Remote Sensing of Environment 241 (2020) 111690

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Fiducial Reference Measurements for validation of Sentinel-2 and Proba-V surface reflectance products

Niall Origo[a,b,*], Javier Gorroño[a], James Ryder[a], Joanne Nightingale[a], Agnieszka Bialek[a]

[a] Earth Observation, Climate and Optical group, National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, UK
[b] Department of Geography, University College London, Gower Street, WC1E 6BT, UK

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
Surface reflectance
Validation
FRM
Uncertainty

ABSTRACT

Many derived Earth Observation products share surface reflectance as a common step in their processing chains. This makes the maintenance and improvement of surface reflectance product quality of fundamental importance to ensure information derived from these downstream products can be trusted. Despite this, the literature is relatively light on the implementation of validation methodologies designed for surface reflectance. In response to the need to improve general EO validation methodologies, the concept of fiducial reference measurements (FRM) was created that would produce validation data that is fully characterised and independent, with associated uncertainties, and traceable to SI.

This paper describes a field campaign designed to produce surface reflectance validation data, in line with the FRM ideology, for validating the Sentinel-2 L2A (S2) and Proba-V S1-TOC (Proba) surface reflectance products. The key methodological procedures are outlined in detail to facilitate uptake of FRM methods in future validation studies. These include the calibration and characterisation of field instruments, the field measurement protocol, and uncertainty propagation.

Comparison of S2 with field measurements showed agreement within the stated uncertainties present for all bands and locations at the pixel and area scales. The Proba comparison results demonstrated a general disagreement within the stated uncertainties. However, there is a lack of uncertainty information provided by the product as well as publicly available product uncertainty requirements which mean that it is difficult to assess this fully. The aim of this practical demonstration of FRM-based surface reflectance validation, that utilises metrological practices for uncertainty characterisation, is to encourage the adoption of these procedures in future land surface reflectance validation studies and operational activities.

1. Introduction

In the terrestrial domain, atmospherically corrected surface reflectance products are the last in a series of processing steps that starts with at-sensor digital counts. These products then form the starting point for numerous application areas such as land cover mapping and the derivation of biophysical essential climate variables (ECV). Therefore, ensuring the quality of surface reflectance products is beneficial to maintaining the integrity of the research coming out of these application areas.

Validation provides a key route to ensuring that these products are performing to their specifications. In this context, an independent data source is used as a reference to which the satellite product is assessed. The prevalence of validation studies and operational validation activities is relatively widespread within the literature for a variety of Earth observation (EO) products (e.g. Liang et al., 2002a; Liu et al., 2009; Guillevic et al., 2012; Camacho et al., 2013) since this provides a means for users to assess the utility of the product for their application. In recognition of the importance of this work, the Committee for Earth Observation Satellites' (CEOS) Land Product Validation (LPV) subgroup was established to coordinate international validation activities across different space agencies and research institutions.

Review of the surface reflectance product validation literature reveals a broad trend towards the assessment of such products against long-term in situ monitoring networks such as the Surface Radiation Budget Network (SURFRAD) (Jin et al., 2003; Salomon et al., 2006; Liu et al., 2009), Aerosol Robotic Network (AERONET) (Vermote et al., 2016) and the AERONET-based Surface Reflectance Validation Network (ASRVN) (Wang et al., 2009; Wang et al., 2010; Sogacheva et al., 2015). These networks are made up of remote ground stations transmitting

[*] Corresponding author at: Earth Observation, Climate and Optical group, National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, UK.
E-mail address: niall.origo@npl.co.uk (N. Origo).

https://doi.org/10.1016/j.rse.2020.111690
Received 7 November 2019; Received in revised form 27 January 2020; Accepted 28 January 2020
Available online 06 March 2020

radiation fluxes (SURFRAD) and multispectral aerosol optical depth (AOD) estimates (AERONET). The ASRVN system provides atmospherically-corrected MODIS surface reflectance for the 50 km^2 surrounding the ground station (Wang et al., 2009). 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" (Vermote et al., 2016). 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.

However, some common themes begin to emerge when analysing the studies using these networks. Firstly, the combined impact of the sensor spatial resolution (MODIS at 500 m in most cases) and the site heterogeneity can cause significant disagreement. These tend to be mitigated in one of two ways: stricter station inclusion requirements (Román et al., 2009; Cescatti et al., 2012) or by using a transfer product with a greater spatial resolution (Liang et al., 2002a; Fan et al., 2014). For example, Jin et al. (2003) found that the increased heterogeneity brought by snow pushed the validation results outside of the product accuracy requirements for the winter months, an effect that was also seen in other studies (Salomon et al., 2006) and attributed this as one of the major issues affecting the validation (Liu et al., 2009). However, with sparse validation datasets often there is little opportunity to be selective. This leads to an alternative approach which attempts to overcome this issue by accounting for that heterogeneity with a finer resolution transfer product. While this is appealing, matching of the product geometries is required (and often not available) and correlation is introduced between the reference and test datasets (i.e. similar processing, atmospheric assumptions, etc.). Furthermore, the transfer product adds its own uncertainty. These three factors make this type of processing challenging and assessment of the results ambiguous.

Secondly, there is little mention or quantitative assessment of the uncertainties associated with the in situ data nor any endeavour to perform the validation in a metrologically robust way. The best attempt appears to be the analysis described in Vermote and Kotchenova (2008) and later used in Vermote et al. (2014) where the authors use the in situ data as a representation of the truth and describe the validation uncertainty as a linear combination of the precision and accuracy results. However, this is a long way off meeting the needs of future applications of satellite-derived data and requires consideration of the reference data uncertainty (Widlowski, 2015). Likewise, reduced ambiguity in the validation results is required.

Thirdly, many of the published activities affecting many sensors (for example: Sentinel-2: Gascon et al. (2017); MODIS: Vermote et al. (2016); and AATSR: Sogacheva et al. (2015)) utilise reference data which cannot be considered independent (e.g. those utilising other satellite products). This is because the main assumption is that the aerosol correction is the dominant error. This may be the case, but by only comparing the impact of the aerosol retrieval, the impact of other errors in the processing chain (radiometric, geolocation, etc.) is lost and the validation is only testing the aerosol retrieval. Validation of satellite surface reflectance data against independent in situ reflectance estimates appears to be more common for assessing the quality of new algorithm developments (e.g. Liang et al., 2002b; Li et al., 2010) rather than operational products.

Given the increasing role that quantitative EO derived products assume in climate and environmental monitoring applications, the quality of this data is coming under increasing levels of scrutiny (Widlowski, 2015; Nightingale et al., 2018; Nightingale et al., 2019). Therefore, detailed assessment of EO data product quality that includes uncertainty characterisation of the retrieval algorithm as well as the in situ data and methods used to validate the algorithm, is essential. Further, provision of the end-to-end quality assessment in a standardised and comprehensible manner is required to help potential data users navigate and understand the nuances between the wealth of similar satellite derived products available to them.

This is particularly important for operational activities, such as the European Union's Copernicus Climate Change Service (C3S) whose mission is to provide authoritative information about the past, present and future climate in Europe and the rest of the world. In a progressive commitment to ensure that all datasets available through the C3S Climate Data Store are traceable, adequately documented and accompanied by quality information so that data users can make informed decisions for their application, C3S has made significant investments in the ongoing development of Evaluation and Quality Control (EQC) functionality (Nightingale et al., 2019). Similarly, the European Space Agency (ESA), in attempting to address some of the issues mentioned above, have created a program of work to address the concept of Fiducial Reference Measurements (FRM). These aim to provide fully characterised and independent in situ data with uncertainties where the measurements are traceable to the International System of Units (SI - from Système international (d'unitès)), or community-agreed standards, to underpin the validation of satellite products. The production of these datasets should follow the Quality Assurance for Earth Observation (QA4EO) guidelines; namely that data provenance is ensured and that uncertainty assessment follows the Guide to the Expression of Uncertainty in Measurement (GUM) approach (JCGM, 2008).

With this in mind, the present paper describes a field campaign designed to produce an FRM dataset for the validation of the Sentinel-2 and Proba-V surface reflectance products, with the aim of developing the concept of an FRM validation. We consider the validation problem as containing three distinct components: the reference sensor, the test sensor, and the measurement conditions. The first two of these are generally considered obvious (and in the context of this study refer to the in situ and satellite sensors respectively), but attention to the latter is rarely formalised in a remote sensing context (i.e. by inclusion into the uncertainty budget). Here we attempt to account for each in turn and formalise the assumptions made throughout, ensuring that the traceability, provenance and uncertainty are at the forefront of the analysis. It is noted that our usage of the terms "uncertainty", "error" and other related terms in this paper is guided by the GUM (JCGM, 2008) and defined by the international vocabulary of metrology (VIM).

2. Materials and methods

2.1. FRM campaign design

Formulating a validation campaign that adheres to FRM principles involves considering the factors influencing the comparison. Table 1 provides a non-exhaustive list of subcategories under each of the distinct validation components.

The goal of collating this information is to determine: which measurements need to be made in the field; the calibration and characterisation experiments that need to take place; the choice of field protocol; and timing of the field data acquisition. Consideration of the measurement quantity actually covers several of the subcategories mentioned in the table. It is paramount that the reflectance quantity measured by the in situ sensor is the same as that measured by the satellite, and this informs the instrumentation that is required and assumptions

Table 1
Subcategories of each validation component to consider.

Reference and test sensor	Measurement conditions
Spatial resolution	Surface BRDF
Field of view	Atmospheric changes
Levelling	Solar angle changes
Measurement quantity	Homogeneity
View angle	
Spectral response	

