Climate Change Initiative+ (CCI+) Phase 1

Sea Surface Salinity

Product Validation Plan (PVP)

Customer: ESA

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

Version: v1.1

Ref. internal: AO/1-9041/17/I-NB_v1r6

Revision Date: 04/12/2019

Filename: SSS_cci-D2.5-PVP-v1.1.docx

Signatures

	Name	Signature	Date
AUTHOR	Antonio Turiel BEC		04-12-2019
REVIEWED BY	Cristina González Haro BEC		
	Rafael Catany (Project Manager) ARGANS		
	Jacqueline Boutin (Science Leader) LOCEAN		
	Paolo Cipollini (Technical Officer) Telespazio		
APPROVED BY	Jacqueline Boutin (Science Leader) LOCEAN		04-12-2019
	Nicolas Reul (Science Leader) IFREMER		
	Rafael Catany (Project Manager) ARGANS		04-12-2019
ACCEPTED BY	Craig Donlon (Technical Officer) ESA	Digitally signed by Craig Donlon Date: 2020.01.13 09:49:32 +01'00'	

DIFFUSION LIST
Sea Surface Salinity Team Members
ESA (Craig Donlon, Paolo Cipollini)

Amendment Record Sheet

	Document Change Record	
Date / Issue	Description	Section / Page
15-07-2019 / v1.0	Deliver to ESA	New document
19-07-2019/v1.0	Corrected some typos	Whole document
01-12-2019/v1.1	Addressed several issues	Whole document

Table of Contents

List of figures

List of tables

1 Introduction

1.1 Purpose and scope

The Product Validation Plan (PVP) contains a list of all reference datasets to be used for validation of each type of SSS product. The PVP is requested in the Statement of Work (Task 2 SOW ref. ESA-CCI-PRGM-EOPS-SW-17-0032). The purpose of this document (PVP, document version v1.1) is to describe the protocol for validation of the Sea Surface Salinity (SSS) products obtained in the course of ESA CCI+ SSS project when compared with other sources of SSS. This document describes the guidelines to use in situ data for the validation of remote sensing products, in accordance with GEO/CEOS Quality Assurance framework for Earth Observation (QA4EO).

One of the mains objectives of this PVP is to define the appropriate Fiducial Reference Measurements (FRM), that is, a suite of independent, fully characterized and traceable ground measurements for the validation of satellite SSS. According to the definition of FRM (https://earth.esa.int/web/sppa/activities/frm), they are "the suite of independent ground measurements that provide the maximum return on investment for a satellite mission by delivering, to users, the required confidence in data products, in the form of independent validation results and satellite measurement uncertainty estimation, over the entire end-to-end duration of a satellite mission". An FRM must:

1. Document evidence of its traceability to the International System of Measurements (ISM).
2. Be independent from the satellite geophysical retrieval process.
3. Detail an uncertainty budget for the instrumentation and measurement process for the range of conditions it is used over.
4. Adhere to community agreed measurement protocols and management practises.

To the moment of writing this document, there was not an established guideline for using FRM compliant in situ measurement to validate SSS satellite retrievals. Therefore, one of the main objectives of this first PVP is to set the attempt to define SSS FRM. In this document, we will concentrate on points 2 and 3 above, point 1 given as granted by construction and point 4 is something that could be achieved by consensus starting from documents as this one (regarding the comparison with in situ salinity measurements, it has to do with the standard practises for measuring sea water conductivity, consolidated during decades of experimentation; other types of comparison, as structural and correlation measurements – to be dealt in future PVPs -, would require consensus on accepted practices and metrics).

Validation of remote sensing products with reference in situ measurements is in general a complex task, since by their characteristics there are significant differences between what can be measured by an instrument onboard an Earth orbiting satellite and any measure of any other kind. Apart from the different nature of instrumental errors, remote sensing measurements differ from other types of measurements because of different spatial resolution, different time scope and different representativeness (e.g. spatial integration). Therefore, the comparison of remote sensing products with any other kind of product is not a simple match up of data: it requires quality control and it also requires accounting for the different sources of uncertainty in ground truth measurements and, if possible, natural SSS variability sampled differently by the various instruments, in order to provide a meaningful assessment of the quality of remote sensing products.

The present version of the PVP, intended for the first year of CCI+ SSS, is focused in the comparison of remote sensing SSS products with in situ data. Therefore, this document contains guidelines to assess the uncertainties in in situ data and account them in the validation process.

Finally, this document provides guidelines for the integration of the proposed validation strategies in SMOS Pilot Mission Exploitation Platform (PI MEP).

1.2 Structure of the document

This document is composed of six sections:

Section 1 gives an introduction to the purpose and scope of this document.

Section 2 presents the overview of the theoretical framework of this PVP.

Section 3 defines a valid scheme of FRM of SSS.

