|  |  |
| --- | --- |
|  | **Validation Methodology for Copernicus Land Surface Reflectance Data Products Over FRM-Compliant Field Sites (VM)**VERSION 1.0National Physical LaboratoryUniversity of SouthamptonEOLab05 June 2020 |
| Related image | This document was produced as part of the ESA-funded project “Fiducial Reference Measurements for Vegetation Phase 2 (FRM4VEG 2)” under ESA contract number: 4000129823/20/I-NS |

##### Authors

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Name** | **Organisation** | **Date** |
| **Written by** | Niall Origo | NPL | 05/06/2020 |
|  | Rosalinda Morrone  | NPL |  |
|  | Joanne Nightingale | NPL |  |
|  | Luke Brown | University of Southampton |  |
| **Reviewed by (consortium)** |  |  |  |
| **Approved by (ESA)** |  |  |  |

##### Version History

|  |  |  |
| --- | --- | --- |
| **Version** | **Date** | **Publicly available or private to consortium** |
| 1.0 | 05/06/2020 | Private Consortium |

##### Table of contents

[Authors 1](#_Toc42263176)

[Version History 1](#_Toc42263177)

[Table of contents 2](#_Toc42263178)

[List of figures 3](#_Toc42263179)

[Acronyms 3](#_Toc42263180)

[1 Introduction 5](#_Toc42263181)

[1.1 Purpose and Scope 5](#_Toc42263182)

[2 Land Surface Reflectance Data Products, Terminology and State-of-the-Art 6](#_Toc42263183)

[2.1 Context 6](#_Toc42263184)

[2.2 Terminology 7](#_Toc42263185)

[2.2.1 Albedo 7](#_Toc42263186)

[2.2.2 Bidirectional Reflectance Distribution Function (BRDF) 7](#_Toc42263187)

[2.2.3 Hemispherical-Directional Reflectance Factor (HDRF) 8](#_Toc42263188)

[2.3 State-of-the art methods for satellite surface reflectance validation 8](#_Toc42263189)

[2.3.1 Direct validation 8](#_Toc42263190)

[2.3.2 Indirect validation 11](#_Toc42263191)

[3 FRM Validation Methodology 12](#_Toc42263192)

[4 Validation configurations 13](#_Toc42263193)

[4.1.1 Scale 13](#_Toc42263194)

[4.1.2 Homogeneity 13](#_Toc42263195)

[4.1.3 Temporal changes 14](#_Toc42263196)

[4.1.4 Viewing configuration 14](#_Toc42263197)

[4.1.5 Illumination conditions 15](#_Toc42263198)

[4.1.6 Spectral considerations 15](#_Toc42263199)

[4.2 In situ data uncertainty and propagation 16](#_Toc42263200)

[4.2.1 Field data uncertainty assessment 16](#_Toc42263201)

[4.2.2 Field data uncertainty propagation 18](#_Toc42263202)

[5 Satellite data pre-processing, quality control and uncertainty 19](#_Toc42263203)

[5.1 Pre-processing and quality control 19](#_Toc42263204)

[5.2 Satellite product uncertainty 20](#_Toc42263205)

[5.2.1 Uncertainty of Sentinel 2 Level 2A reflectance products 21](#_Toc42263206)

[5.3 Satellite surface reflectance product validation gaps and challenges 22](#_Toc42263207)

[6 References 24](#_Toc42263208)

# List of figures

[Figure 1: Diagram of to illustrate the role that field measurements play in the validation of Analysis Ready Data (ARD). The left- hand side of the diagram illustrates the satellite-data process flow from raw data to ARD. The right-hand side shows the field data acquisition and data process flow [2] 7](#_Toc42094322)

[Figure 2: The validation procedure and possible considerations 13](#_Toc42094323)

# Acronyms

|  |  |
| --- | --- |
| **Abbreviation** | **Stands for** |
| ADC | Analog to Digital Converter |
| AERONET | Aerosol RObotic NETwork |
| AOD | Aerosol Optical Depth |
| AOT | Aerosol Optical Thickness |
| ARD | Analysis Ready Data |
| ASD | Analytical Spectral Device |
| ASRVN | Aeronet Surface Reflectance Validation Network |
| BHR | Bi-Hemispherical Reflectance |
| BRDF | Bidirectional reflectance distribution function |
| CEOS | Committee on Earth Observation Satellites |
| ECV | Essential Climate Variable |
| EOLAB | Earth Observation Laboratory |
| ESA | European Space Agency |
| ETM | Enhanced Thematic Mapper |
| FIDUCEO | Fidelity and Uncertainty in Climate Data Records from Earth Observation |
| FPI | Fabry-Perot Interferometer |
| FPI | Fabry-Perot Interferometer |
| FPP | FRM Protocols and Procedures |
| FRM | Fiducial Reference Measurements |
| FRM4VEG | Fiducial Reference Measurements for Vegetation |
| GER | Geophysical and Environmental Research Corporation |
| GPS | Global positioning system |
| GUM | Guide to the expression of Uncertainty in Measurement |
| HCRF | Hemispherical-Conical Reflectance factor |
| HDRF | Hemispherical directional reflectance factor |
| HYPERNETS | Hyperspectral Network |
| InGaAS | Indium gallium arsenide |
| JCGM | Joint Committee for Guides in Metrology |
| L2A | Level 2A |
| MEMs | Micro Electro-Mechanical Structures |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MSI | Multispectral Instrument |
| NASA | National Aeronautics and Space Administration |
| NIST | National Institute of Standards and Technology |
| NPL | National Physical Laboratory  |
| RADCALNET | Radiometric Calibration Network |
| RGB | Red green blue |
| ROI | Region-of-interest |
| S2 | Sentinel 2 |
| S2-RUT | Sentinel-2 Radiometric Uncertainty Tool |
| SR | Surface Reflectance |
| SURFRAD | SURFace RADiation Budget Network |
| SWIR | Shortwave-infrared |
| TOA | Top of Atmosphere |
| UOS | University of Southampton  |
| VM | Validation Methodology |
| VNIR | Visible near infrared |
| WGCV | Working Group on Calibration and Validation |
| WMO | World Meteorological Organization |

