Optical Mission Quality Assessment Guidelines

Author(s): ***Samuel Hunt*[1]*, Javier Gorroño*[1]*, Andrea Melchiorre*[2]**

[1]NPL, [2]Telespazio Vega, Task Experts

Approval: ***Nigel Fox***

NPL, Task Lead

Accepted: ***Clement Albinet***

EOP-GMQ EDAP Technical Officer

EDAP.REP.002 *Issue: 1.0* *16 October 2019*

National Physical Laboratory

AMENDMENT RECORD SHEET

The Amendment Record Sheet below records the history and issue status of this document.

ISSUE	DATE	REASON
1.0-DRAFT	17/1/19	Initial draft for review
1.0	16/10/19	First Issue

TABLE OF CONTENTS

1. INTRODUCTION

In recent years, the increasing range of applications of Earth Observation (EO) data products and availability of low-cost satellites has resulted in an increasing number of commercial EO satellite systems, developed with a view to deliver end to end information services, many of which sense the optical domain. These commercial satellite systems may provide complementary capabilities and services to those of Space Agencies. This evolution in the market place means there is potential for some commercial missions to be considered as candidate European Space Agency (ESA) Third Party Missions (TPMs). TPMs are non-ESA developed/owned missions, for which ESA has taken some formal responsibility for providing access to the data through a relationship with the Mission / Data Provider.

To ensure that the data from these missions can be efficiently exploited and that reliable scientific conclusions can be drawn from it, there is a need for data quality to be assessed including the results of appropriate calibration and validation of the satellite sensors (either by the data provider or ESA). The ESA Earthnet Data Assessment Pilot (EDAP) project is intended to establish a framework/methodology and then to perform this assessment for several missions to ensure the delivered data is fit for purpose.

1.1 Scope

This document is intended to provide specific guidelines for the EDAP mission quality assessments of optical sensors, as part of the implementation of the EDAP high-level mission quality assessment [RD-21] for this domain. Section 2 provides some background information on the EDAP mission quality assessment framework. Section 3 then describes considerations for implementing this for optical satellite missions.

1.2 Acronyms & Abbreviations

AERONET	Aerosol Robotic Network
ATBD	Algorithm Theoretical Basis Document
BOA	Bottom-of-atmosphere
CF	Climate & Forecast (Metadata Convention)
CEOS	Committee on Earth Observation Satellites
DCC	Deep convective cloud
DIMITRI	Database for Imaging Multi-Spectral Instruments and Tools for Radiometric Intercomparison
ECV	Essential Climate Variable
EDAP	Earthnet Data Assessment Pilot
EO	Earth Observation
ESA	European Space Agency
FRM	Fiducial Reference Measurement

FRM4GHG	Fiducial Reference Measurements for Ground-Based FTIR Greenhouse Gas Observations
FRM4SOC	Fiducial Reference Measurement for Satellite Ocean Colour
FRM4STS	Fiducial Reference Measurements for validation of Surface Temperatures from Satellites
GAIA-CLIM	Gap Analysis for Integrated Atmospheric ECV Climate Monitoring
HR	High Resolution (spatial resolution between 5 and 30 m)
IVOS	Infrared and Visible Optical Sensors
L1	Level 1
L2	Level 2
L3	Level 3
LR	Low Resolution (spatial resolution higher than 300 m)
MR	Medium Resolution (spatial resolution between 30 and 300 m)
MTF	Modular Transfer Function
NPL	National Physical Laboratory
PICS	Pseudo-Invariant Calibration Site
QA4EO	Quality Assurance Framework for Earth Observation
QA4ECV	Quality Assurance Framework for Essential Climate Variables
RadCalNet	Radiometric Calibration Network
SAR	Synthetic Aperture Radar
SI	Système International (International System of Units)
TOA	Top-of-atmosphere
TPM	Third Party Mission
VHR	Very High Resolution (spatial resolution lower than 5 m)
WGCV	Working Group on Calibration and Validation

1.3 Reference Documents

RD-1 JGCM, Guide to the Expression of Uncertainty in Measurement (GUM), JCGM 100:2008.

RD-2	QA4EO, A guide to content of a documentary procedure to meet the Quality Assurance requirements of GEO, QA4EO-QAEO-GEN-DQK-002
RD-3	QA4EO, A guide to "reference standards" in support of Quality Assurance requirements of GEO, QA4EO-QAEO-GEN-DQK-003
RD-4	Evaluation and Quality Control for Observations, https://climate.copernicus.eu/node/244
RD-5	J. Nightingale et al., "Quality Assurance Framework Development Based on Six New ECV Data Products to Enhance User Confidence for Climate Applications," Remote Sens., vol. 10, no. 8, p. 1254, 2018.
RD-6	CARD4L, Product Family Specification, Surface Reflectance, Working Draft (2017)
RD-7	Eaton et al., "NetCDF Climate and Forecast (CF) Metadata Conventions", Version 1.7 (2017). See: http://cfconventions.org
RD-8	Infrastructure for spatial information in Europe (INSPIRE), https://inspire.ec.europa.eu
RD-9	QA4ECV, QA4ECV Product Documentation Guidance: Algorithm Theoretical Basis Document, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20ATBD%20Guidance.pdf
RD-10	QA4ECV, QA4ECV Documentation Guidance: Product User Manual, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20PUM%20Guidance.pdf
RD-11	JGCM, International Vocabulary of Metrology (VIM 3rd Edition), JGCM 200:2012
RD-12	J. Gorroño et al., "A radiometric uncertainty tool for the sentinel 2 mission," Remote Sens., vol. 9, no. 2, p. 178, Feb. 2017.
RD-13	E. R. Woolliams et al., "Applying Metrological Techniques to Satellite Fundamental Climate Data Records," J. Phys. Conf. Ser., vol. 972, no. 1, p. 012003, Feb. 2018.
RD-14	C. Merchant et al., "Radiance Uncertainty Characterisation to Facilitate Climate Data Record Creation", to appear in Journal of Remote Sensing 2019
RD-15	QA4ECV, QA4ECV Guidance: Provenance Traceability Chains, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20Traceability%20Chains%20Guidance.pdf
RD-16	QA4EO Task Team, A Quality Assurance Framework for Earth Observation: Concept, Version 4.0 (2010). See: http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf

