Earth Observation Mission Quality Assessment Framework - Optical Guidelines

Author: **_Samuel Hunt_**
NPL, Task Technical Expert

Approval: **_Nigel Fox_**
NPL, Task Lead

Accepted: **_Clément Albinet_**
ESA, Project Technical Officer

EDAP.REP.002 _Issue: 2.1_ _31 October 2021_

AMENDMENT RECORD SHEET

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

ISSUE	DATE	REASON
1.0	16/10/19	First Issue of ESA EDAP project guidelines
2.1	8/3/21	Reworking of EDAP guidelines following ESA/NASA collaboration
2.1	31/10/21	Update for EDAP project close

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

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 optical 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 optical 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 optical 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 4 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
Aeronet	Aerosol Robotic Network
ATBD	Algorithm Theoretical Basis Document
BBR	Band-to-Band Registration
BELMANIP	Benchmark Land Multisite Analysis and Intercomparison of Products
BSRN	Baseline Surface Radiation Network
CF	Climate & Forecast (Metadata Convention)
CEOS	Committee on Earth Observation Satellites
DCC	Deep convective cloud
DDR	Detector-to-Detector Registration
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 Infrared 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
GSCIS	Global Space-based Inter-Calibration System
HCS	Horizontal Cell Size
HR	High Resolution (spatial resolution between 5 and 30 m)
HSI	Horizontal Sampling Interval
IVOS	Infrared and Visible Optical Sensors
L1	Level 1
L2	Level 2
LIME	Lunar Irradiance Model of ESA
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
NASA	National Aeronautics and Space Administration, USA
NEON	National Ecological Observatory Network
NIST	National Institute of Standards and Technology, USA

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
ROLO	Robotic Lunar Observatory
SAR	Synthetic Aperture Radar
SI	Système International (International System of Units)
SNR	Signal-to-Noise Ratio
SSR	Sensor Spatial Response
TOA	Top-of-atmosphere
VHR	Very High Resolution (spatial resolution finer than 5 m)
WMO	World Meteorological Organisation

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)*, 2001, 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 4, 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	Radiometric 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
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.

