Radiometric inter-sensor cross-calibration uncertainty using a traceable high accuracy reference hyperspectral imager

Javier Gorroño a,b,*, Andrew C. Banks a, Nigel P. Fox a, Craig Underwood b

a National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, UK
b Surrey Space Centre, University of Surrey, Guildford, Surrey GU2 7XH, UK

A R T I C L E   I N F O

Article history:
Received 22 December 2016
Received in revised form 22 May 2017
Accepted 6 July 2017

Keywords:
TRUTHS
Inter-sensor cross-calibration
Pseudo Invariant Calibration Site (PICS)
Radiometric uncertainty

A B S T R A C T

Optical earth observation (EO) satellite sensors generally suffer from drifts and biases relative to their pre-launch calibration, caused by launch and/or time in the space environment. This places a severe limitation on the fundamental reliability and accuracy that can be assigned to satellite derived information, and is particularly critical for long time base studies for climate change and enabling interoperability and analysis ready data. The proposed TRUTHS (Traceable Radiometry Underpinning Terrestrial and Helio-studies) mission is explicitly designed to address this issue through re-calibrating itself directly to a primary standard of the international system of units (SI) in-orbit and then through the extension of this SI-traceability to other sensors through in-flight cross-calibration using a selection of Committee on Earth Observation Satellites (CEOS) recommended test sites. Where the characteristics of the sensor under test allows, this will result in a significant improvement in accuracy. This paper describes a set of tools, algorithms and methodologies that have been developed and used in order to estimate the radiometric uncertainty achievable for an indicative target sensor through in-flight cross-calibration using a well-calibrated hyperspectral SI-traceable reference sensor with observational characteristics such as TRUTHS. In this study, Multi-Spectral Imager (MSI) of Sentinel-2 and Landsat-8 Operational Land Imager (OLI) is evaluated as an example, however the analysis is readily translatable to larger-footprint sensors such as Sentinel-3 Ocean and Land Colour Instrument (OLCI) and Visible Infrared Imaging Radiometer Suite (VIIRS). This study considers the criticality of the instrumental and observational characteristics on pixel level reflectance factors, within a defined spatial region of interest (ROI) within the target site. It quantifies the main uncertainty contributors in the spectral, spatial, and temporal domains. The resultant tool will support existing sensor-to-sensor cross-calibration activities carried out under the auspices of CEOS, and is also being used to inform the design specifications for TRUTHS.

© 2017 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

TRUTHS (Traceable Radiometry Underpinning Terrestrial and Helio Studies), is a proposed satellite mission led by the National Physical Laboratory (NPL), UK. This mission is designed to have sufficient accuracy to allow the unequivocal detection of trends, from a background of natural variability, in a number of key indicators of climate change in the shortest time possible, allowing verification of climate forecast models on decadal timescales (Fox et al., 2011). This would be achieved by establishing a fiducial reference data set of spectrally resolved incoming and outgoing solar radiation. In terms of Earth viewing radiance, the characteristics of this data set are: spectrally-resolved—5–10 nm Full Width Half Maximum (FWHM)—Earth radiances, continuously sampled (spectrally and spatially) with a Ground Instantaneous Field Of View (GIFOV) of approximately 50 m over the 320–2400 nm spectral range, and the corresponding solar spectrally-resolved irradiance; both with SI-traceable radiometric uncertainties of <0.3% (Fox et al., 2011). These fiducial data sets establish a high accuracy benchmark of the Earth’s spectral radiation budget in the solar spectral domain in a similar manner to its US-proposed sister mission Climate Absolute Radiance and Refractivity Observatory (CLARREO) against which future change can be detected (Wielicki et al., 2013). The chosen spectral and spatial resolutions are optimum to allow the data sets to be utilised to retrieve many Essential Climate Variables

* Corresponding author at: National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, UK.
E-mail addresses: javier.gorrono@npl.co.uk (J. Gorroño), andrew.banks@npl.co.uk (A.C. Banks), nigel.fox@npl.co.uk (N.P. Fox), c.underwood@surrey.ac.uk (C. Underwood).