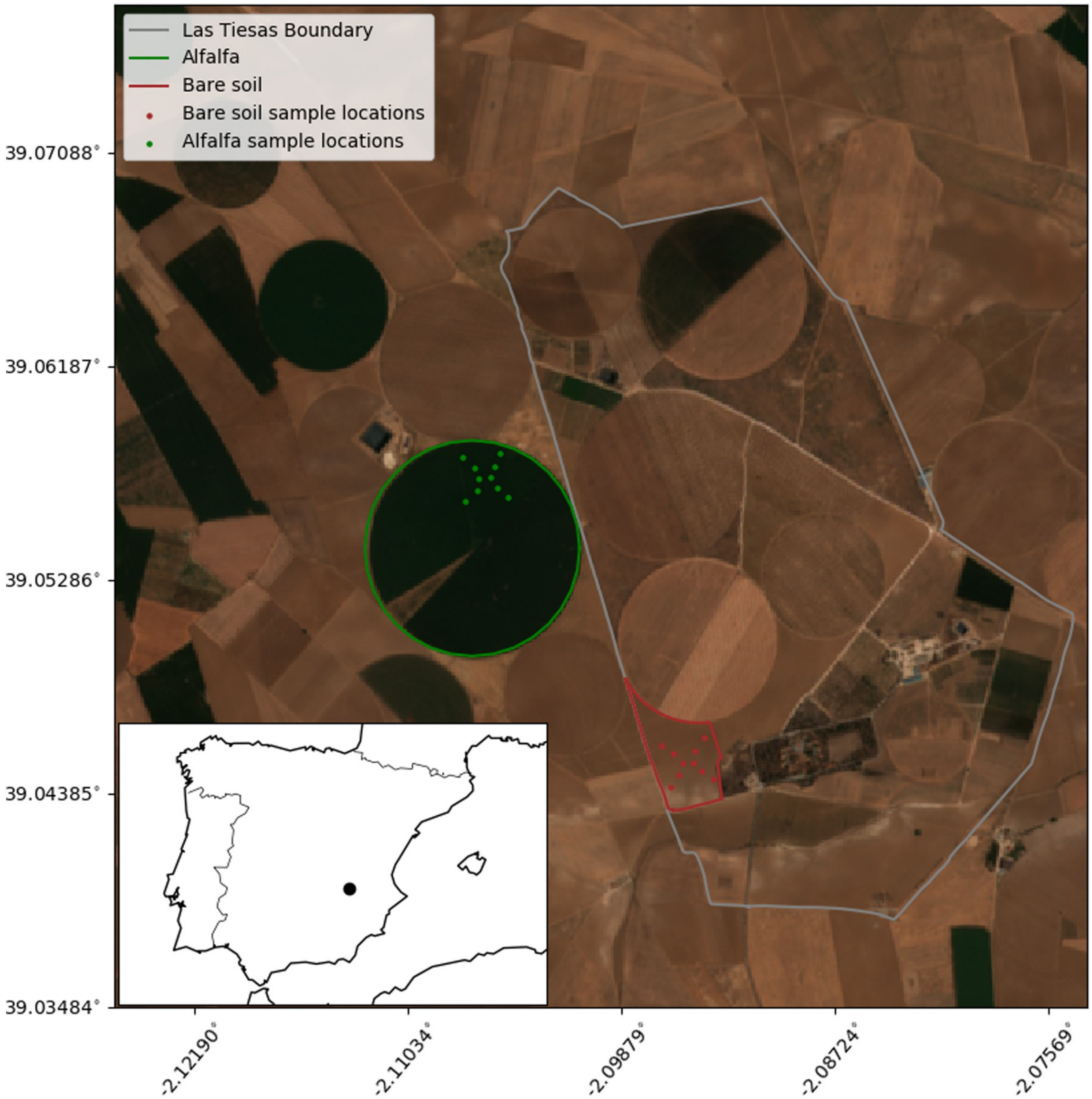


Fig. 1. The location of the field campaign with sample points shown. Las Tiesas farm boundary shown in grey, with the boundaries and sample points of the Alfalfa and bare soil shown in green and red, respectively (GIS data provided by José González Piqueras). The data from the overpass can be found in the file: *S2A_ MSIL2A_ 20180802T105621_ N0208_ R094_ T30SWJ_ 20180802T141714.SAFE.*

that will be made. Likewise, site selection is based on minimising the effect of environmental factors that influence the mismatch between the in situ and satellite sensors. Examples of these include the surface homogeneity (within pixel) and the surface Bidirectional Reflectance Distribution Function (BRDF) which becomes important when assessing results from sensors with different fields of view (FOV). Refinement of methodological practices should be done over these idealised sites before moving to "imperfect" sites where greater location based uncertainties can be expected. Many of the considerations and practicalities involved in conducting a surface reflectance validation campaign are provided in Malthus et al. (2018), including spectrometer operation practices, data collection and storage, and site selection.

These factors determine the field and satellite data processing, as well as formulation of the measurement equation and uncertainty analysis.

2.2. Field site

The validation campaign took place on the 2*nd* August 2018 at Las Tiesas, Barrax, Spain (Fig. 1). Las Tiesas was chosen for the field campaign because it has several desirable features: large and homogeneous crop surfaces, low summer cloud cover likelihoods, and a long history of remote sensing validation. Las Tiesas has a large number of different crop cover types including Alfalfa (*Medicago sativa*), poppy (*Papaver somniferum*), garlic (*Allium sativum*) and wheat (*Triticum aestivum*). The climate in this part of Spain is predominantly hot and dry in the summer months, meaning in-rotation fields require additional water provided by a central-pivot irrigation system. Measurements were made over the Alfalfa and bare soil surface types. Alfalfa was chosen for a number of reasons: the field area was one of the largest at the farm (approximately 1 km diameter); the canopy height does not reach much more than 1 m during a single rotation and is therefore convenient for manual spectrometer measurements; navigating through the crop is relatively simple; and based on prior analysis of S2 images (not shown), the Alfalfa was the least variable (reflectance wise) of all the crops. A bare soil surface area was also chosen to provide a brighter surface in the visible domain. While a number of bare soil surfaces were available, the specific surface used was selected visually and quantitatively based on its homogeneity (satellite and ground scales) and proximity to the Alfalfa field. It is worth noting that the measurements over the Alfalfa were given a higher priority and subsequently measured closer to the satellite overpass.

2.3. Field measurement procedure

The procedure utilised several key apparatus in order to collect the appropriate data for the validation: a spectrometer, Spectralon panel, and sunphotometer. The in situ reflectance values were collected in raw digital number (DN) mode with an ASD FieldSpec 4 spectroradiometer with the 8° foreoptic attachment. The spectrometer has a 400 nm–2500 nm wavelength range with a variable bandwith of 3 nm at 700 nm and 10 nm at 1400/2100 nm (Malvern Panalytical, 2019); internal software interpolates to a 1 nm wavelength interval. The reflectance standard used in the procedure was a 0.4572 m^2, 99% reflectivity Spectralon™, panel produced by Labsphere. A calibrated Microtops sunphotometer was used to retrieve the aerosol optical thickness (AOT) during the campaign.

For each of the cover types, 10 sample points were selected over a 200 × 200 m^2 area according to the pattern shown in Fig. 1 , where it was ensured that every point was at least 50 m from the cover type boundary in order to avoid edge effects. Since the Alfalfa was the primary interest, the measurement window was selected with the aim of conducting measurements no more than 30 min either side of the Sentinel-2A (S2A) overpass. This was done to minimise the change in solar zenith angle (θ_s) during the measurement window and thereby minimise any BRDF effects. Nevertheless, the change in θ_s was expected to be around 6°–7° (Fig. 2) based on the predicted S2 overpass. This meant that the bare soil field was sampled before both overpasses. In any case, attempting to straddle the predicted S2 overpass time proved difficult to execute given the time required to travel between the bare soil and Alfalfa sample locations.

At each sample location a sequence of six measurements of the reflectance panel and surface were made (in the order: panel - 4 x target - panel). Each individual measurement is made up of ten ASD scans, resulting in a total of 60 spectra for each sample location. Each scan was saved separately in order to provide statistical estimates of the variability during the measurement sequence. Before beginning the measurements at each sample location, a white reference measurement was taken of the panel. This tunes the spectrometer settings (integration time, etc.) to the ambient field conditions and ensures that saturation cannot happen while the illumination remains constant. In addition, approximately three AOT measurements were made per sample location.

2.4. Calibration and characterisation

In this study, traceability to SI is maintained via the reference panel. The Bidirectional Reflectance Factor (BRF) was characterised using a Perkin–Elmer Lambda 900 spectrophotometer fitted with a 0°:45° radiance factor accessory (allowing an illumination angle of 0° and a viewing angle of 45°) at the National Physical Laboratory (NPL). The characterisation was conducted over the 345 nm–1055 nm at 5 nm wavelength interval (Fig. 3 - hereafter referred to as hPanel). The remaining wavelengths were extrapolated using a linear interpolation of the calibration data from 700 nm–800 nm, which represented the most stable region of the spectrum before the region with increased uncertainty. Conversely, the uncertainty was linearly extrapolated from the 800 nm–900 nm region and increased by 2% to reflect the increased uncertainty in the extrapolation. Typically, the central portion of the panel is measured during this procedure. However, due to the panel's size, four locations were measured in the centre of each of the four sides.

By using the Spectralon panel as the calibration source, measurements made using the spectrometer can be done in relative mode. This means that absolute calibration of the spectrometer is not required as long as instrument changes between the test and reference measurements are minimal (and mainly due to non-linearity effects). Despite this, wavelength scale and bandwidth checks were conducted to ensure that the instrument was performing to specification.

Since the desired reflectance quantity includes direct and diffuse up and downwelling radiation, as well as off-angle (in this case off 45°) direct beam radiation, characterisation of the panel at multiple angles is required to account for its anisotropic properties. Therefore a separate multi-angular panel characterisation was conducted. The multi-

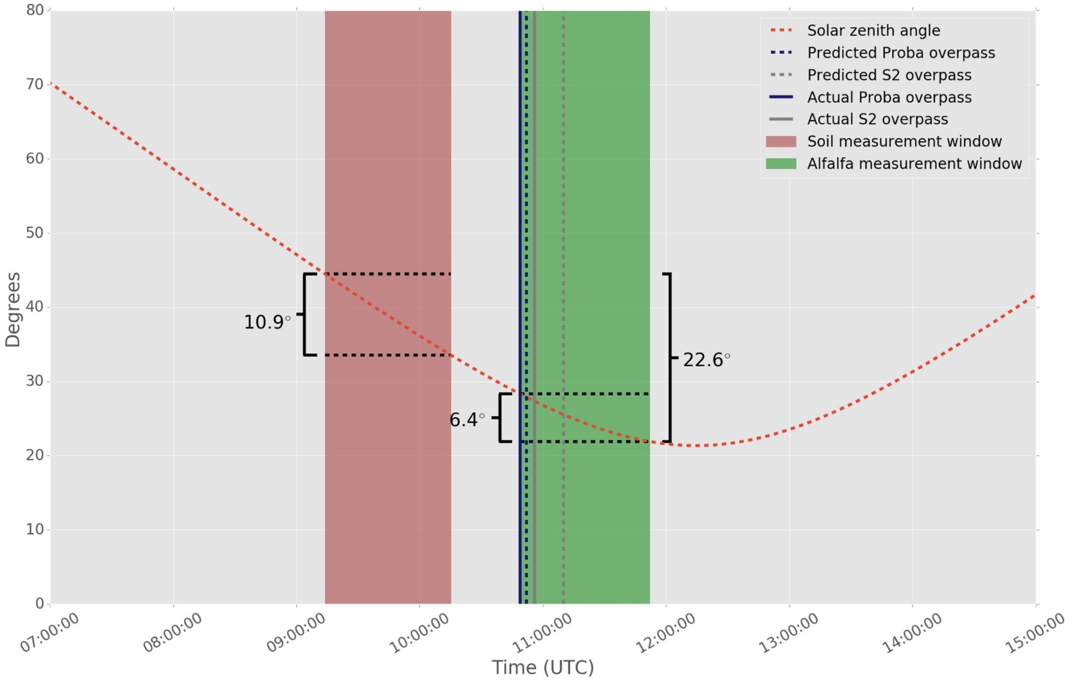


Fig. 2. Solar zenith angle throughout the day on 02/08/2018, as estimated by PySolar (Zebner et al., 2007). The in situ data collection windows are given by the red (bare soil) and green (Alfalfa) filled areas. The black and green dashed lines give the overpass times predicted by the CEOS Visualisation Environment (COVE) tool (Kessler et al., 2013) and used for planning the campaign. The solid lines give the overpass times according to the product.