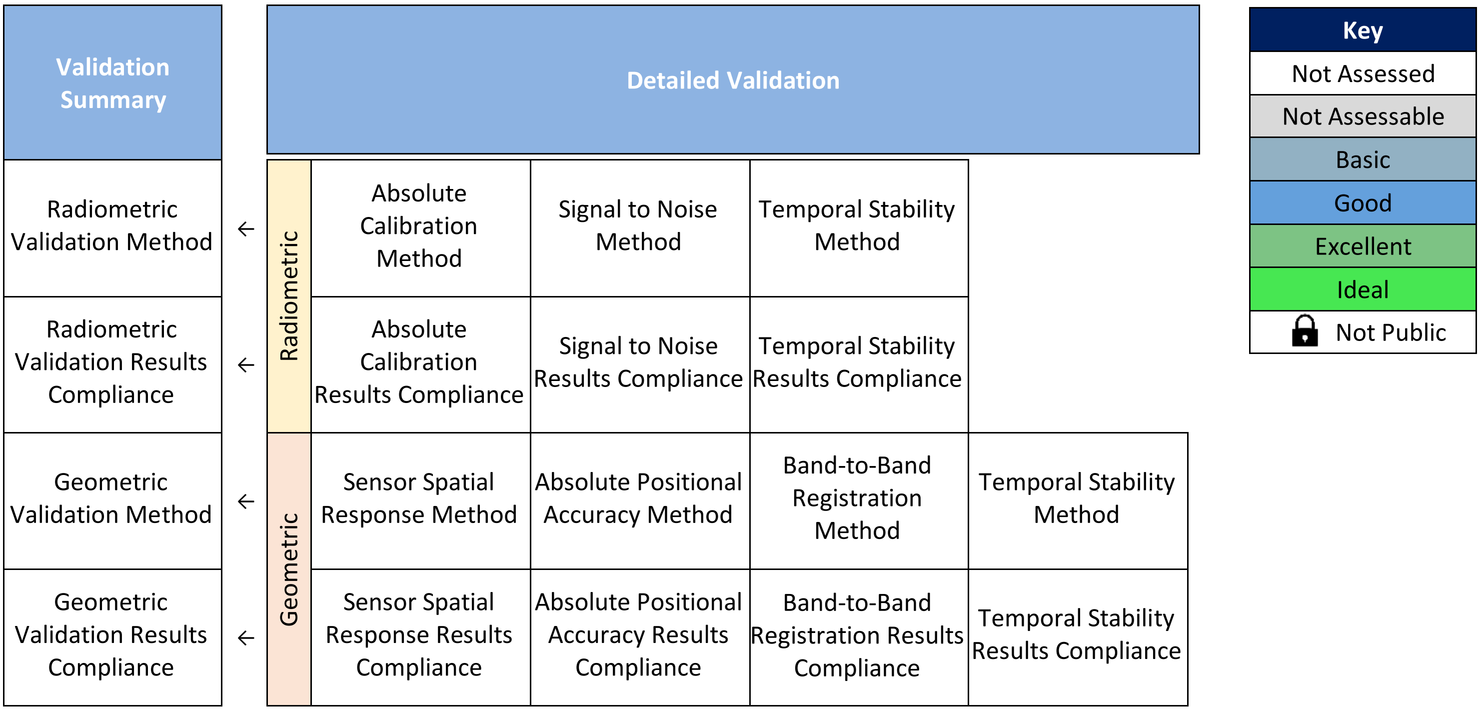

Figure 2 – Validation Cal/Val Maturity Matrix for the optical domain, which includes the Validation Summary column from the Summary Cal/Val Maturity Matrix

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.

2.4 Considerations for the optical domain

Since the optical domain 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 (e.g. multi-channel, hyperspectral) and spatial resolution. The spatial resolution of a sensor may be defined as low resolution (LR; spatial resolution coarse than 300 m), medium resolution (MR; 30 to 300 m), high resolution (HR; 5 to 30 m) and very high resolution (VHR; finer 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”.

3. DATA PROVIDER DOCUMENTATION REVIEW AND VALIDATION SUMMARY

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	Radiometric 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 defined below.

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, with specific requirements added for optical sensors:

- Product name
- Sensor Name
- Sensor Type

    Describe sensor design type, e.g., multi-channel, hyperspectral, interferometer etc., and spectral domains, e.g. visible (VIS), near infrared (NIR), shortwave infrared (SWIR), 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

    Defined for optical sensors as:
    - L0 – uncalibrated instrument counts
    - L1 – time-tagged, geo-located, calibrated top-of-atmosphere radiance, reflectance, or brightness temperature
    - L2a – surface radiance or reflectance
- Measured quantity name

    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.

    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

    Pixel spatial sampling, include if viewing nadir or tilted off-axis. Categorise as either LR, MR, HR or VHR. Wide swath sensors should define the nadir and edge-of-swath pixel size to indicate scan angle effect on pixel size.
- 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 section is about how readily the data are available to those who wish to use them. Does the data set follow the FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles for scientific data management and stewardship (Wilkinson *et al.*, 2016), that provide valuable principles for all applications. These principles 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 provision of the above information relates to the grade a data product 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 toward FAIR principles.
Excellent	The data set meets many of the FAIR principles, has an associated data management plan. The data are available through an easy-to-access licence.
Ideal	The data set fully meets the FAIR principles, has an associated data management plan. The data are available through an easy-to-access 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) (CEOS LSI, 2020) 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 (Eaton *et al.*, 2020), or, for data from the European Union, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive (INSPIRE Drafting Team Metadata and European Commission Joint Research Centre, 2013).
- 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 include 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 that for a given mission, a combination of articles, publications, webpages and presentations provides a similar set of information in place of these documents. To achieve the highest grades this information should be presented as formal documents, and users should not be expected to search for this information.

The QA4ECV project provides guidance for the expected contents of these documents (Scanlon, 2017b, 2017a). The user guide should provide general information on the product, including:

- Description of available products (as specified in Section 3.1.1).
- Description of how to read the products, i.e. product format and metadata.
- Contact information.
- References

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

- Basic overview of the instrument design concept (not necessarily proprietary details), including viewing geometry.
- Description of the radiometric calibration processing, including the sensor measurement function.
- Description of the geometric processing.
- Description of the geophysical retrieval processing, if required
- Description of any other mission specific processing, as necessary.
- Description of the uncertainty analysis performed on this processing.
- Details of assumptions and limitations of the algorithm.

Note that the PUG and ATDB will likely be the source of much of the information required for the other sub-sections of the assessment. In particular, the technical review of the fitness for purpose of the processing algorithms is undertaken in the Product Generation section of the assessment (described in Section 3.3).

Table 3-4 describes how to grade the user documentation of a product within the assessment framework.

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 Radiometric Calibration & Characterisation

The sensor radiometric calibration and characterisation, 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.

