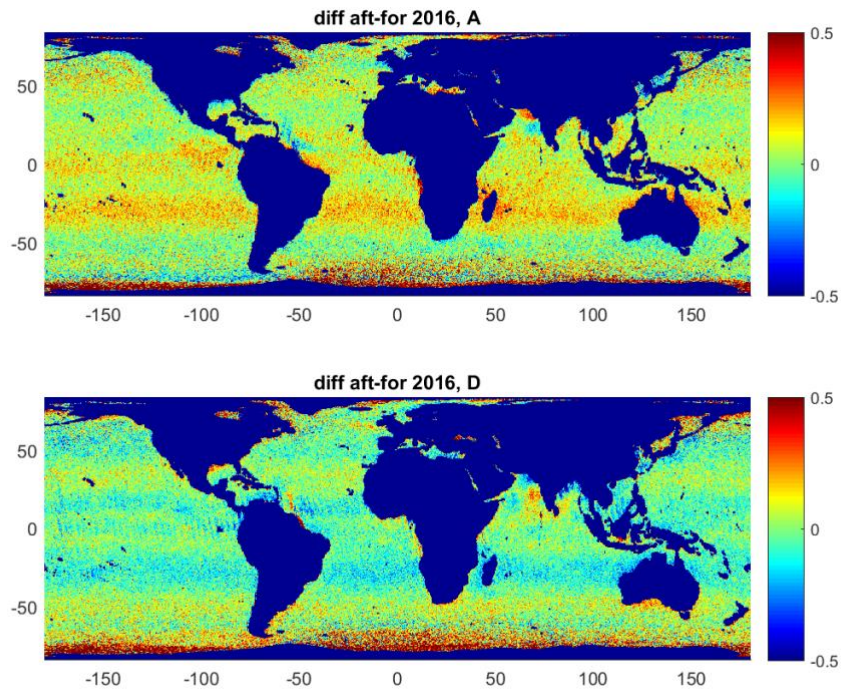
Figure 21: idem Figure 15 for L2C v3.
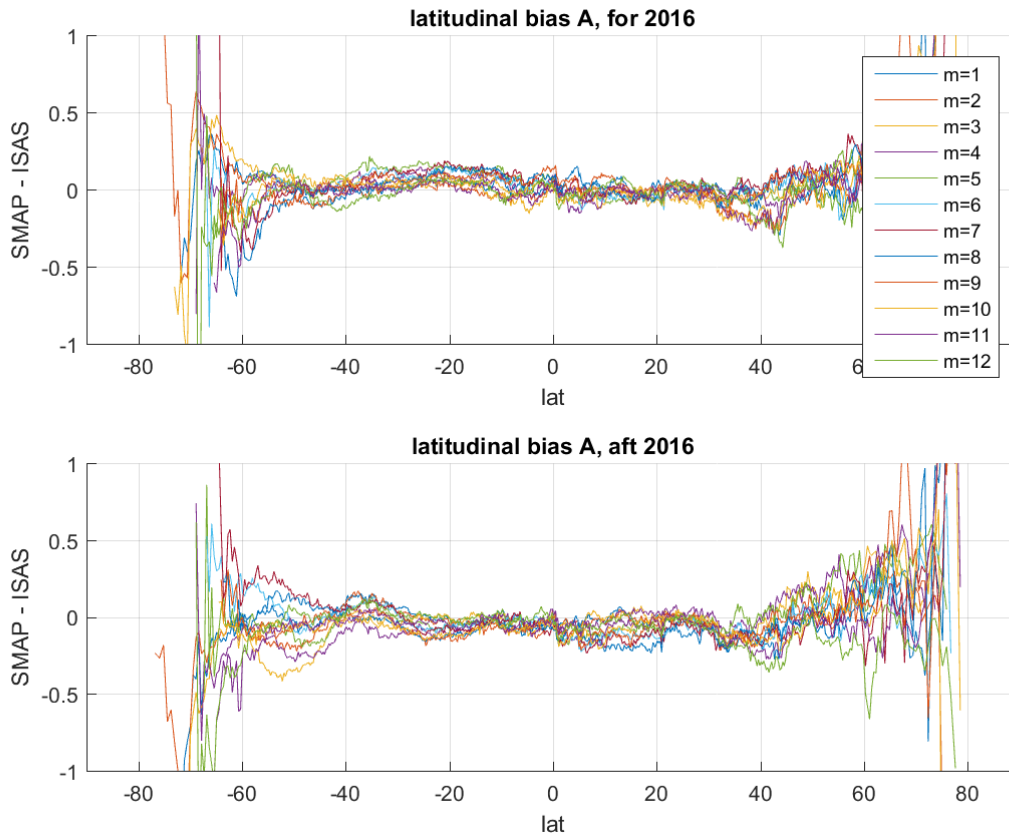

Figure 22: idem Figure 16 for L2C v3.