0924-2716/© 2017 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
(ECVs)—as defined by Global Climate Observing System (GCOS)—and facilitate detailed analysis of attribution effects (GCOS, 2010) and the Earth system’s cycles and processes.

It is thus not surprising that TRUTHS’s observational specifications—spatially and spectrally—match/allow reconstruction of many of the current, and planned, solar domain EO sensors, such as Landsat-8 (L8) Operational Land Imager (OLI). However, the addition of high SI-traceable radiometric accuracy in the reference sensor, maintained throughout the mission lifetime, also provides a powerful opportunity to cross-calibrate other sensors through co-incident viewing of stable target scenes and in particular, the radiometric characterisation of Pseudo Invariant Calibration Sites (PICS). For target sensors, such as Copernicus Sentinel-2 (S2) Multispectral Imager (MSI) and Sentinel-3 (S3) Ocean and Land Colour Instrument (OLCI), TRUTHS allows not only an assessment of performance but also a calibration upgrade towards that needed by many climate studies, and thus leads to the prospect of a space-based climate and calibration observatory as requested by the international community (Dowell et al., 2013).

The existing on-board calibration systems of many sensors such as Sentinel-2 and 3 have significant complementary merit, allowing assessment of any short term performance variation of the sensor over its full orbital path and between reference calibrations. In these cases, TRUTHS provides the in-flight anchor to SI units and the prospect of a regular update of the on-board monitoring systems. However, for sensors whose primary objectives do not warrant an on-board calibration system, such as the UK-DMC (Disaster Monitoring Constellation) series, similar cross-calibration activities would provide the means to achieve radiometric traceability, broadening the scope of application of such sensors, even to the point where these sensors could contribute towards climate studies and services. Following this logic, a constellation of new generation, low-cost Cube-/Nano-Sats could be envisaged, also contributing to the global observing system, radiometrically anchored to a reference sensor such as TRUTHS.

The ideal configuration for vicarious target inter-calibration is that the two instruments should make matched measurements viewing the same target at the same time; with the same spatial and spectral responses at the same viewing geometry. Since these idealized conditions never occur in reality, there will always be some additional compensatory steps needed to allow comparison of the two instruments. The accuracy achievable by the target sensor via the inter-sensor cross-calibration is ultimately limited by the reference sensor accuracy and the inability to fully account for the differences from the ideal comparison conditions. These differences include the instrument spectral response, target site spectral signature and the radiometric properties of the selected target site for the calibration process, including effects of solar illumination and sensor view angles and any variance in the atmosphere transmittance between the observations by the two sensors. Similar conditions apply even when the reference sensor measurements are used only as an input for the radiometric characterisation of PICS. In that situation, the longer term temporal radiometric properties of the site and its atmosphere become relevant factors.

In a recent study by Chander et al. (2013a), the uncertainty introduced by the main effects inherent in the cross-calibration transfer using a calibration target site was assessed to fall well below an uncertainty level of 0.3% (k = 1) with the exception of a spectral shift in SBAF. In that case, the use of filters such as those used in the Moderate Resolution Imaging Spectrometers (MODIS) (often used as a reference sensor) have suggested worst-case tolerances/shifts of 5 nm in the bands would produce larger differences. As a result, the uncertainty associated with the calibration of the reference sensor is now often the dominant component in the final uncertainty achieved for the test sensor.

