



Joint Earth Observation Mission Quality Assessment Framework - Optical Guidelines



Joint Earth Observation Mission Quality Assessment Framework - Optical Guidelines

Signature/Approval Page

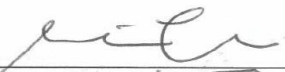
Accepted by:



Katrin Molch
Head of the EO Mission Management & Product Quality Division
EO Mission Management & GS Department
Directorate of EO Programmes
ESA



Date



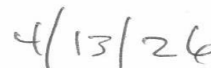
Melissa Yang Martin
Commercial Satellite Data Acquisition Program Manager
Earth Science Division
NASA



Date



Timothy R. Newman
Program Coordinator, National Land Imaging Program
U.S. Geological Survey (USGS)



Date

Preface

This document is part of the Commercial Satellite Data Acquisition (CSDA) Program under the CSDA Project configuration control. Changes to this document shall be verified by a document change notice (DCN) and implemented by change bars or by complete revision.

The material in this document is advisory in nature and is not legally binding. Conformity with this guidance is voluntary only, and nonconformity will not affect the rights or obligations of any entity. This document does not constitute an international agreement.

Abstract

This Joint Earth Observation Mission Quality Assessment Framework – Optical Guidelines document was created for the benefit of the Earthnet Data Assessment Project (EDAP) project and the Commercial Satellite Data Acquisition (CSDA) Program as part of a collaboration between ESA and NASA.

Keywords: configuration management, document control, quality assessment, optical, multispectral

Cover Art: Cover art is AI generated graphic using Microsoft Copilot Designer using term “commercial satellite multispectral radar observation across Atlantic AND Northern Hemisphere AND digital downlink”

Authored and prepared by

Samuel E. Hunt
National Physical Laboratory
United Kingdom

Clément Albinet
European Space Agency / ESRIN
Intergovernmental

Jaime Nickeson
Science Systems and Applications Inc
National Aeronautics and Space Administration
United States of America

Bin Tan
Science Systems and Applications Inc
National Aeronautics and Space Administration
United States of America

Nigel Fox
National Physical Laboratory
United Kingdom

Gary Lin
National Aeronautics and Space Administration
United States of America

Eric Vermote
National Aeronautics and Space Administration
United States of America

Alexei Lyapustin
National Aeronautics and Space Administration
United States of America

Javier Gorroño
National Physical Laboratory
United Kingdom\

Alana Semple
Science Systems and Applications Inc
National Aeronautics and Space Administration
United States of America

Sébastien Saunier
Telespazio
Italy

Valentina Boccia
European Space Agency / ESRIN
Intergovernmental

Change History Log

Revision	Effective Date	Description of Changes
1.0	31/10/21	First public release
1.2	Apr 2024	Updates made after first use, updates to geolocation appendices
2.0	Sep 2024	Restructuring and standardizing layout
2.5	Jul 2025	Incorporating input and feedback from USGS
2.6	Apr 2026	Incorporated updates to signature page, closed and accepted edits to clean document version for signing by all parties.

Table of Contents

1. Introduction	10
1.1. Scope	10
2. EO Mission Quality Assessment Framework Summary	11
2.1. Quality Assessment Report	11
2.2. Data Maturity Matrix	11
2.2.1. Summary Data Maturity Matrix	12
2.2.2. Detailed Validation Maturity Matrix	13
2.3. Approach to Grading	13
2.4. Considerations for the optical domain	14
3. Data Provider Documentation Review	14
3.1. Product Information	15
3.1.1 Product Details	15
3.1.2. Availability & Accessibility	17
3.1.3. Product Format, Flags and Metadata	18
3.1.4. User Documentation	19
3.2. Metrology	20
3.2.1 Radiometric Calibration & Characterisation	20
3.2.2 Geometric Calibration & Characterization	22
3.2.3. Metrological Traceability Documentation	24
3.2.4 Uncertainty Characterization	25
3.2.5 Ancillary Data	26
3.3. Product Generation	27
3.3.1. Radiometric Calibration Algorithm	27
3.3.2. Geometric Processing	27
3.3.3. Retrieval Algorithm – Level 2 Only	28
3.3.4 Mission Specific Processing	29
4. Detailed Validation	31
4.1. Detailed Validation Grading Criteria	31
4.1.1 Validation Method	31
4.1.2. Validation Results Compliance	32
4.2. Performance Metrics	33
4.2.1. Radiometric Validation	33
4.2.2. Geometric Validation	34

4.3. Validation Summary	37
5. References	38
<i>APPENDIX A In-Flight Radiometric Calibration & Validation Methods for Optical Sensors</i>	42
<i>APPENDIX B In-Flight Geometric Calibration Methods for Optical Sensors</i>	47
<i>APPENDIX C Presentation of Absolute Geometric Performance</i>	55

List of Figures

Figure 1. Summary Data Maturity Matrix	12
Figure 2. Validation Maturity Matrix	13
Figure 3. Data Provider Documentation Review Matrix	15

List of Tables

Table 1. Product Information > Product Details – Assessment Criteria	17
Table 2. Product Information > Availability and Accessibility – Assessment Criteria	18
Table 3. Product Information > Product Format, Flags and Metadata – Assessment Criteria.....	19
Table 4. Product Information > User Documentation – Assessment Criteria.....	20
Table 5. Metrology > Sensor Calibration & Characterization – Assessment Criteria	22
Table 6. Metrology > Geometric Calibration & Characterization – Assessment Criteria	24
Table 7. Metrology > Metrological Traceability Documentation – Assessment Criteria	25
Table 8. Metrology > Uncertainty Characterization – Assessment Criteria	26
Table 9. Metrology > Ancillary Data – Assessment Criteria	26
Table 10. Product Generation > Radiometric Calibration Algorithm – Assessment Criteria	27
Table 11. Product Generation > Geometric Processing – Assessment Criteria	28
Table 12. Product Generation > Retrieval Algorithm – Assessment Criteria	29
Table 13. Product Generation > Mission Specific Processing – Assessment Criteria	30
Table 14. Validation > Validation Method – Assessment Criteria	32
Table 15. Validation > Validation Results –Assessment Criteria.....	33

