Journal of Geophysical Research: Oceans

Supporting Information for

Satellite-based ~~Time Series of~~ Sea Surface Salinity designed for Ocean and Climate Studies

J. Boutin[1], N. Reul[2], J. Koehler[3], A. Martin[4], R. Catany[5], S. Guimbard[6], F. Rouffi[7], J.L. Vergely[7], M. Arias[5], M. Chakroun[7], G. Corato[8], V. Estella-Perez[1,*], A. Hasson[1,†], S. Josey[4], D. Khvorostyanov[1], N. Kolodziejczyk[2], J. Mignot[1], L. Olivier[1], G. Reverdin[1], D. Stammer[3], A. Supply[1,2], C. Thouvenin-Masson[1], A. Turiel[9], J. Vialard[1], P. Cipollini[10,‡], C. Donlon[10,‡], R. Sabia[11], S. Mecklenburg[10]

[1]Sorbonne University, LOCEAN/IPSL Laboratory, CNRS–IRD–MNHN, Paris, France.

[2]University of Brest, LOPS Laboratory, IUEM, UBO–CNRS–IRD–Ifremer, Plouzané, France.

[3]Institut für Meereskunde, Centrum für Erdsystemwissenschaften und Nachhaltigkeit, Universität Hamburg, Germany.

[4]National Oceanography Centre, Southampton, UK.

[5]ARGANS Ltd, UK.

[6]Ocean Scope, France.

[7]ACRI-st, France.

[8]Adwaiseo, Luxemburg.

[9]Barcelona Expert Center (BEC) and Institute of Marine Sciences (ICM), CSIC, Spain.

[10]European Space Agency, ECSAT, Harwell, United Kingdom.

[11]Telespazio-UK for ESA, ESRIN, Frascati, Italy.

Corresponding author: Jacqueline Boutin (jb@locean.ipsl.fr)

*Now at UL Services Spain SL

†Now at Mercator Ocean International, France

‡Now at European Space Agency, ESTEC, Noordwijk, the Netherlands.

Contents of this file

Text ~~S4~~S1, S2, S3 and S6

Introduction

This supporting information ~~gives details about~~provides information on the user survey, ~~about the differences between SSS CCI version 1 and version 2 (version 2 is described in the paper), about~~on SMOS data preprocessing, on the methodology used to build L4 SSS fields, to estimate SSS variability and level 2 SSS uncertainties used for the L4 SSS generation~~, PIMEP validation statistics obtained for weekly SSS CCI fields and SSS anomalies computed with ISAS SSS fields~~.

All the ~~information~~material presented here ~~are~~is not essential to the comprehension of the article but ~~bring~~provides more detailed information ~~and details~~ to the reader.

S1. SSS data requirements for ocean and climate studies

To create an SSS data set that satisfies the needs of climate users, both modeling and Earth-observing scientists groups, users of satellite SSS data were consulted through various approaches: personally, via e-mail, mailing lists, or at meetings. They were invited to participate in web surveys and to specify their requirements for satellite SSS data. In our survey, we asked specific questions to find out the user's priorities (typically higher resolution or improved uncertainty estimates).

The survey (available on https://forms.gle/BVDroYrNpVvpxFJu9) gathered 54 answers from various countries of origin/fields (Table S1). Most responses were from the USA (28%), followed by Germany (19%) and the UK (17%).

Table S1: Percentage of respondents to the online survey.

Country	USA	Germany	UK	South Korea	Italy	Spain	Argentina	Australia	Brazil	Norway	Japan	France

%	28	19	17	2	4	7	2	2	2	4	2	13

The questionnaire had four different parts, which were 1) User profile information, 2) User dataset requirements, 3) User dataset quality information, 4) Other user requirements or suggestions.

The majority of users requires global spatial coverage and temporal coverage from at least 9 years. The resolution requirements vary according to the studied phenomena. About 33% of respondents require data with a temporal resolution of 1-3 days, while, for 35% (28%) of respondents, weekly (monthly) averaged data are sufficient (Figure S1a). In terms of spatial resolution, 39% of respondents require data on a 0.25° spatial grid, while 28% of respondents require data on 1° spatial grid (Figure S1b). The majority of respondents would prefer a data product with high spatial and temporal resolution (weekly, 0.25°) on a regular latitude-longitude grid. Interestingly, a majority of users would prefer a product with high temporal and spatial resolution and a lower accuracy rather than working with a product with high accuracy but a lower resolution (Figure S1c). It was also found that the participants are aware of the data set limitations and have realistic expectations.

