

FLEX Validation Strategy, Goals and Methodological Framework

Roberto Colombo¹, Béatrice Berthelot², Alireza Mahmoodi², Louis Sainte-Marie², Gwennaél Matot², Jorge Vicent Servera², Ornela Nanushi², Brice Mora², Mohamed Salah Djellali², Neus Sabater³, Pekka Kolmonen³, Timo H. Virtanen³, Astrid M. Zimmermann⁴, Pieter De Vis⁴, Micol Rossini¹, Giulia Tagliabue¹, Pietro Chierichetti¹, Sergio Cogliati¹, Juliane Bendig⁵, Bastian Siegmann⁵, Uwe Rascher⁵, Egor Prikaziuk⁶, Christiaan van der Tol⁶, Darren Ghent⁷, Jasdeep Singh Anand⁷, Tommaso Julitta⁸, Matthias Drusch⁹, Marco Celesti⁹, Dirk Schütttemeyer⁹, Marin Tudoroiu¹⁰

¹ University Milano Bicocca, Milano, Italy

² Magellium, Toulouse, France

³ Finnish Meteorological Institute, Helsinki, Finland

⁴ National Physical Laboratory, Teddington, UK

⁵ Forschungszentrum Jülich GmbH, Jülich, Germany

⁶ University of Twente, Enschede, The Netherlands

⁷ University of Leicester, Leicester, UK

⁸ JB Hyperspectral Devices, Dusseldorf, Germany

⁹ European Space Agency, ESTEC, Noordwijk, The Netherlands

¹⁰ European Space Agency, ESRIN, Frascati, Italy

Keywords: FLEX mission, FLEX products, Cal/Val plans

Abstract

Using novel technology and dedicated algorithms, the FLuorescence EXplorer (FLEX) will yield information about the health of the world's plants. The information gathered will be used to improve our understanding of how carbon moves between plants and the atmosphere and how photosynthesis affects the carbon and water cycles. The FLEX mission offers two high-level products for exploitation by end-users: the Level-1C (L1C) and Level-2 (L2). In this context, a comprehensive strategy for the validation of FLEX operational products is planned. This strategy ensures the traceability to the mission requirements and guarantees that all parameters relevant for the operational products have been adequately validated. The FLEX mission requirements are highly challenging and they cover the end-to-end Earth observation system including high-level requirements, mission operations, data product development and processing, data distribution and data archiving. Risks associated with the validation of the FLEX products have been identified, possible mitigation scenarios have been outlined and back-up solutions proposed. In this paper, we provide an overview of the FLEX strategy for the calibration and validation (cal/val) of the satellite's operational Level 1C and Level 2 science data products. It also outlines instruments and measurements needed to assess spatial and temporal comparison and possible sampling strategies and tools for upscaling ground measurements.

1. Introduction

1.1 Overview of the FLEX mission

The FLuorescence EXplorer (FLEX) mission, developed by the European Space Agency (ESA) as its 8th Earth Explorer, will provide global maps of vegetation fluorescence as an indicator of photosynthetic activity together with the necessary parameters for deriving the amount of carbon assimilated by plants. FLEX is currently planned to be launched in September 2026. Beside Sun Induced Fluorescence (SIF), FLEX will provide different products, encompassing spectral radiances, reflectance, and vegetation parameters. The FLEX on-board instrument, the FLuORescence Imaging Spectrometer (FLORIS) sensor, will acquire data in the 500 - 780 nm spectral range with a spectral resolution between 0.3 nm (High Resolution, HR) and 1.8 nm (Low Resolution, LR). The spectral sampling interval will be from 0.1 nm in the oxygen absorption bands (748-769 nm and 686-697 nm) up to 2 nm outside the atmospheric absorption bands. The FLEX mission will deliver data at a spatial resolution of 300 meters, with observations scheduled around 10:00 local time. It will have a swath width of 150 km and a repeat cycle of 27 days. FLEX will operate in tandem with Sentinel-3, working synchronously with the Sentinel-3 Camera 4 (nadir-looking), which has a 14-degree field of view. These characteristics make the validation of FLEX products complex and challenging,

especially for those with high dynamism and significant spatial variability.

1.2 Overview of the FLEX products

L1C products consist of radiometrically and spectrally calibrated TOA radiances for the FLORIS (HR and LR), Ocean and Land Colour Instrument (OLCI), and Sea and Land Surface Temperature Radiometer (SLSTR, nadir and oblique) spatially resampled and projected into the FLORIS HR focal plane geometry with appended information (quality flags, surface topography, geometry, spectral characteristics).

L2 products consist of surface apparent reflectance for the FLORIS instruments; at surface solar irradiance (direct and diffuse) at FLORIS spectral resolution; cloud probability mask; Land Surface Temperature; Fluorescence full spectrum (650-800 nm) and surface real reflectance spectra; Biophysical parameters and photosynthesis related parameters. All products are provided with appended uncertainties. Further information on FLEX product is available in Vicent Servera et al., (2026).

The full list of FLEX products, classified according to their categories, their validation strategies and expected validation stage during the mission is provided in Figure 1. The traceability of the validation according to the Mission Requirement Document (MRD) available at <https://earth.esa.int/eogateway/announcement-of-opportunity/flex-cal-val> is also reported.