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
CSDA	Commercial Satellite Data Acquisition Program, NASA
DCC	Deep convective cloud
DDR	Detector-to-Detector Registration
EDAP	Earthnet Data Assessment Pilot, ESA
EO	Earth Observation
ESA	European Space Agency
ESF	Edge spread function
FAIR	Findable, Accessible, Interoperable and Reusable
FOV	Field of View
FRM	Fiducial Reference Measurement
FRM4SOC	Fiducial Reference Measurement for Satellite Ocean Colour
FRM4STS	Fiducial Reference Measurements for validation of Surface Temperature 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

PSF	Point Spread Function
PUG/PUM	Product User Guide/Manual
QA4ECV	Quality Assurance Framework for Essential Climate Variables
RadCalNet	Radiometric Calibration Network
RER	Relative Edge Response
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 Organization

1. Introduction

In recent years, the increasing range of applications of Earth Observation (EO) data products and availability of low-cost launch services has resulted in a large expansion of commercial EO satellite systems, developed to deliver end-to-end information services. 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 currently offered.

To this end, both ESA and NASA have initiated activities aimed at assessing the quality and utility of such data products. On the ESA side, the Earthnet Data Assessment Pilot (EDAP) project (Mannan et al., 2019) performs early data assessments on EO missions in optical, atmospheric and SAR domains. In the meantime, the NASA Earth Science Division initiated pilot study (NASA Earth Science Division, 2020) of 3 commercial vendors, and later formalized a Commercial Satellite Data Acquisition Program, and entered into a sustained use agreements for some commercial data sets.

To ensure that decisions on acquisition of commercial data are made with confidence, an objective framework is needed within which to assess the data quality of commercial satellite data sources. The ESA EDAP project established such an EO mission quality assessment framework (Hunt, 2021), that has developed into a collaboration between ESA and NASA.

This document is developed with the evolution of the marketplace and the advance of Earth sciences and applications of EO data products in mind and thus will be revised as appropriate.

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 (Hunt, 2021) for the optical 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. The appendices provide some of the common practices by science users for radiometric and geometric calibration and validation.

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 mission follows community best practice in a manner that is “fit for purpose”.

This quality assessment framework was initially developed within the ESA Earthnet Data Assessment Pilot (EDAP) project and aims to build on the experience of this previous work.

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			Validation Summary	Key
Product Information	Metrology	Product Generation		
Product Details	Radiometric Calibration & Characterization	Radiometric Calibration Algorithm	Radiometric Validation Method	Not Assessed
Availability & Accessibility	Geometric Calibration & Characterization	Geometric Processing	Radiometric Validation Results Compliance	Not Assessable
Product Format, Flags & Metadata	Metrological Traceability Documentation	Retrieval Algorithm	Geometric Validation Method	Basic
User Documentation	Uncertainty Characterization	Mission Specific Processing	Geometric Validation Results Compliance	Good
	Ancillary Data			Excellent
				Ideal


 Not Public

Figure 1. Summary Cal/Val Maturity Matrix (To be filled in to report the result of assessment.)

2.2.2. Detailed Validation Maturity Matrix

The *Detailed Validation Maturity Matrix* (see Figure 2) provides more complete reporting of the analyses contributing to 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 may be multiple *Detailed Validation Maturity Matrices* produced and reported.

Validation Summary	Detailed Validation				
Radiometric Validation Method	R A D I O M E T R I C	Absolute Radiometric Calibration	Signal-to-Noise Method	Temporal Stability Method	
Radiometric Validation Results Compliance		Absolute Radiometric Calibration Results Compliance	Signal-to-Noise Results Compliance	Temporal Stability Results Compliance	
Geometric Validation Method	G E O M E T R I C	Sensor Spatial Response Method	Absolute Positional Accuracy Method	Band-to-Band Registration Method	Temporal Stability Method
Geometric Validation Results Compliance		Sensor Spatial Response Results Compliance	Absolute Positional Accuracy Results Compliance	Band-to-Band Registration Results Compliance	Temporal Stability Results Compliance

Key
Not Assessed
Not Assessable
Basic
Good
Excellent
Ideal
🔒 Not Public

Figure 2. Validation Cal/Val Maturity Matrix for the optical domain, showing 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 is 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. multispectral, hyperspectral) and spatial resolution. The spatial resolution of a sensor may be defined as low resolution (LR; spatial resolution coarser 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) and above.

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

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 & Characterization	Radiometric Calibration Algorithm
Availability & Accessibility	Geometric Calibration & Characterization	Geometric Processing
Product Format, Flags & Metadata	Metrological Traceability Documentation	Retrieval Algorithm
User Documentation	Uncertainty Characterization	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., multispectral, 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
 - For example, defined for optical sensors as:
 - L0 – uncalibrated instrument counts
 - L1 – time-tagged, geo-located, calibrated top-of-atmosphere radiance, reflectance, or brightness temperature
 - L2 – 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, it 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.
- Geographical Coverage
 - The full swath width and footprint of a scene or single acquisition. Define data geographical coverage, i.e., if 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 organization, including email address)
- Product access (e.g., URL, DOI if applicable)
- Restrictions for access and use, if any

Table 1 shows how provision of data product information relates to the grade it achieves for this subsection of the quality assessment.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Relevant information not made available.
Basic	Many pieces of important information are missing.
Good	Some pieces of important information are missing.
Excellent	Almost all required information is available.
Ideal	All required information is available.

Table 1. Product Information > Product Details – Assessment Criteria

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 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 standardized communication protocol
- The protocol is open, free and universally implementable
- The protocol allows for an authentication and authorization 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 2 shows how provision of the above information relates to the grade a data product achieves for this subsection of the quality assessment.