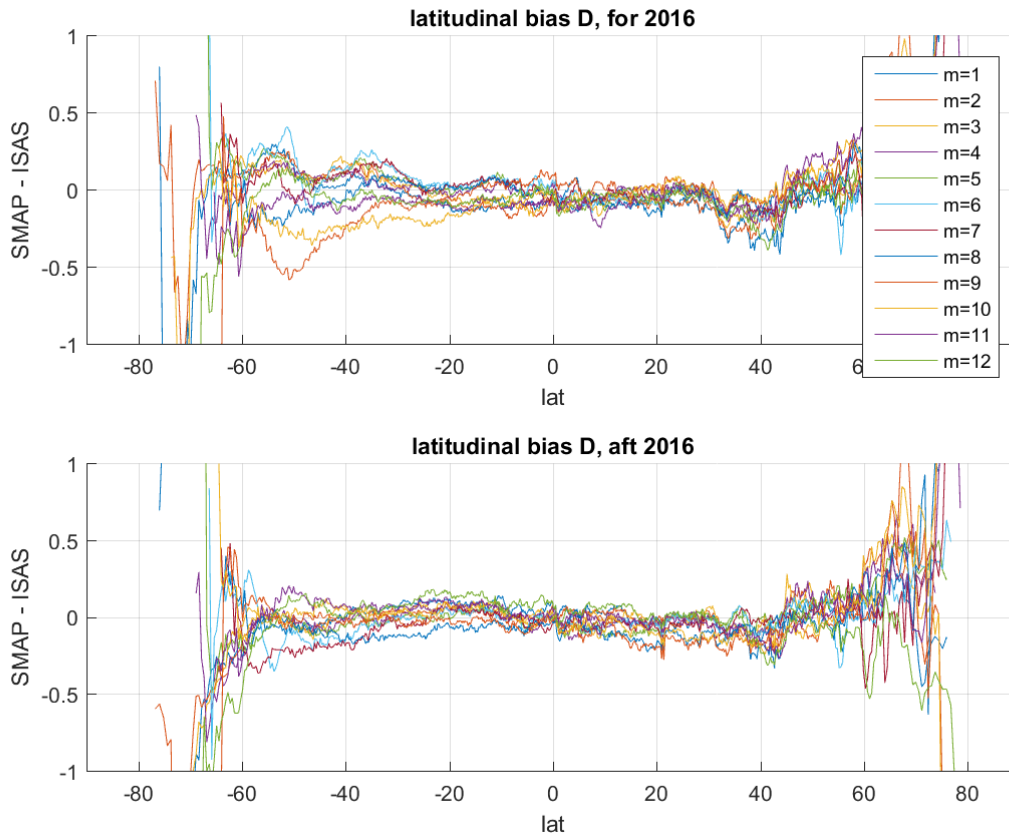Figure 23: idem Figure 18 for L2C v3.

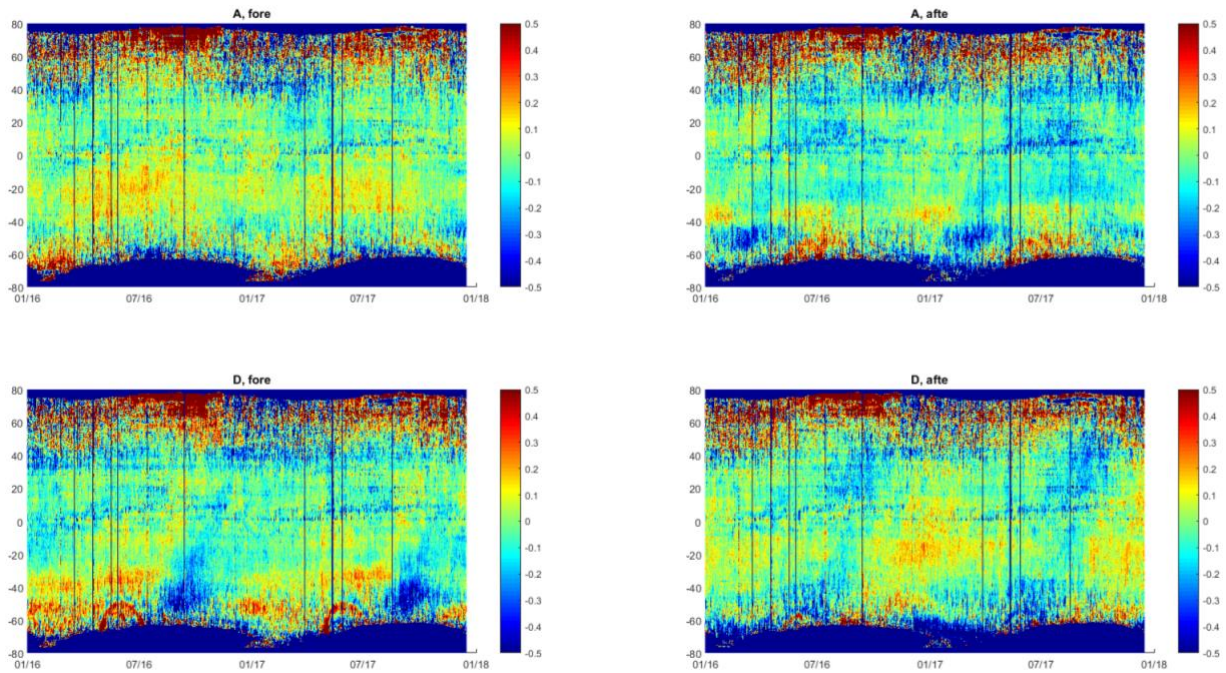*Figure 24: idem Figure 19 for L2C v3.*

Figure 25: idem Figure 20 for L2C v3.

4.4.4.3.4 flagging the data

We used an analogous filtering as the one used by RSS team (RD10):
1. The sun glint angle is less than 50° and the azimuthal look angle lies between 30° and 50° (bit 5 in L2 Q/C flag is set).
2. The moon glint angle is less than 15° (bit 6 in L2 Q/C flag is set).

3. The v/h-pol average of the reflected galactic radiation exceeds 2.0 K (bit 7 in L2 Q/C flag is set).

4. The TB consistency, which is defined as the $\sqrt{\chi^2}$ of the MLE in the salinity retrieval algorithm, exceeds 1.0 K (bit 10 in L2 Q/C flag is set).