Characterisation and calibration should be based on the sensor measurement function, which must include all relevant parameters influencing the sensor measurement. Parameters influencing optical systems 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.
- *Geometric* – covered in 3.2.2.

For a given instrument with a stated performance and application area, the calibration/characterisation activities required should be determined based on its likely performance constraints, e.g. what is the expected performance impact of not correcting for a given effect? The mission assessor may need to apply their expert judgement to review this in their assessment.

The activity, from pre-launch and post-launch commissioning and monitoring, should be documented and available for assessment. This should include the calibration traceability, preferably to SI, and an uncertainty budget, with evidence of the stated performance.

Pre-Launch

For a thorough overview of pre-launch calibration and characterisation of optical sensors, see the U.S. National Institute of Standards and Technology (NIST) best practice guide (Tansock *et al.*, 2015).

Pre-launch calibration can be divided into three key stages, which allow for a full understanding of instrument behaviour (Datla *et al.*, 2011):

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.

Best practice dictates that the pre-flight tests are performed in the same environment as the sensor, i.e. thermally, and under vacuum.

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

Post-Launch

As in the pre-launch case, the post-launch radiometric calibration and characterisation activity should encompass a given sensor's behaviour to an extent that is "fit for purpose" within the context of the mission's stated performance, though the extent to which an instrument can be characterised in-flight is limited compared to a ground campaign. Post-launch calibration and characterisation is divided into two key activities – an initial commissioning phase, followed by on-going monitoring of performance. The frequency of this on-going monitoring of performance on-orbit is dependent upon several factors that need to be considered for each sensor system configuration (Tansock *et al.*, 2015).

For a review of various post-launch radiometric calibration methods see 0. 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). 0 should allow the assessor to judge the extent to which a given on-orbit calibration method can achieve a stated performance. Unfortunately, for some common post-launch calibration methods, rigorous uncertainty analysis and traceability are not always available. For this reason, it is important to use high-quality reference data methods, e.g., where metrological best practices are followed – for example, data from ESA's Fiducial Reference Measurement (FRM) campaigns (e.g. FRM4STS (Fox, 2019) and FRM4SOC (Vendt, 2020) amongst others) or the RadCalNet (Radiometric Calibration Network) sensor network (Bouvet *et al.*, 2019).

Note that though different methods may primarily be suited for either absolute on-orbit calibration or validation/monitoring activity, some are suitable for both. The post-launch calibration and the post-launch validation should be performed independently.

Table 3-5 shows how sensor radiometric calibration and characterisation are graded within the assessment framework.

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

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Pre-flight and post-launch radiometric calibration & characterisation activities are not documented or information not available.
Basic	Pre-flight and post-launch radiometric 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 radiometric 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 radiometric 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

Similar to radiometric calibration and characterisation, geometric calibration and characterisation, both pre-flight and on-orbit, should encompass a sensor's behaviour to an extent and sufficient quality that is "fit for purpose" within the context of the mission's stated performance.

Pre-Launch

Pre-launch engineering, manufacturing, testing and analysis must be performed to the standards needed to build an instrument that has sufficiently stable geometry (including focal length) to produce data with geometric accuracies required for particular scientific research and applications. This can be an issue with small/cube satellites because their optics may not be thermally stable. Because of their small mass, and perhaps a lack of sufficient on-board heating/cooling, it becomes more difficult to maintain the thermal stability needed to maintain consistent spatial resolution, accurate pointing knowledge and band-to-band alignment.

The optical sensor pre-flight calibration and characterisation for geometric performance may be found in (Wolfe *et al.*, 2013; Knight and Kvaran, 2014; Lin and Wolfe, 2016). This includes effects such as spatial resolution, MTF, band-to-band co-registration, alignment and pointing. Additional components of the satellite that influence the geometric processing should also be characterised, such as guidance, navigation and control, star trackers or attitude control systems.

Post-Launch

With any satellite sensor, because of potential long-term changes in sensor characteristics, it is necessary to monitor the instrument's performance over the entire mission to ensure that any changes in performance over time are understood. A long-term trending is performed after early on-orbit checkout and an initial intensive calibration and validation campaign (Storey, Choate and Lee, 2014; Dechoz *et al.*, 2015; Lin and Wolfe, 2016).

The variables that impact the geometric processing accuracy should be monitored to ensure a sufficient quality that is "fit for purpose" within the context of the mission's stated performance. The monitored variables include the status of the star tracker, the accuracy of the satellite attitude, the gyro data, and the quality of the Kalman Filter results (if there is any) etc. A set of quantitative criteria for these variables should be created to flag the quality of the geometric accuracy. A small set of ground control points (GCPs) could be used to assess the geometric accuracy together with the production. This is to provide a preliminary accuracy assessment. A full geometric accuracy assessment will be performed during post-production assessment (Section 4.2.2).

