Copernicus Cal/Val Solution

D3.6 Copernicus CalVal Solution

Ref: CCVS.ACR.D3.6

Version: 1.0

Date: 07/09/2022

GA number: 101004242

Document Reference

	Company
Work Package	3
WP leader	S. Clerc
Deliverable Title	Reference Scenario for Implementation
Delivery Number	3.6
Delivery Type	Report
Dissemination Level	Confidential
Version	1.0
Date of issue	07/09/2022
Lead beneficiary	ACRI-ST
Lead authors	S. Clerc (ACRI-ST)
Contributors	M. De Mazière, S. Compernolle, J.-C. Lambert, M.K. Sha, T. Verhoelst, F. Tack (BIRA-IASB), S. Labroue, G. Hadjuch, P.Vincent (CLS), M. Raynal, C. Tison, A. Meygret (CNES), B. Pflug, S. Holzwarth (DLR), L. Bourg (ACRI-ST), B. Mota (NPL), N. Gobron (EC/JRC), M. Ligi (U. Tartu)
Reviewed by	F. Jacq (Copernicus), H. S. Andersen (EEA)

Changes Log

Version	Date	Changes
1.0	30/08/2022	Version for internal review
1.0	07/09/2022	Submission

This project has received funding from the European Union's Horizon 2020 programme under grant agreement No 101004242.

Table of Content

List of Figures

Aucune entrée de table d'illustration n'a été trouvée.

List of Tables

Glossary

Acronym	Meaning	Proper noun
ACIX	Atmospheric Correction Intercomparison eXercise	yes
ADCP	Acoustic Doppler Current Profiler	no
AOD	Aerosol Optical Depth	no
ARD	Analysis Ready Data	no
ARM	Atmospheric Radiation Measurement	yes
ATBD	Algorithm Theoretical Baseline Document	no
BRDF	Bi-directional Reflectance Distribution Function	no
C3S	Copernicus Climate Change Service	yes
CAMS	Copernicus Atmosphere Monitoring Service	yes
CCI	Climate Change Initiative	yes
CEOS	Committee on Earth Observation Satellites	Yes

CEOS LPV	CEOS Land Product Validation subgroup	yes
CEOS WGCV	CEOS Working Group on Calibration and Validation	yes
CGMS	Coordination Group for Meteorological Satellites	yes
CMIX	Cloud Mask Intercomparison eXercise	yes
CNES	Centre National d'Etude Spatiales	yes
CNRS	Centre National pour la Recherche Scientifique	yes
CR	Corner Reflector	no
DEM	Digital Elevation Model	no
DLR	Deutsches zentrum fuer Luft und Raumfahrt	yes
DOAS	Differential Optical Absorption Spectroscopy	no
DHP	Digital Hemispherical Photography	no
EC	European Commission	yes
ECMWF	European Center for Medium-range Weather Forecast	yes
EDAP	Earthnet Data Assessment Pilot	yes
EEA	European Environment Agency	yes
EO	Earth Observation	no
ERI	European Research Infrastructure	yes
ESA	European Space Agency	yes
EUFAR	European Facility for Airborne Research	yes
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites	yes
EVDC	ESA atmospheric Validation Data Centre	yes
(FA)PAR	(Fraction of Absorbed) Photosynthetically Active Radiation	no
FRM	Fiducial Reference Measurement	no
FTIR	Fourier-Transform InfraRed spectroscopy	no
GBOV	Ground Based Observation for Validation	yes
GCP	Ground Control Point	no
GHG	Green House Gases	no
GNSS	Global Navigation Satellite System	no
GSICS	Global Space-based Inter-Calibration System	yes
GUM	Guide to the expression of Uncertainty in Measurement	Yes
HFR	High Frequency Radar	no
JPL	Joint Propulsion Laboratory	yes
LAI	Leaf Area Index	no
MPC	Mission Performance Cluster	yes
MSL	Mean Sea Level	no
NASA	National Aeronautics and Space Administration	yes
NCEP	National Center of Environmental Prediction	yes
NDACC	Network for the Detection of Atmospheric Composition Change	yes
NIR	Near InfraRed	no
NOAA	National Oceanic and Atmospheric Administration	yes
NRT	Near Real Time	no
NTC	Non Time-Critical	no

OWI	Ocean Wind Fields	no
PDF	Portable Document Format	yes
POCA	Point of Closest Approach	no
POD	Precise Orbit Determination	no
PRF	Pulse Repetition Frequency	no
PSF	Point Spread Function	no
PUM	Product User Manual	no
QA	Quality Assurance	no
QC	Quality Control	no
QWG	Quality Working Group	yes
R&D	Research and Development	no
RAMI	RAdiation transfer Model Intercomparison	yes
RTM	Radiative Transfer Model	no
S1 DH	Dual HH+HV polarization	no
S1 DV	Dual VH+VV polarization	no
S1 EW	Extra-Wide swath	no
S1 IW	Interferometric Wide swath	no
S1 RVL	Ocean Radial Velocity	no
S1 SM	StripMap	no
S1 TOPS	Terrain Observation with Progressive Scans	no
S1 WV	Wave mode	no
S2 SCL	Scene Classification Layer	no
S3 FRP	Fire Radiative Power	no
S3 LST	Land Surface Temperature	no
S3 OLTC	Open Loop Tracking Command	no
S3 SST	Sea Surface Temperature	no
SAR	Synthetic Aperture Radar	no
SDR	Surface Directional Reflectance	no
SIF	Sun Induced Fluorescence	no
SRIX4VEG	Surface Reflectance Intercomparison eXercise for Vegetation	yes
SSB	Sea State Bias	no
SSHA	Sea Surface Height Anomaly	no
SVT	Sentinel Validation Team	yes
SWH	Significant Wave Height	no
SWIR	Short Wavelength InfraRed	no
TIR	Thermal InfraRed	no
TOA	Top Of Atmosphere	no
UAV	Unmanned Aerial Vehicle	no
ULM	Ultra Léger Motorisé	no
VIS	Visible spectral range	no
WMO	World Meteorological Organization	Yes
WV	Water Vapour	no

1 Introduction

1.1 Scope of the document

This document presents the synthesis of activities performed in Task 3 of the CCVS project.

The document provides an overview of the current organization of the Cal/Val within the Copernicus project and a global assessment of its maturity.

A holistic cal/val strategy is proposed for the programme which relies on common principles for all components. To achieve this strategy, the CCVS project formulates a number of recommendations, some which are specific to a particular domain, some which are common to all domains, and some which concern organization aspects and collaboration activities.

Programmatic and sustainability aspects are not addressed in this document (cf. Task 4 documents).

1.2 Structure of the document

The document starts with a description of the current actors involved in cal/val activities in the Copernicus program and their respective contributions.

Chapter 3 is an assessment of the maturity of current cal/val activities for Sentinel missions. The purpose of this analysis is to identify the weak points on which progress is needed to achieve a higher maturity level. This “top-down” approach complements the “bottom-up” approach led in of the work packages where avenues for progress are identified on the basis of current technical bottlenecks.

The next Chapter 4 describes the proposed cal/val solution. This chapter examines in turns all the relevant activities: sensor characterization and calibration, validation of the calibration, data validation and performance assessment, and uncertainty characterization. For each activity, we identify the main gaps highlighted by the different tasks for each observation domain, and we propose a solution to address this gap.

Finally, we address non-technical aspects (organization, processes and standards) for which we propose ideas to improve the overall efficiency of cal/val activities.

2 Current organization for the Cal/Val Activities of the Copernicus program

2.1 Introduction

This chapter provides an overview of the current organization of Cal/Val activities for Sentinel missions. It provides the general context in which the recommendations of the CCVS project are presented.

For each Sentinel mission, a set of cal/val requirements and an associated cal/val plan is initially defined. The implementation of the operational cal/val activities are under the responsibilities of the Space Agencies (ESA/EUMETSAT) in the frame of the missions Ground Segment. The activities rely on source data which come partly from the Copernicus program and partly from other funding sources.

Other entities also contribute to the validation of Sentinel data products through independent activities. Different forums allow exchanges on validation results.

The following presentation does not take into consideration funding aspects: this point is addressed in detail in D4.1.

2.2 ESA

2.2.1 Introduction

The European Space Agency is currently responsible for the Calibration and Validation of the Sentinel missions operated by ESA: Sentinel-1, Sentinel-2, Sentinel-5P. For Sentinel-3, the responsibility is shared with EUMETSAT. The cal/val activities are mostly performed in the frame of the Mission Performance Clusters (formerly Mission Performance Centers).

ESA, together with EUMETSAT, is also responsible for the organization of the Sentinel Validation Teams and Quality Working Groups (see section 2.5.1 and 2.5.2).

ESA is also contributing to the Sentinel validation through supporting activities like the “FRM4” R&D projects and data distribution projects, and through synergies with ESA Earth Explorer and Earth Watch programs.

ESA CCI projects also contribute to the validation of some Sentinel variables, either for new data products or for the evolution of existing data products.

2.2.2 Mission Performance Clusters

With the first set of Sentinel missions, 4 Mission Performance Centers (S1, S2, S3 and S5P) were set-up by ESA to ensure the overall management of the mission data quality aspects, including:

- Quality Control and Data Anomaly management
- Sensor calibration and monitoring

- Processor parameterization and perfective maintenance
- Data product Performance assessment and validation
- User support and product documentation management
- Proposition for product and algorithm evolutions

Starting from 2021, four thematic Mission Performance Clusters (Radar imaging, Optical imaging, Altimetry and Atmospheric Composition) replaced the initial MPCs. The scope of the activity is basically unchanged, with a stronger emphasis on product uncertainty assessment.

To perform the calibration and validation activities, the MPCs rely on:

- On-board calibration data
- Specific ground-based calibration data in some cases (see next section)
- Vicarious acquisitions
- Publicly available ground-based reference data, or data obtained through agreement with the Copernicus program (e.g., Railroad Valley radiometer data provided by NASA).
- Publicly available sensor or model data.

The outputs of the MPC calibration and validation activities include:

- Operational calibration files
- Operational validation results (available through public and internal websites)
- Publicly available periodic data quality and performance reports
- Reference publications and communications, including participation to the Quality Working Groups, Sentinel Validation Teams and CEOS/WGCV meetings.

2.2.3 Operational FRM for Copernicus

ESA operates some operational calibration and validation facilities supporting Sentinel missions:

- The Permanent Altimetry Calibration Facility for the Sentinel-3 altimeter
- The Sentinel-1 calibration facilities (corner reflectors and transponders) operated in the frame of S1 MPC and then SAR MPC. This set of calibration devices is complemented by elements made available through international collaboration (refer to section 2.6)
- The Gobabeb radiometric validation site, part of RadCalNet, is maintained by ESA and is used by Sentinel-2 and Sentinel-3 missions
- Operational services to provide FRM from DOAS and GHG instruments for the Copernicus atmospheric missions is planned for the near future.

2.2.4 FRM and R&D projects

Over the years, ESA has managed several “FRM4” R&D projects which benefit the calibration and validation activities. The “FRM4” projects combine:

- Laboratory Calibration Exercise for ground-based reference sensors
- Field Inter-Comparison Exercises
- Uncertainty analysis
- Recommendations on measurement protocols
- In a few cases, establishment of a central processing facility for the FRM data (e.g., FRM4DOAS central processing for a network of MAX-DOAS instruments) or data distribution services (ST3TART)

More precisely, the following projects are relevant for the sentinel missions:

- FRM4STS for Sentinel-3 SST and LST (follow-on with the EUMETSAT study TRUSTED for SST)
- FRM4SOC for Sentinel-3 Ocean Colour products (follow-on with the EUMETSAT sturdy FRM4ASOC V2)
- FRM4ALT for Sentinel-3 altimetry mission
- FRM4SAR for Sentinel-1
- FRM4DOAS, FRM4GHG and Pandonia/PGN FRM benefit several S5P data products
- FRM4VEG contributes to the validation of Sentinel-3 OGVI/OTCI products, and Sentinel-2 L2A and S3 SYN reflectance products
- St3TART contributes to the validation of the Sentinel-3 in-land water and cryosphere altimetry mission

In addition, some dedicated campaigns for Sentinel data validation can be organized: e.g. FIDEX aerial campaign for Sentinel-3 fire products.

Other R&D activities relevant to Sentinel cal/val are implemented in the frame of the SEOM (Scientific Exploitation of Operational Missions) or EO for Science and Society programs (e.g. SEOM S2RADVAL).

2.2.5 Validation data archiving and distribution services

ESA is providing long-term support to the EVDC (ESA atmospheric Validation Database) service which provides access to reference validation data (collected from field campaigns, from established ground-based measurement networks, and from other fixed instruments) for atmospheric composition missions. Data from EVDC is used for the validation of S5P both in the operational routine validation service of the S5P MPC and for in-depth validation and algorithm evolution studies.

ESA also contributes to the Pandonia Global Network program (PGN, in collaboration with NASA) which processes and distribute atmospheric column data used in the validation several S5P data products.

ESA (with Copernicus funding) also supported the LAW project (Land surface temperature, Aerosols and Water vapour) which contributes to the validation of the corresponding Sentinel-3 products.

2.2.6 Benchmark projects

ESA is also organizing some benchmarking or inter-comparison exercises which are contributing to the validation of the Sentinel products, in particular:

- Atmospheric Correction Intercomparison Exercise (ACIX) contributes to the validation of Sentinel-2 Level 2A (Surface Reflectance) products
- Cloud Masking Intercomparison Exercise (CMIX) contributes to the validation of the Sentinel-2 Level-2A Scene Classification Layer
- Together with JAXA and NASA, ESA develops an international validation protocol for cloud and aerosol profile data (under the umbrella of the CEOS WGCV ACSG).
- Surface Reflectance Intercomparison eXercise for VEGetation (SRIX4VEG), part of the FRM4VEG project, contributes to defining a common protocol for surface reflectance validation using UAV-mounted imagers.

2.3 EUMETSAT

2.3.1 EUMETSAT Cal/Val activities and operational FRM

EUMETSAT is directly responsible for the calibration and validation of Sentinel-6 (altimetry L1 products) and shares activities with ESA on Sentinel-3. EUMETSAT contributes to the organization of the corresponding Sentinel Validation Team and QWG (see 2.5.1 and 2.5.2).

EUMETSAT is also directly responsible for the calibration and validation of the upcoming atmospheric Sentinels 4 and 5 during Phase E2, and shares activities with ESA for the Phase E1 validation of these missions.

EUMETSAT publishes Product Notices and Collection Reports (in case of reprocessing) documenting product evolutions and performance assessments, as well as scientific publications and communications.

A dedicated operational FRM facility for System Vicarious Calibration of Ocean Colour products (OC-SVC) is currently under development under the supervision of EUMETSAT. This facility is a follow-on of the BOUSSOLE facility used for MERIS and the first years of OLCI.

EUMETSAT is currently leading a study for the preparation of the validation of the future CO2M missions.

The EUMETSAT Satellite Application Facilities (SAF) carry out monitoring and validation of current EUMETSAT data satellites. The new cycle 2022-2027 of the Atmospheric Composition SAF (AC SAF)

Continuous Development and Operations Phase (CDOP-4) foresees the validation of non-Copernicus data products from Sentinels 4 and 5, e.g., water vapour column data. These products could possibly become official Copernicus data products at a later stage.

2.3.2 FRM projects and R&D activities

EUMETSAT supports R&D project on reference measurements similar to the ESA FRM4 projects:

- FRM4SOC-2 dedicated to Ocean Colour
- TRUSTED for Sea Surface Temperature – this project will evolve toward an operational data provision service
- EUMETSAT also leads various ad-hoc R&D studies to support the improvement of calibration and validation facilities and methodologies.

2.3.3 Validation Data distribution services

The Ocean Colour Database (OC-DB) supported by EUMETSAT provides validation data for Ocean Colour.

2.4 Copernicus services

2.4.1 Relevant validation activities

Copernicus Services do not contribute directly to the validation of Sentinel core products. However, there is some overlap with activities performed for products of Copernicus Services.

For instance, the GBOV service provides to the CLMS in-situ reference data for the validation of LST from Meteosat satellites which could in some cases be used for Sentinel-3 validation. Similarly, the CMEMS in situ TAC maintains an archive of in-situ data which can be used for validation of Ocean remote sensing products.

To some extent, the same reference data are used for validation of Copernicus Atmosphere Monitoring Service (CAMS) products, Copernicus Climate Change Service (C3S) climate data records, and S5P data products. Recently an agreement was reached between ECMWF (entrusted entity for CAMS) and ESA for a shared support of this data provision activity.

The validation data acquired by C3S for the validation of Ozone concentration Level3 products could be used as well for Sentinel product.

Finally, let us mention that the C3S maintains a long-term archive of ground-based in-situ and remote sensing observations which can be used as a reference for assessing trend estimates from Sentinel missions.

2.4.2 EEA cross-cutting activities

EEA is responsible for the Copernicus In-Situ component. More precisely, EEA:

- Maintains the "state of play": overview of in-situ data requirements, use and challenge within the Copernicus program
- Ensures data access for Copernicus services
- Engaging with data providers, especially at international level
- Supports and advises the EC and entrusted entities regarding in-situ data.

Although not a prime objective EEA may provide support and facilitate the coordination of activities related to improving access to and availability of in situ and FRM observations needed for cal/val of current and future Sentinel missions.

2.5 Collaborative activities

2.5.1 Sentinel Validation Teams

The Sentinel Validation Teams gather experts from different institutions which contribute to the validation of Sentinel products. SVT meetings are held once per year approximately to exchange information, compare results and product quality assessment, discuss initial validation of new data products and inform the go/no-go decision for public release of the new data, and suggest new validation methods and activities.

Participants are selected through an open call and generally include European and International Space Agencies, other governmental agencies, remote sensing companies and scientific research organisations.

Outcomes of the SVT team is used to improve operational cal/val activities. Validation Teams currently exist for S2, S3, S5P, and S6. There is no S1 Validation Team, but similar discussions are held in the frame of CEOS/WGCV for SAR.

2.5.2 Quality Working Groups and Mission Performance Working Groups

The main objective of the Sentinel Quality Working Groups is to advise the mission manager on Product Quality and suggest evolutions and improvements. Participants of the QWG are key users (in practice the representatives of the Copernicus Services) and MPC representatives. There are QWGs for S1, S2, S3 and S5P, while a Mission Performance Working Group exists for Sentinel-6. The S6 MPWG gathers experts from agencies collaborating on the mission (ESA, NOAA, NASA, CNES, EUMETSAT) rather than key users. There is also an L2 expert team for Sentinel-5P L2 products algorithm evolution (L2-ALG) and processor evolutive maintenance (L2-PRO).

2.5.3 National agencies

Member states provide in-kind contributions to the Copernicus program, in particular through the provision of reference in-situ data. Moreover some national space agencies are directly performing CalVal activities for Sentinel missions.

CNES is maintaining a radiometric validation activity on Sentinel-2 and Sentinel-3 optical products. The validation activity relies among other sources on in-situ radiometer measurements (ROSAS system deployed at La Crau and Gobabeb) processed by CNES. In addition, cross-validation with CNES missions (Venus, PLEIADES, SPOT) provide a unique indirect validation source for Sentinel missions.

CNES contributes to in-situ measurement networks for atmospheric composition, such as AERONET, and maintains a data distribution network (Data Terra).