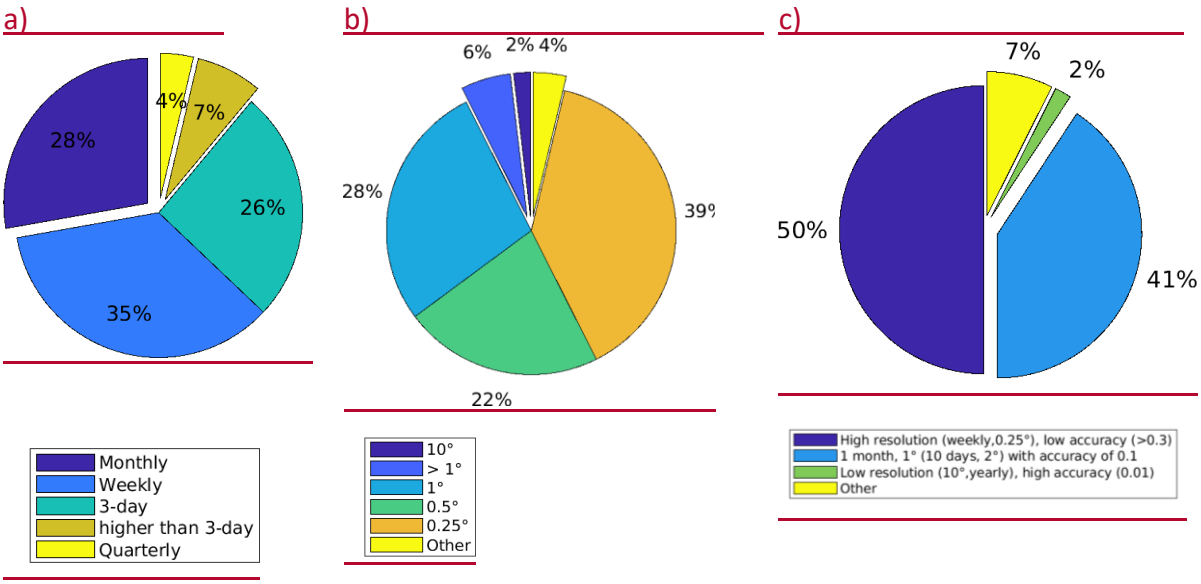

Figure S1: Percentage of required (a) temporal and (b) spatial resolution. c) Preferred SSS product based on the user's spatial and temporal resolution needs.

According to the survey, data should be combined to overcome the weaknesses of individual datasets. 50% prefer a combination of satellite and in-situ measurements, whereas 39%

require the combination of data from different satellite sensors. However, in the CCI L4 SSS products described here, information from in-situ measurements was restricted to a minimum in order to work with measurements having homogeneous spatio-temporal resolution and sampling. By making available the multiple-sensor datasets on different spatial-temporal grids, the needs of different users can be met. The most common requirement is for L4 data (43%), directly followed by requirements for L3 (37%). Some potential users, mainly modelers or scientists investigating rapid SSS changes, require L2 (20%). L3 and L2 data are already available from the original data centers. L2 and L3 datasets including the CCI+SSS systematic corrections are kept as an internal CCI+SSS product.

Uncertainty information for each SSS grid point has to be fully characterized, including random noise and systematic uncertainties of the applied adjustments. Information about bias (systematic uncertainty) correction is most commonly required by respondents. 46% of the respondents would prefer a quality information easy to use, such as a good/bad flag or the probability that a value is good/bad.

User Requirement Survey results show the importance of contacting users and promoting communication between the users and potential users of CCI L4 SSS fields. Users will be regularly contacted to refine requirements, as well as to check their satisfaction with the CCI L4 SSS product. The recommendations regarding resolution, format, quality, and additional information derived from the user consultation are summarized in the User Requirement Document (URD available on https://climate.esa.int/en/projects/sea-surface-salinity/key-documents/).

S2. SMOS SSS data pre-processing

We only consider SMOS SSS satisfying the following criteria (same notations as in *(Boutin et al., 2018)*): normalized χ of the retrieval, $\chi_N < 3$, SSS random uncertainty, $E_{SSS\_L2} < 3$, pixel within +/-400 km from the center of the swath, with small number of Tb outliers (level 2 fg_outlier flag), uncontaminated by ice (level 2 Dg_suspect_ice=0; this flag removes pixels in cold waters (SST<2°C) in which at least one Tb differs by more than 20K from modeled Tb. This is a very stringent filtering that is likely to be removed in future versions), with moderate to strong RFI and ice contamination as detected using SMOS retrieved pseudo-dielectric constant, Acard (|Acard smos – Acard mod|<2 and Acard>42; see more in (Supply et al., 2020b)), wind speed less than 16 m/s, SSS between 2 and 45.

