Earth Observation Mission Quality Assessment Framework - Atmospheric Guidelines

Author: ***Samuel Hunt[1] & Alessandro Piro[2]***

[1]NPL, QA Expert

[2]SERCO, Atmospheric Mission Expert

Approval: ***Nigel Fox***

NPL, Task Lead

Accepted: ***Clément Albinet***

ESA, Project Technical Officer

EDAP.REP.061 *Issue: 0.1* *31 October 2021*

AMENDMENT RECORD SHEET

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

ISSUE	DATE	REASON
0.1	31/10/21	Early issue of atmospheric guidelines

CONTRIBUTING AUTHORS

Clément Albinet, ESA

Valentina Boccia, ESA

Nigel Fox, NPL

Javier Gorroño, NPL

Philippe Goryl, ESA

Alfreda Hall, NASA

Samuel E. Hunt, NPL

Alexei Lyapustin, NASA

Guoqing (Gary) Lin, NASA

Will McCarty, NASA

Jaime Nickeson, SSAI/NASA GSFC

Batuhan Osmanoglu, NASA

Alessandro Piro, Serco

Sébastien Saunier, Telespazio

Bin Tan, SSAI/NASA GSFC

Compton Tucker, NASA

Eric Vermote, NASA

Emma Woolliams, NPL

Robert E. Wolfe, NASA

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 a growing number of commercial EO satellite systems, developed with a view to deliver end-to-end information services, many of which sense the atmospheric domain. This evolution in the marketplace has led to increasing interest from Space Agencies in the acquisition of commercial EO data products, as they may provide complementary capabilities and services to those they currently offer.

To ensure that decisions on commercial data acquisitions can be made fairly and with confidence, there is a need for an objective framework with which their data quality may be assessed. The ESA Earthnet Data Assessment Pilot (EDAP) project therefore defines this EO mission quality assessment framework, within which the project performs quality assessments of commercial satellite missions in the optical, SAR and atmospheric domains. Presented here is the latest evolution of this framework for atmospheric missions, which is now under development as a collaboration between ESA and NASA.

1.1 Scope

This document is intended to provide specific guidelines for mission quality assessment of atmospheric sensors, as part of the implementation of the generic EO mission quality assessment [RD-1] for this domain. Section 2 provides a summary of the mission quality assessment framework. Section 3 provides a review of the optical mission quality, as evidenced by its documentation. Finally, Section 3.1 provides guidelines for verifying the mission data quality is consistent with the stated performance of the sensor.

1.2 Acronyms & Abbreviations

APA	Absolute Positional Accuracy
ATBD	Algorithm Theoretical Basis Document
BBR	Band-to-Band co-Registration
BOA	Bottom-of-atmosphere
CF	Climate & Forecast (Metadata Convention)
CEOS	Committee on Earth Observation Satellites
DCC	Deep convective cloud
DDR	Detector-to-Detector co-Registration
ECV	Essential Climate Variable
EDAP	Earthnet Data Assessment Pilot
EO	Earth Observation
ESA	European Space Agency
FAIR	Findable, Accessible, Interoperable and Reusable

FOV	Field of View
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
FTIR	Fourier Transform InfraRed spectroscopy
FWHM	Full Width Half Maximum
GCP	Ground Control Point
GFOV	Ground Field of View
GSD	Ground Sampling Distance
HCS	Horizontal Cell Size
HR	High Resolution (spatial resolution between 5 and 30 m)
HIS	Horizontal Sampling Interval
IVOS	Infrared and Visible Optical Sensors
L1	Level 1
L2	Level 2
LR	Low Resolution (spatial resolution coarser than 300 m)
LSF	Line Spread Function
MR	Medium Resolution (spatial resolution between 30 and 300 m)
MTF	Modulation Transfer Function
NPL	National Physical Laboratory, UK
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)
SNR	Signal-to-Noise Ratio
SSR	Sensor Spatial Response
TOA	Top-of-atmosphere
TPM	Third Party Mission
VHR	Very High Resolution (spatial resolution finer than 5 m)
WGCV	Working Group on Calibration and Validation

1.3 Reference Documents

[RD-1] S. E. Hunt, "Earth Observation Mission Quality Assessment Framework," 2021.

[RD-2] QA4EO Task Team, "Quality Assurance for Earth Observation Principles," 2010. [Online]. Available: http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf.

[RD-3] J. Nightingale *et al.*, "Ten Priority Science Gaps in Assessing Climate Data Record Quality," *Remote Sens.*, vol. 11, no. 8, p. 986, Apr. 2019, doi: 10.3390/rs11080986.

[RD-4] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, 2016, doi: 10.1038/sdata.2016.18.

[RD-5] CEOS LSI, "CARD4L Product Family Specification - Surface Reflectance," 2020. [Online]. Available: https://ceos.org/ard/files/PFS/SR/v5.0/CARD4L_Product_Family_Specification_Surface_Reflectance-v5.0.pdf.

[RD-6] B. Eaton *et al.*, "NetCDF Climate and Forecast (CF) Metadata Conventions," 2020. [Online]. Available: https://cfconventions.org/latest.html.

[RD-7] European Parliament and Council of the European Union, *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).* 2007, pp. 1–14.

[RD-8] INSPIRE Drafting Team Metadata and European Commission Joint Research Centre, "INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119," 2013. [Online]. Available: https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1.

[RD-9] INSPIRE Thematic Working Group Orthoimagery, "Data Specification on Orthoimagery – Technical Guidelines," 2013. [Online]. Available: https://inspire.ec.europa.eu/id/document/tg/oi.

[RD-10] T. Scanlon, "QA4ECV Product Documentation Guidance: Product User Manual," 2017. [Online]. Available: http://www.qa4ecv.eu/sites/default/files/QA4ECV PUM Guidance.pdf.

[RD-11] T. Scanlon, "QA4ECV Product Documentation Guidance: Algorithm Theoretical Basis Document," 2017. [Online]. Available: http://www.qa4ecv.eu/sites/default/files/QA4ECV ATBD Guidance.pdf.

[RD-12] R. U. Datla, J. P. Rice, K. R. Lykke, B. C. Johnson, J. J. Butler, and X. Xiong, "Best practice guidelines for pre-launch characterization and calibration of instruments for passive optical remote sensing," *J. Res. Natl. Inst. Stand. Technol.*, vol. 116, no. 2, p. 621, 2011, doi: 10.6028/jres.116.009.

[RD-13] M. Bouvet *et al.*, "RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range," *Remote Sens.*, vol. 11, no. 20, p. 2401, Oct. 2019, doi: 10.3390/rs11202401.