The CNES MAGIC campaign was designed to support several greenhouse gases EO products, including Sentinel-5P products.

CNES is also strongly involved in the CAL/VAL of the altimetry missions (Sentinel-3 and Sentinel-6). For Sentinel-3, CNES is responsible of the commissioning phase and acts after as expert during all the mission lifetime. For Sentinel-6, CNES is in charge of several work packages of the commissioning phase which is under the lead of EUMETSAT. DLR performs radiometric validation campaigns such as the Lake Stechlin which has been used by Sentinel-2. Another example of DLR-led campaign is the Methane-to-go aerial campaign for the validation of the Sentinel-5P SO2 products.

Federal Scientific Institutes (FSI) working under authority of the Belgian Federal Science Policy Office (BELSPO) like BIRA-IASB and IRM-KMI contribute to in-situ measurement networks for atmospheric composition: AERONET, COCCON, GO3OS, NDACC UV-VIS, NDACC FTIR, PGN, and TCCON. They operate multi-platform validation systems, some of these being used in an operational environment for routine validation of ESA and EUMETSAT satellites and the Copernicus services C3S and CAMS. In particular, BELSPO co-funds Sentinel-5P validation activities in support to the ATM MPC and AC SAF activities. BIRA-IASB also operates UAV-based and other mobile in-situ instrumentation for atmospheric measurements.

RBINS, another FSI under BELSPO authority, performs marine in-situ measurements over the North Sea, operates the Belgian Marine Data Centre (BMDC), and contributes to international infrastructures like SeaDataNet which are used for validation of several Sentinel products and the Copernicus Marine Environment Monitoring Service (CMEMS).

The Atlantic Meridional Transect (AMT) cruises led by the NERC (UK) contributes to the validation of several Sentinel products (ocean colour, sea surface temperature).

2.6 International collaborations

International collaborations contributing to the cal/val of Sentinel missions include:

CEOS/WGCV activities and its validation subgroups contribute to the advancement of the cal/val of Sentinel missions: (standards, protocols and best practices, guidelines, R&D, communication, inter-comparisons, tutorials, cross-domain synergies etc.)

The GSICS is an international group working under the umbrella of WMO and CGMS and dedicated to the inter-calibration of operational meteorological satellites.

NASA performs independent validation activities on some Sentinel missions; regular exchanges about performance assessment or other topics are held in the frame of QWG, SVT, or dedicated forums. Similar exchanges exist with other US agencies (NOAA, USGS).

Collaborations exist with other non-European space agencies: e.g., operation of SAR calibration networks (Australia), and regular interaction with ground-based measurement networks of atmospheric composition (including their participation in calls of ESA and EUMETSAT).

Opportunistic usage of calibration device is also in place: This includes for instance using the network of corner reflector deployed by Nasa in the Rosamond site (https://trs.jpl.nasa.gov/handle/2014/48764) or the network of buoys from the NOAA/NDBC (https://www.ndbc.noaa.gov/)

3 Calibration and Validation Status Assessment

3.1 Methodology

For the assessment of the cal/val maturity of the different Copernicus missions, we have used a specific maturity matrix defined by the project. This matrix is derived from existing matrices, in particular the WGISS Data Management and Stewardship Maturity Matrix (DMSMM) and the EDAP cal/val maturity matrix. Some modifications were needed to reflect the aims of the CCVS project. The matrix comprises 4 main columns with several sub-items:

Calibration and characterization of the sensor

- Pre-flight characterization and calibration
- In flight characterization and calibration
- Geometric calibration

Product quality information

- Auxiliary data quality
- Prognostic (or ex-ante) Uncertainties
- Quality flags

Product validation

- Reference data representativeness
- Reference data quality

  - Validation method
  - Validation results

Quality management processes

  - QC and anomaly management
  - Data quality reporting
  - Feedback to ref. and aux data providers
  - Feedback from users
  - Products and algorithms evolutions
  - Reprocessing strategy
  - Inter-operability

Product quality and validation statuses can be evaluated for a group of products, for a single product, or for several aspects of one product. For instance, a given product may have a high validation maturity for the instantaneous error but a low maturity for the validation of the long-term drift.

For each item we define four levels of maturity: basic, good, excellent, ideal. The levels are defined in the following tables.

Cal and char	**pre-flight char. and cal**	**In-flight characterization**	**in-flight measurement calibration**	**Geometric calibration**
Poor	No pre-flight char. and cal	no characterization	no calibration	not assessed
Good	Partial, at component level	partial characterization	Periodic calibration	periodic calibration
Excellent	Complete at component level + documented SI traceability	periodic characterization + monitoring	Periodic calibration + monitoring	periodic calibration + monitoring

Ideal	instrument level + SI traceability	Periodic characterization + independent assessment	Periodic calibration + independent assessment	periodic calibration + independent assessment

Product Quality information	**Auxiliary data quality**	**Uncertainties: method**	**Uncertainties: sources**	**Uncertainties: values**	**Quality flags**
Poor	Unsuitable auxiliary data quality	no uncertainty char	no uncertainty char	no uncertainty char	no quality flags
Good	Suitable auxiliary data quality	Partial GUM	Most important sources	Single value	quality metadata
Excellent	id. + basic uncertainties provided	GUM approach	All important sources	Per pixel uncertainty basic components	quality layer (per measurement) - most defective measurements flagged
ideal	Fully characterized uncertainties and SI traceability	GUM + error covariance	All sources	Per pixel + error covariances for all components	idem + all defective measurement flagged

Validation	**Ref data representativeness**	**Ref data quality**	**Validation method**	**Validation results**
Poor	Sparse dataset	No uncertainty specified	No or sporadic validation	no regular validation results
Good	mostly representative	single uncertainty for dataset	simple uncertainty estimate	good agreement
Excellent	well representative	FRM level	assesses sat and ref data uncertainties	excellent agreement
Ideal	fully representative	FRM level and error correlation information	error covariance assessment	uncertainty validated

Quality Management	**QC and anomaly management**	**Reporting**	**Feedback to ref/aux data providers**	**Feedback from users**	**Products and algorithm evolutions**	**Reprocessing strategy**	**Inter-operability**
Poor	no QC	no data quality reporting	No feedback	No feedback	No evolutions / frozen software	No reprocessing	Not managed
Good	Basic QC + major anomalies traced	Occasional publications & reports	Ad-hoc exchanges	User support desk	Bug correction and minor evolutions	Ad-hoc reprocessings for anomaly recovery	Adjustment factors computed and documented
Excellent	All anomalies traced	Regular periodic reports	periodic meeting	Periodic user workshop	evolutions partially managed	Partial reprocessings	Harmonization implemented
Ideal	+ anomaly correction/ mitigation actions	Regular On-line reporting	periodic working groups	Quality working groups	Evolution cycle fully managed	Collection strategy with regular reprocessings	Harmonization implemented and archive reprocessed

For each element in the matrix, we provide a comment explaining which is the most critical item that needs to be improved in order to increase the overall maturity level.

3.2 Optical component

3.2.1 Calibration and characterization

Table 1 : Optical mission characterization and calibration status

	Pre-flight characterization	In-flight characterization	Measurement Calibration	Geometric calibration	Overall	Comment / Weak link

S2 MSI (A and B)	Good	Good	Ideal	Ideal	Excellent	Lack of documentation of characterization uncertainties and traceability No in-flight spectral characterization
S3 OLCI (A and B)	Good	Excellent	Ideal	Excellent	Excellent	High uncertainty on pre-flight characterization
S3 SLSTR (A and B)	Good	Excellent	Excellent	Good	Excellent	High uncertainty on pre-flight characterization Inter-channel geometric calibration issue Saturation values lower than specified

Generally speaking, pre-flight characterization is identified as the main weak point on this aspect. First, there is a lack of documentation on instrument characterization in terms of uncertainties and SI traceability. Second, pre-flight characterization is sometimes inconsistent with observed in-flight performances. This may-be due to errors during pre-flight characterization or underestimated launch effects.

Relevant examples are:

- Significant radiometric biases observed using vicarious methods for some sensors (SLSTR, OLCI-A).
- Seasonal effects observed on radiometric calibration, pointing to solar diffuser BRDF errors (S2, OLCI)
- Inconsistencies between straylight models and observation on lunar acquisitions (OLCI)
- This calls for a characterization strategy relying more heavily on in-flight characterization using e.g.:
- Tandem phases to assess radiometric calibration biases

- Intercomparison with reference sensor (e.g. TRUTHS) for future generation Sentinels, restoring a full SI traceability chain
- Characterization of diffuser BRDF using yaw maneuvers
- Tuning of straylight models using lunar images (a methodology needs to be developed for this)
- The spectral characterization of the S2 MSI relies only on pre-flight assessment and no monitoring of the stability of the spectral response is performed. However, there is no evidence that this is impacting the quality of the S2 products.

Within the limits mentioned above, in-flight operational calibration is generally satisfactory. However due to the impact of launch, the SI traceability chain is inevitably broken.

Regarding geometric calibration, the only limitation identified concerns the TIR channels. Geometric calibration of TIR images is made difficult by the lack of permanent geometric features. Several methodologies have been proposed (use of a database of gas flares, correlation between optical and thermal channels e.g. on land/water interfaces, use of higher resolution reference such as LANDSAT/TIRS) but they are not operationally used today. This issue will be even more critical for LSTM.

3.2.2 S2 MSI product quality information

Table 2 : Sentinel-2 quality information status

	Auxiliary data quality	Prognostic uncertainty method	Prognostic uncertainty sources	Prognostic uncertainty values	Quality Indicator	Overall	Comment / Weak link
L1C	Excellent	Good	Good	Excellent	Excellent	Excellent	Prognostic uncertainties limited to radiometric random – user-side generation Some quality information missing
L2A	Good	Poor	Poor	Poor	Excellent	Good	No prognostic uncertainties Some quality information missing

The S2 MSI products are provided with relatively detailed quality layers. Some information is still missing (traceability of missing source packets and saturated values) but improvements are in progress.

Prognostic uncertainties are provided through a post-processing tool (Radiometric Uncertainty Tool). Improvement of the tool and extension to L2A product are in progress.

3.2.3 S2 MSI products validation status

Table 3 : Sentinel-2 Validation Status

	Ref. Data Representativeness	Ref. Data Quality	Validation Method	Validation Results	Overall	Comment / Weak link
L1 geometry	Excellent	Good	Excellent	Excellent	Excellent	Geometric uncertainty assessment per product needed
L1 Radiometry	Good	Excellent	Good	Excellent	Excellent	No validation of prognostic uncertainties
L2 SDR	Poor	Good	Excellent	Excellent	Good	Lack of SI traceable reference measurements
L2 SCL	Poor	Poor	Good	Good	Poor Good	Validation methodology needs consolidation
L2 AOD	Excellent	Good	Excellent	Good	Good	No prognostic uncertainties
L2 WV	Excellent	Good	Excellent	Excellent	Excellent	No prognostic uncertainties

The overview of the validation maturity for Sentinel-2 L1C and L2A products is summarized in Table 3.

For the Level 1 geometry, high quality GCPs are used for the validation. These points are independent from the ones used for the calibration. The main progress that could be done to reach "ideal" level would be to provide in the products a local assessment of the geometric uncertainty (currently only a global uncertainty is provided). This could be done by performing an on-line validation after the geometric refinement step.

The validation of Level 1 and L2A directional surface reflectances are mainly limited by the lack of SI traceable reference measurements. On the one end, vicarious validation methods provide only partial uncertainty estimates and generally no SI traceability. On the other hand, there is a lack of systematic ground-based reference measurements. A data quality assessment methodology using synthetic data has been proposed in the frame of the ACIX workshop, but the uncertainty associated to this methodology is not quantified. Level 1 radiometric per pixel uncertainties can be computed with the Radiometric Uncertainty Tool, but these uncertainties are not currently validated. Level 2 uncertainties are currently not available. Validation results for Sentinel-2 are excellent at L1 and satisfactory at L2 for most spectral bands and in most conditions. A correction has been applied on S2B radiometry to align it with that of S2A in the VISNIR domain.

The scene classification layer (SCL) is subject to an intensive validation activity, which is in line with higher user expectations on this important information. However, validations activities are currently limited by a) a lack of precisely defined performance targets (including a clear definition of when a pixel should be considered as cloudy or clear) b) lack of a commonly agreed way to express confidence in classification results, which would play a similar role to measurement uncertainties for continuous measurands c) a lack of automatic validation process.

AOD and WV validation benefit from the availability of reference data and show generally satisfactory results, although higher-quality aerosols measurements would be needed ideally (uncertainty below 10%). To progress further, per pixel uncertainty estimates would be needed. Note however that AOD and WV are considered as by-product so a detailed assessment of the uncertainty may not be necessary.

3.2.4 S3 OLCI /SYN Product quality information

Table 4 : OLCI quality information status

	Auxiliary data quality	Prognostic uncertainty method	Prognostic uncertainty sources	Prognostic uncertainty values	Quality flags	Overall	Comment / Weak link
L1b	Good	Good	Good	Excellent	Good	Good	Assessment based on planned deployment Partial saturation information missing

L2 Land products	Good	Good	Poor	Good	Good	Good	Propagation of L1 uncertainties to be done
L2 Marine products	Poor	Good	Poor	Good	Good	Poor	Lack of System Vicarious Calibration data Propagation of L1 uncertainties to be done
SYN products	Poor	Poor	Poor	Poor	Good	Poor	Lack of updated SRF for SYN VGT No prognostic uncertainties Cloud shadow information missing

The OLCI products are provided with detailed quality indicators. Partial detector saturation is currently not traced but this improvement is in progress. For SYN and L2 products, a cloud shadow mask is missing. SYN products

Prognostics uncertainties (per pixel) have been developed for OLCI L1b. The uncertainties are limited to random radiometric terms. Although prognostic uncertainties are provided for L2 products, they do not rely currently on propagation from L1 uncertainties as recommended by the GUM. An empirical L1 uncertainty model is used instead. Another limitation of the L2 water product is the current lack of an operational system vicarious facility, although the deployment of new facility is planned for the near future.

3.2.5 S3 OLCI products validation status

Table 5 : Sentinel-3 OLCI products validation status

	Ref. Data Representativeness	Ref. Data Quality	Validation Method	Validation Results	Overall	Comment / Weak link

L1b geometry	Excellent	Good	Good	Good	Good	Independent set of GCPs to be used for validation
L1b radiometry	Good	Excellent	Good	Good	Good	OLCI-A partially compliant with specifications
L2 Land GIFAPAR	Good	Good	Good	Good	Good	Limited sampling of certain biomes. Lack of validated satellite data uncertainties. In progress
L2 Land OTCI	Good	Good	Good	Good	Good	Validation methodology needs consolidation
L2 Land IWV	Excellent	Good	Good	Good	Good	No uncertainty validation - poor data quality over ocean
L2 Marine RRS/Chl	Good	Good	Good	Good	Good	Lack of validation data on open ocean - partial validation of uncertainties – partial compliance at shorter WL
L2 Marine Other variables	Poor	Poor	Poor	Poor	Poor	Lack of reference data

The validation status of OLCI products is summarized in Table 5.

The geometric uncertainty validation is performed routinely using a set of suitable reference GCPs. However, the same GCPs are used to perform the validation and update the geometric calibration. A possible improvement would be to separate GCPs used for validation from those used for calibration and monitoring.

As for Sentinel-2, L1b radiometry validation is limited by the availability of relevant reference measurements either from vicarious sources or on-ground measurements. Per pixel uncertainties are not currently provided operationally at L1b, but an upgrade is planned for the near future. These uncertainties could be used in future validation activities. Finally, the validation maturity is limited by the partial compliance of OLCI-A radiometric performance due to a relatively high positive bias. An operational correction of this bias would provide an improvement of the maturity level and increase inter-operability.

The validation stage of land products (GI-FAPAR and OTCI) is at an intermediate level and is expected to improve in the near future. New reference measurements are now operationally available for FAPAR thanks to the GBOV service, but they are not yet used on a routine basis for the validation of OLCI products. Some reference measurements for OTCI are available (e.g. from FRM4VEG campaigns) but the validation methodology needs consolidation. In both cases, the current validation relies mostly on comparisons with other satellite or climatology.

The validation of the OLCI marine products has been assessed from the validation reports issued for Collection 3. The validation relies on several data sources, including high quality data from the AERONET-OC network. This data set is limited to coastal areas and therefore not fully representative. Further progress could be made by consolidating (using Level 1 prognostic uncertainties when available) and assessing per-pixel uncertainties. The validation results are generally satisfactory, although some non-compliances have been noted for shorter wavelengths. Finally, no validation results are provided for other marine variables (KD490, PAR, TSM) due to a lack of reference measurements. Intercomparisons with other satellite data are not reported.

3.2.6 S3 SYN products validation status

Table 6 : Sentinel-3 SYN product validation status

	Ref. Data Representativeness	Ref. Data Quality	Validation Method	Validation Results	Overall	Comment / Weak link
SYN L2 SDR Land	Poor	Poor	Poor	Poor	Poor	No operational validation
SYN L2 Aerosols	Poor	Poor	Poor	Poor	Poor	No operational validation
SYN VGT TOA	Poor	Poor	Poor	Poor	Poor	No operational NRT validation
SYN VGT SDR	Poor	Poor	Poor	Poor	Poor	Intercomparison with PROBA-V
SYN AOD aerosols	Excellent	Good	Excellent	Good	Excellent	Partial uncertainty validation

The validation status of OLCI SYN products is summarized in Table 6. This status is highly contrasted, with some variables having no routine validation at all, while the SYN AOD aerosol product presents one of the most advanced levels for all Sentinel optical data products.

No operational validation of SYN L2 SDR product exists today, but a methodology relying on intercomparison with MODIS is in developments. This is mainly due to the low maturity of the product itself, with an on-going evolution of the retrieval algorithm.

The SYN VGT product has been extensively compared with PROBA-V data during the overlap between the two missions. However, since the end of PROBA-V in 2020, no operational validation activity is performed.

The SYN AOD product by contrast has been subject to a very exhaustive validation activity first in the frame of the ESA LAW project and now in the OPT-MPC. The validation uses high quality reference data provided by the AERONET network. SYN AOD products are provided with per pixel uncertainty estimates, and the quality of these estimates has been assessed, which is a rather unique situation among operational Sentinel optical products. However, the uncertainty estimates are only partially representative, and the validation results highlighted discrepancies in some cases. The SYN AOD product contains a global SDR layer which is not validated in an operational way as it is not considered a core data but a by-product of the retrieval. However inter-comparisons with MODIS data have been performed and reported previously.

3.2.7 S3 SLSTR products quality information

Table 7 : SLSTR quality information status

	Auxiliary data quality	Prognostic uncertainty method	Prognostic uncertainty sources	Prognostic uncertainty values	Quality flags	Overall	Comment / Weak link
L1b	Good	Poor	Poor	Poor	Good	Poor	No prognostic uncertainties provided
L2 LST	Good	Good	Good	Excellent	Excellent	Good	
L2 SST	Good	Good	Good	Excellent	Excellent	Good	
L2 FRP	Good	Good	Good	Excellent	Good	Good	F1 channel overshoot not flagged

The SLSTR L1b product does not include per pixel uncertainties, but a post-processing tool allows the estimation of L1 uncertainties which are propagated to L2 products. L2 products are provided with Quality indicators and per pixel uncertainties.