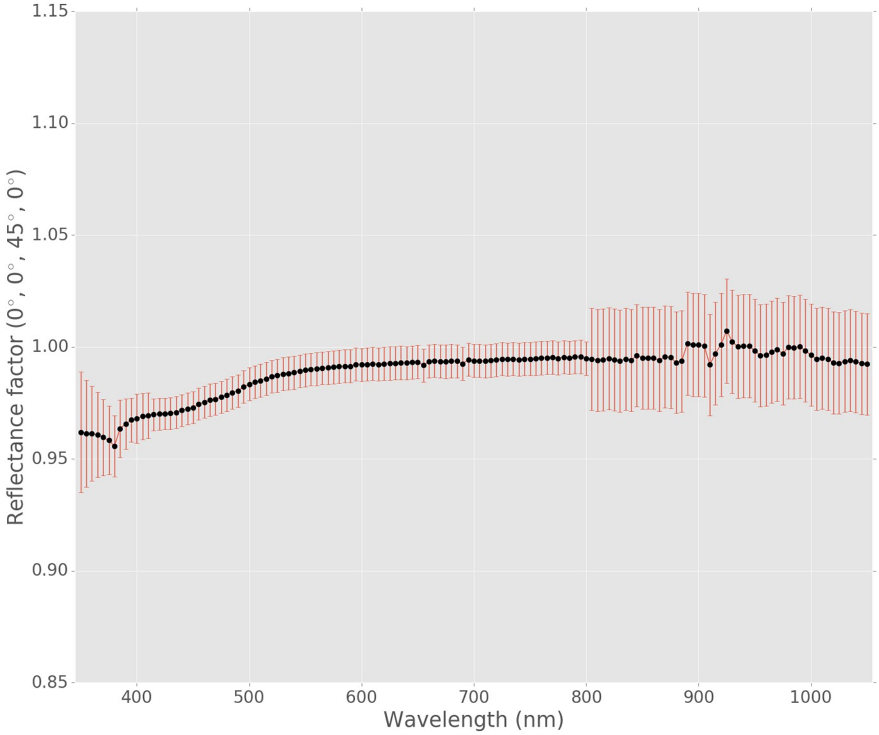


Fig. 3. Direct beam Spectralon panel characterisation (hPanel). The characterisation was conducted from 345 nm–1050 nm. The remaining wavelengths were extrapolated using a linear model from the data between 700 nm and 800 nm (not shown). Error bars represent the uncertainty associated with the measured values with a coverage factor of $k = 2$.

spectral/angular panel characterisation ranges from 400 nm–1001 nm and 10°–85° (in 5° spacing). For the angular range 20°–60° (at 10° spacing) additional characterisation was conducted between 1001 nm and 2300 nm. To estimate the values between 1001 nm and 2300 nm that had not been measured, two procedures were implemented. For the values that could be interpolated, a linear interpolation was used. For the angles outside of the interpolation, an exponential function of the form $y = ae^{-b\theta_v + c} + d$ was used to fit to the data between 400 nm and 1001 nm (where a, b, c, and d represent the function parameters to be solved for, and an initial guess of -0.02279229, -2.75335837, -1.57744657 and 1.05652974 was used for each parameter respectively, and θ_v gives the view zenith angle). The average of these parameters (with wavelength) was used as the parameter set for the exponential function between 1001 nm and 2300 nm. The standard deviations of these parameters (combined according to the function mentioned) was used as an estimate of the additional uncertainty introduced at longer wavelengths in the extrapolated angular range. A scaling factor was then produced which would normalise the difference between the interpolated values and the function used for the extrapolation (within the interpolation range). This was applied to the extrapolation range and all values (i.e. extrapolated regions and interpolated regions) were combined to produce the final data given in Fig. 4. This input is hereafter referred to as mPanel. The measurements were made when the panel was first purchased and therefore reflects a new Spectralon panel surface. Since the panel was used in the field several times before deployment in this field campaign, the hPanel characterisation was undertaken to ascertain the current panel reflectance factor. As a result, the mPanel data were normalised to the data at 45° (i.e. mPanel/mPanel45) and scaled by the hPanel data to provide angularly resolved calibration coefficients. Hence the additional information contained between 1055 nm and 2300 nm that was not provided by the hPanel characterisation is included at the reduced multispectral wavelength steps. It is worth noting that the panel was cleaned, according to the recommended procedures provided by Labsphere, before the hPanel characterisation; we assume that its angular characterisation remains constant. A 2D piecewise linear interpolation function was created so that the appropriate direct calibration coefficient could be retrieved per solar zenith angle. The solar zenith angle in the field ranged from 21°–49° , which is clearly within the interpolation range.

2.5. Field data processing

The in situ reflectance processing combines several data sources including the ASD binary files, multi-angular/spectral panel characterisation data (mPanel), hyperspectral panel calibration (hPanel), and spectral diffuse/direct ratios.

The in situ Hemispherical Directional Reflectance Factor (HDRF; assumed to be equivalent to the Hemispherical Conical Reflectance Factor (HCRF) measured by the ASD) was retrieved according to Bialek et al. (2016):

$$R_{g,HDRF}(\theta_s) = \frac{(R_{p,BRF}(\theta_s)d + R_{p,TDR}(1-d))X_g}{0.5(X_p(t_1) + X_p(t_2))} \tag{1}$$

Where R represents reflectance factor, the surface is represented by the subscript g (ground) or p (panel), and the type of reflectance factor is either the *HDRF*, *BRF* (Bidirectional Reflectance Factor) or *TDR* (Total Diffuse Reflectance factor - i.e. solely from diffuse illumination); θ_s gives the solar zenith angle; d gives the ratio of direct to total irradiance; X gives the DN output by the spectrometer; and t_1 and t_2 give the prior and posterior measurement of the panel. This was common to all wavelengths so, for clarity, the dependence upon wavelength (λ) is omitted. Under this definition it is noted that the HDRF is defined according to Schaepman-Strub et al. (2006), where the HDRF definition given by Nicodemus et al. (1977) is referred to as TDR in line with Bialek et al. (2016).

$R_{p,BRF}(\theta_s)$ was retrieved directly from the interpolation function described in Section 2.4, while $R_{p,TDR}$ was calculated according to Chunnilall et al. (2003):

$$R_{p,TDR} = \frac{\int_0^{\frac{\pi}{2}} R(\theta_v)\sin 2\theta_v d\theta_v}{\int_0^{\frac{\pi}{2}} \sin 2\theta_v d\theta_v} \tag{2}$$

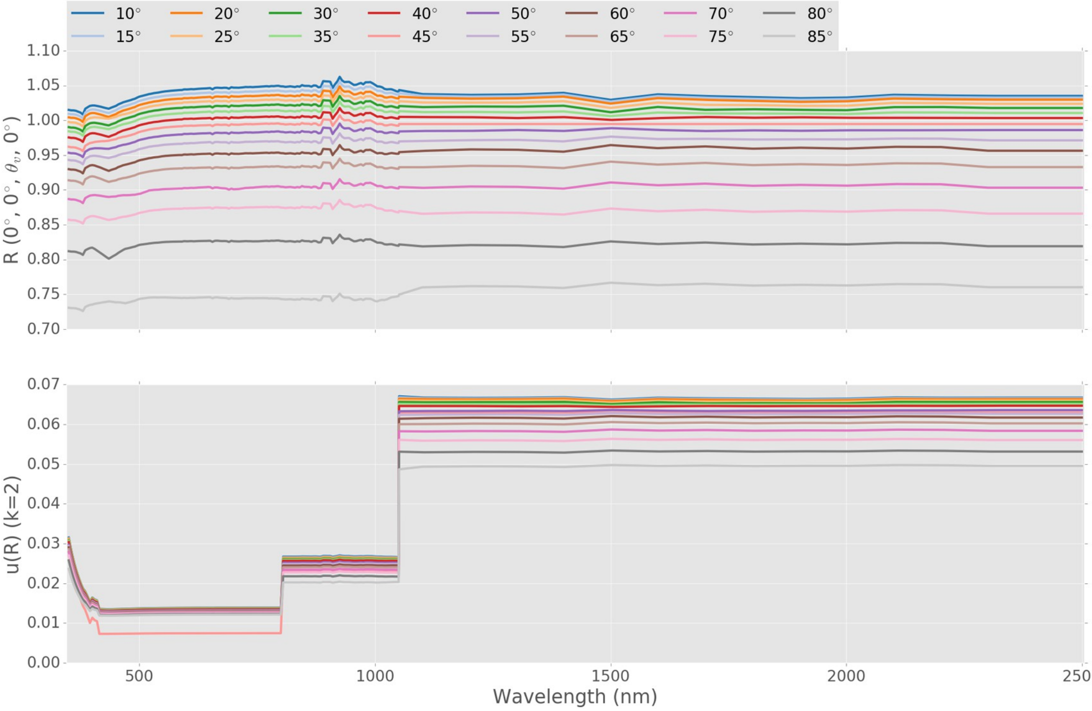


Fig. 4. Multi-angular spectralon panel characterisation (scaled by hPanel mentioned above; top panel) and associated uncertainty $k = 2$). In both cases data point markers have been omitted for clarity.

d is calculated using 6S (Vermote et al., 1997; Wilson, 2013). This involves reading the Microtops data file and retrieving the latitude, longitude, AOT and water vapour content. Ozone was retrieved from local stations in Murcia and Madrid at the daily timestep hosted by a Spanish weather service; this was set to 0.31 atm cm (AEMet, 2018). The AOT values were interpolated to give AOT at 550 nm. Finally, the 6S continental aerosol profile was used to derive the direct and diffuse spectral irradiance at the surface. The solar zenith angles (SZA) derived from the Microtops data were used in this analysis for the calculation of the calibration coefficient for each sample point.

The allocation of the direct-to-total ratio and solar zenith angles to specific measurement windows are made based on the times specified in the ASD files. This means that if there is data between the two end measurements (both of which will be panel scans) then an average of these are used for that location. If there aren't any between the two then an average of all the measurements are used but weighted by the time difference to the middle point of that location. For each of the target measurements the diffuse and direct calibration coefficients are applied, weighted by the direct-to-total ratio.

Once the calibration coefficients have been applied, the averaged values for each sample location are convolved according to the spectral response function of S2A and Proba-V. For comparison with the Proba-V surface reflectance product, an average value corresponding to the pixel of interest is produced.

2.6. Field data uncertainty propagation

As mentioned previously, the terms "uncertainty" and "error" are defined according to the international vocabulary of metrology (VIM; JCGM (2012)) where the "measurement uncertainty" is given as a:

> "non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used" (JCGM, 2012, pp. 25)

And the "measurement error" is given as a:

> "measured quantity value minus a reference quantity value" JCGM (2012, pp. 22).

The uncertainty propagation used here implements the approach described in Merchant et al. (2019) which gives the uncertainty equation as:

$$u^2(y) = \sum \mathbf{C}{\cdot}\mathbf{U}{\cdot}\mathbf{R}{\cdot}\mathbf{U}^{\mathbf{T}}{\cdot}\mathbf{C}^{\mathbf{T}} \tag{3}$$

Where **C** represents a matrix of sensitivity coefficients (along diagonal); **U** is the uncertainty matrix where the diagonal elements are the uncertainty contributions (given with a coverage factor (*k*) of 1 - the coverage factor can be considered as statement of the level of confidence in the uncertainty; when considering an uncertainty with a Gaussian distribution and a coverage factor of 2, the corresponding level of confidence is approximately 95%) and off-diagonals are zero; **R** represents the correlation matrix; and $^{\mathbf{T}}$ denotes the transpose operation. This approach is fully compliant with the Guide to the Expression of Uncertainty in Measurement (GUM) (JCGM, 2008).

This approach separates the correlation and the uncertainty allowing both to be set accordingly and propagated through the equations as required. In certain cases, the output uncertainty matrix is collapsed into a single uncertainty value, especially when dealing with multi-dimensional situations (e.g. ρ(θ,λ)). In the equations following these, the correlation is estimated for the relevant variables. A breakdown of the uncertainty components is shown in Table 2.

The approach to accounting for as many sources of uncertainty as is feasible, as well as correlation, relies on developing a comprehensive measurement equation. This equation attempts to explicitly account for all processes that influence the output value. By taking Eq. (1) as an example, it is implied that X_g, $X_p(t_1)$ and $X_p(t_2)$ are all measured at nadir. A fuller version of the equation would therefore account for the impact of measuring off-nadir:

$$R_{g,HDRF}\left(\theta_s\right) = \frac{\left(R_{p,BRF}\left(\theta_s\right)d + R_{p,TDR}\left(1-d\right)\right)X_g\cos\left(\theta_{v_1}\right)}{0.5\left(X_p\left(t_1\right)\cos\left(\theta_{v_2}\right) + X_p\left(t_2\right)\cos\left(\theta_{v_3}\right)\right)} \tag{4}$$

Here the view zenith angles for each reading are given by θ_{v_1}, θ_{v_2} and θ_{v_3}. Now $u(\theta_v)$ can be accounted for and an appropriate sensitivity

Table 2
Uncertainty component inputs included in the in situ processing. σ_a, σ_b, σ_c and σ_d give the standard deviations of the parameters *a*, *b*, *c* and *d* from the exponential fit given in Section 2.4; σ and σ_w give the standard deviation and weighted standard deviation respectively; *n* gives the number of measurements for that field. All distributions are assumed to be Gaussian based on the uncertainty assessment method.

