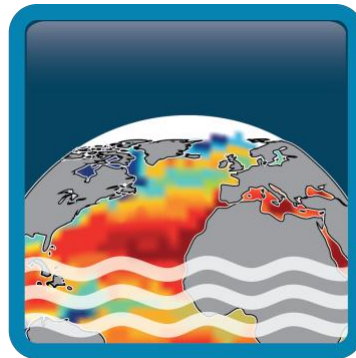# Climate Change Initiative+ (CCI+) Phase 1

Sea Surface Salinity

Algorithm Theoretical Development Basis Document **(**ATBD**)**

Customer: **ESA**

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

Version: v1.3

Ref. internal: AO/1-9041/17/I-NB_v2r2

Revision Date: 19/12/2019

Filename: SSS_cci-D2.3-ATBD-v1.3.docx

Signatures

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

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

Amendment Record Sheet

Date / Issue	Description	Section / Page
15-07-2019 / v1.0	Delivery to ESA	New Document
22-10-2019 / v1.2	Update reference document documents section	Section 1.3.2 / page 11
	Updated acronym table	Section 1.4 / page 12
	Improved text in captions of Figures: 5-9	
	Rephrasing main text	Section 4.3.3 / page 22
	Updated text in caption of Figure: 5-9	Section 4 / page 18 to 36

Table of Contents

List of figures

List of tables

1   Introduction

1.1   Scope of this document

This document holds the Algorithm Theoretical Development Basis Document (ATBD) prepared by CCI+SSS team, as part of the activities included in the [WP230] of the Proposal (Task 2 from SoW ref. ESA-CCI-PRGM-EOPS-SW-17-0032).

1.2   Structure of the document

This document (ATBD v1.3) is composed of 4 sections and presents the CCI+SSS algorithms implemented during year 1. Section 1 is an introduction presenting the scope, reference and applicable documents, acronyms, and the structure of the ATBD. Section 2 present the algorithms of the so-called Level 2 products which are swath retrievals from L-band sensor SMOS and SMAP. Section 3 presents the Level 3 SSS products, which are averaged intermediate products obtained sensor by sensor, without mixing inter-sensor information. Level 4 data set is foreseen to be produced each year and form the core of the CCI+SSS products for year 1 and is described in §4. In Section 5, we provide summary and discuss the next steps.

1.3   References

1.3.1   Applicable Documents

ID	Document	Reference
AD01	Sea Surface Salinity Climate Change Initiative Phase 1 Data Access Requirement Document.	SSS_cci-D1.3-DARD-v1r4
AD02	SMOS Level2 Algorithm Theoretical Baseline Document (ATBD). Available at: https://earth.esa.int/documents/10174/1854519/SMOS_L2OS-ATBD	SO-TN-ARG-GS-0007_L2OS-ATBD v3.13
AD03	CATDS (2017). CATDS-PDC L3OS 2P Algorithm Theoretical Basis Document. Available at: https://www.catds.fr/content/download/78841/file/ATBD_L3OS_v3.0.pdf	ATBD_L3OS_v3.0
AD04	Aquarius Official Release Level 2 Sea Surface Salinity v5.0 ATBD. Available at: ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/docs/v5/	RSS Technical Report 120117
AD05	Aquarius Official Release Level 3 Sea Surface Salinity v5.0. Aquarius L2 to L3 Processing Document.ATBD. Available at: ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/docs/v5/	AQ-014-PS-0017_Aquarius_L2toL3ATBD_DatasetVersion5.0

ID	Document	Reference
AD06	RSS SMAP Level 2 Sea Surface Salinity V3.0 40km Validated Dataset. Available at: ftp://podaac-ftp.jpl.nasa.gov/allData/smap/docs/V3/	RSS Technical Report 101518
AD07	Sea Surface Salinity Climate Change Initiative Phase 1 Product Specification Document	SSS_cci-D1.2-PSD-v1r6
AD08	Sea Surface Salinity Climate Change Initiative Phase 1 Algorithm Theoretical Development Basis Document	SSS_cci-D2.5-PVP-v1.0

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	Liang Hong, Normal Kuring, Joel Gales and Fred Patt (2017), AQ-014-PS-0017_Aquarius_L2toL3ATBD_DatasetVersion5.0	
RD04	Fred Patt, Liang Hong (2017), AQ-014-PS-0018_AquariusLevel2specification_DatasetVersion5.0	
RD05	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.	www.remss.com/missions/smap, doi:10.5067/SMP20-2SOCS
RD06	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
RD07	Yiwen Zhou ; Roger H. Lang ; Emmanuel P. Dinnat ; David M. Le Vine (2017), L-Band Model Function of the Dielectric Constant of Seawater, *IEEE Transactions on Geoscience and Remote Sensing* (Volume: 55 , Issue: 12)	
RD08	Gaillard F. (2015), ISAS-13 temperature and salinity gridded fields. SEANOE.	http://doi.org/10.17882/45945.

1.4 Acronyms

AD	Applicable Document
ATBD	Algorithm Theoretical Basis Document
Aquarius	NASA Sea Surface Salinity mission
CCI	The ESA Climate Change Initiative (CCI) is formally known as the Global Monitoring for Essential Climate Variables (GMECV) element of the European Earth Watch programme
CCI+	Climate Change Initiative Extension (CCI+), is an extension of the CCI over the period 2017–2024
CMEMS	Copernicus Marine Environmental Monitoring Service
DARD	Data Access Requirements Document
DOI	Digital Object Identifier
DPM	Detailed Processing Model
ECMWF	European Centre for Medium Range Weather Forecasts
EASE	Equal-Area Scalable Earth (EASE) Grid
ECV	Essential Climate Variable
EO	Earth Observation
FOV	Field Of View
Hs	Significant Wave Height (see also SWH)
ISAS	In Situ Analysis System
KS	Klein and Swift sea water dielectric constant model
MW	Meissner and Wentz sea water dielectric constant model
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NOP	Numerical Ocean Prediction
NWP	Numerical Weather Prediction

OTT	Ocean Target Transform
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SWH	Significant Wave Height (see also Hs)
TBC	To Be Completed
UCR/CECR	Uncertainty Characterisation Report (formerly known as the Comprehensive Error Characterisation Report)
URD	User Requirements Document
VOS	Volunteer Observing ships
WS	Wind Speed

Climate Change Initiative+ (CCI+)
Phase 1
Algorithm Theoretical
Development Basis Document

Ref.: ESA-CCI-PRGM-EOPS-SW-17-0032
Date: 19/12/2019
Version : v1.3
Page: 15 of 41

2   Level 2 Algorithms

In year 1 of the CCI+SSS project (June 2018 – June 2019), the production of the first CCI+SSS dataset used the input data levels 2 (i.e. non averaged retrieved SSS along each satellite orbit) for SMOS and SMAP or Level 3 (single instrument averaged SSS) for Aquarius sensor. These data are all projected on the same EASE grid at a spatial sampling of 25 km (see section §3.4.1). The L2 products used for each sensors come directly from official space agency dedicated centres (e.g., CATDS, RSS). They are therefore not generated by the CCI+SSS processing chains and are the following for each sensor.

2.1   SMOS Level 2 Algorithm

In year 1, we directly use as input for SMOS data the L2 products generated internally by the CATDS. These are the so-called L2P products based on SMOS ESA L2 v622 algorithm [AD.2]. These products are provided onto an EASE grid at 25km resolution and are classified/filtered according to an ensemble of quality flags. The data are split into ascending and descending products. SSS is also classified according to the distance to the sub-satellite track. The ATDB is not reproduced here as it is already detailed in [AD.3].

2.2   SMAP Level 2 Algorithm

In year-1, we directly use as input for SMAP sensor the RSS Level 2 v3.0 products. These are swath SSS from SMAP provided daily at 40 km resolution. The data are split into ascending and descending products and between the fore and aft views. Details on the processing algorithms can be found in [AD.6]. The ATDB is not reproduced here as it is already detailed in [AD.6].

3   Level 3 algorithms

3.1   Introduction

The level 3 (L3) products are, by definition, time and space-averaged products obtained sensor by sensor, without mixing inter-sensor information. Here, we consider simple averages of swath Level 2 SSS products, which may have been already corrected for some biases (e.g. land sea contamination or spatio-temporal drifts corrections in the brightness temperatures of the instrument such as the Ocean Target Transformation). These products can thus be used as a reference in terms of observed SSS variability since we don't apply to the observed SSS any specific smoothing operation (for example, by allowing introducing representativity errors or variance filtering).

A priori, these products will not be distributed to users in year 1. They will be made available as part of the validation and verification of products and will serve to build up the Level 4 products.

3.2   Input Level 2 data

Input Level 2 products are the swath L2 SMOS and SMAP data generated by the algorithms described in §2.1 and §2.2, respectively.