Section 4 presents an overview of the reference in situ data set.

Section 5 is devoted to the integration of the previous sections in the SMOS PI MEP. Finally,

Section 6 describes the way of implementing this PVP (validation with in situ) for the full duration of the CCI+SSS project

1.3 Applicable Document

DSTD	CCI Data Standards, CCI-PRGM-EOPS-TN-13-0009	V2.1, 25/03/2019
SRD	System Requirement Document	SSS_cci-D3.1-SRD-i1r5
SSD	System Specification Document	SSS_cci-D3.2-SSD-i1r0
URD	User Requirement Document	SSS_cci-D1.1-URD-i1r0
DARD	Data Access Requirement Document	SSS_cci-D1.3-DARD-v1r3
PSD	Product Specification Document	SSS_cci-D1.2-PSD-v1r4
SoW	CCI+ Statement of Work	
ATBD	Algorithm Theoretical Baseline Document	SSS_cci-D2.3-ATBD_L3_L4-i1r0_v1.1

Table 1 – Applicable documents

1.4 Reference Document

ID	Document	Reference
RD01	R. Droghei, B. Buongiorno & R. Santoleri (2018). A New Global Sea Surface Salinity and Density Dataset From Multivariate Observations (1993–2016). *Frontiers in Marine Science* **5**, 84.	
RD02	L.R. Centurioni, V. Hormann, Y. Chao, G. Reverdin, J. Font & D.K. Lee (2015). Sea surface salinity observations with Lagrangian drifters in the tropical North Atlantic during SPURS: Circulation, fluxes, and comparisons with remotely sensed salinity from Aquarius**.** *Oceanography*, **28 (1): 96-105**	
RD03	G. Reverdin, S. Morisset, L. Marié, D. Bourras, G. Sutherland, B. Ward, J. Salvador, J. Font, Y. Cuypers, L.R. Centurioni, V. Hormann, N. Koldziejczyk, J. Boutin, F. D'Ovidio, F. Nencioli, N. Martin, D. Diverres, G. Alory & R. Lumpkin (2015). Surface salinity in the North Atlantic subtropical gyre during the STRASSE/SPURS summer 2012 cruise. *Oceanography* **28** (1): 114-123	
RD04	N. Hoareau, A. Turiel, M. Portabella, J. Ballabrera-Poy & J. Vogelzang (2018). Singularity Power Spectra: A Method to Assess Geophysical Consistency of Gridded Products - Application to Sea-Surface Salinity Remote Sensing Maps. IEEE Transactions on Geosciences and Remote Sensing 56, 5525-5536	
RD05	Boutin, J., Y. Chao, W.E. Asher, T. Delcroix, R. Drucker, K. Drushka, N. Kolodziejczyk, T. Lee, N. Reul, G. Reverdin, J. Schanze, A. Soloviev, L. Yu, J. Anderson, L. Brucker, E. Dinnat, A.S. Garcia, W.L. Jones, C. Maes, T. Meissner, W. Tang, N. Vinogradova, B. Ward (2016b), Satellite and In Situ Salinity: Understanding Near-surface Stratification and Sub-footprint Variability, Bulletin of American Meteorological Society, 97(10), doi: 10.1175/BAMS-D-15-00032.1..	
RD06	In Situ database Analyses Report. PI-MEP Consortium. March 15, 2019.	https://pimep.ifremer.fr/diffusion/analyses/insitu-database/report/pimep-insitu-report_20190315.pdf

Table 2 – Reference documents

1.5 Acronyms

AD	Applicable Document
ADP	Algorithm Development Plan
AOPC	Atmospheric Observation Panel for Climate
AR	Assessment Report (of the IPCC)
AR6	IPCC Scientific Assessment Report 6
ATBD	Algorithm Theoretical Basis Document
BEC	Barcelona Expert Center
C3S	Copernicus Climate Change Service
CAR	Climate Assessment Report
CCI	The ESA Climate Change Initiative (CCI) is formally known as the Global Monitoring for Essential Climate Variables (GMECV) element of the European Earth Watch programme
CCI+	Climate Change Initiative Extension (CCI+), is an extension of the CCI over the period 2017–2024
CDR	Climate Data Record
CEOS	Committee on Earth Observation Satellites
CFOSAT	Chinese French Oceanography Satellite
CGMS	Coordination Group for Meteorological Satellites
CliC	World Climate Research Programme - Climate and Cryosphere Project
CLIVAR	WCRP Climate Variability and Predictability project
CMEMS	Copernicus Marine Environmental Monitoring Service
CMIP	Coupled Model Intercomparison Project
CMUG	Climate Modelling User Group