3.2.8 S3 SLSTR products validation status

Table 8: SLSTR product validation status

	Ref. Data Representativeness	Ref. Data Quality	Validation Method	Validation Results	Overall	Comment / Weak link
L1 geometry	Excellent	Good	Good	Good	Good	Independent set of GCPs to be used for validation Lack of GCPs for thermal channels
L1 radiometry VIS-SWIR	Good	Good	Good	Poor	Good	NC on SWIR channels
L1 radiometry TIR	Poor	Poor	Poor	Poor	Poor	No operational validation
L2 LST	Excellent	Good	Excellent	Excellent	Excellent	
L2 SST	Good	Good	Good	Excellent	Good	No recent publicly available validation report Assessment based on last S3VT presentation
L2 FRP NRT	Good	Poor	Poor	Good	Poor	Inter-comparisons with MODIS/TERRA and MSG/SEVIRI products
L2 FRP NTC	Good	Poor	Poor/Good	Good	Poor	Inter-comparisons with MODIS products + sparse campaign reference data

The L1 geometry validation of SLSTR is performed on the S3 (NIR) channel using a set of high quality GCPs as for OLCI. In addition, inter-channel co-registrations analyses highlighted a systematic offset of approximately 100 m for the TIR channels.

The validation of the L1 radiometry for VIS-SWIR channels relies on vicarious methods similar to those used for OLCI. Per pixels uncertainties are not provided in the products but can be computed with an off-line tool. These uncertainties are not yet validated on a routine basis, but efforts in this direction are in progress. The validation results show a significant non-compliance on the radiometry of SWIR channels. The correction of the offset is not yet implemented operationally. This correction would require a complete reprocessing of the time series and some adaptations for downstream products users.

The L1 radiometry of TIR bands is not validated on a routine basis due to a lack of suitable reference data. However, some inter-comparisons with IASI have been performed in the past and showed satisfactory results.

The L2 LST product is validated using an extensive set of high quality in-situ measurements. The validation results are generally within specifications. Although per pixel uncertainties are provided in the products, there is no systematic validation of these uncertainties yet.

The validation of the L2 SST product benefits from the development of the high-quality buoys from the TRUSTED project. Available results presented during S3VT meetings show a very good agreement of the SLSTR products with reference measurements. However, we note that there is no publicly available validation report.

No consistent validation report for the FRP NRT product is currently available, only limited results and the intent to setup a GEO-LEO cal/val framework based on the comparison with MSG/SEVIRI detections.

The FRP NTC product validation is limited by the scarcity of reference data (FIDEX campaigns). In addition, inter-comparisons with MODIS products are performed on a routine basis. Available results indicate the good quality of the data. As for other classification products, the Active Fire Pixel data would benefit from a better definition of performance targets and validation metrics. Per pixel uncertainties are not provided.

3.2.9 Optical missions Quality management

Table 9: SLSTR product validation status

	QC and anomaly mgmt.	Reporting	Feedback to ref. and aux data providers	Feedback from data users	Products and algorithms evolutions	Reprocessing Strategy	Inter-operability	Comment / Weak link
S2 MSI	Ideal	Excellent	Good	Ideal	Ideal	Excellent	Ideal	
S3 OLCI	Excellent	Good/ Excellent	Good	Ideal	Ideal	Good	Good	Vicarious calibration factors not applied at L1b

								Lack of systematic validation reporting for some variables
S3 SYN	Excellent	Excellent	Good	Ideal	Ideal	Poor	Poor	No “collection” reprocessing
S3 SLSTR	Excellent	Good/ Excellent	Good	Ideal	Ideal	Good	Good	Vicarious calibration factors not applied at L1b Lack of systematic validation reporting for some variables

The data quality management process for optical missions has a very high maturity. In particular, the Sentinel Validation teams and Quality Working Groups ensure an efficient process to manage data quality issues and product evolutions. Inter-comparisons with alternative processing chains are organized for Sentinel-2 processing. In some cases, a “collection” strategy has been devised to ensure a uniform reprocessing of the time series. For Sentinel-2 the Collection 1 reprocessing will implement a radiometric harmonization to improve inter-operability.

However, there is generally a lack of a consistent plan for systematic re-processings to generate uniform “collections”.

We also note that data quality reporting processes differ between ESA and EUMETSAT.

3.3 Altimetry component

3.3.1 Calibration and characterization

The calibration of the instruments reaches a very high level of maturity for the three instruments (radar altimeter, radiometer and POD) on-board altimetry mission, as stated in the table below. Nonetheless, there is still space for necessary improvements as explained in the comments. The final accuracy required on the topography (few mm) is very demanding on the instrument characterization, for which a perfect knowledge is sought.

The document [D3.1] states all the improvements we recommend to reach an even better characterization of the instruments. The table below only remind the major ones. The tandem phase is definitely a key process to reach a high continuity of measurements between the satellites and also a very high knowledge on the instruments.

Table 10 : Altimetry mission characterization and calibration status

	Pre-flight characterization	In-flight characterization	Measurement Calibration	Overall	Comment / Weak link
Altimeters	Ideal	Ideal	Ideal	Ideal	Some room for improvements on the antenna pattern characterization. Extension of the tandem phase to the RF redundant path (completely necessary for climate record in case of nominal path failure).
Radiometers	Ideal	Ideal	Ideal	Ideal	Design of the radiometers can still be improved to minimize and simplify the differences between measurement and calibration RF path.
DORIS/GNSS	Ideal	N/A (except for yaw/flip manoeuvre on Sentinel-6)	Ideal	Ideal	Yaw-flip manoeuvres for the POD (Attitude flips at low beta angles (0°, 10°) should be implemented to disentangle the centre of phase errors from time-tagging, dynamic modelling errors). It has been successfully tested on S6.

3.3.2 Product quality information

For Altimetry Level-2 products, there are relatively few quality indicators available. Indeed, for altimetry, the concept of quality information is subjective and not relevant because closely linked to the nature of the application (ocean topography, waves and wind, hydrology, sea ice, freeboard …). A Sea Surface Height (SSH) measurement considered as "bad" because of swell effect contamination will be of interest for other user communities studying waves. Usage of Level-2 altimetry products thus implies a minimum level of expertise to identify which variables and criteria can be used to determine a quality indicator that fits a given usage. For the most common applications (open ocean surface topography, wind and waves characterization and inland waters topography monitoring) dedicated post-processed turnkey products are generated and distributed by Copernicus Services (CMEMS, Global Land…).

Another information source about the product quality is provided through the delivery of thematic cyclic reports. This aspect is addressed in section 3.3.4.

Table 11 :Altimetry quality information status

	Prognostic uncertainty method	Prognostic uncertainty sources	Prognostic uncertainty values	Quality Indicator	Overall	Comment / Weak link
L1	Poor	Poor	Poor	Poor	Poor	
L2	Poor	Poor	Poor	Poor	Poor	

3.3.3 Product validation status

This section aims at summarizing the maturity of validation techniques used to qualify the altimetry level-2 products. As mentioned in the previous section, because the application fields are wide, and sometimes involve many (instrumental and modelled) geophysical variables, different tables have been defined to cover the main usages of altimetry data. A distinction is made regarding the source of the data used for validation. The following categories are identified:

- "In-situ" refers to on-ground (or close to ground) measurements.
- "Inter-satellites" refers to comparison between different satellites measuring the same geophysical phenomenon.

- "Models" refers to comparison with physical models. Cases of assimilation or non-assimilation of the altimetry measurements are not distinguished.
- "Alternative processing" refers to comparison between different ground processing solutions (from same instrument raw measurements).

Following subsections are organized as follows:

- The first subsection is dedicated to the ocean wind & wave parameters. They are more or less directly retrieved from altimeter measurements and do not involve as many variables as the other variables related to topography information.
- The second one addresses the orbit validation as it is a key parameter, involved in all topography variables whatever the considered surface, application.
- The third one relates to the geophysical corrections retrieved from models or instruments (MicroWave Radiometer (MWR)). Their involvement in topography estimations depends on the kind of application, on the surface type.
- The fourth subsection addresses the validation of ocean SSH measurements. A distinction is made for the different applications (operational oceanography, climate studies…) by splitting the temporal and spatial frequencies by categories.
- The fifth subsection is based on the same principle as the one dedicated to ocean topography and is dedicated to the Level-2 product validation over inland waters.
- The sixth subsection is based on the same principle as the one dedicated to ocean topography and is dedicated to the Level-2 product validation over sea ice regions.
- The seventh subsection is based on the same principle as the one dedicated to ocean topography and is dedicated to the Level-2 product validation over land ice regions.

3.3.3.1 Waves & Wind validation

Table 12 : product validation status for Wave and Wind

	Ref. Data Reprensentativeness				Ref. Data quality				Validation Method				Validation results				Overall
	In-situ	Inter-sat	Models	Alternative processing	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
Wind	Poor	Good	Ideal	Ideal	Good	Ecellent	Good	Good	Good	Good	Good	Good	Good	Ecellent	Ecellent	Ecellent	Good
Wave	Poor	Good	Ideal	Ideal	Ecellent	Ecellent	Good	Good	Good	Ecellent	Good	Good	Good	Ecellent	Ecellent	Ecellent	Good

Results reported are almost similar for both variables. Most important observations to note are the following:

- In-situ measurements constitute a reliable reference, they are however punctual and mainly localized in specific coastal areas. They are not representative of the open ocean sampled by satellites. Methods used to compare with satellite measurements can be refined to better account for the non-synopticity of the measurements. Comparisons with in-situ data are used to complete the results derived from other methods.

- Inter-satellite comparisons provide an almost good representativeness. Methods (based on crossover approach) to perform the comparisons are well established, even if the matchups definition relies on approximate assumptions on the time lag considered, accepted between the two satellite measurements. From the results obtained, it is however difficult to separate the error contributions coming from each instrument. It provides a relative quantification of errors, the absolute part should be estimated from other means (in-situ / models). The specific case of tandem phase is important as it provides a global and representative dataset for comparison, with the certainty (when time lags are low) that the same ocean & atmosphere conditions have been observed.

- Comparisons with models are largely used by the CalVal community. Representativeness is excellent (a reference values is computed for each satellite measurements) and comparisons methods are simple (direct difference). However, models also contain their own errors (limited temporal and spatial resolutions, approximations…) that are not quantified and could not be easily separated from altimetry errors. Moreover, most of the time they assimilate altimetry data, which could induce some error correlation in the comparisons.

- Alternative processings are useful to investigate algorithms errors. Algorithm modifications, evolutions are well mastered and their impact can be evaluated directly. However, it does not give access to the complete error estimation (errors may remain in the algorithm change).

To conclude, several validation methods exist, they present different advantages, drawbacks in each categories (representativeness, quality, methods used, results obtained). The best way to characterized altimetry wind and waves estimations relies on a mix of these different methods. The main complexity, for the validation of these variables lies in the combination of these methods and on the ponderation defined for each result obtained.

Orbit validation

Table 13 : product validation status for orbit

	Ref. Data Representativeness				Ref. Data quality				Validation Method				Validation results				Overall
	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
Altitude	Good	Ideal	Ideal	Ideal	Excellent	Excellent	Good	Excellent	Good	Good	Good	Good	Excellent	Excellent	Good	Excellent	Excellent

Representativeness of the data is excellent, nonetheless, we have not yet reached a full level of quality and some further improvements should be done to improve uncertainties assessment.

Results are excellent regarding their accuracy, but figures of uncertainties should be provided on a more systematic manner.

3.3.3.2 Geophysical corrections involved in topography calculation

Table 14 : product validation status for geophysical corrections

	Ref. Data Representativeness				Ref. Data quality				Validation Method				Validation results				Over all
	In-situ	Inter-sat	Models	Alternative processing	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
Ionosphere correction			Ideal	Ideal			Good	Good			Good	Good			Good	Good	Excellent
Sea State Bias correction				Ideal				Good				Good				Good	Excellent
Dry troposphere correction			Ideal				Good				Good				Good		Excellent
Wet troposphere correction	Poor	Excellent	Ideal	Ideal	Good	Excellent	Good	Good	Good	Excellent	Good	Good	Good	Excellent	Excellent	Excellent	Excellent
ocean tide correction	Poor		Ideal		Good		Good		Good		Good		Good		Good		Good
Solid earth tide correction			Ideal				Good				Good				Good		Excellent
pole tide correction			Ideal				Good				Good				Good		Excellent
Mean Sea Surface	Poor	Excellent	Ideal		Good	Good	Good		Good	Excellent	Good		Good	Good	Good		Good

Different maturity levels are observed. This is mainly related to the level of criticality of the correction (its impact on the final topography) and the way it is computed (from instrumental sensing parameters or simply retrieved from models). This criticality depends on the amplitude of the correction and its temporal and spatial variability. The wet troposphere correction is a particularly sensitive topic as its amplitude, its temporal and spatial variability are strong (that is why most of the altimetry missions onboard a Microwave radiometer). For the CalVal of instrumental corrections (ionosphere, wet troposphere and sea state bias), tandem phase configurations with a low time lag between the two satellites (o(1 min)) is very powerful and enable the detection and characterization of very low residual errors.

The following observation can be noted:

- Main limitation of these validation activities concerns the short spatial scales that are not retrieved by the GIM (Global Ionosphere Maps) global model. They can be partially validated using inter-satellites comparisons, but the final uncertainty remains high due to assumptions used to identify matchups (time lag tolerated at crossovers) and it is often not possible to identify the different error contributors (natural variability during the time delta, errors from the two sensors are mixed and potentially partially correlated).
- The wet troposphere correction is computed either from instrumental measurements (from onboard MWR) or by models (ECMWF, NCEP…). Wet troposphere path-delay estimations from in-situ measurements (GNSS and radiosondes) are mainly available in coastal areas and over continents. This in-situ network is not sufficient to perform a complete validation of the correction. Comparisons between instrumental and in-situ solutions are not trivial in these regions because the accuracy of wet troposphere path-delay measured by MWR decreases significantly when the MWR footprint is contaminated by land surface. These on-ground reference measurements are mainly used to validate innovative algorithms dedicated to coastal areas or to corroborate validation results obtained from comparisons with models. As for other variables, the comparison between MWR and modelled corrections is useful and commonly used for validating the large temporal and spatial scales. Short spatial scales (O(50km)) can be partially validated using inter-satellites comparisons but the final uncertainty remains high due to assumptions used to identify matchups (time lag tolerated at crossovers) and it is often not possible to identify the different error contributors (natural variability during the time delta, errors from the two sensors are mixed and potentially partially correlated).
- Sea State Bias (SSB) correction is an empirical solution based on measured wind and wave parameters. It corrects from the electromagnetic bias induced by the wave crest and troughs distribution “seen” by the radar within its footprint. This is the correction the most complicated to validate as there is no “true” reference to compare with. Its validation relies on comparisons between different solutions and the quantification of the improvements brought on the SSH calculation. These improvements (from one solution to another) are measured using inter-satellite comparisons, mono-mission crossovers, but the uncertainties induced by these methods remain high and limit the correction validation.
- The Mean Sea Surface (MSS) is a reference surface from which Sea Surface Height Anomalies (SSHA) are measured. It is a stationary surface elevation induced by the gravity effect of bathymetry on the water column. It is computed from a large part of the altimetry constellation to maximize the spatial

sampling. Several models are available, computed from different groups (CNES/CLS, SCRIPPS, DTU…). Their inter-comparison is good validation mean to get an overall relative uncertainty estimation but it does not provide a clear uncertainty estimation for each individual solution. Inter-satellite validation can be used for validation purposes, but it requires independent satellites measurements and it is limited to observations that are spatially very closed (o(10 km)). The method developed by Pujol et al. (2018) provides a spectral estimation of the global MSS residual error for a given satellite ground track.

3.3.3.3 Sea Surface Height Validation

Table 15 : product validation status for Sea Surface Height

SSHA (Sea Surface Height Anomaly)		Ref. Data Representativeness				Ref. Data quality				Validation Method				Validation results				Overall
spatial scales	temporal scales	In-situ	Inter-sat	Models	Alternative processing	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
o(1km)	o(1s)		Excellent		Ideal		Good		Good		Excellent		Excellent		Ideal		Ideal	Excellent
o(10km) to o(100km)	o(10s)	Poor	Poor		Ideal		Good		Good		Good		Good		Poor		Poor	Poor
o(100km) to o(10000 km)	o(days)	Poor	Excellent		Ideal	Poor	Good		Good	Excellent	Ideal		Ideal	Good	Excellent		Excellent	Good
o(1000 km) to o(10000 km)	o(months) to o(years)	Good	Ideal		Ideal	Good	Excellent		Good	Ideal	Ideal		Ideal	Excellent	Excellent		Excellent	Excellent

For the SSHA variable, several categories of temporal and spatial scales have been distinguished. It is necessary as these different components of the variable are used by different user communities to address various usages (operational oceanography, biogeochemical processes and air-sea interactions, climate monitoring…). Most important aspects to note are the following:

The very short spatial & temporal scales (o(1km) and o(1s)) of the SSHA mainly contain random error related to the instrument and processing noises. It is easily characterized and compared with respect to other missions, alternative datasets.

The short spatial & temporal scales (from tens to hundred of kilometers) are known for containing a mix of topography signal and correlated errors (Dibarboure et al., 2014). Because of the lack of truth (from on ground measures or from space), their validation is very challenging. Most of the studies performed, have established partial correlations between a part of the SSHA signal at these wavelengths and different phenomena (surface roughness heterogeneity, rain cells contamination, SWH properties…). However, errors and “true” topography signal have never been clearly disentangle. The lack of reference measurements is a clear gap for validation of these SSHA scales.

Medium to long spatial wavelengths over time scales from day to weeks are well characterized. Although the comparison with in-situ measurements provide limited results (mainly concerns coastal areas, in-situ measurements uncertainties remain relatively high over such periods), comparisons between the different altimetry missions with various processing alternatives show a general excellent agreement. These inter-satellites and multi-processing comparisons allow for the characterization of small mission-specific residual errors. The specific case of tandem phases, by maximizing the number of co-located observations and reducing the level of uncertainties specific to ocean and atmosphere variability, improves the method accuracy and thus the errors characterization.

SSHA climate scales (basins spatial scales and long temporal scales) validation is well advanced. Validation methods integrate uncertainty budgets (Ablain et al., 2019) to detect potential SSHA drifts. Some improvements can be brought on the refinement of input error budgets.

3.3.3.4 Inland waters

Table 16 : Product validation status for inland waters

WSH (Water Surface Height)	Ref. Data Representativeness				Ref. Data quality				Validation Method				Validation results				Overall
	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
WSH (Lakes and large rivers)	Good	Good	Poor	Good	Excellent	Excellent	Poor	Excellent	Good	Good	Poor	Good	Excellent	Excellent	Poor	Excellent	Good
WSH (Small lakes and rivers)	Poor	Poor	Poor	Poor	Excellent	Good	Poor	Good	Good	Good	Poor	Good	Poor	Poor	Poor	Poor	Poor

We separated the water surface height variable in two groups by considering the water body size to highlight the differences in the validation approaches for these different types of targets. The global score summarises the better validation status for the larger targets (large lakes and the largest rivers) which is mainly driven by a better representativeness and data quality of the reference data sets that can be used compared to the smaller water bodies. Moreover, smaller rivers are likely to present larger slopes, making the altimetric water level estimates comparisons to nearby in situ gauges more sensitive to biases induced by the height (slope times distance) in between the virtual station and in situ measurement. Smaller water bodies also have a lower contribution in terms of the backscattered signal compared to surrounding echogenic targets (other water bodies, sand benches, anthropic reflective surfaces), making their contribution more difficult to disentangle in the altimetric waveforms, results are much more 'environmental' dependant for small water targets, leading to lower representativeness.