3.3   Input Aquarius Level 3 products

For Aquarius, we use as input the official release products L3 v5.0, which is the official end of mission public data release from the AQUARIUS/SAC-D mission (with DOI: 10.5067/AQR50-3SADS and which are accessible here). Aquarius Level 3 sea surface salinity standard mapped image data contains gridded 1 degree spatial resolution SSS averaged over daily, 7 day, monthly, and seasonal time scales. We use the daily average dataset for generating the CCI+SSS L4 dataset of year 1. An average of ascending and descending products over the 3 radiometer footprint is performed. The ATBD for these Aquarius L3 products is detailed in [AD.4] and [AD.5] and therefore, not reproduced here.

3.4   Methods

3.4.1   Input Data re-gridding

While the SMOS Level 2 data are provided onto the EASE grid at 25 km resolution, the SMAP Level 2 data and the Aquarius Level 3 products are not given on that grid. A first step therefore consists in interpolating both the Aquarius and SMAP SSS onto the EASE grid at 25 km resolution. A bilinear interpolation scheme is used for that purpose.

3.4.2   Space-time aggregation

Assuming that, for a given grid node and over the duration of a month (in case of monthly average), we have a set of level 2 salinities SSS_i affected by errors σ_i^2, the averaged salinity is obtained as follows:

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

This average is calculated for each grid node and on a daily or decade temporal sliding window.

The following error is assigned to the average product:

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

3.5   Products

The monthly L3 products contain the averaged SSS field and the associated error for each L-band sensor: SMOS, Aquarius and SMAP.

4 Level 4 algorithms

4.1 Introduction

The approach adopted to generate Level 4 CCI+SSS products is as follows:

- The bias correction exploits as much information as possible from the data. SSS that seem to be affected by various contaminations (coastal, RFI, galactic, solar, etc.) are kept. It is considered that these effects can, to some extent, be corrected a posteriori, since in most cases these effects lead to systematic biases that can be confused with real geophysical signals. In fact, in some cases, satellite data provide information on strong geophysical signals that can reach several units in the Practical Salinity Scale (pss). In this situation, a bias in the order of the pss does not justify the removal of the data.

- The self-consistency of the measurements (averaged in a monthly time window) over the whole time period (2011-2018), and taking into account the natural variability of the SSS expected in this window, allows the different inter-sensor biases to be corrected relatively to each other. A posteriori 3 sigma filters is then applied to remove outliers. These filters are applied with respect to the natural SSS variability that must be taken into account in the satellite SSS estimation process. Indeed, if a low variability is expected (in comparison to SSS L2 retrieval error), the filters applied must be more severe. Otherwise (e.g., at river mouths or in strong currents where variability is high), data that differ significantly (from more than 3 times the retrieval error on the SSS) from the mean should be retained.

- The self-consistency criteria considered in the algorithms is temporal. A spatial correction of the SSS according to a certain reference (e.g. WOA climatology) could affect spatial and temporal dynamics and could remove some of the interannual signals and mesoscale signatures. We therefore consider coastal/ocean biases that are constant over time. These biases can be corrected without affecting geophysical SSS dynamics. In practice, the SSS correction/estimation is done grid node per grid node considering the inter-sensor self-consistency in SSS. To correct seasonal latitudinal biases, a relative correction is finally also applied, similar to what is described in Boutin et al (2018). It applies to all basins and should not affect the interannual dynamics.

- The different corrections are relative. As a result, SSS anomalies are available at the end of the correction processing. These anomalies are then calibrated against an absolute reference. As far as possible, this shift should be a time-independent correction in order to maintain the temporal dynamics of the SSS. In some cases (e. g. high latitudes), coastal biases are not constant over time due to variations in ice edges. A specific processing shall be found in these areas and will be the subject of future studies.

- In the settings of the various processing parameters (time correlation length, a priori variability), either spatially smooth fields or slowly-evolving time fluctuations are estimated to reduce errors as much as possible. This is why the CCI+SSS L4 products are split into two sub-products: a monthly product and a weekly product.

The aggregation of the L2 products from the different sensors requires a step of homogenization of the data. This homogenization affects the following points:

- qualification of seasonal latitudinal biases
- estimation of representativity errors.

4.2   Input data

The input data are L2 or L3 from the different sensors (L2 for SMOS and SMAP, L3 for Aquarius). These data are all projected on the same EASE grid at a spatial sampling of 25 km (see section §3.4.1). In year 1, the L2 or L3 products used for each sensor come directly from official space agency dedicated centres (e.g., CATDS, RSS, NASA). They are therefore not generated by the CCI processing chains and are the following for each sensor:

- For SMOS: we use the CATDS products described in §2.1
- For Aquarius: we used the official release products L3 v5.0 described in section §3.3
- For SMAP we use the RSS L2 v3.0 products described in §2.2

Note that the Aquarius and SMAP products are corrected from latitudinal temporal biases based on the ARGO products in the course of their L2 processing (see [AD04] and [AD06])

4.3   Method

4.3.1   Bias estimation

The biases affecting the data are of different kinds, which include:

- Instrumental
- Related to errors in the forward emissivity models used for the retrieval (mainly dielectric constant at high latitudes, roughness corrections, etc…).
- Linked to biased auxiliary geophysical data
- Related to measurement contamination (Brightness temperatures) by anthropogenic sources (e.g. RFI).

One of the important sources of bias is related to the contamination of the instrument side lobes around the coasts (for real aperture radiometers) or by the land signal in the reconstruction of oceanic scene (for the SMOS interferometric radiometer, see instrument characteristic description in the E3UB document, section 2).

The analysis of biases and errors (see E3UB, section 4) shows that:

- There are sources of constant bias over time which vary as function of the location on the globe, the instrument and the geometry of observation.
- There are sources of bias that are not constant over time and depend on latitude.

For these reasons, we propose to model the bias as a sum of two components.

We consider that the latitudinal (orbital) biases are independent of the basin (Atlantic, Pacific or Indian Ocean) and that they apply in addition to the coastal/land biases.

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

$$\text{SSS}_{obs}(X, t, orb, lat, lon) = \text{SSS}(t) - bc(X, orb, lat, lon) - bl(X, orb, t\_month, lat) \qquad \text{Eqn 4-1}$$

with

- bc, the coastal bias
- bl the latitudinal bias
- SSS_{obs} is the observed salinity,
- SSS(t) is 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 considered or, in the case of SMOS, the across swath location of the measurement (X_{swath}).
- orb is indicating the pass directions (asc/desc)

bc and bl can be estimated independently starting with the estimation of bl on open sea areas at grid nodes sufficiently far away from the coasts. Then, a latitudinal correction is applied to the coastal pixels. From these latitudinal bias corrected data, we estimate bc. Since the number of independent subsets of data is relatively large, the different biases can be estimated in a self-referenced way, i.e. there is no need for an external reference when considering SSS anomalies rather than absolute salinities. Note that the Eqn 4-1 requires a simultaneous estimation of SSS(t) (or anomalies with respect to an arbitrary reference salinity given by the measurements themselves) and of the biases bc and bl since for the CCI L4 products, we do not use an external reference that gives us SSS(t). The arbitrary reference is taken as the SMOS SSS at the center of the track. At the end of the correction process, all SSS calibrated with respect to this reference are adjusted to the 2011-2018 ISAS SSS mean.

The L4 products, represented by SSS(t) are therefore estimated together with the biases in each product. The error propagation occurs at the time of this estimate using 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 each sensor.

4.3.2 SSS and bias estimation: main processing steps.

Figure 1 summarizes the different processing steps of the algorithm including the generation of intermediate products.
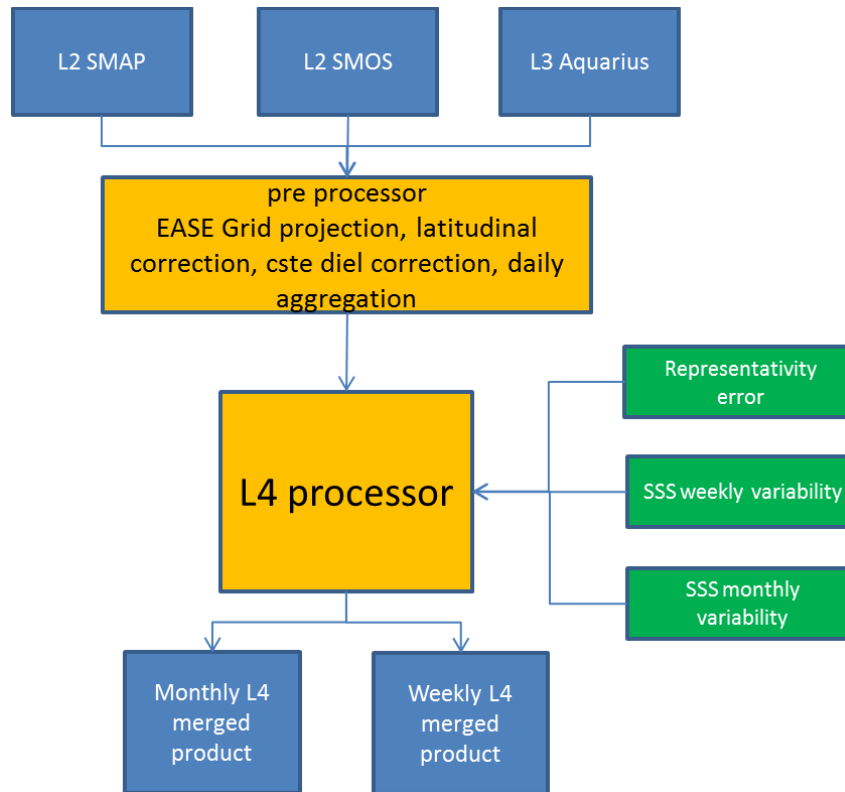
Figure 1: main L4 processing steps.