Data type	Product Level	FLEX L1C/L2 variable	Units	Product Uncertainty (yn)	Validation (yn)	Validation Category	CEOS stage	MRD traceability (v3.0)
Core Science Data	L1B	floris, toa, radiance (from L1B radiance)	-	-	y	B, C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, toa, radiance	mW·m ⁻² ·sr ⁻¹ ·nm ⁻¹	y	n	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1A	floris, apparent reflectance	-	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2B	total, integrated, albedo	nm	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2B	slr_peak, values	nm	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2B	slr_peak, values	nm	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2B	slr_O2, bands, values	nm	y	y	A, C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2B	slr_emission, spectrum	nm	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	direct, irradiance	mW·m ⁻² ·sr ⁻¹ ·nm ⁻¹	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	diffuse, irradiance	mW·m ⁻² ·sr ⁻¹ ·nm ⁻¹	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
Supporting Science Data	L2C	leaf_chlorophyll_content	µg·cm ⁻²	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	fraction_of_absorbed_photosynthetically_active_radiance	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	leaf_chlorophyll_content	µmol·mol ⁻¹	y	y	A, B, C	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
Experience Data	L2A	land surface temperature	K	y	y	A	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	aerosol optical thickness	g·cm ⁻²	y	y	A	3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	columnar water vapor	g·cm ⁻²	y	y	A	3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	reversible energy dissipation	µmol·mol ⁻¹ ·s ⁻¹	y	y	A, B	1/2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	fluorescence quantum efficiency	µmol·mol ⁻¹ ·s ⁻¹	y	y	A, B	1/2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	electron transport rate	µmol·mol ⁻¹ ·s ⁻¹	y	y	A	1/2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	fluorescence escape probability	µmol·mol ⁻¹ ·s ⁻¹	y	y	B	1/2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	foliar PSI relative contribution	-	y	y	A	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	langmuir exponent	-	y	y	A	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	asymmetry parameter	-	y	y	A	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
Ancillary Science Data	L2B	floris, real reflectance	µmol·photon·m ⁻² ·s ⁻¹	y	y	A	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2A	integrated par	-	y	y	A	2	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L2C	cross calibration coefficients	-	y	y	C	-	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, central wavelengths	nm	y	y	B	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, fwhm	nm	y	y	C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, fwhm	nm	y	y	C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, fwhm	nm	y	y	C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, fwhm	nm	y	y	C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, fwhm	nm	y	y	C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090
	L1C	floris, spectral channel, fwhm	nm	y	y	C	2/3	MRD-FLEX-090, MRD-FLEX-090, MRD-FLEX-090

Figure 1. FLEX product categories and main characteristics of the L1C/L2 mission products subject to validation. CEOS validation stage and MRD requirements are also indicated

The FLEX Core Science Data (CSD) constitute the fundamental data products required to meet the mission objectives and scientific goals. These products (Top of Atmosphere – TOA radiances, apparent surface reflectance and fluorescence parameters) are directly traceable to the mission's core objectives and drive the definition of all associated requirements, including performance specifications. The requirements for the CSD products, together with their associated uncertainties, are defined in the MRD and undergo both verification and validation. The algorithms used to retrieve the CSD products have been developed and integrated within the end-to-end simulator. The corresponding Algorithm Theoretical Basis Document (ATBDs) are available and are kept up to date. Within the Payload Data Ground Segment (PDGS), CSD products are generated continuously, their quality is systematically monitored and ESA is responsible for ensuring their timely dissemination. All CSD products are fully described and detailed in the specific Cal/Val Plans and the validation infrastructure designed for the mission is specifically tailored to support these products. The mission

aims to achieve full validation in accordance with the Committee on Earth Observation Satellites (CEOS) Stage 3 during Phase E2.

In summary, the validation activities are related to the following products:

- L1B instrument calibration;
- L1C/L2: Geolocation at L1C and L2;
- L1: Vicarious validation of the radiometric and spectral calibration and FLORIS/OLCI cross calibration coefficients;
- L2A: Atmospheric characterisation and atmospheric correction parameters (surface apparent reflectance and at-surface solar irradiance) and Land Surface Temperature;
- L2B: Fluorescence parameters;
- L2C: Biophysical and photosynthesis-related parameters

2. FLEX Cal/Val framework

2.1 Validation concept

There are no satellite fluorescence products with a better geometric resolution than FLEX, and therefore it is not possible to rely on comparisons with other satellites for an upscaling of the field measurements. In this context, the traditional bottom-up validation approach based on satellite data cannot be fully pursued with FLEX.

The overall FLEX strategy and plan for the validation of the satellite's operational products uses a requirements-based approach to ensure by traceability to the mission requirements that all parameters relevant for the operational product validation have been adequately covered.

The most important points considered for the FLEX cal/val are represented on the definition of i) site requirements according to the specific product (e.g. spatial homogeneity, acceptable topography, meteorological conditions...); ii) ground instruments and available networks (spectral/atmospheric, imaging vs non imaging system); iii) site classification and sites category (e.g. one/two or more components, core vs support sites); iv) sampling approach protocols (e.g. platforms, single/multiple points, image sampling); v) definition of the validation approach (direct, indirect, satellite inter-comparison) and validation stage; vi) set-up of cal/val tools, processing chain, traceability and match-up uncertainty.