# Introduction

## Purpose and Scope

This document forms part of deliverable D-30 of the European Space Agency (ESA) project ‘Fiducial Reference Measurements for Vegetation (FRM4VEG Phase 2)’. The purpose of the document is to provide an overview of methodology for validating Copernicus land surface reflectance data products over vegetated FRM-compliant field sites. The document is organized into 4 key sections:

* **Section 1** provides a summary of the Validation Methodology – Surface Reflectance (VM-SR)
* **Section 2** provides an overview of land surface reflectance data products, the relevant surface reflectance terminology and a review of the state-of-the-art methods for the validation of satellite-derived land surface reflectance data products.
* **Section 3** summarises the FRM-recommended validation methodology for validating satellite derived land surface reflectance data products using FRM-compliant field site data. This section will discuss the requirements and challenges for validating moderate resolution data products as well as demonstrate the handling of various error sources and their relative impact on the validation methods and results.
* **Section 4** discusses the methods used to pre-process the satellite data, including aspects such as quality screening.

# Land Surface Reflectance Data Products, Terminology and State-of-the-Art

## Context

There are a significant number of derived Earth Observation products that share surface reflectance as a common starting point in their processing chains. Subsequently, these products tend to feed into numerous application areas such as land cover mapping and the derivation of biophysical essential climate variables (ECV). All these areas are reliant on a good quality surface reflectance product if the results of analysis utilising any of the downstream products are to be trusted. In order to do this, validation is needed to ensure that each product performs to its given specification [1]. In the context of the validation of surface reflectance products derived from satellites, it is key to make the distinction between the quantity estimated by the satellite product and that measured at the surface. In doing this, two separate processing chains are present which, ideally, ensure that the validation data is independent of the satellite product. This is one of the requirements for an FRM based validation. Figure 1, from [2] gives an example of the processing chains from the validation and satellite perspective.



Figure 1: Diagram of to illustrate the role that field measurements play in the validation of Analysis Ready Data (ARD). The left- hand side of the diagram illustrates the satellite-data process flow from raw data to ARD. The right-hand side shows the field data acquisition and data process flow [2]

As mentioned previously, validation of surface reflectance should be taken in the context of the validation of related products (e.g. albedo and BRDF). [3] lists a broad range of definitions for the most commonly used reflectance quantities. The authors of that paper make the key point that there is a large amount of poorly defined reflectance definitions in the literature. They raise the issue that potential users, such as surface modellers, may go on to utilise the data without reference to particular uncertainties specific to the provided definition. For discussion on the many reflectance definitions the reader is referred to [4] and for definitions related to in situ quantities a greater discussion is presented in the FPP-SR. Nevertheless, some key definitions, particularly for satellite products, are given below.

## Terminology

### Albedo

Albedo, or Bi-Hemispherical Reflectance(BHR),is the main term for energy balance models of the Earth’s surface. It corresponds to the ratio of the integrated energy incident on a surface from any angle, to the integrated energy reflected from that surface. For satellite products, it is generally white and black sky quantities that are generally produced [5], both refer to idealised conditions: white sky, where the illumination field is perfectly isotropic, and black sky, where the illumination is solely from direct radiation (i.e. a point source). Algorithms for the retrieval of surface albedo from satellite data require knowledge of the bidirectional reflectance of the Earth’s surface[6].

### Bidirectional Reflectance Distribution Function (BRDF)

The BRDF of a surface can be described as the function which relates the illumination incident on a surface to that reflected from a surface in any combination of illumination and reflectance angular configurations. BRDF is defined over infinitesimally small angles and therefore refers to the direct illumination and reflectance. BRDF is an intrinsic reflectance property of a surface[7].

### Hemispherical-Directional Reflectance Factor (HDRF)

The HDRF is the “*radiance reflected by the target surface into a specified direction relative to the radiance reflected by a lossless Lambertian reflector into the same beam geometry under ambient illumination conditions”* [8]. Due to the small field of view of satellite pixels, the reflectance products produced by those satellites are often considered HDRF (although strictly they are Hemispherical-Conical Reflectance factors - HCRF) when the diffuse component of the illumination has not been removed.

