Climate Change Initiative+ (CCI+) Phase 2

Sea Surface Salinity

esa
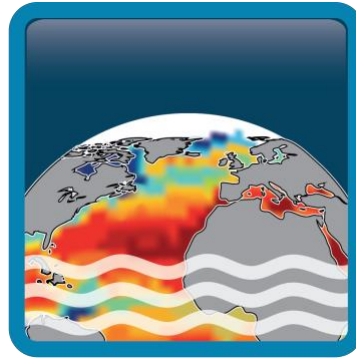
[D2.3]End-to-End ECV Uncertainty Budget (E3UB)

ARGANS

ARGANS · LOCEAN · LOPS · ACRI ST · adwäisEO · OCEANSCOPE · National Oceanography Centre · LEGOS · Met Office · MERCATOR OCEAN INTERNATIONAL

Signatures

Author	Jean Luc Vergely Fabrice Bonjean	*[signature]*	02-05-2023
Reviewed by	Jacqueline Boutin (Science Leader)		
Approved By	Jacqueline Boutin (Science Leader)	*[signature]*	02-05-2023
	Nicolas Reul (Science Leader)		
	Rafael Catany (Project Manager)	*[signature]*	02-05-2023
Accepted by	Susanne Mecklenburg (Technical Officer) ESA		

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

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
08-07-2020 / v2.1	Document update for delivery MS#04 to ESA	Version 2.1 of the document
10-11-2020 / v2.2	Implement ESA's feedback (minor changes)	N/A
13-09-2021 / v3.1 (this document)	Phase 1 final version	All
02-05-2023 / v4.0 (this document)	Phase 2 v4 draft	All

Table of Contents

List of figures

List of tables

1   Introduction

1.1   Scope of this document

This document holds the End-to-End ECV Uncertainty Budget (E3UB) prepared by the CCI+ Salinity team (phase 2, year 1).

At the moment of writing this report, SSS measurements are available from three L-Band radiometer satellite missions, SMOS, Aquarius and SMAP, each with very different instrument features leading to particular measurement characteristics. The Climate Change Initiative Salinity project (CCI+SSS) aimed to produce SSS Climate Data Record (CDR) to include satellite measurements and well-documented uncertainties. In order to ensure a consistent CDR, variations in instruments are closely examined by comparing SSS data and then adjusted using a thorough analysis of the satellite measurements and outside referral sources. This document describes the basis for the uncertainties characterisations.

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 characterisation
- ✓ Section 4: L2 uncertainty characterisation
- ✓ Section 5: L3 and L4 uncertainty budget

L2/L3/L4 data sets will be provided in parts 1 and 2 of the CCI+SSS phase 2 project and are described in the SRD.

This document is the fourth version of the E3UB document addressing Phase 2, part 1 activity.

1.3   References

1.3.1   Applicable Documents

ID	Document	Reference
AD01	CCI+ Statement of Work	SoW
AD02	Product User Guide (PUG)	SSS_cci-D4.3-PUG-v3.0
AD03	User Requirement Document (URD)	SSS_cci-D1.1-URD-v3
AD04	Product Specification Document (PSD)	SSS_cci-D1.2-PSD-v3
AD05	Algorithm Theoretical Baseline Document	SSS_cci-D2.3-ATBD_L3_L4-

ID	Document	Reference
AD06	End-to-End ECV Uncertainty Budget (E3UB), version 1	SSS_cci-D2.3-E3UB_v3
AD07	Algorithm Theoretical Baseline Document, version 2	SSS_cci-D2.3-ATBD_v4
AD08	End-to-End ECV Uncertainty Budget (E3UB), version 2	SSS_cci-D2.3-ATBD_L3_L4-i1r0_v2
AD09	Algorithm Theoretical Baseline Document, version 3	SSS_cci-D2.3-ATBD_L3_L4-i1r0_v3.0

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
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
RD11	Dinnat, E., D. Le Vine, J. Boutin, T. Meissner, and G. Lagerloef. 2019. "Remote Sensing of Sea Surface Salinity: Comparison of Satellite and In Situ Observations and Impact of Retrieval Parameters." Remote Sensing, **11 (7):** 750	
RD12	C. J. Merchant et al. (2017): Uncertainty information in climate data records from Earth observation, Earth Syst. Sci. Data, 9, 511-527	

ID	Document	Reference
RD13	Brodzik, M. J., B. Billingsley, T. Haran, B. Raup, and M. H. Savoie (2012), EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets, *ISPRS International Journal of Geo-Information*, *1*(1), 32-45.	
RD14	Le Vine, D. M., and P. de Matthaeis (2014), Aquarius active/passive RFI environment at L-band, IEEE Geosci. Remote Sens. Lett., 11(10), doi:10.1109/LGRS.2014.2307794.	
RD15	Drushka, K., Asher, W. E., Ward, B., and Walesby, K. (2016), Understanding the formation and evolution of rain-formed fresh lenses at the ocean surface, J. Geophys. Res. Oceans, 121, 2673– 2689, doi:10.1002/2015JC011527.	
RD16	Supply, A., J. Boutin, G. Reverdin, J.-L. Vergely, and H. Bellenger, 2020: Variability of satellite sea surface salinity under rainfall. In: Satellite Precipitation Measurement, V. Levizzani, C. Kidd., D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, F. J. Turk, Eds., Springer Nature, Cham, Advances in Global Change Research, 69, 1155-1176, https://doi.org/10.1007/978-3-030-35798-6_34.	
RD17	Boutin, J., J.L. Vergely, F. Bonjean, X. Perrot, Y. Zhou, E. Dinnat, R. Lang, D. Levine, and R. Sabia, New Seawater Dielectric Constant Parametrization and Application to SMOS Retrieved Salinity, IEEE Transactions of Geoscience and Remote Sensing, 2023, doi : 10.1109/TGRS.2023.3257923.	

1.4 Acronyms

AD	Applicable document
ADB	Actions database
AMOC	Atlantic Meridional Overturning Circulation
ATBD	Algorithm theoretical basis documents
BRO	Brochure
CliC	Climate and Cryosphere
DIR	Directory
DS	Dataset availability
DS-UM	Dataset user manual
DVP	Development and validation plan
EC RTD	European Commission Directorate General for Research and Innovation
EDS	Experimental dataset
EMI	Electromagnetic Interference
EO	Earth Observation
EOEP	Earth Observation Envelope Program
ESA	European Space Agency
FR	Final review
FWF	Freshwater fluxes
GCOS	Global Climate Observing System
IAR	Impact assessment report
ITT	Invitation to tender
IPP	Year of Polar Prediction
KO	Kick-off
MR	Monthly report
MTR	Mid-term review
MV-TN	Modelling and validation technical note
NDVI	Normalised Difference Vegetation Index
OTT	Ocean Target Transform
PAR	Preliminary analysis report
PGICs	Peripheral glaciers and ice caps
PM	Progress meeting
PMP	Project management plan
RD	Reference document
RB	Requirements baseline
SAR	Synthetic Aperture Radar
SIAR	Scientific and impact assessment report

SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
SoW	Statement of work
SR	Scientific roadmap
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
TDP	Technical data package
TDS	Training Data Set
TN	Technical note
VIR	Validation and intercomparison report
VR	Validation report
WCRP	World Climate Research Programme
WP	Work package
WS	Wind Speed
WWRP	World Weather Research Programme
AD	Applicable document
ADB	Actions database
AMOC	Atlantic Meridional Overturning Circulation
ATBD	Algorithm theoretical basis documents
BRO	Brochure
BV	Boutin-Vergely dielectric constant model
BVZ	Boutin-Vergely-Zhou dielectric constant model
CliC	Climate and Cryosphere
DIR	Directory
DS	Dataset availability
DS-UM	Dataset user manual
DVP	Development and validation plan
EC RTD	European Commission Directorate General for Research and Innovation
EDS	Experimental dataset
EMI	Electromagnetic Interference
EO	Earth Observation
EOEP	Earth Observation Envelope Program
ESA	European Space Agency
FR	Final review
FWF	Freshwater fluxes
GCOS	Global Climate Observing System
IAR	Impact assessment report

ITT	Invitation to tender
IPP	Year of Polar Prediction
KO	Kick-off
KS	Klein and Swift dielectric constant model
MR	Monthly report
MTR	Mid-term review
MV-TN	Modelling and validation technical note
NDVI	Normalised Difference Vegetation Index
PAR	Preliminary analysis report
RFI	Radio-Frequency Interference
TB	Brightness Temperature

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 a centre 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 polarisations

- ✓ 3-dB (half power) footprint size: between ~32 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 a centre 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 polarisations

- ✓ 1000 km wide swath.

- ✓ 3-dB (half power) footprint size: ~43 km.

✓ Global coverage: 3 days

2.4 Aquarius

The main Aquarius characteristics are:

- ✓ Radiometer (3 beams) with a centre 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, TV and third Stokes

- ✓ Aligned with a scatterometer (1.26 GHz), both instruments are polarimetric.

3 L1 uncertainty characterisation

3.1 Introduction

In order to better understand the SSS uncertainties at L4, we need to have a global understanding of what happens at L0-L1 in terms of the TB uncertainty budget. L0-L1 processings are very complex and require specific expertise. Indeed, reviewing all L0-L1 processing methods for all satellite missions is outside our scope. Our purpose is to replay L2 from existing L1 products. Therefore, it is not our objective to re-estimate TB uncertainties but to take what comes out of the L0-L1 products and work with them, as long as we have sufficient information to propagate TB uncertainties toward level 2 properly. TBs at the L1 processing output have systematic uncertainties that are not corrected at L0-L1 and that we take into account in the higher levels empirically.

Hence, we develop methods for correcting systematic uncertainties and estimating random uncertainties at the SSS level from the remaining differences between forward model predictions and TB data from SMOS, Aquarius or SMAP. Seasonal patterns of these differences can vary from ascending to descending passes. This is mainly due to uncertainties in the thermal model and instrument monitoring, inaccurate Radiative Transfer Models (RTMs), and input auxiliary EO data.

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 TB in the frequency space domain to the physical space domain. This different processing requires 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 the physical space domain. In the following this operation is called reconstruction

Complex calibrations for thermal drifts based on Noise Injection Radiometer data and several onboard thermistor measurements are used to calibrate the visibilities. Short-term calibration is regularly performed raw to compensate for high variability drifts. In addition, cold-sky calibration is performed several times a year when the satellite sensor is rotated sky upward during dedicated manoeuvres (used for the so-called Flat Target Transformation). However, systematic and seasonal image reconstruction uncertainties are still found in the reconstructed level 1 data despite raw data calibration. This can occur, for example, due to the instrument's 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 uncertainties (noise floor, aliasing, instrument impulse response function, antenna pattern uncertainties).

To compensate for these distortions in the image, a vicarious calibration is performed. This involves comparing the SMOS antenna TBs with a radiative transfer forward model of the

brightness obtained from specific orbits in the middle of the Pacific and evaluating the mean spatial difference in the antenna coordinate frame. 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 uncertainties on TBs against the expected radiometric noise. These uncertainties result from filtering out various unmodeled contributions, such as RFI and sun effect, which align with the expected radiometric noise. On the other hand, there are reconstruction biases that are not corrected by the OTT, which then generates biases in the estimated parameters. For now, L1 processing does not make it possible to avoid such biases.

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.

For SMAP (it is more complicated than Aquarius: the SMAP antenna has some non-negligible emissivity), the antenna temperature predicted from the thermal model has some uncertainties, and a latitudinal correction has to be applied.

In addition, uncertainties in the modelled side-lobes of the radiometer antenna patterns used for 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 coastlines 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 surfaces (from sea to land, for instance).

3.4   RFI filtering

SMOS, Aquarius, and SMAP missions operate in the L-band protected spectrum (1400-1427 MHz), which 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. Several strategies were implemented to filter data from RFI contaminated measurements to alleviate this situation. As SMOS, launched in 2009, was the first satellite to operate in L-band, it does not have any onboard hardware/software to filter RFI, so RFI filtering/mitigation relies only on data post-acquisition processing. This issue is significantly less critical for SMAP (and to the least extent for Aquarius), as they are (were) equipped with onboard 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 a 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 prominent tails spreading from the RFI source. In the case of SMAP and Aquarius, the RFI contamination is different as they operate real-aperture radiometers and onboard 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 works well over the ocean with modest rates for false alarms (5%) and missed detection. However, RFI is still detected in Aquarius (Le Vine and De Matthaeis, 2014). SMAP takes a multidomain approach to RFI mitigation by utilising an innovative onboard digital detector back end with digital signal processing algorithms to characterise the received signals' time, frequency, polarization, and statistical properties. Almost 1000 times more measurements than conventionally necessary are collected to enable the ground processing algorithm to detect and remove interferences.

4   L2OS uncertainty characterisation

4.1   Introduction

SSS random uncertainties are taken directly from L2 products (L2 random theoretical uncertainties), estimated from self-consistency analysis, or by adjusting global metrics of differences between observed SSS with external data (ISAS, Argo). Random and systematic uncertainties can be obtained in a relative way by comparing averaged products from different sensors and orbit types (ascending or descending).

L2 SSS random uncertainty is first derived for open ocean data. In the second step, their estimate is updated in case of land contamination. We extract from L2 data a multiplicative factor to be applied to the SSS random uncertainties. This factor allows taking into account secondary side lobe effects or reconstruction effects affecting TB measurements close to the coast.

In the following, we present the generic methods in section 4.2 and the uncertainty estimates obtained with version 4 of the CCI+SSS processing in section 4.3.

4.2   Methods

4.2.1   Random uncertainty

4.2.1.1   Introduction

In this section, we present three different methods which allow estimating SSS random uncertainties:

- by error propagation
- by comparing measured SSS with a reference (affected by a neglectable error)
- by self-consistency analysis

4.2.1.2   Random uncertainty propagation

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

Level 2 algorithms are used to propagate the TB noise characterised by the radiometric accuracy and the uncertainty of all geophysical parameters (wind speed, surface temperature, etc.) on the salinity. The propagation methods generally assume a Gaussian statistic, a linearisation of the forward model near the solution and least square type retrieval. The theoretical uncertainty obtained depends on the a priori uncertainties on the parameters. In order to homogenise L2 uncertainties from the different sensors (SMOS, SMAP, Aquarius), a review of uncertainty propagation methods has been conducted. We propose a strategy to standardise the uncertainty calculation and the a priori uncertainties assigned to the geophysical parameters. This strategy

of standardising the uncertainty calculation must be in phase with the standardisation of the auxiliary data and the uniformisation of the direct and inverse models on the set of sensors. Note that for the time being, model uncertainties (sun glint, roughness, atmospheric, galactic, dielectric models) are not propagated in L2 SSS. After that, when computing L3 products by combining different L2 SSS, it is possible to use the L2 uncertainty to weight properly during the average.

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

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

This yields, according to a $\frac{\partial \text{TB}}{\partial \text{SSS}}$ (SST) approximation for Stokes1 of $-0.015 . SST - 0.25$

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

SSS uncertainties due to WS and SST uncertainties could be added quadratically to this relation after propagation.

4.2.1.3 Random uncertainty from the external data comparison

The uncertainty balance obtained by uncertainty propagation does not generally include the uncertainties on the models themselves (model of galactic noise, roughness, solar contamination, etc.). In order to estimate the errors and validate the uncertainties obtained by propagation, it is necessary to compare the satellite SSS data with external information (e.g. in-situ measurements). This can be done directly with L2 SSS. We expect the random uncertainties to decrease to zero when averaged over a large space. This is the property we use to validate the random uncertainties estimate and to discriminate it from systematic uncertainties. It is then preferable to average the L2 data before comparison in order to reduce the random uncertainty and to have an estimation of the systematic uncertainty. Hence, the external data is compared on L2 and L3 products. This approach has been performed for each sensor.

4.2.1.4 Random uncertainty from self-consistency analysis

The three sensors provide independent measurements. Three comparisons are made, depending on the period: SMOS-Aquarius over 2012-2015, SMOS-SMAP over the period 2015-2020 and SMOS-SMAP-Aquarius over April to June 2015. Note that if we standardise the direct and auxiliary data, the random and systematic uncertainties on the SSS data will not be completely

independent. On the other hand, this could make it possible to qualify uncertainties related to the unexpected behaviour of the instruments (sensor drift, issues relating to image reconstruction, contamination by RFI, etc.).

4.2.1.5 Qualitative estimation of random uncertainty and identification of outliers

In some cases, we know that estimating the random uncertainties at L2 is inaccurate (primarily if one identifies convergence problems in retrieval algorithms). In this case, it is essential to identify and flag the SSS outliers accordingly. In addition, some statistical indicators for TB residues may show that the uncertainty obtained by propagation is underestimated. It is then possible to empirically re-evaluate the SSS random uncertainties upwards.

4.2.2 Systematic uncertainty

4.2.2.1 Introduction

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

4.2.2.2 Relative systematic uncertainties

The solution to this problem is not to estimate absolute SSS but to analyse salinity anomalies. This approach has been applied to SMOS and has yielded excellent results (Boutin et al., 2016). In particular, specific algorithms allow for correcting the relative across-track systematic uncertainties (Kolodziejczyk et al. 2016) and ascending-descending latitudinal biases (Boutin et al. 2018). This type of algorithm has been extended to the other sensors in previous CCI+SSS CRDP versions (Boutin et al. 2021) and it is thus possible to estimate the relative biases for all the sensors. We recall below the principle of the method and the updates performed in CCI+SSS version 4.

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

In order to characterise the bias, we can distinguish two types of bias:

- a land-sea contamination bias independent of time-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 constant.

- a seasonal latitudinal bias that depends on sun and galactic noise contamination and possibly on other instrumental drifts considered here periodic over one year. In CCI+SSS v3, this correction is applied only on SMOS data but on other sensors in version 4.

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:

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

With bc, coastal bias and bl latitudinal bias. SSS_obs is the observed salinity, and SSS(t) corresponds to the unbiased SSS. 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 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. As mentioned, bl is considered for Aquarius and SMAP in CCI+SSS v4 (but not in v3). For SMOS, a time-dependent correction is added in some RFI regions (see 4.2.2.5).

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 self-referenced, i.e. there is no need for an external reference when considering anomalies and not absolute salinities. Note that Eqn 4-2 requires 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 critical point because, in this situation, we estimate the L4 products represented by SSS(t) simultaneously as we characterise the biases. The uncertainty propagation occurs at the time of this estimate. Because 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 for SMOS.

An improvement of this correction has been proposed, particularly regarding the inclusion of SST. This approach remains valid for all L-band sensors.

In principle, correcting the inter-dwell or latitudinal instrument bias does not depend on geophysical conditions. However, suppose the brightness temperature bias (ΔTB) is generally independent of geophysical conditions, especially the sea surface temperature (SST). In that case, this is not the case for the SSS bias (ΔSSS) ([RD09]). Indeed, the sensitivity of the retrieval (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:

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

The lower the sensitivity $\frac{\partial TB}{\partial SSS}(SST)$, the greater the bias on the SSS for a constant TB bias. This behaviour does not simplify the management of bias correction in SSS since, at a given point, SST can vary significantly 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:

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

$$\equiv \Delta SSS(SST = SST0).\ \mathrm{coeff_{SST0}}(SST1)$$

$\mathrm{coeff_{SST0}}(SST1)$ represents the multiplicative coefficient to calculate the bias when observed at a different SST. So the idea is to compute the bias for a given SST, that is, to reduce each measure to an average SST that does not necessarily correspond to the SST observed at the time of the acquisition.