Overall, FLEX will utilise a wide range of methodologies in calibrating and validating the mission science products, these include: i) in situ networks; ii) tower, drone, and airborne-based; iii) natural targets; iv) satellite products; v) model-based products; vi) field experiments and vii) simulated scenes (in Development Phase only).

Validation schemes and letters in Figure 2, refer to the type of validation approach and, for FLEX missions, we can consider different categories. Regarding the sampling approach, the different shapes in Figure 2 are related to the different type of measurements that can be done encompassing field spectroscopy and image systems. Regarding the instruments and the measurements, we should consider SIF-p: Ground spectrometers for top of canopy radiance, spectral reflectance and fluorescence; At: Instruments for atmospheric characterization; Dr: Unmanned Aerial Vehicle (UAV (or robotic) spectroscopy); Air: Airborne sensors; E: Additional instruments for environmental measurements (spectral, ecological...); EO: Earth Observation data. We designed a constellation of different sites (core, support ...), which provide data suitable for each product and validation scheme. For example, the key in Figure 2 indicates the core site's typology. Finally, the target type is classified according to site requirements and ecosystem shape, evaluating spatial characteristics.

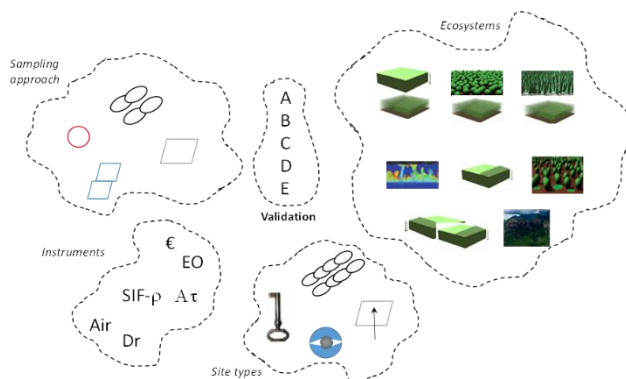


Figure 2. Main components of the FLEX validation process for L2 products

2.2 Reference networks and campaigns

Since FLEX will provide global products, validation activities will be conducted in a wide range of global climate and vegetation conditions. Product accuracy will be assessed over a widely distributed set of locations and time periods via several ground-truth and validation activities. The input data for validation originates from different ground networks, field campaigns, as well as satellite data. The Radiometric Calibration Network (RadCalNet) and VICALOPS sites will be exploited for L1C TOA radiances products, while AERONET (AERosol RObotic NETwork), HYPERNETS and the upcoming International Network of Sun Induced Fluorescence (INSIF) will be exploited for L2 reflectance and fluorescence products. Sentinel 2 will be used for geolocation validation of the L2 products. For vegetation biophysical products and surface temperature, the LAND VALidation (LANDVAL), the Ground-Based Observations for Validation (GBOV), the Surface Radiation Budget Network (SURFRAD) Atmospheric Radiation Measurement (ARM) and the Advanced Surface Temperature Radiometer Network (ASTeRN) and the United States Climate Reference Network (USCRN) will be exploited. In addition to the exploitation of the reference networks, dedicated field/airborne campaigns will be conducted in both Commissioning and Routine Operation phases.

2.3 Field sites nomenclature

For FLEX, we propose a constellation of different sites, which provides data suitable for the different validation approaches: *RadSites*, *SpecSites* and *CrossCalSites*. These are non-fluorescent sites to address radiometric/spectral cal/val of L1B, L1C and L2A and include RadCal network and CEOS VICALOPS sites; Desert locations and sites devoted to determining baseline and offset error (zero fluorescence) and to evaluate the FLORIS performance characterisation and recalibration.

GeolocSites. These sites will be used for co-registration and geolocation validation for L1C FLORIS TOA radiance vs OLCI geolocation and for L2 FLORIS reflectance vs Sentinel 2 geolocation validation.

AtmoSites. These sites correspond to AERONET sites, and they are used for direct validation category A (Section 2.4) of atmospheric parameters at L2A.

SolarSites. These sites are established according to the current worldwide distribution of the HYPERNETS and INSIF (mainly FLOX instruments). Direct comparison according to category A can be performed to evaluate solar irradiance products.

RefSites. These sites are used for the direct validation A of apparent reflectance and currently they correspond to the LAND HYPERNET and INSIF sites.

FluoSingleSites. These sites are core sites for fluorescence products, and they belong to the INSIF network. They are established research sites explicitly dedicated to continuous monitoring of SIF (and reflectance). Core sites are characterized in term of instrument and spatial uncertainties. These sites are also used for satellite data inter-comparison validation activities. They can be homogeneous or heterogeneous and can be sampled by different approaches and scaling techniques (i.e. field spectroscopy, multiple spots, and mapping – Section 2.4). These sites will provide the highest accuracy and best error budget. These sites will allow validation approaches for L2A, L2B and L2C according to the categories A, B (and C for some biophysical parameters) and could match with other networks sites (e.g. ICOS, S3 networks).