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 the FAIR principles
Excellent	The data set meets many of the FAIR principles and has an associated data management plan. The data are available through an easy-to-access license.
Ideal	The data set fully meets the FAIR principles and has an associated data management plan. The data are available through an easy-to-access license.

Table 2. Product Information > Availability and Accessibility – Assessment Criteria

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, 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. 2003), or, for data from the European Union, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive (INSPIRE Drafting Team 2013)
- whether flags and metadata provide an appropriate breadth of information.

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

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

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. Little useful metadata or data flags.
Good	Data exists in a documented standard file format. Non-standard naming conventions used. Reasonable set of documented metadata and data flags.
Excellent	Data is organized in 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 <i>Excellent</i> .

Table 3. Product Information > Product Format, Flags and Metadata – Assessment Criteria

3.1.4. User Documentation

Data products 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 generic guidance for the expected contents of these documents (Scanlon 2017a, [b] 2017). 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 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 ATBD will likely be the source of much of the information required for the other subsections 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.

Table 4 describes how the product user documentation is assessed.

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 as formal documents or from multiple sources, e.g. articles. Documentation is up to date.
Excellent	PUG meets a referenced community standard, includes a reasonable ATBD. Documents are up to date.
Ideal	PUG ATBD available meeting community referenced standard. Documents are up to date.

Table 4. Product Information > User Documentation – Assessment Criteria

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 Section 4.2.2.

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 APPENDIX A. Methods include inter-calibration with other satellite sensors, vicarious calibration to in-situ reference measurements, and calibration to simulated radiances from so-called pseudo-invariant calibration sites (PICS). APPENDIX A should allow the assessor to judge the extent to which a given 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 5 shows how documentation of sensor radiometric calibration and characterisation is graded within the assessment framework.

Grade	Criteria
Not Assessed	Assessment outside of the scope of study.
Not Assessable	Pre-flight and post-launch sensor calibration & characterization not documented or information not available.
Basic	Pre-flight and post-launch sensor calibration & characterization misses some important aspects of instrument behavior and/or is not entirely of a level of quality to be judged fit for purpose.
Good	Pre-flight and post-launch sensor calibration & characterization covers the most important aspects of instrument behavior at a level of quality to be judged fit for purpose.
Excellent	Pre-flight and post-launch sensor calibration & characterization covers all reasonable aspects of instrument behavior to a quality that is “fit for purpose” in terms of the mission’s stated performance. Pre-flight calibration traceable to SI or community reference, characterization methods meet good practice. Post-launch Cal/Val uses appropriate community infrastructure/methods (e.g. RadCalNet).
Ideal	Meets <i>Excellent</i> criteria, additionally, the calibration and characterization 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).

Table 5. Metrology > Sensor Calibration & Characterization – Assessment Criteria

3.2.2. Geometric Calibration & Characterization

As with radiometric calibration and characterisation, geometric calibration and characterisation, both pre-flight and on-orbit, should also 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 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, Lin et al., 2022).

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, Lin et al., 2022)

Table 6 shows how the geometric calibration and characterization is graded within the assessment framework.

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Geometric calibration & characterization 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” in terms of the mission’s stated performance.
Good	Geometric calibration & characterisation covers most important aspects of instrument behaviour at a level of quality to be judged “fit for purpose” in terms of the mission’s stated performance.
Excellent	Geometric calibration & characterization covers all reasonable aspects of instrument behavior to a quality that is “fit for purpose” in terms of the mission’s stated performance. Post-launch characterization uses appropriate community infrastructure/methods (e.g., from CEOS).
Ideal	In addition to meeting <i>Excellent</i> criteria, geometric calibration and characterization 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-user expectations.

Table 6. Metrology > Geometric Calibration & Characterization – Assessment Criteria

3.2.3. Metrological Traceability Documentation

Traceability is defined in the vocabulary of metrology (VIM) (JCGM 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 centered on “uncertainty tree diagram”, more suitable for reconstructed data (and some geophysical products) processing and should be the aspiration for missions in the future (Mittaz, Merchant, and Woolliams 2019).

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 7 shows how the metrological traceability documentation is graded, based on its completeness.

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, a traceability chain documented, identifying all reasonable steps of 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.

Table 7. Metrology > Metrological Traceability Documentation – Assessment Criteria

3.2.4. Uncertainty Characterization

To ensure measurements are both meaningful and defensible, it is crucial that they come with 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 measurement and, 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 products 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 (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 8 shows the uncertainty characterization grading under the assessment framework.

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 included. Uncertainty per pixel provided.
Ideal	Full GUM approach is used to estimate measurement uncertainty, including a treatment of error-covariance. Per pixel uncertainties are provided for appropriate error correlation components, e.g., random systematic.

Table 8. Metrology > Uncertainty Characterization – Assessment Criteria

3.2.5. Ancillary Data

Throughout the processing chain there may be a requirement for external input data, for example, 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 (in terms of factors like surface type). The requirements will be specific to the retrieval method used and may require some expert judgement.

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

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”
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.

Table 9. Metrology > Ancillary Data – Assessment Criteria

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 image formation and 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., multispectral, 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.

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

Table 10 shows how the image formation and calibration algorithm are graded.

Grade	Criteria
Not Assessed	Assessment outside the scope of study.
Not Assessable	Image formation and calibration algorithm not documented.
Basic	Image formation and calibration algorithm somewhat documented. Calibration algorithm too simple to be judged “fit for purpose” in terms of the mission’s stated performance.
Good	Image formation and 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	Image formation and calibration algorithm documented. Calibration used “fit for purpose” in terms of the mission’s stated performance in all expected use cases.
Ideal	Image formation and calibration algorithm well-documented. State-of-the-art calibration algorithm used, easily “fit for purpose” in terms of the mission’s stated performance.

Table 10. Product Generation > Radiometric Calibration Algorithm – Assessment Criteria

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 11 shows how the geometric processing is graded.

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.

Table 6. Product Generation > Geometric Processing – Assessment Criteria

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 12 shows how the retrieval algorithm used to generate L2 products is graded.