4.2.2.3 Absolute systematic uncertainties

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

The absolute systematic uncertainty calculation requires correction based on climatology or in-situ data. At the SSS level, the resolution of Eqn 4-2 gives L4 SSS anomalies. At this level, it is possible to add these SSS anomalies to a constant shift to reach an absolute SSS.

4.2.2.4 Systematic uncertainties due to TB modelling

Framework

We recall that at Level 2 for a given grid point, the SSS retrieval from satellite Tb measurements (L1c) uses a direct (forward) model, a set of auxiliary data from various variables, the most important ones being SST and wind speed, used either as a first guess (in case of an iterative retrieval process as in SMOS L2OS processor) or as true value to estimate various contributions to be corrected at Tb level before retrieving SSS (as in SMAP and Aquarius RSS processing).

Theoretical SSS variations corresponding to changes

Based on the above, we hereafter present the simplified approach that we have developed for estimating the SSS variations induced, for example, by changes in the direct model, OTT calibration, and prior parameters. Such changes have been performed in the SMOS CCI v3/L2OS v671 relative to SMOS/ESA v662 L2OS operational processor. Hence, the SSS differences observed between the two versions can be considered a reference for validating the correction tool.

1/ Change in direct models and/or auxiliary parameters:

They affect Tb and Tb calibration (via the OTT calculation in the case of SMOS). Hence, in the case of SMOS (similar reasoning could be considered for other instruments), the reverted effect of changes on retrieved SSS can be written as:

$$\Delta SSS = -1 \Big/ \frac{\partial Tb}{\partial SSS} \times (\Delta Tb - \widetilde{\Delta Tb_{OTT}})$$

Where ΔSSS is the expected SMOS SSS change at a given grid point, $\frac{\partial Tb}{\partial SSS}$ is the partial derivative of Tb with respect to SSS at the considered SST, ΔTb is the expected Tb change coming from changes in direct models and auxiliary parameters, $\widetilde{\Delta Tb_{OTT}}$ is the median of ΔTb in the OTT region, and the removal of $\widetilde{\Delta Tb_{OTT}}$ mimics the OTT calibration. In our SMOS-CCI example, ΔTb corresponding to the various changes have been estimated as follows:

- Changes due to changes in prior parameters (SST and Wind Speed)

Variations of Tb are given by differentiation of the forward model:

$$\Delta Tb = \frac{\partial Tb}{\partial VAR} \times (VAR_{new} - VAR_{old})$$

Where VAR_{new} corresponds to ERA5 SST or WS, and VAR_{old} to ECMWF forecasts.

- Change of dielectric constant model.

It was switched from KS (Klein and Swift, 1977) to Boutin et al., 2020 (BV).

$$\Delta Tb = Tb_{BV} - Tb_{KS}$$

or from Boutin et al., 2020 (BV) to Boutin et al., 2023 (RD17).

$$\Delta Tb = Tb_{BVZ} - Tb_{BV}$$

2/ Change in OTT adjustment

In L2OS v671, the reference SSS dataset was switched from WOA to ISAS, and the corresponding impact on SSS can be expressed as:

$$\Delta SSS = \frac{\partial SSS}{\partial Tb} \times median \left(\frac{\partial Tb}{\partial SSS} \times \Delta SSS_{OTT} \right)$$

With $\Delta SSS_{OTT} = SSS_{ISAS} - SSS_{WOA}$, the median is estimated over the OTT area.

4.2.2.5 Systematic uncertainties due to RFI

This section presents a method and algorithm developed to mitigate the impact of RFI on SMOS SSS retrievals. RFI sources have intermittently affected SMOS SSS retrievals since the mission's launch in 2010. Our approach focuses on recovering the SSS signal in L2 orbits where intense RFI pollution occurs. The method and algorithm are detailed, including a prototype example of their application.

4.2.2.6 General approach

The impact of RFI pollution on previous SMOS SSS products is evident. Strong coherent RFI signatures in time and space can be effectively extracted from L3/4 datasets around major RFI sources. This can be achieved through principal component analysis (PCA) conjugated with high-pass spatial frequency filtering. If the first PCA mode is dominant (e.g. percentage of explained variance larger than 70%), the RFI signal can be approximately reconstructed from the product of a time series s(t) and a spatial field f(x, y). The time series characterises the timing of the RFI and may resemble a step function activating or deactivating in the middle of the SMOS period. The spatial field characterises the peculiar star-like shape of the RFI perturbations viewed from SMOS. From the analysis of L3 products, we identified useful properties to propose a correction method for orbital L2 SMOS data.

Separating the geophysical signal from the pollution perturbation is essential to correct RFI in SSS data. Our approach is partitioning L2 SMOS orbital data into dwell lines along each orbit across the swath. Each dwell line contains the true SSS signal but is affected differently by RFI and other sources of uncertainty and noise. While constructing a strict dwell difference at any precise time and location is impossible, we overcome this difficulty by building moving average fields of the different dwell lines. By subtracting these aggregated fields from each other, we can effectively characterise the RFI perturbation. We chose an 11-month window width, which is approximately twice the strict repeat cycle of SMOS (149 days). This helps to filter out seasonality and provides temporal variations throughout the entire period for most grid points within a given region. This is necessary for the application of the PCA method mentioned in the previous paragraph.

We will provide a detailed description of the method in the following section.

4.2.2.7 Method and algorithm

The algorithm has two objectives:

1. Extract characteristic time series of RFI contamination in a specific region around a known source.

2. Apply a correction scheme to L2 SMOS data based on the results from step 1.

To achieve these objectives, we implement the following steps.

    A. Dwell-line fields

Climate Change Initiative+ (CCI+) Phase 2

End-to-End ECV Uncertainty Budget

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

Date: 21/06/2023

Version: v4.0

Page: 31 of 72

We divided the SMOS ascending (A) and descending (D) half-orbits into intervals of 50km-averaged dwell lines across the swath. Then, we resampled those intervals using widths of 25km. The intervals are defined by their distance in km from the swath centre line, ranging from -487.5km to +487.5km. The L2 SMOS SSS within each of these intervals is then projected onto the global 0.25° grid and agglomerated in monthly windows. This results in the construction of 78 (2*39) dwell-line fields for all months during 2010-2022. Moving forward, we will refer to these intervals as "dwell-lines" or simply "dwell" for ease of understanding.

Each dwell-line field is formally labelled as $S^1(d_k, x, y, t)$, where S is SSS, «1» stands for «1 month agglomeration», d_k is the kth dwell-line (k=1,...,78), x and y are longitude and latitude, and t is time.

B. 11-month dwell fields

We construct a moving average for each dwell line. Specifically, we calculate the 11-month moving average for the entire period at each data point within the global 0.25° grid. This moving average is denoted as $S^{11}(d_k, x, y, t)$.

We consider the gaps within the 11-month window to calculate the moving average. Thus, at any given point x and y for a given dwell, the average is calculated regardless of any gaps in the data over the 11 months.

We write:

$$S^{11}(d_k, x, y, t) = S^{11}_{true}(x, y, t) + E_{RFI}(d_k, x, y, t) + \varepsilon(d_k, x, y, t)$$

Where S^{11}_{true} represents the actual SSS signal, which is subjected to an 11-month moving average, E_{RFI} is the RFI perturbation within that 11-month window configuration, and ε the residual, including other uncertainty sources and noise.

C. 11-month dwell differences

The difference relative to 2 different dwell lines, i and j, is expressed as:

$$\Delta_{i,j}S^{11}(x, y, t) \equiv S^{11}(d_i, x, y, t) - S^{11}(d_j, x, y, t) \approx \Delta_{i,j}E_{RFI}(x, y, t)$$

Where we assume that we can neglect the residual contribution.

For a given region $\mathcal{R}$ centred around an RFI source, typically within 500km, we test 1) if $\Delta_{i,j}S^{11}(x, y, t)$ effectively characterise some of the RFI perturbation, and 2) if, from the ensemble of $\Delta_{i,j}S^{11}(x, y, t)$ for all distinct i and j, we can extract the RFI characteristic time series relative to $\mathcal{R}$.

The step 2) is achieved through temporal PCA of the $\Delta_{I,j}S^{11}(x, y, t)$ ensemble.

D. Anomalies with respect to time mean.

Our RFI correction procedure is specifically designed to address time varying RFI perturbations and does not account for the steady RFI signal that persists throughout the period. The treatment of the steady bias is deferred to a later stage of the L3 processing. This section focuses solely on anomalies concerning the time mean calculated over the entire period.

We use the following «dot» notation for any variable A to which we subtract the time mean:

$$\dot{A} = A - \frac{1}{N_t}\sum_t A$$

E. Dwell difference PCA

The objective of this step is to extract the first few modes of dwell difference PCA and verify that the first mode is dominant. By dominant, we mean that the first mode is well separated from the higher modes, typically, if the percentage of variance explained by the first mode is at least larger than 70%. If this is the case, we can write the following approximation,

$$\Delta_{i,j}\dot{S}^{11}(x,y,t) = \gamma_0 u_0^{11}(t) v_0(i,j,x,y) + \cdots$$

by omitting the smaller variance terms.

The scalar and functions γ_0, $u_0^{11}(t)$ and $v_0(i,j,x,y)$ are yielded by PCA, that is, through a singular value decomposition (SVD) applied to the matrix composed of time series for each i, j (i ≠ j), x and y, and in which time series are arranged in columns. We keep the superscript «11» in $u_0^{11}(t)$ to recall that this time series is relative to the 11-month moving average configuration. By construction, $u_0^{11}(t)$ is centred and normalised.

To reduce the number of variables we need to treat, we can take advantage of the fundamental property of PCA to reduce any matrix according to its rank, eliminating linearly dependent variables or columns. Instead of treating all 3003 (78*77/2) dwell differences, we can create a matrix for PCA using only 77 non-dependent variables, represented by the (i,j) index couples (1,2), (1,3), (1,4),...,(1,78). This approach allows us to exclude any other difference, such as (4,7), which is linearly dependent on the 77 variables, such as (1,7)-(1,4).

Note that the sign of variables (or columns) in PCA is unimportant, as PCA searches for the optimum linear combination of columns that maximises its variance.

Therefore, we can extract a time series $u_0^{11}(t)$ characteristic of the RFI source relative to the 11-month moving average pre-treatment of the dwell dataset.

In the following paragraphs, we will describe the correction procedure. The first step is to generate the monthly RFI-characteristic time series and their corresponding spatial patterns for all the various dwells. Next, we apply the correction scheme to the L2 SMOS data using these elements. To achieve this, we implement the following steps hereafter.

We consider an extended region around the initially selected region $\mathcal{R}$ from the previous steps, which we refer to as Ω. Ω is our correction domain, and from here on $(x,y) \in \Omega$.

F.   First regression: the spatial structures of correction per dwell.

To calculate the spatial structures of correction per dwell, we project the 11-month moving average dwell fields onto the RFI characteristic time series $u_0^{11}(t)$. Note that we are now working with the dwell fields and no longer with their differences. We can express this as :

$$V_0(d_k, x, y) = \sum_t \dot{S}^{11}(d_k, x, y, t) \times u_0^{11}(t)$$

Note that we can speak indifferently of projection and regression since $u_0^{11}(t)$ is centred and normalised.

We now normalise the result:

$$V_0'(d_k, x, y) = V_0(d_k, x, y) \Big/ \sqrt{\sum_{k,x,y} V_0(d_k, x, y)^2}$$

Hence, $V_0'(d_k, x, y)$ is a spatial field per dwell that characterises the RFI spatial structures for each dwell d_k. This is the first corrective element used in the final correction scheme.

G.   Second regression: the monthly RFI characteristic time series

Until now, we have focused on analysing the 11-month moving average configuration of the dwell dataset. However, if we want to make corrections to the L2 SMOS dataset, we must establish a connection with the monthly agglomeration of the dwell orbital data. To achieve this, we use a secondary regression method, whereby we project the monthly dwell-line agglomerates onto the RFI spatial structures. This allows us to calculate the monthly RFI characteristic time series.

This is given by:

$$U_0^1(t) = \sum_{(k,x,y)\in\Gamma(t)} \dot{S}^1(d_k, x, y, t) \times V_0'(d_k, x, y) \Big/ \sum_{(k,x,y)\in\Gamma(t)} V_0'(d_k, x, y)^2$$

The summation $\sum_{(k,x,y)\in\Gamma(t)}$ is processed for all dwell-lines k and all x, y within the $\Gamma(t)$ set that includes only dwell and locations for which the data exists at time t.

Note that if there were no missing values, the denominator in this formulation would be 1.

Therefore, $U_0^1(t)$ is the RFI characteristic time series relative to the monthly agglomerates of dwell lines and thus constitutes the second and final corrective element associated with $V_0'(d_k, x, y)$.

H.   Construction of the correction to be applied to the ensemble of dwell-line intervals.

From the two corrective elements previously calculated, the corrected SMOS SSS anomaly for a particular dwell-line k, at location x, y and time t is given by:

$$\dot{S}_{corr}(d_k, x, y, t) = \dot{S}(d_k, x, y, t) - V_0'(d_k, x, y) \times U_0^1(t)$$

(This is a formulation that applies to all L2 SMOS data, irrespective of how they are agglomerated, and thus we omit the superscript «1»)

4.3 Uncertainty estimates of SMOS, SMAP and Aquarius SSS.

4.3.1 Introduction

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

The quality of the products is assessed by the development team (round-robin tests) and the validation team, which includes the PI-MEP.

We present in this section the estimates obtained in CCI+SSS version 4 products.

4.3.2 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. In this case, the representativity uncertainties are significant in the comparison mechanism. This representativity uncertainty could be calculated based on high spatial resolution models.

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

The Soil Moisture and Ocean Salinity (SMOS) mission was launched on 2nd November 2009 as the second Earth Explorer Opportunity mission within ESA's Living Planet Programme. It has continuously provided brightness temperature data in L-Band since January 2010, which are used to retrieve Soil Moisture (SM) and Sea Surface Salinity (SSS) data over land and ocean, respectively. This project, funded by ESA, aims at setting up a Pilot Mission Exploitation Platform (Pi-MEP), focussing on ESA's SMOS mission and supporting enhanced validation and scientific process studies over the ocean.

Pi-MEP project objectives:

- 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 and the two other NASA missions. This will enable the following:

1. The user to choose which satellite SSS product is best adapted for their 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 excellent or degraded quality.
3. To converge towards the best approaches and generate less but better satellite SSS products.

A large ensemble **of in situ SSS data distributed by different data centres** 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, PIRATA, 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 is very useful in estimating the quality of the L2OS and L3OS products.

4.3.2.2   ISAS

To cover the entire data record length, we use various ISAS products by order of preference:

-the delayed mode ISAS products created by LOPS laboratory, ISAS17 (reprocessed to avoid drawbacks linked to WOA climatology in the Arctic Ocean, N. Kolodziejczyk, pers. comm.), which includes detailed quality-controlled Argo profiles and other in situ measurements (ships of opportunity, research ships, sailing ships, surface drifters, marine mammals).

-For periods when the LOPS product is not available, the delayed mode ISAS product available on Copernicus Marine Environment Service, derived from the objective analysis of different sources of in situ data, mainly Argo floats: http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=INSITU_GLO_TS_OA_REP_OBSERVATIONS_013_002_b

-For periods when none of these products is available, we use the NRT products available on: http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=INSITU_GLO_TS_OA_NRT_OBSERVATIONS_013_002_a.

4.3.3 Data analysis

4.3.3.1 Introduction

Data analysis is carried out based on the self-consistency of SSS and comparison with ISAS SSS.

A first understanding of the L2 SSS uncertainties comes from each sensor's ATBDs and validation reports.

Random uncertainties must be well known in order to be able to properly weigh the different SSSs when developing L3 and L4 products. We propose to validate the theoretical uncertainties provided in the products or to empirically estimate the uncertainties that affect the SSS estimator (in the case of SMAP L2C products and Aquarius L3 products, such uncertainties are not provided). The theoretical uncertainty (which assumes that the direct model and instrumental response are known) depends on radiometric measurement errors and the sensitivity of TBs to SSS. This sensitivity depends on the sea surface temperature SST. We will therefore characterise the theoretical uncertainty according to SST.

Concerning systematic uncertainties, several causes generate them:

-the instrument, 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 uncertainties result from limited knowledge of the signal and sensor.

The various sensors currently have their own correction strategies at brightness temperature levels.

However, as we will show, there are residual biases (also called systematic uncertainties here) in salinity. Here, we are trying to build a salinity field from salinities from the different sensors. The aim is to mix SSS as homogeneously as possible and thus 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 the bias empirically and to correct it afterwards.

These biases affect the data differently depending on the across-swath position and the orbit type (ascending or descending). More precisely, we know that glint effects depend on the season and the geometry of observation. A solar glint and a galactic glint 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 instance, even if the back lobe gains are very low, a source like the sun can significantly the total signal).

We consider that a constant bias affects the following subsets of measurements:

- for SMOS, the same bias is considered for all SSS coming from the same dwell line and position (lat,lon). The dwell lines are sampled across the track every 25 km. The bias has two components: a time-independent component and a latitudinal seasonal component.

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

SSS from ascending and descending orbits are also differentiated for both sensors.

- we will use the L3 data for Aquarius, ascending and descending orbits separately.

Therefore, we have several datasets from the different sensors, and it is necessary to solve Eqn 4-2 to estimate the biases bc and bl for SMOS. As underlined in section 4.2.2.2, the biases are estimated simultaneously with the SSS optimal interpolation.

4.3.3.2 SMOS L2OS data

During phase 1, particular attention was paid to studying random uncertainties. The theoretical L2 errors multiplied by the chi give a reasonable estimate of the true error far from the coast (section 4.3.3.2.1.1). We have kept the algorithm which increases the theoretical error according to the distance from the coast as developed in phase 1 and presented in section 4.3.3.2.1.2.

In phase 2, we sought to improve the corrections to salinities in order to reduce systematic uncertainties (4.3.3.2.2). For this, we worked with three different salinity estimators (SSSuncorr, SSScorr and SSSacard). The two main salinity estimators provided by the DPGS L2OS are SSSuncorr and SSScorr. SSSuncorr is obtained after the correction of the TBs by the OTT. SSScorr is obtained after the correction of the TBs by OTT correction and LSC correction, also at the level of the TBs. We added a bias study from a new SSS estimator, SSSacard, obtained from Acard estimator based on the BVZ dielectric constant model. Acard is obtained from OTT and LSC corrected TBs but without wind speed inversion during the SSS retrieval, unlike SSScorr and SSSuncorr. The fact of not estimating the wind speed during the extraction of Acard makes the latter more robust, especially in strong wind speed regions at low SST.