FluoMultiSites. These sites correspond to sites where the validation is performed by using multiple points measurements collected by field spectroscopy techniques in dedicated campaigns, with instruments on board of UAV-like platforms (category A based on drones/zip line-like systems).

FluoImageSites and *FluoSatSites*. Correspond to sites where the validation is performed by using an airborne system (category A) in dedicated campaigns or other satellite data. These sites can correspond to FluoSingleSites and FluoMultipleSites in case they are covered by an airborne acquisition. Satellite sites respond to the validation approach by category C (mainly by using TROPOMI) and are selected based on existing networks with different purposes (e.g. Orbiting Carbon Observatory, OCO-2 networks).

BioDirectSites, *BioIndirectSites* and *BioImageSites* dealing with biophysical or functional vegetation parameters (L2C). BioDirectSites correspond to LANDVAL and GBOV networks and allow for Validation A. BioIndirectSites correspond to INSIF sites and are used for validation B. BioImageSites approximately correspond to FluoSatSites and allow for Validation category C.

BioCampSites. Correspond to sites used for the validation of vegetation L2 products collected in specific field/airborne campaigns.

LSTSites are used for L2A - LST validation. Ground instrument uncertainty is provided, and validation A will be conducted.

SupportSites are sites defined on demand after the ESA Announcement of Opportunity. These sites can address different types of validation.

At the current stage, 714 sites belonging to the reference networks described in Section 2.2 and classified according to this nomenclature have been identified for FLEX validation.

2.4 Validation categories, sampling approach and scaling

Different validation categories will be exploited, encompassing direct, indirect approaches and inter-comparison with space products, with the aim to fully validate FLEX products according to the CEOS guidelines. For the FLEX mission, the following categories (Guillevech et al., 2014) are considered:

Category A: Comparison of satellite with in-situ measurements:

This is the traditional and most straightforward approach to validating any given parameter. It involves a direct comparison of satellite-derived value with collocated and simultaneously acquired measurement from ground-based instruments. The approach is basically a direct validation using a bottom-up approach (e.g. Morisette et al. 2006). This type of validation is performed over different sites, using single spot spectral measurements, multiple points or mapping measurements with

well-characterised instruments and defined uncertainty of the retrieval;

Category B: Indirect validation through numerical Radiative Transfer Model (RTMs, i.e., radiance-based validation technique). This approach can be exploited where availability of fluorescence monitoring is limited or absent, but also at core and support sites. This technique uses top-of-atmosphere data in conjunction with a radiative transfer model to propagate TOA reflectance for L1B and to simulate ground SIF using field data and atmospheric profiles of air temperature and water vapour content;

Category C: Inter-comparison with space products. A wide variety of spaceborne instruments collect similar data and many provide operational products. Inter-comparison of products from different satellite instruments can be valuable for determining the quality of a parameter, given that all the systems use the same (or compatible) retrieval algorithms.

Category D: Time series analysis. Analysing time series of satellite data over a temporally stable target site allows for identification of potential calibration drift or other issues of the instrument that manifest themselves over time. Furthermore, problems associated with cloud contamination for example may be identified from artefacts evident in the time series. Care will be taken in distinguishing between instrument-related issues such as calibration drift and real geophysical changes of the target site or the atmosphere.

Category A and B represent the most important approaches to be used for FLEX. Category C is not fully considered for fluorescence, since the main fluorescence products from FLEX do not have similar counterparts in other satellite products to compare against. Lack of similar or higher spatial resolution satellite fluorescence data do not allow to apply a full bottom-up approach strategy. FLEX will collect data at an unprecedented spatial resolution, and it may become a future benchmark for evaluating other fluorescence products derived from other space missions. A network of sites of vegetated areas distributed worldwide could be harmonised with, for example, Tropospheric Monitoring Instrument on Sentinel-5 Precursor (TROPOMI) and OCO-2 estimates for time series/stability comparison. However, this comparison approach can be also done for larger areas (such as corn belt, forest, savannah type...) to evaluate consistency in terms of spatial pattern, bias, completeness etc.

Finally, due to the temporal resolution of FLEX, the Category D may not be fully adequate for L2B product evaluation.

The terminology related to the field sampling approach can be defined according to the following criteria. *Single spot* field spectroscopy measurements (tower-based, at different heights) with continuous non-imaging systems to cover the temporal domain at the core site; *Multiple spot* measurements (e.g., from UAVs spectroscopy or other robotised platforms) would be useful for acquisition of sparse measurements in the spatial domain; *Mapping measurements* (e.g. from airborne or intermediate platforms with imaging systems) to cover relatively large areas. Alternatively, RTMs have been used to build reference data over whole targets by utilising a tool such as the Scene Generator Module (SGM) developed for the FLEX. Region Of Interest (ROI) are instead expected for satellite intercomparison or for the qualitative products evaluation at the Image Sites.