Source	Uncertainty ($k = 1$)
Panel BRF (hPanel) all	See Fig. 3
Panel BRF (mPanel) angle	0.05°
Panel BRF (mPanel) radiometric (400 nm–1001 nm at all angles and > 1001 nm and between 20°≤ θ_v ≤ 60°)	1%
Panel BRF (mPanel) radiometric >1001 nm and θ_v < 20° and θ_v>60°	1% + GUM combination of σ_a, σ_b, σ_c and σ_d
Direct to total irradiance ratio	σ or σ_w of *d* within location period
SZA	σ or σ_w of SZA within location period
Levelling and raw counts	$\frac{\sigma}{\sqrt{n}}$ of panel and target scans respectively

coefficient derived. In this way, a cascade of measurement equations leading from all the inputs to the final output can be created. The reader is referred to Mittaz et al. (in press) for a comprehensive review of the approach.

2.7. Satellite data processing

The Alfalfa field acquisitions were made around the predicted S2A overpass as explained in Section 2.3. The actual overpass occurred 14 min before the predicted S2A and 3 min before Proba-V. In addition, the PROBA-V satellite overpassed the area ~7 min before S2A. The bare soil acquisitions were performed around one hour prior to both satellite overpasses.

The selected S2A product is the Level-2A (L2A) UTM tile T30SWJ acquired on the 2*nd* of August 2019. This product was downloaded from the Copernicus Open Access Hub and represents the output generated by the Sen2Cor processor (Main-Knorn et al., 2017). The product provides atmospherically-corrected reflectance images, approximating HDRF, and several auxiliary retrievals: aerosol optical thickness (AOT), water vapour, scene classification maps and quality indicators for cloud and snow probabilities.

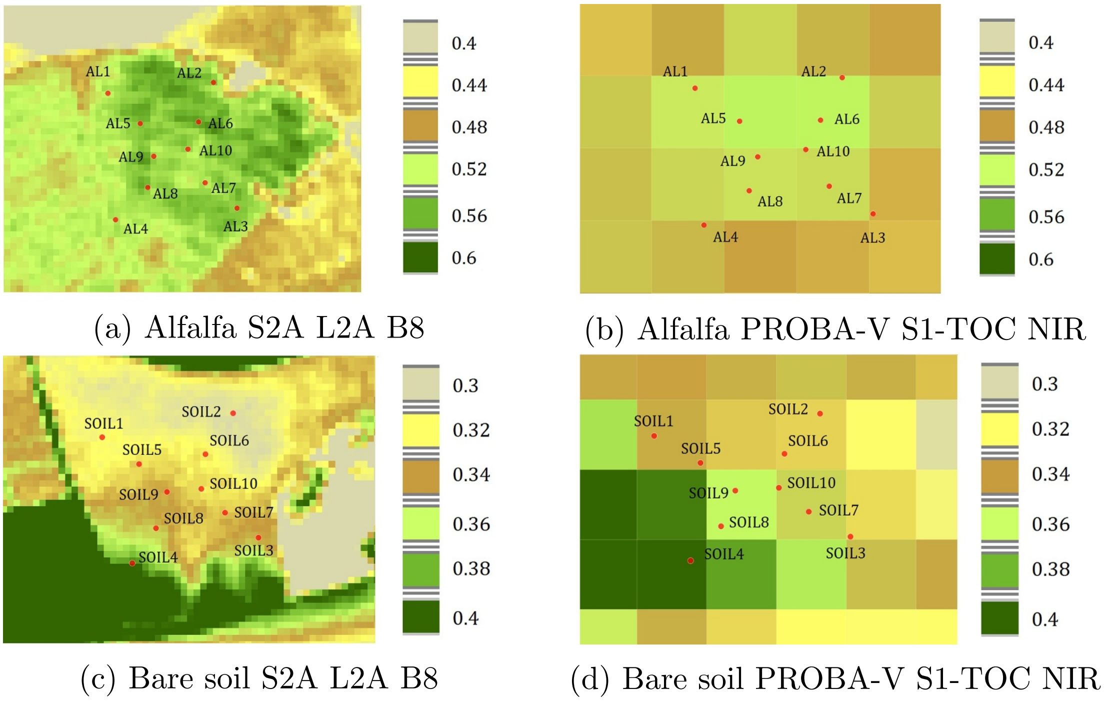
The selected PROBA-V product is the daily top-of-canopy synthesis (S1-TOC) version 101 acquired on the 2*nd* of August 2019, and was downloaded from the VITO product distribution portal. The product is the result of TOA reflectance corrected for the atmosphere and orthorectified at 100 m resolution (Sterckx et al., 2014).

The processing of both products was performed using the Sentinel Application Platform (SNAP) routines. Using the file flags, the Alfalfa and bare soil areas in S2A L2A product were checked for cloud, cirrus, saturated and defective pixels (none present). All pixels in the Alfalfa field were classified as "vegetation" whereas the pixels in the bare soil where classified as "bare soil", indicating that the land cover classification was working as expected. The approximate satellite view zenith angle was 6° in the forward scattering plane for all bands. In addition, for the S2A overpass the Barrax field site lies approximately in the middle of detector 10 (S2A MSI contains 12 staggered detectors) which minimises the slight spectral disturbances experienced close to the edge of the detectors.

The S1-TOC product quality was assessed by checking the status map band integrated within the product. It was verified that the measured areas have a value of 993 which corresponds to a clear land pixel with good acquisitions in all radiometric bands. PROBA-V views Barrax at 2° zenith angle for VNIR bands and 4° for SWIR bands. Both acquisitions are in the forward scattering plane.

2.8. Satellite data selection

One of the most critical steps when comparing the site measurements with the satellite overpass is the selection of the pixels in the satellite product. The measurement campaign was designed so that it



(a) Alfalfa S2A L2A B8

(b) Alfalfa PROBA-V S1-TOC NIR

(c) Bare soil S2A L2A B8

(d) Bare soil PROBA-V S1-TOC NIR

Fig. 5. Plots of Alfalfa S2A L2A band 8 (panels a and c) and the PROBA-V S1-TOC NIR band (panels b and d). The plot contains the positions of each campaign measurement point as well as the equivalent centre position of all the considered positions. The images are North-up oriented with a pixel size of 10 m^2 for S2A L2A band 8 and 100 m^2 for the PROBA-V S1-TOC NIR band.

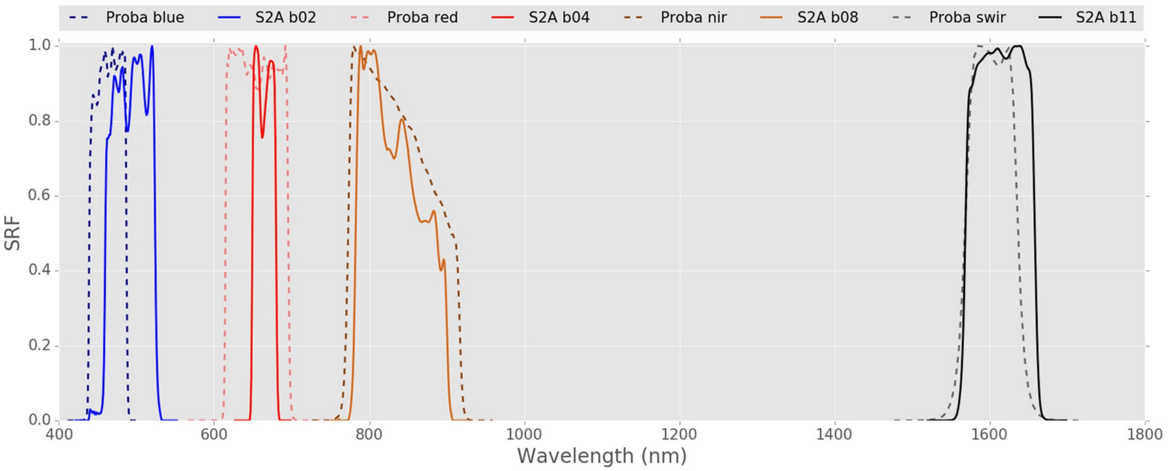


Fig. 6. PROBA-V and S2A spectral response functions for the bands considered for the location offset correction.

could represent a rectangular area, with North-up orientation, of approximately 200 × 200 m^2 on the Alfalfa and bare soil surfaces. The criteria was based on the need to represent one or more 100 m pixels of PROBA-V S1-TOC (see Section 2.3).

Fig. 5 displays a screenshot from the SNAP platform of the Alfalfa and bare soil for the S2A L2A B8 band and the PROBA-V S1-TOC NIR band. Each of the campaign measurement points are shown on both images.

Fig. 5a shows the predominantly random variability inside the Alfalfa, while a sudden decrease of reflectance occurs at the edge of the crop corresponding to the crop/path transition. In addition, other reflectance changes related to harvesting and/or irrigation channels can be seen (see lower right part of the S2A B8 image in Fig. 5a).

For the S2A comparison, two comparisons are tested. The first method is based on comparing the same area that the measurement campaign covers (i.e. point measurements are aggregated). Mathematically, this can be written as:

$$b = 100\cdot\left[\frac{\overline{R}_{S2A\_L2A}}{\overline{R}_{g,HDRF}} - 1\right] \tag{5}$$

Where b gives the bias (in %), $\overline{R}_{S2A\_L2A}$ refers to the S2A L2A surface reflectance in the 200 × 200 m^2 rectangle and $\overline{R}_{g,HDRF}$ refers to the average HDRF over the included sample points; the individual HDRF values are calculated according to Eq. (1). This formulation is given as mean error 1 or (ME1) in Figs. 8 and 9.

The second method compares the HDRF values calculated for each of the 10 locations with the corresponding S2A L2A pixel. The comparison can be defined as:

$$b = 100\cdot\frac{1}{10}\cdot\sum_{i=1}^{10}\left[\frac{R_{S2A\_L2A}}{R_{g,HDRF,i}} - 1\right] \tag{6}$$

Where $R_{S2A-L2A}(b,1pix)$ refers to the S2A L2A surface reflectance for a single pixel. This is given as the mean error 2 (ME2) in Figs. 8 and 9.

Since each of the measured points represents a ~3 m diameter circular area (i.e. based on sampling around the flag point within the four target measurements), this method requires the relatively high spatial resolution of the S2A bands where each measurement point covers a large fraction of a single S2A L2A pixel. The assumption here is that the use of several measurement points (in this case 10) reduces the impact of geolocation errors. That is, it must be assumed that the geolocation error and representativeness error are largely uncorrelated between measurement points (i.e. these effects are not included in the uncertainty budget of the individual estimates). The benefit of using such a comparison is that the campaign does not necessarily need to be performed in a uniform and continuous area. Instead discontinuities (e.g. irrigation canals, shelters, etc.) can be present in the measured area.

The bare soil area in Fig. 5 displays a larger variation showing an increasing reflectance gradiant from North to South. Specifically, it is noted that the South-West area shows a large reflectance area, probably influenced by the nearby dirt road located on the southern boundary of the bare soil. The criteria here is not to consider those measurement points that are close to the high reflectance area since the reflectance variations are very large. Consequently, the points "SOIL3", "SOIL4" and"SOIL8" were disregarded from the comparison and just 7 points were considered. The selected centre in the image was also shifted to account for the lower number of points.