COP	Conference of the Parties
COWCLIP	Coordinated Ocean Wave Climate Project (of JCOMM)
CR	Cardinal Requirement
CRDP	Climate Research Data Package
CRG	Climate Research Group
CSCDA	Copernicus Space Component Data Access System
CSWG	Climate Science Working Group
DARD	Data Access Requirements Document
DEWG	Data Engineering Working Group
DOI	Digital Object Identifier
DPM	Detailed Processing Model
DTBT3	Database for Task 3
DUE	Data User Element
E3UB	End-to-End ECV Uncertainty Budget
EC	European Commission
ECMWF	European Centre for Medium Range Weather Forecasts
ECSAT	European Centre for Space Applications and Telecommunications
ECSS	European Cooperation for Space Standardization
ECV	Essential Climate Variable
EO	Earth Observation
EOV	Essential Ocean Variable (of the OOPC)
ESGF	Earth System Grid Federation
ESM	Earth System Model
EU	European Union

FCDR	Fundamental Climate Data Record
FIDUCEO	Fidelity and uncertainty in climate data records from Earth Observations
FP7	EU Framework Programme 7
FRM	Fiducial Reference Measurements
GAIA-CLIM	Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring
GEO	Group on Earth Observations
GCOS	Global Climate Observing System
GCW	Global Cryosphere Watch
GMECV	Global Monitoring of Essential Climate Variables - element of the European Earth Watch programme.
GNSS	Global Navigation Satellite System
GOOS	Global Ocean Observing System
H2020	Horizon 2020 programme
Hs	Significant Wave Height (see also SWH)
H-SAF	EUMETSAT's Hydrology Satellite Applications Facility
HDD	Hard disk
IOC	Intergovernmental Oceanographic commission (of UNESCO)
IODD	Input Output Data Definition
IP	Implementation Plan
IPCC	Intergovernmental Panel on Climate Change
ISAS	In Situ Analysis System
ISDB	in situ database (of Fiducial Reference Measurements and satellite measurements)
ISM	International System of Measurements
JAXA	Japan Aerospace Exploration Agency

JCOMM	Joint Commission on Oceanography and Marine Meteorology
KO	Kick-off
MOOC	Massive Open Online Course
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NOP	Numerical Ocean Prediction
NWP	Numerical Weather Prediction
Obs4MIPs	Observations for Model Intercomparison Projects
OA	Objective Analysis
ODP	Open Data Portal
OI	Optimal Interpolation
OOPC	Ocean Observation Panel for Climate
PI MEP	Pilot Mission Exploitation Platform
PMP	Project Management Plan
PSD	Product Specification Document
PUG	Product User Guide
PVASR	Product Validation and Algorithm Selection Report
PVIR	Product Validation and Intercomparison Report
PVP	Product Validation Plan
QA4EO	Quality Assurance Framework for Earth Observation
QSR	Quarterly Status Report
R&D	Research and Development
RCP	Representative Concentration Pathways
RD	Reference Document

SAF	Satellite Applications Facility
SAR	Synthetic aperture Radar
SISS	Satellite and In situ [Working Group]
SLP	Sea Level Pressure
SMAP	Soil Moisture Active Passive [mission of NASA)
SMOS	Soil Moisture and Ocean Salinity [satellite of ESA]
SoW	Statement of Work
SRAL	SAR Radar Altimeter (of Sentinel-3)
SRD	System Requirements Document
SSD	System Specification Document
SSS	Sea Surface Salinity
SVR	System Verification Report
SWIM	Surface Waves Investigation and Monitoring (instrument of CFOSAT)
SWH	Significant Wave Height (see also Hs)
TOPC	Terrestrial Observation Panel for Climate
TR	Technical Requirement
UCR/CECR	Uncertainty Characterisation Report (formerly known as the Comprehensive Error Characterisation Report)
UNFCCC	United Nations Framework Convention on Climate Change
URD	User Requirements Document
USB	Universal Serial Bus
USGS	United States Geological Survey
VOS	Volunteer Observing ships
WCRP	World Climate Research Programme

WGClimate	Joint CEOS/CGMS Working Group on Climate
WMO	World Meteorological Programme
WOA	World Ocean Atlas
WWA	World Wave Atlas (of FUGRO)

2 Overview

The validation of a product comprising an oceanic variable requires the comparison between that product and an external, independent reference (condition 2 in a FRM), generally provided by different independent data sources, that could be used as ground truth (i.e. a measure to be assumed to be close to the *true value*). Ensuring the independence of the two datasets (product and ground truth) is not always simple. In some cases, in the derivation of the product some ancillary information closely related to that provided by the ground truth has been used (for instance, in the case of SSS, it is well known that there exists a strong functional relation with SST at the appropriate scales [RD01]). In other cases, the ground truth has been fed by information also used in the generation of the products (this may happen, for instance, if the ground truth is the output of a numerical model with relaxation to climatology, and the same climatology has been used to initialize and/or constrain the retrieval of the product). Thus, to make the comparison between product and ground truth statistically meaningful and unbiased, it is crucial to make sure that the data used for both types are completely independent – which implies that the process for generating these two types of data is well documented (the *traceability* condition in QA4EO) and this is not always the case.