RD-17	I. Barker Snook, Fiducial Reference Measurements for validation of Surface Temperature from Satellites (FRM4STS) Project Brochure (2016). See: http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/D30-FRM4STS-project-brochure-with-cover-16Nov16-signed.pdf
RD-19	QA4ECV, QA4ECV Documentation Guidance: Product Validation and Intercomparison Report, Version 1.0 (2017). See: http://www.qa4ecv.eu/sites/default/files/QA4ECV%20Validation%20Guidance.pdf
RD-20	Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., et al. 2016 The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018. (doi:10.1038/sdata.2016.18).
RD-21	S. E. Hunt et al., Quality Assessment Guidelines, EDAP.REP.001 (2019).
RD-22	Datla et al., Best Practice Guidelines for the Pre-Launch Characterization and Calibration of Instruments for Passive Optical Remote Sensing, NISTIR 7637 (2009)
RD-23	RadCalNet, see: https://www.radcalnet.org
RD-24	FRM4SOC, see: https://frm4soc.org/
RD-25	Vermote, E., Santer, R., Deschamps, P-Y., and Herman, M., “In-flight calibration of large field of view sensors at short wavelengths using Rayleigh scattering” Int. J. Remote Sensing, vol. 13 (1992)
RD-26	B. Fougnie and R. Bach, “Monitoring of radiometric sensitivity changes of space sensors using deep convective clouds: Operational application to PARASOL,” IEEE Trans. Geosci. Remote Sens., vol. 47, no. 3 (2009).
RD-27	K. Thome, N. Smith, and K. Scott, “Vicarious calibration of MODIS using Railroad Valley Playa,” in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, vol. 3, pp. 1209–1211.
RD-28	Y. M. Govaerts, F. Rüthrich, V. John, R. Quast, and V. O. John, “Climate Data Records from Meteosat First Generation Part I: Simulation of Accurate Top-of-Atmosphere Spectral Radiance over Pseudo-Invariant Calibration Sites for the Retrieval of the In-Flight Visible Spectral Response,” *Remote Sens.*, vol. 10, no. 12, p. 1959, Dec. 2018.
RD-29	J. Gorroño *et al.*, “A radiometric uncertainty tool for the sentinel 2 mission,” *Remote Sens.*, vol. 9, no. 2, p. 178, Feb. 2017.
RD-30	G. Chander, T. J. Hewison, N. P. Fox, X. Wu, X. Xiong, and W. J. Blackwell, “Overview of Intercalibration of Satellite Instruments,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 3, pp. 1056–1080, Mar. 2013.

2. EDAP QUALITY ASSESSMENT GUIDELINES

Based on the principles defined in the Quality Assurance framework for Earth Observation (QA4EO) [RD-16], and building on experience developed implementing quality assessment frameworks in other projects (e.g. [RD-4], [RD-5]), the EDAP project has defined a framework to assess the quality of Earth Observation missions [RD-21]. This framework defines the high-level principles and activities that are required in quality assessments for all types of Earth Observation missions, since these are largely common between different domains. This document provides specific guidelines for the assessment of sensors in the optical domain, which is described in Section 3.

The EDAP mission quality assessment framework divides the assessment activity into five areas, these are as follows:

- Product Information
- Product Generation
- Ancillary Information
- Uncertainty Characterisation
- Validation

These sections are themselves divided into sub-sections, which constitute each of the different aspects of the mission that should be assessed and graded, either as of Basic, Intermediate, Good and Excellent. Additionally, a sub-section may receive the grades of Not Assessable or Not Assessed. These grades cover the expected cases that for some EDAP missions, certain aspects of product quality will not be assessed – either because the mission is not yet mature enough to allow the assessment, or because the assessment is currently outside of the scope of this pilot project.

This guide itself provides a generic description of the criteria to achieve a given grade for each sub-section. The starting point for developing these grading criteria is typically to consider what would constitute a near "ideal" scenario, which is hoped to serve as an aspiration to new space providers as well as to space agencies. It is understood that many of the missions the project will assess will only partly comply with these requirements (to different degrees). This is acknowledged in the grading system which is intended to primarily test whether aspects of a given mission are "fit for purpose" within the context of the mission's stated performance and application.

The data product quality assessment for a given mission should finally be reported by populating the template EDAP Data Product Quality Assessment Report (EDAP QA Report). The template is provided to ensure consistency of reporting across the project and to facilitate comparison between the assessments of different missions. The EDAP QA Report itself is a summarising form that covers each section of analysis, linking to supporting and justifying documentation for more detailed information, as well as containing a completed mission quality evaluation matrix (see Figure 1) which diagrammatically presents the results of each sub-section of analysis in a colour-coded table.

Finally, in order to perform their assessment, the mission assessor may be provided with materials from mission operator that are not publicly available, such as documentation or data. While in the ideal case all relevant information would be available to all users, confidentiality may commonly be required for some aspects of missions. In the EDAP Quality Assessment Framework assessment of such confidential materials this is therefore accepted within this context – where this is the case will be indicated in the quality assessment report with a padlock symbol in the corner of the relevant cell of the quality assessment matrix. In general, pertinent key conclusions of confidential documentation should nevertheless still be published openly.

Product Information	Product Generation	Ancillary Information	Uncertainty Characterisation	Validation
Product Details	Sensor Calibration & Characterisation Pre-Flight	Product Flags	Uncertainty Characterisation Method	Reference Data Representativeness
Product Availability & Accessibility	Sensor Calibration & Characterisation Post-Launch	Ancillary Data	Uncertainty Sources Included	Reference Data Quality
Product Format	Retrieval Algorithm Method	If target mission data product is Level 2	Uncertainty Values Provided	Validation Method
User Documentation	Retrieval Algorithm Tuning		Geolocation Uncertainty	Validation Results
Metrological Traceability Documentation	Additional Processing			

Key
Not Assessed
Not Assessable
Basic
Intermediate
Good
Excellent
Information Not Public

Figure 1 – The EDAP Product Quality Evaluation Matrix – uncompleted

3. OPTICAL MISSION QUALITY ASSESSMENT GUIDELINES

In the previous section, the EDAP mission quality assessment framework was introduced. This framework is defined in a high-level document that describes the generic assessment activity required for all types of EO missions [RD-21]. This section provides more detailed guidance for how to implement such an assessment for optical sensors.

The approach here is the same as for the high-level guideline document. Each sub-section of the assessment is discussed in turn, and corresponds to a cell in the mission quality evaluation matrix (see Figure 1). Again, the starting point considers the "ideal" case and grades are based on "fitness for purpose" within the context of the missions' stated performance and application.

Some assessment sub-sections are suitably broad in nature, for example *Product Format*, that the specific guidelines for optical sensors may be no different to the generic case. In this case the section from the original high-level document is repeated in the relevant section of this document for ease for the reader; it is stated when this is the case.

On the other hand, since the optical domain itself covers a broad range of instruments, for some assessment sub-sections different optical sensor types will be handled separately. Distinctions may be drawn in terms of sensor spectral resolution (e.g. multi-channel, hyperspectral) and spatial resolution. The spatial resolution of a sensor may be defined as low resolution (LR; spatial resolution higher than 300 m), medium resolution (MR; 30 to 300 m), high resolution (HR; 5 to 30 m) and very high resolution (VHR; lower than 5 m). This complexity also applies for mission data products of different processing levels, where distinctions may be made for Level 1 (L1) and Level 2 (L2) products.

Finally, it is important to note that these guidelines do not intend to provide absolute criteria on whether any aspect of a given mission attains a given grade – often "expert judgement" is required, especially when considering what is "fit for purpose" in a given context.

3.1 Product Information

The *Product Information* section covers the top-level product descriptive information, product format, and the supporting documentation.