Despite this difference, there is an overall important gap in the reference data availability (both in their representativeness and quality) that also prevents from reaching high scores in the validation results. In addition, there are very few information available on uncertainties both for satellite water level height products and reference data. This global picture should be revisited after S3TART project to measure where improvements have been realised.

Regarding the methods, some improvements have been done in this field by a better assessment of the impact of the slope in the comparison exercise for the rivers but the level of maturity of these methods is still far from the definition associated to the different scores.

The matrix also shows that there is no model accurate enough to rely on for the validation of satellite products.

3.3.3.5 Land ice

Table 17 : Product validation status for land ice

	Ref. Data Representativeness				Ref. Data quality				Validation Method				Validation results				Overall
	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
Ice sheet Elevation	Poor	Good			Good	Good			Poor	Good			Poor	Good			Good

Over the ice sheet/cap surface there are two main geophysical effects that generate errors in the surface elevations derived from altimetry measurements:

- Firstly, the elevation is not estimated at nadir, but at the Point Of Closest Approach (POCA). The measurements are therefore relocated from nadir to POCA using auxiliary DEM and dedicated algorithms. Surface topography variations inside the footprint also impacts the shape of the radar waveform. The altimetry performances therefore depend on the complexity of the surface topography sampled by the altimeter, in terms of steepness and roughness at different scales.
- Secondly, another source of uncertainty is related to the interaction between the Ku band radar wave signal with the snow medium. As the emitted radar wave penetrates the snowpack, the induced volume scattering disrupts the shape of the radar waveforms.

For now, there are no physical models able to account for these two effects at retracking stage. Subsequently, the uncertainties cannot be estimated at a single, or several specific locations, as they greatly vary in space with the topography configuration and snowpack parameters (density, grain size, stratification...)

Hence, in-situ data (airborne acquisitions, GNSS measurements, UAV data) are in general too sparse for providing representative validation data considering the vast areas of the polar ice sheets/caps. In particular, for assessing and discriminating the effects of each geophysical parameters that act on the altimetry performances.

Thanks to a dense sampling and near global coverage, ICESat and in particular ICESat-2 missions from NASA proved to be a good compromise between data quality and representativeness for assessing the altimetry performances (considered as “Inter-sat” in the table above)

3.3.3.6 Sea ice

For this surface, we have distinguished two different variables:

- SSHA over polar oceans which is related to marine applications and products
- Floating Ice Topography which is related to freeboard

Table 18 : Product validation status for sea ice

	Ref. Data Representativeness				Ref. Data quality				Validation Method				Validation results				Overall
	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	In-situ	Inter-sat	Models	Alternative process	global score
Polar SSHA	Good	Good		Good	Poor	Good		Good	Good	Good		Good	Good	Good		Good	Good
Freeboard	Poor	Good		Excellent	Poor	Good		Good	Good	Good		Good	Good	Good		Good	Poor

Polar SSHA validation is conceptually similar to open ocean SSHA validation, with some extra challenges linked to:

- The discontinuous nature of the measurements, originating from leads (cracks exposing open water) in the ice pack,
- The scarcity of in-situ measurements which are difficult to maintain over long time periods due to harsh conditions and difficult access for maintenance crews. This situation is especially obvious for the Southern Ocean.

For the freeboard, the main limitation comes from the lack of in-situ data, both related to their sampling, quality and uncertainty assessment. This global picture should be revisited after S3TART project to measure where improvements have been realised.

3.3.4 Quality Management

Table 19: Altimetry mission quality management status

	QC and anomaly management	Reporting	Feedback to ref. and aux data providers	Feedback from data users	Products and algorithms evolutions	Inter-operability
Copernicus Altimetry missions	Ideal	Ideal	Excellent	Ideal	Excellent	Excellent

Activities related to quality management are very mature. Regarding the products and algorithms evolutions, although the evolution cycle is well managed, the time delay between the endorsement (by QWG for example) of a proposed evolution and its operational implementation can be very long, especially for meeting the needs expressed by the Copernicus Services (the Marine component).

3.4 SAR and microwave imaging component

In this section, only the SAR Calibration and Validation status is assessed, as for now there is no imaging radiometer operated as part of the Copernicus program (CIMR is under setup)

3.4.1 Calibration and characterization

Table 20 : SAR mission characterization and calibration status

	Pre-flight characterization	In-flight characterization	Measurement Calibration	Geometric calibration	Overall	Comment / Weak link
S1 platform	Good	Good	Ideal	Good	Excellent	Performance of attitude knowledge restitution is not sufficient for SAR Level 2 RVL component, probably due to improper definition of error budget during the pre-flight definition of the products.
S1 SAR	Good	Good	Excellent	Excellent	Excellent	Pre-flight characterization of the antenna is partial and does not reflect fully the overall performance while in flight. Alternate in flight method had to be considered

The pre-flight specification and characterisation are identified as the weak point of calibration and characterisation.

First, the required performances of attitude knowledge restitution were probably not accurately performed for all product types, including the Level 2 RVL (Radial Velocity) product. For this specific product, an accurate knowledge of attitude is required to compensate for Doppler shifts induced by variations of attitude of the platform. The method was demonstrated during the era of Envisat/ASAR, while the accuracy of the attitude restitution was high, most probably due to stringent requirements of other instrument on board. The proper budget of error of Doppler shift due to uncertainty in attitude knowledge was probably not performed accurately at the time of the product definition, and then not reflected in the definition of the performances of the platform. The Level 2 RVL products are the only one impacted by this inaccuracy, and alternative way to retrieve accurate knowledge of attitude are under investigation.

Second, the pre-flight characterization of the antenna pattern on ground was partial. Each element of the antenna was characterized individually, and an Antenna Model was derived and validated on ground pre-flight. However, the validation activation performed after launch, demonstrated inaccuracy of this antenna model. While still useful, the characterization of the antenna pattern using the Antenna Model is then complemented by in flight measurements over calibration areas. Specific SAR acquisitions are performed over the Amazon Rain Forest to retro estimate the elevation antenna pattern of each beam of the TOPS acquisition modes.

3.4.2 S1 product quality status

We refer here to the overall information on product quality as accessible to the product users.

Table 21 : Sentinel-1 quality information status

	Prognostic uncertainty method	Prognostic uncertainty sources	Prognostic uncertainty values	Quality flags	Overall	Comment / Weak link
SAR IW Level 1	Excellent	Excellent	Excellent	Poor	Excellent	Product by product quality information Radiometric performance of DH acquisitions
SAR EW Level 1	Excellent	Good	Excellent	Poor	Good	Product by product quality information
SAR SM Level 1	Good	Poor	Good	Poor	Good	Product by product quality information Limitations due to acquisition plan over calibration sites

SAR WV Level 1	Good	Good	Good	Poor	Good	Product by product quality information
SAR IW Level 2	Excellent	Excellent	Excellent	Good	Excellent	Product by product quality information
SAR EW Level 2	Excellent	Excellent	Excellent	Good	Excellent	Product by product quality information
SAR SM Level 2	Poor	Poor	Poor	Poor	Poor	Product by product quality information Limitations due to acquisition plan over ocean
SAR WV Level 2	Excellent	Excellent	Excellent	Good	Excellent	Product by product quality information

For the overall set of products, there is only a limited set of quality diagnostics provided for each product. Those product-by-product diagnostics are of two types:

Basic quality checks on Level 0 data, only made available for Level 1 products (missing lines, corrupted data)

Basic quality checks on Level 1 or Level 2 data, provided as an add-on to the product as a PDF document. There is no advanced quality layer part of the product structure. The PDF cannot be parsed by end-user by automatic methods.

This product-by-product quality information is complemented by statistical analysis of product performance provided in Annual Performance Report, and containing many details on the product quality (for all modes and the various quality aspects o consider).

Considering the product by processing level and acquisition modes, some specificities must be pointed:

For Level 1, the radiometric calibration is ensured through analysis of products acquired over Corner Reflectors (CR) and Transponders, being FRM. The nominal acquisition plan of Sentinel-1 prevents the regular acquisition of products for all modes beyond the Commissioning Phase. Only the IW Dual Polarisation V products are regularly acquired over the Transponders in Europe. The radiometric calibration of other acquisition modes is either performed considering only CR out of Europe, or using geophysical method over ocean (Wave mode) or indirectly performed by extrapolation of the commissioning phase results (Stripmap).

For Level 2, the acquisition plan of stripmap products (only for emergencies, or over volcanic isolated islands) prevents performing in depth validation of the products.

3.4.3 S1 products validation status

Table 22 : Sentinel-1 Validation Status

	Ref. Data Representativeness	Ref. Data Quality	Validation Method	Validation Results	Overall	Comment / Weak link
L1 Basic image validation (IW mode)	Excellent	Excellent	Excellent	Excellent / Good	Excellent	Most of the validation data is available for IW mode only
L1 Basic image validation (other modes)	Good	Good	Excellent	Good	Good	There is less reference data available for other acquisition modes
L1 radiometry (IW mode)	Excellent/Good	Excellent	Excellent	Excellent/Good	Excellent	Availability of data acquisition over FRM only granted for dual polarisation V acquisition. Availability of transponders in dual pol H to be improved
L1 radiometry (other modes)	Excellent/Good	Good	Excellent	Excellent	Excellent	Few acquisitions over Corner Reflectors available For WV an alternative validation methodology is put in place using geophysical method.

L1 geometry (IW and EW modes)	Excellent	Excellent	Excellent	Excellent	Excellent	
L1 geometry radiometry (other modes)	Poor	Excellent	Excellent	Good	Good	Few acquisitions over Corner Reflectors
L2 Wind / OWI	Excellent	Excellent	Excellent	Excellent	Excellent	
L2 Swell (WV mode)	Excellent	Excellent	Excellent	Excellent	Excellent	
L2 Swell (SM mode)	Poor	Poor	Not applied	Not validated	Not validated	Validation of L2 swell on SM mode is not possible due to the nominal acquisition plan of the S1 mission
L2 Radial Velocity	Excellent	Excellent	Excellent	Bad	Not validated	Performance of L2 Radial Velocity does not meet the expectation due to inaccurate attitude restitution

3.4.4 S1 quality management

Table 23: SAR quality management

	QC and anomaly mgmt.	Reporting	Feedback to ref. and aux data providers	Feedback from data users	Products and algorithms evolutions	Reprocessing Strategy	Inter-operability	Comment / Weak link

SAR	Good	Good	Poor	Good	Good	Bad	Good	Link between Quality Disclaimer and product dissemination is not addressed. Organisation of Quality Working Groups. No reprocessing strategy

The QC methodology applied for Sentinel-1 allows to efficiently spot both the basic QC issues and more advanced anomalies. Series of quality disclaimers are issued to inform the users about quality degradation. However, the users are probably not largely aware of the availabilities of the quality disclaimers. A link between the product dissemination service and the quality disclaimer should benefit to the user community (providing list of quality disclaimers impacting one product of user interest).

Reporting on quality of the products is performed through the publication of annual performance report and 2-Cyclic (24 days) performance reports. The publication of 2-cyclic reports is stopped since early 2022 due to technical issues, even if the corresponding monitoring is nominally performed and do not shows product performance degradation.

Feedback to reference and auxiliary data providers is in place. However, the validation of those reference / auxiliary data is under the responsibility of the providers and only basic QC checks are performed through the Sentinel-1 service.

The feedback from users is collected through various means including reception of questions from Sentinel user support, participation to Snap user forum, collection of comments during large EO conferences (Living Planet, Fringe, SeaSAR, IGARSS, EUSAR…). However, in the last years no QWG were organised.

The processing baseline (configuration of the processor, version of the processor…) is regularly updated including both improvement of overall calibration and addition of new measurements. This applies both to Level 1 (SAR images) and Level 2 products (geophysical measurements over ocean). The evolution of products take into account feedback from user community.

There is no reprocessing strategy in place, which is a major complaint from users as expressed in multiple occasions (Living Planet symposium, SeaSAR workshop). The products are nominally processed with the latest processing baseline. Reprocessings are only performed case by case to circumvent to major production issues. However, there is no reprocessing of old data with a consistent processing baseline that could allow users to access long time series of products with similar performances.

No blocking points of interoperability were expressed by users up to now, even if they are pushing for the availability of Analysis Ready Data, that could make the data more accessible to non-advanced users.

3.5 Atmospheric Composition component

In this section, the assessment focuses on Sentinel-5 Precursor TROPOMI, i.e., the only currently operational atmospheric composition Sentinel mission. As the validation work in the Atmospheric Mission Performance Cluster (ATM-MPC) builds substantially on heritage missions (current and historical) from ESA and EUMETSAT, and as it is a stepping stone for the validation frameworks for the upcoming atmospheric Sentinels 4, 5 and CO2M, this assessment has in fact a much wider applicability.

3.5.1 S5P calibration and characterization

For atmospheric missions, calibration is only performed at L1. Bias correction schemes at L2 (e.g., for CO) are not considered to be a calibration.

	Pre-flight characterization	In-flight characterization	Measurement Calibration	Geometric calibration	Overall	Comment / Weak link
S5P L1(b)	Excellent	Excellent	Good	Excellent	Excellent	Ground-based calibration using e.g. in-situ measurements (e.g., RadCalNet and HYPERNETS) and satellite-satellite intercomparisons on natural targets such as Pseudo-Invariant Calibration Sites are usually not part of the official procedures. Current attempts for S5P L1B evaluation promising but needs further R&D.

For S5P-TROPOMI, ground-based calibration methods have been used to limited extent. This is rooted in the focus on in-flight instrument calibration, and scepticism expressed on the possibility to improve calibration key data from on-going ground-based cal/val activities. The main ‘gap’ here could be described as a lack of desire or confidence in using natural targets, intercomparisons with other satellite measurements, and inferences based on level-2 analysis as means to update calibration data.

3.5.2 S5P products quality information status

	Auxiliary data quality	Uncertainties: method	Uncertainties: sources	Uncertainties: values	Quality flags	Comment / Weak link
S5P, entire official L2 portfolio	Excellent	Excellent	Excellent	Excellent	Ideal	Each data product comes with an extensive set of per-pixel quality flags allowing detailed filtering. Uncertainties are also provided on a per-pixel level, derived following the GUM, but some sources are missing, either due to missing uncertainties on underlying aux. data and/or due to the complexity of uncertainty propagation.

3.5.3 S5P products validation status

The product validation status requires differentiation per L2 product as reference data availability and methodological challenges are highly product specific.

S5P product	Ref. Data Representativeness	Ref. Data Quality	Validation Method	Validation Results	Comment / Weak link
Ozone (O3) total column	Excellent	Excellent	Excellent	Excellent	Most mature validation activities in the atmospheric domain. Some improvements still needed in reference data quality, inter-network homogeneity,

					and operational uptake of advanced methods (for completeness of uncertainty budget assessments).
Ozone (O3) tropospheric column	Good	Good	Good	Good	Reference network (O3 sondes and tropospheric lidars) is sparsely distributed with often large data latency and occasional data quality issues. Comparison methods need R&D on harmonization of vertical sensitivity and on uncertainty assessment/propagation.
Ozone (O3) vertical profile	Good	Good	Excellent	Excellent	Reference measurement networks (O3 sondes, lidars, FTIR, MWR) are sparsely distributed geographically, with often significant data latency and occasional data quality issues. Vertical representativeness depends on the vertical resolution of the reference measurements, from very good for ozonesonde (±150 m) to poor for microwave radiometers (8-15 km).
Nitrogen dioxide (NO2) (sub-)columns	Good	Good	Good	Good	Reference data availability and quality is good for stratospheric NO2 (NDACC ZSL-DOAS) and becoming excellent for total NO2 column (PGN) but needs further work for tropospheric NO2 column (MAX-DOAS with profiling capabilities and airborne imaging data). Methodological/instrumental advances are needed to deal with uncertainties introduced by the strong spatiotemporal gradients near pollution hot spots.

Formaldehyde (HCHO) column	Good	Good	Good	Good	Reference networks (FTIR and MAX-DOAS) are sparsely distributed geographically (PGN HCHO is being released and will improve coverage). Uncertainty budget of the comparisons needs further R&D, in particular because of the large noise in the S5P HCHO retrievals.
Glyoxal (CHOCHO) column	Poor	Poor	Poor	Good	Limited heritage from previous missions and consequently needing work on all levels, including instrumentation for acquisition of reference measurements.
Sulphur dioxide (SO2) column	Poor	Poor	Poor	Good	Very little reference data available (a few MAX-DOAS instruments, mainly in low SO2 areas), but situation potentially improving with PGN SO2 measuring capabilities. Work needed on all aspects.
Carbon monoxide (CO) column	Good	Good	Excellent	Excellent	FTIR reference networks (TCCON, NDACC) relatively wide and expanding (COCCON) but large geographical gaps and some calibration issues remain; data quality and timeliness for some still to be improved. Methodology well advanced for current Sentinels but requires further advances for CO2M, adoption by wider community and operationalisation.
Methane (CH4) column	Good	Good	Excellent	Excellent	FTIR reference networks (TCCON, NDACC) relatively wide and expanding (COCCON) but large geographical gaps and some calibration issues

					remain; data quality and timeliness for some still to be improved. Methodology well advanced for current Sentinels but requires further advances for CO2M, adoption by wider community and operationalisation.
Carbon dioxide (CO2) column (CO2M only)	Good	Good	Excellent	NA	FTIR reference network (TCCON) relatively wide and expanding (COCCON) but large geographical gaps remain; data quality and timeliness for some still to be improved. Advanced methods for the upcoming CO2M to be developed, and adoption by wider community and operationalisation to be supported.
Water vapor (H2O) column	Good	Good	Excellent	Excellent	GRUAN radiosonde network of excellent quality, needs geographical expansion. Global PTU radiosonde network better distributed but of unequal quality among stations. FPH sondes excellent but launched too rarely and at only a few stations. AERONET sun photometer network is wide and well calibrated. FTIR reference networks (TCCON, NDACC) relatively wide and expanding (COCCON) but large geographical gaps remain; data quality and timeliness for some still to be improved.
Cloud properties	Poor	Good	Poor	Poor	CLOUDNET focus on Europe. Inclusion of ARM and Asian data (although with different instrumentation) highly wished but requires additional funding for integration and harmonization. Significant R&D still

					needed to make satellite and ground-based truly inter-comparable.
Aerosol optical depth (AOD) and other properties	Excellent (AOD), Good (vertical profile), Basic (other properties)	Excellent/Good (AOD/layer height)	Excellent/Good (AOD/layer height)	NA	Sentinel-5P does not provide an AOD product but an auxiliary Absorbing Aerosol Index (AAI) and an Aerosol Layer Height (ALH). While AAI is difficult to relate to ground-based reference measurements, validation based on EARLINET lidars is being developed for ALH.
Lambertian Equivalent Reflectivity (LER)	-	-	-	-	Not yet operationally validated. Work on all aspects needed. Heritage from optical missions can be applied.
Surface albedo	-	-	-	-	Not yet operationally validated. Work on all aspects needed. Heritage from optical missions can be applied.
Solar-induced fluorescence (SIF)	-	-	-	-	Not yet validated (SIF is not produced operationally for the current Sentinel-5P). For CO2M, development work on all aspects will be needed.