A single PROBA-V S1-TOC pixel represents an area of 100 × 100 m^2. Selection of the pixels corresponding to the campaign measurements can introduce large discrepancies due the representativeness of the area, especially if the in situ sample area overlaps multiple pixels. Since the S2A and PROBA-V satellites overpass the area with a small time difference (~20 min and ~4° zenith angle in the same scattering plane), an empirical correction was calculated based on the ratio of S2A L2A pixels as follows:

$$b = 100\cdot\left(\frac{\overline{R}_{S2A\_L2A}}{R_{S2A\_L2A}(S1\_TOC)}\right) \tag{7}$$

Where $R_{S2A\_L2A}(S1\_TOC)$ represents the S2A L2A 10 m pixels in an area equivalent to the S1-TOC pixels. Consequently the offset location of the S1-TOC pixels can be estimated. The ratio is performed using 4 of the 10 m bands of S2A which are spectrally close to the PROBA-V bands as shown in Fig. 6. The resulting correction values are shown in Fig. 7.

The ratios are, as expected, low for the Alfalfa field since the variation over the measured area is largely random. However, the correction for the bare soil is considerable (at the 5% level) due to the systematic variation that can be seen in Fig. 5c and d.

2.9. *Satellite data uncertainty propagation*

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 three variables are considered uncorrelated. Since there is a lack of uncertainty estimates provided with the latter variables, their error distributions are assumed to be Gaussian with the standard deviations being taken from the S2 L2A uncertainty requirements given in Clerc et al. (2019). In that document the AOT requirement is given as:

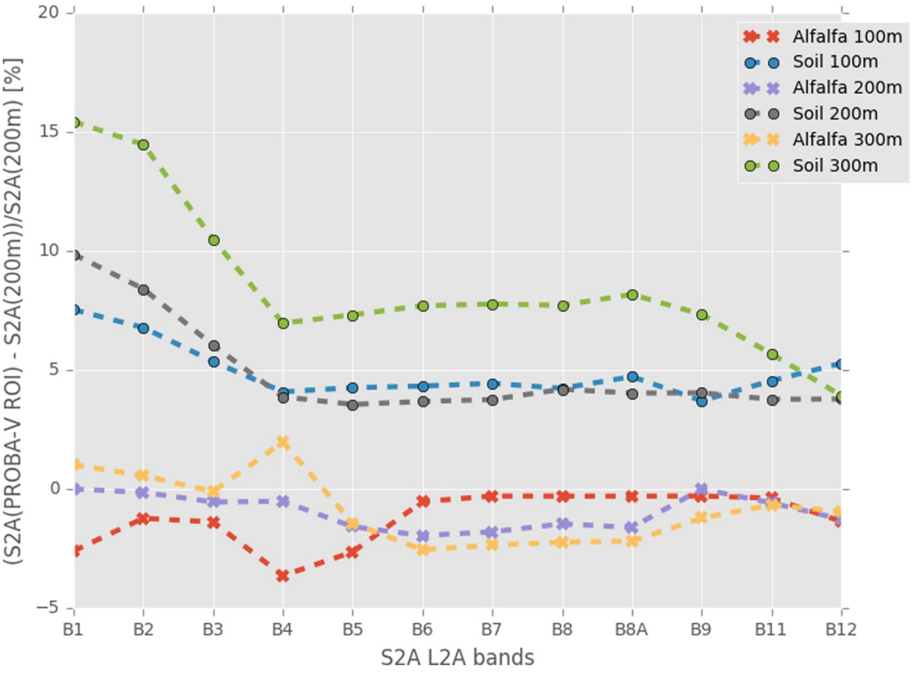


Fig. 7. Offset location correction of PROBA-V S1-TOC pixels based on a S2A ratio for the two areas measured on the 2*nd* of August 2019 in Barrax.

$AOT_{req} = 0.1AOT + 0.03$ while the water vapour requirement is given as $WV_{req} = 0.1WV + 0.2$. The AOT and water vapour error distributions are centred on the values retrieved from the original image by Sen2Cor.

The input TOA reflectance factor uncertainties, derived from the S2 RUT (Gorroño et al., 2018), are used to draw 600 error samples which are saved as separate image files. Traditionally these would be processed by the Sen2Cor atmospheric correction routine, along with the variation associated with the AOT and water vapour. However, since Sen2Cor retrieves the AOT and water vapour directly from the image, doing this would amount to assessing the transformation of the L1 uncertainties through the atmospheric correction procedure *without* accruing the additional uncertainties inherent in the selection of the AOT and water vapour. For this, the 6SV1 atmospheric correction procedure (Vermote et al., 1997; Wilson, 2013) was selected in order to allow the incorporation of these additional uncertainties. 600 samples are taken from the AOT and water vapour error distributions and, combined with the TOA reflectance factor samples, processed using 6SV1 to produce the output distributions. This is a simplified but effective approach given the current limitations. Finally, an uncertainty associated with the quality of the Sen2Cor algorithm is included, this was derived from Richter and Schläpfer (2016), and is given as 0.02 for $HCRF < 0.1$, 0.04 for $HCRF > 0.4$ and 0.03 for $0.1 < HCRF > 0.4$. This value is derived directly for the Barrax field site under similar conditions.

The Proba-V product does not provide uncertainty or any reasonable means in which to estimate it. As such, this has been omitted from the analysis; uncertainty bars from the Proba-V comparison are solely from the in situ data.

3. Results and discussion

3.1. Comparison against field measurements

3.1.1. Sentinel-2

The high spatial resolution of the S2A instrument facilitates the ability to compare individual pixel values against in situ measurements from a single sample location. Comparison of each location against the corresponding S2A pixel values as well as the site averages compared to the 200 × 200 m S2A average is presented in Figs. 8 and 9. Fig. 8 gives the comparison over the Alfalfa field. Here, all bands and locations show good agreement within the stated uncertainty ($k = 2$) with respect to the 1:1 line (dashed), and thereby cover the S2 L2A product requirements (red shaded region; $k = 2$). The variability in the individual mean HDRF values is likely to be explained by a combination of geolocation and directionality issues which are acute at this scale. The higher biases are also generally found in the visible bands where the reflectance signal is much lower, while the bands located beyond the red-edge show much lower biases on account of the increased signal from the target.

The main reason for the conformity seen in Fig. 8 is the large uncertainties given for the S2 data and ASD reference measurements. The larger ASD uncertainties (e.g. for B6–B8A) are a product of including the uncertainty associated with making the two variables equivalent, into the reference measurements (e.g. correction for SZA). Although the overall uncertainty doesn't get much larger than~30% ($k = 2$). However, for low signal bands (B1–B5) the S2 uncertainty dominates, with uncertainties close to 200% found. This is primarily driven by the uncertainty associated with the atmospheric correction (and specifically the accuracy of the atmospheric correction code) which can be up to 158.2% (at $k = 2$; Table 3). For comparison, the uncertainty associated with the top-of-atmosphere data is between ~2 and 6 % for the high signal bands and up to 26.6% for the visible bands (all at $k = 2$). The large S2 uncertainties are particularly problematic for B1–B5 and B9–B12 where the uncertainty associated with the average value is larger than the S2 requirements. Here the uncertainty region defined by the ellipse area marginally extends beyond the lower and upper limits of the requirement.

The change in window size of the S2A region of interest (ROI; tested at 60 m, 100 m, 180 m, 200 m; results not shown) has a negligible effect on the result showing that the Alfalfa site variability is stable at various scales below the full measurement area. This is particularly important for the visible bands where the low signal level increases the variability; this averages out over several pixels.

The reasonable results shown in the Alfalfa are not upheld over the bare soil (Fig. 9). Here the S2 L2A product consistently overestimates the HDRF values relative to the in situ estimates, evidenced by the presence of most sample points in all bands being above the 1:1 line and

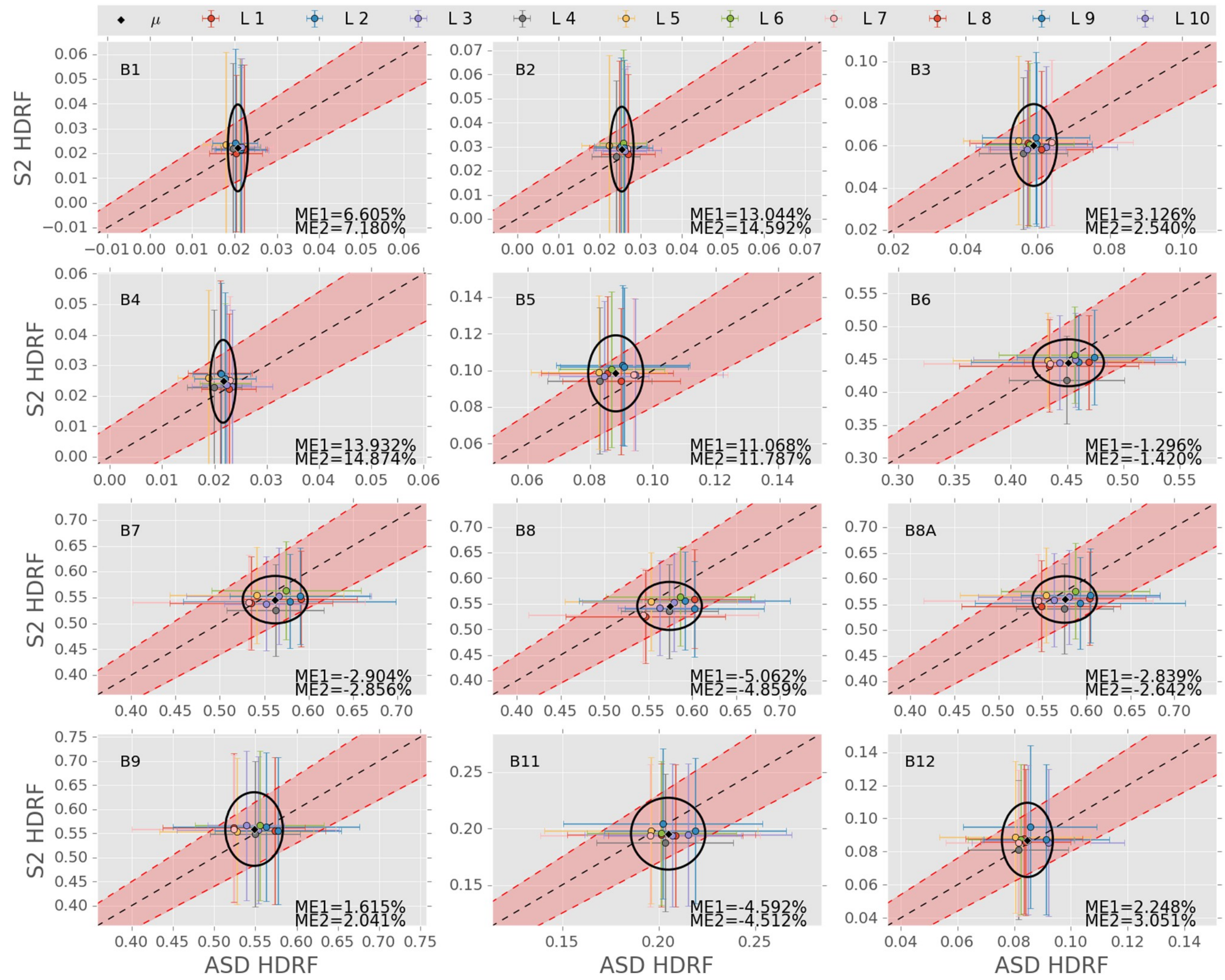


Fig. 8. Comparison of ASD and S2 surface reflectance over Alfalfa for each location (circles) and the average over all locations and 200 × 200 m^2 (diamond). The ellipse gives the $k = 2$ uncertainty of the average. The red shaded region gives the S2A L2A requirements on the pixel level uncertainty ($k = 2$; ±5% + 0.005 (at $k = 1$); Clerc et al., 2019). The error bars on the individual points give their individual uncertainty ($k = 2$). ME1 and ME2 refer to the calculations given in Eqs. (5) and (6) respectively.

in some cases outside of the the S2A requirements, despite a general agreement within the uncertainties. The main reason for the disagreement is likely to be the mismatch in measurement time, since the campaign was designed to cover the Alfalfa during the S2 overpass. We were also not able to correct for bare soil BRDF effects, which may be significant given the change in solar zenith angle between our measurements and the satellite overpass, as well as the slightly off-nadir satellite viewing geometry.