Deficiencies in the dielectric constant model leads us to adjust SMOS SSS with a polynomial SST function derived from comparisons between Aquarius SSS (retrieved with similar dielectric constant model and atmospheric model as SMOS processing models), and Argo SSS (blue dotted curve in Figure 16 of (Dinnat et al., 2019)). A correction for seasonal latitudinal varying biases is also applied, similar to what is described in (Boutin et al., 2018):

$$SSS_{obs}(t, \phi, \lambda, x_{swath}, x_{orb}) = SSSsmos_{ref}(\phi, \lambda, m) - b_{lat}(\phi, x_{swath}, x_{orb}, m)$$

where SSS_{obs} is the observed SSS, t is the time of the measurement, ϕ, and λ, are respectively the latitude and the longitude of the considered pixel over the ocean, x_{swath} corresponds to the pixel location across the swath, x_{orb} indicates the satellite orbit direction (ascending or descending), b_{lat} is a latitudinal correction that varies seasonally as a function of the month, m, and $SSSsmos_{ref}$ is a reference SMOS SSS taken at a given x_{swath} and x_{orb}, chosen so that $SSSsmos_{ref}$ interannual variability for the considered month and the corresponding pixel is the closest with that of in situ interpolated (In-situ Analysis System, ISAS) SSS after 5° latitudinal smoothing. b_{lat} is estimated through a least square minimization approach, and through a series of iterations. In order to avoid land-sea contamination, b_{lat} is derived from SMOS measurements further than 1200 km from coast except at the high northern latitudes where the distance to coast is reduced to 600km in order to get enough measurements. It is computed over 2012-2018 to avoid large RFIs in the North Atlantic in 2010 and 2011. b_{lat} is then removed from all SMOS SSS whatever their distance to coast and before estimating the land-sea contamination correction.

~~S1.~~S3. Generation of level 4 SSS fields: detailed algorithm

The algorithm is ~~looking~~searching for solutions SSS(t) and bc that both minimize 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~~.~~, so that 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 the vectors that contain the biases for each type of acquisition (ascending/descending,dwell lines, etc) that can be grouped into a vector bc. We take as a priori bc=0 for all sensors and acquisition types.

The vector parameter a priori is written:

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

SSSprior is the initial SSS value, this value is constant over time. It is taken equal to the ~~SSS of the SMOS central dwell line ascending orbits when possible. Otherwise, the~~ median of the observed SSS ~~is used~~.

We call H, the matrix of partial derivatives:

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

where: SSSsensor=F(m)=SSS-bc_sensor

with "sensor" = smos (SMOS), aqua (Aquarius) or smap (SMAP).

This matrix is calculated on the observation points.

The covariance matrices used are defined as follows:

- ~~Cd~~Co the error matrix with data uncertainties derived,

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

- Cr the matrix of representativity uncertainties.

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

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

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

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 (either between two observational times or between an observational time and a sampled time of the OI) is written:

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

with ξ=25 days and 6 days for monthly and weekly products respectively. "sigSSS" is interpolated temporally to t1 and t2 from seasonal variability.

~~"sigSSS" is interpolated temporally to the acquisition times from seasonal variability.~~

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

The Cr matrix corresponds to representativity uncertainties:

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

With, in CCI v1 and v2, "Cr_smos" and "Cr_smap" set to 0.

In addition to measurement uncertainties, representativity uncertainties are added:

$$Ct=\sout{Cd}Co+Cr$$

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

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

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

$C(SSS,bc)=(SSSobs-F(m))^{T}\cdot Ct^{-1}\cdot(SSSobs-F(m))+(m-m\_prior)^{T}.Cm^{-1}\cdot(m-m\_prior)$

with:

$$F(m)=SSS-bc$$

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

$$m\_est=m\_prior+\sout{Cm\cdot HT\cdot(H\cdot Cm\cdot HT+Ct)^{-1}}Cm'\cdot H^{T}\cdot(H\cdot Cm\cdot\sout{HT+}H^{T}+Ct)^{-1}\cdot(SSS_{obs}-F(m\_prior))$$

where "T" indicates the transpose operator~~.~~, Cm is the matrix of variability operating in the observational space and Cm' the matrix of variability between observational time and regular sampled time of the OI.