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, easily “fit for purpose” in terms of the mission’s stated performance, full uncertainty budget derived and validated.

Table 7. Product Generation > Retrieval Algorithm – Assessment Criteria

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 and based on the criteria described in Table 13, and then a combined score determined.

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

Table 8. Product Generation > Mission Specific Processing – Assessment Criteria

4. Detailed Validation

In this section we provide guidelines for a *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 2), which are then summarised across all performance metrics in the *Validation Summary*. This *Validation Summary* is the same summary presented in the *Summary Cal/Val Maturity Matrix* shown in Figure 1.

The remainder of this section includes:

- The criteria for grading the quality of the validation method used and validation results compliance are given in Section 4.1.
- Each of the radiometric and geometric performance metrics to be assessed are described in Section 4.2.
- Finally, in Section 4.3 the approach for synthesising the results of the *Detailed Validation* into the *Validation Summary* is described.

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 of the Radiometric and Geometric performance metrics.

4.1.1. Validation Method

Generally, satellite validation attempts to demonstrate compliance of mission data products with respect to some claimed performance level (e.g., documented specifications) by comparison 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., 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 14 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 APPENDIX A for measurement validation and APPENDIX B for geometric validation, which is intended to act as the basis for such assessment.

Table 9. 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 characterized 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 data performance claimed by the data provider and with the possibly more stringent standards from the user community.

Table 15 shows how the validation results are graded within the assessment framework.

Table 10. 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 Measurement 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 APPENDIX A.

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. APPENDIX A details these methods in more detail

4.2.1.2. *Signal-to-Noise*

Measurement noise, occurring in the satellite sensor detector and processing chain, provides a fundamental limit to the achievable quality of a given instantaneous observation. 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 3.2.2, 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. APPENDIX A 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 registration (BBR)
- Temporal stability

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 the 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. For gridded images, an LSF may be constructed in a cross-row or cross-column direction. Alternatively, an LSF may be derived from an edge spread function (ESF), which can be constructed from an image over a natural or man-made sharp edge feature. From the LSF, we can determine image quality parameters such as the footprint size at the full width at half maximum (FWHM), and the modulation transfer function (MTF). Alternatively, from an ESF, relative edge response (RER) can be determined as an image quality parameter. In general, we want the MTF to be at least 0.25 or greater at the Nyquist frequency (one cycle every

per two sampling intervals). Note that for gridded products, the MTF can be improved by aggregating or downsampling 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*

As agency and commercial satellite sensors become more advanced and numerous, with many providing high resolution or very high resolution (VHR) 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, or within 0.5 of the sensor's footprint size measured at the full width at half maximum (FWHM) of its PSFs if that is available. The GCPs should be as evenly distributed geographically 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 or downsampling 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 footprint areas 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 area 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.

4.2.2.4. *Temporal Stability*

As described in Section 3.2.2, 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. The validation stages defined by the CEOS Land Product Validation subgroup include requirements for geographical and temporal consistency. This consistency cannot be assessed without adequate geometric temporal stability. Ideally, the satellite data products are evaluated over globally representative locations. APPENDIX B details some of the absolute positional accuracy methods that can be used to quantify the positional stability of sensor products. These can be applied multiple times over a season and/or years to assess the temporal stability of satellite data products.

It is a challenge to achieve sub-pixel accuracy for images at very high resolution. It is also recognized that there are limited globally distributed points of absolute ground truth. High resolution or VHR images are often used as reference for calibration and validation of geolocation performance, but users should be aware that the uncertainties of these reference images can exceed the pixel size of VHR images. Users of EO data are often in seek of temporal stability at study sites for time series analyses and thus the temporal stability is an important aspect of geolocation accuracy.

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 2). 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.