## State-of-the art methods for satellite surface reflectance validation

Satellite surface reflectance validation has generally been tackled in two broad, separate ways within the literature: direct validation using coincidentally collected ground data, and indirect validation using comparisons to other surface reflectance products. It can also be noted that there are some studies which have borrowed aspects of both (e.g.[9]). This section primarily focuses on the direct validation case because the goal of FRM4VEG Phase 2 is to a) work with ground data, and b) ensure that the ground data is independent of the satellite product processing. As such, a detailed description of direct validation activities and a shorter summary of indirect validation is given below.

### Direct validation

Direct validation involves comparison of physical measurements at the surface level (i.e. in-situ measurements) with the reflectance measurements provided by the satellite product. The measurements at ground level may be taken directly by research teams at a nominated location (i.e. during the field campaign) or might be derived from an automated network of sensors. The collection and processing of in situ data for validating surface reflectance products is discussed within the FPP-SR.

#### Field campaign measurements

Devices for the measurement of reflectance in the field have been under development since the 1970s. The most recent generation of such devices use portable and robust optoelectronic innovations such as micro electro-mechanical structures (MEMs), as charted in [10]. The most popular instruments that are manufactured for measuring spectral reflectance in the field in recent years are spectroradiometers, with VNIR-SWIR wavelength ranges, such as the ASD (Analytical Spectral Devices) range of products, though alternative instruments are produced by the Geophysical and Environmental Research Corporation (GER), Licor and CIMEL. An example of the operational use of the ASD Fieldspec 4, is detailed by [3]. Calibration of the ASD is conventionally carried out at the dedicated facility of the manufacturers using verifiable NIST (the US National Institute of Standards and Technology) -traceable standards of irradiance, reflectance and wavelength [11].

More review to follow…

#### Field campaign measurements with drones

A further innovation is the development of lightweight instruments capable of being mounted on drones. From a practicality perspective, the use of drone mounted sensors allows users to capture reflectance measurements over difficult terrain (e.g. water, tall canopies, etc.) as well as over normal terrain without disturbing the surface. Such measurements would otherwise only be possible using tower access above the canopy, and so would be very limited in scope. While the quality of the drone-mounted instruments is generally considered to be lower when compared to conventional manually operated spectrometers on the ground, improvements have been noted. For example, [11] found that measurements using the Fabry-Perot Interferometer (FPI) hyperspectral camera, when well-calibrated, showed promise against those of a calibrated irradiance spectrometer. A review outlining the development of drone-mounted surface reflectance instruments is provided by [11].

More review to follow…

#### Automated sensor measurements

To validate surface reflectance products across a range of sites simultaneously, use is made of networks of automated sensors, such as those described below. A running theme in such networks is to provide freely available open data with broad global coverage. These networks are made up of remote ground stations transmitting radiation fluxes (SURFRAD) and multispectral aerosol optical depth (AOD) estimates (AERONET). The general premise behind this approach is to compare the surface reflectance derived from atmospheric correction using the image derived AOD against the

equivalent (or MODIS as in the ASRVN case) using the station derived AOD. This assumes that the surface reflectance derived from the station AOD is the “truth”. SURFRAD, on the other hand, derives surface reflectance (in this case albedo) directly from the up and downwelling fluxes measured at the stations. These can then be compared against the product derived surface reflectance [1].

More review to follow…

* AERONET: A Federated Instrument Network and Data Archive for Aerosol Characterisation

The AERONET (AErosol RObotic NETwork) validation system [12] is a federation of ground-based remote sensing aerosol networks established by NASA and PHOTONS. It provides globally distributed observations of spectral aerosol optical depth (AOD), inversion products and precipitable water in diverse aerosol regimes. Sun photometer measurements of the direct solar radiation provide information to calculate the AOD. The network provides an open database or aerosol microphysical and radiative properties, and can be used to validate surface reflectance by atmospheric characterisation [12]. The key instrument is an automated spectral radiometer (the CIMEL CE-318 instrument) that is low maintenance, independently solar powered and transmits measurements for real time processing.

* RADCALNET: A Radiometric Calibration Network

RADCALNET [13] is a network of instrumental sites to allow calibration, inter-calibration and validation of sensors. It was established in 2013 by the Committee on Earth Observation Satellites Working Group in Calibration and Validation (CEOS-WGCV). The network consists of four test sites that automate in-situ measurements, together with estimates of top-of-atmosphere (TOA) reflectance. These sites are located at desert and agricultural sites across four continents – in the USA, France, Namibia and China. RADCALNET, which is a free and open access service, provides a continuously updated archive of TOA reflectances derived over a network of sites, with associated uncertainties, at a 10 nm spectral sampling interval, in the spectral range from 380 nm to 2500 nm. There are admittedly limitations as to how these data can be used for the in-flight assessment of spaceborne sensors as they are provided only at a 30-min interval for a nadir viewing configuration only. Although this limits the possibilities of matching up the data with spaceborne sensors with off-nadir viewing angles, it should allow for intercomparison with sensors viewing any of the sites with low viewing angles [14].