For further discussion of the various in-flight geometric calibration and characterisation 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. It is recognized that sensors with higher resolution images have more cumbersome work to perform geometric characterisation, calibration and validation (Storey, Choate and Lee, 2014). In general, images of higher resolution have higher geolocation accuracy, i.e., smaller errors in linear ground distance from "truth," that has to be established in a more fundamental way (Storey and Choate, 2000). Once the images achieve high geolocation accuracy, they could be used as "truth" for geometric calibration of imagery at the same or lower resolution (Wolfe and Nishihama, 2011; Wolfe *et al.*, 2013; Storey, Choate and Lee, 2014; Dechoz *et al.*, 2015).

Table 3-6 shows how to grade geometric calibration and characterisation within the assessment framework.

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) (JGCM, 2012) as the

"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."

Traceability is therefore a key aspect of achieving reliable, defensible measurements. In this definition, an important part of measurement traceability is that it is well documented. Various diagrammatic approaches have been developed to present the traceability chains for EO data products (e.g. the QA4ECV guidance includes a traceability chain drawing tool (Scanlon, 2017c)). A traceability diagram should be included in the documentation for every EO mission. Guidance for a detailed measurement function centred “uncertainty tree diagram”, more suitable for Level 1 (and some Level 2) processing can be found in (Mittaz, Merchant and Woolliams, 2019) and should be the aspiration for missions in the future.

It is important that traceability documentation remains up to date. It is common that aspects of a sensor’s calibration may be modified or completely changed over the course of a mission, which changes the sensor’s traceability chain, and such updates should be documented.

Table 3-7 shows how the metrological traceability is graded.

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) (JCGM, 2008). 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 (Gorroño *et al.*, 2017). There have also been some initiatives, like the previously mentioned FIDUCEO project, that have applied metrology to historical sensor data records (e.g. Taylor *et al.*, 2019).

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 Radiometric 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.). The mission assessor should apply their expert judgement to determine for a given instrument (e.g., multiband, hyperspectral), if the form of the measurement function applied is appropriate (i.e., all the necessary corrections are applied).

This should be based on the same reasoning applied to the pre- and post-launch calibration assessment and review based on the ATBD.

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

Table 3-10 – Product Generation > Radiometric 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

Several different geometric processing methodologies may be applied to optical imagery data depending on the application of the data product. These may include selection of the Earth model (National Imagery and Mapping Agency, 2000), terrain surface model (Wolfe *et al.*, 2013), correction to ground control points (GCPs), resampling or orthorectification amongst others. Processing may vary between products for a given mission, for example, based on number of available GCPs or geolocation references (Gutman *et al.*, 2013; Storey, Choate and Lee, 2014; Dechoz *et al.*, 2015).

The 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 3-10 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 from 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.

Additional processing steps performed on optical mission products may include the following:

- Cloud masking
- Pan sharpening

The algorithm for these additional processing steps should be documented, including assumptions made and relevant process specific details.

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 and geometric validation activities. Within these two 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 1.4.104.3

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 and Geometric performance metrics to be assessed are described in Section 0.
- Finally, in Section 4.3 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	←	Radiometric	Absolute Calibration Results Compliance	Signal to Noise Results Compliance	Temporal Stability Results Compliance	
Geometric Validation Method	←	Geometric	Sensor Spatial Response Method	Absolute Positional Accuracy Method	Band-to-Band Registration Method	Temporal Stability Method
Geometric Validation Results Compliance	←	Geometric	Sensor Spatial Response Results Compliance	Absolute Positional Accuracy Results Compliance	Band-to-Band Registration Results Compliance	Temporal Stability 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 0 for measurement validation and APPENDIX B 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 and Geometric sections.

4.2.1 *Radiometric Validation*

Different classes of optical satellite sensors are aimed at a broad range of applications and are subject to different design and performance trade-offs to meet mission goals. The performance characteristics of different types of sensors may be very different. Here we assess sensor measurement compliance with performance specifications.

Performance metrics are defined to characterise different aspects of radiometric integrity, 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 analyses, 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 optical satellite sensors:

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

For a discussion of the various in-flight methods used to perform radiometric calibration and validation see 0.

4.2.1.1 Absolute Calibration

The potentially SI-traceable calibration of optical 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 optical calibration systems are not always available, and while providing the means to maintain instrument performance to some extent, they 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 is needed 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. 0 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 in a measurement system.