5. The gain weighted land fraction exceeds 0.01.

6. The gain weighted sea ice fraction exceeds 0.001.

4.4.4.4 AQUARIUS L2OS data

We use the end of mission dataset: RSS L2 v5

the data description is done in the RD05 document.

The errors are provided in the L2 product but not in the L3 R7 product.

(lat lon) positions are given for each acquisition (L2 products are not given over a earth fixed grid).

4.4.4.4.1 Estimation of random error

The SSS theoretical error is given in the L2C v5 product.

The different contributions are added quadratically:

- SSS_unc_EIA: Estimated SSS retrieval uncertainty due to boresight Earth Incidence Angle random error.
- SSS_unc_galact_Ta: Estimated SSS retrieval uncertainty due to galactic effect systematic error.
- SSS_unc_IU_coupling: Estimated SSS retrieval uncertainty due to IU coupling systematic error.
- SSS_unc_moon_Ta: Estimated SSS retrieval uncertainty due to moon effect systematic error.
- SSS_unc_NEDT_X, X = {V, H, 3}: Estimated SSS retrieval uncertainty due to radiometer Noise Equivalent Delta Temperature random error at polarization X.
- SSS_unc_RFI_level: Estimated SSS retrieval uncertainty due to RFI level systematic error.
- SSS_unc_ran: Estimated random component of uncertainty in SSS.
- SSS_unc_surface_temp: Estimated SSS retrieval uncertainty due to sea surface temperature systematic error.
- SSS_unc_sys: Estimated systematic component of uncertainty in SSS.
- SSS_unc_TbV_land_contam: Estimated SSS retrieval uncertainty due to ocean surface Tb land contamination systematic error.
- SSS_unc_TbV_ice_contam: Estimated SSS retrieval uncertainty due to ocean surface Tb sea-ice contamination systematic error.
- SSS_unc_wind_speed_rand: Estimated SSS retrieval uncertainty due to wind speed random error.
- SSS_unc_wind_dir_rand: Estimated SSS retrieval uncertainty due to wind direction random error.
- SSS_unc_wind_speed_syst: Estimated SSS retrieval uncertainty due to wind speed systematic error.

The total uncertainty is given by SSS_unc. It contains random and systematic errors.

The assesment of Aquarius SSS (Hsun-Ying Kao et al. 2018, Remote Sensing, 10, 1341) shows that overall errors on L2 and L3 products are in the order of 0.17 and 0.13 psu respectively. In the paper, the collocated SSS with instantaneous rainfall > 0.25 mm/h are removed.

Large positive biases (up to 0.2 locally) are observed in the sub-polar North Pacific and in South (lat of about 40°). Large negative biases (up to -0.2 psu) are observed in the subtropical South Pacific and along the land boundaries.

We can estimate empirically the error by comparing successive revisit passes for each beam (8 day period). We make the assumption that SSS doesn't vary in 8 day time interval.
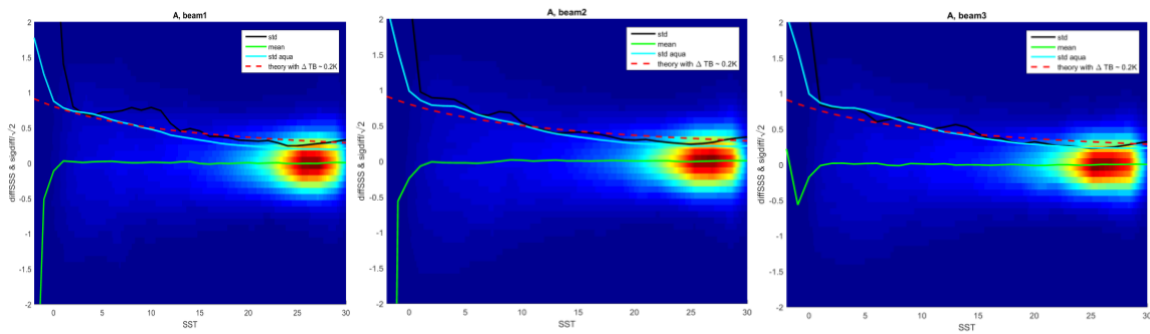
Figure 26: SSS difference between 2 passes (collocated SSS) for Aquarius beam 1 (left), beam 2 (middle) and beam 3 (right). Ascending orbits. In black, the std of the difference; in green, the mean of the difference; in dashed red, the theoretical error expected for a radiometric uncertainty of 0.2K, in cyan the Aquarius L2c SSS error. January 2013.
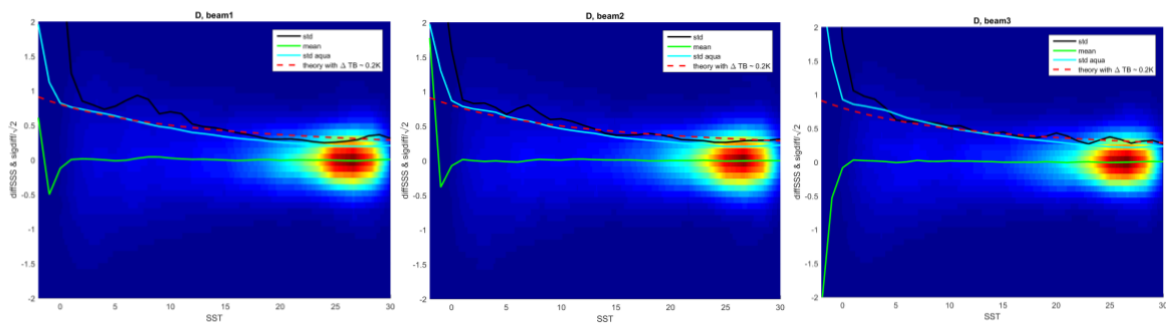

Figure 27: same as Figure 26 for descending orbits.