* HYPERNETS (not currently operational)

AERONET and RADCALNET are based on multispectral instruments requiring modelling associated uncertainties to cover all spectral bands of all sensors; they produce valuable data, but the instruments are expensive to acquire, install and maintain, and they require modelling, therefore incur additional uncertainties, to cover all spectral bands of all sensors. The objective of HYPERNETS [15] project is to develop a new lower cost hyperspectral radiometer and associated pointing system and embedded calibration device for automated measurement of water and land bidirectional reflectance and, subsequently, for validation of all optical bands on all satellite missions. Quality controlled data with associated uncertainty estimates will be provided automatically for the validation of all optical satellite missions. HYPERNETS is an attempt to take advantage of the progress in optoelectronic components to develop a network of less expensive, low maintenance sensors, with miniaturized hyperspectral spectrometers and with pointing systems adapted from those used for the cameras in advanced home security systems.

### Indirect validation

Indirect validation refers to the comparison of two, or more, different satellite-derived products. In recent studies (e.g. [16]) these indirect validations have not been conducted in isolation, but are supported by direct validation through other means.

[17] conducted an evaluation of the Landsat-5 TM and Landsat-7 ETM+ surface reflectance products that also follows a direct and indirect approach, firstly using the Aerosol Robotic Network (AERONET) dataset over 489 sites and secondly conducting a cross-comparison of Landsat and MODIS data collected on the same day. From the first comparison they found a broadly positive agreement, where the largest errors were in the blue band, which is dominated by atmospheric correction uncertainty. The cross-comparison between Terra MODIS and Landsat-7 ETM+ data produced the largest uncertainty in northern latitudes and in tropical evergreen forests. Furthermore, there were discrepancies between the Landsat 5 red band, and Landsat 7 ETM infra-red band, that they attributed to the radiative transfer model being not absolutely suited to account for forest canopies.

More review to follow…

# FRM Validation Methodology

Developing FRM methodologies is about ensuring that the processing chain of the satellite being validated is well understood in order to ensure that the FRM processing chain is independent of this. Likewise, by starting from the quantity of interest (i.e. the satellite product measurand), it is possible to begin laying out a processing chain which reaches the same quantity including the ancillary measurements that are required. Once this has been designed, uncertainty propagation is needed to transfer the uncertainties of the inputs, via the operations they undergo, to an uncertainty in the output (see FRM4Veg Overview Document).

Validation consists of three broad components: the test sensor (i.e. the quantity derived from a sensor we wish to validate), the reference sensor (i.e. the same quantity derived from a sensor which is known to be of better quality) and the comparison conditions [1]. Table 1 from [1] lays out some considerations which are reproduced in Figure 2.



Figure 2: The validation procedure and possible considerations

An FRM methodology may have to consider any number (or more) of the factors presented in Figure 2 to ensure that the appropriate quantities match. For example, ensuring that the validation data are collected at the same time as the satellite overpass means that the site surface and atmospheric conditions are consistent with the satellite overpass. However, differences in the nature of the measurements (e.g. scale, viewing configuration, etc.) mean that sampling strategies need to be employed to cover similar areas, while different fore-optics may be needed to more closely resemble the pixel viewing geometry of the satellite. Nevertheless, some of these factors may not be adequately accounted for, it is under these circumstances that additional uncertainty sources should be considered in the validation data uncertainty budget to include their impact on the comparison. In ensuring that the two quantities match, it is possible that the uncertainty of the validation data will be larger than that of the satellite sensor (depending on the uncertainty sources considered by the satellite product). This is simply because the uncertainty associated with the comparison conditions are attached to the in situ sensor. An example of this would be in needing to interpolate the Spectralon panel calibration illumination angles to a different angle to reflect field conditions (e.g. in [1]).

There are three sensors’ surface reflectance products being validated in FRM4VEG Phase 2: Sentinel-2, Sentinel-3 and Proba-V. The first and last ones were validated in Phase 1, while Sentinel-3 did not have a surface reflectance product at that time. In terms of the issues mentioned in Figure 2, the following subsections will discuss these in relation to these three sensors (which most the information summarised from[1]).

# Validation configurations

### Scale

There are two important considerations to make. The first is the number of pixels to be validated (e.g. single-pixel or average over several). The second is the spatial resolution of the satellite sensors. The answers to both of these determines the size of the area that needs to be measured on the ground. Sentinel-2, -3 and Proba-V all have different spatial resolutions meaning that if all are to be validated at the same site then it needs to be large enough to validate the sensor with the largest spatial resolution (Sentinel-3). Consequently, a sampling scheme must be derived in order to representatively cover the area of interest in a minimal time such that other conditions (e.g. atmospheric conditions, surface changes, illumination angle, etc.) have not changed substantially.

### Homogeneity

Homogeneity is not strictly a requirement of a surface area, in fact since many areas are heterogeneous, validation activities should consider these sites as well. However, the more heterogeneous the site, the greater the amount of sampling that must be done in order to capture a representative estimate of the surface reflectance. For products with large spatial resolutions (e.g. Sentinel-3) homogeneous surfaces are desirable if sampling time is likely to be prohibitive.

### Temporal changes

The measurement of surface reflectance for a single satellite pixel is near-instantaneous when compared to the measurements being acquired in the field for the same area. As such, the ground measurements need to be conducted as quickly as possible and as closely as possible to the overpass time. There are several factors that can change within the ground sampling time such as the solar zenith angle, atmospheric conditions, and the surface state and BRDF. Some of these are discussed in more detail in the remaining sections.