5. References

- Bouvet, M. et al. (2019). RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range, Remote Sensing, Multidisciplinary Digital Publishing Institute, 11(20), p. 2401. doi: 10.3390/rs11202401.
- Cantral, S. J., Christopherson, J. B. (2024). Joint Agency Commercial Imagery Evaluation (JACIE) Best Practices for Remote Sensing System Evaluation and Reporting, Open-File Report 2024-1023, <https://doi.org/10.3133/ofr20241023>
- CEOS LSI (2020). CARD4L Product Family Specification - Surface Reflectance. Available at: https://ceos.org/ard/files/PFS/SR/v5.0/CARD4L_Product_Family_Specification_Surface_Reflectance-v5.0.pdf.
- Chander, G. et al. (2013). Overview of Intercalibration of Satellite Instruments, IEEE Transactions on Geoscience and Remote Sensing (2013) 51(3): 1056-1080. DOI: 10.1109/TGRS.2012.2228654
- Chen, T.; Liu, Y., (2021). A Quick Band-to-Band Mis-Registration Detection Method for Sentinel-2 MSI Images. Remote Sens. 2021, 13, 3351. <https://doi.org/10.3390/rs13173351>
- Datla, R. U. et al. (2011). Best practice guidelines for pre-launch characterization and calibration of instruments for passive optical remote sensing, Journal of Research of the National Institute of Standards and Technology, 116(2), p. 621. doi: 10.6028/jres.116.009.
- Dechoz, C. et al. (2015). Sentinel 2 global reference image, in Bruzzone, L. (ed.) Image and Signal Processing for Remote Sensing XXI. SPIE, pp. 94–107. Available at: <https://doi.org/10.1117/12.2195046>.
- Eaton, B. et al. (2020). NetCDF Climate and Forecast (CF) Metadata Conventions. Available at: <https://cfconventions.org/latest.html>.
- Fang, D. T. and Puschell, J. (2010). Imagery spatial performance throughput correction methodology, Remote Sensing System Engineering III, SPIE 7813, doi: 10.1117/12.860740.
- Fox, N. (2019). FRM4STS D-180 Final Report. Available at: <http://www.frm4sts.org/wp-content/uploads/sites/3/2020/01/OFE-D-180-V1-Iss-1-Ver-1-signed.pdf>.
- Franch, B. et al. (2017). A 30+ year AVHRR land surface reflectance climate data record and its application to wheat yield monitoring, Remote Sensing, 9(3), pp. 1–14. doi: 10.3390/rs9030296.
- Gorroño, J. et al. (2017). A radiometric uncertainty tool for the Sentinel-2 mission, Remote Sensing, Multidisciplinary Digital Publishing Institute, 9(2), p. 178. doi: 10.3390/rs9020178.
- Govaerts, Y. M. et al. (2018). 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 Sensing*. Multidisciplinary Digital Publishing Institute, 10(12), p. 1959. doi: 10.3390/rs10121959.
- Gutman, G. et al. (2013). Assessment of the NASA–USGS Global Land Survey (GLS) datasets, *Remote Sensing of Environment*, 134, pp. 249–265. doi: <https://doi.org/10.1016/j.rse.2013.02.026>.
- Hassanpour, M., Dadras Javan, F., and Azizi, A., (2019). Band to Band Registration of Multi-Spectral Aerial Imagery - Relief Displacement and Miss-Registration Error, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4218, pp. 467–474, 2019. doi:10.5194/isprs-archives-XLII-4-W18-467-2019.
- Hearn, D.R., Mendenhall, J.A., Willard, B.C. (1999). Spatial calibration of the EO-1 Advanced Land Imager, *Proc. SPIE 3750, Earth Observing Systems IV*; doi: 10.1117/12.363503
- Holmes, T. R. H., Poulter, B., McCorkel, J., Jennings, D. E., Wu, D. L., Efremova, B., et al. (2024). On-orbit spatial performance characterization for thermal infrared imagers of Landsat 7, 8, and 9, ECOSTRESS and CTI. *Journal of Geophysical Research: Biogeosciences*, 129, e2023JG007506. <https://doi.org/10.1029/2023JG007506>
- Hunt, S. E. (2021). Earth Observation Mission Quality Assessment Framework.
- INSPIRE Drafting Team Metadata and European Commission Joint Research Centre (2013). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. Available at: <https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1>.
- JCGM (2008). Evaluation of measurement data - Guide to the expression of uncertainty in measurement, JCGM, 100. Available at: https://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf (Accessed: 6 February 2018).
- JGCM (2012). International vocabulary of metrology – Basic and general concepts and associated terms (VIM), JGCM, 200.
- Knight, E. J. and Kvaran, G. (2014). Landsat-8 Operational Land Imager Design, Characterization and Performance, *Remote Sensing*, 6(11), pp. 10286–10305. doi: 10.3390/rs61110286.
- Lin, G., R. E. Wolfe, M. Nishihama (2011). NPP VIIRS Geometric Performance Status, *Proc. of SPIE 8153, Earth Observing Systems XVI*, edited by James J. Butler, Xiaoxiong Xiong, Xingfa Gu, 81531V, doi: 10.1117/12.894652.
- Lin, G. and Wolfe, R. E. (2016). JPSS-1 VIIRS at-launch geometric performance, in Butler, J. J., Xiong, X. (Jack), and Gu, X. (eds) *Earth Observing Systems XXI*. SPIE, pp. 496–505. Available at: <https://doi.org/10.1117/12.2238804>.

- Mannan, R. et al. (2019). ESA's Earthnet data assessment pilot: paving the way for new space players, in Neeck, S. P., Martimort, P., and Kimura, T. (eds) *Sensors, Systems, and Next-Generation Satellites XXIII*. SPIE, pp. 422–428. doi: 10.1117/12.2532818.
- Mittaz, J., Merchant, C. J. and Woolliams, E. R. (2019). Applying principles of metrology to historical Earth observations from satellites, *Metrologia*. IOP Publishing, 56(3), p. 032002. doi: 10.1088/1681-7575/ab1705.
- NASA Earth Science Division (2020). Commercial Satellite Data Acquisition Program Pilot Evaluation Report. Available at: https://cdn.earthdata.nasa.gov/conduit/upload/14180/CSDAPEvaluationReport_Apr20.pdf.
- National Imagery and Mapping Agency (2000). Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems. 3rd edn. St. Louis, MO, USA: National Imagery and Mapping Agency. Available at: http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html.
- Scanlon, T. (2017a). QA4ECV Product Documentation Guidance: Algorithm Theoretical Basis Document. Available at: http://www.qa4ecv.eu/sites/default/files/QA4ECVATBD_Guidance.pdf.
- Scanlon, T. (2017b). QA4ECV Product Documentation Guidance: Product User Manual. Available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_PUM_Guidance.pdf.
- Scanlon, T. (2017c). QA4ECV Product Documentation Guidance: Provenance Traceability Chains. Available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_Traceability_Chains_Guidance.pdf.
- Scanlon, T. (2017d). QA4ECV Product Documentation Guidance: Validation and Intercomparison Report. Available at: http://www.qa4ecv.eu/sites/default/files/QA4ECV_Validation_Guidance.pdf.
- Semple, A., B. Tan, and G. Lin. (2024). Assessment of Sensor Footprint Size and Comparison of Commercial Satellite Images, JACIE 2024 Workshop, <https://calval.cr.usgs.gov/apps/sites/default/files/jacie/2024-CalVal2-Alana-Semple.pptx>
- Storey, J. C. and Choate, M. (2000). Landsat 7 on-orbit geometric calibration and performance, in Shen, S. S. and Descour, M. R. (eds) *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*. SPIE, pp. 143–154. Available at: <https://doi.org/10.1117/12.410335>.
- Storey, J., Choate, M. and Lee, K. (2014). Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance, *Remote Sensing*, 6(11), pp. 11127–11152. doi: 10.3390/rs6111127.
- Tan, B., J. J. Dellomo, C. N. Folley, T. J. Grycewicz, S. Houchin, P. J. Isaacson, P. D. Johnson, B. C. Porter, A. D. Reth, P. Thiyanaratnam, and R.E. Wolfe, (2020). GOES-R series image navigation and registration performance assessment tool set, *Journal of Applied Remote Sensing* 14(3), 032405 (8 April 2020). <https://doi.org/10.1117/1.JRS.14.032405>