3.1.1 Product Details

Certain basic descriptive information should be provided with any EO data product, and so is required for EDAP assessments of all mission domains. The list of this required information is as follows, with specific requirements for optical sensors added where required:

- Product name
- Sensor Name
- Sensor Type
  For optical sensors – should describe sensor design type, e.g. multi-channel, hyperspectral, interferometer etc., and spectral domain, e.g. visible (VIS), near infrared (NIR), shortwave infrared (SWIR) or thermal infrared (TIR).
- Mission Type
  Either single satellite or constellation of a given number of satellites.
- Mission Orbit
  For example, Sun Synchronous Orbit with Local Solar Time.
- Product version number
- Product ID

- Processing level of product
  For optical sensors – defined for optical sensors as:
  - L0 – uncalibrated instrument counts
  - L1 – calibrated top-of-atmosphere radiance, reflectance or brightness temperature
  - L2a – surface radiance or reflectance
- Measured quantity name
  For optical sensors – radiance or reflectance or brightness temperature, describing spectral bands. Where applicable, include units.
- Stated measurement quality
  To provide context to the reader for the rest of assessment, provide the product "quality" as specified by the provider.

  For optical sensors – this should cover both radiometric and geometric quality. In the radiometric case, quality could be given as a typical per-pixel uncertainty, though, typically providers only give a single mission uncertainty value, which may even be the sensor's required accuracy from its specification.
- Spatial Resolution
  For optical sensors – define full swath and pixel width, include if viewing nadir or tilted off-axis. Categorise as either LR, MR, HR or VHR.
- Spatial Coverage
  Define if data is provided globally or regionally. If regionally define regional coverage.
- Temporal Resolution
  Define repeat time, i.e. time between successive observations of a given location.
- Temporal Coverage
  Define period of mission operation (expected if current mission), including any periods of inactivity during the mission.

Also recommended is the following (based on INSPIRE metadata):

- Point of contact (Responsible organisation, including email address)
- Product locator (e.g. URL, DOI if applicable)
- Conditions for access and use
- Limitation on public access
- Product abstract (summary of resource)
  Short description of product.

Table 3-1 shows how a data product's provision of the above information relates to the grade it achieves for this sub-section of the quality assessment.

Table 3-1 – Product Information > Product Details – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Some pieces of required information missing.
Intermediate	Any required information missing.
Good	All required information available, some recommended information missing.
Excellent	All required and recommended information available.

3.1.2 Product Availability & Accessibility

Note:
The assessment activity and criteria for this sub-section are suitably generic that it is the same for all mission types. The advice is therefore the same as appears in the original high-level mission quality assessment guidelines [RD-21]. This is repeated here for ease for the reader.

This is about how readily the data are available to those who wish to use them. It does not necessarily require cost-free access but is more about following the FAIR Data Principles for scientific data management and stewardship [RD-20], which provide valuable principles for all data applications. These state that:

Data should be **findable**

- Metadata and data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include the identifier of the data it describes
- Metadata and data are registered or indexed in a searchable resource

Data should be **accessible**

- Metadata and data are retrievable by their identifier using a standardised communications protocol
- The protocol is open, free and universally implementable
- The protocol allows for an authentication and authorisation procedure where necessary

Data should be **interoperable**

- Metadata and data use a formal, accessible, shared and broadly applicable language for knowledge representation
- Metadata and data use vocabularies that themselves follow FAIR principles
- Metadata and data include qualified references to other (meta)data

Data should be **reusable**

- Metadata and data are richly described with a plurality of accurate and relevant attributes
- Metadata and data are released with a clear and accessible data usage license
- Metadata and data are associated with detailed provenance
- Metadata and data meet domain-relevant community standards

Table 3-2 shows how a data product's provision of the above information relates to the grade it achieves for this sub-section of the quality assessment.

Table 3-2 – Product Information > Product Availability and Accessibility – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Relevant information not made available.
Basic	The data set does not appear to be following the FAIR principles
Intermediate	The data set meets many of the FAIR principles and/or there is an associated data management plan that shows progress towards the FAIR principles
Good	The data set meets many of the FAIR principles and has an associated data management plan and is available either free of cost or through an easy-to-access commercial licence.
Excellent	The data set fully meets the FAIR principles and has an associated data management plan and is available either free of cost or through an easy-to-access commercial licence.

3.1.3 Product Format

Note:
The assessment activity and criteria for this sub-section are suitably generic that it is the same for all mission types. The advice is therefore the same as appears in the original high-level mission quality assessment guidelines [RD-21]. This is repeated here for ease for the reader.

An important aspect of EO data products that ensures they are most easily accessible to the widest variety of users is their file format. CEOS, through initiatives like CARD4L (CEOS Analysis Ready Data for Land) [RD-6], is promoting the concept of Analysis Ready Data (ARD), which attempts to ensure that data are processed to minimum set of requirements to allow immediate analysis of interoperable datasets.

In the ideal case, an assessed mission product format should meet any appropriate CEOS ARD guidelines, for example CARD4L requirements in the case of SAR and high-resolution optical sensors. In the case where these requirements are not met, product formats are graded based on the following:

- the extent to which they are documented;
- whether standard file formats are used (e.g. NetCDF);
- If they comply with standard variable and metadata naming conventions, such as CF Conventions [RD-7], or, for data from the European Union, the INSPIRE directive [RD-8].

Table 3-3 shows how a given EO data product should be graded for its format.

Table 3-3 – Product Information > Product Format – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Non-standard, undocumented data format.
Basic	Non-standard or proprietary data format, or, poorly-documented standard file format.
Intermediate	Data in documented standard file format. Non-standard naming conventions used.
Good	Data in well-documented standard file format, meeting community naming convention standards.
Excellent	Analysis Ready Data standard if applicable, else as *Good*.

3.1.4 User Documentation

Data products should be accompanied with the following minimum set of documentation for users, which should be regularly updated as required:

- Product User Guide/Manual (PUG/PUM)
- Algorithm Theoretical Basis Document (ATBD)

It may be for a given mission that in place of these documents some combination of articles, publications, webpages and presentations provide a similar set of information. For the highest grades however, they should be presented as a formal document, since users should not be expected to search the information out.

The QA4ECV project provides generic guidance for the expected contents of these documents [RD-9, RD-10], which they can be evaluated against. The PUG should provide general information on the product, including:

- Description of available products.
- Description of how to use the products, i.e. product format.

More specifically for optical sensors, the ATBD should include the following for the L1 processing:

- Overview of the instrument design concept (not necessarily proprietary details).
- Description of the of the radiometric calibration processing, including the sensor measurement function.
- Description of additional processing steps, such as orthorectification.
- Description of the uncertainty analysis performed on this processing.

If a mission’s product is L2 the ATBD should also include:

- Description of the of the retrieval algorithm processing steps.
- Details of assumptions the retrieval algorithm makes.
- Description of the uncertainty analysis performed on the L2 processing.

There are a variety of relevant technical details of varying significance which are important to include in such processing descriptions, for example if the product is in units of reflectance defining the used solar irradiance model. The mission assessor should apply expert judgement to decide the extent to which necessary details are included.