Figure 26 and Figure 27 show that the theoretical error provided in the L2C product is a little bit underestimated at high SST for beam 1 and 2.

We will consider these differences as neglectable and use the theoretical error as input for L3 and L4 computation.

4.4.4.4.2 Estimation of systematic error

Large positive biases (up to 0.2 locally) are observed in the sub-polar North Pacific and in the Southern Ocean (lat of about 40°). Large negative biases (up to -0.2 psu) are observed in the subtropical South Pacific and along the land boundaries (see below figures from Hsun-Ying Kao et al. 2018, Remote Sensing, 10, 1341).
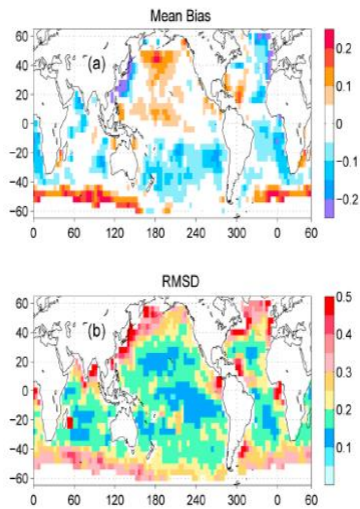
Figure 3. Geographical distribution of (a) mean spatial bias (psu), and (b) root-mean-square deviation (RMSD) (psu) between the Aquarius weekly Level-3 SSS product and Argo float observations. The error statistics were computed by comparing Argo float measurements for a given week with SSS values at the same locations obtained by interpolation of the corresponding Level-3 SSS maps. The geographical distributions are computed in 8°-longitude by 8°-latitude bins.
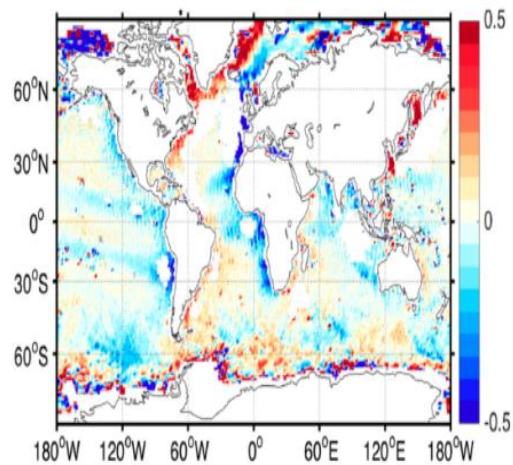
Figure 9. Forty-five-month average of ascending–descending data. Major bias regions are described in the text. The "white" regions in the North Atlantic, Western Pacific (China, Japan) or Indonesia are masked out due to suspected undetected radio frequency interference (RFI). See flagging/masking tables in [9] for details.

Comparing Aquarius with ISAS, we found that the latitudinal systematic error is relatively weak and relatively independent of the month (Figure 28 and Figure 29).
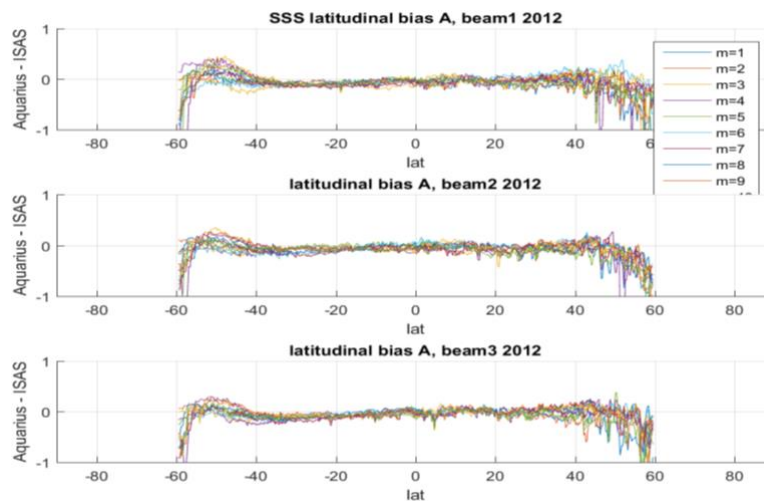