The algorithm steps to perform the corrections in order to estimate the unbiased SSS are as follows:

1. Correction of the bias related to the modelling of the L-band sea water dielectric constant at low SST (section 4.3.3).
2. Correction of the latitudinal biases if necessary (section 4.3.4). We found that latitudinal bias is low for SMAP and Aquarius data that are already latitudinally adjusted via ARGO in-situ data (see E3UB). For SMOS data, a seasonal correction is applied that depends on the latitude.
3. Estimation of the inter-sensor biases and monthly SSS (section 4.3.5). The inter-sensor biases are considered constant for each month and evaluated assuming SSS varies slowly over a month (Figure 2). This computation is carried out by an optimal interpolation whose cost function is described in section 4.3.7.
4. Estimation of errors of the monthly SSS (section 4.3.9). Computation of outliers, std of bias...etc.
5. Correction of individual SSS and computation of a weekly-averaged SSS field. In this step, the bias correction is fixed and the 30-day SSS field is taken as a priori. We estimate fluctuations around this monthly field to achieve a time resolution of 7 days. This computation is carried out by an optimal interpolation whose cost function is described in section 4.3.7.

6. Estimation of errors of the weekly SSS (section 4.3.9).

7. Absolute calibration of salinities (section 4.3.8).
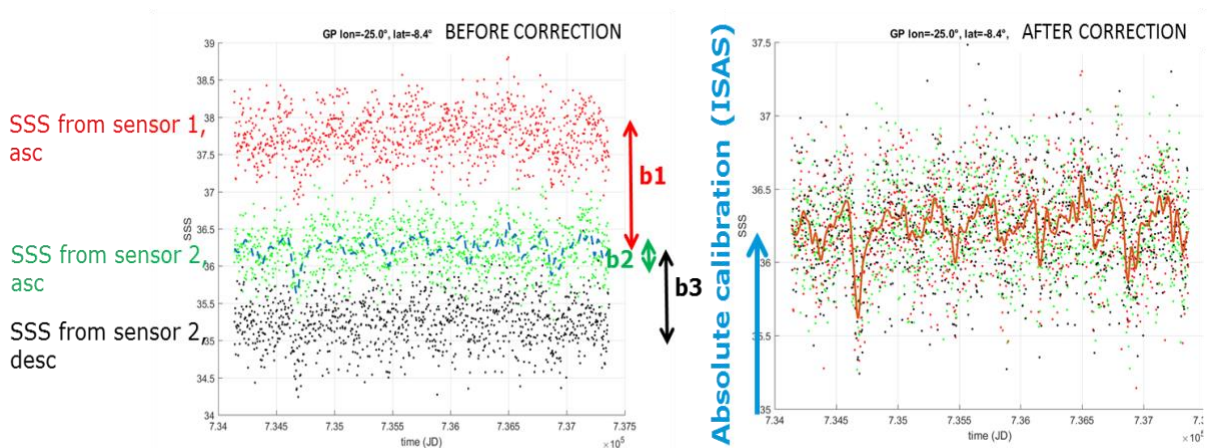
Figure 2: principle of the self-consistency approach. Example of a grid point processing and time series analysis. The different time series coming from three different sensors or acquisition types are corrected from b1, b2, b3 biases. Left: observed SSS; right: corrected SSS. Red curve shows the mean SSS obtained from the measurements (clouds of points).

4.3.3   Sea Water dielectric constant model correction

L2 and/or L3 data are first filtered to discard outliers. In the case of SMOS, an empirical correction of the retrieved SSS at low SST is applied to correct for Klein and Swift (KS) dielectric constant model biases at low temperatures (Dinnat et al. 2017). A filter is also applied to remove data which exhibit a cost-function retrieval Chi-parameter (given by the L2 products) which is strictly larger than 3.

The correction related to the dielectric constant has been empirically derived from a fit of Figure 7 from Zhou et al. 2017 as follows:

$$bias(SSS)=0.0136.SST^2-0.2553.SST+1.1874.$$

It must be subtracted and applied over the SST interval from -2°C to 8.5 °C.

The effect is significant at high latitudes as shown in Figure 3.
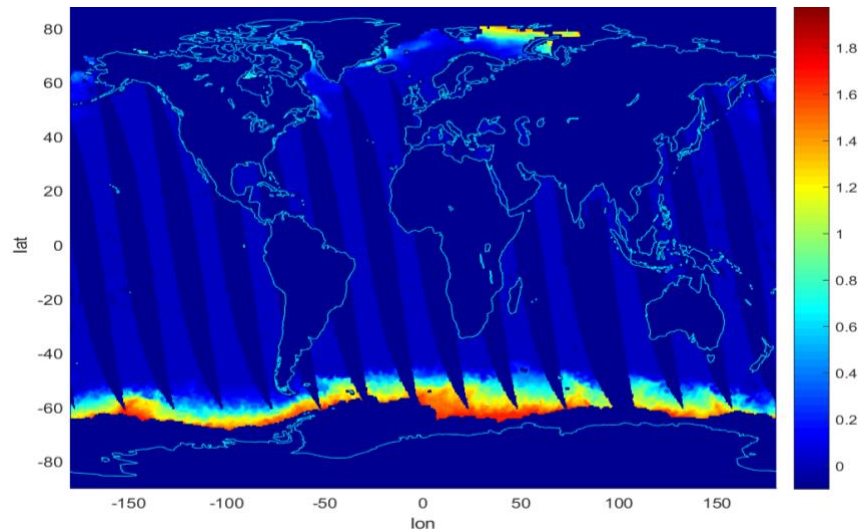
Figure 3 : Example of bias correction due to dielectric constant. Correction reaches 1.5 pss close to ice edge.

This bias is considered to have been at least partially corrected in the SMAP and Aquarius data. SMAP and Aquarius algorithms use an intermediate dielectric constant model between KS and Meissner et Wentz (MW). The MW model behaves better than the KS model at low temperatures.

An update of the dielectric constant model is under study that will be deployed in future versions.

4.3.4   Seasonal latitudinal correction

A seasonal latitudinal correction is further applied to the SMOS data to ensure best consistency between SMOS, SMAP and Aquarius data. The latitudinal seasonal correction applied to the SMOS data depends on the across track location and the orbit pass direction (ascending or descending). The purpose is to estimate the term bl in Eqn 4-1.

Further than 800 km from the coastline, land-sea contamination is not detectable but seasonally-varying latitudinal biases are still observed. For SMOS data, the latter mostly depend on x_{swath}, orb and the month of the year. At first order, the systematic errors are independent from the year as shown by two examples over the Pacific Ocean on Figure 4

Figure 4 : Examples of seasonally-varying latitudinal systematic error. SSS averaged over the Pacific Ocean further than 800 km from coast: green: ISAS, blue: SMOS ascending orbits; red: SMOS descending orbits: a-d) November; middle of the swath (0-50 km from the center of the swath); e-h) January; edge of the swath (350-400 km from the center of the swath); a & e) 2011; b & f) 2012; c & g) 2013; d & h) 2014.

bl (Eqn 4-1) is determined separately for ascending and descending orbits, on a monthly basis, and is assumed to be independent of the longitude and of the year. We therefore negle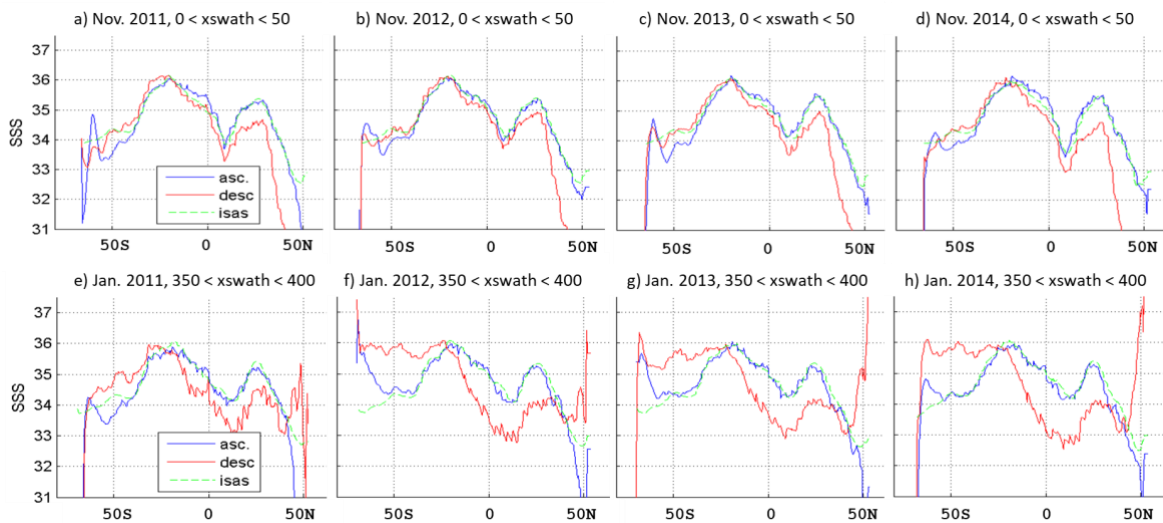ct interannual variations that could result from variation in sun activity, as they appear to be an order of magnitude smaller than the seasonal biases. The correction is estimated from Atlantic Ocean orbits further than 800 km from continental coasts, in order to avoid land-sea contamination (bc in Eqn 4-1 vanishes in this case) and because it is possible to reach very high latitude in comparison with Pacific Ocean.