Note that these documents will likely be the source of much of the information required for the other sub-sections of the assessment.

Table 3-4 describes how EDAP grades a products user documentation.

Table 3-4 – Product Information > User Documentation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No user documentation provided, or, documentation out-of-date.
Basic	Limited PUG available, no ATBD. Documentation up-to-date.
Intermediate	Some PUG and ATBD-type information available. May be as formal documents or made up of e.g. articles. Documentation up-to-date.
Good	PUG meeting QA4ECV standard, reasonable ATBD. Documentation up-to-date.
Excellent	PUG ATBD available meeting QA4ECV standard**.** Documentation up-to-date.

3.1.5 Metrological Traceability Documentation

Traceability is defined in the vocabulary of metrology (VIM) [RD-11] as a,

"property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty"

and reinforced in the QA4EO procedures. Traceability is therefore a key aspect of achieving reliable, defensible measurements. In this definition an important part of measurement traceability is highlighted – that it is well documented. This of course must be the case for EO data products too.

Various diagrammatic approaches have been developed to present the traceability chains for EO data products (e.g. the QA4ECV guidance, which includes a traceability chain drawing tool [RD-15]). Such a diagram should be included in the documentation for every EO mission. The FIDUCEO project [RD-13] has provided guidance for a more detailed measurement function centred "uncertainty tree diagram" which is ultimately more suitable for Level 1 (and some Level 2) processing and should be the aspiration for missions in the future.

It is important that such documentation remains up to date. For optical missions, it is common that aspects of a sensor's calibration may be modified or completely changed over the course of a mission. For example, a pre-flight calibration may be updated to one established with an in-flight method. This entirely changes the sensor's traceability chain and should be documented.

Table 3-5 shows how the EDAP grades the product traceability documentation, based on its completeness.

Table 3-5 – Product Information > Metrological Traceability Documentation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No traceability chain documented.
Basic	Traceability chain diagram and/or uncertainty tree diagram included, missing some important steps.
Intermediate	Traceability chain and/or uncertainty tree diagram documented identifying most important steps and sources of uncertainty.
Good	Rigorous uncertainty tree diagram, with, where appropriate a traceability chain documented, identifying all reasonable steps of and accompanying sources of uncertainty.
Excellent	Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes traceability to SI.

3.2 Product Generation

The Product Generation section covers the processing steps undertaken to produce the data product itself. This starts with an assessment of the calibration of the instrument measurements to L1. If the mission under assessment produces a L2 data product, then additional steps of assessment must be undertaken.

3.2.1 Sensor Calibration & Characterisation Pre-Flight

The pre-flight calibration and characterisation campaign should encompass a given sensor's behaviour to an extent and sufficient quality that is "fit for purpose" within the context of the mission's stated performance. For an overview of optical sensor pre-flight calibration and characterisation see the best practice guide located on the CEOS Cal/Val portal [RD-22].

This guide divides the pre-flight calibration into three key stages, which allow for a full understanding of instrument behaviour:

1. Determination of the mission and calibration requirements.
2. Component/subsystem characterisation and sensor performance modelling.
3. System level end-to-end testing and comparison with model.

The way in which this calibration is traceable, preferably to SI, should be identified and an uncertainty budget and associated evidence established that justifies the stated performance.

Characterisation and calibration should be based on the sensor measurement function, which must include all relevant influencing parameters on the sensor measurement. Influencing parameters for optical systems this may be divided into three categories:

- *Radiometric* – including, but not limited to, effects such as linearity, stability, cross talk, polarisation sensitivity, stray light, temperature sensitivity.
- *Spectral* – including, but not limited to, effects such as spectral responsivity, stability, spectral stray light.
- *Spatial* – including, but not limited to, effects such as spatial resolution, FoV, geometric location, MTF, image quality.

The mission assessor should apply their expert judgement to determine for a given instrument (e.g. multiband, hyperspectral), with stated performance and application area, which of these factors are required to be characterised.

Finally, best practice dictates that ideally one should "test as you fly". This means that the tests should be performed in the same environment the sensor will operate in, i.e. thermally and under vacuum.

Note that some aspects of the instrument calibration and characterisation may be determined with additional tests in-flight, however they should still ideally also be tested on-ground. In particular, many aspects of sensor behaviour are limited or impossible to characterise in-flight, such as the spectral response function, therefore it is key these are determined as part of the pre-flight campaign.

Table 3-6 shows how EDAP grades pre-flight sensor calibration and characterisation.

Table 3-6 – Product Generation > Sensor Calibration & Characterisation – Pre-Flight – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Pre-flight calibration & characterisation not documented or information not available.
Basic	Pre-flight calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose.
Intermediate	Pre-flight calibration & characterisation covers most important aspects of instrument behaviour at a level of quality to be judged fit for purpose.
Good	Pre-flight calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Calibration traceable to SI or community reference, characterisation meets good practice.
Excellent	As *Good*, additionally calibration and characterisation includes the measurements needed to assess uncertainties at component level and their impact on the final product.

3.2.2 Sensor Calibration & Characterisation Post-Launch

As in the pre-flight case, the post-launch calibration and characterisation activity should encompass a given sensor's behaviour to an extent and sufficient quality that is "fit for purpose" within the context of the mission's stated performance.

Also, as in the pre-flight case, post-launch characterisation and calibration should be based on the sensor measurement function – though the extent to which an instrument can be characterised in-flight is limited compared to an on-ground campaign.

For a review of the various in-flight radiometric calibration methods see APPENDIX A. Methods include inter-calibration with other satellite sensors, vicarious calibration to in-situ reference measurements and calibration to simulated radiances from so-called pseudo-invariant calibration sites (PICS). APPENDIX A should allow the assessor to judge the extent to which a given in-flight calibration method can achieve a stated performance. It should be noted, that unfortunately for some common post-launch calibration methods rigorous uncertainty analysis and justified traceability is not currently available. For this reason, it is recommended that the methods where metrological best practices are followed are used – for example, the RalCalNet (Radiometric Calibration Network) sensor network [RD-23] or the data from ESA's Fiducial Reference Measurement (FRM) campaigns (e.g. FRM4STS [RD-17] and FRM4SOC [RD-24] amongst others).

Note that though different methods may primarily be suited for either absolute in-flight calibration or validation/monitoring activity, some are suitable for both. The same method or at least the same team/site in the case of an FRM, should not be used for both the post-launch calibration and the post-launch validation, these must be independent.

For a discussion of the various in-flight geometric calibration methods see APPENDIX B. This should allow the assessor to judge the extent to which a given in-flight geometric calibration method can achieve a stated performance. The methods largely depend on whether the sensor is LR, MR, HR or VHR.

Table 3-7 shows how EDAP grades post-launch sensor calibration and characterisation.