Another difficulty arisen when comparing product and ground truth comes from the presumption that the ground truth is a perfect reference. Regardless that this assumption is rarely true the comparisons with ground truth is always subject to two kinds of uncertainties:

- Class 1 Uncertainty: Accuracy and precision errors
- Class 2 Uncertainty: Representativity errors

Regarding the first class of uncertainty (accuracy and precision errors), it concerns the quality in the acquisition or generation of the ground truth. This is usually well-documented, and it comprises the following:

- In the case of instruments, the instrument specifications provided by the manufacturer (absolute accuracy, granted lifetime, etc).
- In the case of interpolated products (for instance by means of Objective Analysis, OA, or Optimal Interpolation, OI) the error matrix constructed out of the propagation of source data errors by the interpolation scheme.
- In the case of numerical models, the model errors estimated by the model error propagation scheme.

However, those estimates of Class 1 Uncertainty are theoretical and correspond to an ideal situation. Real-world additional sources of Class 1 Uncertainty comprise:

- Fouling, drifts, poor quality control, etc in the case of instrumental in situ data.
- Sampling inhomogeneities, sampling biases, poor estimation of correlation radii, poor estimation of correlation matrices, unability to describe exceptional events, etc in the case of interpolated data.

- Numerical instabilities, spin-up effects, poor description of physical processes, etc in the case of numerical models.

Having an estimate of the actual magnitude of Class 1 Uncertainties implies performing an appropriate pre-processing of ground truth data (self-consistency, error assessment). There is not a standardized procedure for making this king of estimates. We will present some recommendations to evaluate Class 1 Uncertainties for SSS in situ measurements in Section 3.3.1.

Regarding the second class of uncertainty (representativity errors), it concerns the mismatch between the values of the products and the ground truth due to the difference in the spatial and temporal scales (sampling and integration) represented by each type of data. Remote sensing SSS products typically represent the average value of SSS on relatively wide spatial areas repetitively sampled at several days interval (except for level 2 products, which have acquisition times of few seconds). In contrast, in situ data are always referred to a very small spatial area (few centimetres in the horizontal and vertical, at most) and are essentially instantaneous measurements (the value is acquired in times of seconds or less). Due to the geophysical variability of SSS, it is expected that the range of variability of the difference between an in situ value and the corresponding remote sensing SSS value with typical spatial and temporal resolution to be of 0.2 psu or greater [RD02,RD03]. A proper characterization of Class 2 Uncertainty is thus crucial in order to decide when the observed differences between in situ and remote sensing SSS can be considered significant.

3 Definition of the validation protocol

3.1 Introduction to SSS product validation under the frame of QA4EO guidelines

According to QA4EO guidelines, SSS product validation with ground truth data should be performed by:

1) standardization of the used reference data;
2) ensured traceability of the products and validation datasets;
3) well-characterized uncertainty of reference data; and
4) meaningful quality indicators.

According to the definitions given in the guideline we have:

- *Reference (measurement) standard:* realization of the definition of a given quantity, ideally with a stated uncertainty, which can be used as a reference; it can be individual or community defined.
- *Traceability:* property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations each contributing to the measurement uncertainty.
- *Uncertainty:* non-negative parameter characterizing the dispersion of the quantity values that are being attributed to a measure and (quantity), based on the information used. Where possible this should be derived from an experimental evaluation but can also be an estimate based on other information, e.g. experience.
- *Quality Indicators:* a means of providing a user of data or derived product (which is the result of a process) with sufficient information to assess its suitability for a particular application. This information should be based on a quantitative assessment of its traceability to an agreed reference or measurement standard (ideally ISM), but it can be presented as numeric or a text descriptor, providing the quantitative linkage is defined.