4.3.5   Static bias correction

At this stage and after SMOS latitudinal correction (section 4.3.4), it is considered that there are still possible coastal biases affecting the SMOS-SMAP-Aquarius data. These biases are considered constant over time and specific to each type of acquisition (term bc in Eqn 4-1). The types of acquisition include: the orbit direction (ascending/descending, except for the L3 Aquarius) and the geometry of acquisition (across-track location for SMOS and fore/aft viewing direction for SMAP). The correction of relative biases assumes that sensors observe the same SSS on average over 2010-2018. This makes it possible to co-locate the SSS of the different sensors and to assume that, on average, the SSS should be the same (Figure 2).

SSS fields are derived as follows:

-first, for each grid node, a monthly SSS(t) is calculated for each grid node and over the observation time interval (2010-2018). The relative inter-sensor temporal biases are calculated (see section 3.3.2). Daily SSS(t) anomalies are then obtained over the observation interval. Daily

interpolation and estimation of coastal biases are finally carried out as part of an optimal analysis applied independently for each grid node. In particular, we take into account:

- o data errors
- o SMOS-SMAP-Aquarius representativeness errors
- o the monthly variability calculated from the SMOS SSS.
- o The cost function integrates all these covariances (see section 3.3.2).

In addition, a 3 sigma filtering is applied to avoid outliers. We proceed in 2 steps:

1) first estimation of the SSS and biases and 3-sigma filtering,

and,

2) then a second estimation of the SSS and biases is performed after filtering.

-Finally, the L2 SSS are corrected for biases and the weekly SSS is estimated. The weekly SSS is obtained by adding weekly fluctuations to the average monthly field. The introduction of 7-day to 30-day variability in the cost function (through representativity error covariance) allows to limit noise on areas with low natural variability. A new 3sigma filtering is applied before the SSS weekly retrieval.

4.3.6   Representativity errors and natural variability.

In the optimal interpolation, we consider that the SSS field has some monthly variability. This variability is actually an RMSD estimated from an independent SSS global average on each grid node. The variability is seasonal and the SSS is allowed to fluctuate more or less around its mean value with the seasons.

In order to obtain monthly variability at all points and all seasons, we used SMOS filtered SSS fields for low latitudes and ISAS SSS fields [RD08] at the highest latitudes or at some points on the globe where RFI contamination is intermittent. This variability has been increased as detailed below in order to always leave the possibility of unexpected fluctuations.
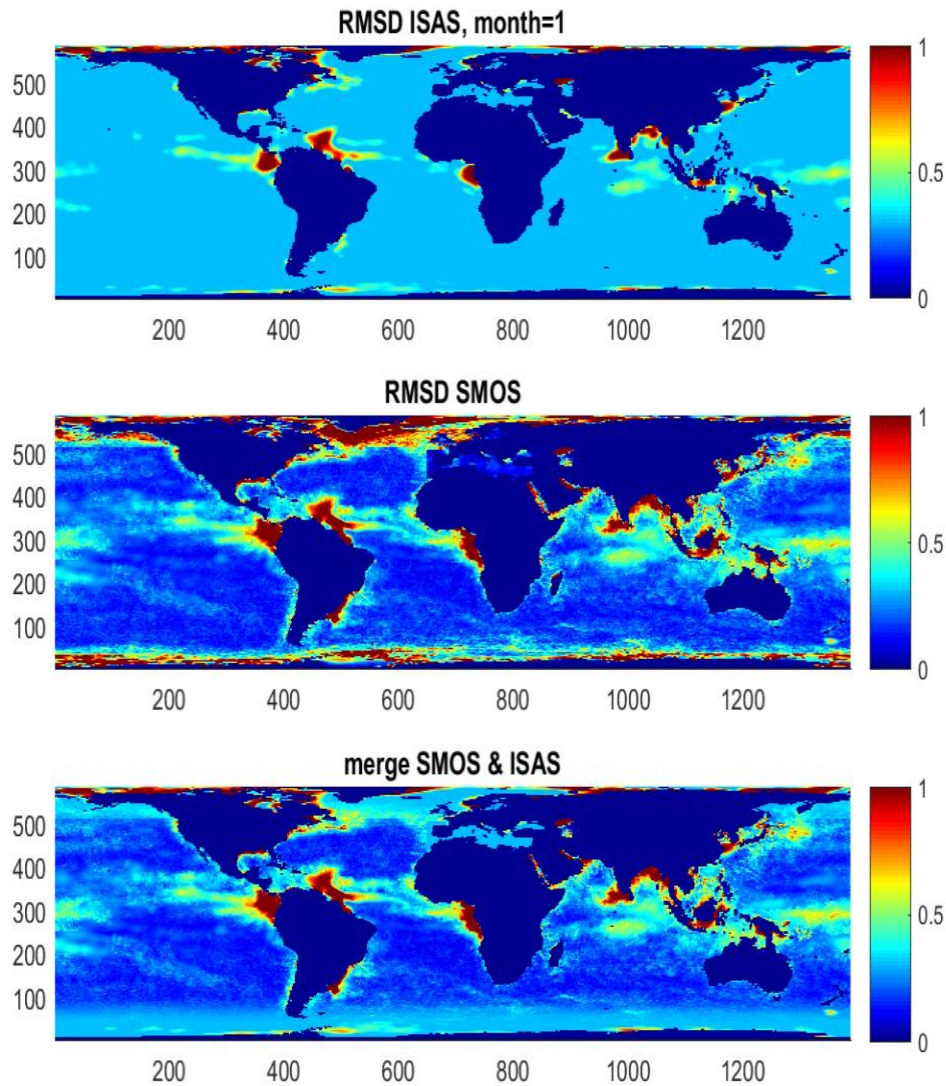
Figure 5: SSS RMSD obtained from the merged SMOS and ISAS information. January month. x and y axis units in pixel number for longitude and latitude respectively.

For weekly CCI fields, we start from the monthly average fields. We estimate the weekly fluctuations around these monthly average fields. Mercator model simulated SSS are used to calculate the relative seasonal variability between monthly and weekly salinity fields (Figure 6).