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

4.2.1.3 Temporal Stability

As described in Section 3.2.1, 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 and various vicarious calibration methods allows for the identification and correction of such performance drifts. 0 details these methods.

4.2.2 Geometric Validation

There are three main aspects of assessing geometric performance in remote sensing data: 1) instrument sensor spatial response (SSR); 2) geolocation accuracy on the Earth's surface, or absolute positional accuracy (APA); and 3) multispectral sensor band-to-band registration (BBR). In geometric assessment, it is also important to consider temporal stability and global consistency in all aspects.

For geometric assessment, 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 observations that have been resampled to a fixed Earth grid with a fixed pixel interval and may be orthorectified to correct for terrain distortions.

Swath products must be accompanied by additional information regarding geometry of the observations in the product, either within the product or as a separate geolocation product. This additional information usually includes time-tagged geodetic latitude and longitude of each observation (sample or pixel), and for many data sets, the terrain height. It may also include information such as the solar zenith and azimuth angles, quality flags, satellite position and its velocity and attitude, and the satellite zenith and azimuth angles. This data may be available for each observation or at a coarser resolution, e.g. at the scene centre. For multispectral instruments there may be additional information about relative alignment of the individual bands, such as the band-to-band offsets.

Gridded products 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:

- Sensor spatial response (SSR)
- Absolute positional accuracy (APA)
- Multispectral sensor band-to-band co-registration (BBR)

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

The *Geometric Validation* assessment combines geometric specification and the uncertainty criteria in one evaluation matrix for each metric. Details of the alternative reporting style for geometric validation for each of the metrics described above is detailed in APPENDIX C.

4.2.2.1 Sensor Spatial Response (SSR)

A sensor or detector spatial response is a function describing overall system response to a point impulse that is spatially located at every possible position. This spatial response function is called the system point spread function (PSF). A PSF is a spatial weighting function describing the responsivity of a detector to energy from a scene. A PSF may be constructed by two orthogonal line spread functions (LSFs), one in the along-track direction and another in the cross-track direction, for either a pushbroom, whiskbroom or frame sensor instrument. A PSF is usually tested and analysed pre-launch and verified on-orbit. From the PSF, we can determine parameters such as the field of view (FOV) at the full width at half maximum (FWHM), and the modulation transfer function (MTF). In general, we want the MTF to be at least 0.25 or greater at the Nyquist frequency (two times the horizontal spatial sampling interval or ground sample distance). Note that for gridded products, the MTF can be improved by gridding the data at a larger pixel size. For multispectral instruments, these measurements should be made separately for each spectral band. Also, the spatial response may vary by position within the focal plane, e.g. by detector, so measurements should be made to understand any detector-specific variation that may be present.

4.2.2.2 Absolute Positional Accuracy (APA)

As agency and commercial satellite sensors become more advanced and numerous, with many providing very high-resolution imagery, it is important to evaluate the positional accuracy of the products against the accuracy specifications and typical user needs.

Geolocation accuracy assessment typically involves evaluation of the positional accuracy of the data using ground truth with a known geolocation accuracy, typically ground control points (GCPs). For many applications, the geolocation accuracy should have a circular error at the 90th percentile (CE90) to within 0.5 of the product pixel size for gridded products, and within 0.5 of the ground sample distance for swath products. The GCPs should be as evenly distributed spatially as possible, to ensure consistency in the geolocation accuracy assessment globally. For sensors with numerous detectors acquiring data simultaneously, to ensure an unbiased assessment due to image distortion, GCPs should be evenly distributed over the entire detector array.

For swath data, the accompanying geolocation information in the geolocation product is used to compare the geolocated observations to the ground truth. Note, that for multi-spectral data, the geolocation accuracy may be assessed using a single band, but may also be done for individual bands, and so may be impacted by band-to-band registration.

Should the data in a single scene be used for object identification, for example, a geolocation error of a few pixels may not be significant, and thus further geolocation error correction may not be required for the application. However, should the data be used for time series analyses, these same geolocation errors will result in unusable data for this purpose. Relative geolocation errors could be reduced by aggregating the data to a larger pixel size.

4.2.2.3 Band-to-Band Registration (BBR)

For multispectral data sets, it is important for many applications that the individual bands are in alignment with one another. This is referred to as “band-to-band registration” (BBR).