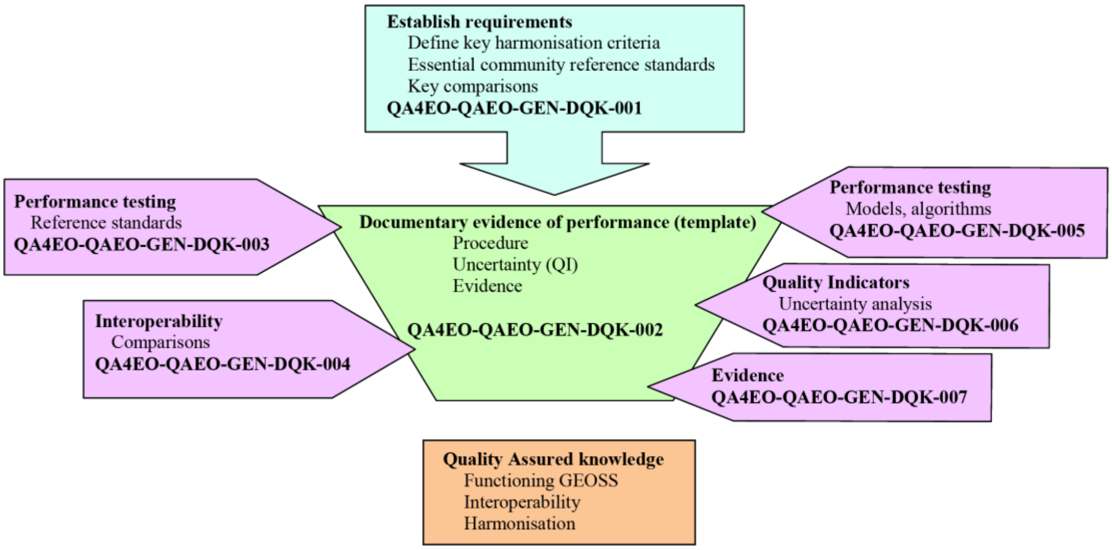

Figure 1 Summary of data quality aspects in QA4EO.

We will define our validation strategy in order to fulfil the 4 criteria mentioned above.

3.2 Validation strategy overview

- **Adhere to community standards:** Data to be used should have been endorsed by the appropriate community. Due to the task performed by PI-MEP, we recommend simply adhere to the in situ datasets recorded by PI-MEP and according to its quality control. If necessary, recommendations to include new datasets in PI-MEP should be used (see "Interaction with PI-MEP"). For other sources of data, we should discuss case by case which ones are the most appropriate; those ones should have at least recorded traceability and uncertainty estimates (see section 3.3).
- **Ensure traceability:** The datasets of in situ and other sources of data to be used for validation should have a well-documented traceability record on how the data have been generated. Examining that record, data which has been used directly or indirectly for the generation of the remote sensing SSS product should not be, to the extent of possible, included in the validation. This includes:
  - Data included in WOA SSS Climatologies used in some remote sensing products, either for filtering or for defining an appropriate reference.
  - Data that have been generated using a third source of data that was also used in the remote sensing product (e.g., SST reference field).

  If it is necessary to include data that may have contributed to the generation of the remote sensing SSS product, such a circumstance should be noted in the validation metrics.
- **Characterization of uncertainties:** Before computing the quality metrics, it is required to assess the uncertainties (both of Class 1 and Class 2), as they will provide the level of significance of the metrics.

- **Compute quality metrics:** The appropriate methods to compare ground truth data and remote sensing products will be used. In this version of PVP (PVP v1.0) we will just compute simple statistics from the differences between ground truth data and remote sensing data.

3.3 Metric for validation diagnosis

3.3.1 Assessing Class 1 Uncertainties in ground truth

Any source of ground data used for validation that is compliant with QA4EO has a traceable record of uncertainties (what comprises biases and random errors). These are the identified errors. However, other errors could be present due to instrument degradation, incomplete statistical description or model limitations, as commented above. These errors are, by definition, unidentified errors and as such we cannot know them from beforehand. The only way to quantify these errors is by intercomparison of comparable ground truth datasets, if enough of them are available.

There is not a single, correct way to intercompare different datasets representing the same variable; for the purposes of this PVP we propose to proceed to the intercomparison of ground truth datasets in the following way, depending on the type of data:

- *In situ data:* The samplings of different sets of in situ data will never be completely equivalent, so in the intercomparison there will always be a certain distance in space and in time between data. Considering Class 2 Uncertainties (see next section) this additional source of error should be added. For instance, if we are interested in computing the standard deviation of the errors, the intercomparison error of two datasets, Dataset 1 and Dataset 2, with identified standard deviations of errors σ_1 and σ_2, will be given by:

$$\varepsilon_{12}^2 = \sigma_{12}^2 - \sigma_1^2 - \sigma_2^2 - r_{12}^2$$

  where the ε_{12} denotes the intercomparison error of datasets 1 and 2, σ_{12} is the error of the difference (in this case, the standard deviation of the difference of values from 1 and from 2), and r_{12} is the standard deviation of the representativity error (as described in the next section), because all those errors are considered to be independent of each other.
- *Interpolated data:* The same considerations as stated above apply if they represent different spatial or temporal scales. In the case that the intercompared data are taken to represent the same scales the representativity error r_{12} can be equated to zero.
- *Outputs from numerical models:* Same as for interpolated data.