For L2 products, different aggregation schemes and appropriate transfer functions are currently implemented to scale up field measurements at pixel level. A kernel of 2x2 or 3x3 pixels is considered in the match-up of L2 products. L2 sites are characterised in terms of spatial homogeneity and representativeness. The spatial homogeneity is quantified computing the coefficient of variation (Cv) of the Near-Infrared Radiance of vegetation (NIRv_rad) from Sentinel-2, in selected

ROIs. The spatial representativeness is instead quantified computing the relative coefficient of variation (Rcv) according to Roman et al. (2009) and spatial representativeness index, comparing internal and external ROIs with regard the measurement point. The knowledge of the spatial homogeneity and representativeness allow to determine the scaling approach and to define the area to be sampled and to characterize the spatial uncertainty. For defining the location on ground to install the instrument for single point measurements we consider the most representative location within hypothetical 300 × 300m areas. A single point per area is identified using a Centroid-Weighted Mean Representative Point algorithm based on NIRv_rad (e.g. from Sentinel-2 at 10 m).

In case of multiple points sampling, a simple random sampling (hypothesis: normal distribution), is defined finding the minimum number of sampling points (n) according to Cochran (1977):

$$n = \left(\frac{Z_{\alpha/2} * \sigma}{E} \right)^2, \quad (1)$$

where $Z_{\alpha/2}$ = the desired confidence level (e.g., 1.96 for 95% confidence)

σ = the standard deviation of the selected area

E = the error on the mean value (e.g. for SIF 0.1 mW m⁻² sr⁻¹ nm⁻¹)

For optimized spatial sampling, giving a fixed number of sampling points, an optimal location of the points according to an optimization approach based on genetic algorithms has been defined (Holland, 1975). Further information on sampling strategy, temporal uncertainty, guidelines and protocols for multiple point measurements is available in Rossini et al., (2026).

2.5 Uncertainty analysis

The ground measurements and the overall uncertainty is following a metrological approach (SI-traceable) exploiting the Fiducial Reference Measurements (FRM) protocols by a suite of independent, fully characterized, and traceable ground measurements outlined by the GEO/CEOS Quality Assurance framework for Earth Observation and endorsed by CEOS (QA4EO, <https://qa4eo.org/>). QA4EO provides a metrological framework, which includes comparisons such as those used in satellite validations. These criteria include the establishment of documented evidence of uncertainty with respect to a community-agreed reference and specific tailoring to the needs of a satellite mission and are good cal/val practices adopted in ESA satellite missions.

The validation of FLEX products consider that both the FLORIS products and the ground truth measurements have inherent uncertainties and variances due to several factors. These include intrinsic instrument accuracy caused by random and systematic noises, the retrieval approach, natural spatial variability of target properties at different spatial scales at the time of acquisition, environmental conditions during ground truth measurements and finally error propagation in retrieval algorithms or radiative transfer models. The comparison metric are based on the difference between the satellite product and the reference measurement by considering their combined uncertainties and exploiting the Equivalence Ratio (EN). Spatial homogeneity and representativeness of the point measurements will be assessed by using statistical metrics and the scaling strategy at FLEX scale will be based on the use of transfer functions. This uncertainty approach will be exploited within the cal/val tools (Section 4.1).

The FLEX matchup is evaluated considering three primary sources of uncertainty:

- 1- Uncertainties associated with the FLEX product (u_x), which include errors originating from different sources and they are typical for of each product;
- 2- Uncertainties in the reference data (u_y) (instrument precision, measurement protocols, retrieval approach...);
- 3- Uncertainties introduced by the match-up process itself (u_{comp}) (spectral, spatial, angular and temporal discrepancies between satellite and in situ measurements and consider differences in sampling geometry and scene heterogeneity).

When validating satellite products, a match-up often refers to the comparison between satellite data and independent field measurements acquired under conditions as similar as possible in time, spatial location, and observational geometry. The match-up uncertainties arise from multiple factors and should be minimized wherever feasible.

The most fundamental comparison metric is to check whether the difference between the satellite product and the reference measurement is within the combined uncertainties:

$$|x - y| < k\sqrt{u_x^2 + u_y^2 + u_{comp}^2}, \quad (2)$$

where x = measurand (i.e. the quantity intended to be measured, here the FLEX products)
 y = is the measurand derived from the reference product

u_x, u_y = standard uncertainties on those measurands associated with the comparison itself

u_{comp} = standard uncertainty associated with the comparison itself

k = the coverage factor for the appropriate confidence interval (for FLEX $k = 2$ to have a confidence interval of ~95 %).

Equation (2) can also be rewritten as an Equivalence Ratio and at 95% of confidence level, the FLEX product and the in-situ data are in statistical agreement if:

$$E_N = \frac{|x - y|}{\sqrt{u_x^2 + u_y^2 + u_{comp}^2}}, \quad (3)$$

An equivalence ratio $E_N < 1$ (for $k = 2$) suggests that the two measured values agree within their uncertainties, while a larger E_N ratio suggests that at least one uncertainty is underestimated.

In the absence of uncertainty information, there are a wealth of validation metrics that have been used in the literature (e.g. Loew et al. (2017) and that will be used to provide useful validation information (i.e. Root Mean Square Error, mean residual and variance or standard deviation around mean measured value).