[RD-14] N. Fox, "FRM4STS D-180 Final Report," 2019. [Online]. Available: http://www.frm4sts.org/wp-content/uploads/sites/3/2020/01/OFE-D-180-V1-Iss-1-Ver-1-signed.pdf.

[RD-15] R. Vendt, "FRM4SOC D-290 Final Report," 2020. [Online]. Available: https://frm4soc.org/wp-content/uploads/filebase/parentdir/techreports/temp_pic/D-290-FRM4SOC-FR_30.06.2020.pdf.

[RD-16] JGCM, "International vocabulary of metrology – Basic and general concepts and associated terms (VIM)," *JGCM*, vol. 200, 2012.

[RD-17] T. Scanlon, "QA4ECV Product Documentation Guidance: Provenance Traceability Chains," 2017. [Online]. Available: http://www.qa4ecv.eu/sites/default/files/QA4ECV Traceability Chains Guidance.pdf.

[RD-18] J. Mittaz, C. J. Merchant, and E. R. Woolliams, "Applying principles of metrology to historical Earth observations from satellites," *Metrologia*, vol. 56, no. 3, p. 032002, Jun. 2019, doi: 10.1088/1681-7575/ab1705.

[RD-19] JCGM, "Evaluation of measurement data - Guide to the expression of uncertainty in measurement," *JCGM*, vol. 100, 2008, Accessed: Feb. 06, 2018. [Online]. Available: https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.

[RD-20] J. Gorroño *et al.*, "A radiometric uncertainty tool for the Sentinel-2 mission," *Remote Sens.*, vol. 9, no. 2, p. 178, Feb. 2017, doi: 10.3390/rs9020178.

[RD-21] M. Burgdorf, I. Hans, M. Prange, J. Mittaz, and E. Woolliams, "FIDUCEO D2.2 (Microwave): Report on the MW FCDR Uncertainty," 2019. [Online]. Available: https://cordis.europa.eu/project/id/638822/results.

[RD-22] F. Rüthrich, E. Woolliams, Y. Govaerts, R. Quast, and J. Mittaz, "FIDUCEO D2.2 (MVIRI): Report on the MVIRI FCDR Uncertainty," 2019. [Online]. Available: https://cordis.europa.eu/project/id/638822/results.

[RD-23] G. Holl, E. Woolliams, and J. Mittaz, "FIDUCEO D2.2 (HIRS): Report on the HIRS FCDR Uncertainty," 2019. [Online]. Available: https://cordis.europa.eu/project/id/638822/results.

[RD-24] M. Taylor, J. Mittaz, M. Desmons, and E. Woolliams, "FIDUCEO D2.2 (AVHRR): Report on the AVHRR FCDR Uncertainty," 2019. [Online]. Available: https://cordis.europa.eu/project/id/638822/results.

[RD-25] T. Scanlon, "QA4ECV Product Documentation Guidance: Validation and Intercomparison Report," 2017. [Online]. Available: http://www.qa4ecv.eu/sites/default/files/QA4ECV Validation Guidance.pdf.

[RD-26] 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, doi: 10.1109/TGRS.2012.2228654.

[RD-27] E. Vermote, R. Santer, P. Y. Deschamps, and M. Herman, "In-flight calibration of large field of view sensors at short wavelengths using Rayleigh scattering," *Int. J. Remote Sens.*, vol. 13, no. 18, pp. 3409–3429, Dec. 1992, doi: 10.1080/01431169208904131.

[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, doi: 10.3390/rs10121959.

[RD-29] A. Lyapustin *et al.*, "Scientific impact of MODIS C5 calibration degradation and C6+ improvements," *Atmos. Meas. Tech.*, vol. 7, no. 12, pp. 4353–4365, 2014, doi: 10.5194/amt-7-4353-2014.

[RD-30] 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, pp. 851–861, 2009, doi: 10.1109/TGRS.2008.2005634.

[RD-31] 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, doi: 10.1109/IGARSS.2001.976794.

[RD-32] T. C. Stone, H. Kieffer, C. Lukashin, and K. Turpie, "The moon as a climate-quality radiometric calibration reference," *Remote Sens.*, vol. 12, no. 11, pp. 1–17, 2020, doi: 10.3390/rs12111837.

2. EO MISSION QUALITY ASSESSMENT FRAMEWORK SUMMARY

This section outlines the overall EO mission data product quality assessment framework. The evaluation is primarily aimed at verifying that mission data has achieved the claimed mission performance and, where applicable, reviews the extent to which the missions follow community best practice in a manner that is "fit for purpose".

The approach taken to assess data product quality is based on the QA4EO principle [RD-2] and builds on the structure and reporting style developed in other similar work (e.g. [RD-3]). This quality assessment framework, developed within the ESA Earthnet Data Assessment Pilot (EDAP) project, aims to build on the experience of this previous work targeting the satellite Cal/Val context.

The assessment itself is conducted in two parts, as follows:

- *Documentation Review* – review of mission quality as evidenced by its documentation.
- *Detailed Validation* – quantitative assessment of product compliance with stated performance.

These parts of the assessment, along with their grading criteria, are described in Sections 3 and 3.1, respectively. The activities are divided into sections and subsections constituting each of the different aspects of data product quality that are assessed and graded. Assessment results are provided in a separate Quality Assessment (QA) Report and are also summarised in a colour-coded Cal/Val maturity matrix.

It is expected that all relevant mission information needed to perform the assessment would be available to all users, however it is understood that confidentiality may be required for some aspects of a mission. Where this is the case, it will be indicated as confidential in the quality assessment report. In general, pertinent key conclusions of confidential documentation should nevertheless be published openly.

2.1 Quality Assessment Report

The quality assessment for a given mission is reported using the QA Report template. The template ensures consistency of reporting and facilitates comparison between the assessments of similar missions. The QA Report covers each section of analysis, providing more detailed information, as well as including a completed mission Cal/Val maturity matrix (see following subsection) presenting the results of each sub-section of analysis in a colour-coded table.

2.2 Cal/Val Maturity Matrix

A Cal/Val maturity matrix provides a high-level colour-coded summary of the quality assessment results. The matrix contains a column for each section of analysis, and cells for each subsection of analysis. Subsection grades are indicated by the colour of the respective grid cell, which are defined in the key. A padlock symbol in the corner of given cell indicates that the information used to assess the respective subsection is not available to the public. The reporting of assessment results is divided between two Cal/Val maturity matrices, as follows:

- *Summary Cal/Val Maturity Matrix*
- *Detailed Validation Cal/Val Maturity Matrix*

These matrices are described below.