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~~,~~ is performed spatial grid node by spatial grid node ~~is performed~~,

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

The aim here is to identify any outliers against the returned SSS field. Outliers can be linked to intermittent RFIs. It is ~~considered~~assumed 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.

Estimation of weekly SSS

To estimate the weekly SSS, the biases calculated at the monthly SSS generation step 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 value. We estimate the weekly fluctuations around this a priori, taking into account the acceptable SSS variability between weekly and monthly fields that was derived as a monthly climatology from the Mercator model. A 3-sigma filter is used in order to eliminate outliers that deviate too far from what is expected. Here, $\sigma = \sqrt{error\_L2^2 + variability^2}$. The weekly SSS field estimate is done in a single step.

Absolute SSS correction

At the end of the inter-sensor bias correction step, the salinities ~~obtained are set on average of the SSS of all sensors. However, the SMOS ascending central across swath location that was taken as reference can itself be affected by a bias.~~are obtained in relative values, i.e. they are known within one additive constant. This is corrected by adjusting a quantile of the CCI and ISAS SSS statistical distributions in each grid node over the period considered. The dynamics of ~~the~~ SSS

variability are not affected by this adjustment as only one constant value, grid node per grid node, is added for the entire period. In regions where SSS variability is low, we assume that high frequency variability better sampled by CCI than by ISAS does not affect significantly the median of the SSS and we therefore adjust ~~both~~the SSS median (50% quantile). In regions with larger variability, given that intermittent freshening is much more frequent than intermittent over-salting, we expect the high part of the SSS distribution to be less affected by the higher frequency sampling by satellite than by ISAS. Hence in case of high weekly variability, we perform the calibration of CCI SSS on ISAS SSS, not by using the median but a ~~high~~higher quantile, in order to promote the calibration on the high SSS values. A high quantile is not used everywhere as in case the SSS error is greater than the variability, the high quantile of the satellite SSS is expected to differ (be higher) from the one of ISAS.

If the variability is greater than 0.8, the quantile is taken as 80%. If the variability is between 0.6 and 0.8, we take a quantile intermediate between 50% and 80% that varies linearly with the SSS variability. The map of quantiles used for the absolute calibration of the SSS is given in Figure S4.

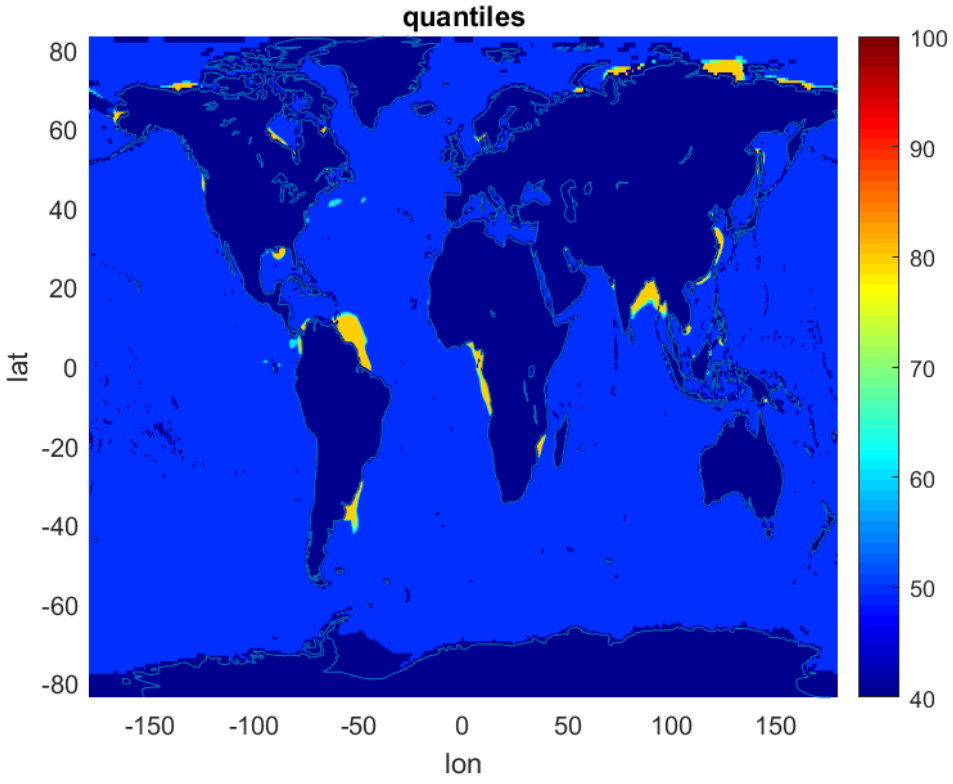

Figure ~~S4~~S3: Quantile map used for the SSS absolute calibration. x and y axis units in pixel number for longitude and latitude respectively.