Figure 28: latitudinal difference between Aquarius ascending and ISAS. Top: beam 1; middle: beam 2; bottom: beam 3. Each curve represents a month. Year 2012.
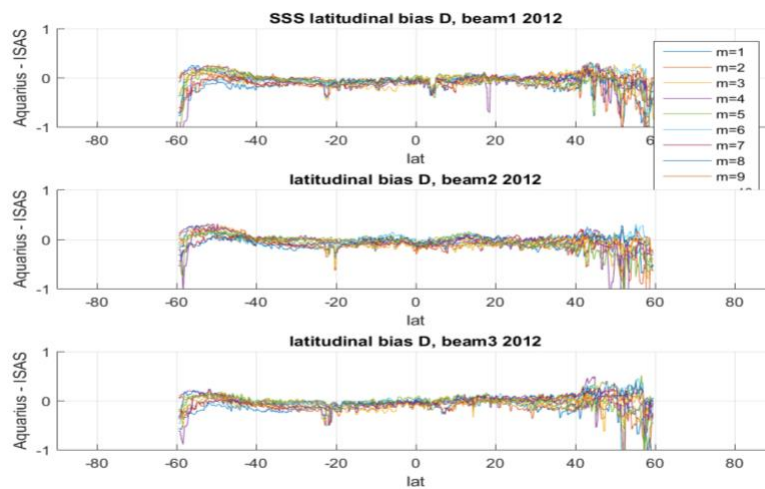
Figure 29: same as Figure 28 for descending orbits. Year 2012.
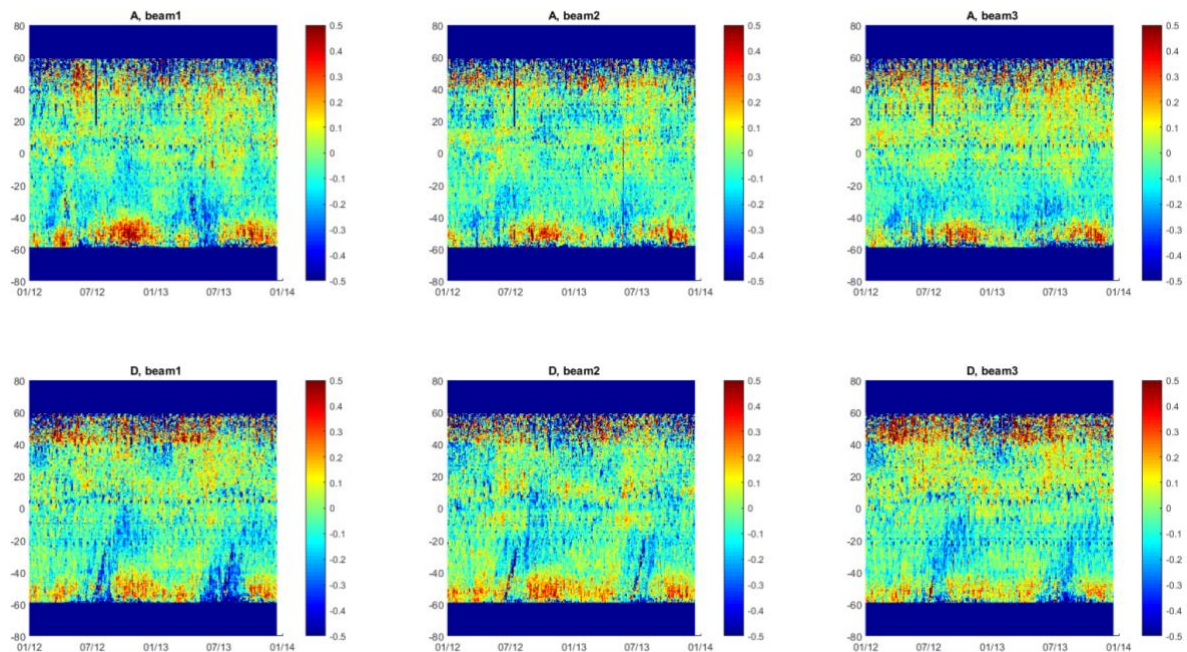
Figure 30: Hovmoller plot Aquarius-ISAS. Top left : ascending orbits, beam 1; top middle : ascending orbits, beam 2; top right : ascending orbits beam 3; bottom left : descending orbits, beam 1; bottom middle : descending orbits, beam 2; bottom right : descending orbits, beam 3. dcoast>400km.

4.4.4.4.3 flagging the data

Filtering on the flags recommended by RSS (Table 1 of ATBD) : land contamination (severe), sea ice contamination (severe), non-nominal navigation (roll, pitch, yaw), roughness, moon (severe), gal (severe), RFI (severe).

No filter over WS, SST and the rainfall flag.

5 L3 and L4 error budget

5.1 Introduction

The use of L3 data allows comparing global maps provided at different month and for different orbit types. In year 1, the project does not provide specific L3 data. In this section, we just explain how we have computed L3 SSS errors.

5.2 L3 data

5.2.1 Introduction

The SMOS, SMAP and Aquarius L3 products are not provided by the project in year 1.

In year 1, we use Aquarius L3 product as input of L4 merging. This means that we need to estimate the SSS error for Aquarius L3 SSS because this error is not provided in the product. So we start from Aquarius L2 error in order to estimate the L3 error.

5.2.2 AQUARIUS L3OS error

In year 1, L3 Aquarius products are used in the L4 aggregation scheme (Aquarius 7 days running products).

The advantage of using these products is that, unlike L2, they cover the entire globe (no interbeam missing data). The disadvantage concerns revisit times and the correlation of SSS from one day to another. It is therefore not trivial to handle the errors associated with this product.

Moreover, in L3 7 days running products, there is no information about error.