2.2.1 Summary Cal/Val Maturity Matrix

The *Summary Cal/Val Maturity Matrix* provides an overall summary of the quality assessment results (see

Figure 1). The matrix on the left (in dark blue) summarises the results of the *Documentation Review*, while the additional column on the right (in light blue) summarises the results of the *Detailed Validation*. The *Validation Summary* column is separated from the main table to make clear the results can come from multiple assessment sources.

Data Provider Documentation Review		
Product Information	**Metrology**	**Product Generation**
Product Details	Radiometric Calibration & Characterisation	Calibration Algorithm
Availability & Accessibility	Geometric Calibration & Characterisation	Geometric Processing
Product Format, Flags & Metadata	Metrological Traceability Documentation	Retrieval Algorithm
User Documentation	Uncertainty Characterisation	Mission-Specific Processing
	Ancillary Data	

Validation Summary
Radiometric Validation Method
Radiometric Validation Results Compliance
Spectral Validation Method
Spectral Validation Results Compliance
Geometric Validation Method
Geometric Validation Results Compliance

Key
Not Assessed
Not Assessable
Basic
Good
Excellent
Ideal
Not Public

Figure 1 - Summary Cal/Val Maturity Matrix and Key. To be colour-coded to report results of assessment.

2.2.2 Detailed Validation Cal/Val Maturity Matrix

The *Detailed Validation Cal/Val Maturity Matrix* (see

Figure 2) provides more complete reporting of analysis behind the *Validation Summary* – breaking down the validation methodologies used and the results. This section is aimed at the more technically focused reader. Since, for a given mission multiple validation studies may be performed – for example, by the mission/vendor and/or by independent assessors – there can be multiple *Detailed Validation Maturity Matrices* produced and reported.

Validation Summary		Detailed Validation			
Radiometric Validation Method	←	Radiometric	Absolute Calibration Method	Signal to Noise Method	Temporal Stability Method
Radiometric Validation Results Compliance	←	Radiometric	Absolute Calibration Results Compliance	Signal to Noise Results Compliance	Temporal Stability Results Compliance
Spectral Validation Method	←	Spectral	Absolute Calibration Method	Temporal Stability Method	
Spectral Validation Method	←	Spectral	Absolute Calibration Results Compliance	Temporal Stability Results Compliance	
	←	Geometri	Absolute Positional Accuracy Method	Temporal Stability Method	Line of Sight Method

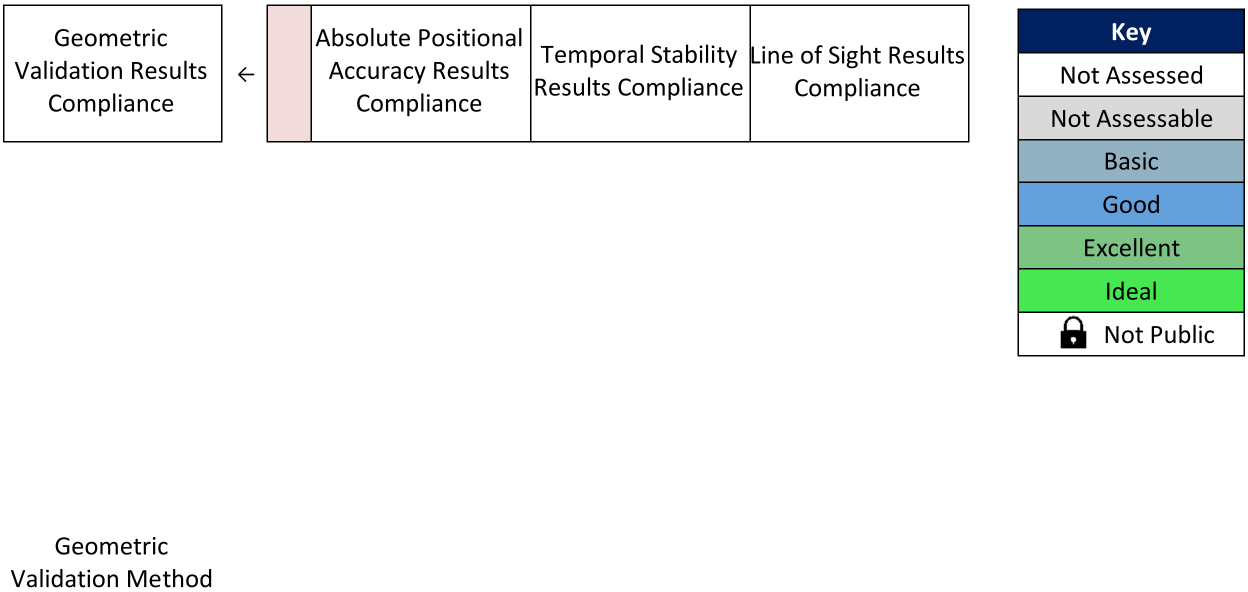

Figure 2 – Validation Cal/Val Maturity Matrix for the atmospheric domain, which includes the Validation Summary column from the Summary Cal/Val Maturity Matrix

2.3 Approach to Grading

The assessment framework is aimed at verifying the claimed mission performance, and that the mission follows community best practice to an extent that is "fit for purpose". The grading criteria for each category are determined based on a logical interpretation of this principle. For example, pre-launch calibration quality grading is based on the comprehensiveness of activity with respect to the target instrument performance.

Grades of Basic, Good, Excellent, or Ideal may be given. The Ideal grade level is generally reserved to provide recognition for achieving the highest standard of quality with respect to community best practice. This high bar of quality may be aspirational but is the benchmark that EO data providers should aim for. Note that a grade of Basic can also be considered acceptable in a given context.

Additionally, a subsection may also indicate Not Assessable or Not Assessed. These cover the cases where certain aspects of product quality will not be assessed – either because there insufficient information available to make an assessment, or because it is out of scope of the assessment.

3. DATA PROVIDER DOCUMENTATION REVIEW

In this section we provide detailed guidelines for *Documentation Review*. This assessment aims to review mission quality as evidenced by its documentation. It is divided into the follow sections:

- Product Information
- Metrology
- Product Generation

In the following we look at each of these sections in turn and discuss the grading criteria.

The results of the *Documentation Review* are reported on the left portion of the *Summary Cal/Val Maturity Matrix*. This portion is shown in Figure 3.

Data Provider Documentation Review		
Product Information	**Metrology**	**Product Generation**
Product Details	Radiometric Calibration & Characterisation	Calibration Algorithm
Availability & Accessibility	Geometric Calibration & Characterisation	Geometric Processing
Product Format, Flags & Metadata	Metrological Traceability Documentation	Retrieval Algorithm
User Documentation	Uncertainty Characterisation	Mission-Specific Processing
	Ancillary Data	

Figure 3 – Data Provider Documentation Review Matrix

3.1 Product Information

The *Product Information* section covers the top-level product descriptive information, product format, and the supporting documentation. Its subsections are now defined.