The computation of theoretical uncertainties is obtained directly from the pseudo Hessian matrix.

$$Cpost=\sout{Cm-Cm\cdot HT\cdot(H\cdot Cm\cdot HT+Ct)^{-1}}Cm''-Cm'\cdot H^T\cdot\left(H\cdot Cm\cdot H^T+Ct\right)^{-1}\cdot\sout{H\cdot Cm}H\cdot Cm'^T$$

Where single apostrophe (') indicates covariance defined between observational time and regular sampled time of the OI. Double apostrophe (") indicates covariance acting over regular sampled time of the OI.

The problem ~~turns~~becomes into inverting the "$H\cdot Cm\cdot H^T+Ct$" matrix over the entire period, which is rather computationally heavy. We therefore prefer to ~~make~~take a sliding window over a large time interval and invert the matrix on this time domain (the computation being similar to the one we could perform over the entire period).

Note that the *a posteriori* uncertainty is necessarily lower than the variability introduced in the operator Cm. In the monthly case, this variability corresponds to the expected monthly fluctuations with respect to the whole time series. In the weekly case, the variability is calculated relative to the monthly field. The latter is generally lower than the monthly variability. The a posteriori uncertainty obtained on the weekly fields should therefore be lower than that obtained on the monthly fields. However, this is only true if the weekly fields are derived from noise-corrected monthly fields, which is not the case. The propagation of uncertainties on the weekly fields must therefore take into account uncertainties on the monthly field. Thus, for the monthly fields, we have:

$$\sout{Cpost\_month=Cm\_month-Cm\_month\cdot HT\cdot(H\cdot Cm\_month\cdot HT+Ct)^{-1}\cdot H\cdot Cm_{month}}$$

$$Cpost_{month}=Cm''_{month}-Cm'_{month}\cdot H^T\cdot(H\cdot Cm_{month}\cdot H^T+Ct)^{-1}\cdot H\cdot Cm'^T_{month}$$

and for the weekly fields:

$$\sout{Cpost\_week=Cpost\_month+Cm\_week-Cm\_week\cdot HT\cdot(H\cdot Cm\_week\cdot HT+Ct)^{-1}\cdot H\cdot Cm_{week}}$$

$$Cpost_{week}=Cpost_{month}+Cm''_{week}-Cm'_{week}\cdot H^T\cdot(H\cdot Cm_{week}\cdot H^T+Ct)^{-1}\cdot H\cdot Cm'^T_{week}$$

with "~~Cm_month~~Cm_{month}", the monthly variability and "~~Cm_week~~Cm_{week}" the weekly variability relative to the monthly variability.

The a posteriori uncertainties on the monthly and weekly fields are therefore obtained as follows:

$$\sout{\sigma SSS_{month}=\sqrt{diag(Cpost\_month)}}$$

$$\sout{\sigma SSS_{week}=\sqrt{diag(Cpost\_week)}}$$

$$\sigma SSS_{month}=\sqrt{diag(Cpost_{month})}$$

$$\sigma SSS_{week} = \sqrt{diag(Cpost_{week})}$$

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.

S6. SSS random uncertainties of L2 satellite SSS

The SMAP random uncertainties are derived from the std of the difference between SSS retrieved from fore and aft acquisitions (Figure S6.1, red). They are very close to a modeled error with a 0.45K radiometric noise.

The Aquarius random uncertainties are derived from comparisons of ~~SSS sampled at~~ successive (7-~~day interval~~days apart) Aquarius SSS measurements and are fitted with an SST dependency. (Figure S6.1, blue).

The SMOS random uncertainties are taken from the theoretical error multiplied by the Chi of the retrieval provided in SMOS L2 files which ~~are found to give~~provide a reasonable estimate (Figure S6.2).

In the above estimates, only pixels further than 800 km from coast have been considered in order to avoid land-sea contamination and very large representativity uncertainties.