4.3.3.2.1 Random uncertainty

4.3.3.2.1.1 Random uncertainty propagated in the L2OS processor

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

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

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

The theoretical a posteriori variance (uncertainty) $\sigma_{P_i}^2$ 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:

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

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 uncertainty. Both are considered in the antenna reference frame. The radiometric uncertainty $\sigma_{A_{meas\_n}}^2$ is already given in the antenna reference frame. The model uncertainty $\sigma_{A_{model\_n}}^2$ is given in the ground reference frame and propagated to the 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 critical for several applications, 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 on 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 the incident angle and, therefore, has a cross-track dependency/variation and dwell line dependency. In the first

approach, we took the model uncertainty $\sigma^2_{A_{model\_n}}$ to be constant and equal to 0.5 K for H and V polarisation 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 a 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.

However, the theoretical uncertainty, as described above, is independent on real data; it only relies on the expected characteristics of the measurements and retrieval. To get an uncertainty estimate closer to the real observations, a first improvement has been made by multiplying SSS a posteriori uncertainty by χ, which is the normalised square root of the cost function χ^2 after convergence.

4.3.3.2.1.2   Random uncertainty estimated by using an empirical approach

In order to estimate the actual uncertainties 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 is about four 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 uncertainty. However, the uncertainty depends on the position of the measurement according to the centre of the swath. In order to normalise the data, we create a reduced dimensionless centred variable X (normalising by the theoretical uncertainty a posteriori σ_{SSS} multiplied by the χ):

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

If the theoretical uncertainty is realistic, this new random variable should follow a Gaussian distribution of mean 0 and standard deviation one. If the theoretical uncertainty is underestimated, the std of this new variable is greater than 1. In this way, it is possible to homogenise SSS affected by different theoretical uncertainties.

The histograms of the reduced centred variable are presented in Figure 1, 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 the theoretical uncertainty weighted by the χ gives results very close to the error for the open sea. For the coast, there are still high uncorrected contaminations. They are dominant near 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 and river plumes). In this case, the stability assumption of the SSS is not ensured.
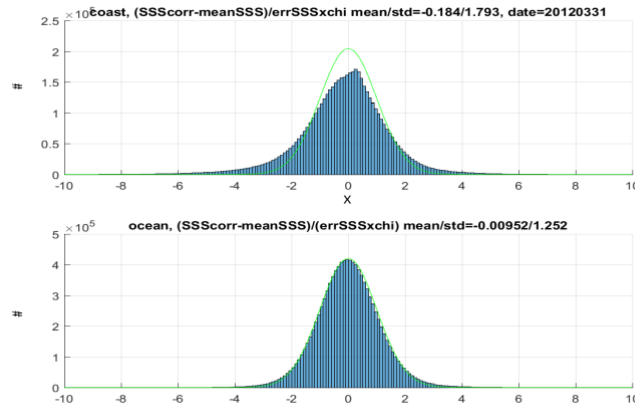
Figure 1: Histogram of the new random variable X (reduced centred SSS) after applying a coastal correction. Top: pixels near the coast (dcoast<400 km); bottom: pixels in the open ocean. March 2012.

So, to conclude, the SMOS theoretical uncertainty multiplied by the χ adjustment is, in some cases, underestimated, especially close to the Asian coast and to places heavily contaminated by RFIs. This indicates the presence of outliers. The algorithm of Boutin et al. 2018 allows us to eliminate some of these outliers by adding a 3-sigma filter from data intercomparison.

A specific CCI processing is implemented close to the coast by multiplying the theoretical uncertainty by a factor depending on the distance to the coast f(dist) (see Figure 2).
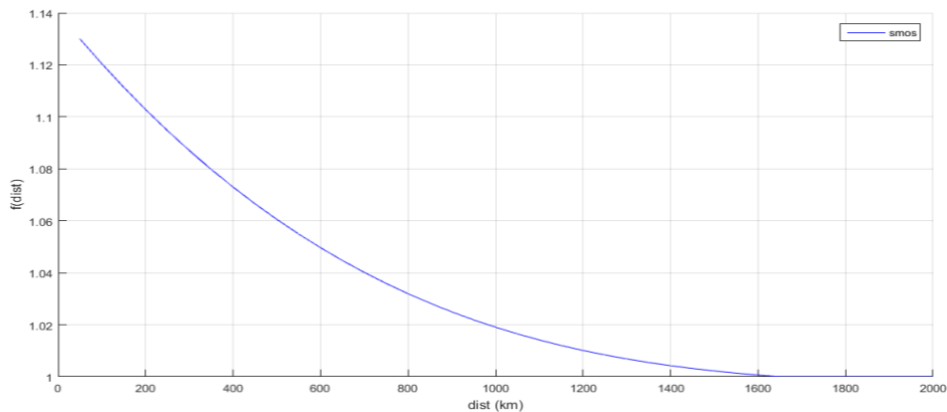
Figure 2: uncertainty factor according to the distance to the coast.

4.3.3.2.2 Systematic uncertainty

4.3.3.2.2.1 Characterisation of systematic uncertainties related to rainfall events

The objective of the CCI is to provide "bulk" SSS, i.e. not affected by the instantaneous rainfall. To correct this effect, a first model linking surface freshening and the instantaneous rain rate has been proposed by LOCEAN (Supply, 2020). However, this model does not integrate wind effects.

Indeed, vertical mixing is very efficient, especially in strong winds, and the surface effects related to rain are quickly attenuated.

In order to integrate the effect of the wind in the rain correction, we propose to start from the LOCEAN model, with an additional weight by the wind, following an adaptation of the model proposed by (RD15):

$$dSSS=a0(RR).WS^{-1.1}$$

With a0(RR), the freshening corresponds to an instantaneous rain rate (RR) for a wind speed (WS) of 1m/s. The original model predicts freshenings an order of magnitude lower than the ones observed on SMOS SSS as a function of IMERG RR. In order to calibrate the Drushka formalism with satellite observations, we will take as a0(RR) the correction obtained by A. Supply from freshening data taken on the ITCZ region after integrating it on all wind speeds.

The relation proposed by Drushka et al. (RD15) does not fit well with satellite observations, likely due to inhomogeneous conditions in satellite pixels integrated in ~50km. Moreover, the correction can reach very large (infinite) non-physical values for very low wind speeds. Therefore, it is necessary to choose a wind speed cutoff below which the correction no longer depends on the wind speed. The new relation which integrates this WSmin cutoff is written as follows:

$$dSSS(RR,WS>=WSmin) = a0(RR).WS^{-1.1}/<WS^{-1.1}>$$

$$dSSS(RR,WS<WSmin) = a0(RR).WSmin^{-1.1}/<WS^{-1.1}>$$

$<WS^{-1.1}>$ is the weighted average of WS over the ITCZ region, equaling 0.17. Note that this average is only slightly dependent on RR. With this formulation, we ensure the continuity of the correction according to the wind and avoid over-corrections at very low winds. Moreover, it allows us to recover the a0(RR) correction of A. Supply was obtained by averaging the freshening in the ITCZ region overall wind speeds (see A. Supply, 2020 PhD thesis, https://theses.hal.science/tel-04024079)).

To choose the WSmin cutoff, we selected heavy rainfall events. We compared the corrected SMOS SSS with the ISAS SSS (over areas of low variability) under different wind cutoffs and over the period [2015, 2016]. Figure 3 shows the histograms of the SMOS-ISAS residuals after the rain correction for regions of high SST.

Figure 4 shows the median(SMOS-ISAS) and std(SMOS-ISAS) as a function of the WSmin cutoff. The WSmin cutoff is chosen such as the one which minimises the std and the absolute value of the median differences.
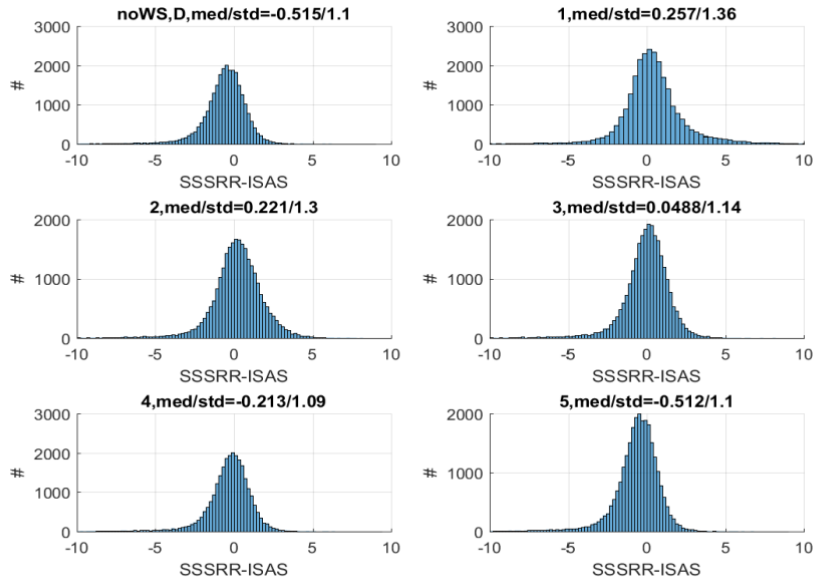
Figure 3: Corrected SSS – ISAS SSS with different WSmin cutoffs in case of heavy rain and high SST (>25°C). Top left, WS-independent a0(RR) correction; then, from top to bottom and from left to right, WSmin cutoff taken at 1, 2, 3, 4, 5m/s respectively. Median and std differences are reported in the title of the figures for each WSmin cutoff.
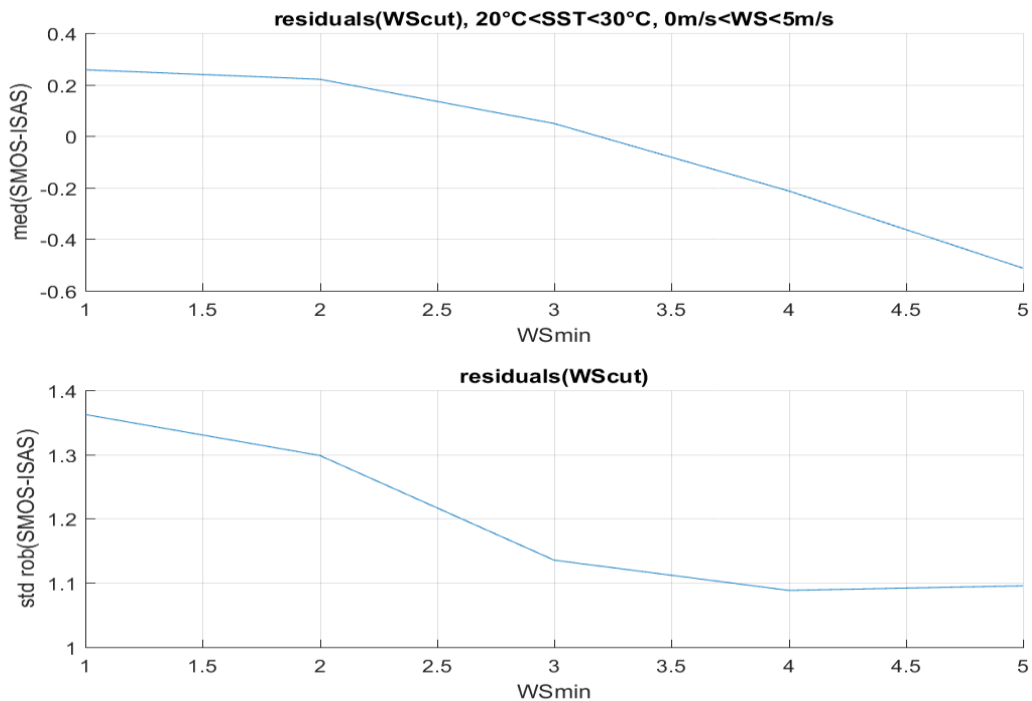

Figure 4. Top: median(SMOS-ISAS) as a function of the WSmin parameter chosen as cut-off. Bottom: std(SMOS-ISAS).

We did the same for regions with a low SST (Figure 5 et Figure 6).
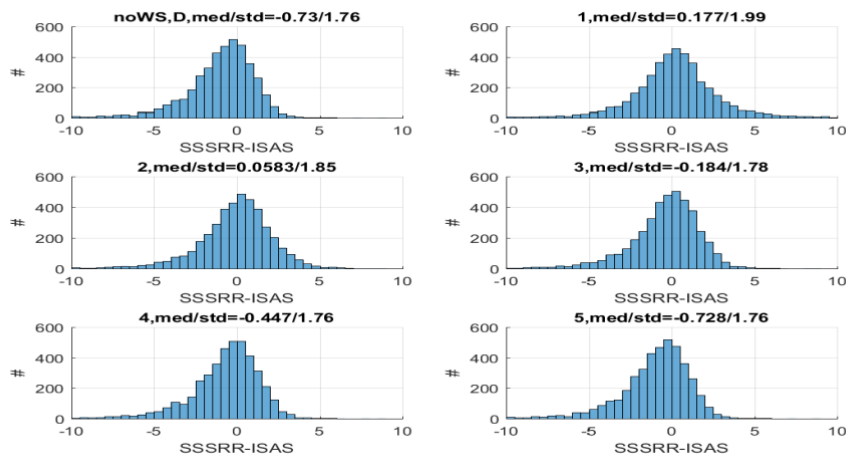
Figure 5: idem Figure 3 with an SST between 10 and 20°C
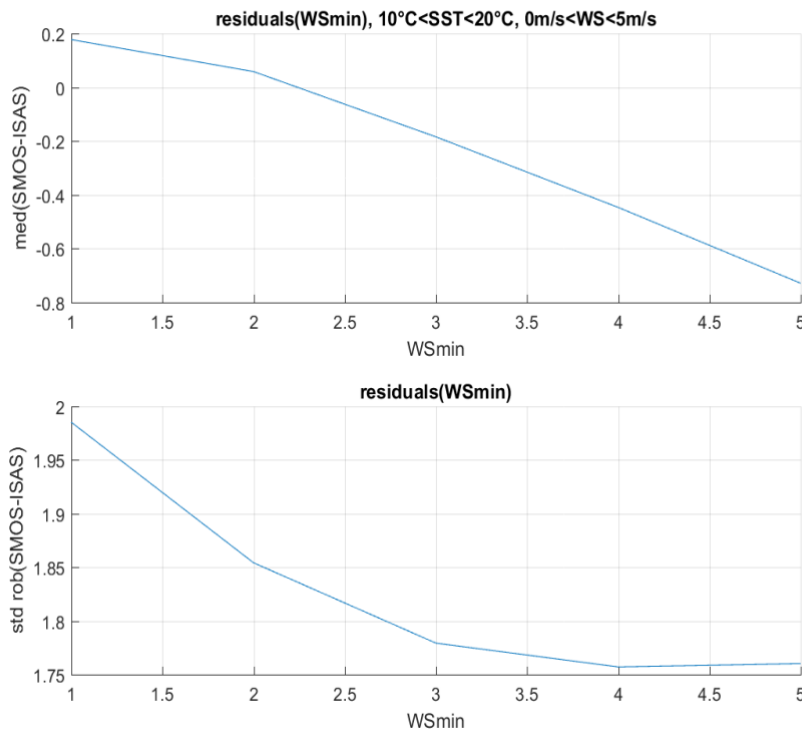
Figure 6: idem Figure 4 with an SST between 10 and 20°C

In the two SST cases presented, the minimum of std(SMOS corrected - ISAS) is around 4m/s. The bias, on the other hand, has a minimum around 2 or 3 m/s.

Based on these results, we took a cutoff value of WSmin=3.5m/s.

The freshening correction according to the RR for various wind speeds is presented in Figure 7:
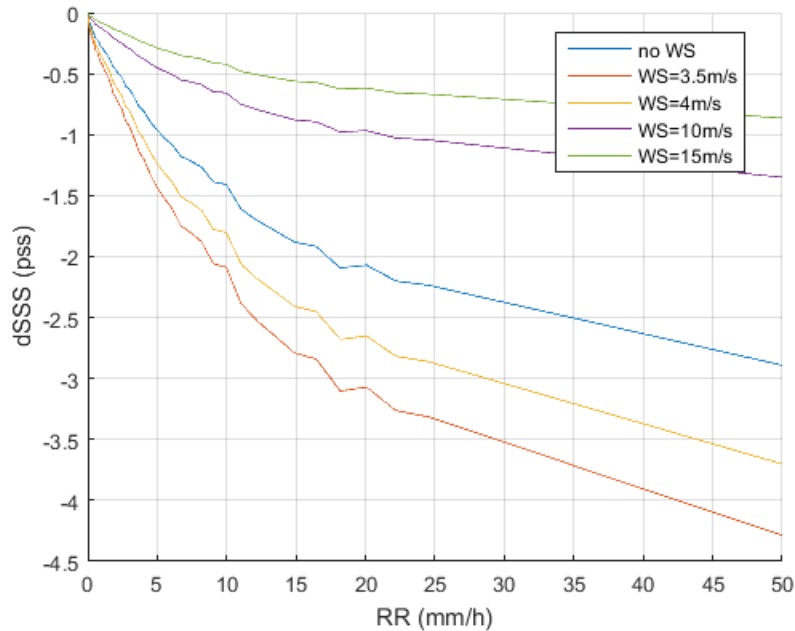
Figure 7: correction of the effect of rain on SSS according to various WS. The correction integrated overall wind speeds is shown as a blue curve.

In Figure 7, the blue curve shows the a0(RR) correction, which is applied on CCI+SSS version 3 based on the work of Alexandre Supply (RD16). The other curves correspond to the CCI+SSS V4 correction for winds ranging from 3.5 to 15m/s. The curve at 3.5m/s corresponds to the maximum possible correction given the choice of WSmin=3.5m/s.

4.3.3.2.2.2   Characterisation of systematic uncertainties related to the dielectric constant

The L2 SMOS SSS are retrieved with the BV dielectric constant model (Boutin et al., 2020) implemented in the L2OS processor version 7. This model has been updated recently with an adjustment of the Debye model based on laboratory measurements instead of SMOS measurements. The new parametrisation is called BVZ (RD17). In CCI+SSS V4, we correct the SSS level for systematic differences due to the change in the dielectric constant model.

There are at least two effects:

- An effect comes from the estimation of the OTT. Indeed, a change in the dielectric constant model modifies the calibration of the brightness temperatures at the time of the OTT calculation. The mean brightness temperature deviation between the two models averaged over SST encountered in the OTT area (SST between 10°C and 25°C) was estimated to be 0.1K (Stokes 1, St1). This systematic effect depends on the SST and is written as a bias on the SSS:

$$dOTTsss=0.1/(dSt1/dSSS)=0.1/(0.015*SST+0.25)$$

This formula shows that an OTT bias has a greater impact on the SSS the lower the SST is.

- An effect related to the change of the model itself on the SSS estimator. The change from BV to BVZ generates the following first-order correction:

$$SSS_{BVZ} = SSS_{BV} + (St1_{BVZ} - St1_{BV})/(0.015*SST + 0.25)$$

Since both effects are cumulative, the total correction to be applied is, therefore:

$$SSS_{BVZ} = SSS_{BV} + (St1_{BVZ} - St1_{BV})/(0.015*SST + 0.25) + 0.1/(0.015*SST + 0.25)$$

The SSS differences induced by the model change and approximated as described above are very close to SSS differences obtained with full L2OS processor retrievals (RD17).

The SSSacard is obtained directly using the BVZ model, so there is no need to apply the second effect.