3.1.1 Product Details

Certain basic descriptive information should be provided with any EO data product and is required for assessment of all mission domains. The list of this required information is as follows:

- Product name
- Sensor Name
- Sensor Type
- 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
- Measured quantity name
- Stated measurement quality
  To provide context to the reader for the rest of assessment, provide the product "quality" as specified by the provider.
- Spatial Resolution
- Spatial Coverage
  The full swath width and footprint of a scene or single acquisition. Define if data's spatial coverage, i.e., if provide global or for specific regions.
- Temporal Resolution
  Define repeat/revisit time, i.e., time between successive observations of a given location.
- Temporal Coverage
  Define period of mission operation (expected if current mission)
- Point of contact (Responsible organisation, including email address)
- Product access (e.g., URL, DOI if applicable)
- Restrictions for access and use, if any

Table 3-1 shows how provision of data product information relates to its grade 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	Many pieces of important information missing.
Good	Some pieces of important information missing.
Excellent	Almost all required information available.
Ideal	All required information available.

3.1.2 Availability & Accessibility

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 (Findable, Accessible, Interoperable, Reusable) Data Principles for scientific data management and stewardship [RD-4], 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 > 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
Good	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
Excellent	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.
Ideal	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, Flags and Metadata

An important aspect of EO data products that ensures ease of access to the widest variety of users is their format. Product metadata and flags offer users important extra layers of useful descriptive information, in addition to the measurements themselves, that can be crucial to their analysis.

In the ideal case, the product format would meet the appropriate Committee on Earth Observation Satellites (CEOS) Analysis Ready Data (ARD) metadata guidelines, such as CEOS ARD for Land (CARD4L) [RD-5] requirements in the case of surface reflectance products.

In the case where such a standard does not exist, product format is graded based on the following:

- the extent to which it is documented
- whether a standard file format is used (e.g., NetCDF)

- whether it complies with standard variable, flag and metadata naming conventions, such as the Climate and Forecast (CF) metadata Conventions [RD-6], or, for data from the European Union, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive [RD-7].
- whether flags and metadata provide an appropriate breadth of information

If product is derived from a constellation of satellites, the specific satellite used should be included in the product metadata.

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

Table 3-3 – Product Information > Product Format, Flags and Metadata – 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. Minimal useful metadata or data flags provided.
Good	Data exist in a documented standard file format. Non-standard naming conventions used. Includes a good set of documented metadata and data flags.
Excellent	Data are organized a well-documented standard file format, meeting community naming convention standards. Comprehensive set of metadata and data flags.
Ideal	Analysis Ready Data standard if applicable, else as *Excellent*.

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 guidance for the expected contents of these documents [RD-8], [RD-9], which they can be evaluated against.

Table 3-4 describes how the assessment framework 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. Information is up-to-date.
Good	Some PUG and ATBD-type information available. These may be formal documents or from multiple sources. Documentation is up-to-date.
Excellent	PUG meets QA4ECV standard, reasonable ATBD. Documents are up-to-date.
Ideal	PUG and ATBD available meeting QA4ECV standard. Documents are up-to-date.

3.2 Metrology

Metrology is the science of measurement. This section covers the aspects of the mission related to measurement quality, including calibration, traceability and uncertainty. The Metrology subsections are now defined.

3.2.1 Measurement Calibration & Characterisation

The sensor's calibration and characterisation for measurement, pre-launch and on-orbit, should encompass a given sensor's behaviour to an extent and quality that is "fit for purpose" within the context of the mission's stated performance, based on its measurement function.

What this requires is specific to given instrument types, which will be discussed in the sensor specific assessment implementation guidelines and will require a degree expert judgement. However, in general, for post-launch calibration and characterisation, where a CEOS endorsed method/test-site is available this should as a minimum be used to monitor sensor performance throughout the mission.

Table 3-5 shows how the assessment framework grades pre-flight sensor calibration and characterisation.

Table 3-5 – Metrology > Measurement Calibration & Characterisation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Pre-flight and post-launch measurement calibration & characterisation activities are not documented or information not available.
Basic	Pre-flight and post-launch measurement calibration & characterisation documentation does not include important aspects of instrument behaviour and/or is not entirely of a level of quality to be judged fit for purpose.
Good	Pre-flight and post-launch measurement calibration & characterisation documents cover most important aspects of instrument behaviour at a level of quality to be judged fit for purpose.
Excellent	Pre-flight and post-launch measurement calibration & characterisation efforts cover all reasonable aspects of instrument behaviour to a quality that is “fit for purpose” in terms of the mission’s stated performance. Pre-flight calibration is traceable to SI or standard reference, characterisation methods meet good practice. Post-launch Cal/Val uses appropriate community infrastructure/methods (e.g. RadCalNet).
Ideal	In addition to meeting *Excellent* criteria, calibration and characterisation include the measurements needed to assess uncertainties at the component level and their impact on the final product. Post-launch Cal/Val uses appropriate community infrastructure/methods traceable to SI (e.g. FRMs, RadCalNet).

3.2.2 Geometric Calibration & Characterisation

As for measurement calibration and characterisation, geometric calibration and characterisation, pre-flight and on-orbit, 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.

Pre-flight this includes the calibration and characterisation of the geometric aspects of the sensor, such as field of view, as well as other components of the satellite that influence the geometric processing should be characterised, such as star trackers or attitude control systems. Post-launch relevant performance parameters should be temporally monitored.

Again, what this requires is specific to given instrument types and calibration methods, which will be discussed in the sensor specific assessment implementation guidelines and will require a degree expert judgement. However, for post-launch calibration and characterisation where a CEOS endorsed method/test-site is available this should as a minimum be used.

Table 3-6 shows how the assessment framework grades post-launch sensor calibration and characterisation.

Table 3-6 – Metrology > Geometric Calibration & Characterisation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Geometric calibration & characterisation not documented or not available.
Basic	Geometric 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.
Good	Geometric calibration & characterisation covers most important aspects of instrument behaviour at a level of quality to be judged fit for purpose.
Excellent	Geometric 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. Post-launch characterisation uses appropriate community infrastructure/methods (e.g., from CEOS).
Ideal	In addition to meeting *Excellent* criteria, geometric calibration and characterisation includes the measurements needed to assess uncertainties at the component level and their impact on the final product. The quality is "fit for purpose" in terms of the mission's stated performance, and meets the science users expectations.

3.2.3 Metrological Traceability Documentation

Traceability is defined in the vocabulary of metrology (VIM) [RD-10] 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-11]). Such a diagram should be included in the documentation for every EO mission. The FIDUCEO project 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 [RD-12].