Figure 6: SSS variability between SSS fields taken at 50km/7days and 50km/30days. January. From Mercator 1/12th of degree. x and y axis units in pixel number for longitude and latitude respectively.
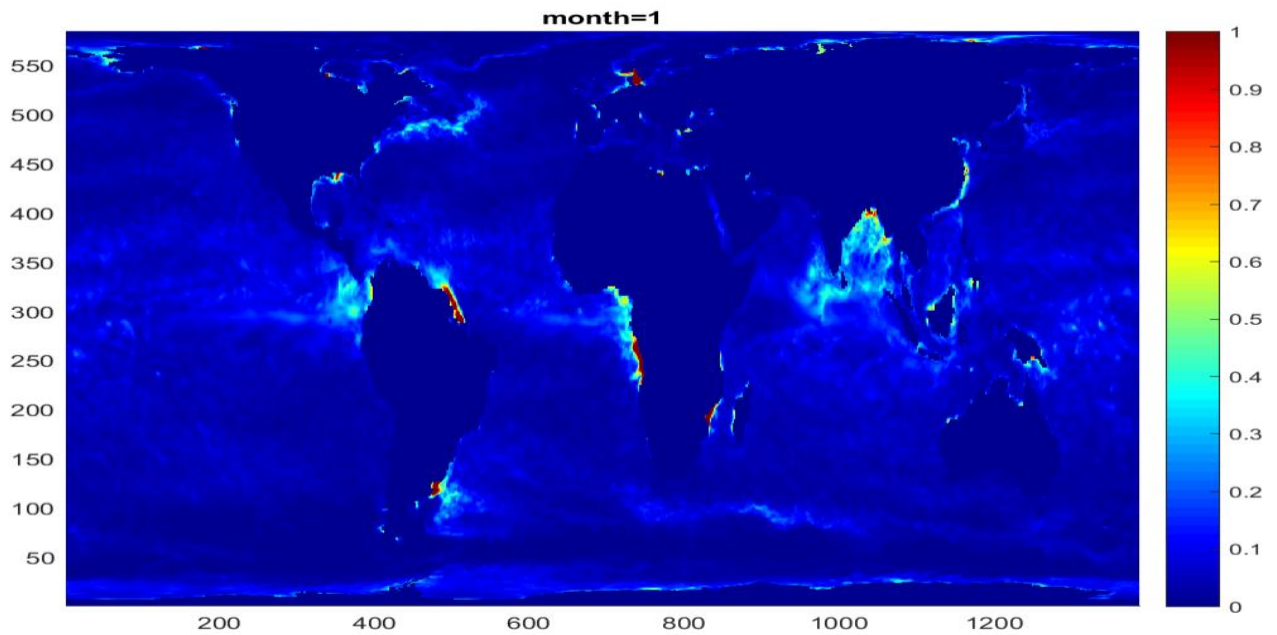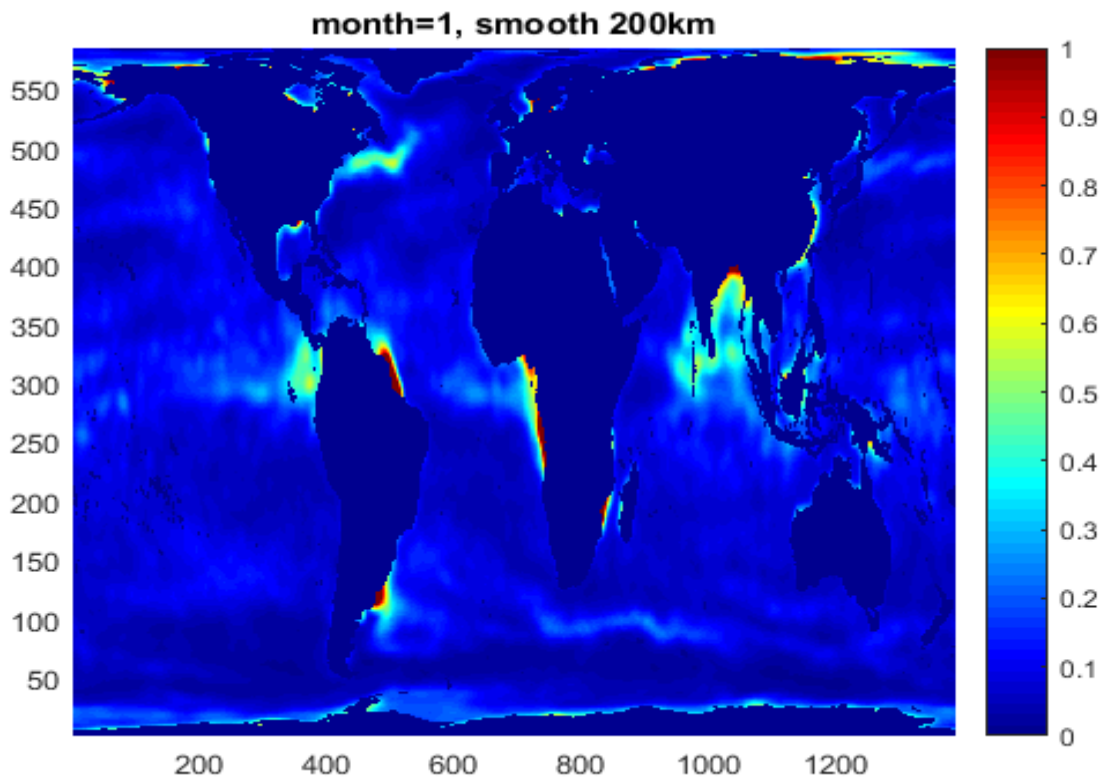

Figure 7: same than Figure 6 after rescaling and spatial smoothing.x and y axis units in pixel number for longitude and latitude respectively.

Figure 6 shows the variability in January between a 50km/7-day field and a 50km/3-day field obtained from the Mercator fields at 1/12th of a degree. After some testing, it appears that this variability is underestimated in river mouths and plumes. Since river discharges used in Mercator model are based on river discharge climatologies, interannual variations in runoff flows are not properly taken into account. In fact, we have tried to spatially smooth this variability over a length of 200km. However, smoothing introduces a significant decrease in amplitude. We therefore decided to multiply the variability by a factor 2 and then smooth this variability over a length of 200km. Figure 7 shows the variability map actually used in the L4 processing.

In principle, when averaging data from different sensors, it makes sense if the acquisitions have the same spatial and temporal resolution.

The temporal resolution of the Aquarius product we used (7days) is much longer than the temporal resolution of SMOS and SMAP L2 SSS (on the order of 2 seconds that corresponds to the SMOS acquisition time) so that a representativity error corresponding the the SSS variability within 7 days should be taken into account. However, this representativity error remains low, partly because most SSS structures at 150km resolution do not vary much within 7days, compared to SMOS and SMAP measurement errors (representativity errors are quadratically added to measurement errors).

The spatial resolution of Aquarius SSS is 150km, 50km for SMOS and SMAP, hence we must take into account a representativity error for Aquarius, corresponding to the variance of the SSS between 50 and 150km^2. The aim is to increase the error associated with the Aquarius data by a representativity error that can be calculated from a model such as Mercator. This information, calculated off-line, is used in the cost function. Note that this representativity error depends on the season and dynamics of the SSS field (Figure 8). Indeed, in some parts of the world, the SSS does not vary over large distances and durations: in this case, the error of representativity tends towards 0. Here again, tests have shown that the error of representativity is not sufficient, particularly near the coast. We proceed as for the weekly variability: the variability was doubled and spatially smoothed.
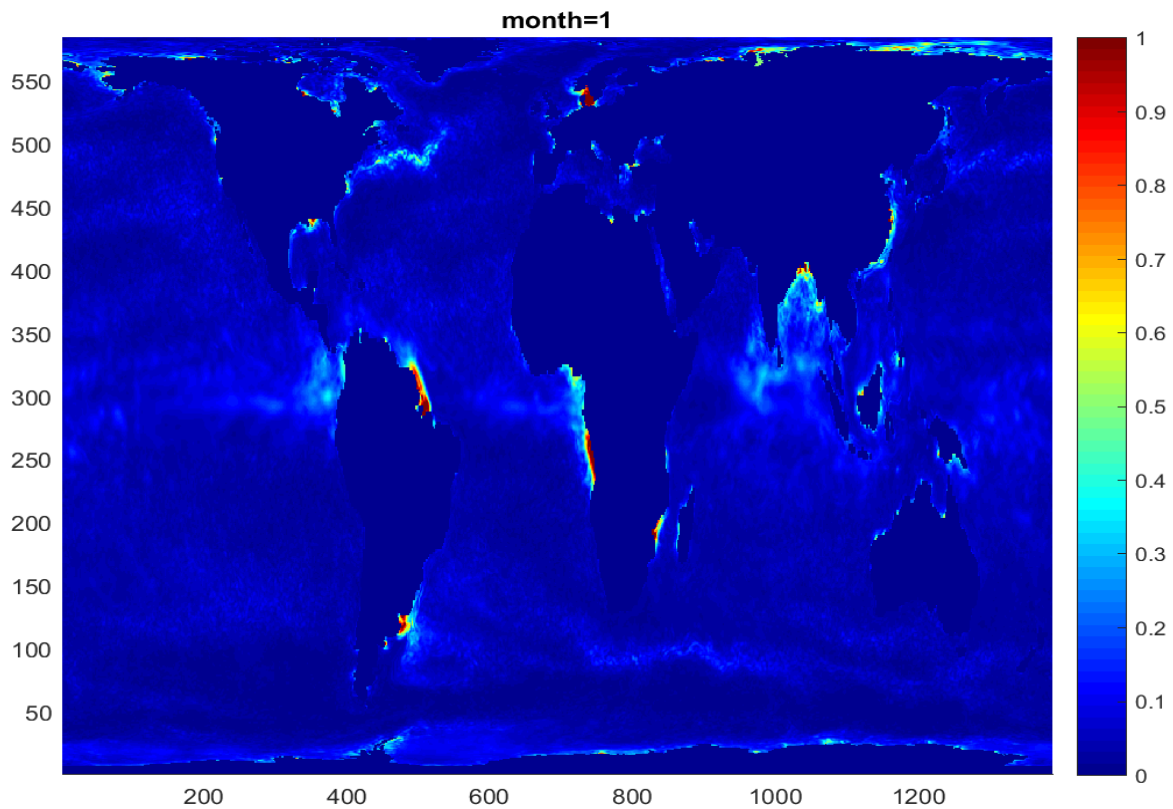
Figure 8: representativity error Aquarius vs. SMOS (150km/7jours Aquarius, 50km/30jours). January. From Mercator 1/12th of degree. x and y axis units in pixel number for longitude and latitude respectively.