SSScorr and SSSuncorr are corrected for the two effects mentioned above.

For data from the beginning of the period covered by the CATDS, the SSSacard is corrected based on the SSScorr DPGS.

4.3.3.2.2.3 Characterisation of systematic uncertainties related to wind speed

In order to characterize systematic uncertainties related to wind speed, the starting point is the SSS corrected for rain and dielectric constant effects.

To characterize the systematic effects related to the wind model, we select observation conditions not strongly affected by other uncertainty sources by considering the following criteria:

- SSS variability < 0.1
- Descending orbits (in order to avoid strong ice contamination in the Southern Ocean)
- abs(xswath)<200km
- Fg_sunglint = 0
- Data outside the eastern equatorial Pacific region [abs(lat)<5° for -180°<lon<-80°] known to experience large positive biases, likely related to strong currents.

According to the investigations performed by N. Reul in the frame of the ESA L2OS studies, the correction is expected to depend on the sea states. Here, two classes of sea states are considered: a class dominated by the wind sea, called "wind" hereafter, and a class dominated by swell, called "swell". These two classes are identified by sea state flags in the L2OS V7 product (see table below).

The occurrence of the sea states is given as a function of time in Figure 8. The abrupt temporal variations in these occurrences are not physical and correlate with the ECMWF processing change. When gathering the sea states into two classes (wind and swell), these abrupt variations

almost disappear, with nevertheless a remaining effect associated with the 45R1 change early 2018 (Figure 9). Following these results, we decided to merge the six classes into two classes.

Table 1: L2OS sea states

|	
Fg_sc_sea_state_1[Ngp]	Sea state class 1: waves wind dominated, old sea. Flag set if swell fraction < Tg_swell and omega < Tg_old_sea
Fg_sc_sea_state_2[Ngp]	Sea state class 2: waves swell dominated, old sea. Flag set if swell fraction >= Tg_swell and omega < Tg_old_sea
Fg_sc_sea_state_3[Ngp]	Sea state class 3: waves wind dominated, medium sea. Flag set if swell fraction < Tg_swell and omega >= Tg_old_sea and omega <= Tg_young_sea
Fg_sc_sea_state_4[Ngp]	Sea state class 4: waves swell dominated, medium sea. Flag set if swell fraction >= Tg_swell and omega >= Tg_old_sea and omega <= Tg_young_sea
Fg_sc_sea_state_5[Ngp]	Sea state class 5: waves wind dominated, young sea. Flag set if swell fraction < Tg_swell and omega > Tg_yound_sea
Fg_sc_sea_state_6[Ngp]	Sea state class 6: Waves swell dominated, young sea. Flag set if swell fraction >= Tg_swell and omega > Tg_young_sea
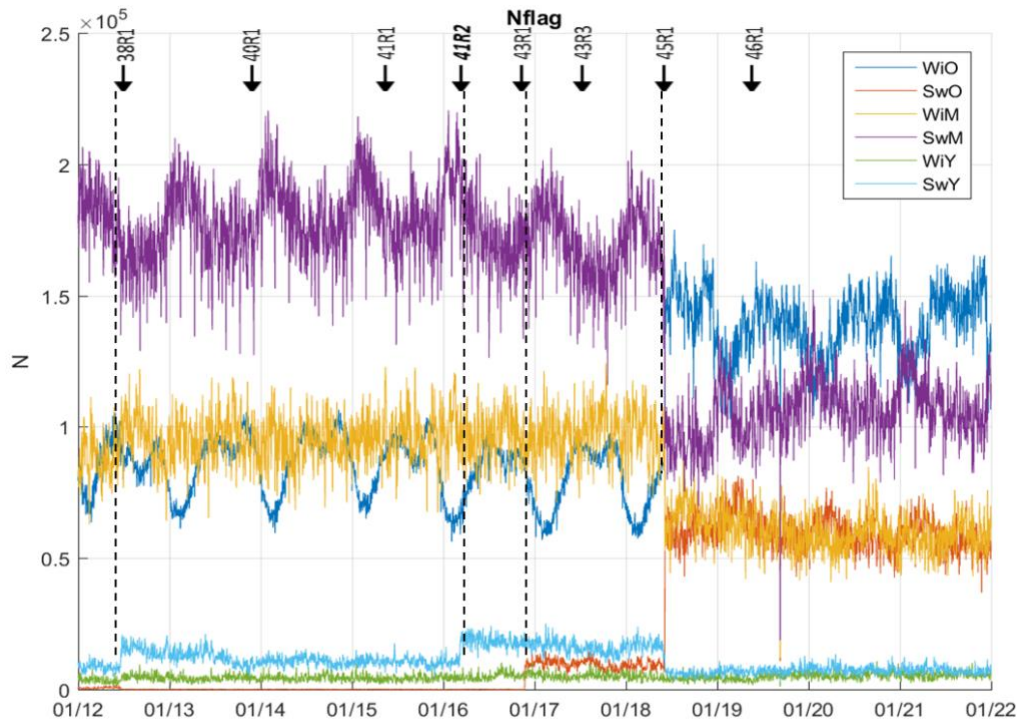
Figure 8: Number of flags raised as a function of time for the 6 L2OS classes. ECMWF model versions are indicated at the top of the figure. Model changes in phase with abrupt changes in the number of flags are indicated as dotted vertical lines
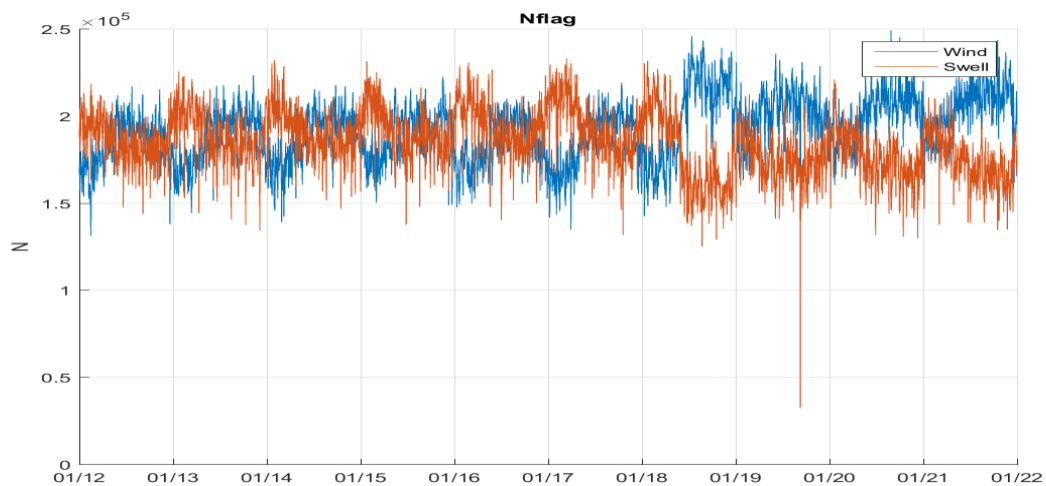
Figure 9 : number of flags for the two classes, swell and wind.

The bias on the SSS is characterised in the (WS, SST) plane: bSSSobs(WS, SST). If the bias of the modelled TB is independent of the SST, then the effect on the estimated SSS depends on the SST due to the dependence of ∂Tb/∂SSS with SST. It is estimated with the following relationship:

$$bSSStheo(WS,SST)=bSSS(WS,15°C).(0.015*15+0.25) /(0.015*SST+0.25) \qquad \text{Eqn 4-3}$$

where bSSS(WS,15°C) is the estimated wind speed related bias at SST=15°C. According to this equation, for a given bias in TB, the SSS bias should not change of sign if the model bias in TB is independent of SST. Figure 10 shows that this is not the case for WS<4m/s in case of a swell, suggesting that for low wind speeds, the Tb bias is not independent of SST. For this reason, the correction at WS<4m/s is done without following Eqn 4-3.

The SSS biases in the (WS and SST) plans were calculated for SSScorr, SSSuncorr and SSSacard. There are no significant differences between the SSScorr and SSSuncorr biases. However, the SSSacard behaves differently, which can be explained by the fact that the wind components are not retrieved during the retrieval, while this retrieval may be affected by an error in the wind model and/or in the dielectric constant model. The correction applied to Acard will therefore be specific. The BSS (WS,15°C) correction is obtained from the bias given in the (WS,SST) plane based on a least squares algorithm. It is estimated separately for the wind and swell class (Figure 11).
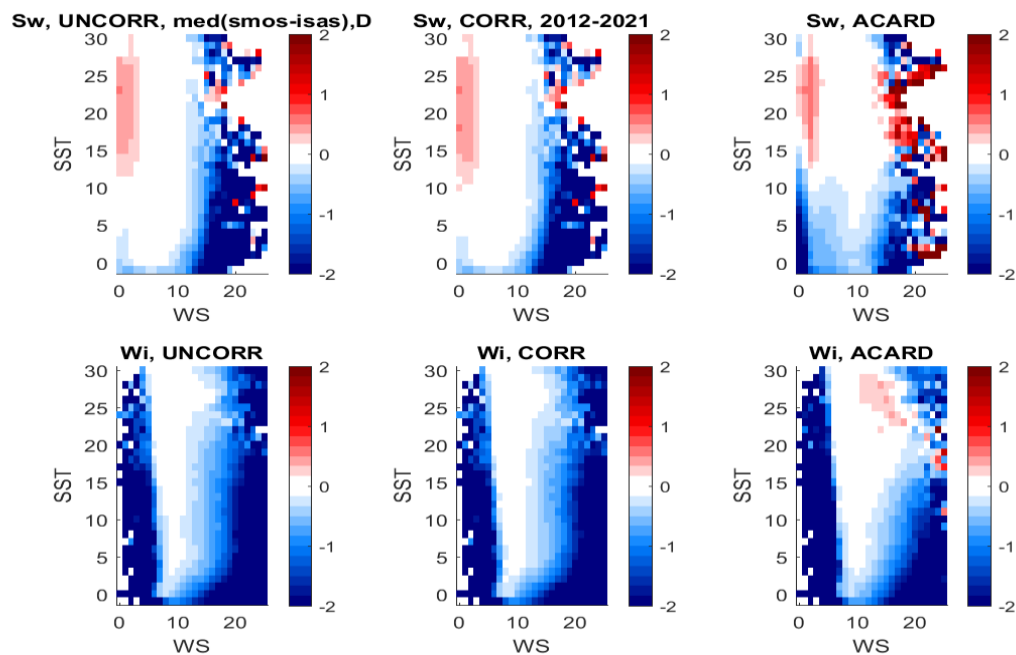
Figure 10: SSS biases in the (WS,SST) plane (descending orbits). Top row: roughness dominated by the swell; bottom row: roughness dominated by the wind sea.
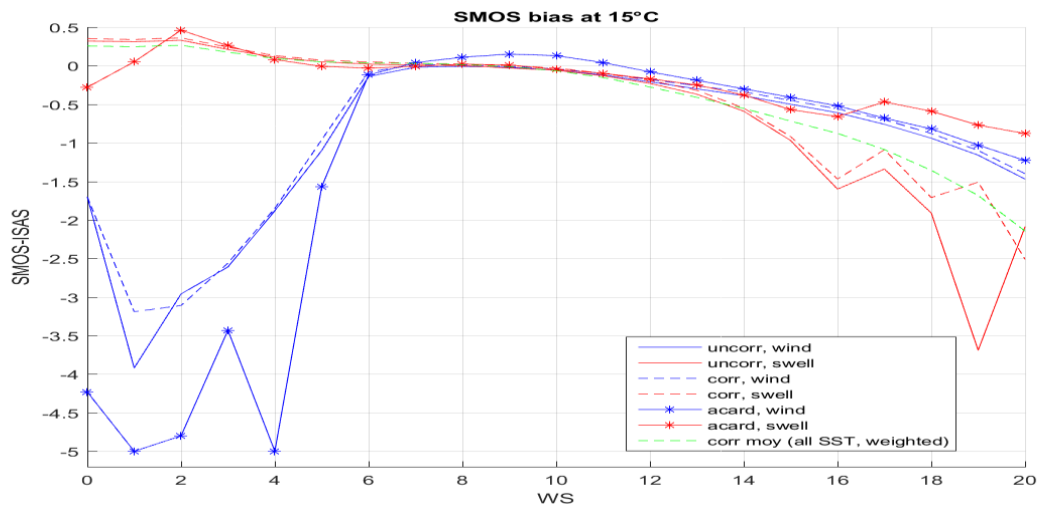
Figure 11: Correction at 15°C for the 3 SSS (uncorr, corr and acard) and the two sea classes (swell and wind).

Given these results, the correction algorithm for a satellite pass over a given grid node (to which we associate three estimators of the SSS: SSScorr, SSSuncorr and SSSacard) follows the following steps:

-identification of the class to which the SSS belongs according to the sea state flags (wind or swell)

-if the SSS belongs to the wind class, the correction calculated at 15°C is applied to the effective SST (bSSStheo) for the SSScorr and SSSacard cases.

-if the SSS belongs to the swell class with a WS>4m/s, the correction calculated at 15°C is applied to the effective SST (bSSStheo) for the SSScorr and SSSacard cases.

-If the SSS belongs to the swell class with a WS<4m/s, the bSSS_obs correction is applied.

The same correction is applied to SSScorr and SSSuncorr.

For the data from the beginning of the period covered by the CATDS, we cannot classify the sea states (there are also sea state flags, but they are not defined in the same way as in DPGS). We apply an average WS correction independent of SST and sea state. This correction is obtained from the DPGS data averaged and weighted by the number of measurements.

4.3.3.2.2.4 Characterisation of latitudinal seasonal systematic uncertainties

A bias depending on latitude and season was quantified (Boutin et al. 2018). Therefore, we are in the situation of Eqn 4-2 with 2 biases to manage (latitudinal and inter-dwell biases).

SMOS data is affected by latitudinal and coastal systematic uncertainties (biases).

The latitudinal bias was estimated under the following conditions:

• Period [2013-2021].

• Pseudo-dielectric constant (Acard) filtering (|Acard smos – Acard mod| <4). This filtering allows removing part of the ice and RFI contaminations.

• Filtering of SSS with large random uncertainty (L2 SSS random uncertainty < 20)

• Filtering of SSS which pertains to the outside reasonable interval (SSS > 42 or SSS< 5)

• Filtering of SSS with WS>16 m/s

• Distance to the coast > 600km. Exclusion of some regions strongly affected by RFIs. The selected area for latitudinal computation is given in Figure 12.

• Smoothing of the correction with a 5° latitude window width (instead of 10° in version 1)

• Reference SSS from ISAS: ISAS 17 + delayed and NRT products

• Selection of ISAS SSS explaining a significant amount of variance (with PCTVAR < 80%)

In order to characterise the latitudinal seasonal systematic uncertainties, SSS corrected for rain, dielectric constant and wind effects are used.

It is assumed that the LSC does not contaminate the open ocean data. Data are selected at 600km from the coast (if a too large limit is taken, it is impossible to calculate the correction north of 45°N). A mask is added over the North Pacific and the Indian Ocean, which RFI contamination affects.

The latitudinal biases depend on the season (solar and galactic contamination, eclipse period), the type of orbit (ascending or descending) and the acquisition geometry. The latter is characterised by the distance of the acquisition from the satellite track expressed in km and called swath in the following.

In order to estimate this bias, we use ISAS as a reference. The ISAS data are filtered on a PCTVAR criterion (< 80) and a variability criterion (<0.1pss). In order to cover all latitudes, in case of coverage problems, the filtering criteria of the ISAS data are relaxed.
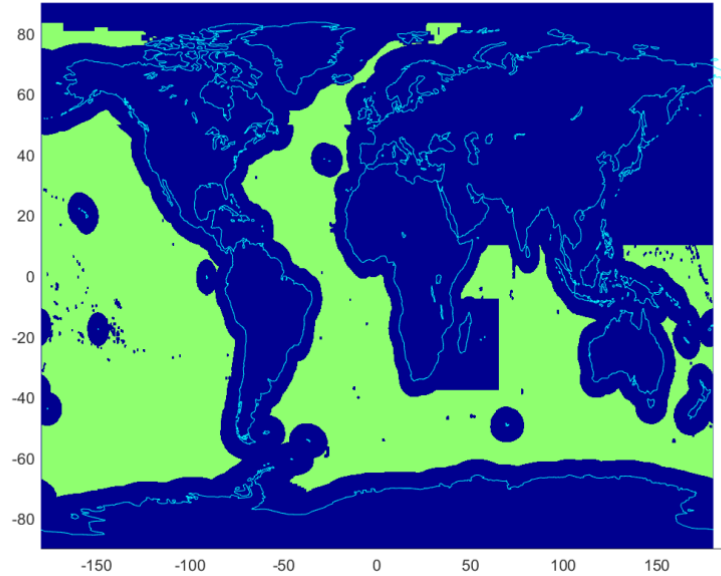
Figure 12: Mask for seasonal latitudinal correction

We keep the CATDS latitudinal correction already applied for the beginning of the period.

The biases estimated from SSScorr are used to correct SSSuncorr. This reduces the effects of coastal contamination in calculating the latitudinal correction.

The following figures (Figure 13) show, for some cases, the estimated latitudinal biases for SSSuncorr (not used for correction), SSScorr (used for correction of SSSuncorr and SSScorr) and SSSacard, the month of January.
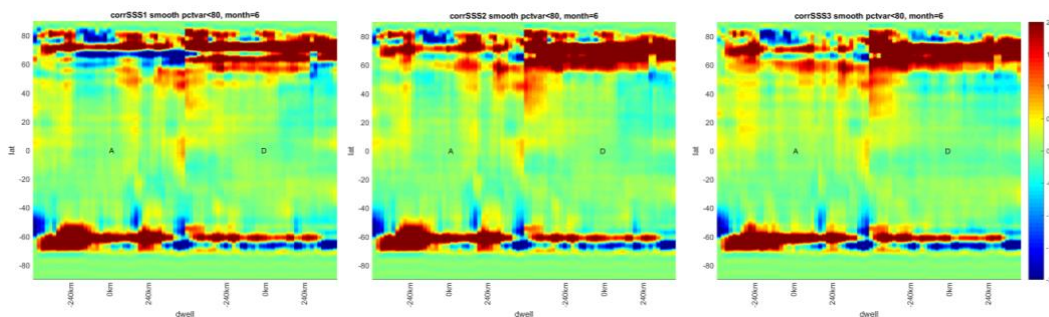
Figure 13: example of latitudinal correction for the 3 SMOS SSS (SSSuncorr on the left, SSScorr in the middle and SSSacard on the right). Ascending dwells are shown on the left-hand side of each figure, descending dwells on the right-hand side. Month of June. The y-axis gives the distance from the track in km.