Table 3-7 shows how the assessment framework grades the metrological traceability documentation, based on its completeness.

Table 3-7 – Metrology > 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.
Good	Traceability chain and/or uncertainty tree diagram documented identifying most important steps and sources of uncertainty.
Excellent	Rigorous uncertainty tree diagram, with a traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty.
Ideal	Rigorous uncertainty tree diagram and traceability chain documented, identifying all reasonable steps and accompanying sources of uncertainty. Establishes traceability to SI.

3.2.4 Uncertainty Characterisation

To ensure measurements are both meaningful and defensible, it is crucial that they include rigorously evaluated uncertainty estimates. 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-13]. The GUM approach should be applied to all EO missions.

The application of Earth Observation metrology has progressed greatly in recent years. Increasingly, providers of operational and reprocessed data products are applying different approaches to evaluate and distribute metrologically rigorous error-covariance information for L1 and L2 product at the per pixel level, as required by climate studies. For example, ESA's Sentinel-2 mission has developed an on-the-fly, pixel-level uncertainty evaluation tool [RD-14]. There have also been some initiatives, like the previously mentioned FIDUCEO project, that have applied metrology to historical sensor data records [RD-15].

With that said, it is typical for uncertainties (or performance estimates) to be evaluated in a manner that does not comply with the GUM, for example, the performance specification value or single offset from a comparison sensor may be quoted as the uncertainty.

Table 3-8 shows the uncertainty characterisation grading under the assessment framework.

Table 3-8 – Metrology > Uncertainty Characterisation – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No uncertainty information provided.
Basic	Uncertainty established by limited comparison to measurements by other sensor/s.
Good	Limited use of GUM approach, and/or, an expanded comparison to measurements by other sensors. Most important sources of uncertainty are included.
Excellent	Full GUM approach is used to estimate measurement uncertainty, all important sources of uncertainty are included. Uncertainty per pixel provided.
Ideal	Full GUM approach is used to estimate measurement uncertainty, including a treatment of error-covariance. Per pixel uncertainties in components, e.g., random systematic – as appropriate for the error-correlation structure of the data

3.2.5 Ancillary Data

Throughout the processing chain there may be a requirement for external input data, for example, atmospheric state information, a digital elevation model or reference data for algorithm tuning. The ancillary datasets used during the processing should be identified to the user (where possible due to commercial sensitivity). Ideally this should be traceable on a per product level.

Ancillary datasets must be of a sufficient quality, including the application of suitably rigorous metrology, for example, in the form of SI traceability.

The suitability of the ancillary data for its application must also be considered, with respect to the mission's stated performance requirements. For example, the quality, size and representativeness of algorithm input data. The requirements will be specific to the retrieval method used and may require some expert judgement.

Table 3-9 shows how the ancillary data are graded under the assessment framework.

Table 3-9 – Metrology > Ancillary Data – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Use of ancillary data undocumented.
Basic	Ancillary data used in product generation, specified to some extent, though incomplete. Not entirely of a sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.
Good	Ancillary data used in product generation, specified, though not necessarily on a per product basis. Mostly of a sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.
Excellent	Ancillary data used in product generation, fully specified per product, and traceable. Ancillary data used are of sufficient quality to be judged "fit for purpose" in terms of the mission's stated performance.
Ideal	Ancillary data used in product generation, meets the Excellent criteria, and are traceable to SI where appropriate.

3.3 Product Generation

The Product Generation section covers the processing steps undertaken to produce the data product. This starts with an assessment of the application of 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.

3.3.1 Calibration Algorithm

The applied L1 calibration algorithm, or measurement function, 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 and scene types (e.g. land, ocean, etc.). What this requires is specific to the sensor-domain and will require a degree of expert judgement. This should be based on the same reasoning applied to the pre-launch and in-flight calibration assessment and reviewed based on the ATBD.

Table 3-10 shows how the calibration algorithm is graded within the assessment framework.

Table 3-10 – Product Generation > Calibration Algorithm – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Calibration algorithm not documented.
Basic	Calibration algorithm somewhat documented. Calibration algorithm is too simple to be judged "fit for purpose" in terms of the mission's stated performance.
Good	Calibration algorithm documented. Reasonable retrieval algorithm used, judged "fit for purpose" in terms of the mission's stated performance for most expected use cases.
Excellent	Calibration algorithm documented. The calibration applied is considered "fit for purpose" in terms of the mission's stated performance for all expected use cases.
Ideal	Calibration algorithm well-documented. State-of-the-art calibration algorithm applied and considered "fit for purpose" in terms of the mission's stated performance.

3.3.2 Geometric Processing

A number of different geometric processing methodology may be applied to remote sensing data depending on the domain and application of the data product. The applied geometric processing should be of a sufficient quality that is "fit for purpose" within the context of the mission's stated performance for all mission products. Again, this constitutes a technical review of the ATBD from the data provider.

Table 3-11 shows how geometric processing is graded.

Table 3-11 – Product Generation > Geometric Processing – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Geometric processing not fully documented.
Basic	Geometric processing documented. Missing all or part of the calibration parameters. Calibration algorithm too simple to be judged "fit for purpose" in terms of the mission's stated performance. Confidence in the calibration quality is minimal.
Good	Geometric processing documented. Missing part of the input calibration parameters. Reasonable retrieval algorithm used. Confidence in the calibration quality is considered sufficient.
Excellent	Geometric processing documented. All input calibration parameters exist. Methodology used is considered "fit for purpose" in terms of the mission's stated performance for all expected use cases. Quality flags indicate good geometric accuracy with less than 5% exceptional.
Ideal	Geometric processing well-documented. State-of-the-art methodology used, easily "fit for purpose" in terms of the mission's stated performance. Quality flags indicate excellent geometric accuracy.

3.3.3 Retrieval Algorithm – Level 2 Only

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 3-12 shows how the assessment framework grades the retrieval algorithm used to generate L2 products.

Table 3-12 – Product Generation > Retrieval Algorithm – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Retrieval algorithm not documented.
Basic	Retrieval algorithm somewhat documented. Retrieval algorithm too simple to be judged “fit for purpose” in terms of the mission’s stated performance.
Good	Retrieval algorithm documented. Reasonable retrieval algorithm 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.
Excellent	Retrieval algorithm documented. Retrieval algorithm “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.
Ideal	Retrieval algorithm documented. State-of-the-art retrieval “fit for purpose” in terms of the mission’s stated performance, full uncertainty budget derived and validated.

3.3.4 Mission Specific Processing

Additional processing steps are separate to the main sensor calibration or retrieval processing. These may include processes like 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.