- Tansock, J. et al. (2015). Guidelines for Radiometric Calibration of Electro-Optical Instruments for Remote Sensing, NISTHB 157. Gaithersburg, MD. doi: 10.6028/NIST.HB.157.
- Taylor, M. et al. (2019). FIDUCEO D2.2 (AVHRR): Report on the AVHRR FCDR Uncertainty. Available at: <https://cordis.europa.eu/project/id/638822/results>.
- Thome, K., Smith, N. and Scott, K. (2001). 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). IEEE, pp. 1209–1211. doi: 10.1109/IGARSS.2001.976794.
- Tilton, J. C., Wolfe, R. E., & Lin, G. (2017). On-orbit line spread function estimation of the SNPP VIIRS imaging system from Lake Pontchartrain causeway bridge images. *J. of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 5056–5072. <https://doi.org/10.1109/JSTARS.2017.2729879>
- Tilton, J., Wolfe, R., Lin, G., and Dellomo, J. (2019). On-Orbit Measurement of the Effective Focal Length and Band-to-Band Registration of Satellite-Borne Whiskbroom Imaging Sensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 12. 4622-4633. 10.1109/JSTARS.2019.2949677.
- USGS, (2025). Landsat Missions, Landsat 8-9 Calibration Validation Algorithm Description Document, LSDS-1747, Available at: <https://www.usgs.gov/media/files/landsat-8-9-calibration-validation-algorithm-description-document>
- Vendt, R. (2020). FRM4SOC D-290 Final Report. Available at: https://frm4soc.org/wp-content/uploads/filebase/parentdir/techreports/temp_pic/D-290-FRM4SOC-FR_30.06.2020.pdf.
- Wenny, B. N., Helder, D., Hong, J., Leigh, L., Thome, K. J. & Reuter, D. (2015). Pre- and post-launch spatial quality of the Landsat 8 Thermal Infrared Sensor, *Remote Sensing*, 7, 1962-1980. <https://doi.org/10.3390/rs70201962>
- Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3(1). doi: 10.1038/sdata.2016.18.
- Wolfe, R. E. et al. (2013). Suomi NPP VIIRS prelaunch and on-orbit geometric calibration and characterization, *Journal of Geophysical Research: Atmospheres*, 118(20), pp. 11,508-511,521. doi: <https://doi.org/10.1002/jgrd.50873>.
- Wolfe, R. E. and Nishihama, M. (2011). Accurate MODIS global geolocation through automated ground control image matching, in Le Moigne, J., Netanyahu, N. S., and Eastman, R. D. E. (eds) *Image Registration for Remote Sensing*. Cambridge University Press, pp. 437–455. doi: 10.1017/CBO9780511777684.022.

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 that 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 Maxar 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 (in 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 a 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 uniform. 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),
- 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 dependent.

Quality

These measurements can have traceability chains and quantified uncertainties, though are not ubiquitous across the field. The 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 a more accurate method to characterize absolute temporal stability, this method can be used to extend the scope of the analysis to a wider range of sites. It may also be used for a wider range of 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, coarser than 300 m; MR, 30 – 300 m; HR, 5 – 30 m; VHR, finer than 5 m) and their target geometric accuracy.

For VHR images, ground sampling distance (GSD) or grid size could differ from footprint size dramatically due to optical effects. Since footprint size is a criterion for grading positional accuracy, and footprint area is a criterion for grading BBR (see Appendix C), the methods for sensor spatial response are described first to estimate of footprint size and area, and image quality parameters.

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, crosstalk, 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 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. 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 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.

Two methods are described below to estimate sensor spatial response in the form of LSF or ESF. The associated image quality parameters such as FWHM (which defines GFOV), MTF or relative edge response (RER) can be estimated.

B.1.1 Bridges for LSF construction

Description

One method is similar to using multiple, uniformly spaced, and phased slits in pre-launch ground tests under controlled ambient or thermal vacuum conditions (Fang and Puschell, 2010; Lin et al., 2011). The straight bridges over relatively dark and cold open water bodies in natural settings act like slits, providing line sources. When a bridge is slightly tilted with along-track or cross-track direction in a non-gridded swath image, the tilted angle provides a phase shift along the bridge direction. The uniform and extended areas on both sides of a bridge could be used to construct an LSF, see examples in (Tilton et al., 2017) for LR VIIRS images. For gridded images, slightly tilted bridge against either the row or column direction may provide line sources for LSF construction in the cross-row or cross-column direction, see examples in (Homes et al., 2024) for MR images from Landsat 7, 8, and 9, the ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) and the Compact Thermal Imager (CTI). Image quality parameters such as FWHM, MTF will then be derived from the LSFs.

There are other methods to assess the sensor performance from the on-orbit satellite data not described here. For example, if L0 data is available, a more comprehensive Monte Carlo simulation method can be applied. The method was developed for Landsat optical sensors and is described in full in the Landsat 8-9 Calibration Validation Algorithm Description Document (USGS, 2025). Through this method, an optimized optical transfer function, the FWHM, RER, and MTF at Nyquist frequency can be estimated based on an image of a bridge. The advantage of this method is that the assessment results are more accurate than the method above. The disadvantage is that some required input parameters are not available to the regular users and this method can only work with Level 0 raw image in the scan coordinates, which is often not available to users working with commercial data.

Scope of Representativeness

Visible to longwave infrared

Quality

The width of a bridge usually causes over-estimate of FWHM of the constructed LSF from the true sensor LSF. When the bridge width is less than 20% of the sensor's FWHM, the over-estimate is negligible. The impact is less than 10% if the bridge width is less than half the sensor's FWHM. A caution needs to be exercised to examine the uniformity of the open water background around the section of the bridge used for LSF construction.

Geometric Calibration/Validation Methods

Validation of sensor spatial response.