Over a week window, the number of acquisitions is N. The L2 individual SSS error corresponds to a 0.2K radiometer equivalent error with SST dependence as shown in section 4.4.4.4. :

```
sig_SSS_aqua=(0.2/sqrt(N))./(0.015.*SST0+0.25);
```
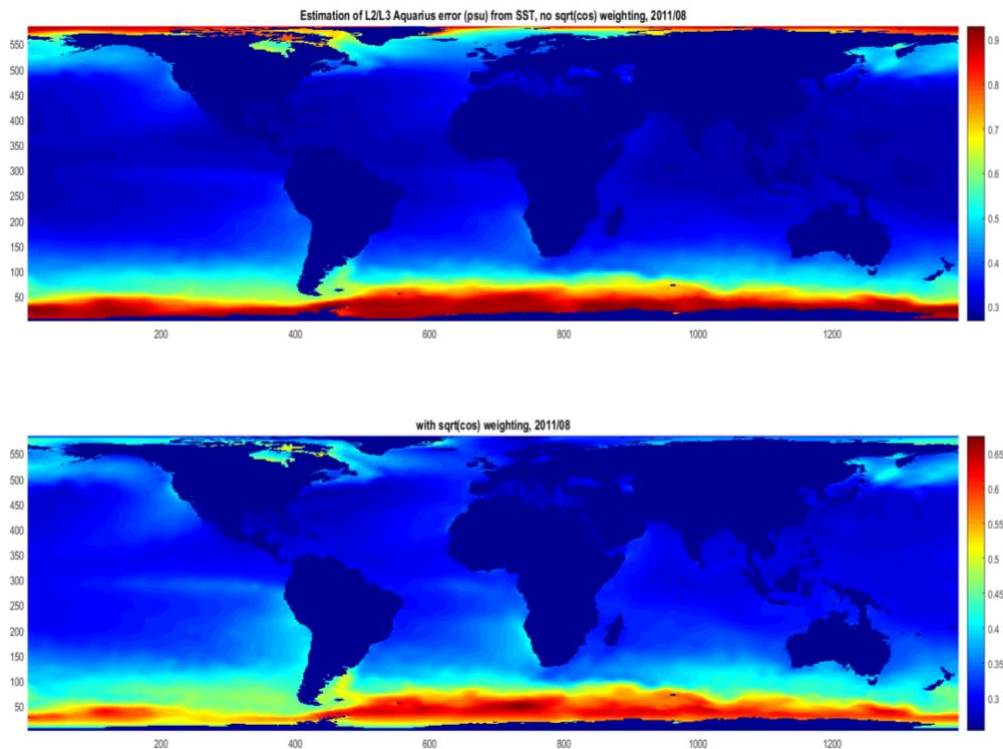
Figure 31: Aquarius L2/L3 error estimation from SST and number of revisit. Top, error of L2 according to the SST; bottom, error of L3 according to SST and acquisition number.

It is clear that on a 7 day running average there is a strong error correlation. These correlations have to be taken into account, for instance by oversampling the daily data (see Figure 32). This sampling effect will be analysed further in the next release of the document.
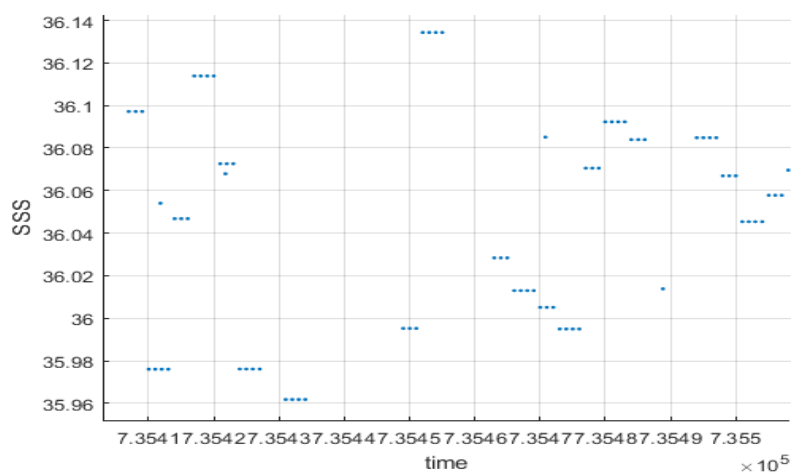

Figure 32: Aquarius 7 days running product. Time series of SSS for a grid point in the Atlantic Equatorial. The SSS is highly correlated from one day to the other day. Every dot is a daily value.

The problem is to know the number of acquisitions that generates a daily product averaged over a window of 7 days.

In one week Aquarius spends 28 orbits (14 ascending + 14 descending) * 7 days. Its swath width is ~ 400km.

So it observes a cumulative distance of: 28 * 7 * 400 = 78400km ~ 80000km.

Because the length of the earth's turn at the equator is 40000km, we have 80000/40000 = 2 passes in 7 days

At another latitude, the length of the earth's turn is 40000*cos (lat), so the number of passes is 2/cos (lat), eg at lat=45 °, there must be 8/(4cos (45)) = 2.8 passes, and at lat=55 °: 3.5 passes.

So we have to take the individual error Aquarius (for an acquisition) and divide by sqrt(2/|cos (lat)|).

We use a data every 3.5 days, which makes us oversampling by a factor of 2 (compared to the window 7 days) and so we must multiply the error by sqrt(2).

Result of the analysis:

-1 day every 3.5 days.

-The error should be multiplied by sqrt (|cos(lat)|)

```
sig_theo=(0.2*sqrt(|cos(lat)|))./(0.015.*SST0+0.25);
```

5.3   L4 data

5.3.1   Introduction

The L4 products v1.6 delivered at the end of the first-year exercise consist in two Level-4 datasets:

-        A monthly mean product centred the 1st and 15th day of each month.

-        A 7-days running mean at one day time sampling

These products contain:

1/ SSS field

2/ random error

3/ systematic error + std(bias)

4/ flag good/bad computed from different indicators ($\chi 2$, number of outliers...)

5/ number of outliers

6/ PCTVAR (ratio between the a posteriori variance and the a priori variance).

The random error can be computed in different ways (see here-after). The algorithm used to estimate the L4 error effectively provided in the L4 product is described in the ATBD.