3. The FLEX Collaborative Platform

Dedicated tools and resources are being developed to support the validation and quality assessment processes for L1C and L2. An interactive portal (i.e. the FLEX Collaborative Platform, CP), is under development and it is expected to provide important validation support functionalities. The FLEX CP will provide analysis tools to enable assessment of quality indicators from specific products and address any specialized data processing requirements. The functional architecture of the FLEX CP will provide analysis of cal/val tools that enable generation and

assessment of quality indicators from specified products and by accommodating special data processing needs (Figure 3). External ancillary data including Cal/Val data from field campaigns, in situ networks, and special target data sets provided by the validation teams are ingested into the Cal/Val Database.

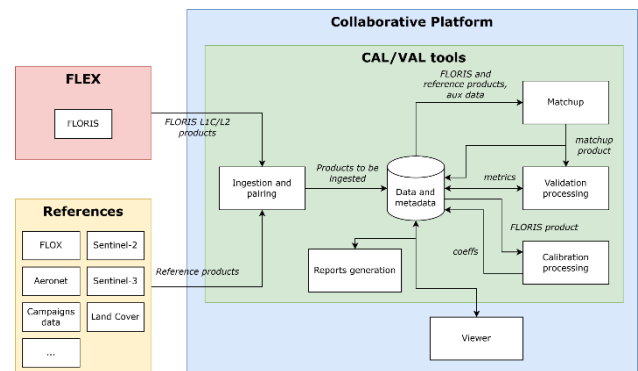


Figure 3. Architecture of the Collaborative Platform

4. Validation timeline

The FLEX Cal/Val Plan includes pre-launch and post-launch activities starting in Phase C/D and continuing after launch and commissioning through the end of the mission (Phase E).

4.1 Validation during Commissioning

The nominal duration of the Commissioning phase will be 3 months. The objective of FLEX Validation during the Commissioning Phase (E1) is to perform an assessment of the quality of FLEX Level 1B, C and selected Level 2 products by comparing them to independent data. It is expected that at least a preliminary quality assessment of most of L2 products shall be given at the In-Orbit Commissioning Review (IOCR).

Overall, the objectives of the Cal/Val activities during the Commissioning phase are:

- to calibrate all instruments;
- to tune all L1 processing parameters;
- to verify all L1 processing functions (i.e., geolocation, datation etc.);
- to perform a validation of all L1 products;
- to provide a preliminary quality assessment of L2 products.

The FLEX products will be evaluated using different approaches and by employing various types of data and reference networks. Figure 4 shows the validation activities in E1 for L1 and L2 products. i) Systematic Quality Control (QC) tests and Quality Indicators (QI) for L1C/L2; ii) Ad-hoc Quality Assessment (QA) for L2; iii) Validation by cal/val tools (using L1B tools, reference networks and campaigns data) for L1B/L1C and L2; iv) In-flight Performance Assessment Modules used for the assessment of L0/L1B performance with in-orbit data.

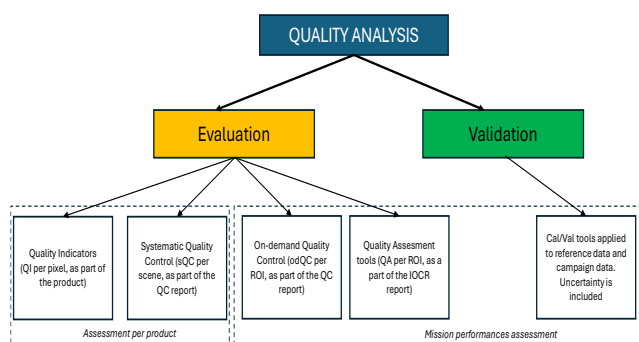


Figure 4. Approaches to evaluate the quality and to validate FLEX products during Commissioning

During the commissioning phase, when using real data, the QI and QC are expected to fine tune thresholds and define ROI where some of these tests are expected to monitor product values. Although most configurations are expected to happen prior to launch, some fine tuning is needed during commissioning and further during operation. During the commissioning, the QI and QC will be investigated, and their analysis will provide support in anomaly management.

In addition to QI and QC, it is planned to perform several QA tests to check that the values for reflectance, fluorescence, or biophysical parameters over certain surfaces are consistent with expectations. Although tools for ad hoc QA are expected to be run on the CP to take advantage of data accessibility, they are not expected to become part of the industrialized tools used systematically by the operators. The *ad hoc* quality assessment together with the systematic QC and early validation by the cal/val tools, allows to quickly assess the product quality.

Finally, the cal/val tools are automated tools that enable the validation of FLEX products within the CP. Currently 18 automatic cal/val tools for L1C and L2 products are in place and will be integrated in the CP and available from the commissioning phase (Figure 5). Cal/val tools can operate on reference networks as well as field/campaign data. As previously mentioned, cal/val tools include uncertainty estimates and follow a metrological approach. Spatial, temporal, angular and spectral uncertainties are considered wherever possible.