Table 3-7 – Product Generation > Sensor Calibration & Characterisation – Post-Launch – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Post-launch calibration & characterisation not documented or not available.
Basic	Post-launch calibration & characterisation misses some important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose.
Intermediate	Post-launch calibration & characterisation covers most important aspects of instrument behaviour at a level of quality to be judged fit for purpose and uses appropriate community infrastructure/methods (CEOS/FRMs).
Good	Post-launch calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance and uses appropriate community infrastructure/methods (CEOS/FRMs).
Excellent	Post-launch calibration & characterisation covers all reasonable aspects of instrument behaviour to a quality that is "fit for purpose" in terms of the mission's stated performance. Measurements fully traceable to SI or community reference at an uncertainty commensurate with the product specification and carried out regularly across the full range of observational conditions of the product and dynamic range.

3.2.3 Retrieval Algorithm Method – Level 2 Only

Note:
The assessment activity and criteria for this sub-section are suitably generic that it is the same for all mission types. The advice is therefore the same as appears in the original high-level mission quality assessment guidelines [RD-21]. This is repeated here for ease for the reader.

For many types of L2 products there are typically a variety of potential retrieval methods that may be used to derive them. These may vary in ways such as model complexity and computational efficiency – resulting in higher or lower quality final products.

As with the L1 sensor calibration, the L2 retrieval method should be of a sufficient quality that is "fit for purpose" within the context of the mission's stated performance across all stated use cases (e.g. scene types). What this requires is specific to a given variable's retrieval methods and will require a degree of expert judgement.

Table 3-8 shows how EDAP grades the algorithm retrieval method used to generate L2 products.

Table 3-8 – Product Generation > Retrieval Algorithm Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Retrieval method not documented.
Basic	Retrieval method too simple to be judged "fit for purpose" in terms of the mission's stated performance.
Intermediate	Reasonable retrieval method used, judged "fit for purpose" in terms of the mission's stated performance for most expected use cases, with at least a sensitivity analysis carried out.
Good	Sophisticated retrieval method used, "fit for purpose" in terms of the mission's stated performance all expected use cases and validated performance against similar algorithms or with empirical evidence.
Excellent	State-of-the-art retrieval, easily "fit for purpose" in terms of the mission's stated performance, full uncertainty budget derived and validated.

3.2.4 Retrieval Algorithm Tuning – Level 2 Only

Note:
The assessment activity and criteria for this sub-section are suitably generic that it is the same for all mission types. The advice is therefore the same as appears in the original high-level mission quality assessment guidelines [RD-21]. This is repeated here for ease for the reader.

Level 2 retrieval algorithms often require some initial "tuning" or calibration against reference data, such as in-situ measurements. The reference datasets used in this process must be of a sufficient quality, size and representativeness (in terms of factors like scene type or dynamic range) in order to achieve the mission's stated performance across all stated use cases. What this requires is specific to the retrieval method used and may require some expert judgement.

Table 3-9 shows how EDAP grades a mission's retrieval algorithm tuning.

Table 3-9 – Product Generation > Retrieval Algorithm Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Retrieval algorithm tuning not documented.
Basic	Algorithm tuned, but to data that is not of a sufficient quality or sufficiently representative to be judged "fit for purpose" in terms of the mission's stated performance.
Intermediate	Algorithm tuned to data that is of a quality or representativeness that is "fit for purpose" in terms of the mission's stated performance for most use cases.
Good	Algorithm tuned to data traceable to SI, potentially through an FRM.
Excellent	Algorithm tuned to data traceable to SI, potentially through an FRM. Representative of all stated use cases and all input parameters fully traceable with robust uncertainties.

3.2.5 Additional Processing

Additional processing steps are separate to the main sensor calibration or retrieval processing. These may include processes like resampling or the generation of classification masks. Additional processing steps must themselves be assessed for quality based on their "fitness for purpose" in the context of the mission.

Typical additional processing steps performed on optical mission products include the following:

- Orthorectification
- Resampling
- Cloud masking

The algorithm for these steps should be documented, including assumptions made and relevant process specific details (such as digital elevation model used in orthorectification). In the case of additional processes where the measurement data itself is transformed in some manner, such as orthorectification, uncertainties from the original data must also be propagated through the processing for them to remain meaningful, as well as introducing appropriate additional uncertainty components caused by the processing itself.

Each additional processing step should be separately assessed and based on the criteria described in Table 3-10, and then a combined score determined.

Table 3-10 – Product Generation > Additional Processing – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Additional processing steps not documented.
Basic	Additional processing steps documented. Some important additional processing steps may not be fit for stated purpose.
Intermediate	Additional processing steps documented. All significant additional processing steps are fit for stated purpose.
Good	Additional processing steps documented. All additional processes steps fit for stated purpose.
Excellent	All additional processing steps fully documented and state-of-the-art.

3.3 Ancillary Information

In addition to its core measured variables, the assessment of which is covered by the *Product Generation* section, EO data products typically contain a variety of additional ancillary information to facilitate interpretation and further analysis of the data. This section of the mission quality assessment evaluates this ancillary information both in terms of its quality and completeness (i.e. do users have access to all the relevant information they need).

3.3.1 Product Flags

Product Flags offer users important extra layers of useful descriptive information on top of the measurements themselves. They can include information on the performance of the instrument, such as indicating periods of unusual instrumental behaviour where the data should not be used, or classification information as to the type of pixel.

For the user is it important that flags are clearly named and documented and that they cover an appropriate breadth of information. What is exactly required will depend on the instrument type and the intended use case.

As an example, for optical sensors, a minimum set of a flags may include:

- Data dubious flag
- Data unusable flag

A reasonable set of flags for optical sensors may also contain:

- Classification masks for e.g. land and ocean

A comprehensive set of flags for optical sensors may also contain:

- Flags for more detailed aspects of instrument behaviour, e.g. saturated pixels.
- Cloud mask (though more commonly expected at L2 than L1). Can be as a boolean flag or as data with percentage likelihood.

For the EDAP criteria for grading a product's flags see Table 3 11.

Table 3-11 – Ancillary Information > Product Flags – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Product flags not available or not documented.
Basic	A limited set of product flags of poorly documented product flags available.
Intermediate	A limited set of well documented product flags available, but mostly binary in nature e.g. relative to a threshold.
Good	A reasonable set of well documented product flags available, including meaningful gradation i.e. % of clouds
Excellent	A comprehensive set of well document product flags full gradation where appropriate and many provided or calculable at pixel level.

3.3.2 Ancillary Data

Ancillary Data provides users with vital additional data layers to properly define, interpret and analyse a product's measurement data. Where appropriate this data should be uncertainty quantified.

As a minimum all information required to properly define the primary measured data should be included in the data product. For optical sensors this includes information such as:

- Sensor spectral response function data
- Viewing and illumination angles
- Longitude, latitude, altitude

Other information, though not strictly required to define the measurement, may be useful to interpret the measurements or for further analysis. Inclusion of this kind of data, though it may be available or derivable from other sources, is convenient for users and considered advantageous. For optical sensors this may include information such as:

- Meteorological data, such as wind speed, humidity etc.
- In-band solar irradiance.

For the EDAP criteria for grading ancillary data provision see Table 3-12.