Notice that the intercomparison error ε_{12}^2 is the sum of the unidentified errors of Dataset 1 and Dataset 2, $\varepsilon_{12}^2 = x_1^2 + x_2^2$. It is impossible to know which is the precise contribution to the intercomparison error from each one of the Datasets, so it is proposed that this error is attributed proportionally to the identified error. For instance, in the example above we will say that the unidentified errors for each dataset are:

$$x_1^2 = \frac{\sigma_1^2 \varepsilon_{12}^2}{\sigma_1^2 + \sigma_2^2} \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad ; \quad x_2^2 = \frac{\sigma_2^2 \varepsilon_{12}^2}{\sigma_1^2 + \sigma_2^2} \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

If several unidentified errors are estimated for the same datasets, the arithmetic mean of all will be taken.

The final total error for a given dataset will be given by the sum of the identified and unidentified errors, $\varepsilon^2=\sigma^2+x^2$.

3.3.2 Assessing Class 2 Uncertainties in ground truth

The absolute amplitude of geophysical variability of in situ SSS data over the time-space scale of remote sensing products depends completely on the particular spatial resolution and time window defining the remote sensing products, but also on the region at which this variability is estimated (inter-regional variability being quite significant). However, recent analyses of the spatial and temporal power spectra of SSS provide evidence that allow relating the total variability of SSS with the variability at those scales not resolved by remote sensing products.

In [RD04] it was shown that the spatial power spectra of SSS consistently exhibit a spectral slope of -2.4 in a range going from a few kilometres to basin scale (~10.000 km), disregarding the zone of interest over monthly maps of SSS gridded products of different origin (remote sensing, interpolated in situ and numerical model outputs). Looking at the northern subtropical Atlantic Ocean, Kolodziejczyk et al. (JGR 2015) found that this slope vary seasonally but remains between -2. and -3. between 10km and 100km wavelengths. It has been verified at Barcelona Expert Center (BEC) that the same spectral slope is observed even with shorter time windows, with an estimate error of ±0.2 (private communication by Nina Hoareau). Thanks to Plancherel's equality, we can relate the integral of the power spectra density $S(k)=\beta\ k^{-2.4}$ in a given range of wavenumbers with the geophysical variability (comprised by the variance of the signal) in the corresponding range of scales. The variance contained between the spatial frequency k_L and k_l (respectively, between the scales l and L) is given by the double integral

$$\sigma^2(k_L, k_l) = \iint_{k_L<k<\boldsymbol{k_l}} d\boldsymbol{k}\ S(\boldsymbol{k}) = B \int_{k_L}^{\boldsymbol{k_l}} kdk\ k^{-2.4} = -\bar{B}\left\lfloor {k_l}^{-0.4} - {k_L}^{-0.4} \right\rfloor = A\lfloor L^{0.4} - l^{0.4} \rfloor$$

where we have assumed elliptic symmetry (common in geophysical flows, as the zonal and meridional components are dominant) and A is an appropriate positive constant. Therefore, the variance $\sigma^2(d)$ contained by all scales greater or equal to d is given by

$$\sigma^2(d) = \sigma_0^2 \left[1 - \left(\frac{d}{L}\right)^{0.4}\right]$$

where L is the size of the considered area and $\sigma_0^2 = \sigma^2(d=0)$ is the variance contributed by all scales.

Let us now assume we have three scales: let g be the scale for ground truth measurements, r the scale for the remote sensing product and L the basin scale (recall that, as shown in [RD04], the slope is the same even at basin scale). The variability described by the ground truth which is not described by the remote sensing product is thus:

$$\Delta\sigma^2(g,r) = \sigma_0^2 \left[\left(\frac{r}{L}\right)^{0.4} - \left(\frac{g}{L}\right)^{0.4}\right]$$

If we have g<<r<<L, we have $\sigma_0^2 \approx \sigma^2(r)$ and

$$\Delta\sigma^2(g,r) \approx \sigma_0^2 \left(\frac{r}{L}\right)^{0.4} \approx \sigma^2(\mathrm{r}) \left(\frac{r}{L}\right)^{0.4}.$$

That is, we can estimate the uncertainty at the scale of the ground truth from the variability of the remote sensing product at the basin scale and the ratio of the remote sensing scale to the basin scale.

For example, if we compute the variability in the North Atlantic basin (L = 5000 km) as compared to a 25 km SSS product, the variance of ground truth is expected to be a fraction which is $(1/200)^{0.4}$ = 0.12 of the variance of the remote sensing product. In terms of standard deviations, the standard deviation of the ground truth is expected to be a 34% of the standard deviation of the remote sensing product over the full basin. This estimate fits well with observed variability (for instance, the time variability observed in the North Atlantic during the SPURS campaign was found to be 0.2-0.3, [RD03]).