Representativity errors are generally low (compared to observational errors) for grid nodes in the open sea and have significantly high values near river mouths. Moreover, near rivers, large interannual variations show that a Gaussian statistical approach leads to some drawbacks and the estimation of fluctuations in these areas can be approximate.

In conclusion:

- representativity errors are generally low in the open sea (compared to measurement errors) except in some areas of high gradients and in rainy areas.

- close to the coast and at the rivers mouths, these errors are difficult to assess because of the high interannual variability that puts the statistical approach into default. These errors can be greater than 1 pss, which becomes dominant in the error budget (SMOS L2, SMAP L2 and Aquarius L3 errors are, at an SST of 25°C, about 0.8, 0.7 and 0.3 (at the equator) pss, respectively. Under these conditions, it can be seen that Aquarius data are much more affected by representativity errors than SMOS and SMAP data.

The algorithm take into account Aquarius representativity errors due to the difference in spatial resolution and the low relative amplitude of Aquarius errors.

4.3.7   Cost function

4.3.7.1   Introduction

In order to estimate the SSS at a given time, the algorithl is using an optimal interpolation. This interpolation is applied grid node per grid node, without spatial smoothing. Indeed, we want to preserve the temporal dynamics of the SSS and to avoid a spatial catch-up made to a reference climatological field (this catch-up would then be done month by month) for which the temporal dynamics of the observations are reduced to the temporal dynamics of the reference field, which might remove important interannual variations.

To better understand the process, let's consider SSSobs observations from a single sensor. First, the bias is ignored. To estimate a time series SSSest of spatial resolution R1 and temporal resolution T1, knowing that the observed data SSSobs are at spatial resolution r1 and temporal resolution t1, the cost function to be minimized is written (as a scalar product <X|Y>):

$$C(SSS) = < SSSobs - SSS | C1^{-1} \cdot (SSSobs - SSS) > +$$
$$< SSS - SSSmoy | C2^{-1} \cdot (SSS - SSSmoy) >$$

Eqn 4-2

The SSSmoy field is the SSS field a priori that is given and which, in the case of the monthly SSS estimate, is constant and corresponds to the average of the SSS of the SMOS central dwell line. For the weekly SSS, it is the monthly SSS obtained in the previous step (the monthly SSS estimate) from SMOS, SMAP and Aquarius data. Note that Eqn 4-2 only applies if the time sampling of the data is greater than the time resolution T1.

The algorithm seeks SSSest values that minimizes the cost function. To do so, the algorithm evaluates C1, the covariance of the data (which contains the error on the data) to which a representativity error Cr has been added that takes into account the difference in resolution of fields R1-T1 and r1-t1 :

$$C1 = Cd + Cr(R1-T1, r1-t1)$$

An example of covariance Cr (representativity error) is given for the month of January in Figure 8. This covariance is seasonal and specific to each grid node. Cd is the covariance on the observed data, i.e. measurement errors.

C2 is the variability of field R1-T1 relative to the SSSmoy field. An example of variability applied in the monthly case is shown in Figure 5, again, this variability is seasonal and different for each grid node.

However, each sensor must be considered independently in the cost function considering the space/time resolution of each product. We therefore have, in the case where we estimate the monthly field strength and if we consider the three sensors SMOS, SMAP and Aquarius:

$$C(SSS) = \langle SSSobs\_smos - SSS \,|\, C1\_smos^{-1} \cdot (SSSobs\_smos - SSS) \rangle +$$

$$\langle SSSobs\_smap - SSS \,|\, C1\_smap^{-1} \cdot (SSSobs\_smap - SSS) \rangle +$$

$$\langle SSSobs\_aqua - SSS \,|\, C1\_aqua^{-1} \cdot (SSSobs\_aqua - SSS) \rangle +$$

$$\langle SSS - SSSmoy \,|\, C2^{-1} \cdot (SSS - SSSmoy) \rangle$$

Here, the average SSS field SSSmoy is given at a monthly resolution and a spatial resolution of 50km.

In principle, we should also include in the SMOS covariance the representativity error that corresponds to the transition from acquisition time (about one second) to monthly resolution (30 days). This same error of representativity should be applied to SMAP. However, we consider here that SMOS and SMAP measurement errors are generally dominant, which is true in 90% of cases. For Aquarius, which has much smaller errors, the algorithm takes into account representativity errors.

Finally, the cost function also contains the estimation of biases. For each type of acquisition, a different relative bias is considered. Only the bc bias of Eqn 4-1 is taken into account, the latitudinal bias being corrected beforehand.

4.3.7.2   Parameter estimation

The algorithm is therefore looking for solutions SSS(t) and bc that both minimizes the cost function. Each grid node is processed separately. All available SSS data associated with the grid node considered are used by the algorithm. The problem is linear. To minimize the cost function, a classic Raphson-Newton descent is used.

SSSobs is the observation vector that contains SMOS, SMAP and Aquarius data:

$$SSSobs = \begin{pmatrix} SSSsmos \\ SSSaqua \\ SSSsmap \end{pmatrix}$$

The parameter vector is written:

$$m = \begin{pmatrix} SSS \\ bc\_smos \\ bc\_aqua \\ bc\_smap \end{pmatrix}$$

bc_smos, bc_aqua,bc_smap are vectors that contain the biases for each type of acquisition (ascending/descending,dwell lines, fore-afte ...etc) that can be grouped into a vector bc. We will take as a priori bc=0 for all sensors and acquisition types.

The vector parameter a priori is written:

$$m\_prior = \begin{pmatrix} SSSmoy \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

SSSmoy is the starting SSS used as a priority (this value is constant over time). It is taken equal to the SSS of the central dwell SMOS, ascending orbits when possible. Otherwise, the median of the observed SSS is used.

If we call H, the matrix of partial derivatives:

$$H = \begin{bmatrix} \dfrac{\partial SSSsmos}{\partial SSS} & \dfrac{\partial SSSsmos}{\partial bc\_smos} & \dfrac{\partial SSSsmos}{\partial bc\_aqua} & \dfrac{\partial SSSsmos}{\partial bc\_smap} \\ \dfrac{\partial SSSaqua}{\partial SSS} & \dfrac{\partial SSSaqua}{\partial bc\_smos} & \dfrac{\partial SSSaqua}{\partial bc\_aqua} & \dfrac{\partial SSSaqua}{\partial bc\_smap} \\ \dfrac{\partial SSSsmap}{\partial SSS} & \dfrac{\partial SSSsmap}{\partial bc\_smos} & \dfrac{\partial SSSsmap}{\partial bc\_aqua} & \dfrac{\partial SSSsmap}{\partial bc\_smap} \end{bmatrix}$$

where   :

$$SSSsensor = F(m) = SSS - bc\_sensor$$

with "sensor" = smos, aqua or smap.

This matrix is calculated on the observation points.

The covariance matrices used are as follows:

- Cd the error matrix,

- Cm the matrix of SSS variability and a priori error on bc,

- Cr the matrix of representativity errors.

$$Cd = \begin{bmatrix} Cd\_smos & 0 & 0 \\ 0 & Cd\_aqua & 0 \\ 0 & 0 & Cd\_smap \end{bmatrix}$$

$$Cm= \begin{bmatrix} CSSS & 0 \\ 0 & Cbc \end{bmatrix}$$

CSSS is a time smoothing operator that contains the expected variability that is provided as auxiliary data. Thus, the covariance of the SSS that links two times t1 and t2 is written:

$$CSSS(t1,t2)=sigSSS(t1)sigSSS(t2)\exp\left(-\frac{(t1\text{-}t2)^2}{\xi^2}\right)$$

with ξ=25 days and 6 days for monthly and weekly products respectively.

"sigSSS" is interpolated temporally to the acquisition times.

"Cbc" is a diagonal matrix that contains the a priori standard deviation of biases. This standard deviation is set at 4pss.

The Cr matrix corresponds to representativity errors:

$$Cr= \begin{bmatrix} Cr\_smos & 0 & 0 \\ 0 & Cr\_aqua & 0 \\ 0 & 0 & Cr\_smap \end{bmatrix}$$

With, in year 1, "Cr_smos" and "Cr_smap" set to 0.

In addition to measurement errors, representativity errors are added:

$$Ct=Cd+Cr$$

Representativity errors are reported monthly. They are interpolated temporally to the acquisition times.

In this formalism, the cost function is written, for each grid node:

$$C(SSS,bc)=<SSS_{obs}\text{-}F(m)|Ct^{-1}\cdot(SSS_{obs}\text{-}F(m))>+<m\text{-}m\_prior|Cm^{-1}\cdot(m\text{-}m\_prior)>$$

with:

$$F(m)=SSS\text{-}bc$$

We look for SSS_est and bc_est that minimizes C (SSS,bc). The solution of minimization is written:

$$m\_est=m\_prior+Cm\cdot H^{T}\cdot(H\cdot Cm\cdot H^{T}+Ct)^{-1}\cdot(SSS_{obs}\text{-}F(m\_prior))$$

where "T" indicates the transpose operator.