Once the correction has been applied, a check on the effect of the correction is required, in particular, to identify inter-annual biases that could not be corrected because of the seasonal assumption. The largest part of the biases is well corrected except in certain situation where the

bias has some inter-annual signature, especially during descending orbits in 2014 and 2015, north hemisphere (Figure 15, Figure 17 and Figure 19). Ascending orbit correction behaves more homogeneously (Figure 14, Figure 16 and Figure 18). Some inter-annual signatures appear at high latitudes.
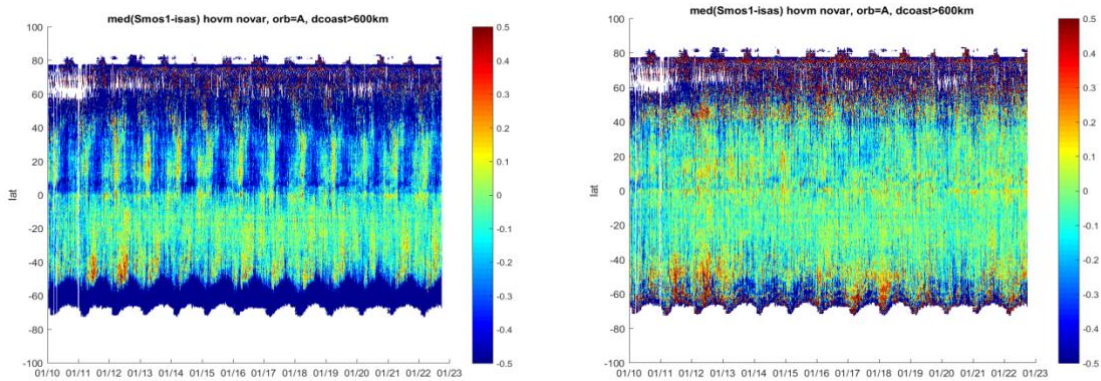
Figure 14: Hovmoller of the difference between SSSuncorr and ISAS, ascending orbits. Left: before latitudinal correction; right: after latitudinal correction.
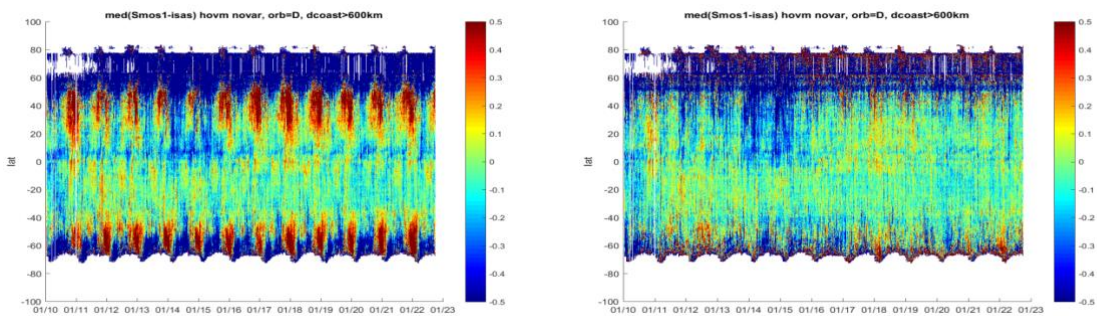
Figure 15: idem than Figure 14 for descending orbits.
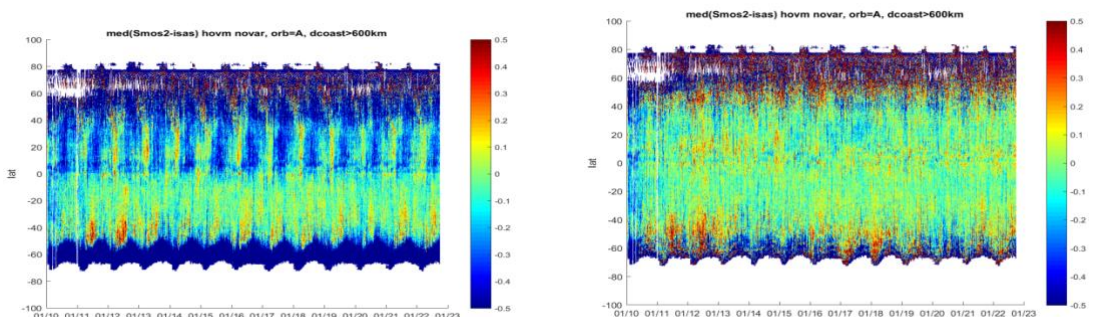
Figure 16: idem than Figure 14 for SSScorr.
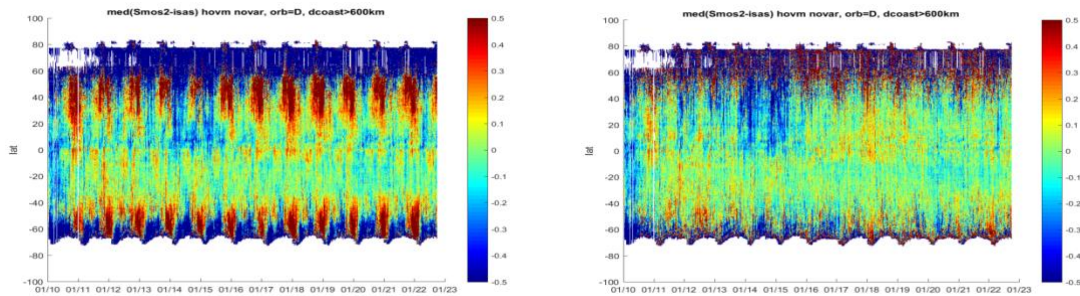
Figure 17: idem than Figure 14 for SSScorr and descending orbits.
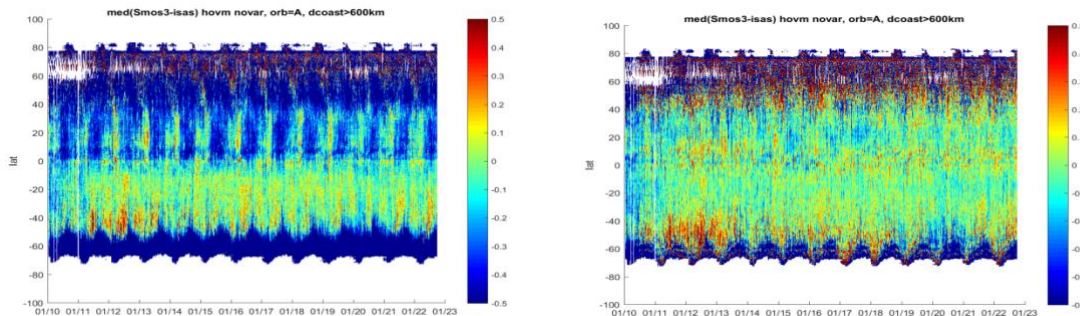
Figure 18: idem than Figure 14 for SSSacard.
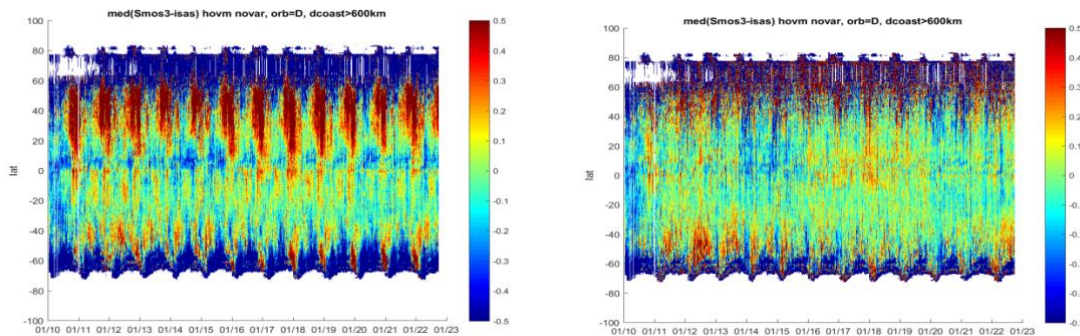
Figure 19:: idem than Figure 14 for SSSacard and descending orbits.

4.3.3.2.2.5 Characterisation of RFI systematic uncertainties.

We present results for the Samoa region (178.255°E, 198.255°E, 23.874°S, 3.874°S) in the southwestern tropical Pacific, but the RFI filtering method has also been applied to other polluted areas like the Barbados islands (69.55°W, 49.55°W, 3.234°N, 23.234°N) and the Guinea Gulf (13°W, 7°E, 5°S, 15°N). In section 4.2.2.5, we described the RFI correction process, and here we highlight its key steps.
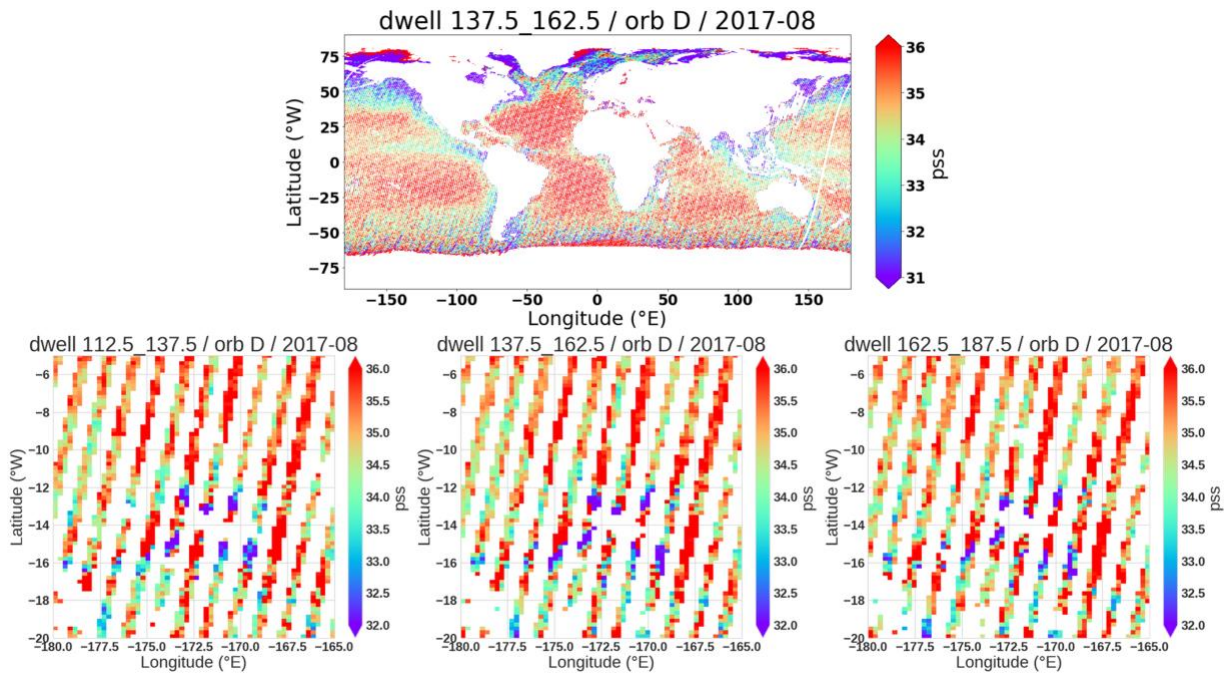
Figure 20: (Top) Monthly global map including dwell-line interval [+137.5km,+162.5km] of SSS for all descending half orbits in August 2017. (Bottom) Zoom in over the Samoa region for [+112.5km,+137.5km], [+137.5km,+162.5km], [+162.5km,+187.5km] intervals of descending half orbits for the same month.

Monthly agglomeration maps in Figure 20 for Samoa exhibit low salinity levels indicating RFI pollution. Figure 21 displays dwell differences $\Delta_{i,j}S^{11}$ with a star-like structure characteristic of RFI disturbance as seen in the maps, and where Hovmöller plots emphasise the start and impact of the RFI source. The first mode of PCA in Figure 22 explains about 60% of the variance for the dwell-line differences, with time series $u_0^{11}(t)$ characterising RFI timing. Spatial structures of correction per dwell $V_0'(d_k, x, y)$ are shown in Figure 23 for a few cases, zoomed in to a small area around the source to clearly see the features.
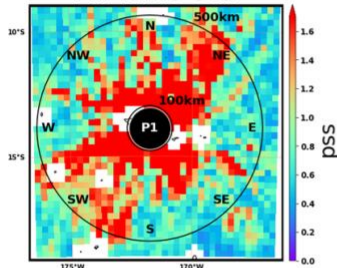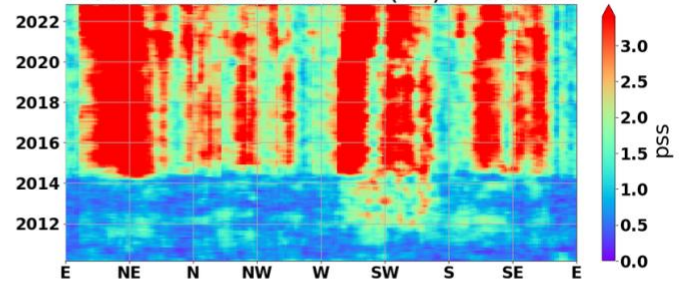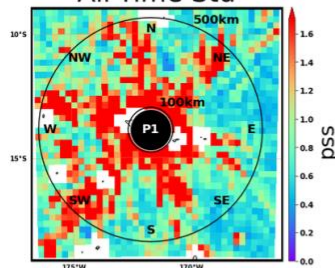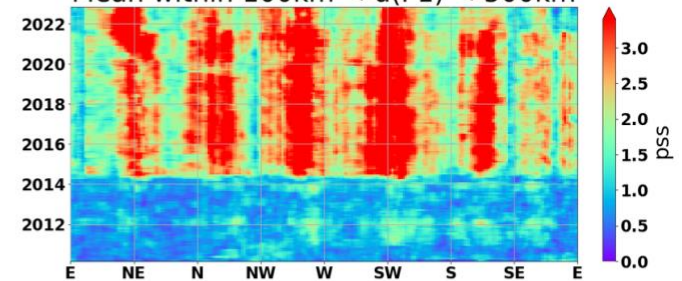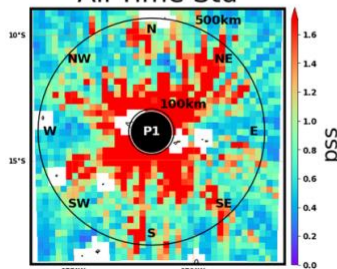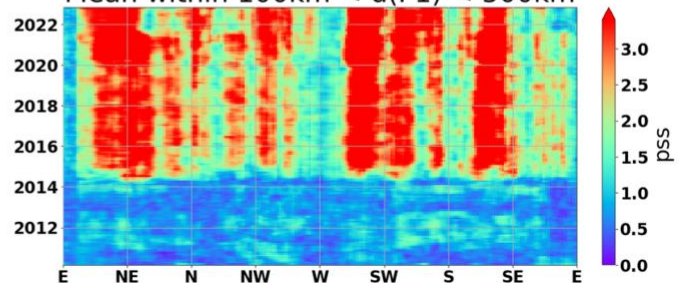
Figure 21: Examples of dwell-line differences calculated from the 11-month moving averaged SSS dwell-lines ($\Delta_{i,j}S^{11}$). (Left) Maps of the std over the entire period of the absolute difference of various 11-month m.a. dwell-line fields. (Right) Angular hovmöller plots where directions E, NE, … are indicated in the rose around the source (P1) on the Left and where SSS is averaged over the angular sectors diverging from the source.

These maps provide a signed weight that will be applied to the unique corrective time series shown in Figure 24 to build the space and time correction. Figure 25 displays the applied correction on a monthly agglomerate dwell field, highlighting the characteristic