For swath products, a band-pair BBR is a collection (in some statistical sense) of detector-to-detector registration (DDR), the co-registration of corresponding detectors between a pair of different spectral bands. The DDR is defined as the overlap of footprints of corresponding detectors for a pair of different spectral bands. The full representation of DDR should be its point spread function (PSF). We usually want the BBR performance to be > 80% of the footprint overlap of the corresponding band detectors, 99.73% of the time (3-sigma). Note that this measurement may be made individually in the along-track and cross-track directions by measuring the offset in each direction and then combining the offset with the LSF in each direction to calculate the footprint overlap. Also, the DDR may vary by position within the focal plane, e.g. by detector, so measurements should be made to understand any detector-specific variation that may be present.

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

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 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 top-of-atmosphere (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.

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 (Govaerts *et al.*, 2018).

Radiometric Calibration/Validation Metrics

Absolute calibration.

A.2 Pseudo-invariant Calibration Sites (PICS)

Description

Pseudo-invariant calibration sites (PICS) are 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 low 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.

The methodology developed by Lyapustin *et al.*, 2014, became a standard part of MODIS calibration protocol. This approach can be used to remove calibration trends among different sensors and allows for cross-calibration to a common reference. This is currently being applied to remove calibration trends and achieve cross-calibration among the DigitalGlobe constellation.

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 (Govaerts *et al.*, 2018).

Radiometric Calibration/Validation Methods

Temporal 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) 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, 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 (Govaerts *et al.*, 2018).

Radiometric Calibration/Validation Methods

Inter-band calibration and stability monitoring.

A.4 In situ Measurements

Description

Satellite sensors should be calibrated or validated against field measurements, at Level 1 and above. At Level 1, comparison can be made against field measurements that are propagated from bottom-of-atmosphere to top-of-atmosphere (TOA) with radiative transfer modelling (RTM). Field measurements may either be from:

- Field measurement campaigns. For example, the ESA Fiducial Reference Measurement (FRM) projects (e.g. FRM4STS (Fox, 2019) and FRM4SOC (Vendt, 2020), FRM4VEG amongst others).
- Permanently instrumented, autonomous sites or networks of sites. For example:
    - Radiometric Calibration Network (RadCalNet) (Bouvet *et al.*, 2019)
    - Aerosol Robotic Network (AERONET) (Holben *et al.*, 1998)
    - National Ecological Observatory Network (NEON) (Li *et al.*, 2021)
    - Baseline Surface Radiation Network (BSRN) (Driemel *et al.*, 2018)
    - Amongst many others.

For radiometric calibration, RadCalNet is the most notable measurement network. RadCalNet consists of four instrumented sites located in the USA, France, China, and Namibia. Top-of-atmosphere nadir-viewing reflectance data with associated uncertainties are available at 10 nm intervals over the 400 nm to 1000 nm spectral range at 30 min intervals. This network is used widely by space agencies and commercial mission vendors for both L1 calibration and validation.

Scope of Representativeness

Network dependant.

Quality

These measurements can have traceability chains and quantified uncertainties, though are not ubiquitous across the field. The aforementioned RadCalNet and FRM campaigns are designed to be metrologically rigorous and thus are recommended. For RadCalNet instrumented sites, typical achievable satellite sensor calibration uncertainty can be < 5 % (e.g. (Thome, Smith and Scott, 2001)).

Note that RadCalNet provides free data for 4 sites at nadir view in 30-minute intervals and at 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 observations 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 Gorroño et al. 2017 (Gorroño *et al.*, 2017) 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.

A.6 Satellite-to-Satellite Intercomparison over Reference Sites

Description

This method entails comparing different satellite measurements, over a period of time, at an agreed set of reference site locations, such as those defined by Benchmark Land Multisite Analysis and Intercomparison of Products (BELMANIP) initiative (Baret *et al.*, 2006). BELMANIP sites are over mostly flat terrain and are homogeneous over a 10x10 km2 area, with a minimum proportion of urban area and permanent water bodies.

Comparison of products over these sites over time can be used to their monitor temporal stability. Such an approach can be useful to complement other direct validation studies, by extending the sampling of sites over both space and time. The site selection was performed for each band of latitude (10° width) by keeping the same proportion of biome types within the selected sites as within the whole band of latitude. Additionally, BELMANIP sites are collocated with ground measurement sites where possible for further comparison.

Scope of Representativeness

Visible to shortwave infrared, depending on reference satellite sensor. Land surface products.

Quality

Although other methods such as PICS provide more accurate method to characterise absolute temporal stability, this method can be used extend scope of the analysis to a wider range of sites. It may also be used for a wider range of products land surface products.

Radiometric Calibration/Validation Methods

Temporal stability monitoring.

A.7 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 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 driven by the sensor's stated spatial resolution (LR, MR, HR, VHR) and their target geometric accuracy.