Reference data representativeness

Reference data availability and representativeness varies tremendously between atmospheric L2 products. Building on a very long heritage, it is close to ideal for total ozone validation, with large networks (1) distributed over most areas of the globe and (2) sampling adequately a wide range of the measurement influence quantities, and some of those providing also the data just a few days after measuring. Horizontal representativeness of total ozone validation is also well documented and the error budget of such data comparisons is well understood. At the other extreme, such as for SO2, aerosol properties and glyoxal, there are only a handful of ground-based or airborne reference instruments, and those still need major development work to produce datasets meeting FRM-level criteria. Spatiotemporal representativeness for several of these products is far from being fully documented, both at the level of the individual instruments and at the network level. Comparability of the satellite and reference measurements is for several products an issue (e.g., satellite and ground-based instruments have different perception of the cloud properties, or different spatial representativeness), requiring extensive harmonization operations which introduce significant additional uncertainty. Timely availability of the reference data would also require further elaboration of (and hence support to) the validation data distribution services. Also the spatial distribution of the reference measurements needs improvement in order to cover different latitudes, albedo types, pollution levels, etc. To give two examples: 1) TCCON sites are now mostly focused on background sites in the Northern hemisphere, and 2) airborne campaigns mostly take place in Europe/US during the summer period for practical reasons.

Reference data quality

Reference data reach (almost) FRM status for some products (e.g., for total ozone columns), but much development is needed for others, such as SO2, glyoxal, cloud and aerosol properties (significant errors sources in the retrievals, limited uncertainty characterization, far from operational and timely data provisions). FRM4*** and similar programs have been very valuable but need continuation. Involvement of the ground-based data providers in the validation work is observed to be a strong driver for improvements in the ground-based data quality. For several products, various types of reference instruments are available but these often don’t yield fully consistent validation results. Intercomparison projects are necessary to find the causes of these discrepancies (either on a campaign basis or at super sites) and funding is required to support the corresponding improvements in data quality.

Validation methodology

- For the vertically resolved species (O_3 profile and tropospheric column, and to some extent also the NO_2 subcolumn products), vertical harmonization between satellite and reference measurements remains a challenge: how (and to what extent) can and should different assumed (prior) profiles, different vertical sensitivity, different definitions of the tropopause, … be taken into account? This requires both research and - based on those outcomes - the definition of validation protocols including best practices for data harmonization and comparison.

- For several species, the maturity is overall very low. This is either due to a lack of heritage, e.g., as a consequence of the unavailability of reference measurements (glyoxal and SO2), or due to the very different nature of the satellite and ground-based products (cloud properties). For glyoxal and SO2, support for methodological development should follow that for the procurement of reference measurements. For cloud properties, dedicated research is required to assess and improve intercomparability. Airborne imaging data, providing valuable reference data sets for tropospheric species, requires consolidation and standardisation at instrument/algorithm level.

- For some species, no systematic validation is performed (yet) in the context of the ATM-MPC, SAFs, or C3S/CAMS: L1b (planned with the ATM-MPC for S5P), LER, surface albedo, aerosol properties (besides AOD), and SIF. Consequently, the methodology needs to be defined from the bottom up.

3.5.4 S5P quality management

The quality management for S5P is performed homogeneously for all official products (in particular in the operational work in the ATM-MPC), hence no differentiation is required.

	QC and anomaly mgmt.	Reporting	Feedback to ref. and aux data providers	Feedback from data users	Products and algorithms evolutions	Reprocessing Strategy	Inter-operability	Comment / Weak link
S5P, entire official L2 portfolio	Ideal	Ideal	Good	Ideal	Ideal	Excellent	Good	Because of clear benefits, interaction with reference data providers deserves more (sustained) support and formalization. Interoperability would benefit greatly from more structured satellite-satellite intercomparison studies, both at level 2 and at level 1b.

4 A Copernicus Cal/Val Solution

4.1 Overall calibration and validation strategy

A comprehensive cal/val strategy for the Sentinel missions should cover all these activities, ensuring a long-term support for collection of reference data, state-of-the-art processing (cal/val activities) and publication of results. It is useful to recall the different activities involved and their specific needs in terms of reference data and processing.

4.1.1 Calibration of the sensor outputs and characterization of influence parameters

The calibration strategy shall ensure that all parameters of the measurement equation are known with an uncertainty compatible with the overall mission parameters, either from pre-flight characterization or through in-flight calibration with a temporal sampling compatible with the temporal stability of the parameters to be measured.

Calibration can rely from a) pre-flight calibration b) in-flight calibration using on-board devices c) calibration using ground reference sources d) inter-calibration using a reference sensor.

The calibration sources and characterization measurements need be of the highest possible quality and ideally SI-traceable, i.e. Fiducial Reference Measurements in case of ground-based and airborne measurements.

In addition to instrument calibration activities, continuous instrument monitoring is needed throughout the lifetime to better understand its characteristics and detect any anomaly or degradation. This monitoring can also detect the need for special maintenance operations such as decontamination.

4.1.2 Data validation and performance assessment

The objective is to establish and associate with every data product the set of quality indicators enabling all users to readily evaluate the fitness for purpose of the data product in a reliable manner. In particular, the main product performance criteria shall be assessed using independent reference data and compared to requirements. The validation strategy shall be representative of the actual usage of the products, and therefore cover the full range of observation conditions and influence parameters.

The validation activity may rely on reference data with a lower quality than for the previous one (i.e. non-FRM data). Obviously, the level of confidence established by the validation depends on the quality of the reference data which should have an uncertainty lower than the product under test. A common challenge is to collect and harmonize data from heterogeneous sources. The main objective is to cover the whole range of applications of the data products with a sufficient temporal sampling.

Quality assessment is the only way to build confidence in the data products and foster their use for science and society. Moreover, an accurate assessment of the measurement uncertainty can be critical for some downstream applications such as data assimilation or climate change studies. It is also a prerequisite for inter-operability of Sentinel missions among each other and with other non-Copernicus missions.

4.1.3 Uncertainty characterization

In addition to the evaluation of the final uncertainty, another objective of the validation activity is to assess the different uncertainty sources affecting the data product (e.g., parameters used in the retrieval algorithms, instrumental degradation and drifts, orbital effects, ...) and evaluate their impact on the final uncertainty of the product. The aim is to ensure that the measurement uncertainty is correctly understood and/or modelled.

This activity can lead to product improvements and can be useful to support investigations of potential data quality anomalies. It does not necessarily require high quality data, provided that the uncertainty of the reference measurement is not correlated to the influence parameter under study.

In particular, the following activities are particularly useful to better understand uncertainties:

- Intercomparison and benchmarking exercises for data processing algorithms (for satellite and ground-based measurements)
- “Round-robin” field campaigns
- inter-satellite comparisons (including analysis of tandem data) and comparisons with models
- analysis of long-term time series, Level-3 products, comparison with climatology, and statistical analyses

These activities can often be performed in a generic multi-mission framework.

4.2 Optical Component

4.2.1 Improve on-board calibration

Reflecting Sun diffusers are used by several Sentinel missions for absolute radiometric calibration. While the devices are clearly useful to ensure a stable (in time) and uniform (in the field) radiometric response, they show some limitations regarding absolute uncertainty. Sensor inter-comparison methods have highlighted differences of several percent between similar sensors (e.g. OLCI-A vs. OLCI-B). To address these issues, it is recommended:

- To improve diffuser ground characterization methods and generalize cross-characterization efforts.
- Ensure a repeatable and accurate alignment of the diffuser with the instrument and the AOCS reference frame
- To consider alternative calibration strategies (e.g. transmission diffusers, on-board reference sources)
- Implement in-flight characterization of Sun diffuser angular dependence during the commissioning phase

On-board spectral calibration or characterization for optical sensors is considered a weak point for current missions. More specifically:

- Sentinel-2 MSI has no spectral characterization/calibration device. This is a difficulty for filter-based broad-band sensors.
- Tunable laser diodes are limited to short wavelengths (UV-blue)
- Doped integrating spheres do not provide accurate references in the SWIR domain

The CCVS project recommends a continued R&D effort to improve on-board calibration devices for optical imagers.

4.2.2 Improve traceability of instrument characterization

The CCVS project recommends improving the traceability of the instrument pre-flight characterization. More precisely, the instrument main characteristics shall be provided together with their **measurement uncertainties and evidence of SI traceability**:

- PSF
- spectral response (including assessment of its in-the-field variations)
- linearity up to saturation or Lmax + 10 % (whichever is lower)
- straylight characterization
- polarization sensitivity.

4.2.3 An operational, hyperspectral, surface reflectance measurement network

Surface reflectance validation is a critical activity for optical missions. While directional surface measurements are generally not considered as scientifically relevant in themselves, an accurate assessment of the uncertainty of the satellite measurement is needed to ensure the validity of downstream processing of geophysical variables. For this, there is a need for:

- Ground based radiometric measurements associated with a well-characterized BRDF from ground-based angular measurements
- The measurement site shall have homogeneous surface and atmospheric properties at the scale representative of the satellite measurement (typically over an area covering 5 x 5 spatial sampling distances), and the instrument shall be installed on a tower with adequate height
- Concurrent assessment of atmospheric properties (in particular aerosol optical depth)
- Measurements shall be acquired within 20 minutes of the satellite overpass, which can be ensured if automatic measurements are performed at high frequency

- Hyperspectral measurements covering the UV-to-SWIR range are needed (cf. D3.1 and D3.3 for detailed specifications).

The measurements should be processed to provide validation products taking into account the satellite viewing direction, solar direction and spectral response function.

There is a need for an operational network of such radiometer measurements covering:

- Bare soils with a range of different surface albedo values (from snow/ice to dark surfaces)
- Vegetated sites with a range of different surface albedo values (decid. /conif. forests, agricultural fields, …)
- Coastal and in-land waters

The raw measurements should be further processed to generate SDR validation products:

- computation of reflectance for each sensor channel
- correction of directional effects using a BRDF model
- conversion to TOA reflectance using an RTM for L1 validation.

This processing is sensor specific. It may be performed by the measurement network itself or another entity. In any case the processing should be uniform across the network, documented and traceable.

Currently, this ideal goal is not achieved:

- The RadCalNet network does not meet the spectral sampling requirements. Moreover, only nadir viewing measurements are provided, while directional surface reflectance validation products are needed. The representativity in terms of surfaces is not sufficient (mostly desertic sites with high reflectance).
- The Aeronet-OC network does not meet all spectral requirements. Moreover, its long-term sustainability is not ensured.
- The candidate Hypernets network could potentially meet the requirements but is not yet operational. There is also a lack of overage for the current sensor in the SWIR domain.
- Other technical solutions (e.g. using hyperspectral instruments mounted on UAVs) are currently explored but have not yet reached an operational maturity. Airborne campaigns can bring very useful data, especially for the development and commissioning phases of the missions. But their cost limit their applicability for routine validation.

The project recommends:

- To maintain the operations of current networks (RadCalNet and AERONET-OC)
- To progressively develop an operational network providing directional reflectance validation products for Sentinel missions.

4.2.4 Improved models for natural scenes

The validation of optical missions relies extensively on comparison against models of stable natural targets such as Ocean sites (Rayleigh method), Pseudo-Invariant Calibration Sites, the Moon, or Deep Convective Clouds (statistically stable). Several models have been developed to predict the mean directional reflectance over the target so it can be compared with the sensor measurements. Generally speaking, **R&D efforts should be sustained to improve the models** by:

- Extending the range (especially for the SWIR domain) and spectral resolution of existing models
- Consolidating uncertainty budgets
- Performing Inter-comparison of models and identify potential sources of discrepancies

In addition, we recommend supporting the **development of advanced simulation models** for non-uniform and/or non-stable scenes using 3D radiative transfer models. In the thermal domain, a dynamic simulation could also provide a valuable reference for validation.

4.2.5 Improvement of RTMs

Radiative Transfer Models have a major impact on optical remote sensing products and validation methods. The project recommends to:

- Support the development of open-source RTMs (e.g. eRadiate initiative)
- Support inter-comparison exercises for RTM (e.g. RAMI4ATM)
- Develop commonly accepted guidelines on modelling to be used according to different needs.

This point is further elaborated in D3.5.

4.2.6 Improvement of cloud and cloud shadow masks

The reliability of cloud and cloud shadow masks is today a limiting factor for the quality of optical surface measurements. It is therefore important to keep investing in R&D efforts to develop new and improved masking algorithms, but also on validation methods. Indeed, a common agreement on performance metrics is needed to drive algorithm improvements in an efficient way. Some specific recommendations are:

- Support the CMIX inter-comparison exercise, with the objective of defining a commonly accepted validation framework and metrics
- Develop open-source reference datasets. This requires some R&D and coordination activities to define an appropriate methodology.
- Develop ground-based validation methods e.g. using hemispherical sky cameras, and eventually an operational measurement network

In addition, there is a need to investigate the impact of cloud mask quality on the metrology of surface measurements. The following questions should be addressed:

- How can we define and validate in a robust way a confidence level for the cloud/cloud shadow mask?
- How could such a confidence level be used by downstream users (e.g. for data assimilation)?

4.2.7 Improving validation methodologies for Fire Products

According to the CEOS LPV subgroup, the maturity of the validation of fire products is considered low. These parameters are set to become more and more important in the context of a warming world. There is a need to improve validation methodologies, standards and protocols for fire products through dedicated R&D activities ("FRM4FIRE" project).

4.2.8 Cal/Val methods for surface reflectance

Surface reflectance products are evolving toward Analysis Ready Data (ARD) products, i.e. providing measurements of intrinsic surface properties which are not dependent on observation conditions. This requires the correction of atmospheric effect, angular anisotropy and adjacency effects. R&D activities on retrieval and correction algorithms are needed to progress in this domain. They must be also accompanied with activities on validation methods, in order to assess rigorously the efficiency of the various correction methods. In this respect, the development of UAV-based reflectance measurements could be an opportunity to evaluate these effects. Initiatives like SRIX4VEG should be pursued and further elaborated.

4.3 Altimetry Component

4.3.1 Structure of the proposed solution

The altimetry community recognizes that the validation activities cannot rely on a single mean such as in-situ data. For years now, the CalVal teams are used to cross compare all results obtained through the comparison with other altimetry satellites, in-situ data and models. To account for this specificity, the sections dedicated to altimetry component solution are organized by sensors and variables and for each of these, the different CalVal means that could be further improved and/or set-up are addressed.

This organization of the solution also helps:

- to constitute a whole, self-consistent and fully detailed set of recommendations,
- to directly link the proposed solution and recommendations to either a specific mission requirement and/or user needs,
- to ease their future potential exploitation by stakeholders who would like to address separately one specific surface or one specific part of the uncertainty budget, etc.

4.3.2 Altimeter Sensor

This section addresses the different topics inherent to the altimeter sensor such as on-board calibration activities, long-term monitoring with and without external data, validation of Level-1 processing and innovative algorithms.

- Perform tandem phase for each Sentinel mission. Such tandem phase should not be restricted to missions with climate objectives since this kind of configurations is a powerful tool to assess the performance of sensors and products, whatever the considered spatial and temporal scales. **Criticality – high, effort needed – high**
- Calibrate the altimeter nominal and the redundant RF paths during the tandem phase (done for the first time with Sentinel-6 mission but it is essential for the climate data records). This should be systematically studied when preparing the Sentinel commissioning phase. **Criticality – high, effort needed – medium**
- Trade-offs to optimize the Altimeter CAL1/CAL2/AutoCAL number of acquisitions and durations to be made (based on return of experience from in-flight missions). **Criticality – low, effort needed – low**

Regarding the Level 1 validation and instrument's calibration (delay, backscattering coefficient), several recommendations have been identified:

- Improve transponders to calibrate both the phase and the backscattering coefficient so that they offer simultaneous calibration of altimeter range and the backscattering coefficient. **Criticality – medium, effort needed – medium**
- Develop transponder capability to validate C-band measurements, both in range and backscattering coefficient. A first experiment is conducted by JPL teams for Sentinel-6 mission on Catalina transponder. The validation of C-band for Sentinel-3 mission should be also investigated. **Criticality – medium, effort needed – high**
- Develop transponders for Ka-Band for CRISTAL mission. This is an important tool to understand the subtilities of the Delay Doppler processing in Ka-band which has never been used so far. **Criticality – medium, effort needed – high**
- Use corner reflectors to characterize quite easily both the phase and the backscattering (for SAR mode only). **Criticality – medium, effort needed – low**

For the range validation, several transponder sites are available for the purpose of CalVal activities. Similarly, some corner reflectors experiments are also emerging. Therefore, it appears reasonable to discuss the pros and cons of the different facilities to better understand and outline their complementarity in the whole CalVal Sentinel framework. Improvements can be reached in the instrument's calibration (delay, backscattering coefficient) with corner reflectors and transponders. The following items shall be addressed for the range and backscatter validation:

- Identify the outcome expected from several transponders operating for the same mission and during the same period. In other words, how many transponders do we need to calibrate Ku-band and C-band at the requested level of accuracy and which ultimate accuracy can be reached with all the transponders.

- Identify and document the uncertainty of the different transponders to describe which ultimate accuracy can be reached individually
- Assess the uncertainty between the different transponders and corner reflectors
- Assess the transponders capability to detect altimeter drifts (on range and backscatter coefficient) which means assess the facility uncertainty in stability over different considered time series length.
- Investigate the monitoring capability of altimeter range drift arising from massive deployment of corner reflectors.
- Assess the coupling of corner reflectors with other sensors to reach a reasonable individual uncertainty. Main drawback relies in the estimation of the wet troposphere delay that could be tackled by systematic coupling of corner reflector with GNSS station.

4.3.3 Radiometer Sensor

- We recommend considering the importance of the tandem phase for the radiometer:

○ This is essential to intercalibrate MWR with the highest possible accuracy (radiometric sensitivity). **Criticality – high, effort needed – medium**

○ Need to follow Sentinel-3 feedbacks (30 s gaps between both satellites/duration≥ 2 months). **Criticality – high, effort needed – medium**

- Radiometer absolute calibrations (cold sky/hot target): keep short sequences but frequent ones to limit the impact of a short-term drift of the instrumental gain (AltiKa, …, S3 feedback). **Criticality – medium, effort needed – low**

4.3.4 Orbit Validation

This section is identified on its own since this a major and critical variable for all altimetry missions. Ocean surface topography accuracy is the main driver and the more demanding variable, but this quality also benefits to the topography measurements over other surfaces.