Since the biases shown in Fig. 9 are of a systematic nature, no improvement to the agreement is expected when aggregating (diamond marker in plot). As a result, the closer agreement in bands 9 and 12 shown in Fig. 9 are also seen here. As with the Alfalfa, changing the size of the ROI in the S2 data did not alter the results significantly.

Unlike the Alfalfa results, the uncertainty is consistently dominated by the S2 uncertainty (i.e. in all bands). This is particularly pronounced for the visible bands where the effect of the atmospheric correction is largest (see Table 3). Once again, the problems related to the size of the uncertainty relative to the S2A requirements highlighted for the previous figure are shown to hold here.

The contrast in variability between the two sites is reflected in the proportion of individual measurement points that fall within the uncertainty ellipse defined by the average HDRF. Fig. 8 (Alfalfa) has the majority of points falling within the uncertainty ellipse, while Fig. 9 has fewer for all bands. This reflects the systematic nature of the variation seen in the bare soil plot.

3.1.2. Proba-V

Table 4 provides the comparison against the Proba-V S1-TOC product over Alfalfa and bare soil.

Similar to the area based comparison for S2, the window size has limited effect on the Proba-V comparison (not shown) partly due to the empirical correction using S2 images (see Section 2.8). All bands are within 10% for bare soil (Table 4) while the NIR band is within 10% for Alfalfa (Table 4). However, the blue and red bands show a disagreement at the 85–90% level for the Alfalfa measurements (Table 4) while for the SWIR band this drops to ~18%. The lowest biases can be found for the bands that are measuring high signal radiance (bare soil and NIR in the Alfalfa). This may be caused by a baseline error which produces larger relative errors when there is low signal (examples of potential sources would be spectral stray light at the instrument level and residual atmospheric correction errors). In contrast to S2, the best results here are found over the bare soil region, despite the mismatch in measurement and overpass time, and the steep reflectance gradient over the measured area.

These results suggest that the atmospheric correction of PROBA-V S1-TOC might retrieve surface reflectance with an absolute offset.

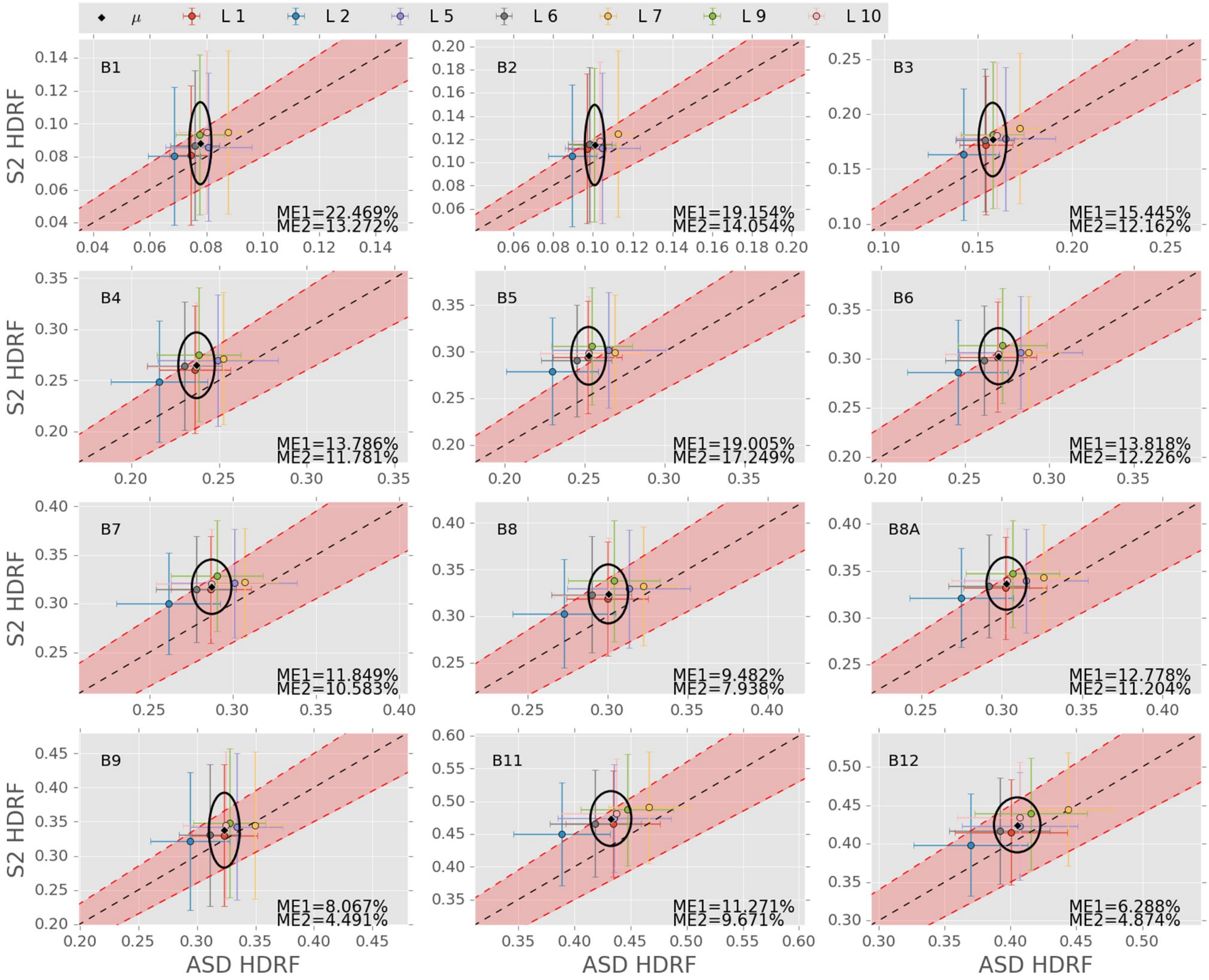


Fig. 9. Comparison of ASD and S2 surface reflectance over bare soil for each location (circles) and the average over all locations and 200 × 200 m^2 (diamond). The ellipse gives the $k = 2$ uncertainty of the average. The red shaded region gives the S2A L2A requirements on the pixel level uncertainty ($k = 2$; ±5% + 0.005 (at $k = 1$); Clerc et al., 2019). The error bars on the individual points give their individual uncertainty ($k = 2$). ME1 and ME2 refer to the calculations given in Eqs. (5) and (6) respectively.

Table 3

Summary results of the uncertainty retrieved from the procedure described in Section 2.9. The table includes, for both the Alfalfa and soil cases, the theoretical dispersion at the retrieved surface reflectance due to the dispersion of TOA reflectance values (*TOA* column), the dispersion of TOA reflectance values plus AOT and WV requirements (*TOA + AOT + WV* column), and finally including an allocation for the accuracy of the atmospheric correction code (*TOA + AOT + WV + AC* column). Values given as $k = 1$.

S2A	$\sigma(\rho^{6S}_{L2A})_{Alfalfa}$			$\sigma(\rho^{6S}_{L2A})_{Soil}$		
	TOA	TOA + AOT + WV	TOA + AOT + WV + AC	TOA	TOA + AOT + WV	TOA + AOT + WV + AC
B1	13.3%	19.6%	79.1%	5.4%	6.7%	26.1%
B2	9.6%	14.3%	61.4%	4.0%	4.8%	28.8%
B3	4.6%	6.6%	32.0%	2.5%	2.9%	18.4%
B4	6.3%	9.7%	54.9%	1.8%	1.9%	11.9%
B5	2.7%	3.7%	21.2%	1.6%	1.7%	10.3%
B6	1.6%	1.7%	8.0%	1.6%	1.7%	9.3%
B7	1.4%	1.4%	8.4%	1.6%	1.7%	8.7%
B8	1.6%	1.8%	8.6%	1.9%	2.0%	9.6%
B8A	1.7%	1.8%	8.1%	1.9%	1.9%	8.2%
B9	2.8%	11.2%	13.7%	3.0%	11.5%	15.7%
B11	2.2%	2.3%	16.3%	1.8%	1.8%	8.7%
B12	3.2%	3.3%	25.85%	2.0%	2.1%	8.3%

Table 4

Comparison of ASD and Proba-V surface reflectance over Alfalfa over the 200 × 200 m^2 area given as $\frac{Proba - ASD}{ASD}$. Uncertainty, from the in situ data only, is given at $k = 2$.

Band	Alfalfa	Bare soil
Blue	89.82 ± 20.55 %	4.21 ± 0.42 %
Red	85.15 ± 19.46 %	4.47 ± 0.39 %
NIR	-8.99 ± 1.15 %	9.06 ± 0.74 %
SWIR	18.12 ± 3.43 %	8.13 ± 0.69 %

However, another confounding factor is the inability to assess the Proba-V product uncertainty which means that the uncertainty on the biases shown in Table 4 is underestimated (potentially significantly). Likewise, and unlike the S2A L2A product, pixel level uncertainty requirements are not publicly available to assess whether the satellite and in situ data agree within those requirements.

3.2. Ancillary characterisation

We have introduced two substantial components into the validation procedure designed for surface reflectance products which include the

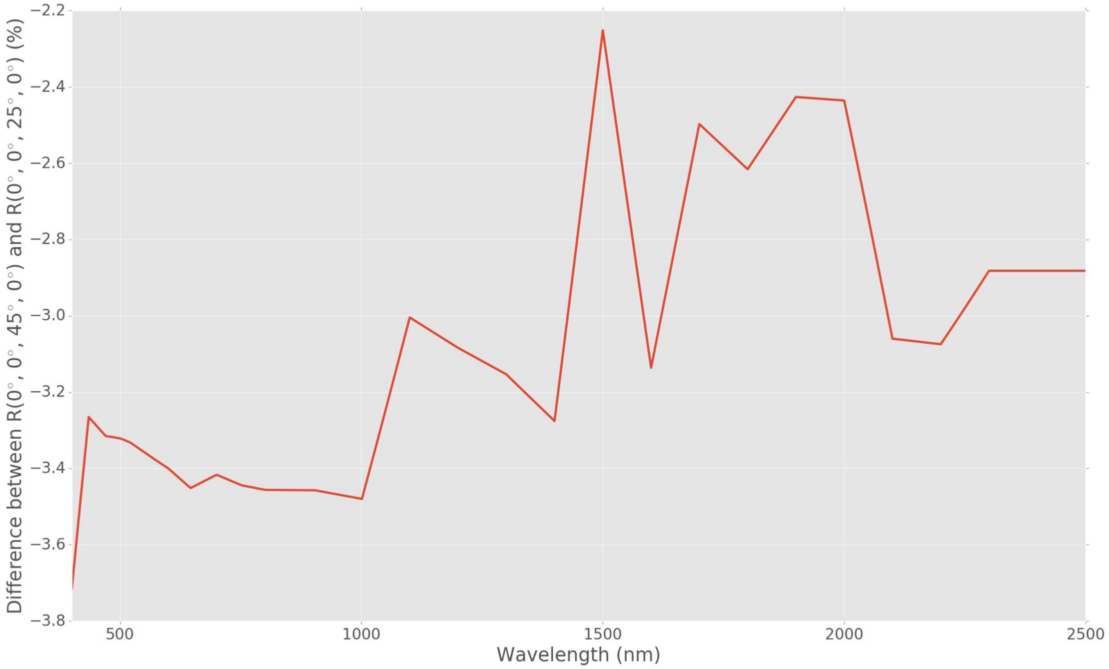


Fig. 10. Percentage difference between the BRF at 45° (standard laboratory calibration angle) and 25° view zenith angle (average solar zenith angle during the field campaign). Refer to Fig. 4 for values.

diffuse component. The first is a correction for the offset in the panel characterisation illumination configuration and the solar zenith angle experienced in the field. Fig. 4 gives the BRF for various viewing configurations. This data formed the basis of solar zenith angle dependent direct beam calibration coefficients. The importance of this correction is reflected in Fig. 10 where the minimum error that would be incurred is greater than 2%. However, below 1500 nm this is generally greater than 3% which accounts for a significant proportion of the validation requirements.