B.1 Sensor Spatial Response

Description

The parameters of the sensor spatial response indicate how the instrument's optical and related mechanical and electronic systems affect image quality, including possible anomalies such as aberration, stray light, cross-talk, and sample electronic transfer effectiveness. These parameters should be available in pre-launch ground tests and analysis.

If these parameters are not available from pre-launch ground tests, estimates must be made from post-launch data. One such estimate is the "resolution" defined by the ground sampling distance (GSD) or horizontal sampling interval (HSI) and the detector ground FOV (GFOV). Under-sampling occurs if

$$\frac{GFOV}{GSD} < 1. \quad (1)$$

And over-sampling occurs if

$$\frac{GFOV}{GSD} > 1. \quad (2)$$

Over-sampling usually make images blurry, while under-sampling makes the images sharp, but leaves gaps on the ground undetected between observations.

When the pixels are aggregated, say 3x3, then

$$\frac{GFOV}{GSD} \approx 1, \quad (3)$$

and sampling tends to become Nyquist. This scenario could happen when we want to bring geolocation errors to within 0.5 pixels, by aggregating native pixels where the geolocation errors are relatively too large, c.f., Table 6 in Appendix CB.3. We often desire Nyquist sampling. But some over-sampling or a little under-sampling is acceptable, e.g.,

$$75\% < \frac{GFOV}{GSD} < 125\%. \quad (4)$$

The ground sampling distance (GSD) may appear in other forms, such as horizontal sampling interval (HSI) and horizontal cell size (HCS). They may be derived from native samples or re-sampled or aggregated pixels.

While the angular detector FOV for a specific instrument is usually a constant, the ground projected footprint GFOV varies with range from the sensor to the Earth surface, terrain relief, off nadir angle, and possible orbital decay over time. The same can be said of the GSD, HSI or HCS.

Scope of Representativeness

Visible to longwave infrared.

Quality

Pre-launch LSF measurements use a target (reticle) and optics that produce a good-quality apparent image of a slit that is less than 10% of the FOV of the detector. On-orbit LSF measurements can be performed with astronomical objects (e.g. moon) or using higher-resolution imagery that contains targets with sharp edges or spatial features near the Nyquist frequency.

Geometric Calibration/Validation Methods

Sensor spatial response.

B.2 Band-to-Band Registration

Description

A band pair BBR is a collection (in some statistical sense) of detector-to-detector registration (DDR), the co-registration of corresponding detectors between a pair of different spectral bands. DDR is defined as the overlap of footprints of corresponding detectors for a pair of different spectral bands. The full representation of a footprint of a detector should be its point spread function (PSF). Thus, DDR is expressed as below,

$$DDR_{i,j} = \oiint_{(x,y)} \left| PSF_i - PSF_j \right| / \mathbf{2}\, dxdy \tag{5}$$

where, PSF is normalized, i.e.,

$$\oiint_{(x,y)} PSF\, dxdy = 1. \tag{6}$$

Since the exact PSF is usually difficult to obtain, we tend to measure the DDR (BBR) through geolocation differences, such as

$$DDR_{i,j} \approx \begin{cases} 0, & \Delta S_{i,j} > L_S \;\; or \;\; \Delta T_{i,j} > L_T \\ \left(1 - \frac{\Delta S_{i,j}}{L_S}\right)\left(1 - \frac{\Delta T_{i,j}}{L_T}\right), & \text{else} \end{cases} \tag{7}$$

where $\Delta S_{i,j} \Delta T_{i,j}$, and L_s and L_T are the length scales approximating the equivalent area in the cross-track (*S*) and track (*T*) directions, respectively, with the assumption that the PSF is evenly distributed. These length scales may or may not be the same as the sampling intervals (GSD or HSI or HCS), but should be within that order of magnitude0. Note that DDR in Equation (7) has the value between 0 and 100%. A criterion of 80% or greater is often the threshold set for a good DDR (or BBR). BBR could be improved by aggregating the data into a larger pixel size.

B.3 Absolute Positional Accuracy (APA)

B.3.1 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 a resolution of 3 m, and each GCP needs to be well defined in the object space in order to achieve a subpixel pointing accuracy. 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 documentation of the methodology and instrumentation used to acquire the GCPs, uncertainty information from the GPS receiver, and the definition of the GCP.