RFI   *spatial*   *structure.*
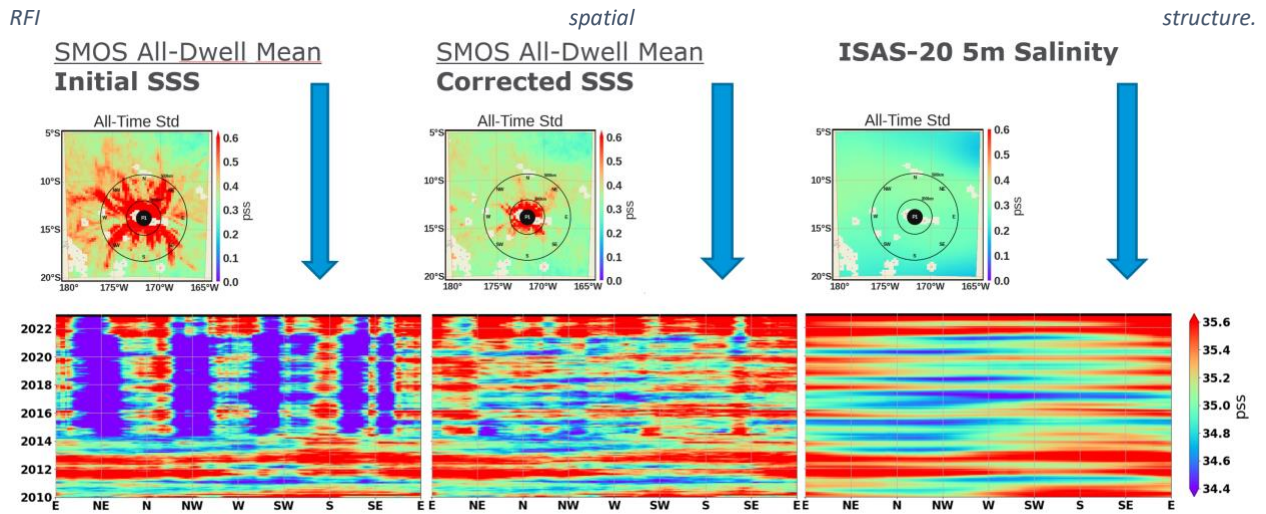
Figure 26 demonstrates the improved performance of the correction method over time. The corrected space/time field is closer to ISAS20, and SSS variability is significantly recovered from below the initial RFI pollution. Finally, Figure 27 depicts the evolution of a comparison metric with respect to ISAS20, revealing that the RFI perturbation has been effectively filtered out, resulting in corrected SSS that more closely matches ISAS20.
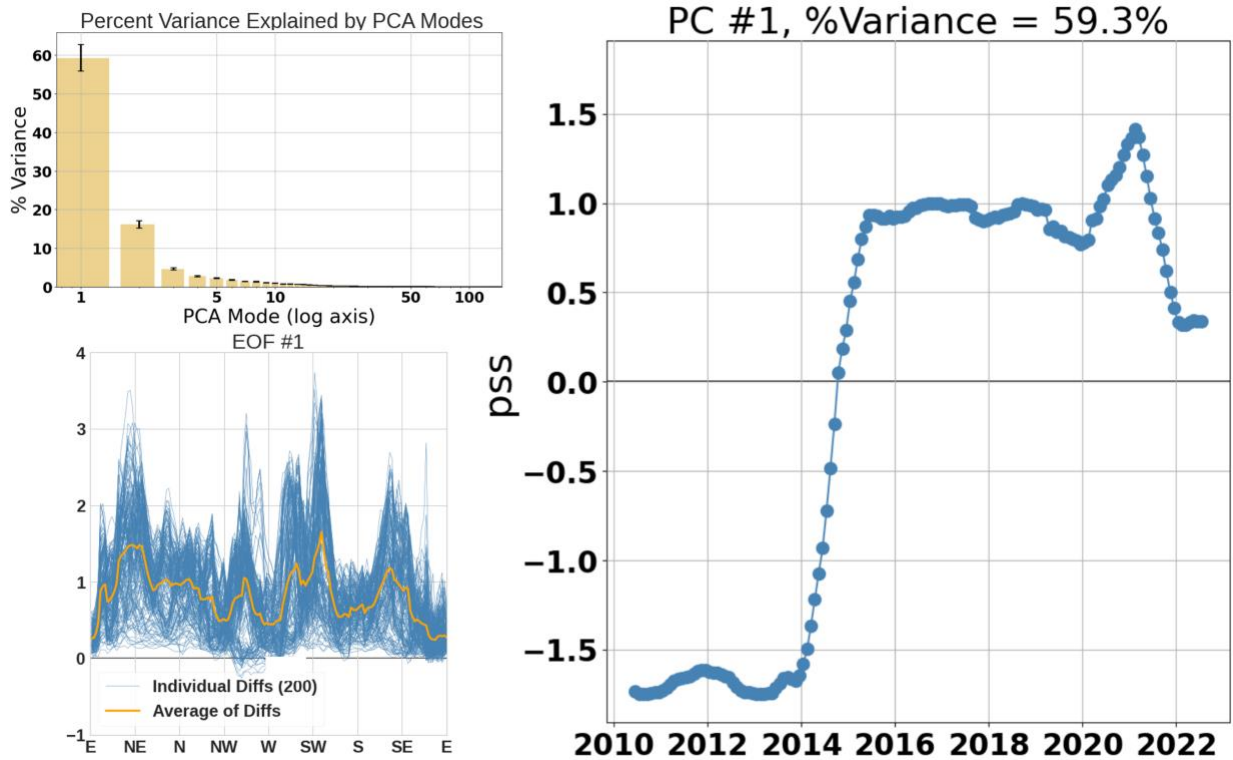
Figure 22: 1st mode results of the PCA applied to the dwell-line differences. (Top-Left) Percentage of variance explained by the PCA modes as a function of the mode number. (Top-Bottom) Spatial structures of the first mode are given in angular sectors around the RFI source. (Right) The time series is associated with the first mode (in pss).
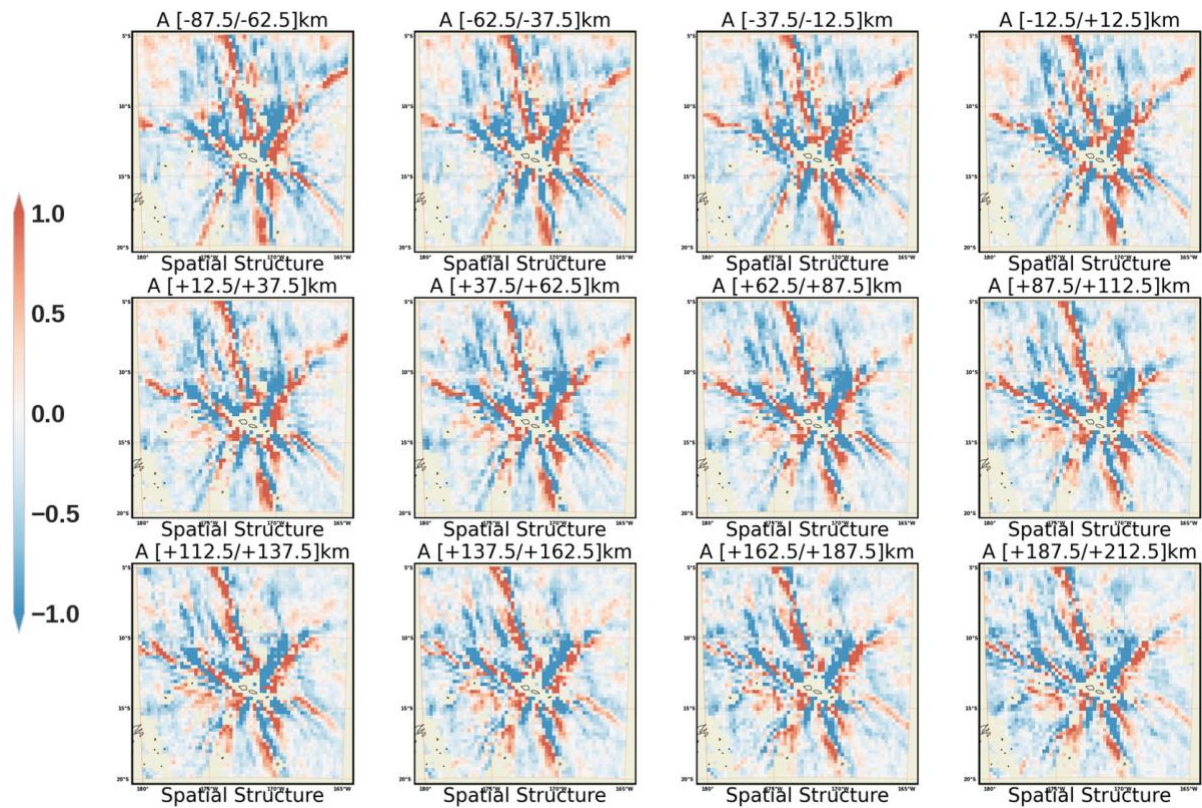
Figure 23: Portion of the spatial patterns related to the initial regression utilized to compute $V_0'(d_k, x, y)$, which is the spatial element of the correction package.
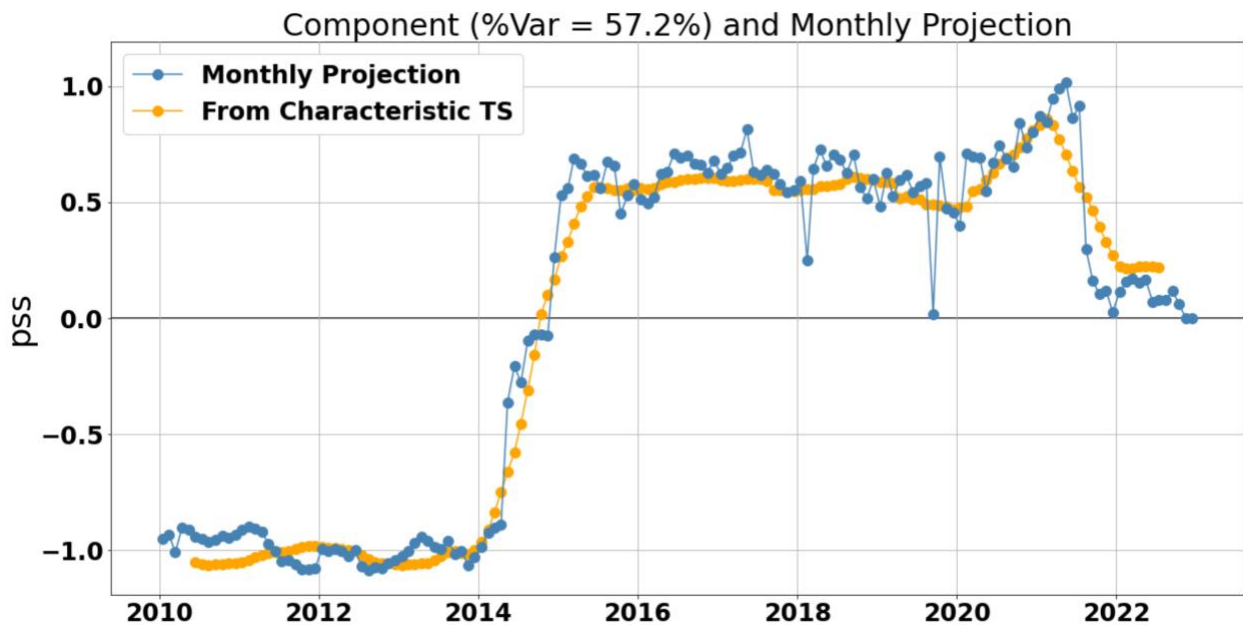
Figure 24: (Blue) Monthly RFI characteristic time series $U_0^1(t)$ obtained by projecting the monthly dwell-line agglomerates onto the RFI spatial structures. (Orange) The (11-month m.a.) characteristic time series previously calculated is superimposed for comparison.
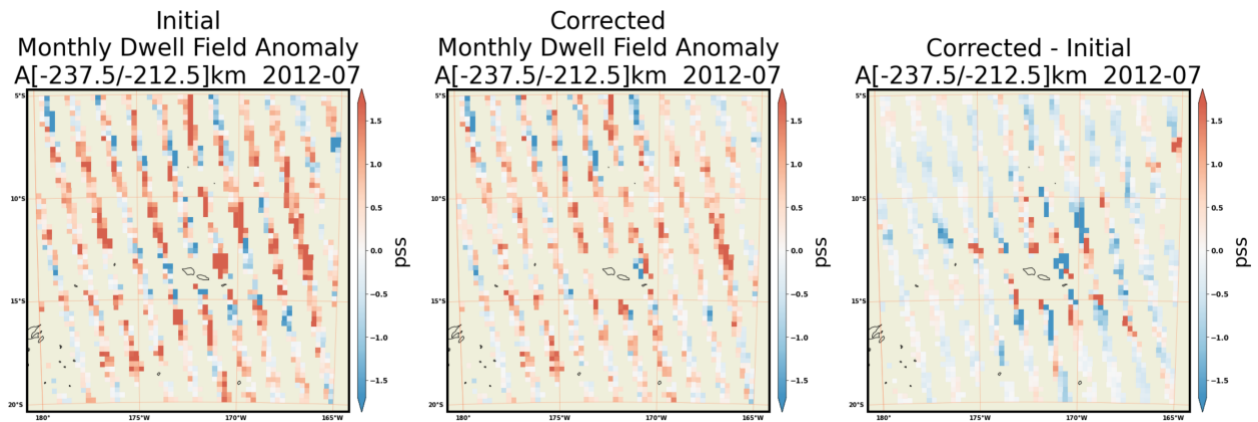
Figure 25: Example of correction for a specific dwell line (Ascending, -237.5km,-212.5km) in July 2012. (Left) Initial dwell field SSS anomaly (with respect to time mean). (Middle) Corrected dwell field anomaly. (Right) Difference between corrected and initial fields.
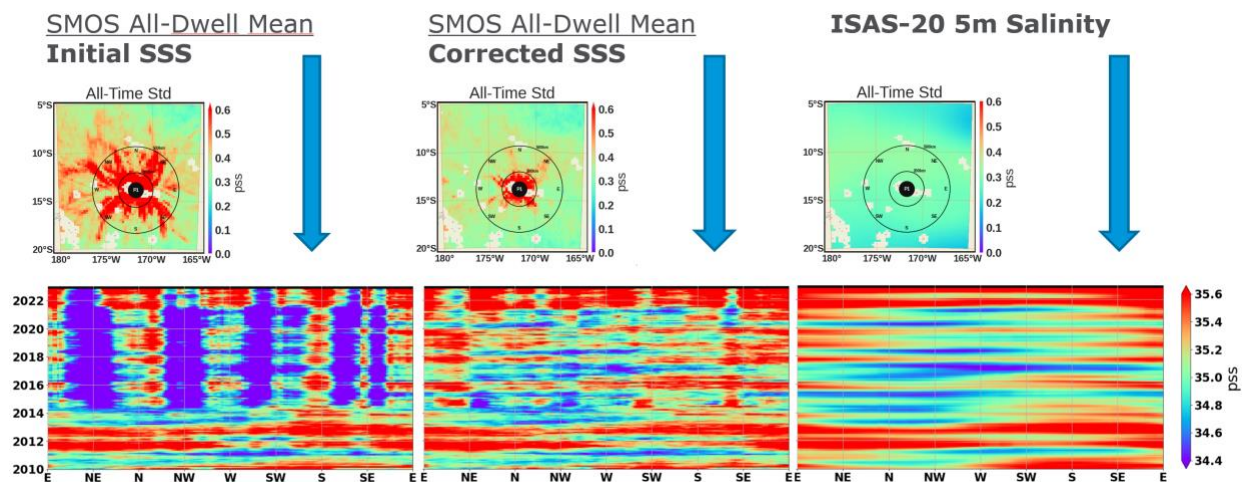
Figure 26: Performance of the correction method with respect to ISAS-20 5m salinity. Here we average SSS over all the dwell lines, add a constant time-mean field, and look at angular hovmoller variations around the source (Left) All-dwell mean of initial SSS. (Middle) All-dwell mean of corrected SSS. (Right) ISAS20 5m salinity.
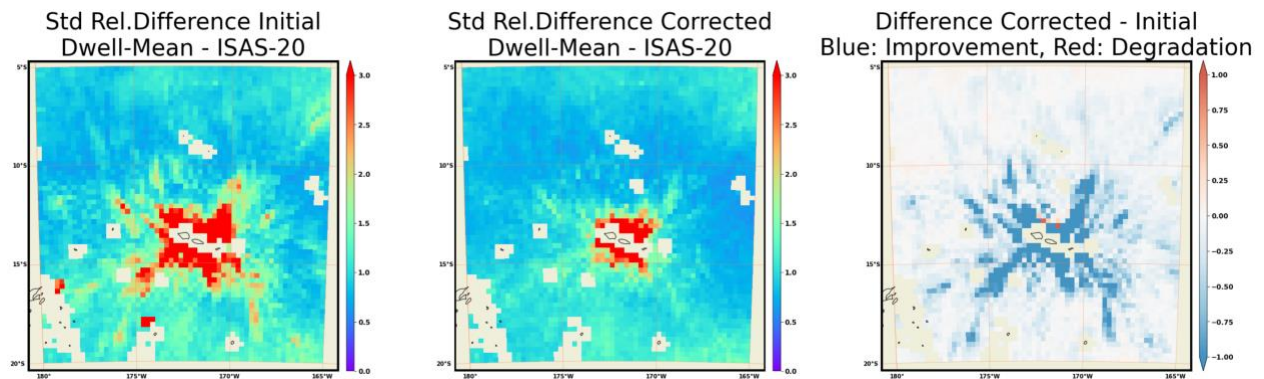
Figure 27: Performance of the correction method with respect to ISAS-20 5m salinity. Here we calculate and plot a performance metric with respect to ISAS20 at each grid point. All dwell SSS are preliminarily averaged. (Left) Plot of the metric $std(sat - isas)/std(isas)$, for the initial SSS dwell data. (Middle) Plot of the same metric for the corrected SSS dwell data. (Right) Difference (Middle)-(Left), that highlights improvement (blue) or degradation (red) with respect to ISAS20.

4.3.3.2.3 flagging the data

We chose to apply the SSS filtering before merging and before correcting for relative and absolute biases so that to remove most of the pixels with SSS outside the reasonable oceanographic range, contaminated by ice or by RFI, or contaminated by high wind speed:

- SSS associated with wind speed larger than 16 m/s has been removed.

- SSS associated to pseudo-dielectric constant (Acard) is kept only if |Acard smos – Acard mod|<2 and Acard>42. This filtering allows removing part of the ice and of the RFI contamination.

- SSS is kept if SSS > 2 & SSS<45

- SSS is removed if chi > 3 or SSS random uncertainty > 3.

- SSS is removed if the acquisition occurs too far from the cente of the track: abs(xswath)>400km.

- SSS is kept if fg_outlier = 0 (fg_outlier is raised at L2OS if the number of TB outlier data is lower than a given threshold).

4.3.3.3   SMAP L2 data

We used SMAP L2C v5.0 products (SMAP_RemSSS_Release_V5.0), SSS at 40 km resolution.

4.3.3.4   Random uncertainty

An empirical estimation is performed for the SSS L2 uncertainty estimate (see [AD 06]) by using self-consistency approach. For SMAP, it is possible to compare SSS coming from aft and fore acquisition and to compute the std of the difference which should be an estimator of the SSS random uncertainty multiplied by sqrt(2) (assuming that the random uncertainty according to SST is the same for aft and for acquisition). This yields the relationship of the SSS uncertainty according to the SST (Figure 28). This uncertainty is derived far from coast to avoid land and RFI contamination effects. This empirical uncertainty is not used during the L4 optimal analysis (L2C theoretical random uncertainty is used).

As for SMOS, specific processing is implemented by multiplying the previous random uncertainty by a factor depending on the distance to the coast f(dist) (Figure 29). This factor was obtained by comparing the open ocean uncertainty with an error estimation according to the coast distance. Indeed, coast contaminations are expected due to secondary lobe and RFI.
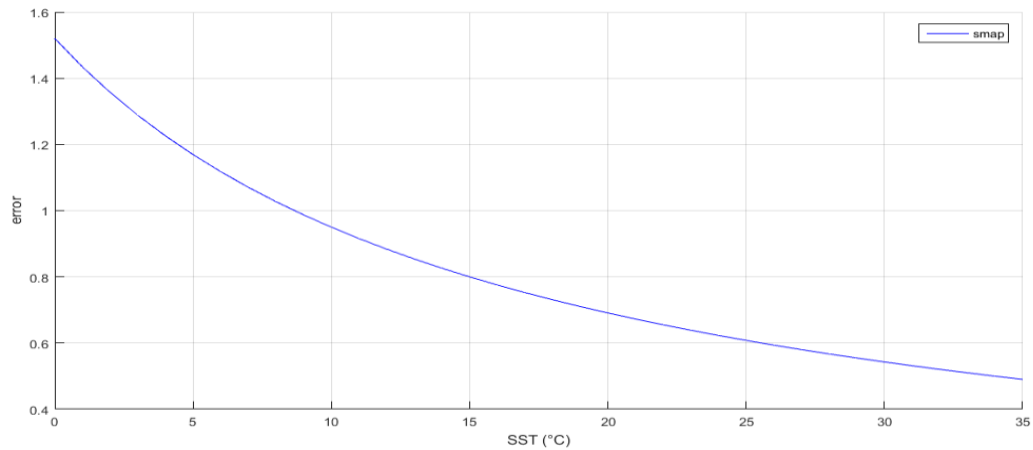
Figure 28: SSS SMAP uncertainty obtained by comparing aft and fore acquisitions.
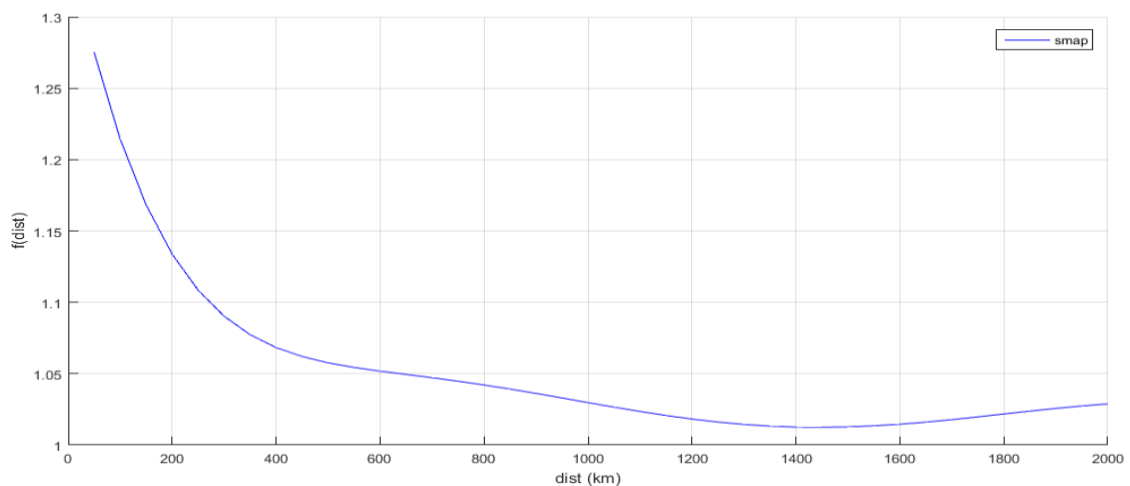

Figure 29: uncertainty factor according to the distance to the coast. SMAP.