B.1.2 Slanted Edges for ESF construction

Description

This method is similar to using a straight knife edge in pre-launch ground tests under controlled ambient or thermal vacuum conditions (Hearn et al., 1999; Wenny et al., 2015). Knife edge coastal border areas or the edges created in man-made checkerboard features for spatial calibration (https://calval.cr.usgs.gov/apps/spatialsites_catalog) are manmade knife edges. When these edges are slightly slanted away from along-track or cross-track directions in swath images, or along-row or along-column directions in gridded images, they provide phase shift along the edge direction. The uniform and extended areas on both sides of the edge could be used to construct edge spread functions (ESFs) (Wenny et al., 2015, Semple et al., 2024). A relative edge response (RER) can be derived from an ESFs as an image quality parameter. LSFs can then be derived from ESFs. Image quality parameters such as FWHM, MTF can then be derived from the LSFs.

Scope of Representativeness

Visible to longwave infrared

Quality

When the length of an edge extends a multiple of a targeted imager's footprint size, an ESF may be properly constructed. In natural settings, perturbation of the edge straightness less than 20% of the sensor's footprint size has negligible impact on the ESF construction. However, a caution needs to be exercised to examine the uniformity of the radiometric responses on both sides of the edge.

Geometric Calibration/Validation Methods

Validation of sensor spatial response.

B.2 Absolute Positional Accuracy (APA)

B.2.1 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 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.2.2 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.2.3 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.

B.3 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,

$$DDR_{i,j} = \iint_{(x,y)} |PSF_i - PSF_j|/2 \, dx dy \quad (5)$$

where, PSF is normalized, i.e.,

$$\iint_{(x,y)} PSF \, dx dy = 1. \quad (6)$$

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

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

where $\Delta S_{i,j}$ and $\Delta T_{i,j}$ are the absolute offsets between detector i and j in the cross-track and along-track direction, 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 are the footprint size, and they may not be the same as the GSD). 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 BBR (the statistics of DDR measurements). BBR could be improved by aggregating or downsampling the data into a larger pixel size.

In on-orbit operations, DDR is rarely measured. Instead, BBR is measured based on the assumption that SSR for very detector is the same in either swath products or gridded images.

Sometimes, BBR could be expressed as geolocation offsets in orthogonal directions between bands, i.e., $\Delta S_{i,j}, \Delta T_{i,j}$ themselves while the length scale of sensor spatial response as describe in Section B.1 are still to be determined, where i and j are corresponding band pair i and j .

In gridded products, BBR offsets are measured in the cross-row and cross-column directions.

Three methods are described below to estimate on-orbit multispectral BBR.

B.3.1 Crosshair Assessment

Description

The most precise method would be to find a crosshair target where the spectral response changes sharply across a boundary. Valid locations would include checkerboard Cal/Val sites, road intersections, or regular sharp terrain boundaries (i.e. farmlands, clear cut forest). Perform an edge-spread assessment across all bands in a single image of the crosshair. With edge spread functions calculated for the boundary, shifts in the location of the center value are precise measurements of band misalignment. Although this is the most precise method, the limitation is that this method can only be assessed over few point locations.

Scope of Representativeness

Visible to longwave infrared.

Quality

Full traceability and uncertainty quantification for this method requires the methodology uncertainty information. Valid evaluation sites are limited and vary based on image pixel size. The results only apply to the evaluated point in the image, so image warping cannot be identified here. It is the most precise method and should be used in compliment to at least one of the two methods described in Sections B.3.2 and B.3.3.

Geometric Calibration/Validation Methods

Band-to-band registration, sensor spatial response.

B.3.2 Direct Comparison of Two Bands

Description

Image matching between two spectral bands may also be used for assessing band misalignment. This method assumes that the spectral responses of two bands to be assessed are similar to each other, resulting in similar features for the assessment. Usually, the BBR is not performed by matching the whole image pair but a few sub-image pairs. To achieve the sub-pixel assessment accuracy, the image matching between the two corresponding subsets is not performed at the original image resolution but at a finer resolution. The subset pairs are resampled to a finer resolution by the interpolation methods before performing the image matching. In addition, a curve

fitting process on the correlation plane may be used to improve the assessment accuracy further [Chen and Liu, 2021; Hassanpour et al., 2019; Tan et al., 2020]. As this method 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).

Scope of Representativeness

Visible to longwave infrared.

Quality

Full traceability and uncertainty quantification for this method requires the methodology uncertainty information, and any post-processing applied including outlier removal. For bands that are acquired sequentially, uncertainties introduced by illumination differences, and/or temporal decorrelation (i.e., the acquisition time difference) between the band pairs must be reported. Additionally, uncertainties introduced by spectral response differences between the bands must be reported.

Geometric Calibration/Validation Methods

Band-to-band registration.

B.3.3 Comparison of Band pair through matching with reference images

Description

BBR can be assessed using the image matching method between a target and a reference image, (see Section B.2.3 for more details in image matching, and Tilton et al., 2019 in applications). The intent here is to use reference image pair with similar spectral response to the target image pair. In this method, it is assumed the reference image has negligible BBR errors. The target images are assessed against the corresponding reference image pair. BBR is calculated as the difference of the determined offsets between two assessments in the corresponding images. The reference could also be one or more similar spectral bands as the target image pair. The reference images can also be used to bridge the target image pair where the images are too dissimilar to analyse directly. Using a reference image in the analysis can increase uncertainties unless it is of significant finer resolution than the band pair being assessed. The method is based on the use of a reference raster dataset of known geometric accuracy and preferably finer resolution than the image being examined. As this method 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 band pair and the third-party 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 a data user to see a mission’s assessed validation results in an absolute context, with respect to current state-of-the-art. For geometric performance, missions typically have a similar overall objective in this area. Therefore, this appendix gives a complimentary means of presenting geometric validation results to the *Detailed Validation Review* Cal/Val matrix described in Section 4, to provide an absolute context.