In the case of additional processes where the measurement data themselves are transformed in some manner, such as orthorectification, the uncertainties from the measurement data must be propagated, as well as introducing appropriate additional uncertainty components caused by the processing itself. This is required for the uncertainties to remain meaningful.

Each additional processing step should be separately assessed based on the criteria described in Table 3-13, and then a combined score determined.

Table 3-13 – Product Generation > Mission Specific 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. Additional processing steps not considered fit for stated purpose.
Good	Additional processing steps documented. All significant additional processing steps are fit for stated purpose.
Excellent	Additional processing steps documented. All additional processes steps considered fit for stated purpose.
Ideal	All additional processing steps are fully documented and considered state-of-the-art.

4. DETAILED VALIDATION

In this section we provide guidelines for the *Detailed Validation* assessment. The overall goal here is to verify that the mission performance is consistent with the sensor stated performance.

The detailed validation assessment is broadly divided into radiometric, spectral, and geometric validation activities. Within these three sections are paired sub-sections describing each of the assessed performance metrics, each of which are evaluated both in terms of the quality of the validation method used and the validation results compliance. The results are reported as part of the *Detailed Validation Cal/Val Maturity Matrix* (

Figure 4) and are then summarised across all performance metrics in the *Validation Summary*. This *Validation Summary* is the same summary presented as a column in the *Summary Cal/Val Maturity Matrix* shown in

Figure 1.

The remainder of this section includes:

- The criteria for grading the quality of the validation method used and validation results compliance is given in Section 4.1.
- The Radiometric, Spectral and Geometric performance metrics to be assessed are described in Section 4.2.
- Finally, in Section 0 the approach for synthesising the results of the *Detailed Validation* into the *Validation Summary* is described.

Validation Summary		Detailed Validation			
Radiometric Validation Method	←	Radiometric	Absolute Calibration Method	Signal to Noise Method	Temporal Stability Method
Radiometric Validation Results Compliance	←		Absolute Calibration Results Compliance	Signal to Noise Results Compliance	Temporal Stability Results Compliance
Spectral Validation Method	←	Spectral	Absolute Calibration Method	Temporal Stability Method	
Spectral Validation Method	←		Absolute Calibration Results Compliance	Temporal Stability Results Compliance	
Geometric Validation Method	←	Geometric	Absolute Positional Accuracy Method	Temporal Stability Method	Line of Sight Method
Geometric Validation Results Compliance	←		Absolute Positional Accuracy Results Compliance	Temporal Stability Results Compliance	Line of Sight Results Compliance

Figure 4 – Detailed Validation Cal/Val Maturity Matrix and Validation Summary

4.1 Detailed Validation Grading Criteria

This section describes how, in generic terms, the criteria for grading the quality of the Validation Method and Validation Results Compliance subsections the Radiometric and Geometric performance metrics.

4.1.1 Validation Method

Generally, satellite validation attempts to demonstrate the compliance of data products with respect to some claimed performance level (e.g., documented specifications) by comparison of the product data with independent reference data. A metrologically-rigorous validation of measurements goes a step further, attempting to verify both the satellite measurements and their associated uncertainties. Validated uncertainties provide evidence of the credibility of the uncertainty estimate given. Commonly used metrics such as the statistical spread of differences may be used to estimate the uncertainty, however this often may not provide a realistic estimate of the actual uncertainty.

A rigorous validation must compare mission data products with independent reference data that are fully representative of the satellite measurements being validated (e.g. point to pixel scaling considerations), over the full extent of measurements the satellite may make (e.g. biomes, dynamic range, seasonal variation). This may require the use of a variety of different reference datasets to cover different observation conditions.

In the same way, these guidelines describe how to assess the quality of satellite mission data. Similar considerations must be made for the quality of reference data used to validate the satellite mission data. The highest quality validation reference data provide uncertainty-assessed validation reference data traceable to SI, and come from activities, such as the ESA Fiducial Reference Measurement (FRM) projects (e.g. (Fox, 2019; Vendt, 2020)). .

Table 4-1 shows how the validation methods are graded. The specific interpretation of these criteria in the quality assessment of a particular validation activity depends on a number of factors, for example the particular method used or the sensor target performance, therefore some level of expert judgement may be required when determining the grading. A review of potential validation methodologies is provided in **Error! Reference source not found.** for measurement validation and **Error! Reference source not found.** for geometric validation, which is intended to act as the basis for such assessment.

Table 4-1 – Validation > Validation Method – Assessment Criteria

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Methodology is simple comparison, covering a limited range of satellite measurements. Uncertainty information not available for reference data.
Good	Methodology covers a range of satellite measurements that represents typical use cases, using representative reference measurements. Uncertainty information not available for reference data.
Excellent	Methodology assesses satellite measurements and reference data with respect to their characterised uncertainties. Reference measurements are assessed to be well representative of the satellite measurements.
Ideal	Methodology assesses satellite measurements and reference data with respect to their error-covariance and attempts to validate those uncertainties. Reference measurements independently assessed to be fully representative of the satellite measurements.

4.1.2 Validation Results Compliance

This section assesses the actual results of the validation activities themselves. In the best case these will show both validated satellite measurements and their associated uncertainties and will have been obtained by a group independent of the satellite data provider.

The results should be documented in a Validation report from a user community, see the QA4ECV guidance for expected content (Scanlon, 2017d).

Grading for this subsection is based on the compliance of the validation results with the performance claimed by the data provider and with the possibly more stringent standards from the user community..

Table 4-2 shows how the validation results are graded within the assessment framework.

Table 4-2 – Validation > Validation Results – Assessment Criteria

Grade	Criteria

Not Assessed	Assessment outside the scope of study.
Not Assessable	No validation activity performed.
Basic	Claimed mission performance shows some agreement with validation results.
Good	Claimed mission performance shows good agreement with validation results.
Excellent	Claimed mission performance shows excellent agreement with validation results. Analysis performed independently of the satellite mission owner.
Ideal	Claimed mission performance shows excellent agreement with validation results, measurement uncertainties also validated. Analysis performed independently of the satellite mission owner.

4.2 Performance Metrics

This section describes the performance metrics that define the *Detailed Validation Cal/Val Maturity Matrix* structure. This is divided into the Radiometric, Spectral, and Geometric sections.

4.2.1 *Radiometric Validation*

Different classes of atmospheric satellite sensors are aimed at a broad range of applications using a variety of observational techniques, and so are subject to various design and performance trade-offs in order to meet their goals. The domain spans from instruments designed for imagery analysis (detections of specific features, i.e., clouds) to chemical composition (spectrometers, radiometers) with different geometrical acquisitions strategies (nadir viewing, limb scanning, profiling with occultations) to long term quantitative analysis (e.g., climate observation). The performance characteristics of these different types of sensors may in general be very different. Here we assess their compliance with their claimed performance, which in absolute terms is mission/application specific.