### Viewing configuration

There are several components to matching the viewing configurations that need to be considered. Firstly, the view zenith angle of the pixels covering the area of interest for the satellite sensor is determined by the location of the field site in the detector array, while the range of possible view zenith angles is determined by the sensor’s orbital characteristics and sensor design. This means that in many cases, satellite pixels over a validation area will be in off-nadir, which does not match the standard protocol for in situ sensors (e.g. ASD) which are to be measured at nadir. In this case, it is best to ensure that the satellite will overpass close to nadir or that the surface BRDF is known so that a correction can be applied. New ways of measuring off-nadir (e.g. filtering drone pixel arrays) would help to alleviate this.

The second factor to consider is that the surface reflectance measured by the satellite is coming from a small angular field of view for the whole area. By virtue of their proximity, in situ sensors will not be able to measure the whole area effectively in that same field of view. Drone mounted sensors offer a way of getting closer to the satellite estimate but ultimately it must be assumed that the BRDF is constant over the difference in field of view. Therefore, avoiding areas of pronounced change (like the hotspot) is preferred.

### Illumination conditions

As mentioned previously, from the perspective of the field measurements, the satellite acquisition of the site can be considered near-instantaneous. The solar zenith angle is another factor which changes the greater the difference in measurement time between the in situ sensor and satellite. Furthermore, the rate of that change depends on the latitude of the field site. Therefore, measurements that are made as close to the satellite overpass as possible will minimise this effect.

However, a second factor should also be considered and depends on the manner used to determine the surface reflectance in the field. For surface reflectance derived from panel measurements (e.g. comparison between measurements of a Lambertian panel and the surface), the calibration of the panel reflectance is usually determined at a specific illumination angle. If the solar zenith angle and the panel calibration illumination are different, then errors related to the non-Lambertianess of the panel will be incurred. These can be significant, as shown in [1].

For panel-based validation, the diffuse component of the illumination must also be characterised. Likewise, a diffuse panel calibration must also be undertaken in order to characterise the reflectivity of the panel from non-direct illumination sources. This means that additional sensors (to measure e.g. diffuse-direct ratio sensors, AOD, etc.) may be needed in order to acquire additional data during the overpass).

### Spectral considerations

The spectral characteristics of the satellite sensor must also be considered when attempting to match with the target quantity. This means that the in situ instruments must have a suitable number of spectral bands within the spectral region acquired by the satellite sensor. The in situ data must then be convolved with the satellite spectral response.

The specifics of the in situ data collection, instrument operation, etc. are discussed in detail in the FPP-SR.

## In situ data uncertainty and propagation

To derive in-situ measurement uncertainties and ensure traceability, the approach adopted within the FRM4VEG project is based on that proposed by *Working Group 1 (WG1) of the Joint Committee for Guides in Metrology (JCGM)* [18]*.* Basic concepts and vocabulary aboutthe expression of uncertainty in the measurement of a physical quantity may be found in the *FRM4Veg Overview document*.

To derive a reasonable estimate of the uncertainty of the surface reflectance quantity retrieved from the field it is necessary to create a measurement function. The concept of the measurement function used in this project is borrowed from the Fidelity and Uncertainty in Climate Data Records from Earth Observation (FIDUCEO) project [19]. Since this is described in detail in the FPP-SR document it will only be summarised here. The measurement function should ideally completely describe the path from all the data inputs and the operations they undergo to create the final surface reflectance estimate. There are usually some unknown components or incomplete modelling which means that this cannot be done exactly, but that is the aim at least.

If the previous sections have been considered and the validation has been designed based on the specific issues raised then it is likely that most of the measurement function can be created prior to making the measurements, with only the parameters being filled in after. This leaves two components to be determined, the parameterisation of the uncertainty and the uncertainty propagation method. These will be discussed below.

### Field data uncertainty assessment

Uncertainty assessment goes beyond determining a single value for the uncertainty, it should also capture the way that uncertainty should be treated when it is combined with other uncertainties (e.g. due to correlation of the underlying effects, the probability distribution it describes, etc.). For field measurements there is a distinct separator between uncertainties brought about from pre-field activities (mainly calibration and characterisation), those from in-field activities (mainly sampling), and post-processing (e.g. interpolations, etc.).

The uncertainties derived from the calibration and characterisation of instruments prior to field deployment are likely to be from type-B sources (i.e. from calibration certificates, existing data, etc.). Examples of calibration and characterisation experiments might relate to the spectrometer (radiance/irradiance calibration, wavelength calibration, spectral resolution/bandwidth/etc. characterisation, stray light, etc.), the reference panel (reflectance calibration, multi-angular reflectance characterisation, etc.), or other instruments being used (e.g. GPS for location based uncertainty of the sample, accuracy estimates of the instruments assessing atmospheric conditions, etc.).

In-field uncertainty assessment is typically focused on sampling, but at various scales. For example, assessment of the uncertainty from spectrometer noise would be determined from multiple measurements of the dark current such that a standard deviation of these can be derived. Likewise, assessment of the uncertainty from the levelling of the viewing optics can be determined by making repeated measurements over each surface. On a larger scale, uncertainty related to the variability of the site can be determined through the number of samples within a single pixel. These are considered type-A assessments.