Geometric Calibration/Validation Methods

Absolute geometric accuracy

B.3.2 Ground Control Points from 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 imagery 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

Visible to longwave 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. In addition, seasonal effects and solar geometry may impact the usability of GCPs. Ideally, GCPs from the same season should be used and from a similar solar geometry as the image data being assessed. Also, GCPs with a finer resolution than the image data being assessed enhances sub-pixel matching accuracy (see Appendix B.3.3).

Geometric Calibration/Validation Methods

Multitemporal geometric stability, relative geometric accuracy.

B.3.3 Image Matching

Description

Image matching of sensor images may be used to assess the absolute geolocation accuracy and monitor the evolution of geometric accuracy within a product, e.g. 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 applications. Generally, intensity correlation methods, such as normalized cross correlation (NCC) produce good results (Wolfe and Nishihama, 2011). However, if the images have illumination changes, pre-processing of the optical products may be necessary, such as applying an edge-enhancement operation.

Scope of Representativeness

Visible to longwave infrared, depending on the test product and reference product.

Quality

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

Geometric Calibration/Validation Methods

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

APPENDIX C PRESENTATION OF ABSOLUTE GEOMETRIC PERFORMANCE

In general, the objective of the quality assessment framework is to grade various aspects of satellite mission quality relative to the mission's claimed performance in these areas. It is recognised, however, that it may be of interest to the reader to see a mission's assessed validation results in an absolute context, with respect to current state-of-the-art. For radiometric performance, it is felt that this is not feasible as each aspect performance is closely linked to the very specific application of the mission or user, such that no absolute performance scale could be determined that is widely useful. For geometric performance, however, this is more feasible as missions typically have a similar overall objective in this area. Therefore, this appendix gives a complimentary means of presenting of geometric validation results to the *Detailed Validation Review* Cal/Val matrix described in Section 3.1, to provide an absolute context.

Table 16 shows a template *Geometric Performance Assessment Matrix,* which is designed to provide this view of the three discussed dimensions of performance together – i.e. the observed and claimed performance in the absolute context. The table effectively plots observed performance against the mission's claimed performance for each of the geometric performance metrics (SSP, BBR and APA – see Section 4.2.2 and APPENDIX B for more information), in terms of four absolute performance classes – *Not Assessed*, *Basic*, *Intermediate* and *Goal*.

Table 17 provides an example of a completed geometric performance assessment matrix. In this example, the observed SSR performance meets the claimed *Basic* performance level, the observed APA performance fails to meet the claimed *Goal* level, and the observed *Intermediate* BBR is outperforming the claimed *Basic* level.

Finally, Table 18, Table 19, and Table 20 give the specific quantitative grading criteria for the absolute assessment required to complete *Geometric Performance Assessment Matrix*.

Table 16 - Template geometric performance assessment matrix

Performance Grade		Observed			
		Not Assessed	Basic	Intermediate	Goal
Claimed	Not Assessable				
	Basic				
	Intermediate				
	Goal				

Table 17 - Example of a completed geometric performance assessment matrix

Grade		Observed Performance			
		Not Assessed	Basic	Good	Excellent
Specification	Not Assessable				
	Basic		SSP	BBR	
	Good				
	Excellent			APA	

Table 18- Geometric Characterization > Sensor Spatial Response – Quantitative Grading Criteria (3-σ)

Grade	FWHM Criteria	MTF Criteria
Basic	The ratio is less than 1.5.	MTF@Nyquist is greater than 0.20.
Intermediate	The ratio is less than 1.3.	MTF@Nyquist is greater than 0.25.
Goal	The ratio is less than 1.1.	MTF@Nyquist is greater than 0.30.

Specification criteria based on the ratio of FWHM of the line spread functions in the along-track and cross-track directions to the GSD in swath products or pixel size in gridded products. This criterion may be used in both pre-launch and post-launch characterization. The MTF at the Nyquist frequency is usually measured pre-launch. The confidence level is usually levied on the success of all detectors, with exception of one failed out of a few hundred detectors. That may be translated to a confidence level of 3-σ or 99.73%.

Table 19 - Geometric Characterization > Absolute Positional Accuracy – Quantitative Grading Criteria (1-σ)

Grade	Criteria
Basic	Circular error smaller than 80% of the ground sample distance or grid pixel size.
Intermediate	Circular error smaller than 50% of the ground sample distance or grid pixel size.
Goal	Circular error smaller than 30% of the ground sample distance or grid pixel size.

Table 20 - Geometric Characterization > Band-to-band Registration – Quantitative Grading Criteria (3-σ)

Grade	Criteria
Basic	Band overlap area is 60% of a sample area.
Intermediate	Band overlap area is 80% of a sample area.
Goal	Band overlap area is 90% of a sample area.