Similarly, the inclusion of a calibration coefficient which treats the diffuse sky irradiance explicitly is important due to the difference between the direct and diffuse reflectivity of the panel (Fig. 11; top panel), and because the proportion of the diffuse irradiance present in the field varies spectrally (Fig. 11; middle panel). Overall, the panel TDR is lower than the direct beam reflectance factor at 25°, and for 45° at greater than 1000 nm (Fig. 11; bottom panel) so represents an important (and erroneous) increase in the reference panel reflectivity that would not be captured otherwise and would lead to an overestimation of the target HDRF. At longer wavelengths this difference can be substantial (between 2 and 13 %) depending on the angle. When considering both the diffuse proportion and reflectance difference between BRF and TDR this effect would be negligible at longer wavelengths but could lead to errors of up to ~0.75% for bands 1–3 (S2) and the blue band of Proba-V.

3.3. Procedural improvements

While the validation approach outlined in this paper represents a significant advance, there are several improvements that can be made to 1) remove biases and/or 2) improve the uncertainty budget.

We propose two key instrument-based improvements. The first requires a much narrower field-of-view fore-optic with added pointing capabilities. The basis for this stems from the difference between the FOV of a single S2 pixel and that of the ASD (~0.008° (10 m bands) and 8° respectively) and the placement of the relevant pixels within the instrument FOV. This means that the ASD fore-optic is closer to the overall S2 FOV (20.6°) than it is to an individual pixel and is therefore requiring a significant portion of the BRDF to be homogeneous. By combining this with nadir pointing, there is also a significant proportion of the signal coming from the scattering hemisphere opposite to that experienced by the spaceborne pixel, potentially introducing larger errors. These two components are magnified when the BRDF of the background is expected to deviate substantially from the Lambertian assumption, as in the case of vegetation. For the ASD there is limited scope for improvement since any significant constraint in the FOV of the fore-optic will require the measurement to be made at a height greater than possible when operated by a human. However, a gimbal with optical mount could be feasible to address the pointing issue. Another solution would be to use an Unmanned Aerial Vehicle (UAV) mounted imaging spectrometer, this way a single 10 × 10 m^2 pixel would represent an aggregate of several UAV-based pixels with a FOV approaching 1° being feasible. With this, removal of unsuitable viewing geometries can be implemented to retrieve the pixels best suited to the validation of the specific product.

A second improvement would be to derive the spectral diffuse-to-total irradiance ratio from a dedicated instrument. This would have the benefit of removing the requirement for atmospheric radiative transfer model simulations based on minimal data and make the in situ data fully independent of the satellite surface reflectance products which also use similar codes in their derivation. Superior to this would be a spectral sky radiance image which would allow a better calculation of the TDR reference value.

Additionally, incorporation of BRDF knowledge would facilitate an improved understanding of the impact of the angular uncertainty. Currently we assess the optic levelling uncertainty (combined with instrument noise) by retrieving the standard deviation of multiple scans and samples at each location. By including knowledge of the BRDF it would be possible to include extra terms into the measurement equation to account for this at the individual scan level while having appropriate sensitivity coefficients with which to scale them by. Admittedly, this may be difficult for plant canopies such as Alfalfa since they are heliotrophic, meaning that the plant moves in response to the changes in sun angle (Ehleringer and Forseth, 1980; Strub et al., 2003).

A further improvement would be to characterise the reference panel at longer wavelengths at a higher spectral resolution and with a lower uncertainty.

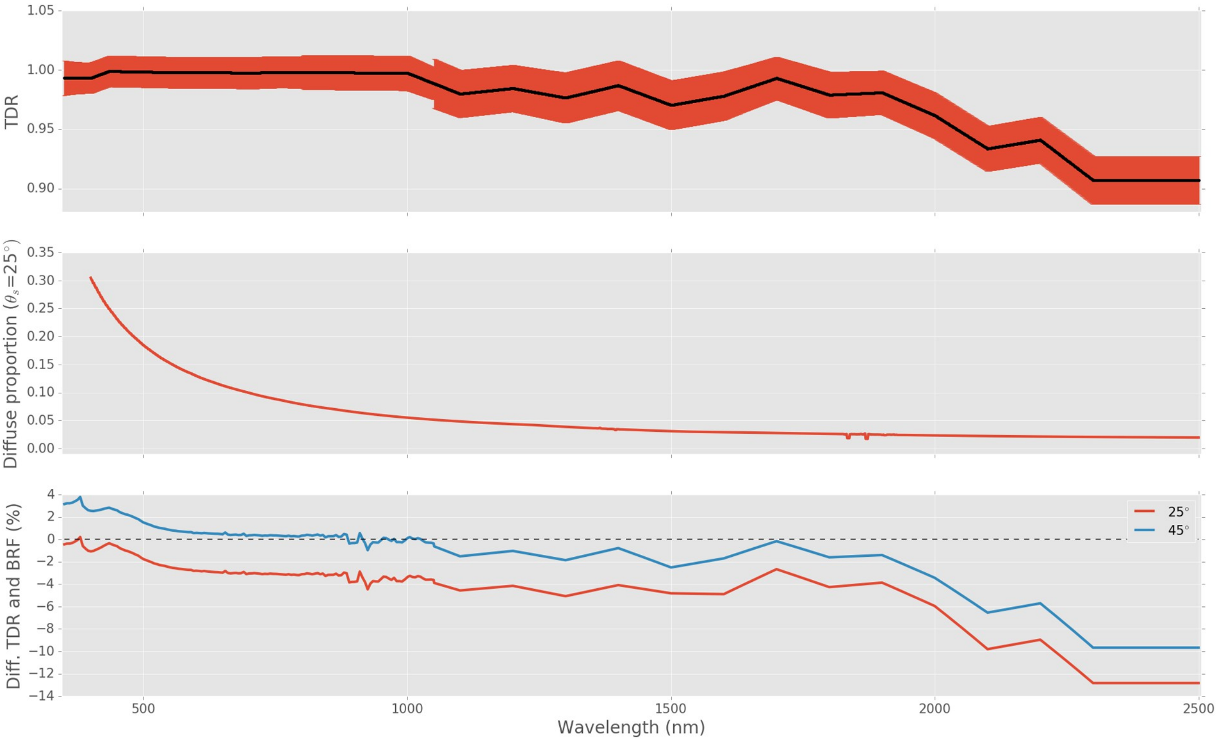


Fig. 11. The spectral total diffuse reflectance factor of the Spectralon panel where error bars indicate the uncertainty at $k = 2$ confidence interval (top); spectral diffuse-to-total irradiance ratio at 25° (middle; markers omitted for clarity; the uncertainty for the diffuse-to-total irradiance proportion is dependent on the time window, and therefore not included here); and the difference between TDR and BRF (bottom; again, markers omitted for clarity; positive values indicate TDR > BRF).

It is worth mentioning that improvements to the geolocation error, associated with the S2 Global Reference Image (Gascon et al., 2017), are likely to lead to improved comparison uncertainties as a result of better colocation between pixel values and field measurements.

4. Summary and conclusions

The present paper has demonstrated the fundamental steps required to perform an FRM validation of Sentinel-2 and Proba-V surface reflectance products. Firstly, we use a metrological approach that informs the calibration and characterisation of the reference panel based on the illumination conditions experienced in the field. Secondly, the products are assessed at the pixel level, from which the product requirements are made, for a high and high/medium resolution satellite instrument; as well as at the area level (as is done with conventional validation activities). Thirdly, we provide (where possible) an end-to-end uncertainty characterisation of the in situ and satellite data acquisitions. Lastly, we present suggested improvements to the methods we performed in order to aid users wishing to perform FRM grade validations in the future.

The results showed that the Sentinel-2 surface reflectance product agreed (within the stated uncertainty) with the ground data collected over the Alfalfa field at both the pixel and area scales. There was full agreement for both surfaces in terms of the product requirements and 1:1 line. The comparison was less convincing over the bare soil, however there is sufficient reason to believe that this is due to the mismatch between the overpass time and the in situ data collection. However, the large uncertainties encountered when factoring in the atmospheric correction uncertainty are a major driver in the agreement and are greater than the S2A requirements.

The results for the Proba-V showed general disagreement over all surfaces, but particularly large errors were seen over dark areas (i.e. Alfalfa in the blue and red band) and in the SWIR band. Part of this is likely due to the unavailability of Proba-V product uncertainty, which leads to underestimation of the comparison uncertainty. Similarly, publicly available product uncertainty requirements are currently not available to assess against.

In this paper, the L1 to L2 uncertainty (i.e. uncertainty associated with the atmospheric correction) is characterised using a simple Monte Carlo scheme, which benefits the interpretation of the validation results. Nonetheless, it should be developed further and released with the operational product in order to pass on to derived products, inform users of the true product uncertainty, and aid FRM-based validation.

This paper highlights the importance of adopting FRM principles into the validation design since the final comparison allows unequivocal determination of the product conformity (depending on the conformity criteria; see Widlowski (2015)). This allows users to easily assess whether a product meets their requirements. Similarly, by designing validation activities that minimise the overall comparison uncertainty, increased scrutiny is placed on the individual elements, meaning that improvements to the product and in situ data collection and processing can be easier to identify and implement.

CRediT authorship contribution statement

Niall Origo: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Javier Gorroño:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **James Ryder:** Conceptualization, Methodology, Investigation. **Joanne Nightingale:** Conceptualization, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Agnieszka Bialek:** Formal analysis, Investigation, Resources, Data curation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Fernando Camacho de Coca, José González Piqueras and Alonso Garrido Lamiñana for their support in arranging the fieldwork; Paul Green for his support in conducting the fieldwork; and, Jadu Dash and Luke Brown for supplying some of the equipment used in the fieldwork. The research leading to these results was partially funded by the EU and European Space Agency through the FRM4Veg project and the National Measurement System of the UK government's Department for Business, Energy and Industrial Strategy. The views expressed herein can in no way be taken to reflect the official opinion of the European Space Agency or the European Union.

References

AEMet, 2018. Radiation and Ozone. http://www.aemet.es/en/eltiempo/observacion/radiacion/ozono?l, Accessed date: 10 October 2018.

Bialek, A., Greenwell, C., Lamare, M., Marcq, S., Lacherade, S., Meygret, A., 2016. Namibia Field Campaign Technical Note. In: Technical report. National Physical Laboratory June.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., Weiss, M., 2013. GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: validation and intercomparison with reference products. Remote Sens. Environ. 137, 310–329. https://doi.org/10.1016/j.rse.2013.02.030.