Post-field uncertainty assessments related to factors such as interpolation are generally harder to address because these operations are typically needed when data does not exist. However, experiments using simulated data might be considered to assess the potential impact of specific interpolation, integration, or other routines.

The assessment of the probability distribution being described by an uncertainty is key to understanding the way in which it can be combined with other uncertainties because scaling coefficients are required to convert non-Gaussian distributions into Gaussian distributions. An example of a common non-Gaussian distribution is the rectangular (or uniform) distribution which can arise when considering the uncertainty from an instrument’s digitisation of the signal. The uncertainty in this case should be scaled by $\frac{1}{\sqrt{3}}$.

In an ideal case, where all the inputs are known and the measurement function includes every source of uncertainty, partial correlation is not present between the different uncertainty sources. This means that any two or more inputs are either perfectly correlated or independent. This is beneficial because any correlation between quantities derived thereafter (e.g. later in the processing chain) can be described by the processing the data has undergone. If this cannot be done (e.g. due to lack of knowledge), estimation of the initial correlation structure between the input variables is important.

Once the uncertainty has been assessed for all the input quantities, the propagation of these uncertainties through the processing chain is the next step. With the development of a measurement function this should be relatively straightforward using the techniques discussed in the next subsection.

### Field data uncertainty propagation

Uncertainty propagation is the process of combining the input uncertainties into an uncertainty in the output quantity. Data processing of field measurements, which combines and converts information from one form into another, uses computer code to make the processing manageable in a reasonable period of time. While it would be possible to work out what the output surface reflectance should be manually, since the processing is a set of mathematical calculations, we will assume this is never done practically. The use of computers for this task means that Monte Carlo techniques, which run many simulations based on new, randomised samples, can be used as well as or alongside analytical uncertainty propagation. There are distinctive positives and negatives with each technique.

Analytical uncertainty propagation involves deriving a secondary equation, which is related to the measurement function, which describes the way the uncertainty should be combined. The standard equation for this comes from [18]:

$$u^{2}\left(y\right)= \sum\_{i=1}^{N}\left(\frac{∂f}{∂x\_{i}}\right)^{2}u^{2}\left(x\_{i}\right)+2\sum\_{i=1}^{N-1}\sum\_{j=i+1}^{N}\frac{∂f}{x\_{i}}\frac{∂f}{x\_{j}}u\left(x\_{i},x\_{j}\right)$$

Where $u\left(y\right)$ describes the uncertainty of the output, $u\left(x\right)$ describes the uncertainty in the input and the partial differential terms are the sensitivity coefficients relating the input to the output quantities. This may also be rewritten into matrix form. The first component of the equation gives the combination of uncertainties due to random effects, while the second component describes the combination of uncertainties from systematic effects. It should be noted that this form does not describe highly non-linear functions well, in this case higher order terms would be required.

Using this form of uncertainty propagation requires the user to be able to derive the sensitivity coefficients associated with each uncertainty, which can be impossible (e.g. if a function is not differentiable) or difficult to achieve. Recent advances in automatic differentiation software means that often users do not have to manually calculate the sensitivity coefficients. The benefit of the analytical solution is speed, once the uncertainty propagation equation has been developed, the uncertainty can be computed quickly.

The second technique (Monte Carlo) uses just the measurement function and the input parameter probability distributions (which are described by the uncertainty). To compute the output uncertainty the measurement function output is computed many times using random draws from the probability distribution of each input. By doing this, not only is a single function required but the internal correlation is accounted for automatically which can be particularly beneficial if the processing chain is complex. The main disadvantage of this process is the computational expense (and therefore speed) in calculating the output since the time taken is the product of the time taken for a single calculation of the measurement function and the number of samples. Determining the number of samples may also be difficult to assess.

# Satellite data pre-processing, quality control and uncertainty

## Pre-processing and quality control

It is worth making a distinction between the pre-processing that is required to retrieve the satellite surface reflectance data for the area of interest in the validation and the pre-processing that is required to create a surface reflectance product. The former will be discussed below and is akin to data preparation, while the latter omitted since it is highly detailed and sensor dependent. For users wishing to obtain information on the latter, the sensor design and algorithm theoretical basis documents can usually be found on the space agency websites.

The preparation of the surface reflectance products requires us to recognise that there are two categories of product: those that provide surface reflectance in the sensor’s grid, and those which provide it in a normalised geometry according to a cartographic reference. The distinction is important because the latter combines different pixels which means that correlation between neighbouring pixels on a cartographic grid is not just related to the imaging mechanism. This will be discussed later on.

In order to prepare the test dataset, the user is required to isolate the area/pixel of interest. For products in the native sensor imaging geometry there is typically a coordinate grid which gives the central (although sometimes other) coordinate of that pixel. From the validation perspective, the GPS coordinates used to mark the measurement locations can be used to filter the pixels. Immediately there may be issues with this since if the measurement location falls on a pixel boundary, and is not sufficiently large enough, then additional measures, such as those taken in [1] are required. For higher resolution sensors (e.g. ~<10 m), sampling on the ground may be designed so that a single measurement reflects a single pixel. For products with cartographic grids it is possible to know before the measurements are made where the pixel location on the ground is (at least in theory) and thereby target a specific pixel. In both cases, geolocation errors in the satellite product must be considered so that a sufficient area around each pixel is sampled.