Particular performance metrics are defined to characterise different aspects of measurement radiometric quality, which may be of different relative importance depending on the intended application. For data products intended for quantitative analysis the validation of radiometric calibration is clearly necessary to provide credibility to the measurements. For temporal analysis calibration stability of the data record must be demonstrated. Finally, low measurement noise performance may be important for data where instantaneous images are analysed, but less import in long term data where it will tend to average out.

For the *Radiometric Validation* section, the following metrics are used to validate atmospheric satellite sensors:

- Absolute calibration
- Signal-to-noise
- Temporal stability

For a discussion of the various in-flight methods used to perform of radiometric calibration and validation see APPENDIX A.

4.2.1.1 Absolute Calibration

The potentially SI-traceable calibration of atmospheric satellite sensors established in the laboratory pre-flight is not preserved on-orbit, due to the rough conditions of launch and subsequent instrument degradation, exacerbated by the space environment. On-board calibration systems are not always available (in few cases not possible, i.e., radio-occultation instruments), and while providing the means to maintain instrument performance to some extent, are unable to re-establish SI-traceability as they are also subject to similar degradation. Thus, the need for external validation of satellite absolute calibration performance once the instrument is on-orbit.

Many approaches have been developed to validate satellite absolute calibration performance, including comparison with other sensors, comparison with on-ground measurements, and comparison with simulated observations. APPENDIX A details these methods in more detail.

4.2.1.2 Signal-to-Noise

Measurement noise, occurring in the satellite sensor detector and processing chain, provides a fundamental limit to the achievable quality of a given instantaneous observations. In the instrument uncertainty budget, noise will generally be the key contributor to the random component of uncertainty. The signal-to-noise ratio (SNR) is a common measure used to quantify noise performance of a measurement system.

SNR is usually part of the pre-launch instrument characterisation campaign. This performance may then be routinely validated on-orbit in a number of ways, all of which look at the statistical spread of observations for repeated measurements, for example shuttered acquisitions or pseudo homogenous Earth scenes. A full analysis should look at SNR should evaluate how it varies across the detector and as a function of detector temperature. The evolution of SNR overtime may be monitored with statistics.

4.2.1.3 Temporal Stability

As described in Section **Error! Reference source not found.**, validation of instrument absolute on-orbit calibration performance is required to monitor the relative evolution of sensor performance over time. On-board optical calibration systems may only partially compensate for instrument degradation, which leads to declining performance, data record instability, and increasing inconsistency with other sensors.

Comparison with other satellite sensors or ground measurements, and various vicarious calibration methods allows for the identification and correction of such performance drifts. APPENDIX A details these methods in more detail.

4.2.2 Spectral Validation

Spectral calibration is performed using standard measurements from the atmosphere. Particular spectral lines are retrieved in the observed spectrum and the known values of their wavenumbers are used to establish the assignment of the wavenumber to the index of spectral data points. The spectral calibration is used for the wavenumber assignment of all subsequent scene and gain measurements.

For the *Spectral Validation* section, the following metrics are used to validate atmospheric satellite sensors:

- Absolute calibration
- Temporal stability

These are each described in turn below. For a discussion of the various in-flight methods of geometric assessment, see APPENDIX B.

4.2.3 Geometric Validation

Geometric performance assessment of atmospheric remote sensing data typically is not a key factor since there are different typologies of sensors that are focused on different aspects to verified. It’s clear that for optical/atmospheric sensors, geometric validation is important, for profiling instruments the geolocation of profiles is an important parameter, especially in case of coincident measurements with ground based instruments. The geolocation accuracy on the Earth’s surface, or absolute positional accuracy (APA) is the main parameter. In geometric assessment, it is also important to consider temporal stability and global consistency in all aspects.

For geometric assessment, first it is important whether the data are provided in a swath or gridded format. Swath data products have not been resampled and have the original time-tagged observations as sampled by the instrument. Gridded products typically contain data that have been resampled to a fixed Earth grid with a fixed pixel interval and in some case (for the imager) may be orthorectified to correct for terrain distortions.

More frequent are gridded products: they are typically provided as scenes (or tiles) and may be accompanied by additional information such as acquisition time and solar and viewing geometry. This information may be provided as single values for the entire scene or multiple values within a scene, typically at a resolution coarser than the product resolution.

For *Geometric Validation* of satellite imagery, we define the following metrics used for evaluation:

- Absolute positional accuracy (APA)
- Line of sight (LOS)
- Temporal Stability

These are each described in turn below. For a discussion of the various in-flight methods of geometric assessment, see APPENDIX C.

4.3 Validation Summary

The *Validation Summary* provides a synthesis of the per performance metric assessments provided in the *Detailed Validation Cal/Val Maturity Matrix* (

Figure 4). It is also presented as part of the *Summary Cal/Val Maturity Matrix*.

Each row in the *Detailed Validation Cal/Val Maturity Matrix* is represented by one cell in the *Validation Summary* column. Thus, there are four summary cells in total – *Radiometric Validation Method*, *Radiometric Validation Results Compliance*, *Geometric Validation Method* and *Geometric Validation Results Compliance*.

The grade for each of these summary cells represents a combination of the grades of the contributing cells. The approach is to effectively average the grades of the contributing cells, where each grade is valued as follows: Basic is 1, Good is 2, Excellent is 3, and Ideal is 4.

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 atmospheric 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, Chander *et al.*, 2013 and Tansock, et al. 2015.

Radiometric calibration is the process of assigning absolute values in radiance units to the intensity with a specified accuracy and typically uses the approach described by Revercomb et al. (1988). It is performed using typically at least two known radiation sources (a hot source and a cold source). The hot source is an internal calibration blackbody, while the deep space serves as cold source.

The following sections of this appendix each describe a commonly used calibration and validation method, by defining the following:

- **Description** – general outline of method, with appropriated references.
- **Scope of Representativeness** – The types of observations the method can be used to calibrate/validate.
- **Quality** – best uncertainty achievable with this method, according to literature.
- **Radiometric Calibration/Validation Metric** – metrics from the *Detailed Validation* maturity matrix the method can be used for.

A.1 Deep Space

Description

Deep space measurements followed by scene measurements to correct the scene for self-emission of the instrument. Deep space measurements are done frequently to account for changing self-emission of the instrument due to temperature variations along the orbit.

Scope of Representativeness

Scenes are dark. For use in VIS/NIR/TIR.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method depends by the instrument and frequency range and resolution.

Radiometric Calibration/Validation Methods

Absolute calibration.