4.3.3.4.1 Systematic uncertainty

4.3.3.4.1.1   Characterization of systematic uncertainties related to the dielectric constant

The dielectric constant model used in the SMAP processing is that of Meissner & Wentz. The impact of model differences (Klein & Swift, BV, BVZ) on the SSS plays out mainly at low SST. Moreover, the biases at low SST are much less critical on SMAP than on SMOS. We proceeded to a systematic correction in this situation according to the SST. This correction is empirical and uses ISAS as a reference. This correction is shown in Figure 30. It is very similar for ascending and descending orbits except at very low SST, where a discrepancy occurs because of the need for more data.
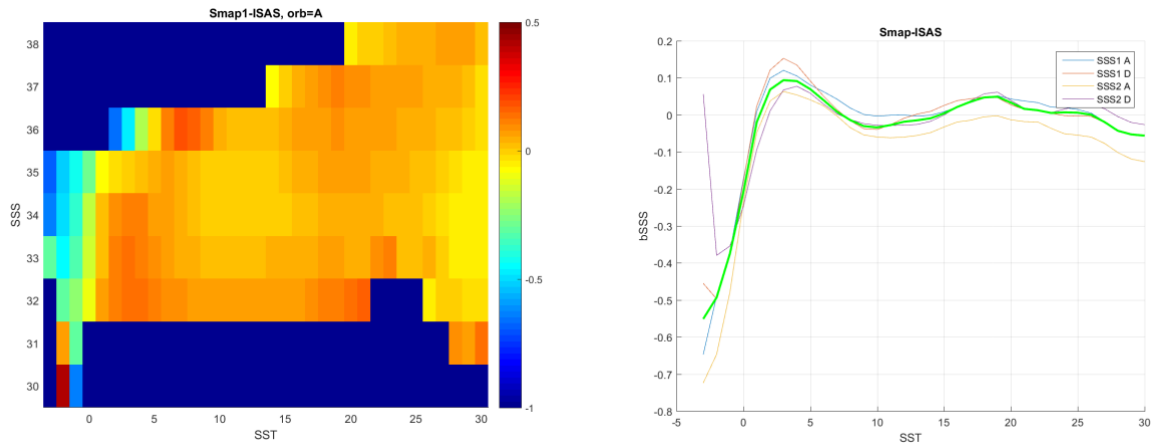
Figure 30. Left: SSS bias in the (SSS,SST) plan. Fore acquisition, A orbits. Right: average bias according to SST. In green, the average relation for the corrections. SMAP.
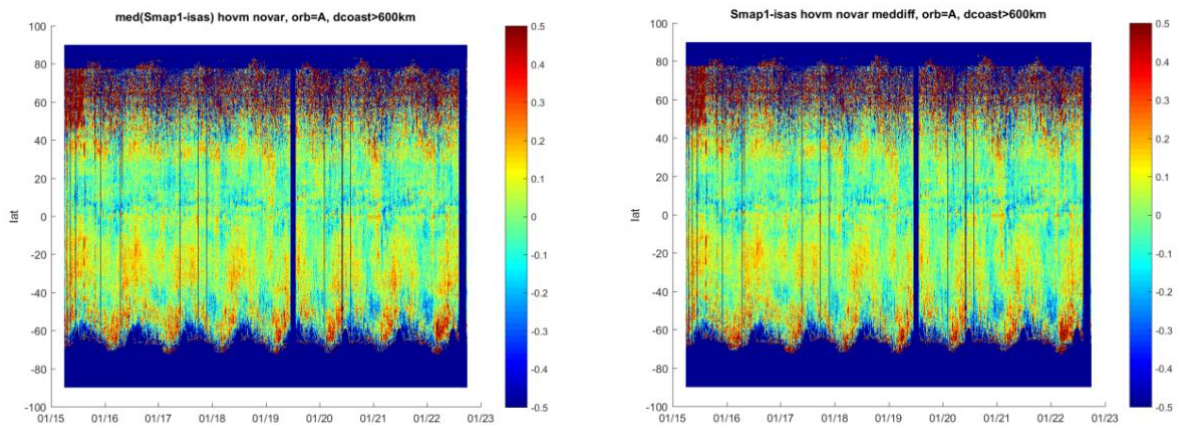
Figure 31: Hovmöller of the SMAP-ISAS difference. Left, uncorrected for SST; right, corrected for SST. Ascending orbits for acquisition. Slight correction at intermediate latitudes and the ice edge in the South.

4.3.3.4.1.2 Characterisation of latitudinal seasonal systematic uncertainties

In order to correct for latitudinal biases, we do not apply any particular mask. The pixels contaminated by land have been filtered out (see 4.3.3.4.2). The latitudinal biases are characterized by orbit type (ascending or descending) and acquisition direction (aft and fore).

Once the correction has been applied, a check on the effect of the correction is required, in particular, to identify inter-annual biases that could not be corrected.
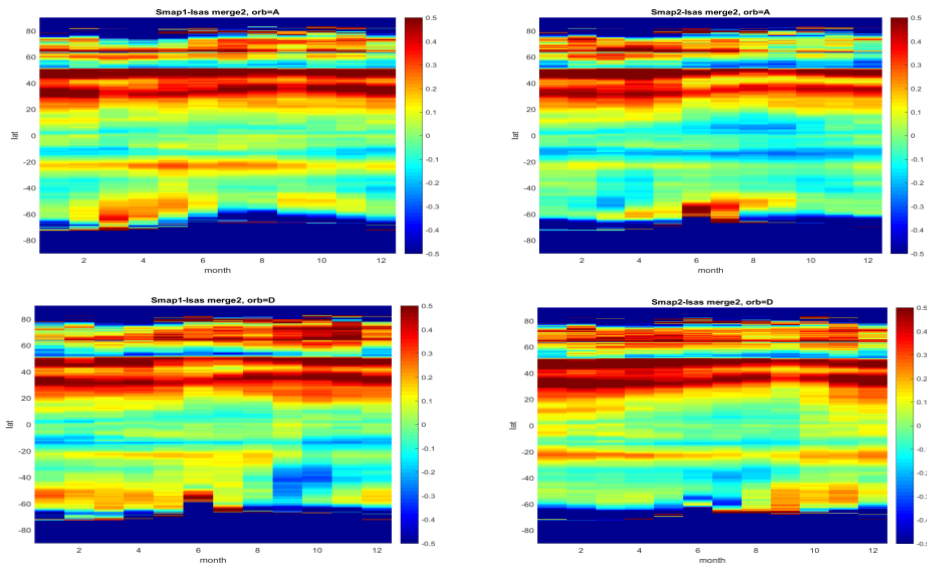
Figure 32: Seasonal latitudinal correction for SMAP. Top, ascending orbits; bottom, descending orbits.
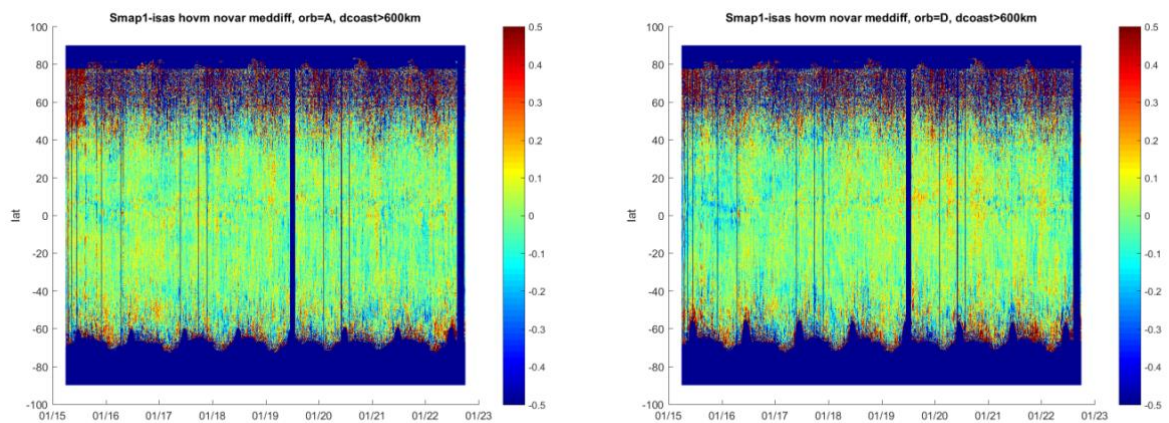

Figure 33: SMAP-ISAS after SST and latitudinal fore acquisition correction. Left: ascending orbits; right: descending orbits.
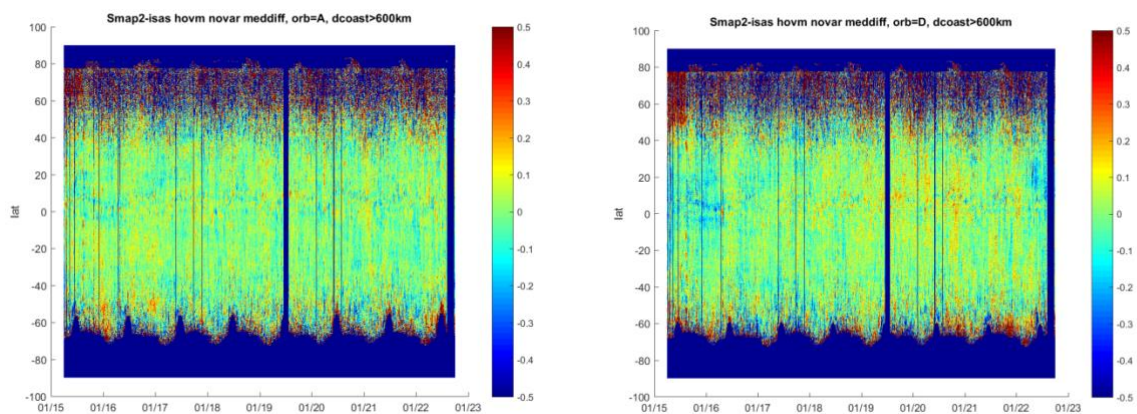

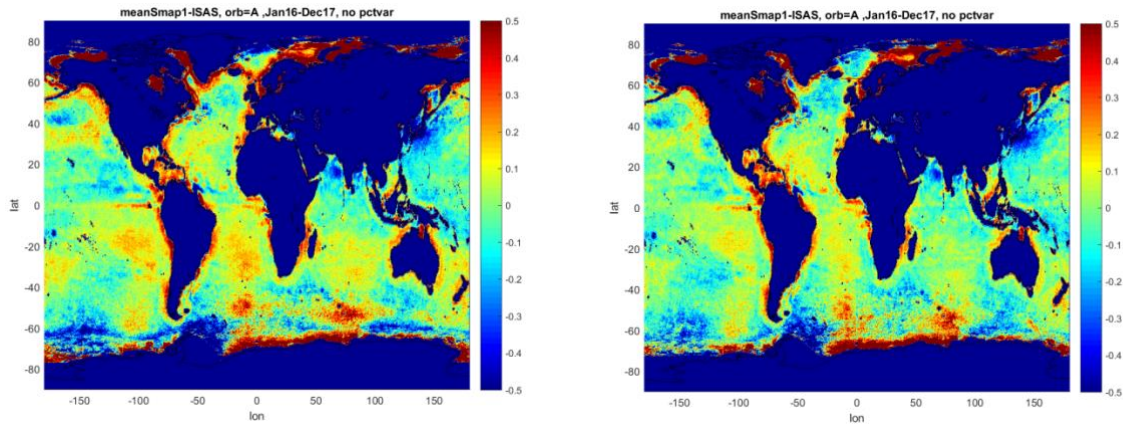*Figure 34: idem Figure 33 for afte acquisition.*

Figure 35: SMAP-ISAS over the period 2016-2017. Left, uncorrected SSS; right, corrected from SST and seasonal latitudinal biases. Fore acquisition, ascending orbits.
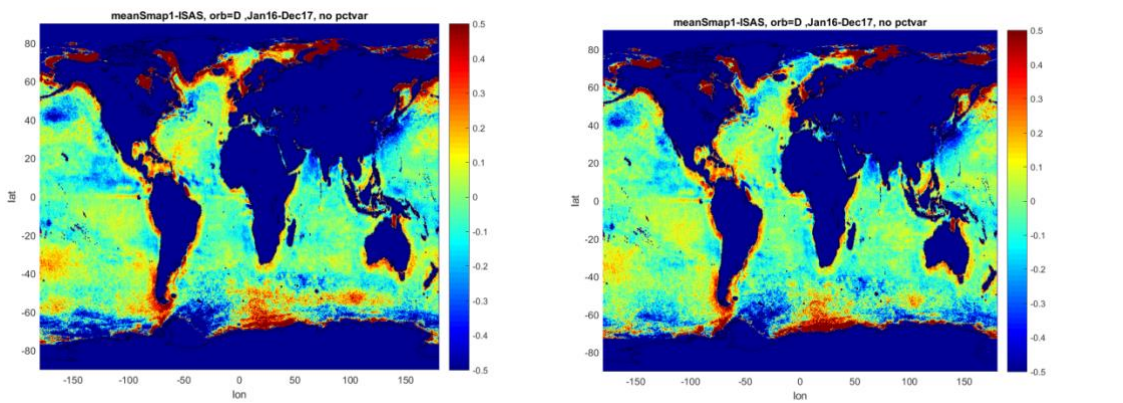

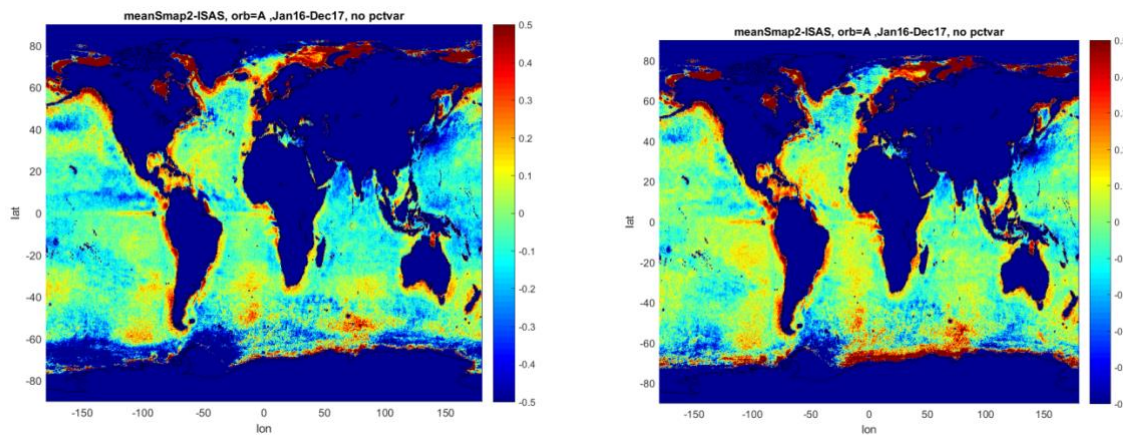*Figure 36: idem Figure 35 for descending orbits.*
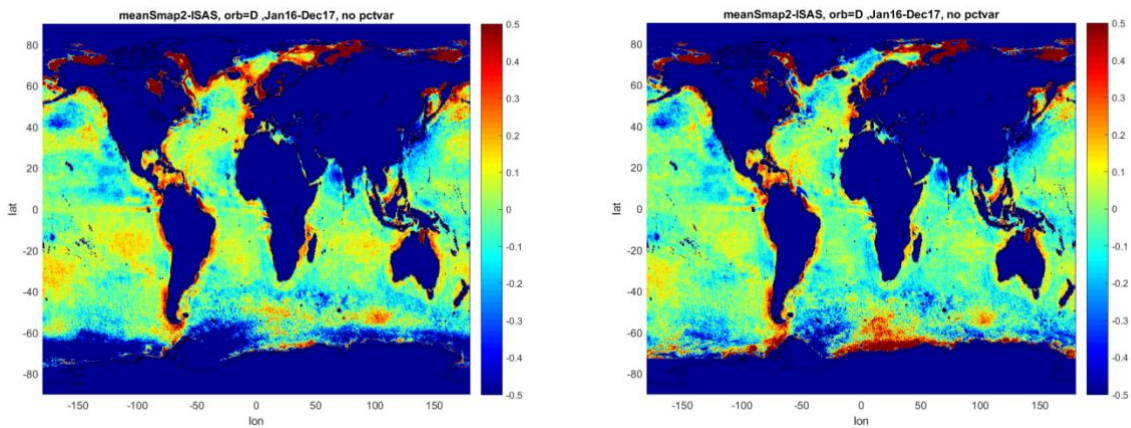

Figure 37: idem Figure 35, afte acquisitions.

Figure 38: idem Figure 35, afte acquisitions, descending orbits.

4.3.3.4.2 flagging the data

We describe below the flags we apply on the SMAP L2 SSS at 40 km. We use analogous filtering as the one used by RSS team. We exclude all the SSS at 40 km which have not been included for the 70-km smoothed product :

- 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).
- The moon glint angle is less than 15° (bit 6 in L2 Q/C flag is set).
- The v/h-pol average of the reflected galactic radiation exceeds 2.0 K (bit 7 in L2 Q/C flag is set).
- moderate land contamination: gain weighted land fraction gland exceeds 0.04 or land fraction in 3-dB footprint fland exceeds 0.005 (bit 8 in L2 Q/C flag is set)
- moderate sea ice contamination (bit 9 in L2 Q/C flag is set)
- TB consistency, defined as the $\sqrt{\chi^2}$ of the MLE in the salinity retrieval algorithm, exceeds 1.0 K (bit 10 in the L2 Q/C flag is set).
- WS<15m/s
- SSS removal is associated with an instantaneous rain rate larger than 0.5 mm/h (no correction applied for SMOS).

4.3.3.5 Aquarius L3 data

The following dataset is used: RSS L3 v5, ascending and descending separated products.

4.3.3.6 Random uncertainty

The random uncertainties can be calculated by comparing the SSS at time t with the SSS at time t+7 days due to the 7-day periodicity of the satellite orbit. SSS on the open ocean is assumed to stay mostly the same in 7 days. The standard deviation of the difference then gives a fairly precise

idea of the random uncertainties (multiplied by sqrt (2)). Figure 39 shows the standard deviation as a function of SST obtained far from coast.