Cal/Val tool name	Cal/Val type	Product Level	Target	Cal/Val Description	Input Data
VICALOPS	V	L1B	L1B - Radiometric Validation	L1B Radiometric relative calibration and monitoring (FLORIS TOA reflectance via PICS and DCC)	PICS and DCC sites
L1C_SPECA	V	L1C	L1C - Spectral Validation	L1C spectral (barycentre and FWHM) validation through vicarious methods	RadCalNet, PICS
L1C_RADCAL	C	L1C	L1C - FLORIS TOA Radiance cross calibration	L1C FLORIS/OLCI cross calibration through vicarious methods	PICS
L1C_GEOLOC	V	L1C&L2	L1C - Coregistration	L1C FLORIS TOA radiance vs OLCI geolocation validation	Sentinel-3
L2_GEOLOC	V	L2C&L2	L2 - Geolocation	L2 FLORIS reflectance vs Sentinel-2 geolocation validation	Sentinel-2
L2A_AtmChar	V	L2A	L2A - Atmospheric parameters	L2A atmospheric characterisation parameters (CWV, AOD, ANG, ASY)	AERONET
L2A_AtmCorr	V	L2A	L2A - Solar Irradiance and apparent reflectance	Single point comparison	HYPERNETS, INSIF, S2
L2A_LST	V	L2A	L2A - Land Surface Temperature	Validating LST using in-situ data from international networks	In-situ data: SURFRAD, USCRN, ARM, LAW, ASTeRN, emissivity
L2B_SIF_FullSpec	V	L2B	L2B - Fluorescence products	Processing ground reference point data from radiances to L2B products.	Radiance data (e.g. from INSIF)
L2B_SIF_SinglePoint	V	L2B	L2B - Fluorescence products	Single point comparison	Output of L2B SIF tool or external SIF data, Sentinel-2 L1C
L2B_SIF_MultiplePoints	V	L2B	L2B - Fluorescence products	Multiple sampling points tool	FLOX, AirFLOX, Sentinel-2 L1C
L2B_SIF_Airborne	V	L2B	L2B - Fluorescence - Red&FarRed	Airborne tool	Airborne image data
L2B_SIF_Satellite	V	L2B	L2B - Fluorescence - FarRed	Comparison between FLORIS L2B data and TROPOMI L3 SIF products	Sentinel-5P TROPOMI
L2C_Bio_Campaign	V	L2C	L2C - Vegetation Biophysical	Validating FAPAR, LCC, LCAI, LAI using in-situ from field campaigns	In situ measurements, Sentinel-2 L1C data, ROI
L2C_Bio_Network	V	L2C	L2C - Vegetation Biophysical	Validating LAI and FAPAR using in-situ from international networks	GBOV LP03 and GBOV LP04 reference data
L2C_Photosyn_Campaign	V	L2C	L2C - Novel Photosynthesis Parameters	Validating FQE, ETR, RED using in-situ from field campaigns	PAM measurements, FioX measurements, ROI
L2C_Veg_Satellite	V	L2C	L2C - Vegetation Biophysical - LAI and FAPAR	Validating L2C vegetation parameters using other satellite data	CMS, Sentinel-2 L2A
L2C_Veg_RTM	V	L2C	L2C - Vegetation Biophysical - LAI, FAPAR, REP, FQE	Validating L2C vegetation parameters using reference data generated by RTMs	INSIF

Figure 5. Cal/val tools: names, target variables and short description

4.2 Validation during Routine Operations

After the completion of the commissioning, i.e. in Exploitation/Routine Phase (E2), FLEX platform and instrument settings are expected to have stabilised, and mainly routine characterization, calibration and monitoring is expected to be required nominally. The Phase E2 validation program will focus on the completion of the initial validation of products that could not be validated during phase E1 (typically the L2 products), and

on the long-term validation where more statistical relevance is obtained.

For the Long-term Validation Objectives in Phase E2, the first objective is to complete the preliminary validation of all FLEX data products. As statistics build up, the second objective will be achieved, namely, to obtain consolidated long-term validation results. A third objective applies after any improvement of algorithms followed by (priority) reprocessing of the collocated FLEX data, namely, to perform a “delta validation2 by repeating the inter-comparison with collocated correlative observations but now with the improved subset of collocated FLEX products (reprocessed with the new algorithms) and the (consolidated) correlative data. During Phase E2, the frequency of reporting will be reduced.

In phase E2, automatic monthly reports are planned for all variables, while every four months more complete “manual” reports will be produced.

5. Examples of Cal/Val tools for Fluorescence

Since it is not possible in this paper to describe all the cal/val tools that have been developed, only one example will be provided, illustrating the work carried out for the validation of fluorescence, which represents the mission’s primary CSD. Due to the scale mismatch between satellite and in situ measurements, acquiring in situ SIF measurements representative of a satellite footprint is challenging. FLEX will generally observe a variety of land covers within a footprint, with associated temporal and spatial variations of SIF, due to the spatial heterogeneity of the land surface at these scales. We therefore propose different strategies and different cal/val tools to perform the validation activities of L2B FLEX products. These strategies are based on direct validation category and comparison and rely on single points, multiple points or image based and satellite comparison by using direct scaling or exploiting transfer function to scale up field measurements to FLEX spatial scale. We have essentially defined four cal/val tools, each of which may include specific variants. These tools operate using different types of input data and provide a validation output that includes an associated uncertainty. Further information is available in Rossini et al., (2026)

Single Point cal/val tool. i) Single-point validation: direct comparison between temporally aggregated in situ single-point SIF measurements and FLORIS L2B data; ii) Transfer function-based validation: single-point SIF measurements are scaled using a transfer function derived from Sentinel-2 images to generate a high-resolution reference map, then aggregated and compared with spatially matched FLORIS L2B data.