The calibration accuracy of sensors measuring in the visible/near infrared (VNIR) and shortwave infrared (SWIR) spectral regions increased notably in the last decades. MODIS on board the Terra and Aqua satellites, or the recently launched S3 OLCI, have requirements for calibration accuracy of below 2% (k = 1) relative to the sun (Donlon et al., 2012; Xiong and Barnes, 2006). Instruments such as the Clouds and the Earth’s Radiant Energy System (CERES) have even more stringent calibration accuracy requirements—calibration accuracy below 1% (k = 1)—have highlighted the need for a reliable inter-calibration with an instrument like TRUTHS or CLARREO to overcome the data gap between the CERES mission instruments, to maintain the demanding stability requirements needed for climate (Loeb et al., 2016). Even if these well-calibrated instruments are used for cross-calibration their accuracy levels remain the dominant contribution to the total uncertainty in the cross-calibration process compared to the ones described in Chander et al. (2013a). Thus, the possibility of a reference instrument like TRUTHS or CLARREO with a radiometric uncertainty below 0.3% (k = 2) would be of a large benefit to reduce the total uncertainty in a cross-calibration over PICS.

This paper addresses the uncertainty contributions affecting typical CEOS WGCV recommended land-based reference sites in its use for cross-calibration of satellite imagers in the three main domains: spectral, spatial, and temporal. The aims of this paper are to: (1) evaluate the “inherent” uncertainty contributions with case studies (2) set up a suite of tools and methodologies useful for the exploitation and design of missions like TRUTHS or CLARREO, and (3) define the uncertainty contributions in a cross-calibration using rigorous metrology. Spectral, spatial and temporal contributors are all considered separately in Section 2.

For the latter point, the uncertainty propagation is based on the Monte-Carlo Method (MCM) as described in Supplement 1 to the Guide to the Expression of Uncertainty in Measurement (GUM) (BIPM et al., 2008b), the use of which is explicitly encouraged in the Quality Assurance Framework for Earth Observation (QA4EO) (http://www.QA4EO.org). Thus, the cross-calibration uncertainty estimates are presented in terms of a probability distribution function (pdf) of the associated parameters. The uncertainty is reported as the interval around the best estimate that approximates a coverage of 68.27% (which is expressed as k = 1). The coverage factor, k, is a numerical factor that multiplies the combined standard uncertainty in order to specify the fraction of the probability distribution that the uncertainty represents.

The MCM uncertainty propagation is a well-described technique which has historically been limited by the computing resources available. The rapid development of computing capabilities in recent years has made it more accessible to the EO community. The quantification and analysis of the uncertainty contributors developed as a software tool here require access to a large amount of memory and CPU time and have thus utilised the UK’s JASMIN supercomputer facility (Lawrence et al., 2013). The high-performance of the computer nodes permits the management of large quantities of memory, while a cluster of virtual and physical machines sharing a dedicated network, permits the parallel processing of the MCM algorithm.

The terms uncertainty, error and bias appear throughout this paper and are extensively analysed. We briefly define these terms here for clarity. Uncertainty expresses the degree of doubt around the measured value and can be reduced by thorough identification and correction of measurement errors. Error is the effect of measurement imperfection and can be systematic or random in nature. The random error can be minimised by using a large statistical sample. Bias is an estimate of a systematic error. These two terms will be used through the document and, in many cases, the difference between them will lead to a slightly different interpretation. The major biases in satellite cross comparison are introduced by
differences in measurement conditions or instrument specifications between the two considered sensors. If the bias is corrected for, only its residual must be taken into account in the uncertainty budget (BIPM et al., 2008a).