Once the validation site has been located from the wider product array it is necessary to check that the isolated product pixels are valid. This means checking the quality control flags that are present in the product metadata. The exact flags that exist will depend on the sensor and the product but they will generally relay information about the pre-processing steps used to generate the product and the performance for that pixel. Examples of the information contained in these flags will be related to the instrument performance (e.g. bad pixels (hot or dead)), scene classification (how the scene is classified may determine the atmospheric properties, BRDF model or other operations), quality of the atmospheric retrieval, cloud and snow detection, etc. to name a few. For validation, it is generally advised to be conservative since these actions may be the cause of any differences between the in situ data and the satellite data. As well as the quality control masks, there may also be confidence maps related to those masks which describe the confidence in the classifications that have been made. Once again, it is generally advised that a conservative approach is used and any data that does not pass this threshold is removed. The satellite product may also contain additional information that may be useful in diagnosing potential differences in the validation. For example, where aerosol and other atmospheric parameters are given it may be useful to compare these with the same information measured on the ground.

## Satellite product uncertainty

Most satellite surface reflectance products do not provide uncertainties on the output values. This is an issue since in order to assess whether a sensor’s product is compliant there should be a statement of uncertainty associated with both sides of the comparison. Where uncertainty estimates are not present there are a few options that can be taken. One solution is to use the product requirements specified in the algorithm theoretical basis document (where these exist). Alternatively, if sufficient information about the processing chain of the product exists it may be possible to create a reasonable estimate of the uncertainty. This was the approach used by [1] for the Sentinel-2 product, which is summarised below.

### Uncertainty of Sentinel 2 Level 2A reflectance products

For Sentinel-2, S2-RUT (Sentinel-2 Radiometric Uncertainty Tool) was developed in order to estimate the uncertaintyin the per-pixel TOA reflectance factors. The S2-RUT [20] considers several uncertainty contributions and integrates these:

* Instrument noise
* Out of field stray light systematic and random part
* Crosstalk
* Diffuser reflectance temporal and absolute knowledge
* Deconvolution residual
* ADC quantisation
* Straylight in calibration mode - residual
* Dark signal knowledge and stability
* Non-linearity and non-uniformity knowledge and spectral residual
* L1B image quantisation
* Angular diffuser knowledge – cosine effect

These contributions are combined using the standard GUM uncertainty propagation, relying on a linearized measurement model. The analytical procedure was validated against a multivariate Monte Carlo analysis to ensure it was accurate. Despite the uncertainty contributions given, several uncertainty sources are not considered such as the deconvolution residual and other sources of straylight, the polarisation error, the orthorectification uncertainty propagation, and the geometrical and spectral knowledge.

When propagating pixel-level TOA radiance or reflectance factor products to higher levels in a processing chain, the pixel-level radiometric uncertainty must be treated carefully. Many applications of higher-level products aggregate data from different pixels in space and/or time using a simple, or weighted mean. To determine the uncertainty associated with the mean, it is also necessary to consider whether there are systematic effects leading to common errors between different pixels. Similarly, higher level products also often involve combining data from different spectral bands. [20] considers different regions of interest (ROIs) to estimate the error correlation structure in spatial, temporal and spectral dimensions and then averaging over a specific ROI in order to reduce the effects of noise and/or to allow comparisons of sensors to references with a different pixel size. This approach can be used to estimate uncertainty through the S2-RUT.

The satellite uncertainty propagation and the derivation of the surface HCRF uncertainties for the S2 L2A product is based on a Monte Carlo approach where input variations are taken from the TOA reflectance factors as well as the AOT and water vapour. The error distributions can be assumed to be Gaussian because there is a lack of information, while the standard deviations are taken from the output uncertainties produced by the S2-RUT and the uncertainty guidelines for the aerosol and water vapour retrievals.

## Satellite surface reflectance product validation gaps and challenges

The information contained in this document has discussed the general considerations for validating satellite-derived surface reflectance products from the perspective of the satellite processing as well as the collection and processing of the data collected in the field. Nevertheless, there remain some existing gaps and challenges to address in future activities:

* **Indirect validation might not highlight systematic errors**

Indirect validation is clearly an effective way to evaluate a product on a global scale, against one that has already been validated against in-situ measurements. However, it should not be taken in isolation – the validation studies cited here have complemented indirect with direct approaches to present a fuller picture. Relying too heavily on indirect validation risks matching systematic errors in regions or surface types that have not been subject to ground measurements, resulting in a group of products that might agree with each other, but are not an accurate representation on surface conditions in those specific regions.

* **Automated validation networks are costly and high maintenance**

The challenge for the community is to continue to innovate and find solutions through projects such as HYPERNETS that can maintain continuous measurements in a cost-effective way, supported by field campaigns

* **Limitations of field campaigns (homogenous fields, broadly cloud-free areas)**

In-situ validation lends itself to summertime periods in cloud regions where the surface vegetation is reasonably homogenous. Expanding field campaigns to less practical locations (e.g. more cloudy regions, areas with less homogenous land cover) is a key future measurement challenge, as well as the need for more validation sites and over a greater variety of land covers.