Multiple Points cal/val tool. i) Two-point validation: SIF measurements from 2 distinct land cover types within the same FLORIS footprint are temporally aggregated and weighted by land cover fractions, then compared to FLORIS L2B data; ii)

Multi-point validation: spatially distributed SIF measurements from multiple sampling points are aggregated and compared with spatially matched FLORIS L2B data; iii) Transfer function-based validation: two- or multi-point SIF measurements are scaled using a transfer function derived from Sentinel-2 images to generate a high-resolution reference map, then aggregated and compared with spatially matched FLORIS L2B data.

Airborne cal/val tool. i) Image-based validation: airborne high spatial resolution SIF products are spatially aggregated and compared with spatially matched FLORIS L2B data.

Satellite cal/val tool. Satellite comparison: comparison between FLORIS L2B data and TROPOMI SIF products.

An example of results obtained by the UAV Multiple Points schemes with transfer function applied on HyPlant data collected

on 11-21 June 2025 in Italy, aggregated at FLEX spatial resolution and using fluorescence data acquired in different crops with FLOX instrument is presented in Figure 6.

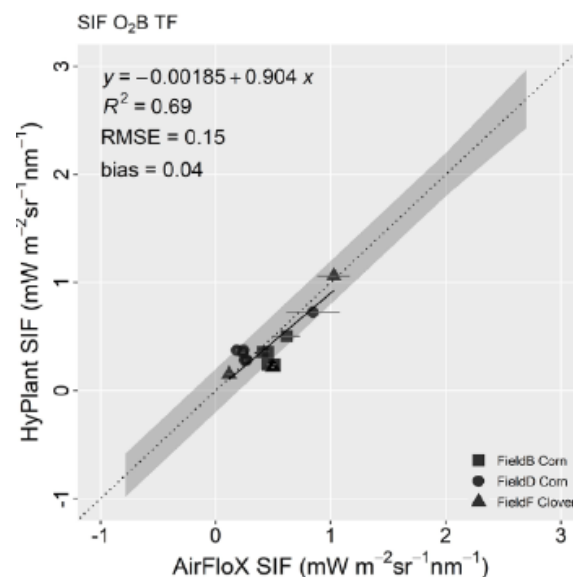


Figure 6. Multiple-Point validation for SIF O₂B using optimized UAV sampling (i.e. AirFloX system), including the application of transfer functions and daily corrections.

6. Conclusions

The FLEX mission provides a wide range of products, supported by calibration and validation plans that have already been prepared and made available to the user community. For each product, a dedicated validation strategy has been designed to ensure that the approach is tailored to the specific characteristics of the data.

During phases E1 and E2, different validation approaches will be implemented, covering QC, QI, and QA activities and making use of the various cal/val tools developed for the mission. These tools will be integrated into the Collaborative Platform and accessible to users.

Validation sites have been identified based on established reference networks, ensuring robust and representative coverage. The reference field measurements collected at these sites are SI-traceable and aligned with FRM best practices, guaranteeing high-quality and reliable reference data.

Finally, extensive documentation has been produced to support the FLEX community in validating the mission products and in making effective use of the available procedures and tools.

Acknowledgements

This study was promoted by ESA within the framework of the project the FLEX Data Innovation and Science Cluster (DISC).

References

- Cochran, W. G., 1977: Sampling Techniques (3rd ed.). John Wiley & Sons.
- Colombo, R., 2026: FLEX Calibration/Validation Plan Overview, available at

<https://earth.esa.int/eogateway/announcement-of-opportunity/flex-cal-val>.

Guillevic, P. C., Biard, J., Hulley, G. C., Privette, J. L., Hook, S. J., Oliso, A., Göttsche, F.-M., Radocinski, R., Román, M. O., Yu, Y., and Csiszar I., 2014: Validation of Land Surface Temperature products derived from the Visible Infrared Imager Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements. *Remote Sensing of Environment*, 154, 19–37.

Holland, J. H., 1975: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. *University of Michigan Press*, Ann Arbor.

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

Morisette, J. T., Privette, J. L. Justice, C. O., 2002: A framework for the validation of MODIS land products. *Remote Sensing of Environment*. 83: 77-96

Roman, M. O., Schaaf, C. B., Woodcock, C., Strahler, A. H., Yang, X., Braswell, R., Goulden, M. L., 2009: The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes. *Remote Sensing of Environment*, 113, 2476–2498.

Rossini M., et al., 2026: L2B Calibration/Validation Plan, available at <https://earth.esa.int/eogateway/announcement-of-opportunity/flex-cal-val>.

Vicent Servera J., et al., 2026: L1C & L2 IPF Product Format Definition, available at <https://earth.esa.int/eogateway/announcement-of-opportunity/flex-cal-val>.