2. Uncertainty assessment

2.1. Spectral domain

2.1.1. Spectral domain: methodology

This section assesses the effect of spectral mismatch between a TRUTHS-like sensor and a target sensor (Sentinel-2 MSI) in the context of the chosen test-site’s spectral properties. Specifically, it studies the capacity of a TRUTHS-like sensor to derive a continuous Top-Of-Atmosphere (TOA) reflectance factor/radiance spectrum and the effect that it introduces in a cross-calibration with a sensor like S2 MSI. The effect of such differences between the band spectral response functions (SRFs) for the reference and target sensors is traditionally compensated for using the Spectral Band Adjustment Factor (SBAF), which is calculated from the known SRFs for each sensor and the spectral radiance of the test site being measured (Chander et al., 2013b). Here, this approach has been adapted to understand the achievable accuracy of the TRUTHS sensor in a cross-calibration with a target sensor. The process is similar to that applied in Green (1998) and is illustrated in Fig. 1: a reference TOA radiance spectrum is generated and convolved with the spectral bands of the TRUTHS sensor, the values from each band are binned as required, then used to reconstruct a hyperspectral curve via interpolation, and this reconstructed curve is then convolved with the target sensor bands. Rather than using a specific simulation, the whole range of potential cases are studied to derive the uncertainty introduced in the spectral dimension. That means that the simulations cover different wavelength positions of the TRUTHS sensor SRF, the reconstruction is set using different interpolation techniques in Section 2.1.2, and the central wavelength and bandwidth of each spectral band have an associated uncertainty in Section 2.1.3.

In order to study the spectral error introduced, a simulated TOA radiance spectral profile was generated and used as a reference. The simulation was initiated with the following conditions: Viewing Zenith Angle (VZA) = 4.60°, Sun Zenith Angle (SZA) = 21.44°; Relative Azimuth Angle (RAA) = 179.22°, Relative Zenith Angle (VZA) = 4.60° for day number 173 (summer solstice), and time = 8:54:53 GMT—similar to a particular Landsat 8 OLI overpass of Libya-4. The spectral resolution was set to the highest MODTRANv5.3.3 spectral resolution of 0.1 cm⁻¹ in order to extract the maximum information (Berk et al., 2005).

The TOA spectral radiance from MODTRAN was further interpolated to 0.0005 nm using linear interpolation over the VNIR and SWIR range. Such a low resolution—0.002 nm at around 450 nm—can capture reasonably detailed information relating to atmospheric and solar features. The sampling is more than twice the original MODTRAN output and uses a linear interpolation meaning that oversampling does not alter the original absorption line structure. The requirement for such a fine resolution derives from the possibility of describing the instrument spectral knowledge uncertainty as a distribution of errors in Section 2.1.3. In addition, the benefit of this high resolution for the impact of sampling/resolution will be discussed in Section 2.1.2.

The sampling and binning are set to values representative of the preliminary design of the TRUTHS satellite-borne imaging spectrometer, see Fig. 2.

The proposed native sampling and resolution of the instrument is used to generate triangular response functions, to model the real TRUTHS spectrometer response. The very low aberration of the candidate spectrometer means the SRF actually achieved is likely to be very close to this idealised triangular response; ultimately, the instrument spectral line characterisation will determine the precise shape. The preliminary design uses matching slit and pixel width dimensions; hence it is possible to approximate the pixel spectral bandwidth by the native sampling interval.

The TRUTHS SRF is then convolved with the site TOA radiance, to produce an instrument “as-measured” TOA radiance. The instrument response is further binned—the binning is set by design to achieve the optimum spectral sampling and Signal to Noise Ratio (SNR)—to emulate the TOA measurement of the TRUTHS bands. The process initiates at around 400 nm and iteratively moves up to the SWIR range stopping at around 2500 nm.

In order to derive a continuous TOA radiance spectrum a fitting or an interpolation can be used. Without any further information

![Fig. 1](image-url). Illustrative method of the TOA TRUTHS spectral profile generation. The red stars are the measurements at the native (We used the word “native” through the paper to indicate the measurement, spectral resolution and spectral sampling prior to any binning or post-processing.) spectrometer bands and the green stars are the result of the merging to a design specified bin. The merged measurements are sampled at a specified interval to obtain the reconstructed TOA as measured by TRUTHS (“TRUTHS TOA 85 x 10⁻⁶”). For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
and given an extremely high accuracy of the sensor, an interpolation is used here. Nonetheless, other sophisticated fitting methods as described in McCorkel et al. (2013) have been successfully applied and could, if prior spectral shape is accurate, reveal further information of the calibration site.