Assessing temporal uncertainty is something more complex, because we have not a simple expression for the time spectrum. We can however use the expression above for the spatial uncertainty to have an informed guess about how the time variability is. Let us denote u the typical (average) speed (i.e., velocity modulus) at scale r (meaning the average of speeds at the specific location during the time period of reference) and by U the typical speed at basin scale (again, the average speed for the same period but averaged over the full basin). We therefore obtain that the uncertainty associated to acquisition times t_g and t_r is:

$$\Delta\sigma^2\left(t_g, t_r\right) \approx \sigma_0^2 \left(\frac{u}{U}\right)^{0.4} \approx \sigma^2(\mathrm{r}) \left(\frac{u}{U}\right)^{0.4}.$$

The main inconvenience of the calculation of temporal uncertainties is that they require some information from an external source of data, that of sea surface speeds. For the typical temporal and spatial scales of remote sensing SSS data, this information can be well approximated by geostrophic currents obtained from altimetric currents (e.g., AVISO dataset) – notice however that very close to equator or under consistently high winds the ageostrophic terms could be important even at the temporal and spatial scales of remote sensing SSS, and under such circumstances other products should be explored.

Finally, the spatio-temporal representativity error variance can be computed as the sum of both spatial and temporal representativity error variances.

We have not discussed about vertical representativity errors, although they are sometimes important, because they are very difficult to estimate. Strong vertical stratification in the range of few centimeters (making a difference between satellite SSS and close-to-surface salinity from buoys) can happen due to persistent weak winds or the presence of freshwater lenses. However, trying to characterize this stratification will require to have very detailed information about surface wind stress and ocean currents, which remains excessively complex as far as no dedicated product exists. The problem with rain lenses can however be avoided in a relatively easy way: the largest effect is the freshening just after (within less than 1hr) rainfall. It could be avoided by discarding satellite SSS that follows at less than 0.5hr a large (>1mm/hr as detected by IMERG) rainfall event and discard as well in situ SSS if IMERG detects RR>1mm/hr [RD05].

3.3.3 Quality metrics

The proposed quality metrics are:

- Bias o mean: Estimated from the average of the difference between the value of the remote sensing product and the value of the associated ground truth dataset. This average should be computed from at least 30 independent samples in order to consider it as statistically significant (criterion according to Central Limit Theorem). The mean is considered to be informative about the systematic biases in the remote sensing product.
- Random error or standard deviation: Estimated from the standard deviation of the difference between the value of the remote sensing product and the value of the associated ground truth dataset. At least 30 independent samples must be taken into account in that mean to be considered as statistically significant (criterion according to Central Limit Theorem). The standard deviation is considered to be informative about the random errors in the remote sensing product.
- Root Mean Squared Error (RMSE) or total error: The square root of the sum of the squares of the two metrics above. The RMS is considered to be informative about the total error (systematic and random) in the remote sensing product.
- Correlation coefficient (Pearson and Spearman): The correlation coefficient, calculated as the ratio of the covariance of the remote sensing and ground truth data and the product of the standard deviations of each data type independently. If the quantities used are the values of each data type, we call the correlation coefficient Pearson coefficient; if we use for the quantities the rank of the values for each data type, we call it Spearman coefficient. Both coefficients are informative about the degree of linearity in the relation between the two variables; the best possible correlation coefficient in this case is 1. Pearson coefficient is the most standard metric, but can be affected by the presence of clusters of points in one or the other side of the distribution that would yield a false impression of good predictability. Spearman coefficient is more robust but its value is usually significantly lower than that of Pearson, and gives no information about the actual size of the error. Therefore we recommend to use both correlation coefficients and to consider significant linearity when Pearson coefficient is above 0.8 (error variance below 36%) and Spearman is above 0.5 (error rank variance below 75%).
- Linear regression of the error vs signal: The error can be, on itself, correlated to the value of the remote sensing product, something which is in fact quite common (and leads to the use of percentage errors in some cases). In order to assess this dependence, we propose to estimate the correlation coefficient of the difference between the remote sensing product and the ground truth data versus the value of the remote sensing product. While this regression would not provide a direct information on the product quality, knowing the slope, intercept and Pearson correlation coefficient would be quite informative about the connection with both. Having a regression slope of reasonable value with correlation coefficient above 0.8 would imply that the error is well characterized as a linear function of the signal.

All quality metrics must be quantized according to the error associated to the ground truth. If the error as reflected by a quality metric is below to the error estimated for the ground truth, the difference between remote sensing product and ground truth must be considered statistically non-significant, which means that from the statistical point of view they are equivalent.

3.3.4 Choice of ground truth

The choice of ground truth to be used for the validation of remote sensing SSS products depends on the kind of product we want to validate, to which purpose and in what region. When using in situ datasets, errors due to spatial and time mismatch and due to different time and spatial integration/undersampling in in situ and in satellite products must be accounted for. The use of sources of interpolated ground truth is only recommended if all their errors are well characterized. Operational numerical models with data assimilation may be considered for L4 validation but only if the L4 product is not generated using the same numerical model and their absolute errors being well characterized.