A.2 Blackbody and Lamps

Description

Blackbody (BB) measurements followed by an equivalent number of deep space measurements to calculate the radiometric gain function.

Scope of Representativeness

Scenes are defined by the characteristics of lamps and BB on board, depending by the sensor. For use in UV/VIS/NIR/TIR.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method depends by the instrument and frequency range and resolution.

Radiometric Calibration/Validation Methods

Absolute calibration.

A.3 Ocean Targets – Night Observations

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.

Scope of Representativeness

Scenes are dark. For use in the visible and NIR.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method are currently not available and depends by the sensitivity of the sensor.

Radiometric Calibration/Validation Methods

Absolute calibration.

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 measurement campaigns.

- Permanently instrumented, autonomous sites or networks of sites.

For the radiometric calibration the CEOS RadCalNet (Radiometric Calibration Network) it the most notable measurement network [RD-13]. It consists of four instrumented sites located in the USA, France, China, and Namibia. Top-of-atmosphere reflectance data with associated uncertainties are available at 10 nm intervals over the 400 nm to 1000 nm spectral range at 30 min intervals for a nadir-viewing geometry. It is in wide use by space agencies and commercial mission vendors for both L1 calibration and validation.

Several measurement networks are operational to validate derived geophysical and physical/chemical products. For example, big networks (including AERONET (Holben *et al.*, 1998), TCCON, IGRA, NDACC, GAW).

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.

Radiometric Calibration/Validation Methods

Absolute calibration and stability monitoring.

A.5 Simultaneous Overpasses

Description

This method involves calibrating a given satellite sensor using another reference satellite sensor. This is accomplished by locating events called simultaneous overpasses, 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 observations between the pair of satellites, or calibration campaigns with sounders and.

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.

Radiometric Calibration/Validation Methods

Absolute calibration.

A.6 Lunar Observations

Description

The Moon provides a photometrically stable source for calibration of earth observation sensors, within the range of the Earth radiometric levels and is free from atmospheric interference. In order to utilize the moon as a radiometric calibration target its disk integrated irradiance, provided by a lunar model, is compared to radiometric measurements taken by the observing instrument to be calibrated (Stone *et al.*, 2020).

The USGS robotic Lunar Observatory (ROLO) (Kieffer and Wildey, 1996) has developed one such lunar irradiance model (Kieffer and Stone, 2005), which has been an invaluable tool for relative radiometric monitoring. Recent efforts are working towards the development of an SI traceable Lunar irradiance model, such as LIME (Lunar Irradiance Model of ESA), to enable the use of the Moon for traceable absolute radiometric calibration.

Scope of Representativeness

Typically, visible to shortwave infrared

Quality

The ROLO model can predict variations in lunar irradiance to a precision of <1%, with an uncertainty of 5 – 10% (Stone and Kieffer, 2004). Recent lunar observations contributing to models are providing full traceability and rigorous uncertainty analysis. The LIME model targets a typical uncertainty of approximately 2%. Through the WMO's GSICS (Global Space-based Inter-Calibration System) and collaborations between ESA and NASA, inter-comparisons of models are taking place to ensure quality and consistency of lunar models and to test their uncertainties.

Radiometric Calibration/Validation Methods

Relative radiometric calibration. Absolute calibration with new models in development.

APPENDIX B IN-FLIGHT SPECTRAL CALIBRATION AND VALIDATION METHODS FOR ATMOSPHERIC SENSORS

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

Spectral calibration is performed using standard measurements from the atmosphere. Spectral lines are retrieved in the observed spectrum and the known values of their wavenumbers are used to establish the assignment of the wavenumber to the index of spectral data points. The spectral calibration is used for the wavenumber assignment of all subsequent scene and gain measurements. The spectral calibration includes the spectral shift and the appropriate spectral lines are identified and the value of their wavenumber is available for ground processing.

As in Appendix A, the following sections each describe a commonly used calibration and validation method, by defining the following:

- **Description** – general outline of method, with appropriated references.
- **Scope of Representativeness** – The types of observations the method can be used to calibrate/validate.
- **Quality** – best uncertainty achievable with this method, according to literature.
- **Radiometric Calibration/Validation Metric** – metrics from the *Detailed Validation* maturity matrix the method can be used for.

B.1 Deep Space

Description

Deep space measurements followed by scene measurements to correct the scene for self-emission of the instrument. Deep space measurements are done frequently (once every four elevation scans) in order to account for changing self-emission of the instrument due to temperature variations along the orbit.

Scope of Representativeness

Scenes are dark. For use in VIS/NIR/TIR.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method depends by the instrument and frequency range and resolution.

Radiometric Calibration/Validation Methods

Absolute calibration.

B.2 Solar Irradiance

Description

The solar spectrum is measured with a mirror diffuser combination to obtain a solar reference spectrum.

Scope of Representativeness

Scenes are spectral referenced. For use in VIS/NIR/TIR.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method depends by the instrument and frequency range and resolution.

Radiometric Calibration/Validation Methods

Absolute calibration.

B.3 Blackbody (BB) and Lamps

Description

Blackbody (BB) measurements followed by an equivalent number of deep space measurements to calculate the radiometric gain function. The gain calibration is done once per week.

Scope of Representativeness

Scenes are defined by the characteristics of lamps and BB on board, depending by the sensor. For use in UV/VIS/NIR/TIR.

Quality

Fully metrologically rigorous traceability and uncertainty analysis for this method depends by the instrument and frequency range and resolution.

Radiometric Calibration/Validation Methods

Absolute calibration.

APPENDIX C IN-FLIGHT GEOMETRIC CALIBRATION AND VALIDATION METHODS FOR ATMOSPHERIC SENSORS

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

As in Appendix A, the following sections each describe a commonly used calibration and validation method, by defining the following:

- **Description** – general outline of method, with appropriated references.
- **Scope of Representativeness** – The types of observations the method can be used to calibrate/validate.
- **Quality** – best uncertainty achievable with this method, according to literature.
- **Radiometric Calibration/Validation Metric** – metrics from the *Detailed Validation* maturity matrix the method can be used for.

C.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 high, within 10% of a pixel size, that is 30 cm for data at resolution of 3 m, 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 longwave 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.

C.2 Line of sight (LOS) Calibration

Description

LOS calibration is required for the in-flight determination of the line of sight pointing biases and variations. It is based on the position and tracking of stars moving in the field of view and is strictly correlated with the attitude and positioning.

Scope of Representativeness

Visible to longwave infrared, limb scanning instruments.

Quality

Full traceability and uncertainty quantification for this method requires the methodology and instrumentation used to acquire stars, with high attitude control.

Geometric Calibration/Validation Methods

Absolute geometric accuracy.