Fig. 3 provides the “as measured” TOA radiance by TRUTHS around each of the Sentinel 2 bands. The figure also overlays the MODTRANv5.3.3 TOA radiance used as a reference.

### 2.1.2. Spectral domain: systematic sampling/ resolution error results

Once the TOA radiance “as-measured” is obtained, the measurements of the TRUTHS bands are interpolated at 0.0005 nm resolution to match the original TOA spectral distribution generated using MODTRANv5.3.3 as described previously. The “true” radiance spectrum, generated using the MODTRAN reference, and this “reconstructed” spectrum are then used to study the impact on the cross-calibration of the sampling/resolution of the TRUTHS bands by convolving each with the SRFs for S2 MSI. The resulting difference between the convolved values for the “true” and “reconstructed” spectra are shown in Fig. 4. This process is undertaken for all the S2 bands except bands 9 and 10. These are centred on the atmospheric water absorption bands but are not intended to provide accurate radiometric measurements of the water vapour level, but rather the detection of clouds in the scenes (Drusch et al., 2012).

The starting wavelength of the TRUTHS bands can be set at a specific spectral position—referred here as wavelength position shift—from the starting wavelength of 410 nm—to simulate all the potential positions of the spectrometer bands. That is, it represents the alignment position of the detector array at the focal plane. This has an impact on the results where spectral structure is found in the observed scene. The results of simulations in Fig. 4 shows the error for all the positions in between two native spectral bands separated by the maximum sampling provided in Fig. 1. Several interpolation method and combinations have been used in the reconstruction of the TOA radiance from the TRUTHS binned measurements and the Sentinel 2 bands. Cubic spline interpolation represents the values between partition points—knots—by a polynomial of third degree with first and second derivatives of the interpolation function continuous at all points of the interpolation range (Dierckx, 1993). The Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) method uses Hermite interpolation conditions that define function values and derivatives at each nodal point (Fritsch and Carlson, 1980). The linear interpolation does not provide continuity in the derivatives at the interpolation knots. Thus, this selection criteria represents the three potential levels of continuity at the knots and all the potential combinations are presented in Fig. 4.

The error remains at the 0.1%-level for all the studied bands with the exception of B1 and B6 for which the error raises up to 0.5% level and B5 which increases up to 0.2%. Bands B5 and B6 have narrower bandwidths and are placed in a spectral region largely defined by atmospheric absorption of water (720.5 nm) and oxygen (687.5 nm and 761 nm) as seen in Fig. 2. For B6 it is shown how the use of an interpolation with continuity at the knots, provides an improvement up to 0.2% due to the better fitting of these atmospheric absorptions. B1 is not affected by large atmospheric absorption peaks but by a large solar irradiance variation. The spectral region in 410–440 nm is one of the most challenging regions in the solar irradiance models as described in Thuillier et al. (2003) and it contains a large spectral irradiance variability.

The bands show a very noticeable oscillation that remains constant for the band B5 onwards and for the different interpolation methods used here. This oscillation has an impact below 0.1% peak-to-peak as consequence of the movement of the bands position (wavelength shift), in combination with rapid variations of TOA radiance. In addition, the sampling/resolution are not fixed values in the range of shift but have a small variation—see Fig. 2—and produce this interference pattern in the image. As part of the validation process, the periodicity was found to be proportional to the slope of the sampling increase across the S2 bands. Furthermore, if the simulation keeps the same sampling/resolution across the band convolution, the systematic error remains at the same amplitude but the period is equal to the applied sampling/resolution. For the lower bands, B1 and B2, a clear periodicity of 4–5 nm can be found. At these bands, several native bands are binned and the period is dominated by the sampling period after binning. The discontinuity due to a linear interpolation in the binning requirement—see Fig. 2—introduces discontinuities for the lower bands B1 and B2.