5.3.2 Error computation

In order to merge the products from different sensors, the relative uncertainties between the different products are first calculated. The bias at each point is modelled into 2 contributions: a seasonal latitudinal contribution and a time-independent component related to coastal contamination. These biases depend on the orbit (ascending or descending), the position on the swath and the viewing orientation.

For SMOS, we consider that the bias depends on the dwell line: we sample the dwell lines in 25 km steps. The dwells are kept between -415 and 415 km.

For Aquarius, a relative bias is considered (L3 Aquarius data is used).

For SMAP, a different bias is considered for fore and after acquisitions.

In fact, it is possible to calculate the inter-sensor calibration bias.

For Aquarius, we do not consider latitudinal bias.

Once the SSS data have been corrected for inter-sensor biases, it is possible to average the data over time in order to obtain a daily product on a sliding window of around 10 days.

The aggregation of SSS is done by weighting them by L2 errors. Several algorithms have been tested:

-averaging on a rectangular window

-average on a Gaussian window

-Bayesian approach (in principle very close to a Gaussian window averaging).

In the case of a rectangular time window, we have:

$$\widetilde{SSS} = \frac{\sum_{i=1}^{n} \frac{SSS_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

a posteriori error is written as follows :

$$\sigma\widetilde{SSS} = \sqrt{\frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}}$$

In the gaussian case, we have:

$$\widetilde{SSS}(t) = \frac{\sum_{i=1}^{n}\frac{\varphi(t-t_i)SSS_i(t_i)}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{\varphi(t-t_i)}{\sigma_i^2}}$$

where $\varphi(t,t_i)$ is a Gaussian filter:

$$\varphi(t) = e^{\left(-\frac{t^2}{\sigma^2}\right)}$$

A posteriori error is written as follows:

$$\sigma\widetilde{SSS}(t) = \frac{\sqrt{\sum_{i=1}^{n}\frac{\varphi^2(t-t_i)}{\sigma_i^2}}}{\sum_{i=1}^{n}\frac{\varphi(t-t_i)}{\sigma_i^2}}$$

The Bayesian case, which is more complicated, involves a matrix inversion:

$$\widetilde{SSS} = SSS0 + Cm.G.H^{-1}(SSS_{L2OS} - SSS0)$$

$$H = G.Cm.G + Cd$$

With G the identity operator, Cm the temporal a priori covariance that is modulated with natural variability and Cd the covariance of L2OS salinities. This approach, which is close to an objective analysis, is described in the papers (Kolodziejczyk et al., 2016, RD02) & (Boutin et al., 2018, RD09).

In any case, a linear operator A can be defined that connects the estimator $\widetilde{SSS}$ to the L2OS data (A is an operator of average):

$$\widetilde{SSS} = A.SSS_{L2OS}$$

The propagation of the error is then written:

$$\sigma_{\widetilde{SSS}} = \sqrt{diag(A.Cd.A^t)}$$

One way to calculate the error is to consider the standard deviation of SSS taken in a time window under the assumption that the SSS does not vary. Dividing the standard deviation obtained by the square root of the number of measurements gives an high estimate of the error on the mean.

Indeed, this standard deviation is an estimator of the average error that affects individual SSS. This type of empirical approach roughly validates the errors actually estimated by the previous equation. Figure 34 shows the errors on the running day averages calculated for the different estimators previously described (simple error-weighted mean, Gaussian filter and Bayesian approach). We see that the errors obtained are very close to each other and also close to the empirical error.
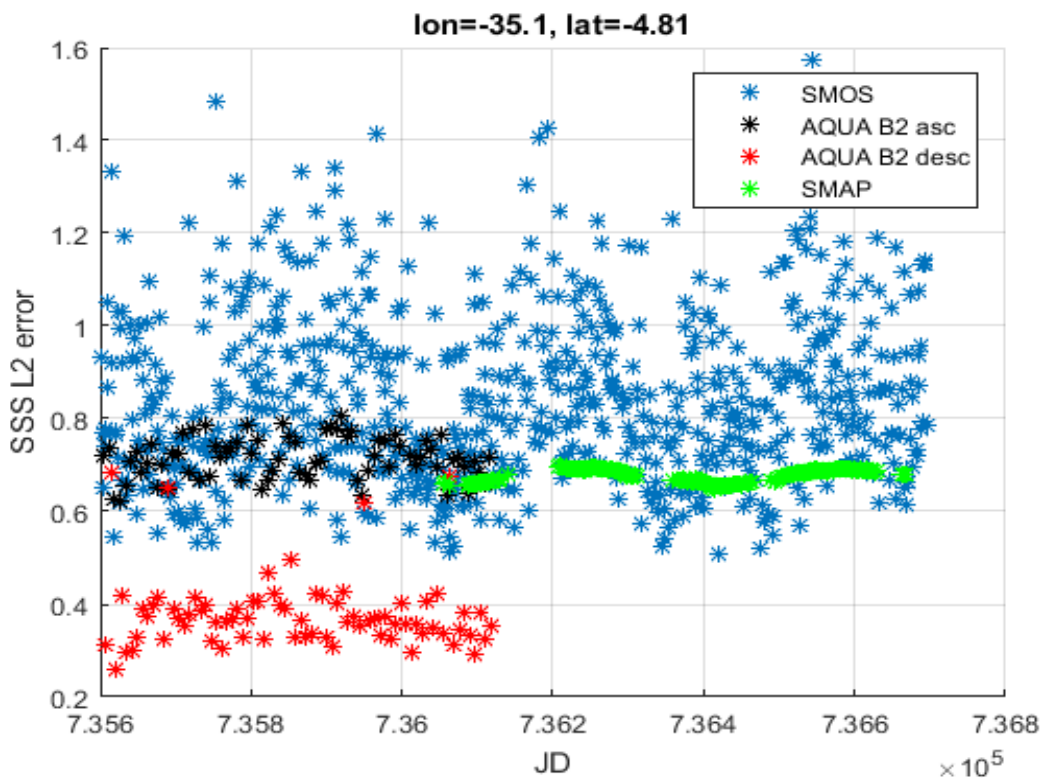