Table 3-12 – Ancillary Information > Ancillary Data – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Ancillary data not available.
Basic	Key ancillary data provided that is required to define measurement.
Intermediate	All ancillary data provided that is required to define measurement.
Good	All ancillary data provided that is required to define measurement, uncertainty quantified where appropriate. Some additional data provided required to interpret measurements.
Excellent	All ancillary data provided that is required to define measurement, uncertainty quantified where appropriate. All key additional provided required to interpret measurements.

3.4 Uncertainty Characterisation

To ensure measurements are both meaningful and defensible it is crucial that they come with rigorously evaluated uncertainty estimates. This section of the mission quality assessment evaluates the methodology used to estimate uncertainty values for a given mission, the extent of the mission's analysis and how the values are provided.

3.4.1 Uncertainty Characterisation Method

Note:
The assessment activity and criteria for this sub-section are suitably generic that it is the same for all mission types. The advice is therefore the same as appears in the original high-level mission quality assessment guidelines [RD-21]. This is repeated here for ease for the reader.

A comprehensive description of how to evaluate sources of uncertainty in a measurement and propagate them to a total uncertainty of the final measurand is provided by the metrological community in the Guide to the Expression of Uncertainty in Measurement (GUM) [RD-1]. This is the approach that should be taken by all EO missions.

A rigorous treatment of uncertainty in EO data should consider the error-covariance between product pixels. Pixel-level errors are often highly correlated on scales that are very relevant to the kind of analysis typically performed, for example, the combination data from different spectral bands or spatial binning. Additionally, many scientific applications, such as data assimilation or optimal estimation retrieval algorithms, can exploit data error-covariance information to achieve more accurate results.

The field of Earth Observation metrology has progressed greatly in recent years. Operational missions are developing different approaches to evaluate and distribute metrologically rigorous uncertainties for L1 and L2 product. For example, ESA's Sentinel-2 mission has developed an on-the-fly, pixel-level uncertainty evaluation tool [RD-12]. There have also been some initiatives, like the FIDUCEO project, that attempt to apply metrology to historical sensor data records [RD-13].

That said, Earth Observation metrology is still a developing field and it is still more common for uncertainties to be evaluated in a manner that does not comply with the GUM. It is still typical for values like the specification performance value or single offset from a comparison sensor to be quoted as the uncertainty.

Table 3-13 describes how EDAP grades a mission's uncertainty characterisation methodology.

Table 3-13 – Uncertainty Characterisation > Uncertainty Characterisation Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Uncertainty characterisation not performed or method not documented.
Basic	Uncertainty established by limited comparison to measurements by other sensor/s Not by independent assessment and then comparison.
Intermediate	Limited use of GUM approach, and/or, an expanded comparison to measurements by other sensors.
Good	GUM approach to estimate measurement uncertainty with full breakdown of components and separated as Type A or B classification.
Excellent	GUM approach to estimate measurement uncertainty, including a treatment of error-covariance.

3.4.2 Uncertainty Sources Included

In addition to the methodology used to determine the uncertainty caused by given error sources the breadth of different error sources analysed must also be assessed. This again is judged on the basis of what is "fit for purpose" in the context of a mission's stated performance. All contributions relevant at the required level of uncertainty should be included in the mission's uncertainty budget. Again, what is required is specific to given instruments and will require a degree of expert judgement.

This requires reviewing each of the uncertainty components estimated during calibration and characterisation activity pre-flight and post-launch, relying on appropriate testing being carried out in these stages.

Table 3-14 describes how EDAP grades the extent of uncertainty sources included in a mission's uncertainty characterisation analysis.

Table 3-14 – Uncertainty Characterisation > Uncertainty Sources Included – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Uncertainty characterisation not performed or sources analysed not documented.
Basic	Some important sources of uncertainty missing.
Intermediate	Most important sources of uncertainty included.
Good	All important sources of uncertainty included.
Excellent	All reasonable sources of uncertainty included.

3.4.3 Uncertainty Values Provided

As described in Section 3.4.1, uncertainty values should be provided in EO data products per-pixel, in a manner that describes the pixel error-covariance. Since it is not practical to provide a full error-covariance matrix for an EO data product due to their data volume various approaches have been developed to approximate this. For example, the FIDUCEO project FCDRs contain three components of uncertainty describe the three typical scales of error correlation [RD-14], these are:

- *Independent uncertainties* – uncertainties that arise from errors that are uncorrelated, i.e. random, between pixel measurements. For optical mission products typical causes are error effects such as noise.
- *Structure random uncertainties* – uncertainties that arise from errors that are correlated in a structured manner over spatial or temporal scales (e.g. image column, image row or scanline). For optical mission products typical causes may be a regular calibration cycle. This causes errors that are regionally systematic (over pixels in a given calibration cycle), but independent between regions (between pixel in different calibration cycles).
- *Common uncertainties* – uncertainties that arise from errors that fully correlated, i.e. systematic, over a full mission. For optical missions this will typically be caused by calibration or characterisation data used for the whole mission, such as non-linearity characterisation.

It is also important to quantify the interband error-correlation for optical satellite missions. This is important as many retrievals use a combination of data from different bands.

It is still typical however for uncertainty values to be provided, if at all, on a per-product or, more often, a per-mission basis – losing a great deal of information significant to users. Table 3-15 shows how EDAP grades missions for the extent of uncertainty information they provide. Note where the mission is a constellation, this assessment should be related to how the data provider assesses or provides their data and evidence on how sensor to sensor variation is accounted for if applicable.

Table 3-15 – Uncertainty Characterisation > Uncertainty Values Provided – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No uncertainty information provided.
Basic	Single uncertainty value provided for whole mission.
Intermediate	Single uncertainty value provided for subsets of data, e.g. per product.
Good	Total uncertainty per pixel is provided, with basic breakdown of key components no error-covariance.
Excellent	Uncertainties per pixel provided with error-covariance information for all appropriate components.

3.4.4 Geolocation Uncertainty

Geolocation uncertainty is typically described as a circular error associated to a certain confidence level (e.g. 95%). It is a less common for geolocation uncertainty to be described in a more detailed manner. For example, the geolocation error might be dependent on the latitude position, time of the year.

Similar to the measurement uncertainty in 3.4.1, the uncertainty associated to the geolocation requires a description of an error-covariance matrix when the product information is processed. This can be estimated by considering the correlation length scales of the errors observed in the reference points measured to test geolocation error. Furthermore, the inaccurate geolocation of a pixel can result in an incorrect estimation of the measurement irrespectively of the uncertainty associated to the data product. Thus, the measurement uncertainty in 3.4.1 and the geolocation one are interrelated.

For optical sensors the geolocation error should not only be considered for the product as whole, but also relative between bands, i.e. the inter-band mis-registration.

Table 3-16 gives the EDAP grading criteria for geolocation uncertainty.

Table 3-16 – Uncertainty Characterisation > Geolocation Uncertainty – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No uncertainty information provided
Basic	Single uncertainty value provided for whole mission.
Intermediate	Uncertainty value provided includes dependency on several variables.
Good	Uncertainty value provided includes dependency on several variables. Includes error-covariance information between pixels
Excellent	Uncertainty value provided includes dependency on several variables. Includes error-covariance information between pixels and impact on measurement uncertainty.