4.3.7.3   Estimation of monthly SSS

In order to estimate the monthly SSS, we proceed in 3 steps:

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

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

 The aim here is to identify any outliers against the returned SSS field. Outliers can be linked to intermittent RFIs. It is considered here that stable RFI contamination can be corrected.

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

The relative biases used to derive monthly SSS are estimated taking the averaged SSS from the SMOS central across swath location as a priori.

4.3.7.4   Estimation of weekly SSS

To estimate the weekly SSS, the biases calculated at the monthly SSS are frozen (it is assumed that the biases will not be better estimated from a weekly smoothing). We start from the monthly SSS as a priori. We try to estimate the weekly fluctuations around this a priori. First, a 3sigma filter is used. Here, sigma = sqrt(error_L2OS ^{2} + variability^2). The variability is given by Mercator model. This eliminates outliers that deviate too far from what is expected.

That being said, the SSS field estimate is done in a single step.

4.3.8   Absolute correction

At the end of the inter-sensor bias correction step, the salinities obtained are set on average on those of the SMOS ascending central dwell (or on the average of the SSS of all sensors if the central dwell is not observed). The central dwell can itself be affected by a bias that must be corrected by using the global average of the ISAS data over the period considered (thus, the dynamics of the SSS SMOS are not affected: only one constant value, grid node per grid node, is added for the entire period covering 2010-2018). This step uses the median or the quantile of the SMOS and ISAS time series. The correction allows matching the SMOS quantile and the ISAS quantile on the 30-day average data. Moreover, it appears that the dynamics of the SMOS SSS is much stronger than that of ISAS. And, it turns out that this dynamic is not symmetrical and that intermittent freshening is much more frequent than intermittent over-salting. In fact, ISAS data contain much less information than SMOS in terms of freshening. This is why, in case of high weekly variability, we propose to perform the calibration of SMOS on ISAS, not by using the median but a high quantile (80%), in order to promote the calibration on the strong values of the SSS.

We therefore propose the following empirical approach:

- if the variability is less than 0.6 pss, the quantile is taken at 50% (the median)

- if the variability is greater than 0.8 pss, the quantile is 80%.

- between 0.6 and 0.8 pss, we take a quantile that varies linearly: quantile(%)=(1.5 x variability - 0.4) x 100.

The map of quantiles used for the absolute calibration of the SSS is given in Figure 9.

In fact, if a high quantile is used for calibration, bias can be generated if the SSS error is greater than the variability. In the latter case, it is better to use the median.
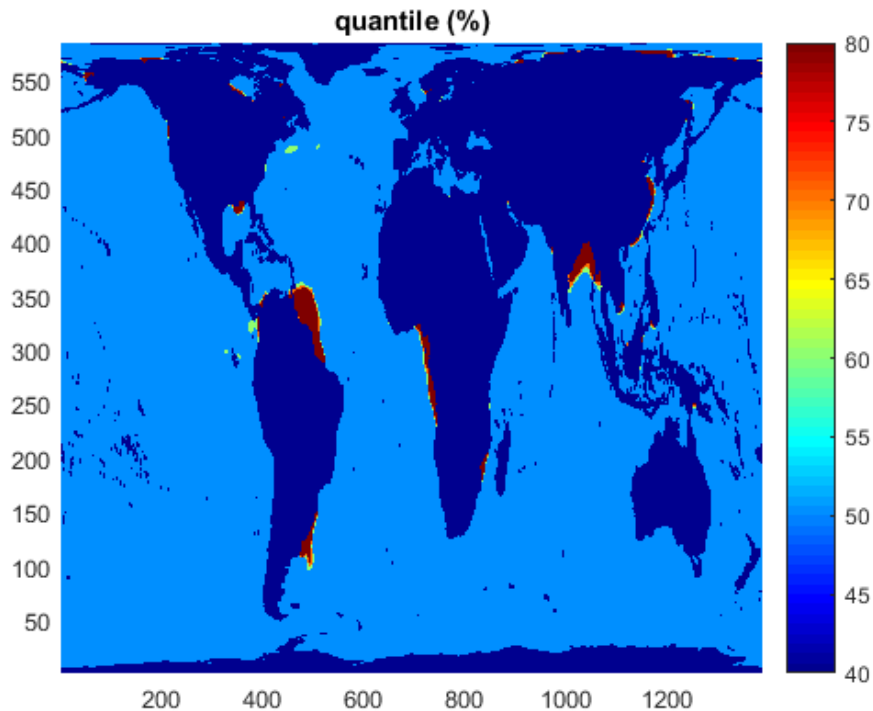
Figure 9: quantile map used for the SSS absolute calibration.x and y axis units in pixel number for longitude and latitude respectively.

4.3.9   Error budget

The computation of theoretical errors is obtained directly from the pseudo hessian matrix.

$$Cpost=Cm-Cm \cdot H^T \cdot (H \cdot Cm \cdot H^T+Ct)^{-1} \cdot H \cdot Cm$$

The problem turns into inverting the "H·Cm·H^T+Ct" matrix over the entire period, which is rather heavy. We therefore prefer to make a sliding window over a large time interval and invert the

Climate Change Initiative+ (CCI+)
Phase 1
Algorithm Theoretical
Development Basis Document

Ref.: ESA-CCI-PRGM-EOPS-SW-17-0032
Date: 19/12/2019
Version : v1.3
Page: 36 of 41

matrix on this time domain (the computation being similar to the one we could perform over the entire period).

Note that the error *a posteriori* is necessarily lower than the variability introduced in the operator Cm. In the monthly case, this variability corresponds to the expected monthly fluctuations shown in Figure 5. In the weekly case, the variability is calculated relative to the monthly field. The latter is generally lower than the monthly variability (Figure 7). The a posteriori error obtained on the weekly fields should therefore be lower than that obtained on the monthly fields. However, this is only true if, to obtain the weekly fields, we started from noise-corrected monthly fields, which is not the case. The propagation of errors on the weekly fields must therefore take into account errors on the monthly field. Thus, for the monthly fields, we have:

$$\text{Cpost\_month} = \text{Cm\_month} - \text{Cm\_month} \cdot \text{H}^{\text{T}} \cdot (\text{H} \cdot \text{Cm\_month} \cdot \text{H}^{\text{T}} + \text{Ct})^{-1} \cdot \text{H} \cdot \text{Cm\_month}$$

and for the weekly fields:

$$\text{Cpost\_week} = \text{Cpost\_month} + \text{Cm\_week} - \text{Cm\_week} \cdot \text{H}^{\text{T}} \cdot (\text{H} \cdot \text{Cm\_week} \cdot \text{H}^{\text{T}} + \text{Ct})^{-1} \cdot \text{H} \cdot \text{Cm\_week}$$

with "Cm_month", the monthly variability and "Cm_week" the weekly variability relative to the monthly variability.

The a posterior errors on the monthly and weekly fields are therefore obtained as follows:

$$\sigma\text{SSS}_{\text{month}} = \sqrt{\text{diag}(\text{Cpost\_month})}$$

$$\sigma\text{SSS}_{\text{week}} = \sqrt{\text{diag}(\text{Cpost\_week})}$$

The number of outliers is also calculated on this same basis as well as the number of data available. The window sizes used are respectively +/- 30 days and +/- 10 days for monthly and weekly products respectively.

4.3.10 Level 4 products

The aim of this part of the algorithm is to estimate the monthly and weekly SSS.

User requests converge on a product that contains the following fields (AD07):

-monthly and weekly SSS fields : obtained from OI algo (section 4.3.7).

-SSS error : obtained from OI algo (section 4.3.9).

-SSS mean bias : mean of the biases applied over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).

-SSS std bias : mean of the biases applied over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).

-number of outliers over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).

-number of data over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).

-quality flag =0 if no data present over the considered time interval (+/-30 days for monthly data and +/-10 days for weekly data).

-pct_var : $100(SSS\ error)^2/variability$ (%).

These values are given for each grid node and sampled every two weeks for monthly products and every day for weekly products.

5   Conclusions and way forward

In year 1, the input data to the CCI processing chains are level 2 for SMOS and SMAP, Level 3 for Aquarius, and they are generated outside of the CCI processor. These data are all projected on the same EASE grid at a spatial sampling of 25 km. The L2 products used for each sensors come directly from official space agency dedicated centres (e.g. CATDS, RSS). They are therefore not generated by the CCI processing chains. The next steps will therefore consider the generation of the CCI level 2 SMOS, SMAP and Aquarius products which shall be debiased to be for example, potentially best assimilated in ocean circulation models. Plans forward will also include refined Land Contamination and RFI filtering, and applying different weights for fore/aft SMAP views and for ascending/descending passes for all sensors during the optimal interpolation processing.