Figure 33 example of the errors of the SSS L2OS (v662) on a grid node near the Amazon plume (time series 2014-2016). In blue, SMOS errors; in black, Aquarius errors (v5.0) ascending orbits, in red, Aquarius descending orbits errors, in green, SMAP errors (v2.0).
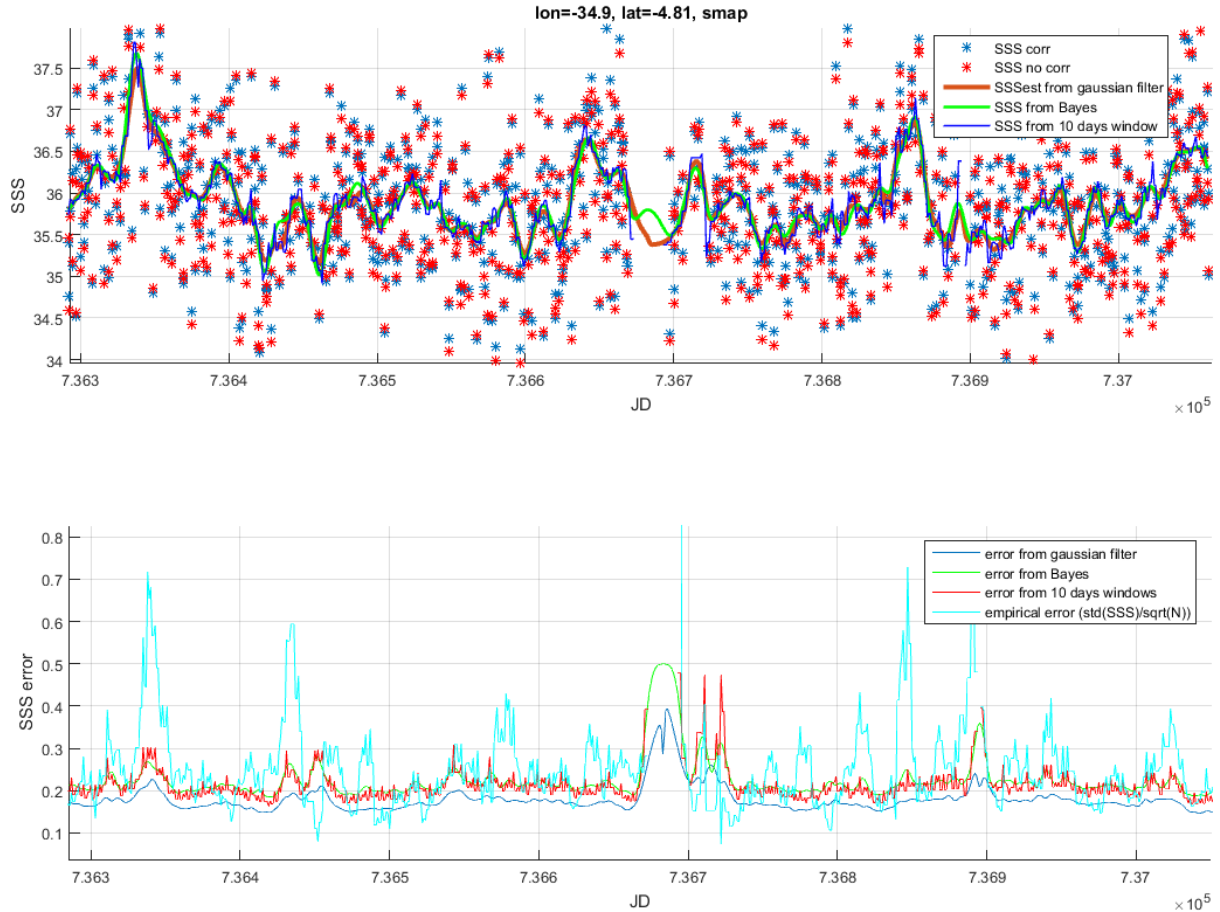
Figure 34: Time series (2016-2018) on a grid node. Use of SMOS, SMAP and Aquarius data. At the top, different averaging solutions obtained with different algorithms: red, Gaussian filter; in green, Bayesian method; in black, average over a 10-day rectangular time window. The blue dots correspond to the SSS corrected for the relative inter-sensor bias, the red dots to the uncorrected SSS. At the bottom, the associated errors.

6 Conclusions and way forward

A relatively precise qualification of the errors was possible by intercomparison of the products coming from the various sensors. This qualification has been performed from level 2 salinity products. It is essential to reach a L2 error as realistic as possible because it makes it possible to obtain by propagation a realistic error on the merged products. Moreover, it allows a relative intercalibration of the instruments at the SSS level which allows making consistency between the measurements coming from the different sensors.

The perspectives are based on user feedback on L4 products, particularly with regard to the estimation of weekly L4 product errors (see ATBD for a description of the propagation algorithms). Indeed, it turns out that these are overestimated. So a peculiar effort shall be made for this error estimation.

End of document