Depending the specific application (for instance, characterizing seasonality or assessing the presence of anomalies), one could prefer a different matching up of ground truth. Notice however that each application genuinely calls for the use of an appropriate type of remote sensing (for instance, L3 products of about 1-month time resolution for assessing seasonality, L2 products for assessing anomalies); using a not so-well suited remote sensing product for a given application would therefore imply some post-processing (e.g., temporal low-pass filtering or detailed representativity error accounting). It could however happen that for the given application the appropriate brand of remote sensing product is not available, and therefore the closest one should be used, trying to account for the expected deviations.

4 Reference data sets

The definition of the reference data sets to be used in the validation is a key component within the validation protocol of new dataset. The reference datasets need to be quality controlled and should verify the QA4EO guidelines in order to be acceptable as FRMs. As we are still defining the conditions for SSS FRMs, we recommend adhering to an existing quality-control facility as PI-MEP for in situ data. For interpolated maps and outputs of numerical models our recommendation is to use those already contained in Copernicus Marine Core Services.

4.1 In situ measurements including SSS FRMs

The main in situ datasets can be classified as follows:

- Close-to-surface Argo
- Thermosalinographers
- Surface drifters
- Sea mammals with mounted temperature and conductivity sensors
- Moorings

To ensure the quality of the products, we recommend to download the data from PI-MEP (https://www.smos-pimep.org/). An extensive report on their quality and limitations can be found at [RD06].

4.2 Interpolated data sets

The main interpolated sets of in situ SSS that we suggest to be used for validation are the ISAS SSS derived at 5m depth. Different ISAS products are available, we suggest by order of preference:

-the delayed mode ISAS products created by LOPS laboratory, which includes thorough quality controlled Argo profiles and other in situ measurements (ships of opportunity, research ships, sailing ships, surface drifters, marine mammals). The version up to 2015 is available on https://www.seanoe.org/data/00412/52367/; an upgraded version up to 2017 will be made available in Fall 2019.

-For periods when the LOPS product is not available, the delayed mode ISAS product available on Copernicus Marine Environment Service, http://marine.copernicus.eu/. We recommend using CORA OA SSS at surface level, 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, 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 Outputs from numerical models

As in the case of interpolated data sets, we recommend to access Copernicus Marine Environment Service, although other products could be used depending on the region and the application. Although there are several products with different resolutions, we consider that the most adequate reference is 0.25˚ daily GLORYS reanalysis, http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=GLOBAL_REANALYSIS_PHY_001_025, as this one has been shown to provide accurate structural and spectral representation of SSS [RD04].

5 Integration in the Pilot Mission Exploitation Platform (PI-MEP)

5.1 Testing, adequacy and fitness of SSS validation activities under the PI-MEP

The metrics proposed in this PVP are very similar to the ones already being used by PI-MEP, the main difference being the proper accounting for representativity and unidentified errors. Therefore, once those errors are accounted (which is important to assign significance levels to the statistical tests), standard PI-MEP validation can be used.

5.2 Standardization of SSS validation protocols to cope with PI-MEP quality control

The main difficulty with the standardization is to carry out the appropriate tests to verify the adequacy of the protocols been used for estimating the errors in the ground truth. According to PI-MEP quality control, those errors could be used either as a threshold on the significance level or to determine the confidence interval of the correspondence between ground truth and remote sensing product.

5.3 Integration of standardized quality-controlled SSS validation procedures into PI-MEP validation system

Integration is straightforward, once the procedures for estimating the errors in ground data are validated by PI-MEP system.

5.4 End-user assessment

It is necessary to count on the assessment by users who are expert in oceanography to contrast if the provided error and quality metrics are in agreement with their expectations. If significant deviations are observed, it would be convenient to run a survey in which the situations of potential conflict/deviations have been observed. The goal of the survey is to quantify those problems in the products that are deemed significant, in order to better focus the search for the origin of the observed problems. Questions to that survey should include estimates about expected absolute values, expected gradients, expected time and space position of frontal zones, observed biases, etc.

6 PVP implementation

6.1 Temporal planning

Implementation of quality metrics: two weeks to one month, once the in situ dataset is identified and compiled.

Testing of metrics and fine tuning: Two weeks.

Producing validation reports for the full period: Two weeks

6.2 Resources

There is no need for additional resources with respect to the original validation plan in this ITT.

6.3 Contingency Plan

No contingency is foreseen on data access, as it has been granted for years now. The only identified difficulty comes from obtaining negative estimates for the standard deviation of unidentified errors; in such a case, they must be taken as zero (negligible or impossible to estimate; notice that by construction, the standard deviation of the difference of two independent measurements should equal the sum of their standard deviations, being increased if other non-accounted sources of independent error – representativity, intercomparison – are present).

Regarding the PVP itself, there will be a need to review it if the criteria for defining SSS FRM are finally given (for instance, through a dedicated white paper).