3.5 Validation

CEOS Working Group on Calibration & Validation (WGCV) defines validation as,

“the process of assessing, by independent means, the quality of the data products”

Validation therefore should assess the consistency between both the data values and their uncertainties with those of independent reference data. For optical sensors both radiometric and geometric (particular for high-resolution sensors) validation is required.

3.5.1 Reference Data Representativeness

By the representativeness of the set of reference data we refer to the extent to which the measurements reflect the satellite measurements that they are being used to validate (e.g. point to pixel considerations for radiometric validation using in-situ reference data), over the full extent of measurements the satellite may make (e.g. dynamic range, seasonal variation, global coverage). This may in general require the use of a variety of different datasets to cover different observation conditions.

For a discussion of radiometric reference data representativeness in common optical sensor validation methods see APPENDIX A.

For a discussion of geometric reference data representativeness see APPENDIX B.

Table 3-17 describes how EDAP grades the extent of validation reference data representativeness.

Table 3-17 – Validation > Reference Data Representativeness – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Reference measurements assessed to be somewhat representative of the satellite measurements, covering a limited range of satellite measurements. Typically a one-off campaign.
Intermediate	Reference measurements assessed to be mostly representative of the satellite measurements, covering a primary range satellite of measurements and at ad hoc opportunities (no formal documented regular timescale).
Good	Reference measurements assessed to be well representative of the satellite measurements, covering a reasonable range of the satellite's measurements and carried out using FRM or community approved methods. Carried out on a regular timescale of approximately annual basis but not necessarily based on need.
Excellent	Reference measurements independently assessed to be fully representative of the satellite measurements, covering the satellite's full range of measurements and with full assessment of uncertainties and carried out on a regular basis determined by product performance.

3.5.2 Reference Data Quality

In the same way these guidelines describe how to assess the quality of satellite mission data, similar considerations must be made of the reference data used to validate them. Primarily, this concerns the following:

- Is uncertainty and error correlation information provided with the data?
- Have the data uncertainties been estimated with the GUM methodology?
- Is the data traceable to SI or a community reference standard?

As mentioned in 3.2.2, often the reference data for some validation methods lack rigorous uncertainty analysis and justified traceability at this time. For this reason, it is recommended that the datasets where metrological best practices are followed are used – for example, for radiometric validation, RadCalNet (Radiometric Calibration Network) [RD-23] or the data from ESA's Fiducial Reference Measurement (FRM) campaigns (e.g. FRM4STS [RD-17] and FRM4SOC [RD-24] amongst others).

For a discussion of reference data quality in common optical sensor validation methods see APPENDIX A.

For a discussion of geometric reference data quality see APPENDIX B.

Table 3-18 describes how EDAP grades validation reference data quality.

Table 3-18 – Validation > Reference Data Quality – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Uncertainty information not available for reference data.
Intermediate	Reference data comes a single uncertainty for the entire dataset.
Good	Reference data comes with full uncertainty information, assessed following the GUM and traceable to community reference or SI (e.g. FRM)
Excellent	Reference data comes with full uncertainty and error-correlation information, assessed following the GUM and traceable to SI (e.g. FRM).

3.5.3 Validation Method

A metrologically-rigorous validation should assess both the satellite measurements and their associated uncertainties. Commonly used values such as the statistical spread of the results may be used to estimate the uncertainty, however this often may not provide a realistic estimate of the actual uncertainty.

Validated uncertainties provide evidence of the credibility of the uncertainty estimate given. For optical missions, radiometric validation can often also come in the form of instrument performance temporal stability monitoring.

For a discussion of common optical sensor radiometric validation methods see APPENDIX A. As mentioned in 3.2.2 though different methods may primarily be suited for either absolute in-flight calibration or validation/monitoring activity, some are suitable for both. The same method should not be used for both the post-launch calibration and validation, these must be independent.

For a discussion of geometric validation methods see APPENDIX B. The method required for a given mission will largely be driven by the sensors spatial resolution class – LR, MR, HR or VHR.

Table 3-19 shows how EDAP grades validation methodology.

Table 3-19 – Validation > Reference Data Representativeness – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Methodology is simple comparison, uncertainties not considered.
Intermediate	Methodology assess satellite measurements, simple uncertainty estimated e.g. from statistical spread for results.
Good	Methodology assesses satellite measurements and reference data w.r.t. their uncertainties.
Excellent	Methodology assess satellite measurements and reference data w.r.t. their error-covariance and validates those uncertainties.

3.5.4 Validation Results

Note:
The assessment activity and criteria for this sub-section are suitably generic that it is the same for all mission types. The advice is therefore the same as appears in the original high-level mission quality assessment guidelines [RD-21]. This is repeated here for ease for the reader.

This final sub-section of the validation quality assessment deals with the results of the validation activities themselves. In the best case these will show both validated satellite measurement and uncertainties and will have been obtained by a group independent of the satellite mission owner.

The results should be documented in a Validation report, prepared following the QA4ECV guidance for expected content [RD-19].

Table 3-20 how EDAP grades validation results.

Table 3-20 – Validation > Validation Results – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Validation results show some agreement between satellite and reference measurement.
Intermediate	Validation results show good agreement between satellite and reference measurements within uncertainties in most cases.
Good	Validation results show excellent agreement between satellite and reference measurements, within uncertainties. Analysis performed independently of satellite mission owner.
Excellent	Validation results show excellent agreement between satellite and reference measurements, within uncertainties. Uncertainty validated. Analysis performed independently of satellite mission owner.

APPENDIX A IN-FLIGHT RADIOMETRIC CALIBRATION AND VALIDATION METHODS FOR OPTICAL SENSORS

This appendix offers a short summary of some of the most common methods for optical satellite sensor in-flight radiometric calibration and validation. These methods can broadly be categorised as follows:

- calibration to simulated radiances from so-called pseudo-invariant calibration sites (PICS);
- vicarious calibration to in-situ reference measurements;
- inter-calibration with other satellite sensors.

Different methods are primarily suitable for either absolute in-flight calibration or validation/monitoring activity, though some are suitable for both.

For a more detailed review of satellite calibration methodologies see, for example, [RD-30].

A.1 Ocean Targets – Rayleigh Scattering

Description

Clear open ocean scenes are selected for this method, with low wind and aerosol. In this case up to 90 % of the TOA signal in the visible part of the spectrum comes from Rayleigh scattering in the atmosphere, which may be accurately modelled along with other smaller components of signal for the absolute calibration of a satellite sensor. The method was first developed in Vermote et al. 1992 [RD-25].

Scope of Representativeness

Scenes are dark, relatively bright in the blue. For use in the visible.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method are currently not available. Recent work suggests that state of the art application of this technique can achieve uncertainty of around 5 % for the simulated radiances [RD-28].

Radiometric Calibration/Validation Methods

Absolute calibration.