- Yaw-flip manoeuvres for the POD (Attitude flips at low beta angles (0°, 10°) should be implemented to disentangle the centre of phase errors from time-tagging, dynamic modelling errors). **Criticality – medium, effort needed – medium**
- Build altimetry platforms with the 2 available techniques GNSS and DORIS on-board since this is the combination of the 3 techniques (GNSS and DORIS on-board and LASER on-ground) that enables to measure the accuracy of the estimated altitude and associated uncertainty. **Criticality – medium, effort needed – medium**
- Pursue the efforts of the altimetry obit validation community which is very well organised on the sharing of approaches, results and metrics. This organisation is supported by the OSTST POD group and by the Sentinel POD group set-up by ESA and these efforts should continue in the future. **Criticality – high, effort needed – medium**

Regarding the laser network issues that are used for precise orbit validation, improvements are needed:

- Laser Network that is identified as of today as the on-ground mean with the lowest uncertainty for orbit validation

- In terms of noise (precision) and biases (accuracy): today, only 5 to 8 SLR stations have a satisfactory accuracy (less than 5 mm biases and 1 cm noise) to validate the short-term orbit accuracy. **Criticality – medium, effort needed – high**
- In terms of long-term stability: laser measurements need to reach new standards of stability to address the challenges of long-term radial orbit stability at regional scales (0.1 mm/y per decade). **Criticality – high, effort needed – medium**
- Stations are needed at high latitudes, especially for polar altimeter missions above 66°: currently no SLR station above 60° N and 30° S! **Criticality – high, effort needed – high**
- This points toward the need for automatic SLR stations. **Criticality – high, effort needed – high**

4.3.5 Ocean sea state

Validation of sea state variables (wind speed and wave height) relies on comparison with other altimetry satellites, specific missions (CFOSAT, IceSAT2), models (Near Real Time and reanalysis) and in-situ data such as buoys networks.

In altimetry, the reduction of uncertainty is certainly the key question to better benefit from the existing on-ground observations. Indeed, their level of uncertainty is often too large to support the validation activities in a very efficient way and quantitative assessment.

We recommend to:

- Decrease the level of uncertainty on SWH acquired from buoys since the validation activities in altimetry need to detect very low signals, from 2 cm to 10 cm on altimeter SWH, when comparing with buoy data. This is today the order of magnitude that we observe between different altimeter processing and we have no external data accurate enough to discriminate those different approaches. **Criticality – medium, effort needed – high**
- Explore the possibility to optimise the network localisation, especially in the Southern hemisphere which coverage is too poor for validation. **Criticality – low, effort needed – high**
- Provide accurate observations of mean wave period since this is a key parameter used for altimetry sea state bias correction. Primary need of on-ground observations on this wave parameter is mainly on long-term trend and stability that would need to be characterised at a global scale. **Criticality – medium, effort needed – high**
- LIDAR is a key technology that should be further investigated for validation of SWH which use has been very limited so far. Recent results with LIDAR on-board a drone led to promising issues which shall be encouraged. **Criticality – medium, effort needed – medium**
- Develop the systematic use of wave models which do not assimilate altimeter data (Wave Watch III model for instance) to better ensure the independence of the comparison. **Criticality – high, effort needed – low**
- Maintain a strong effort on R&D activities on delay doppler processing and retracker algorithms over ocean to better understand the effect of orbital velocity on SARM SWH observations but also patterns on topography correlated with wind speed. **Criticality – high, effort needed – medium**
- Propose alternative processing and associated long time series to support the validation of operational Sentinel products. Thanks to alternative data sets, validation of Sentinel products is

reinforced, and it shortens the time needed to detect errors on SARM variables in L2 products (and even on LRM for Sentinel-6 mission). **Criticality – high, effort needed – medium**

4.3.6 Ocean topography corrections

In this section, we address the different corrections that are used to compute the ocean topography observations. We have identified different recommendations for the following items: ionosphere correction, sea state bias correction, ocean tide correction, wet tropospheric correction..

4.3.6.1 Ionosphere correction

The validation of this correction relies on the comparison with GIM model which is estimated from GPS observations. Nonetheless, the dual frequency correction derived from the altimeter is more accurate to capture short time varying signals compared to this model. Therefore, this validation mainly allows addressing a global bias on the altimetry correction and potentially some geographically correlated errors when their magnitude is larger than the GIM uncertainty.

Nevertheless, to go further in the validation activities, we recommend the following items:

- Assess the uncertainty of GIM correction at different time and spatial scales, liaising with the teams in charge of its computation. This will help identifying the relevancy of any difference spotted between the dual frequency and the GIM corrections. **Criticality – high, effort needed – medium**
- Link the Catalina transponder calibration results in C-band with the validation of the correction. The transponder should at least provide an estimate of the absolute range bias in C-band and possibly provide insights in C-ban waveforms that be further injected into retracker algorithms of C-band data. This is done for Sentinel-6 mission and could be further explored for Sentinel-3 mission. **Criticality – medium, effort needed – medium**

4.3.6.2 Sea State Bias correction

A major weakness in the validation of the topography corrections is identified for the sea state bias correction. Indeed, there are no external data that can be used for its validation. This is because the on-ground content is not fully consistent with altimeter related correction and that by nature, this correction is very much dependent on the type of sea state (swell, wind sea, mixed sea). An exhaustive validation would need to have access to on-ground observations of these different situations and therefore spread all around the globe to sample different conditions.

Recommendation is to revisit the different on-ground means that could support the validation of the sea state bias correction. These means should answer to the need of assessing the mission requirement thanks to punctual validation experiment rather than setting-up permanent means for the monitoring during the mission lifetime. **Criticality – high, effort needed – high**

The verification of the mission requirement relies either on the analysis of the short scale content of the correction (below than 7 km and directly linked to the SWH noise and/or spectral content at scales below 7 km) or on the difference with other altimetry missions that mainly leads to the identification of basin-scale correlated errors for a given mission (accuracy of the correction).

4.3.6.3 Wet tropospheric correction

In altimetry, the reduction of uncertainty is certainly the key question to better benefit from the on-ground observations. Indeed, their level of uncertainty is often too large to support the validation activities in a very efficient way and quantitative assessment. We recommend to:

- Assess the uncertainty of radio sonde observations used for radiometer wet tropospheric correction validation
- Assess the uncertainty of GPS observations used for radiometer wet tropospheric correction validation
- Revisit the validation method of wet tropospheric correction with GPS for Sentinel-6 radiometer that allows getting closer to the coast thanks to its additional channels
- For validation of the radiometer wet troposphere stability, the approach using alternative data sets such as FCDR ones should be considered, along with a precise description of the FCDR uncertainties

4.3.6.4 Ocean tide model

Tide gauge networks are used for ocean tide model validation and also to improve the quality of the tidal models through the assimilation of the in situ data. For the specific purpose of validation of the ocean tide models, we recommend to:

- Increase the number of tide gauges in the polar regions especially in the Arctic Ocean and in the Antarctic region which is very poorly sampled at present time.
- Some other specific regions are still poorly equipped by in situ sensors: Patagonian shelf, China seas, Australian coasts, Indian Ocean, Amazon shelf and coasts ...
- Other in-situ sensors could be used such as
  - Bottom pressure gauges which can be deployed in deeper ocean regions
  - Mesurements of tidal currents are also valuable as they allow validating the modelling of tidal currents and thus smaller scales structures: ADCP , HFR are both interesting

4.3.7 Ocean topography signal

For altimetry Calibration and validation activities no critical gaps or limitation are identified. Most of the methods used are mature and well established. They mainly require minimal improvements and evolutions. The main criticality for altimetry observation over the ocean is the uncertainty definition at all spatio-temporal scales (i.e. individual observations, scales ranging between 10 and 100 kilometers e.g. the mesoscale, and global climate trends).

In addition, the ground truth for the Ocean cannot be reached by in-situ only. The meso and large oceanic scales can be accessed only through a combination of in-situ, models and satellite data.

Different spatio temporal scales must be considered for the ocean topography validation activities and recommendations are separated between each of these scales.

4.3.7.1 Climate scales

As mentioned in the previous deliverables, requirements on ocean topography stability are becoming more and more stringent and therefore constitute a major piece of the validation activities. **The current status is that we have no external means to detect the stability at the level of 0.3 mm/year requested by GCOS.** Note that this value is being reviewed and could be further reduced to 0.1 mm/year over a 10 year-time series.

Hence, we recommend to further improve the different means for drift detection of topography altimetry observations by:

- Maximise the benefits from the comparison with tide gauges networks. This is further detailed in section 4.3.7.1.1). **Criticality – high, effort needed – high**
- Explore other techniques provided by in-situ means for validation of topography global stability. **Criticality – medium, effort needed – medium**
- Continue to use third party altimetry missions to detect possible drifts on Sentinel missions. This implies to rely on third party missions which provide detailed and accurate validation information to identify their possible limits in stability (ideally, the associated uncertainty on Mean Sea Level trend both at global and regional scales). **Criticality – high, effort needed – medium**
- Further refine the uncertainty description of MSL for Sentinel missions at global and regional scales, specifically the qualitative and quantitative assessment of the different sources of uncertainties. This is a critical aspect to properly assess the capability to detect any drift on altimetry topography observations. **Criticality – high, effort needed –** high
- Provide alternative solutions for Level 1 and Level 2 altimeter processing, especially for SARM missions, to allow cross comparisons and discrimination of possible errors in the official products. This effort should not be limited to access to a given software and should rather allow access to the whole Sentinel time series. **Criticality – low, effort needed – high**
- Explore alternative solutions using different sensors observations such as sea level closure budget approach. This should include a detailed uncertainty description of the different contributors. CNES initiative will provide a more thorough assessment of the uncertainties of this method that should rank its potential. **Criticality – low, effort needed – high**
- Continue efforts to maintain different validation approaches. This is crucial to reach the requested level by GCOS which is very low and cross comparisons of the results obtained by several methods are needed to reach some conclusions, of course along with their respective uncertainties to consider the appropriate balance between the results. **Criticality – high, effort needed – medium**

4.3.7.1.1 Focus on tide gauge networks

Tide gauge data are identified as of today as the on-ground mean with the lowest uncertainty to detect drift on SSH I.e. on altimetry payload (as a whole without the possibility to distinguish between altimeter, radiometer or GNSS sensor). In addition, they are also used for the validation of ocean tide models used in the ground segment for SSH computation.

Nonetheless, they suffer from weaknesses that leave the comparison results with a still too high uncertainty level to detect drifts below 0.5 mm/year over the last 25 years of altimetry era. We propose hereafter several recommendations to revisit the approach and increase the benefit that could come out from the use of these networks for altimetry stability assessment:

- Refine the tide gauges uncertainty assessment, including error sources and possible improvements for the tide gauges network.
- Revisit the uncertainty allocation of the altimetry/tide gauges comparison methods.
- Identify the limits in the altimetry/tide gauges comparison methods arising from the tide gauge uncertainty (including vertical land motion error source).
- Consider methods to better link offshore satellite tracks with coastal tide gauges. While this can work on some absolute super site with a large panel of calibration means, the main challenge is to extend these at a global scale for all the available tide gauges. We identified different possible sources of enhancement:
    - Explore the capability of Sentinel-6 payload to get closer to the coast (both for altimeter and radiometer) and thus quantify the error reduction within the overall error of the comparison method.
    - Revisit the method thanks to SWOT capability to link more precisely offshore tracks and tide gauges with a simultaneous and wide coverage topography observation.
- Elaborate technical recommendations for an optimised scenario of the network addressing:
    - Determine the ideal tide gauge network needed to meet the requirements (Ball park number on how many stations would be required, optimised location, etc.).
    - Consider different quality and associated uncertainty between the sites and optimise their combination for drift detection.
    - Investigate the feasibility of a network dedicated to altimetry validation (close enough to the ground tracks).
- Focus shall be done on Sentinel-6 mission but higher latitudes covered by Sentinel-3 might also be considered (this imposes constraints on ground tracks and time periods to be considered).
- Strengthen the link between the altimetry and tide gauge communities. Even if there has always been a good collaboration between both communities, the need for regular sharing of the respective work is present and should be further improved and set-up on a sustainable manner. We identified the EuroGoos TG task team as a good forum for this collaboration.

4.3.7.1.2 Other in-situ techniques

Considering the high level of uncertainty of the comparison method with tide gauge networks, we recommend exploring other in-situ means such as:

- GNSS interferometric reflectometry sensors.

  This should be further investigated since it could provide measurements with a content very close to the altimetry topography, with a global coverage through the already existing network. Nonetheless, there are some limitations of the technique related to strong waves situations that can limit the signal acquisition and quality. This certainly limits to situations of calm seas more often associated to coastal areas. The main advantage would be to rely on a network dense enough to provide robust statistics to decrease the overall uncertainty of the comparison exercise with possible more suitable geometric configurations regarding the distance between the in situ and satellite tracks.
- Surface drifting buoys.

The interest of this technique for monitoring MSL was quantified by Elipot (2020). The proposed system could measure global mean sea level with an accuracy comparable to altimetry and an homogeneous sampling of the oceans. Hence this approach could be of interest for drift detection in altimetry. We recommend to further push this approach by:

- Set-up the comparison method with altimetry observations to understand the pros and cons of such an approach.
- Investigate the technical feasibility on the drifters to add the measurement of altitude to the horizontal position which would require more electric power than currently available on a drifter

- Improvement of Argo Temperature and Salinity profiles.

The validation of the steric sea level component strongly relies on Argo in-situ data. We need to get deeper Argo data (< 2000 m) to improve the validation of this aspect. This could further improve sea level closure budget approach.

4.3.7.2 Short spatial scales

This section addresses the most challenging component of the ocean topography content e.g. scales between 10 and 100 km.

- The lack of ground truth to support and validate current calval activities at those scales has been identified as the largest gap for the topography mission.
- Acquisition of such an in-situ ground truth is extremely challenging because it requires synoptic observations over the 10 to 100 km range for an extended period of time (at least on the order of several tens of satellite passages).
- In the framework of the next NASA/CNES SWOT mission, such kind of in-situ network will be deployed for the first time, providing an unprecedented opportunity to advance our understanding of the content of the short scale altimetry signal and its errors.
- Even if the design (location, in-situ instrument spatial configuration etc.) is not optimized for the validation of Copernicus missions, all the work and results in the framework of the SWOT CALVAL mission (https://www.swot-adac.org) should nonetheless closely monitored to assess the feasibility of such kind of deployments for current and future Copernicus missions.
- Contact with the SWOT ADAC community should be taken to potentially used the collected in-situ observation for opportunity CALVAL analysis on current Copernicus missions. Contact shall also be taken for the next LIDAR campaigns scheduled in the frame of SWOT validation activities to explore the possibility to overfly Sentinel-3 and/or Sentinel-6 ground tracks in the same area.
- If the combination of a fast-repeating phase and intense in-situ deployments (100-km mooring array; lidar overflights; etc.) proves to be a success for the characterization of the topography signal at the 10 to 100 km scales, we recommend considering similar configurations future Copernicus altimetry missions for the best possible characterization of their signal at those scales. A SWOT-based characterization could also be a solution, albeit much less precise than one based on in-situ observations.

4.3.8 Non-ocean surfaces

This section addresses the recommendation for the validation activities over ice sheets, sea ice and in-land waters surfaces.

While ocean variables have been validated since many years and have thus very well-established methods and protocols, land surfaces represent a partial exception as the exploitation of altimetry observations over these surfaces have only recently started.

The most critical issue for in-situ data deals with in-land waters and ice surfaces coverage as no FRM data exist to verify the requirements. Nonetheless, some hydrological sites and/or networks could already be labelled and developed.

The needs and demands on in-land and ice are increasing, and the validation is even trickier than for ocean parameters. Therefore, the in-situ data have an important role to play in the CAL/VAL activities over these surfaces, pushing for more sites and networks. In addition, contrary to ocean surfaces, we cannot rely on models to validate the products over in-land and ice (or only on very small areas).

ESA has already identified this gap and has started the STR3TART project in 2021 to address these needs and establish the basis for in-situ facilities deployment for inland waters, sea ice and land ice parameters.

We recommend the following items for inland waters:

- Improve error budget/uncertainty characterisation and quantitative assessment for Sentinel missions (S3 and S6 and future CRISTAL mission) in order to better separate the difference uncertainty sources affecting water level over lakes and rivers. This is essential to 1/verify the water level requirements in a quantitative way, 2/identify the weaknesses and situations where requirements are not met, 3/understand the current limitations and 4/propose possible enhancements in the validation methods and/or in Level-2 algorithms. **Criticality – high, effort needed – medium**

- Improvement of geoid models in order to have a better representation of spatial scales. This is of importance for validation over lakes as the geoid variation (wrt ellipsoid) can vary of several tens of centimetres. Reconstruction of the local geoid over lakes is currently being tested with IceSat-2 data in the perspective of SWOT CalVal activities. The St3tart project also recommends using heights referenced to the ellipsoid (ellipsoidal height) rather than orthometric heights when possible - insitu sensor close to the satellite pass – to avoid adding geoid errors into the uncertainty budget. **Criticality –high, effort needed – high**

- Better define the constraints on the ground track drift between repeat cycles in order to limit the errors on water height correlated with the slope within the across track displacement. The precise requirement on the largest acceptable width shall be carefully assessed for future missions (e.g. CRISTAL mission and possibly S3 C/D along with S3 NG mission should nadir technology be selected). This will improve the reconstruction of the timeseries at the virtual station position, nevertheless for accuracy comparison with insitu, slope correction will always have to be accounted for – unless the sensor is right at the VS localisation. **Criticality – high, effort needed – low**

- Investigate synergies with Sentinel-6 mission regarding in-situ validation means. This effort is not considered in the frame of STR3TART project since this latter is dedicated to Sentinel-3 missions but Sentinel-6 being an important mission for the Land Copernicus Service, the same level of efforts on product validation, uncertainties assessment, etc. should be put on this mission, of course considering strong synergies between both missions. CNES already supports installation of micro stations on Sentinel-6 ground track but such effort shall be sustained in Copernicus programme. **Criticality – high, effort needed – low**

- Investigate synergies with SWOT mission regarding in-situ validation means. This effort is already considered in the frame of STR3TART project, but it should be sustained on a long-term basis through appropriate framework. **Criticality – high, effort needed – low**

- Investigate synergies with SWOT mission regarding validation methods. Indeed, the simultaneous SWOT observations of heights and slopes over an entire swath should support a complete revisit of validation approaches of the water level estimated from nadir altimetry with Sentinel missions, also including uncertainty estimates. **Criticality – high, effort needed – medium**

- Maintain a strong R&D activity on retrieval algorithms from the altimeter waveforms since this is the essential piece of the quality of water heights. Focus shall be done on the exploitation of the different techniques over the rivers to progress on understanding where the signal comes from and reduce the uncertainty on the retrievals. **Criticality – high, effort needed – low**

- Pursue efforts on altimeter tracking mode through regular OLTC enhancements (both improving accuracy of the existing targets and adding new ones) so that the water bodies are acquired by the altimeter which constitutes a prerequisite to measure water heights. This scheme already implemented in Sentinel-3 missions should be continued with CRISTAL mission. **Criticality – low, effort needed – low**

- Pursue the analysis started in ESA ST3TART project with different aspects related to on-ground means, including in-situ instrument capabilities, network densification, new sensors etc. Below are the main outcomes and recommendations from STR3TART project for inland waters at this stage of the project: **Criticality – high, effort needed – medium**

- Some basins (Europe, North America) are well monitored, whereas others are not covered. Hence there is a discrepancy between the different lakes and rivers typology that can be validated.

- In-situ sensors can be very different from one site to another. Amongst the fixed sensors several means are used: limnometric gauges, pressure sensors, bubbler sensor, ultrasonic or radar sensors, fixed GNSS sensors, lidar microstations. Their strengths, weaknesses and automation possibilities have been listed by the STR3TART project. Moving sensors are also of use during campaigns and their capabilities summarized in STR3TART project.