First validation results using co-localized CCI and ARGO upper SSS over 2010-2018 are available online from the Pi-MEP platform, and can be found here for the monthly and weekly CCI products. PVASR (AD08) presents some further investigations.
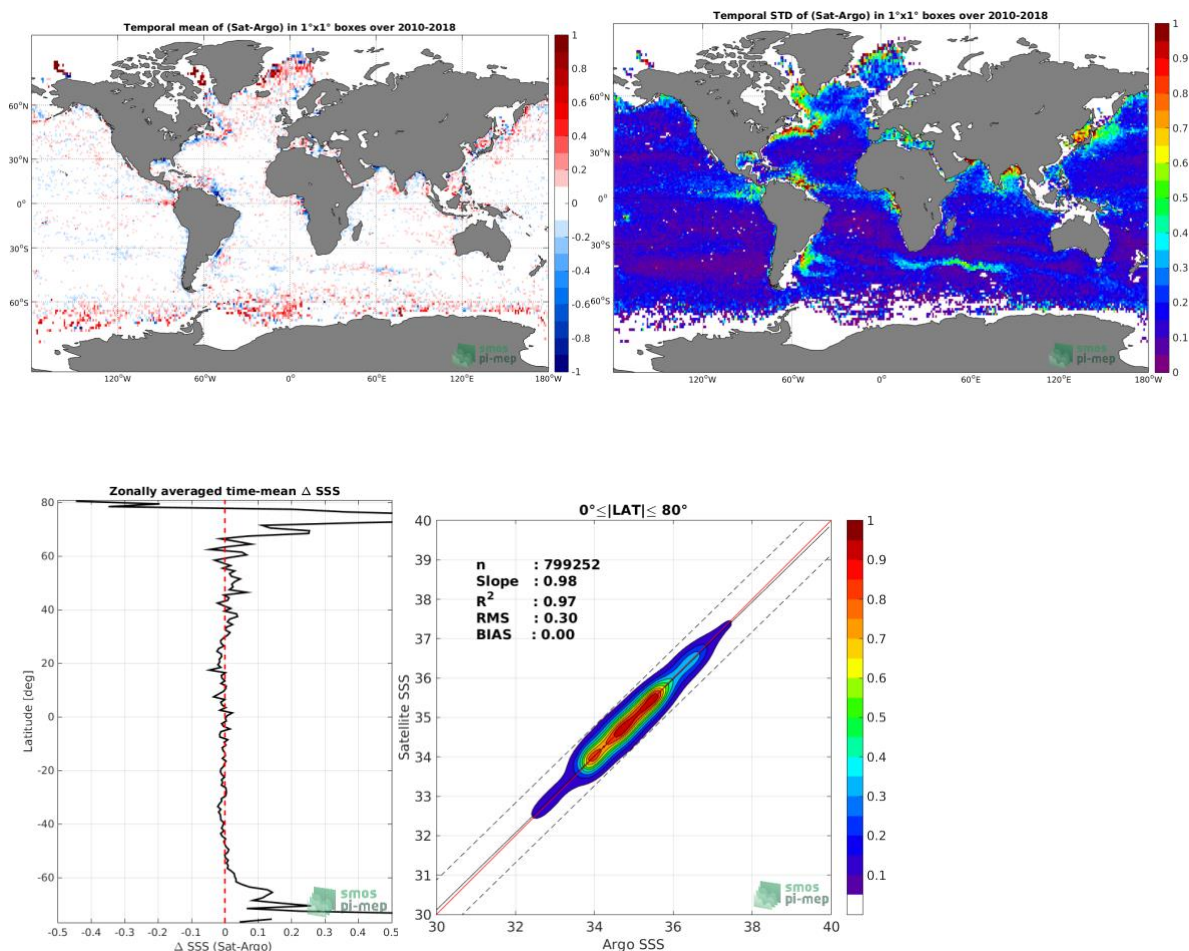

Figure 10 Temporal Mean bias (Top Left) and standard deviation (Top right) maps of the differences between CCI and ARGO SSS values. These maps are obtained by averaging the differences between co-localized CCI and ARGO SSS pairs over 1°x1° boxes and for the period 2010-2018. Bottom left: zonally averaged time-mean ΔSSS (CCI - Argo) for all the collected Pi-MEP match-up pairs at latitudes less than 80°. Bottom right: contour maps of the concentration of CCI-L4-ESA-MERGED-OI-V1.5-MONTHLY SSS (y-axis)

versus Argo SSS (x-axis) at match-up pairs for different latitude bands. For each plot, the red line shows x=y. The black thin and dashed lines indicate a linear fit through the data cloud and the 95% confidence levels, respectively. The number of match-up pairs n, the slope and R^2 coefficient of the linear fit, the root mean square (RMS) and the mean bias between satellite and in situ data are indicated.

Table 1 statistics of the differences between CCI and ARGO delayed mode SSS values. The conditions C1 to C9c are described in the text.

Condition	#	Median	Mean	Std	RMS	IQR	r^2	Std*
all	509675	0.00	0.00	0.25	0.25	0.21	0.96	0.16
C1	186489	0.00	0.00	0.16	0.16	0.18	0.97	0.13
C2	348988	0.00	0.00	0.24	0.24	0.20	0.97	0.15
C3	5040	0.01	0.03	0.28	0.28	0.26	0.95	0.20
C4	91522	0.04	0.05	0.40	0.40	0.27	0.96	0.20
C5	283946	0.00	0.00	0.19	0.19	0.18	0.97	0.13
C6	218757	-0.01	0.00	0.32	0.32	0.26	0.96	0.20
C7a	30490	0.00	-0.02	0.56	0.56	0.36	0.98	0.27
C7b	200829	0.00	0.00	0.28	0.28	0.23	0.94	0.17
C7c	277856	0.00	0.00	0.17	0.17	0.18	0.97	0.14
C8a	34673	0.02	0.02	0.28	0.28	0.25	0.71	0.18
C8b	109674	0.01	0.01	0.24	0.24	0.20	0.97	0.15
C8c	365251	-0.01	-0.01	0.26	0.26	0.21	0.96	0.15
C9a	21585	0.06	0.08	0.47	0.48	0.30	0.98	0.22
C9b	472734	0.00	0.00	0.23	0.23	0.20	0.92	0.15
C9c	15356	-0.04	-0.05	0.37	0.38	0.24	0.79	0.18

We reproduce in Figure 10 and Table 1 some summary statistics for the monthly CCI SSS match-ups with ARGO SSS.

Table 1 shows classical statistics (Median, Mean, Std, RMS, IRQ, r^2) allowing comparisons between SSS products and reference SSS. Std* refers to robust std computation : Std*(X)=Median(|X-Median(X)|)/0.6745 , with X, a random variable. If X follows a gaussian law, Std*(X) converges to the classical standard deviation of X. In case of some large value contamination, Std*(X) also converges to the classical standard deviation.

Figure 10 show the validation results for using both near-real time and delayed mode ARGO floats. The results in Table 1 only considered the best quality 'delayed-mode' ARGO float SSS data. The conditions in table 1 are defined as follows:

- all: All the match-up pairs satellite/in situ SSS are used to derive the statistics
- C1: only pairs where Rain Rate (RR)=0 mm/h, 3< U10 <12 m/s, SST>5°C, distance to nearest coast > 800 km

- C2: only pairs where RR=0 mm/h, 3< U10 <12 m/s
- C3: only pairs where RR>1mm/h and U10 <4m/s
- C4: only pairs where Mixed Layer Depth (MLD)<20m
- C5: only pairs where WOA2013 SSS Std<0.2
- C6: only pairs where the standard deviation of the WOA2013 SSS climatology >0.2
- C7a: only pairs where distance to coast is < 150 km.
- C7b: only pairs where distance to coast is in the range [150, 800] km.
- C7c: only pairs where distance to coast is > 800 km.
- C8a: only pairs where in situ SST is < 5°C.
- C8b: only pairs where in situ SST is in the range [5, 15]°C.
- C8c: only pairs where in situ SST is > 15°C.
- C9a: only pairs where in situ SSS is < 33.
- C9b: only pairs where in situ SSS is in the range [33, 37].
- C9c: only pairs where in situ SSS is > 37.

As found (Figure 10), the global RMSD of the monthly SSS differences with all validated ARGO available at global scale is 0.3. When considering only the best validated ARGO data (delayed mode data) the global RMSD drops to 0.25. In optimal conditions (conditions C1), the RMSD is minimal reaching 0.16. Degraded CCI product quality is nevertheless observed at latitudes higher than 60° (Figure 10, bottom left plot), in highly variable SSS zones (Figure 10, top right panel and condition C6), at low SST (conditions C8a), in the 150 km band around the coasts (condition C7a) and in the freshest oceanic waters (condition C9a).

In the next steps, efforts will be conducted to improve the CCI products at high latitudes and low SST. The large difference between CCI and Argo in low SSS value areas (mostly associated with large tropical river plume and rainy ITCZ areas) is probably dominated by representativity errors as monthly average large-scale satellite data can not represent the local measurements performed in zone highly dynamical SSS zones. The addition of SSS data in year 2 from C-band sensor in these region will expectedly increase the robustness of the CCI product in these freshwater surface area.

End of Document