A.2 Pseudo-invariant Calibration Sites (PICS), desert sites

Description

Pseudo-invariant calibration sites (PICS) temporally stable and spatially homogeneous sites which can be radiometrically modelled to simulate TOA radiances to monitor and calibrate satellite sensors. Many desert sites are ideal PICS due to their high spatial homogeneity and lack of cloud cover. Six desert sites have been identified by CEOS as reference sites – Libya 4, Mauritania 1, Mauritania 2, Algeria 3, Libya 1 and Algeria 4.

These sites may also be used to transfer the calibration from one satellite sensor to another without the need for simultaneous nadir overpasses.

Scope of Representativeness

Visible to shortwave infrared.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method are currently not available. Recent work suggests that state of the art application of this technique can achieve uncertainty in the region of 5 % for the simulated radiances [RD-28].

Radiometric Calibration/Validation Methods

Radiometric stability monitoring, absolute calibration.

A.3 Deep Convective Cloud Targets

Description

Deep convective clouds (DCCs) are very bright, almost white (from the visible to near-infrared) and relatively Lambertian clouds commonly found in the tropics. Due to how well DCCs behave as solar diffusers they may be used for accurate inter-band calibration and stability monitoring relative to reference band. See, for example, Fougnie & Bach 2009 [RD-26] for an example of the use of this methodology.

Scope of Representativeness

Scenes are bright and spectrally flat. For use in the visible to near-infrared.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method are currently not available. Recent work suggests that state of the art application of this technique can achieve uncertainty of around 5 % for the simulated radiances [RD-28].

Radiometric Calibration/Validation Methods

Inter-band calibration and stability monitoring.

A.4 In-situ Measurements

Description

Satellite sensors can be absolutely calibrated against field measurements that are propagated from bottom-of-atmosphere (BOA) to top-of-atmosphere (TOA) with radiative transfer modelling (RTM). Field measurements may either be from:

- One off/regular field measurements campaigns.
  For example, ESA's Fiducial Reference Measurement (FRM) campaigns (e.g. FRM4STS [RD-17] and FRM4SOC [RD-24], FRM4VEG amongst others).
- Permanently instrumented, autonomous sites or networks of sites.
  For example, the CEOS RalCalNet (Radiometric Calibration Network) [RD-23].

Scope of Representativeness

Typically, visible to near-infrared (dependent of field instrument).

Quality

These measurements can have traceability chains and quantified uncertainties, though is not ubiquitous across the field. The aforementioned RadCalNet and FRM campaigns are designed to be metrological rigorous as so are recommended. For RadCalNet instrumented sites typical achievable satellite sensor calibration uncertainty can be < 5 % (e.g. [RD-27]).

Note that RadCalNet provides free data for the sites for nadir view in 30-minute intervals and for 10 nm spectral resolution. For sensors aiming for uncertainties below 10 % these RadCalNet data will need careful interpretation to ensure that these assumptions are useful. The RadCalNet site owners can also provide data with higher temporal and spectral resolution and in some cases for other viewing angles.

Radiometric Calibration/Validation Methods

Absolute calibration and stability monitoring.

A.5 Simultaneous Nadir Overpasses

Description

This method involves calibrating a given satellite sensor using another reference satellite sensor. This is accomplished by locating events called simultaneous nadir overpasses (SNOs), where the given sensor and reference sensor view the same place on the Earth at the same time (within given temporal and spatial tolerances). The uncertainty of the calibration achievable by this method is improved by using many SNO observation between the pair of satellites.

Scope of Representativeness

Visible to shortwave infrared, depending on reference satellite sensor.

Quality

Full traceability and uncertainty quantification for this method requires the reference satellite sensor data to come with uncertainty information and justified traceability.

Level 1 uncertainties, though still not available for many satellite missions, are beginning to become more common. For example, a software tool described in Gorrono et al. 2017 [RD-29] provides L1 per pixel uncertainties for Sentinel-2 images – typical values are around 2 %.

Full traceability to SI for satellite sensors is currently not available, though is planned in the proposed TRUTHS and CLARREO missions.

Radiometric Calibration/Validation Methods

Absolute calibration.

APPENDIX B IN-FLIGHT GEOMETRIC CALIBRATION METHODS FOR OPTICAL SENSORS

This appendix offers a short summary of some of the most common methods for optical satellite sensor in-flight geometric calibration and validation.

The driver behind which reference dataset and analysis method is appropriate for a given mission is largely the sensor's stated spatial resolution (LR, MR, HR, VHR) and their target geometric accuracy.

B.1 Field Survey Ground Control Points

Description

Ground control points (GCP) collected from a field survey can be used as reference points of known location. The accuracy of each GCP needs to be extremely high (30 cm) and each GCP needs to be well defined in the object space in order to achieve a subpixel pointing. Once all GCPs in the set have been identified, true location and predicted location can be compared statistically. This method is very accurate but also relatively time consuming. It is useful for accuracy analysis.

Scope of Representativeness

Visible to shortwave infrared, depending on the number and quality of *in situ* GCPs.

Quality

Full traceability and uncertainty quantification for this method requires the methodology and instrumentation used to acquire the GCPs, uncertainty information due to the GPS receiver and the definition of the GCP at different resolution.

Geometric Calibration/Validation Methods

Absolute geometric accuracy.

B.2 Reference Raster Dataset

Description

The method is based on the use of a reference raster dataset of known geometric accuracy. Generally, this method is based on the extraction of the same GCP from the reference dataset and the target product of unknown accuracy. Generally, this method still provides good results; however, the selection of GCPs from both raster products can be time consuming and subject to inaccuracies due to GCP selection and illumination changes. If images have illumination changes, pre-processing of the optical products is often necessary.

Scope of Representativeness

Near to shortwave infrared, depending on the test product and reference product.

Quality

Full traceability and uncertainty quantification for this method requires the methodology uncertainty information, any post-processing applied including outlier removal. Also, uncertainties introduced by different spatial resolution and/or temporal decorrelation (i.e. the acquisition date and time difference) between the test and reference product has to be reported.

Geometric Calibration/Validation Methods

Multitemporal geometric stability, relative geometric accuracy.

B.3 Image Matching

Description

Image matching of sensor images may be used to monitor the evolution of geometric accuracy within a product, and is used to investigate band-to-band misregistration.

The method is based on the use of a reference raster dataset of known geometric accuracy. Generally, this method is more straightforward than obtaining field survey reference data or reference image GCPs as it compares the overlapping extent of two raster data products, it is repeatable, scalable and it can be used for different scopes. Generally, this method still provides good results; however, intensity correlation methods such as Normalised Cross Correlation (NCC) do not work well for calibration especially if images have illumination changes, thus pre-processing of the optical products is often necessary.

Scope of Representativeness

Near to shortwave infrared, depending on the test product and reference product.

Quality

Full traceability and uncertainty quantification for this method requires the methodology uncertainty information, any post-processing applied including outlier removal. Also, uncertainties introduced by different spatial resolution and/or temporal decorrelation (i.e. the acquisition date and time difference) between the test and reference product has to be reported.

Geometric Calibration/Validation Methods

Multitemporal geometric stability, relative geometric accuracy, band-to-band registration.