Table C-1 shows a template *Geometric Performance Assessment Matrix*, which is designed to provide this view of the four 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, APA, and TS – see Section 4.2.2 and APPENDIX B for more information), in terms of five absolute performance classes – *Not Assessed*, *Basic*, *Good*, *Excellent* and *Ideal*.

Table C-1 – Template and example of a completed geometric performance assessment matrix

Performance Grade		Observed			
		Basic	Good	Excellent	Ideal
Not Assessed				TS	
Claimed	Basic	SSR	BBR		
	Good				
	Excellent		APA		
	Ideal				

Table C-1 also 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 level, and the observed BBR is outperforming the claimed *Basic* level. The example also shows temporal stability of Time Series (TS) geolocation accuracy is *Excellent* in observed performance, but the image vendor does not provide data in their specifications.

Finally, Table C-2a, Table C-2b, Table C-3, and Table C-4 give the specific quantitative grading criteria for the assessment required to complete *Geometric Performance Assessment Matrix*.

The quality criteria for Sensor Spatial Response are based on the ratio of the FWHM of the line spread functions, constructed in pre-launch tests or on-orbit operations, to the ground sampling distance (GSD) or gridded image pixel size may be regarded as primary criteria as list in Table C-2a.

Table C-2a- Geometric Characterization > Sensor Spatial Response – Quantitative Grading Criteria in FWHM to GSD or pixel size ratio (90th percentile)

Grade	FWHM Criteria
Basic	The ratio is greater than 2.
Good	The ratio is less than or equal to 2 but greater than 1.5.
Excellent	The ratio is less than or equal to 1.5 but greater than 1.25.
Ideal	The ratio is less than or equal to 1.25 but greater than 0.75.

The primary criteria above do not take the into account the shape of an LSF. The Modulation transfer function (MTF) does that and is commonly used in the remote sensing community to measure image spatial resolution quality. It is derived from an LSF by Fourier transform from spatial domain to spatial frequency domain and its modulus is normalized to its value at 0 frequency. The form of the LSF varies. We assume a Gaussian function for simplicity to represent a typical LSF. We also use MTF at Nyquist frequency as a typical MTF criterion. Spatial Nyquist frequency is a wave cycle every 2 GSD length or 2 grid pixel length.

Sometimes relative edge response (RER) is used as a measure of image spatial resolution. RER is defined as the slope of an edge spread function (ESF) 0.5 pixel away at the right and left of center. Here again, we assume the ESF is formed by convolving a Gaussian function with a step function. The MTF at Nyquist frequency and RER may be regarded as secondary criteria as listed in Table C-2b. The MTF and RER criteria in Table C-2b are equivalent to FWHM criteria in Table C-2a when the LSF is in the form of a Gaussian function.

Table C-2b- Geometric Characterization > Sensor Spatial Response – Quantitative Grading Criteria in MTF or RER (90th percentile)

Grade	RER Criteria	MTF at Nyquist Criteria
Basic	RER is less than 0.44.	MTF is less than 0.03.
Good	RER is greater than or equal to 0.44 but less than 0.55.	MTF is greater than or equal to 0.03 but less than 0.13.
Excellent	RER is greater than or equal to 0.55 but less than 0.65.	MTF is greater than or equal to 0.13 but less than 0.25.
Ideal	RER is greater than or equal to 0.65 but less than 0.9.	MTF is greater than or equal to 0.25 but less than 0.6.

A satellite vendor may provide spatial resolution in FWHM, MTF or RER form. Assessment of image spatial resolution quality in FWHM, MTF or RER form is often tedious. Thus, there are often just a small number of data points. The statistics of 90th percentile often indicates worst case scenario.

Table C-3 - Geometric Characterization > Absolute Positional Accuracy and TS – Quantitative Grading Criteria

Grade	Criteria	VHR Criteria
Basic	Circular error at 90 th percentile (CE90) is greater than or equal to 1.0 footprint size.	Circular error at 90 th percentile (CE90) is outside vendor’s specification.
Good	CE90 is less than or equal to 1.0, but greater than 0.6 footprint size.	CE90 is within vendor’s specification, but greater than 2 pixels.
Excellent	CE90 is less than or equal to 0.6 but greater than 0.3 footprint size.	CE90 is less than or equal to 2 pixels, but greater than 0.6 footprint size.
Ideal	CE90 is less than or equal to 0.3 footprint size.	CE90 is less than or equal to 0.6 footprint size, suitable for temporal analyses.

A science and applications data user usually works with data in pixel space without needing to know the sensor’s spatial response (SSR) and its associated footprint size. A user performing time series analysis could see significant impacts of geolocation inaccuracy of greater than 0.5 pixels from one moment to the next across boundaries between various land cover and land use (edge of farmland, water bodies, etc.) retrieved from the images. For VHR imagery, a time series analysis could see significant impacts of geolocation inaccuracy of greater than 0.5 pixels A pixel shift could improve the user’s product quality.

Table C-4 - Geometric Characterization > Band-to-band Registration – Quantitative Grading Criteria (90th percentile)

Grade	Criteria
Basic	Band pair overlap area is less than or equal to 25% of footprint area.
Good	Band pair overlap area is less than or equal to 64% but greater than 25% footprint area.
Excellent	Band pair overlap area is less than or equal to 90% but greater than 64% footprint area.
Ideal	Band pair overlap area is greater than 90% footprint area.

A science and applications data user usually works with data in pixel space without needing to know the sensor’s spatial response (SSR) and its associated footprint area. The user typically performs a pixel-based analysis (NDVI, time series, etc) and are often looking for boundaries between various land cover and land use (edge of farmland, water bodies, etc) retrieved from the images. The impact of a BBR offset greater than 0.5 pixels between two bands used in bio-geophysical parameter retrievals could be significant. A pixel shift could improve BBR and the user’s product quality.

The grade for a summary geometric performance represents a combination of the grades of the contributing factors from Tables C-2 to C-4. The approach is to effectively average the grades of the contributing factors, where each grade is valued as follows: Basic is 1, Good is 2, Excellent is 3, and Ideal is 4.