- Regarding CalVal super sites the St3tart project is emitting recommendations, this can be summarized as follow:

○ Sites must be chosen without strong reflectors that would contaminate the altimetry waveforms (other water bodies, metallic surfaces…).

○ Install a device to identify changes in surface roughness (wind sensor or imager) as this will affect the altimetric waveform and should be taken into account in the inversion.

○ Measure the river slope within the satellite footprint, at its variation as a function of water height temporal variations.

○ For rivers, discharge validation (that is derived from WSH by means of models or rating curves) will also require a sensor measuring water velocity.

○ Install weather stations allows to quantify the errors made on the dry and wet tropospheric corrections from models and then better isolate and quantify the errors linked to waveform retracking.

- The campaigns with drone or ULM are very good opportunities. They are quite easy to deploy and reach a high accuracy. The drone can follow the exact path of a river, which enables to capture the high spatial variability of rivers slopes. It has been noticed that over the duration of the drone flight (a few hours) the river level can evolve (several cm or tens cm), which can affect the slope estimate, as well as the comparison with a satellite pass the same day but at a different time. A balance / complementarity between in-situ permanent sites and drone campaigns should be further investigated to optimise the different on-ground means.

- There is no real harmonization between sites at the moment. A FRM label could help to federate several sites. How to harmonize national initiatives?

4.3.9 FRM labelling

As explained in task 3.3 deliverable, FRM sites already exist for ocean even if they are not labeled as such. It is important to perpetuate them for the validation of the Copernicus Sentinel missions. Only the Crete site is labeled as FRM at the present time.

We identified the following facilities and/or sites to be labelled as FRM sites:

- Corsica Site
- Harvest Site
- Bass Strait Site
- Issykkul Lake Site
- Catalina transponder Site
- Leonessa transponder Site

4.3.10 Synergies with other domains

This is generally difficult for altimetry to identify synergies with other domains since measurements content are quite specific. Nonetheless, we have identified four main fields where validation activities could benefit from synergies:

- Transponder deployment that are used both for altimetry and SAR missions. Even if there are different constraints on both sensors, feasibility studies could be initiated to explore the requested techniques needed to set-up common sites for SAR and altimetry missions.
- Wave validation which could benefit from comparisons with optical and SAR sensors. Synergies between altimetry and SAR are already considered at the level of the Marine Service but not at the Level 2 validation step. This should be further investigated to identify the expected outcomes and potential limitations.
- Sea ice validation which could benefit from comparisons with optical and SAR sensors. Some experiments have been done in the past, mainly for quality assessment of identification of floes and leads. This activity should be further explored to analyse the feasibility to systematize such use of Sentinel data for validation.
- Inland waters validation that would benefit indirectly from determination of water delineation obtained by other sensors (SAR and/or optical sensors). This is not a direct validation input but a valuable source of information that would help refining the validation methods over the water bodies detected by altimetry.

4.3.11 Data distribution

Maintain and extend the availability of cal/val reference measurements through the CMEMS in-situ TAC service.

Maintain a sustainable operational data distribution service for altimetry data over in-land water and land ice.

The CAL/VAL activities shall cope with the heterogeneities of the in-situ networks. It is already integrated in the CAL/VAL methods. This will continue as it is a utopia to try to harmonize all these large sets of data. The need is more on the access to the data. The INSTAC initiative of CMEMS should be extended to other surfaces

4.4 SAR imaging component

4.4.1 Provide easier access to Quality Information

The quality control of Sentinel-1 product is split into two parts:

1) the online quality control consisting of very basic checks of products contents based on inspection of a limited part of the product meta data. The result of the online quality control is a pdf report attached to the Sentinel-1 product. There is no easy way for the end-user to make an automatic check of the content of this pdf report.

2) the offline quality control consisting of more advance analysis of the products meta data and to some of the data. This offline QC allows to spot anomalies that are further investigated by the SAR MPC team working on their resolution. Depending on the anomalies the corrections are either performed by updates of instrument configuration, change of upstream production configuration, update of the SAR processor (Level 1 and/or Level 2) or change of its configuration. The QC anomalies are tracked as a series of Quality Disclaimers that are publicly available. The Quality Disclaimers contains the description of the anomaly (cause, impact), how to identify the impacted products (by set of products characteristics like acquisition time, production date, configuration baseline, etc). However, there is for now no easy way for a user to check the exhaustive list of quality degradation that may impact a single product.

Those two QC methodologies (online and offline) are complementary as the online QC is only able to spot anomalies during the nominal production flow that were previously defined and characterized whereas the offline QC is able to characterize anomalies on historical products.

At the very beginning of the Sentinel-1 mission, the Sentinel-1 data were planned to be disseminated to the end-users using a specific tool able to take into account the quality disclaimers generated by the offline QC. The rationale was to us the quality disclaimers to tag the products with their corresponding discrepancies. Then the user could be able to select and filter the products depending on the types of product degradation. This mechanism was finaly never put in place.

We consider that such a service enabling to tag the products with their known quality degradation in the dissemination service, or a service enabling one user to ask for the quality degradation of a single product based on its main characteristics could be beneficial to the user community.

4.4.2 Ensure access to FRM data for all acquisition modes and configuration

The main FRM used for the Sentinel-1 Calibration and Validation are composed of Corner Reflectors and Transponders operated for the SAR MPC by DLR in Germany. Over Germany, the nominal acquisition scenario of the Sentinel-1 mission applies. Then the nominal acquisition mode over the DLR FRM is the Interferometric Wide Swath (IW) operated in Dual Polarisation V (DV) configuration, enabling to achieve the mission objective over the European territory.

However, Sentinel-1 is able to acquire data on other modes (Extra Wide Swath, Strip Map, Wave Mode) and with other polarimetric configuration (Dual Polarization H, and single polarization H or V), that are actively used over other regions of the globe, still for the need of the objective of the mission. For their proper calibration, those other modes and polarimetric configurations require specific acquisition over the DLR FRM that were very infrequently activated since the end of commissioning of the S1A and S1B spacecraft, and in fact not anymore since the last operational acquisition of S1B in December 2021.

Such acquisition out of the IW/DV configuration over the DLR FRM is in fact conflicting with the nominal acquisition plan over Europe. It is as well conflicting with the now available subsidence product of the Copernicus Land Motion Service, requiring consistent time series of data with a constant acquisition configuration.

Going forward would require either to operate European FRM out of European territory, or to build international cooperation with any other agency deploying SAR FRM primarily for their own constellation.

4.4.3 New geophysical models for observation of ocean with L-Band

The availability of FRM like transponders and corner reflector allows having access to limited validation data with very reference accuracy. However, those measurements are only available on limited geographic areas and to specific acquisition modes and configuration. Even if the network of SAR FRM is increased (refer to recommendation of section 4.4.2), there is the need to perform a concurrent statistical validation and calibration on a larger set of data, most specifically for the radiometric calibration.

For Sentinel-1, this alternative geophysical/statistical radiometric calibration is twofold:

- An absolute radiometric calibration using the NRCS measurement of the sensor over open ocean to be compared to predicted NRCS measurement using wind model and geophysical model in C-Band (knowing the observation conditions)

- A relative radiometric calibration using the Gamma measurement of the sensor over Rain Forest canopy aiming the fine calibration of radiometric continuity near sub swath edges and to a limited extent to cross-check the relative calibration of the various acquisition modes.

Considering the lack of SAR Transponders out of Europe, the radiometric calibration of WV acquisition mode is performed only using this geophysical method.

The geophysical model functions (GMF) related to wind speed over ocean versus NRCS in C-Band are well known and benefit from a large community of scientist from the scaterrometer community. New versions of GMF are produced regularly benefiting from active research of this community.

The literature of GMF in L-Band is scarce and a limited to one single publication in 2008 related to ALOS/PALSAR data, that need to be confronted to more recent and larger set of measurements. More specifically, the impact of the instrument noise on the derivation of such GMF must be closely investigated.

4.4.4 L Band Transponders

The Sentinel-1 mission benefits from set of transponders deployed by DLR in Germany. Those transponders are specifically designed for C-Band sensors and Sentinel-1.

The ROSE-L mission, now in preparation will operate an L-Band sensor for which the calibration/validation will require series of transponders now under development. The availability of L-Band transponder is challenging due to the lower frequency.

4.4.5 Flagging of ocean scenes phenomena

The Sentinel-1 Level 2 products (ocean products) are composed of three main measurements corresponding to wind over ocean, swell over ocean, and surface radial velocity over ocean. Part of the inversion process for those measurement is based on assumptions about separated influence of wind, swell and current on the surface roughness at a given scale. In some specific condition, this assumption is not valid and there may be interaction for instance between wind and wave, wave and current, etc.

In such situation the performances of wind / swell / radial velocity estimation are not optimal. There is thus the need to flag the different parts of the Level 2 products to distinguish between pure wind observation, pure swell observation, etc., or combination of wind / swell at the same scale. Preliminary studies were conducted in this field but need to be consolidated.

This is both important for the performances of the Level 2 products and the ones of the Level 1 products, as the radiometric calibration of some specific modes is currently only performed using a geophysical methodology.

4.4.6 Multi-sensor Cal/Val verification

The Sentinel-1 mission is operated since 2014 with the advent of Sentinel-1A. The products of Sentinel-1B spacecraft was launched in 2016. The two instruments and products are calibrated and validated independently and the performances are cross checked. Any potential bias between the two spacecraft is further investigated to understand its cause and to the maximum extent to align the performances of the two sensors and products.

The Sentinel-1 mission is part of a long series of C-Band SAR, including ESA units (ERS-1, ERS-2, ENVISAT/ASAR), Canadian ones (Radarsat-1, Radarsat-2, RCM), Chines (Gaofen-3). For now there was no dedicated activity performed on cross-validation of cal/val for all those instrument, thus making it difficult to quantify the performances of multi-sensor time series, especially for long period of time.

This multi-sensor Cal/Val verification would require for instance:

- to access reference SAR measurements over the dedicated FRM for each mission, and/or specific measurement of one SAR using the FRM dedicated to another mission
- to access large set of SAR observation of the other mission over ocean (for the geophysical cal/val methodology of absolute radiometric calibration)
- to access large et of SAR observation of Rain Forest canopy (for the cal/val methodology of relative radiometric calibration). For this specific set of data, comparison of methodology between all the mission could be performed as they are using either a “flat gamma” or “non flat gamma” assumption.

This requires both accessing large set of historical SAR data (for the European historical mission) and international collaboration.

4.4.7 Formalize the Cal/Val methodology

The Cal/Val methodology of the SAR sensors is well known, defined, with approach discussed inter alias in the CEOS CWG. The Cal/Val plans for the SAR sensor during its commissioning and operation phase are well defined as well. However, it appears that the Cal/Val algorithms must be better documented compared to the other types of sensors presented in this document.

4.4.8 Formalize feedback from user community

The Sentinel-1 product performances are regularly presented through series of reports and presentation in generic EO conferences and workshops (IGARSS, LPS, SeaSAR, Fringe…). However, in the recent years, no Quality Working Group were organized.

The main difficulty of organization of QWG for SAR is related to the extremely broad range of applications it covers, making it probably difficult to setup QWG groups with consistent representativity of users.

Thematic QWG may be organized to overcome this difficulty.

4.5 Atmospheric composition component

The atmospheric domain is already well advanced on many aspects related to quality information and validation, e.g., in terms of prognostic per-pixel uncertainty estimates produced by the level-2 retrieval processors, the availability of extensive networks for ground-based correlative measurements for several products, an operational validation benefitting from an automated validation server, etc. Nevertheless, the level of maturity of quality assessment and validation depends significantly on the atmospheric variable. Significant gaps and challenges remain, as detailed below.

4.5.1 Mission and user requirements

Mission requirements for level-1 data are closely related to instrument performance requirements and are usually well detailed. Mission requirements for Sentinel level-2 data products are often simplistic and limited to one or two threshold values for the systematic and/or random components of the uncertainty. User requirements can be more detailed but are not always easy to find. Nevertheless, both (mission and user) types of data quality requirements are required for their translation into validation requirements, for the appropriate design of the validation process, and for the verification of the compliance of the Sentinel data against (mission and user) requirements. More detail is provided in CCVS D1.4 (“Atmospheric Composition Missions Cal-Val Requirements”).

4.5.2 Satellite product quality information

Quality information associated with the satellite data products is usually comprehensive and detailed: quality flags, traceable prognostic (ex-ante) uncertainty estimates, quality information documented in Product User Guides (PUM), Algorithm Theoretical Basis Documents (ATBD), validation reports, and Product Readme Files (PRF). However, data content aspects like the impact of data filtering based on these quality flags, for instance in terms of data loss or introduced biases, deserve more attention in the validation process.

4.5.3 Reference data representativeness

There is a need for research on *network design*, in particular for the emerging networks, but also for the sustainability of existing networks. Geographical distribution (global or European) of the measurement stations is important, as well as the coverage and sampling they offer regarding the main influence quantities: do they address high and low column/concentration regimes, the full range of solar and viewing angles, low and high surface albedo and temperature, representative aerosol loads including clean and volcanic extrema, etc. Also needed is an assessment of the *spatio-temporal representativeness of measurements* at each individual station for the definition of optimal co-location criteria and the closure of the error budget of the data comparisons, especially for products with strong spatiotemporal variability like NO_2. This requires both a characterization of the atmospheric variability (e.g., through high-resolution campaign measurements with aircraft or UAVs) and a detailed metrological assessment of the measurements, quantifying their actual 3/4-dimensional spatio-temporal sensitivity.

4.5.4 Reference data quality

As the Sentinel missions represent another step change in accuracy and sensitivity, there is a clear need for parallel improvements in the quality of the reference data: e.g., better spectroscopic parameters and their characterization, further developments of more sophisticated retrieval approaches (e.g., spectral fitting using absorption cross sections at different temperatures for an accurately derived effective temperature or improved a priori input in the RTM), and full uncertainty characterization (random, systematic, mixed, correlations). FRM4xxx and similar programs supported by ESA and EUMETSAT play a pivotal role here. Related is the need for improved inter- and intra-network homogeneity (in particular consistency between UV-Vis and IR measurements of the same species, e.g., NO2, but also within networks of similar instruments, such as the ozonesonde or FTIR networks). This can be supported both through (1) funding for tasks on validation consistency assessments within the satellite validation service contracts (as was done through a supporting extension to the S5P-MPC routine operation validation service contract), and (2) dedicated intercomparison field campaigns at sites (temporarily) hosting multiple types of ground-based and airborne instruments. Validation work carried out in an organized framework like a campaign or a service contract is the ideal context for a feedback loop between satellite and reference data providers, stimulating data quality evolution on both sides as well as mutual understanding of the needs and limitations of each party.

4.5.5 Reference data availability

There is a very strong need for operational and timely reference data processing and QA/QC, and streamlined data access (through well-maintained data distribution centres). For S5P, reference data are available within a couple of days only for total O3 and stratospheric NO2 through the CNRS LATMOS NRT processing facility for the SAOZ network (i.e., a network of automated ZSL-DOAS instruments certified for the NDACC). A dedicated CAMS contract with NDACC (CAMS-27) supports delivery of quality NDACC data within one month for selected targets (O3, CO, CH4, NO2 and HCHO) from a.o. LIDAR, FTIR, UVVIS. Timeliness of availability for ozonesonde data varies with the station, depending often on local manpower. The advent of upgraded calibration and retrieval and of central processing facilities for networks like ACTRIS, PGN and MAX-DOAS should improve in the near future the availability for e.g., NO2, HCHO and SO2 total columns. For other species or types of instruments, reference data latency can still be of the order of months to years and data access may require ad hoc interaction with station PIs. This is a critical weak link in the path towards operational validation of the atmospheric Sentinel missions. Beyond the problems related to reference instrument operation and data availability, the role of well-maintained data centres with harmonised data formats, version control and dedicated QA/QC functions is crucial.

4.5.6 Next generation instrumentation

Instrumentation technology is an essential part of Cal/Val solutions. It corporates the technology used in pre-flight characterisation, on-board calibration and in-situ validation. The technology used must be of the highest possible quality to serve as reference measurements and cover the aspects of Cal/Val activities. With new space instruments and missions, the existing technology needs updating, and new developments are expected to cover the gaps. Also, existing prototypes, e.g. airborne imaging instruments, require more R&D to be deployed in an operational context for Cal/Val. New technologies often incorporate available or slightly modified existing instruments on new platforms like automated floats or unmanned aerial vehicles like drones. The latter is an inexpensive and flexible solution for supporting Cal/Val activities, especially during campaigns.

Efforts must be made to bring technology used in Cal/Val activities to the FRM level and increase its maturity. That includes 1) providing best practice documents, standards, protocols, etc., for operation and data acquisition, and 2) harmonising data analysis based on centralised processors and automatic independent validation tools.

More systematic measurements for airborne (e.g., UAVs, AirCores, etc.) and mobile (car, tram, ship, etc.) observations are needed. The development of off-the-shelf operational instruments should be encouraged. This might complement stationary measurements and increase the spatial coverage of the Cal/Val networks.

If new technology is being developed or maturized with EU funding, there should be an obligation to test in known sites to compare measurements with other reference measurements.

More details and some examples on next generation instrumentation are provided in CCVS D3.1 (“Recommendations for R&D activities on instrumentation technologies”).

4.5.7 Validation methodology

Advances in **validation methodology** are needed on many aspects. While for some the necessary research exists and the remaining step is implementation in an operational system, for others the best way forward deserves a detailed assessment and more fundamental work. Detailed descriptions of the challenges and potential solutions are available in CCVS D3.2 ("Recommendations for R&D on Cal/Val Methods"). Eventually, the atmospheric Copernicus missions would benefit from the right combination of generic validation protocols and product-specific protocols that bring together best practices in for each step in the validation process, like those that were recently defined in other domains (e.g., the CEOS Land Product Validation protocols). Below, the key methodological challenges are briefly summarized:

- **Harmonization approaches**. This concerns, among others, methods to minimize the differences in vertical resolution/sensitivity (the known "vertical smoothing" issues) and vertical coverage (e.g., for ground-based instrumentation on mountain sites), the differences in horizontal resolution (some ground-based measurements being more point-like and thus more sensitive to local gradients than the satellites, while other types of ground-based measurement extending over larger areas and smoothing local gradients differently than the satellites), and the differences in measurement times (in particular an issue for photochemically active species such as NO2 and middle atmosphere O3). While a large body of research on such methods exists, the much needed transfer into the operational validation systems requires first an evaluation of the merits and limitations of each method, and then a (CPU-) efficient and robust implementation.

- Sentinel atmospheric products come with a fairly complete ex-ante (prognostic) uncertainty assessment that is an integral part of the level-2 processors. These uncertainties are provided on a per-pixel level (or pixel-cluster level if pixels are averaged for better signal-to-noise ratio) and a differentiation according to uncertainty source and statistical properties (random vs. Systematic) is often made. **Validation of these prognostic uncertainties** is today done only in a research context and only for a few products (e.g., total ozone columns). Various experimental methods exist, but work is needed to identify strengths and weaknesses, to adopt best practices consistent among Copernicus missions, and to bring these into an operational framework. A prerequiste, in particular for those methods that rely on the comparisons to the ground-based reference data, is knowledge on the complete uncertainty budget of the validation (cfr., the next recommendation).