* **Misunderstanding of reflectance definition within the community**

[21] raise an important point about definitions of reflectance amongst the broader community. There are multiple definitions of ‘reflectance’. In other communities using these products the nuances between different definitions might not be universally understood, and so there is scope for erroneous use of data, without full appreciation of uncertainties.

# References

[1] N. Origo, J. Gorroño, J. Ryder, J. Nightingale, and A. Bialek, “Fiducial Reference Measurements for validation of Sentinel-2 and Proba-V surface reflectance products,” *Remote Sens. Environ.*, vol. 241, no. November 2019, p. 111690, 2020, doi: 10.1016/j.rse.2020.111690.

[2] T. Malthus *et al.*, “A community approach to the standardised validation of surface reflectance data.,” 2019.

[3] T. Hakala *et al.*, “Direct reflectance measurements from drones: Sensor absolute radiometric calibration and system tests for forest reflectance characterization,” *Sensors (Switzerland)*, vol. 18, no. 5, May 2018, doi: 10.3390/s18051417.

[4] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V Martonchik, “Reflectance quantities in optical remote sensing — definitions and case studies,” *Remote Sens. Environ.*, vol. 103, pp. 27–42, 2006, doi: 10.1016/j.rse.2006.03.002.

[5] F. Baret *et al.*, *VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite products*. Avignon, France: Institut National de la Recherche Agronomique, 2005.

[6] J. A. Coakley, “REFLECTANCE AND ALBEDO, SURFACE,” *Elsevier Sci.*, 2003.

[7] J. R. J. J. R. Shell, “Bidirectional Reflectance : An Overview with Remote Sensing Applications & Measurement Recommendations,” Rochester, New York, 2004.

[8] C. P. Ball *et al.*, “Hemispherical-Directional Reflectance (HDRF) of Windblown Snow-Covered Arctic Tundra at Large Solar Zenith Angles,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5377–5387, 2015, doi: 10.1109/TGRS.2015.2421733.

[9] E. Vermote, C. Justice, M. Claverie, and B. Franch, “Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product,” *Remote Sens. Environ.*, vol. 185, pp. 46–56, 2016, doi: 10.1016/j.rse.2016.04.008.

[10] M. Danner, M. Locherer, T. Hank, and K. Richter, *EnMAP Field Guides: Technical Report: Spectral Sampling with the ASD FieldSpec 4*. Potsdam, Germany: GeoForschungsZentrum Data Services, 2015.

[11] H. Aasen, E. Honkavaara, A. Lucieer, and P. Zarco-Tejada, “Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows,” *Remote Sens.*, vol. 10, no. 1091, pp. 1–42, 2018, doi: 10.3390/rs10071091.

[12] E. F. Vermote and S. Kotchenova, “Atmospheric correction for the monitoring of land surfaces,” *J. Geophys. Res. Atmos.*, vol. 113, no. D23, 2008.

[13] CEOS - Committee on Earth Observation Satellites, “RadCalNet Portal.” https://www.radcalnet.org/#!/ (accessed Apr. 09, 2020).

[14] M. Bouvet *et al.*, “RadCalNet: A radiometric calibration network for earth observing imagers operating in the visible to shortwave infrared spectral range,” *Remote Sens.*, vol. 11, no. 20, 2019, doi: 10.3390/rs11202401.

[15] HYPERNETS, “HYPERNETS.eu.” https://www.hypernets.eu/from\_cms/summary (accessed Apr. 09, 2020).

[16] F. Gascon *et al.*, “Copernicus Sentinel-2A calibration and products validation status,” *Remote Sens.*, vol. 9, no. 6, 2017, doi: 10.3390/rs9060584.

[17] M. Claverie, E. F. Vermote, B. Franch, and J. G. Masek, “Evaluation of the Landsat-5 TM and Landsat-7 ETM + surface reflectance products,” *Remote Sens. Environ.*, vol. 169, pp. 390–403, 2015, doi: 10.1016/j.rse.2015.08.030.

[18] Working Group 1 of the Joint Committee for Guides in Metrology, "*Evaluation of measurement data — Guide to the expression of uncertainty in measurement"*. Paris, France: Bureau International des Poids et Mesures, 2008.

[19] FIDUCEO, “Example Uncertainty Analysis Tree.” http://www.fiduceo.eu/tutorial/example-uncertainty-analysis-tree (accessed Feb. 22, 2019).

[20] J. Gorroño *et al.*, “Providing uncertainty estimates of the Sentinel-2 top-of-atmosphere measurements for radiometric validation activities,” *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 650–666, 2018, doi: 10.1080/22797254.2018.1471739.

[21] U. Beisl, G. Strub, C. Dickerhof, and R. S. Laboratories, “Validation of hyperspectral imaging data from the Barrax test site with BRDF ground measurements in the reflective wavelength range,” in *Second EARSeL Workshop on Imaging Spectroscopy, Enschede, 2000 VALIDATION*, 2000, pp. 2–10.