As for SMOS and SMAP, specific processing is implemented by multiplying the previous empirical uncertainty (obtained far from the coast) by a factor depending on the distance to the coast f(dist) (Figure 40). This factor was obtained by comparing the open ocean uncertainty with an error estimation according to the coast distance. Indeed, coast contaminations are expected due to secondary lobe and RFI.
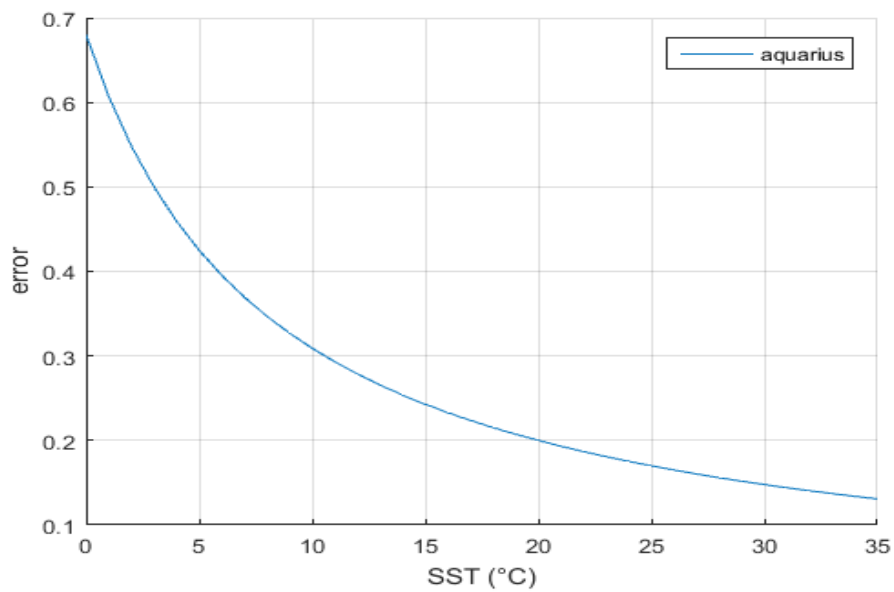
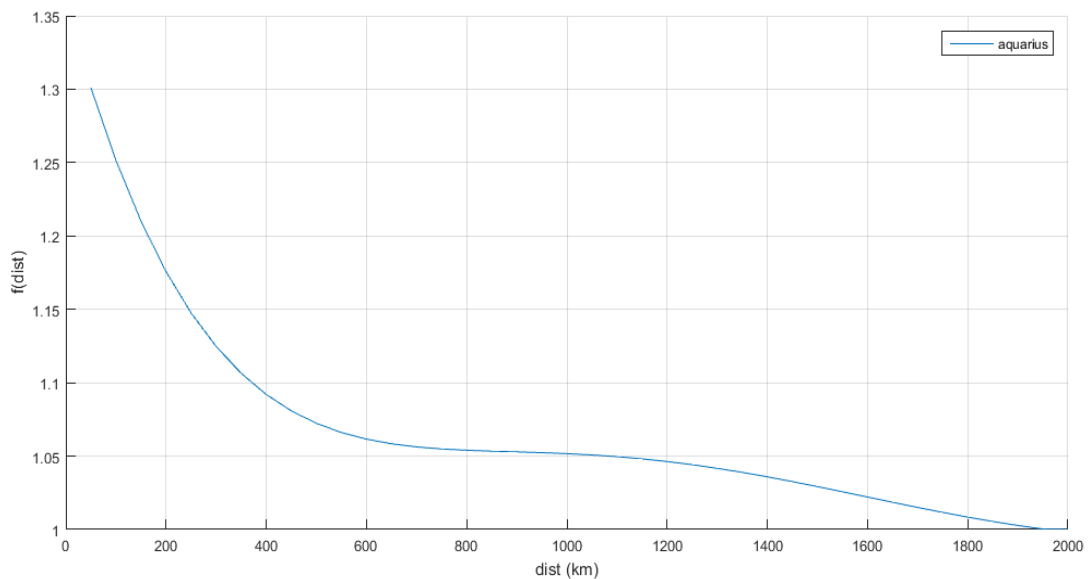Figure 39: SSS Aquarius uncertainty obtained by comparing 7 day acquisitions.

Figure 40: uncertainty factor according to the distance to the coast. Aquarius.

4.3.3.7   Systematic uncertainty

4.3.3.7.1.1   Characterisation of systematic uncertainties related to the dielectric constant

An error in the dielectric constant model has an effect expected to depend on the SST. For this reason, we sought to characterize the bias in SSS as a function of SST. In the southern Indian Ocean, we observe a positive anomaly that seems to be correlated with the SST front (Figure 41). In the (SSS SST) plan, the SSS bias follows a law essentially driven by the SST (Figure 42). Finally, when plotting the bias as a function of SST, the ascending and descending orbits provide the same trend with an overestimation of the Aquarius SSS at low SSTs and a slight underestimation of the SSS around 5°C (Figure 42). Based on these results, an SST-dependent correction was applied (green curve in Figure 42).

The effect of this correction on the Aquarius-ISAS hovmöller and SSS maps averaged over two years (2013 and 2014) is shown in Figure 43. It can be seen that the strong biases are very much attenuated after correction. However, this does not remove all the latitudinal seasonal biases.
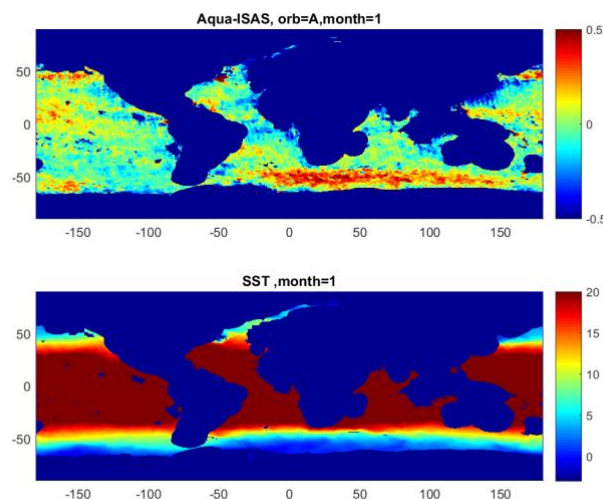

Figure 41: Example of the Aquarius bias for January (climatology) to be compared with the SST field. The positive SSS anomaly in the Southeastern Atlantic Ocean and Indian Ocean is correlated with a northern SST front.
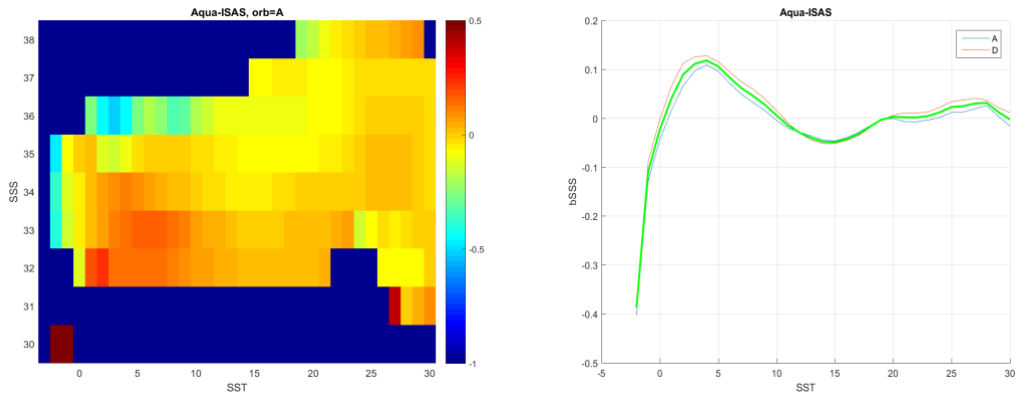
Figure 42. Left: Aquarius bias in the (SSS,SST) plan. Right: bias as a function of SST. In green, the average relationship for the corrections; in blue, the relationship obtained with ascending orbits and in red, with descending orbits.

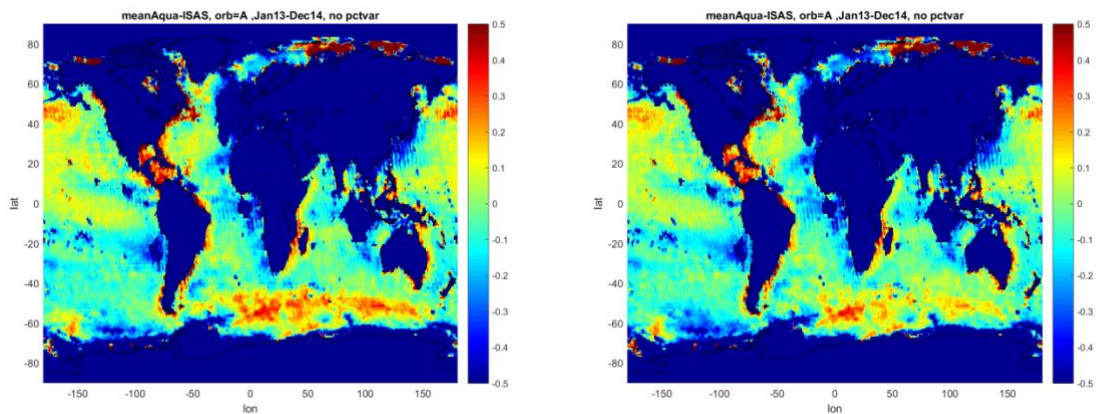There is a small dependence of the bias on SSS, which we did not consider.

Figure 43: Aquarius-ISAS differences over 2013-2014. Left, uncorrected for SST; right, corrected for SST. Ascending orbits.
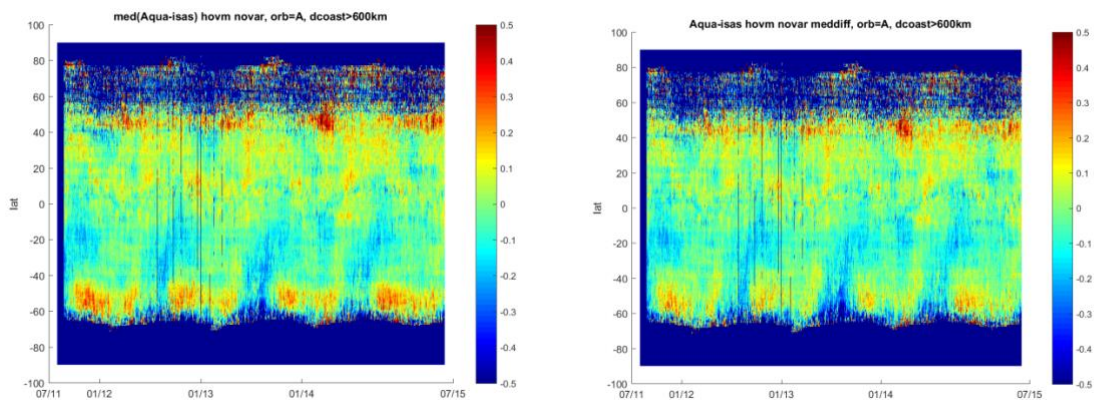

Figure 44: Hovmöller of the Aquarius-ISAS difference. Left, uncorrected for SST; right, corrected for SST. Ascending orbits.

4.3.3.7.1.2 Characterisation of latitudinal seasonal systematic uncertainties.

During phase 1 of the CCI+SSS project, we observe that latitudinal biases on L4 SSS appeared at high latitudes. These biases observed up to CCI+SSS V3.2 products originate partly from Aquarius, especially at high southern latitudes. After correcting for a bias related to the SST, non-zero Aquarius - ISAS residuals still show a latitudinal seasonal signature. To estimate the correction for these latitudinal biases, we apply a 600km mask. The latitudinal biases are characterized by orbit type (ascending or descending).
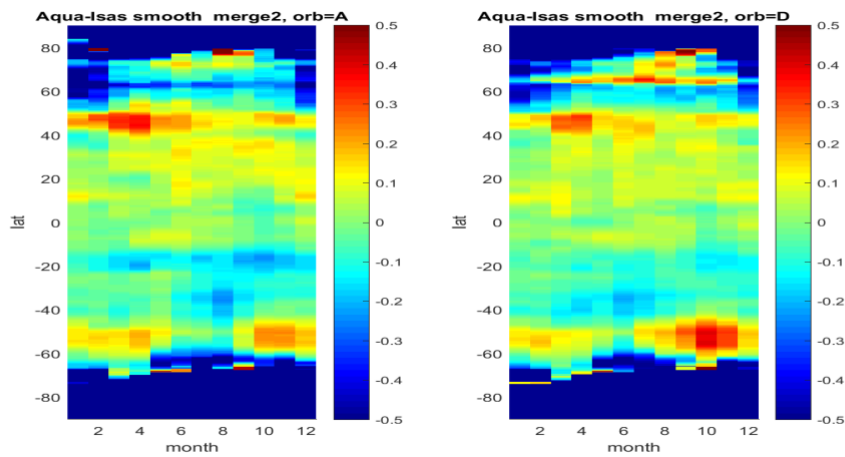

Figure 45: Latitudinal correction applied for Aquarius as a function of month and latitude. This correction is given at an average SST. Left, ascending orbits, right, descending orbits.
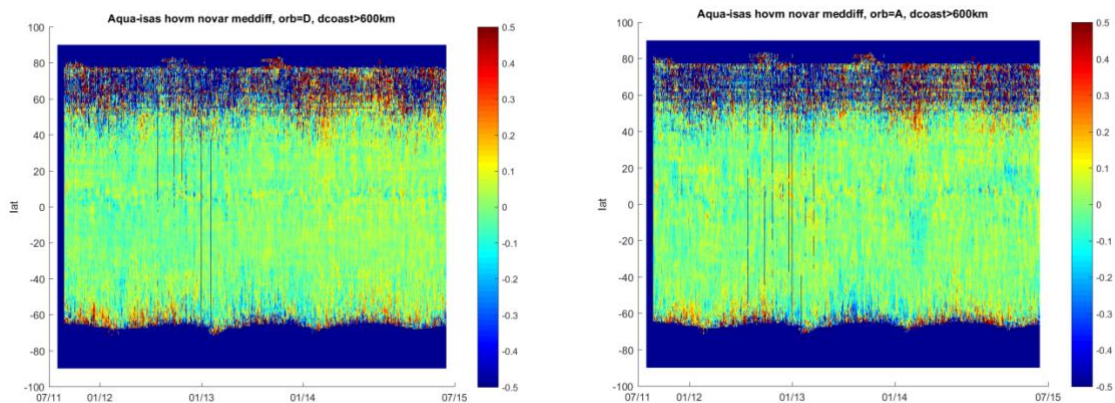

Figure 46: Hovmöller of SSS Aquarius - SSS ISAS. SSS Aquarius corrected for SST-dependent bias and latitudinal seasonal bias. Left, ascending orbits; right, descending orbits.
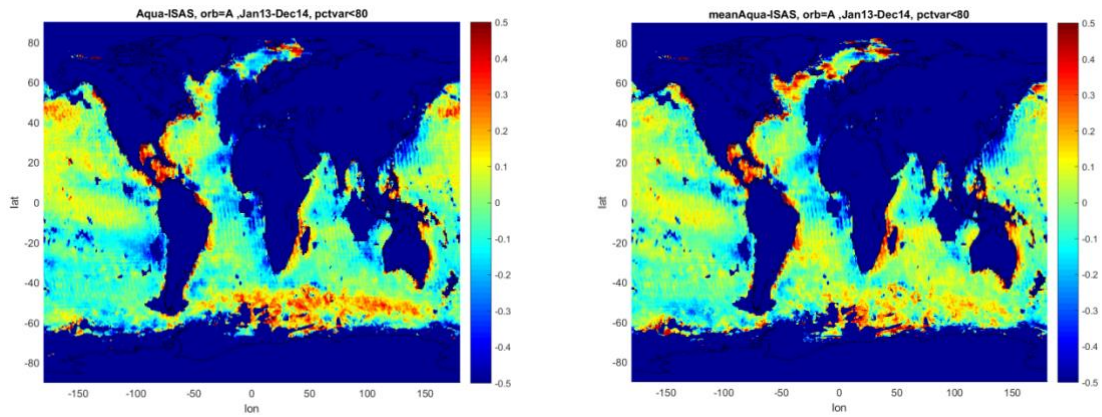
Figure 47: Aquarius SSS – Isas SSS over 2013 and 2014. Left, uncorrected SSS; right, corrected SST+latitudinal correction. Ascending orbits
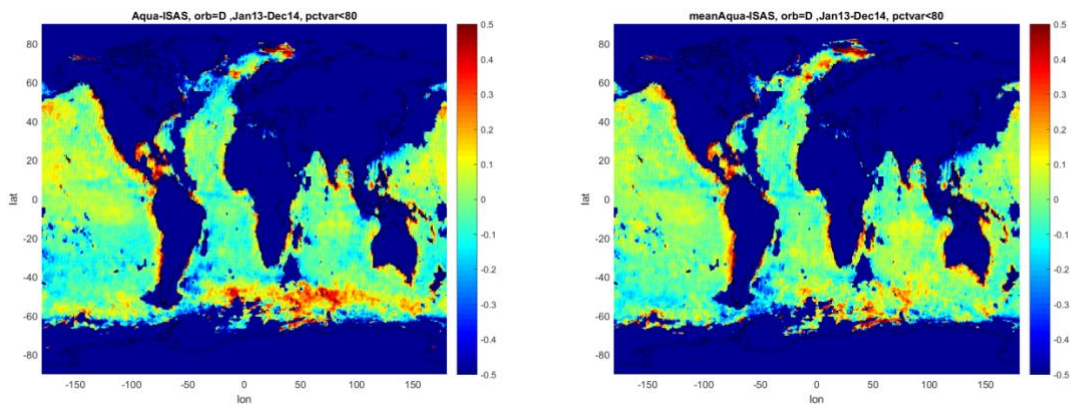

Figure 48: idem Figure 47 for descending orbits.

4.3.3.7.2 flagging the data

Aquarius L3 products are already filtered. No extra filtering is applied before merging.

5 L3 and L4 uncertainty budget

Main CCI products are L4 SSS products estimated at a weekly and a monthly resolution (with a daily and 15-day sampling, respectively). In these products, uncertainty estimation is performed using a classical least square uncertainty budget. The L2 SSS input random uncertainties are presented in this document's first part for SMOS, SMAP and Aquarius sensors. The relative systematic uncertainties are estimated simultaneously with the monthly SSS L4 estimation. The SMOS, SMAP and Aquarius latitudinal systematic uncertainties are corrected before the L4 merging.

Using L3 data allows for comparing global maps for different months and orbit types. The project provides specific L3 data with corrected and uncorrected SSS. These data are used by the validation team which is in charge the product assessment.

6 Conclusions and the way forward

Phase 2 CCI+ SSS project is dedicated to SMOS-SMAP-AQUARIUS synergy. The merging of different products from different sensors (radiometer for SMAP and AQUARIUS, interferometer for SMOS) shall consider a realistic random and systematic uncertainty model.

The self-consistency between satellite SSS measured by the various sensors and under various geometries has been used for correcting and/or estimating systematic and random uncertainties.

Due to remaining seasonal biases in CCI+SSS v3.2, systematic latitudinal uncertainties have been estimated for SMOS, SMAP and Aquarius at the beginning of phase 2 (in phase 1, only SMOS seasonal latitudinal biases have been corrected). Part of the seasonal biases has been removed but some inter-annual residuals remain.

Specific RFI correction algorithms using PCA have been implemented and tested. This kind of correction allows taking into account the time variation of the RFI intensity. It will be generalised in the previous version.

End of document