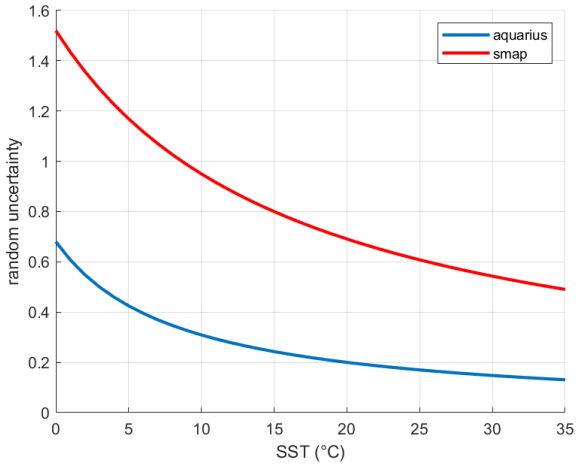

Figure S6.1. SSS uncertainties derived for SMAP (red) and Aquarius (blue) as a function of SST.

We check the reasonable behavior of estimated random errors, σ*SSS*, by ~~looking at~~considering the statistical distribution of the centered reduced SSS, SSSc:

$$SSSc = (SSSobs - SSSref)/\sigma SSS \qquad (1)$$

where SSSobs is the retrieved SSS possibly corrected from systematic uncertainties, SSSref is a reference SSS. Figure S6.2 shows an example obtained with SMOS data further than 800 km from coast compared with a 20-day SSS average. In that case SSSc is quite close to the expected Gaussian law. The slightly large value of std(*SSSc*) (1.25 instead of 1) is partly due to the presence of outliers.

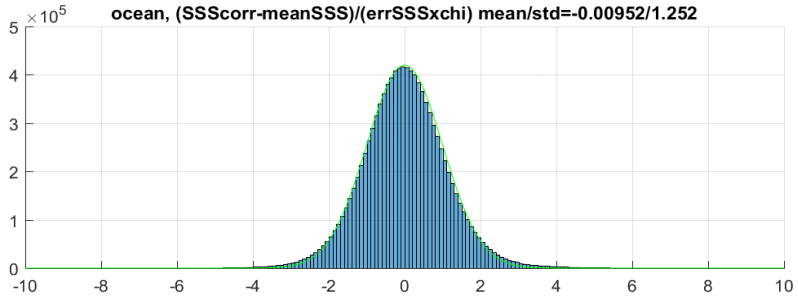

Figure S6.2. Example of the distribution of the centered reduced SSS for grid points in open ocean (further than 800 km from coast) in March 2012. SSScorr represents SMOS L2 SSS corrected from systematic uncertainties. meanSSS is an estimation of the true SSS obtained by averaging SSScorr over a 20day period.

Closer to the coast, std(*SSSc*) deviates more significantly from 1. Part of this difference can be associated with the variability of ~~the~~ salinity. In order to verify this, we sought to quantify std(*SSSc*) in regions with low variability. For grid nodes with variability lower than 0.2 on SMOS-CCI SSS rmsd, we compute a robust std, and we observe it to increase towards the coast. We apply the same process to SMAP and Aquarius data. We then derive multiplicative factors which are function to the distance to the coast, f(dcoast), that will be applied to σ*SSS* of each instrument (Figure S6.3), so that the reduced random variables normalized with the L2 random uncertainties multiplied by these factors, have a standard deviation equal to 1.

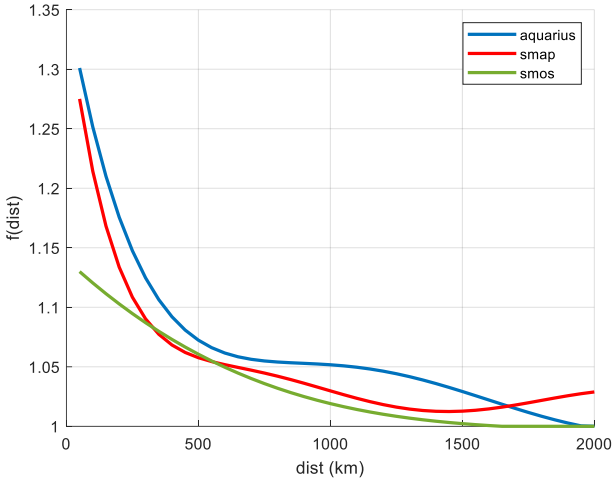

Figure S6.3: Multiplicative factor applied to the σSSS as a function to the distance to the coast.