Cescatti, A., Marcolla, B., Vannan, S.K.S., Pan, J.Y., Román, M.O., Yang, X., Ciais, P., Cook, R.B., Law, B.E., Matteucci, G., Migliavacca, M., Moors, E., Richardson, A.D., Seufert, G., Schaaf, C.B., 2012. Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network. Remote Sens. Environ. 121, 323–334. https://doi.org/10.1016/j.rse.2012.02.019.

Chunnilall, C.J., Deadman, A.J., Crane, L., Usadi, E., 2003. NPL scales for radiance factor and total diffuse reflectance. Metrologia 40 (1), 192–195. https://doi.org/10.1088/0026-1394/40/1/003.

Clerc, S., Devignot, O., Pessiot, L., Team, M.P.C., 2019. S2 MPC. Level 2A Data Quality Report. In: Technical Report S2-PDGS-MPC-L2ADQR. European Space Agency. https://sentinel.esa.int/documents/247904/685211/Sentinel-2-L2A-Data-Quality-Report, Accessed date: 19 June 2019.

Ehleringer, J., Forseth, I., 1980. Solar tracking by plants. Science 210 (4474), 1094–1098. https://doi.org/10.1126/science.210.4474.1094.

Fan, L., Berger, F.H., Liu, H., Bernhofer, C., 2014. Validating MODIS land surface reflectance products using ground-measured reflectance spectra a case study in semi-arid grassland in Inner Mongolia, China. Int. J. Remote Sens. 35 (5), 1715–1728. https://doi.org/10.1080/01431161.2014.882031.

Gascon, F., Bouzinac, C., Thepaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Tremas, T., Cadou, E., De Bonis, R., Isola, C., Martimort, P., Fernandez, V., 2017. Copernicus Sentinel-2A calibration and products validation status. Remote Sens. 9 (584), 1–81. https://doi.org/10.3390/rs9060584.

Gorroño, J., Peters, M., Fomferra, N., Fox, N., Gascon, F., 2018. A second version of the radiometric uncertainty tool for the sentinel-2 mission. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 6460–6463. https://doi.org/10.1109/IGARSS.2018.8518866. July.

Guillevic, P.C., Privette, J.L., Coudert, B., Palecki, M. a., Demarty, J., Ottlé, C., Augustine, J. a., 2012. Land surface temperature product validation using NOAA's surface climate observation networks - scaling methodology for the Visible Infrared Imager Radiometer Suite (VIIRS). Remote Sens. Environ. 124, 282–298. https://doi.org/10.1016/j.rse.2012.05.004.

JCGM, 2008. Evaluation of measurement data - guide to the expression of uncertainty in measurement. In: Technical Report. 100.

JCGM, 2012. Internationl vocabulary of metrology - basic and general concepts and associated terms (VIM). In: Technical Report. 200.

Jin, Y., Schaaf, C.B., Woodcock, C.E., Gao, F., Li, X., Strahler, A.H., Lucht, W., Liang, S., 2003. Consistency of MODIS surface bidirectional reflectance distribution function and albedo retrievals: 2. Validation. J. Geophys. Res. 108 (D5), 1–15. https://doi.org/10.1029/2002JD002804.

Kessler, P.D., Killough, B.D., Gowda, S., Williams, B.R., Chander, G., Qu, M., 2013. CEOS Visualisation Environment (COVE) tool for intercalibration of satellite instruments. IEEE Trans. Geosci. Remote Sens. 51 (1), 1081–1087. https://doi.org/10.1109/TGRS.2012.2235841.

Li, F., Member, S., Jupp, D.L.B., Reddy, S., Lymburner, L., Mueller, N., Tan, P., Islam, A., 2010. An evaluation of the use of atmospheric and BRDF correction to standardize landsat data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3 (3), 257–270. https://doi.org/10.1109/JSTARS.2010.2042281.

Liang, S., Fang, H., Chen, M., Shuey, C.J., Walthall, C., Daughtry, C., Morisette, J., Schaaf, C., Strahler, A., 2002a. Validating MODIS land surface reflectance and albedo products: methods and preliminary results. Remote Sens. Environ. 83, 149–162. https://doi.org/10.1016/S0034-4257(02)00092-5.

Liang, S., Fang, H., Morisette, J.T., Chen, M., Shuey, C.J., Walthall, C.L., Daughtry, C.S.T., 2002b. Atmospheric correction of landsat ETM + land surface imagery: II . Validation and applications. IEEE Trans. Geosci. Remote Sens. 40 (12), 1–10. https://doi.org/10.1109/TGRS.2002.807579.

Liu, J., Schaaf, C., Strahler, A., Jiao, Z., Shuai, Y., Zhang, Q., Roman, M., Augustine, J.A., Dutton, E.G., 2009. Validation of Moderate Resolution Imaging Spectroradiometer (MODIS) albedo retrieval algorithm: dependence of albedo on solar zenith angle. J. Geophys. Res. 114, 1–11. https://doi.org/10.1029/2008JD009969.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2cor for sentinel-2. In: Image and Signal Processing for Remote Sensing XXII. Proceedings of SPIE. vol. 10427https://doi.org/10.1117/12.2278218.

Malthus, T., Ong, C., Lau, I., Fearns, P., Byrne, G., Thankappan, M., Chisholm, L., Suarez, M., Clarke, K., Scarth, P., Phinn, S., 2018. A community approach to the standardised validation of surface reflectance data. A technical handbook to support the collection of field reflectance data. In: Technical Report 1.0. CSIRO, Australia.

Malvern Panalytical, 2019. ASD FieldSpec 4 Standard-Res Spectroradiometer. https://www.malvernpanalytical.com/en/products/product-range/asd-range/fieldspec-range/fieldspec-4-standard-res-spectroradiometer, Accessed date: 11 June 2019.

Merchant, C.J., Holl, G., Mittaz, J.P.D., Woolliams, E.R., 2019. Radiance uncertainty characterisation to facilitate climate data record creation. Remote Sens. 11 (474), 17. https://doi.org/10.3390/rs11050474.

Mittaz, J.P.D., Merchant, C.J., Woolliams, E.R., 2019. Applying principles of metrology to historical Earth observations from satellites. Metrologia 56 (3), 1–28. https://doi.org/10.1088/1681-7575/ab1705.

Nicodemus, F., Richmond, J., Hsia, J.J., Ginsberg, I., Limperis, T., 1977. Geometrical considerations and nomenclature for reflectance. In: Technical Report October. National Bureau of Standards, U.S. Department of Commerce, Washington, DC.

Nightingale, J., Boersma, F., Muller, J.-P., Compernolle, S., Lambert, J.-C., Blessing, S., Giering, R., Gobron, N., De Smedt, I., Coheur, P., George, M., Schulz, J., Wood, A., 2018. Quality assurance framework development based on six new ECV data products to enhance user confidence for climate applications. Remote Sens. 10. https://doi.org/10.3390/rs10081254.

Nightingale, J., Mittaz, J., Douglas, S., Dee, D., Ryder, J., Taylor, M., Old, C., Dieval, C., Fouron, C., Duveau, G., Merchant, C., 2019. Ten priority science gaps in assessing climate data record quality. Remote Sens. 11. https://doi.org/10.3390/rs11080986.

Richter, R., Schläpfer, D., 2016. Atmospheric/Topographic Correction for Satellite Imagery (ATCOR-2/3 User Guide, Version 9.0.2, March 2016). In: Technical report. ReSe Applications Schläpfer.

Román, M.O., Schaaf, C.B., Woodcock, C.E., Strahler, A.H., Yang, X., Hollinger, D.Y., Kolb, T.E., Meyers, T.P., Munger, J.W., Privette, J.L., 2009. The MODIS (Collection V005) BRDF/albedo product: assessment of spatial representativeness over forested landscapes. Remote Sens. Environ. 113 (11), 2476–2498. https://doi.org/10.1016/j.rse.2009.07.009.

Salomon, J.G., Schaaf, C.B., Strahler, A.H., Gao, F., Jin, Y., 2006. Validation of the MODIS bidirectional reflectance distribution function and albedo retrievals using combined observations from the aqua and terra platforms. IEEE Trans. Geosci. Remote Sens. 44 (6), 1555–1565. https://doi.org/10.1109/TGRS.2006.871564.

Schaepman-Strub, G., Schaepman, M.E., Painter, T.H., Dangel, S., Martonchik, J.V., 2006. Reflectance quantities in optical remote sensing-definitions and case studies. Remote Sens. Environ. 103, 27–42. https://doi.org/10.1016/j.rse.2006.03.002.

Sogacheva, L., Kolmonen, P., Virtanen, T.H., Rodriguez, E., Sundström, A., Leeuw, G.D., 2015. Determination of land surface reflectance using the AATSR dual-view capability. Atmos. Meas. Tech. 8, 891–906. https://doi.org/10.5194/amt-8-891-2015.

Sterckx, S., Benhadj, I., Duhoux, G., Livens, S., Dierckx, W., Goor, E., Adriaensen, S., Heyns, W., Hoof, K.V., Strackx, G., Nackaerts, K., Reusen, I., Achteren, T.V., Dries, J., Roey, T.V., Mellab, K., Duca, R., Zender, J., 2014. The Proba-V mission: image processing and calibration. Int. J. Remote Sens. 35 (7), 2565–2588. https://doi.org/10.1080/01431161.2014.883094.

Strub, G., Schaepman, M.E., Knyazikhin, Y., Itten, K.I., 2003. Evaluation of spectrodirectional alfalfa canopy data acquired during DAISEX ' 99. IEEE Trans. Geosci. Remote Sens. 41 (5), 1034–1042. https://doi.org/10.1109/TGRS.2003.811555.

Vermote, E., Kotchenova, S., 2008. Atmospheric correction for the monitoring of land surfaces. J. Geophys. Res. 113 (D23S90), 1–12. https://doi.org/10.1029/2007JD009662.

Vermote, E., Tanré, D., Deuzé, J.L., Herman, M., Morcrette, J.J., 1997. Second simulation of the satellite signal in the solar spectrum, 6s: an overview. IEEE Trans. Geosci. Remote Sens. 35 (3), 675–686. https://doi.org/10.1109/36.581987.

Vermote, E., Justice, C., Csiszar, I., 2014. Early evaluation of the VIIRS calibration, cloud mask and surface reflectance Earth data records. Remote Sens. Environ. 148 (2014), 134–145. https://doi.org/10.1016/j.rse.2014.03.028.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008.

Wang, Y., Lyapustin, A., Privette, J.L., Morisette, J.T., Holben, B., 2009. Atmospheric correction at AERONET locations: a new science and validation data set. IEEE Trans. Geosci. Remote Sens. 47, 2450–2466. https://doi.org/10.1109/TGRS.2009.2016334.

Wang, Y., Lyapustin, A.I., Privette, J.L., Cook, R.B., SanthanaVannan, S.K., Vermote, E., Schaaf, C.L., 2010. Remote sensing of environment assessment of biases in MODIS surface reflectance due to Lambertian approximation. Remote Sens. Environ. 114 (11), 2791–2801. https://doi.org/10.1016/j.rse.2010.06.013.

Widlowski, J.L., 2015. Conformity testing of satellite-derived quantitative surface variables. Environ. Sci. Pol. 51, 149–169. https://doi.org/10.1016/j.envsci.2015.03.018.

Wilson, R.T., 2013. Py6S: a Python interface to the 6S radiative transfer model. Comput. Geosci. 51, 166–171.

Zebner, H., Zambelli, P., Taylor, S., Nwaogaidu, S.O., Michelson, T., Little, J., 2007. Lahmeyer International Pysolar: staring directly at the sun since 2007. https://pysolar.readthedocs.io/, Accessed date: 4 October 2019.