- The **end-to-end uncertainty budget of the validation process** needs to be established. This includes not only the uncertainties of the satellite and reference measurements but also those introduced by imperfect spatio-temporal co-location (differences in sampling and in smoothing properties), by harmonization (e.g., in vertical gridding and resolution and/or assumed vertical distribution), by the use of auxiliary data (e.g., temperature data needed for various unit conversions), etc. All of these uncertainty sources need to be quantified, and their amplitude will determine the value of the data quality feedback that can be obtained with the validation.

- In the atmospheric domain, the infrastructure that is in place for the validation of the level-2 geophysical products can (and is) also (be) used for **the validation of level-3** products, which are spatio-temporal aggregations of level-2 data onto a regular and usually contiguous grid. This is for instance done for the satellite data products that are delivered to the C3S Climate Data Store. The validation of level-3 adds another layer of complexity, primarily related to (differences in) spatio-temporal representativeness of the measurements within the level-3 grid cells, and the characterization and validation of the ensuing uncertainties. While some exploratory research is published for a few products (ozone and NO2), extensive R&D is still needed, both on the side of the level-3 data producers and on the validation side (see CCVS D3.5, "Performance Analysis and Impact on Level-3").

- In general, there is a large gap to be bridged between methods that have been developed and demonstrated in a pure research context and what is implemented in the operational systems addressing the needs of the space agencies and Copernicus. This gap is even wider when considering methodologies developed in other domains but with potential application in the atmospheric domain (and vice versa). Initiatives to foster **cross-domain transfer of advanced methods** have proven to be fruitful (see e.g. Loew et al., Reviews of Geophysics, v55, 2017) and deserve further sustained support.

4.5.8 Satellite-satellite and satellite-model intercomparisons

These are in general only performed on an ad hoc basis (exception: the feedback on satellite data quality in the operational assimilation systems such as CAMS). While another satellite or model data set cannot be considered as a fiducial reference, such intercomparisons are often very sensitive to small error sources with a specific spatio-temporal structure, and as such highly valuable. They also bring a geographically wider inspection of the Sentinel data quality than what can be achieved by ground-based validation at a set of stations (the detection of striping structures in the Sentinel data, e.g., hardly detectable by point-like ground-based observations). Development work is needed on methodology for harmonization (making intercomparable), for uncertainty validation (e.g., through triple co-location) and for operational implementation.

4.5.9 Quality management

For the Sentinel atmospheric products, data quality management as organized in the ATM-MPC is overall excellent, with quality-assured prototyping and implementation of the retrieval processors, near real time quality control and anomaly tracking of level-1 and level-2 data, routine operational validation coupled to in-depth scientific studies, and extensive and timely documentation. As argued earlier in this document, the feedback loop between satellite data producers, validators, and reference data providers is essential and deserves continued support, and the interaction with the reference data providers further formalization (with associated funding for Sentinel-specific FRM developments and services where necessary).

4.6 Cross-mission recommendations

4.6.1 Improve uncertainty assessment for Sentinel data

The analysis of the cal/val maturity matrix revealed the need to progress on prognostic (ex-ante) uncertainties for some Sentinel variables in order to reach higher maturity levels. Several variables do not have any prognostics uncertainties associated with them while for other variables only some error sources are accounted for in the prognostic uncertainty estimates. Exchanges with Copernicus services also highlighted the importance of robust prognostics uncertainties for some downstream applications such as climate change assessment and data assimilation in simulation models.

The validation of prognostics uncertainties is a challenging task but can be extremely valuable to reveal potential issues in our understanding of the underlying measurement equations.Obviously, the provision of uncertainty values has a strong impact on data processing and distribution, especially if rich information is provided (per pixel uncertainties, multiple components, covariance matrices…) Thus, alternative approaches should be explored: user-side or on-demand generation (using a post-processing tool), download options, short-term rolling archive for uncertainties…

The CCVS project recommends **continuing and reinforcing efforts to develop prognostic uncertainty models for all Sentinel data products** based as far as possible on GUM principles. This effort should be guided by the needs of data users, which could be clarified during dedicated Working Groups (see also next session). It shall be applied in a consistent way from L1 to L3 data.

To achieve this objective, a more systematic documentation of instrument characterization and uncertainties is also needed (see section 4.2.2).

Another important aspect concerns the assessment of the quality of auxiliary classification layers (especially cloud and cloud shadow mask). Indeed, if a pixel is contaminated by a cloud or cloud shadow and not flagged as such, the measurement error is not consistent with prognostics uncertainties. This fact has practical consequences on validation results for optical data products. To better address this issue, a reliable confidence estimate of classification layer is needed. See section 4.2.6 for more on this subject.

The validation of (the effect of) quality flags and other filters and data usage guidelines provided with the data products also deserves more attention than usually given.

4.6.2 Improve and update mission specifications

The Mission Requirements are important during the mission design and implementation phase. But their use does not end at the start of the operational phase: indeed they are used as inputs for the cal/val activities. The documents ensure that cal/val activities are actually serving user needs in terms of validation, reporting and quality improvement.

The Sentinel Mission Requirement Documents should follow some good practices:

- Performance requirements should be specified using proper metrology terminology (e.g. referring to measurement uncertainty and not “error”, and clearly stating the k value to be considered). As far as possible the performance criteria should be unambiguously defined (e.g. statistical processing to be applied).

- The specification should be feasible: the relative measurement uncertainty cannot be bound by a constant percentage if the measurand can reach 0. The specification should mention the range of value over which it is applicable.

More generally, there is a need to periodically revisit the Mission Requirement during the operational phase:

- To add specifications for new data products
- To modify specifications according to new usage of data products
- To amend requirements if needed (e.g. in case of instrument anomaly)

For this purpose, the CCVS project recommends periodic working groups with Copernicus services dedicated to Mission Requirements, in parallel to existing QWG. This working groups should also clarify user needs in terms of prognostic uncertainties and quality flags (see also section 4.6.1). Another mission of this working group could be to ensure a rigorous propagation of uncertainties from L1 up to L3, L4 and in data assimilation for models (see D3.5 for details).

4.6.3 UAV-based measurements

The project considers that UAVs should play a major role in the future of satellite cal/val activities. Indeed, they complement automatic measurements in so far as:

- They can provide additional characterization of instrumented sites (e.g. spatial homogeneity) and support intercomparison between sites
- They can be used on sites where a systematic instrumentation is not available or not possible.

There are today a number of existing platforms and sensors which could be used for validation activities for different type of missions. However, there is a need to develop methodologies and protocols. The project recommends to initiate R&D studies for the different Sentinel parameters addressing:

- Identification of the most suitable sensors for validation activities
- Sensor characterization and calibration protocols
- Uncertainty budgets taking into account UAV geometric uncertainty
- Satellite-UAV intercomparison methodologies
- Comparison between UAVs and ground-based measurements

4.6.4 Cal/Val Methods

4.6.4.1 Uncertainty estimates

There is a general need to improve uncertainty assessment at all levels:

- Not all Sentinel data products provide prognostic uncertainties. Efforts are still needed to generalize prognostics uncertainties following GUM recommendations. In some cases, sensor characterization is missing to estimate some sources of uncertainty (see also 4.2.2)

- Reference data do not always come with measurement uncertainty estimates, or with state-of-the-art measurement uncertainty estimates

- Few vicarious validation methods provide uncertainty estimates.

- Uncertainties coming from intercomparison processes (e.g., temporal and spatial collocation, regridding and smoothing operations) are also not generally assessed.

Assessing uncertainties is a long and difficult effort especially for complex systems like remote sensing measurements. Space Agencies and measurement networks should continue to support efforts in this direction.

There is also a need for training and education on uncertainty management for validation and measurement experts, as well as end users. Training programs (e.g. summer schools) could contribute to this objective.

4.6.4.2 Tandem formation

The CCVS project recommends implementing tandem phases for satellite inter-comparisons whenever possible. This should be done in particular for satellites flying in constellation the same orbit (S1, S2 and S3), typically during the commissioning phase. End-of-life tandem (as foreseen for S2A) is also expected to provide useful information to ensure data continuity over the full mission lifecycle. Tandems with non-Copernicus satellites (e.g. FLEX with Sentinel-3) can provide opportunities for useful inter-comparison activities.

4.6.5 Multi-disciplinary activities

CalVal activities are usually led by experts specialized in on type of observation missions. As a result, there are very few validation activities based on inter-comparison of data coming from different domains. The CCVS project recommends implementing a pilot multi-disciplinary validation activity on the comparison of Wave Height measurements from SAR, optical and Altimetry mission (see section 4.3.10). This activity would require the participation of experts from different domains. If successful, similar activities could be attempted for other variables.

4.7 Organization, coordination, processes and standards

NB: some of the recommendations listed in this section extend beyond the scope of the Copernicus programme.

4.7.1 Labelling of FRM data

The CCVS project recommends implementing a formal procedure for labelling FRM datasets, that ideally should be defined by the CEOS WGCV. This procedure should be decorrelated from considerations regarding eligibility for Copernicus funding.

The procedure could be similar to the Analysis Ready Data (ARD) certification process developed by CEOS WGCV with LAND data products as a pilot (CARD4L). The certification should require FRM data providers to submit an application file providing evidence of their application of the recommended measurement protocols, example data products and metadata and associated documentation. A certification board appointed by CEOS could then review the documentation and grant the certification if appropriate. As for the CEOS ARD certification, different levels could be defined in order to encourage a gradual evolution to higher quality standards.

The expected benefits from this process would be:

- A better visibility for FRM data providers, which should help for data uptake, feedback and sustainability
- A guideline for FRM data users
- By providing an incentive for FRM compliance, it should generally increase the quality of in-situ measurements and promote CEOS-endorsed protocols and standards

In addition, the CCVS project notes that FRM protocols should be considered as living documents which need to evolve as scientific practices and technologies progress. Questioning and criticism of existing protocols as part of a scientific review process should be encouraged.

4.7.2 Collaboration with European Research Infrastructures

European Research Infrastructures will be key in the development of a pan-European environment monitoring network. In this respect, there is a very good complementarity with the Copernicus programme. Some collaborations are already in place (e.g. with ACTRIS) or in discussion (ICOS). The CCVS project recommends to make these collaborations more formal and systematic. ERIs could recognize the provision of cal/val data for the Copernicus programme as one of their objectives, and appoint a cal/val contact point for this activity. The Copernicus programme on the other hand could ensure a regular feedback on data provided and possibly support on data acquisition and processing.

4.7.3 Support to in-situ measurement networks

In-situ measurement networks – generic name including also ground-based remote sensing networks for atmospheric composition - would benefit from centralized facilities to improve the quality of their measurements:

- Travelling SI standards to ensure a common SI traceability throughout a measurement network or between different networks
- Central instrument calibration, characterization and maintenance facilities whenever industrial instrument providers are not able or likely to perform those operations
- Measurement site characterization means, e.g. airborne campaigns to characterize site homogeneity

These activities could be facilitated by dedicated funding sources at European level.

4.7.4 Copernicus Cal/Val Data Distribution service

The acquisition, processing, QA/QC, archiving and distribution of Cal/Val data involves many different entities: space agencies, Copernicus services, European projects, measurement networks, etc. This means that cal/val experts need to collect data from different sources, with generally heterogenous data transfer protocols, formats and terminologies. As highlighted in D3.4, some cal/val data services are performing this activity for the benefit of all cal/val experts (e.g. EVDC, GBOV, CMEMS in-situ component, etc.) The CCVS project recommends to continue and expend these services, and to coordinate their activities in order to build a distributed Copernicus Cal/Val Data Distribution service. This service should work in close interaction with the Copernicus In-Situ component.

D3.4 also highlights the need to build long-term partnerships with other international players (eg CEOS or NASA) and to have a governing body to ensure optimal usage of available cal/val data within the different Copernicus entities.

4.7.5 Field and aerial campaign coordination

With few exceptions, validation campaigns are organized at national level or through mission-specific funding (e.g. ESA campaigns). This limits somehow the valorization of these campaigns for cal/val activities.

For the planning of campaigns:

- Sharing information in advance about planned campaigns could generate opportunities for collaboration or synergies. (e.g. satellite overpass scheduling, inter-comparisons, opportunity payload…)

- Having a European-level policy regarding scientific campaigns could help in the planning and authorization of campaigns. For instance, getting authorizations for UAV flights in urban areas for scientific purposes, or negotiations with third party countries for field campaigns outside Europe.

Dissemination of campaign data is currently a major limitation to its use for cal/val. In many cases, campaign data are not distributed or archived at all, and only used by the campaign team. Often, data are published but with a very long latency, either as a result of scientific embargo, or through lack of operational commitment. Finally, even when campaign data is made available, there is very little effort at harmonization and standardization of the data and metadata (e.g. PANGEA archive). This situation is in contrast with data from systematic measurement networks, for which several data harmonization and distribution exists.

Problems of data availability could be partially solved by negotiating data sharing agreement with data providers. For instance, data could be provided with a short latency for validation activities only with restriction on publications rights. But other efforts (data curation, harmonization and preservation) will not occur without a specific funding.

Within CEOS WGCV space agencies can exchange information about upcoming campaigns. However, this is one on a “best effort” basis with variable regularity. A more constraining process within European agencies could be needed in order to make sure the best return on investment at European level.

Note that such a coordination effort exists for airborne campaigns thanks to EUFAR, which aims at being the reference pan-European portal and network for airborne research infrastructures. This initiative should be supported and consolidated. In addition, there is a need for a simple funding mechanism allowing any site in EU to request the use of EUFAR airborne facilities, with a selection based on the scientific merit and the relevance for European programs such as Copernicus.

4.7.6 Altimetry calibration coordination

For the different services one of the key aspects is the spatial and temporal sampling of the observations assimilated. Today 10 altimetry missions (including 4 Copernicus missions) are flying and providing operational products. To assimilate non-Copernicus mission, the Copernicus services have to rely on non-operational services from collaborative agencies without SLA & long-term commitment.

Extra-activity of homogenezing and completing the Cal/Val metrics of collaborative missions is crucial and is not funded by the Copernicus Program. There is no formal agreement between collaborative agencies and the Copernicus program to coordinate their CalVal activities.

To tackle this issue, the CCVS project recommends to:

Setup multi agency forum between EUM/ESA (Copernicus Program) and Collaborative agencies (CNES/NASA for SWOT, NSOAS/CNES for Hy-2 missions, CNES for SARAL, ESA for Cryosat-2) to ensure CalVal activities are consistent and coordinated.

NRT communication channel is required to meet the timeliness requirements.

Define, set up and fund complementary Collaborative Cal/Val activities needed to meet the Service requirements independently from Collaborative Agencies (e.g. near real time monitoring on essential multi-mission metrics, delayed time comparisons between Sentinels and Collaborative altimeters).

Criticity: high, effort: high

4.7.7 Reporting and communication

It is recommended to publish L2 product validation results with a suitable regularity. Harmonization between EUMETSAT and ESA reporting procedures would be beneficial for users.

4.7.8 Reprocessing strategy

The Copernicus program lacks a systematic reprocessing strategy based on collections with a uniform processing baseline and consistent calibration. This is a limitation in the uptake of Sentinel missions as references especially for climate data records. It is recommended to **plan systematic re-processings whenever substantial changes in product format or quality are implemented** (either through calibration or evolution of the processing).

4.7.9 Improving SI traceability and uncertainty assessment for ground measurement networks

The “FRM4” projects have been very successful at developing guidelines and methodologies to improve the quality and reliability of reference measurements. However, the uptake of these practices by operational networks is still limited. This is due to several factors.

Instrument calibration and characterization require specific expertise and facilities which are not always accessible to PIs on the measurement sites. There is a need to maintain a network of reference laboratories in Europe to provide this type of service. Finally, new sensing technologies should be cross-validated against existing ones with inter-comparison exercises on reference sites.

Assessing measurement uncertainties is often seen as an overwhelming task by some measurement experts. To facilitate their adoption, the project recommends:

- To set-up regular training sessions (e.g. summer schools) explaining the principles of metrology and the practice of uncertainty assessment, especially for young researchers. This could be of interest for satellite cal/val experts as well.

- To encourage the development of community processors to estimate uncertainties in all fields, in collaboration with CEOS. Efforts have already been initiated in this direction (e.g in the frame of FRM4SOC in collaboration with NASA). The CCVS project is currently investigating the feasibility of this for the estimation of LAI using DHPs, in a collaboration with the GBOV project.

5 Conclusion

This document includes recommendations at different levels:

- Actions which can be readily implemented within current organizations, or which should be implemented more systematically

- R&D activities either on instrumentation technology or methods and software, to improve the provision of reference data and their use for the calibration and validation of current and future missions. In particular, we have highlighted the need for high quality reference data beyond the current state of the art for some future missions (CO2M, ROSE-L, CRISTAL…)

- Specific funding efforts needed to bridge reference data gaps in some area (e.g. surface reflectance)

- Organization and coordination suggestions which could make cal/val activities more efficient

The analysis performed by the CCVS team showed the organization of the Cal/Val activities for the Sentinel mission is generally appropriate to meet the scientific and operational objectives of the programme. Mastering measurement uncertainties at all levels, from space sensor to downstream products, as well as for reference data, shall be the overall principle structuring these activities.

One of the most critical aspects concerns the provision of reference in-situ measurements, which involves many contributors (International, European and national Space Agencies, Copernicus Services, International, European and national environmental research institutions and projects).

Mission-critical calibration infrastructures are clearly identified by the Copernicus programme. The development and operations are generally supported directly by the programme. Although a few gaps remain currently (lack of calibration facilities outside Europe for Sentinel-1, lack of an operational system vicarious facility for Ocean Colour and for L1B calibration validation of the atmospheric Sentinels), solutions are actively pursued.

However, for almost all other variables, cal/val activities need to rely on data supported by other entities (generally environmental research infrastructures). In Europe, these infrastructures are primarily developed and supported by national research programs in application of the subsidiarity principle. European Research Infrastructures bring very significant benefits in terms of coordination, data collection, processing and distribution, but they do not support directly the development and operation of the sites. This point has been mentioned several times during exchanges with in-situ measurement experts. This has consequences in terms of sustainability (which are further elaborated in D4.1), as well as in terms of data gaps. This limits both the scientific objectives of the research networks and their usefulness for cal/val activities. Another limitation is that the measurements acquired by these networks are not always “FRM” because they lack well established protocols, SI traceability or uncertainty estimates. A consistent effort has to be deployed to encourage the adoption of FRM standards when they exist or develop them.

6 References

Loew, A., et al. (2017), Validation practices for satellite-based Earth observation data across communities, *Rev. Geophys.*, 55, 779– 817, doi:10.1002/2017RG000562.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Rivier L., et al., ENVRI-FAIR Deliverable 8.1 Atmosphere subdomain FAIRness Assessment (2019). https://envri.eu/wp-content/uploads/2020/06/ENVRI-FAIR_D_8-1_Atmospheric-subdomain-FAIRness-assessment.pdf

Group on Earth Observations, Data Management Principles Implementation Guidelines (2015). https://earthobservations.org/documents/open_eo_data/GEO-XII_10_Data%20Management%20Principles%20Implementation%20Guidelines.pdf

End of Document