Differences between all possible interpolations are small with maximum variations of around 0.2% peak-to-peak for specific bands and spectral regions such as for B1 and B6. The larger difference occurs between the TOA radiance interpolation using cubic spline and linear one whereas the PCHIP remains a middle case. For all bands but B4, the cubic spline interpolation provides the minimum error since it is able to adapt to the rapid variations of the TOA radiance. At 650 nm, there is a discontinuity of binning.
requirements from 2 native spectral bands to just 1. The differences between interpolations are mainly generated by the type of interpolation of TOA spectral radiance whereas the type of interpolation of the S2 SRF has a negligible impact. The assumption is that the S2 SRF sampled at 1 nm captures sufficient information regarding the spectral variations across the spectral band. The majority of the variations in the S2 spectral band—see Fig. 4—have a period of variation greater than twice the SRF sampling. The impact of the S2 SRF interpolation represented a small variation below 0.05%. In order to visualise the impact, the previous error of B1 in Fig. 4 has been normalised by the mean at each wavelength shift position and plotted in Fig. 5. 

Fig. 4 has shown that the error depends on where the focal plane array is situated. These errors are inter-dependent in between bands of the VNIR focal plane and in between bands of the SWIR focal plane. That is, the “shift” must be applied to all...
the bands of the focal plane since it is the whole array position that matters. The approach here is calculating the root mean square (RMS) error between the different bands error and for all positions of the array as follows:

\[ \text{rms}_{\text{VNIR}}[s] = \sqrt{\sum_{i=[B1-B8A]} \left( \frac{L_{\text{S2TRUTHS}}[i,s] - L_{\text{S2MODTRAN}}[i,s]}{L_{\text{S2MODTRAN}}[i,s]} \right)^2} \]  

\[ \text{rms}_{\text{SWIR}}[s] = \sqrt{\sum_{i=[B11-B12]} \left( \frac{L_{\text{S2TRUTHS}}[i,s] - L_{\text{S2MODTRAN}}[i,s]}{L_{\text{S2MODTRAN}}[i,s]} \right)^2} \] 

where \( s \) is the position in the array and \( i \) are the S2 bands. The term \( L_{\text{S2TRUTHS}} \) refers to the TOA spectral radiance as measured by TRUTHS and convolved with the S2 band whereas \( L_{\text{S2MODTRAN}} \) refers to the TOA spectral radiance using the MODTRAN reference and convolved to the S2 band.

Fig. 4. TOA error in estimating the Sentinel-2 MSI equivalent radiance for VNIR bands (above, B1-B8A) and SWIR bands (below B11 & 12) due to the TRUTHS sampling bands and preliminary resolution of the detector bands. The errors are plotted for different types of interpolation to reconstruct the TOA radiance and Sentinel 2 bands.
The result of the rms calculation produces a set of values for the different potential positions of the array. The results in Fig. 6 shows the spectral sampling error for each of the bands in the case of best and worst case rms position.

The results in Fig. 6 describe an error in the 0.1% range for all types of interpolations and focal plane alignment with the exception of B1, B5 and B6 bands. Potential methods to reduce the error introduced in these bands are separately studied in Sections 2.1.4 and 2.1.6 for B1, B5 and B6 respectively.

The criteria used here serves as an example of the design process and can be adapted to other scenarios. For example, the design of TRUTHS focal plane proposes a certain level of overlapping between the VNIR and SWIR focal plane. In that context, the method described in Section 2.1.1 might need to be adjusted and it is likely that the error for bands like B8 and B8A will reduce further.

2.1.3. Spectral domain: spectral knowledge uncertainty

The effect relating to “knowledge” (centre wavelength/bandwidth) of the TRUTHS SRF has been studied by reconstructing the TOA radiances spectrum $n$ times with different centre wavelength and/or bandwidth each time, before convolving it with the S2 bands. In a simplified model, intended to model likely instrumental errors, the central wavelength and bandwidth of the TRUTHS triangular bands (see Fig. 2) are modelled as a normal distribution, with the wavelength shift constant in sign and magnitude for all wavelengths across the spectrum. This is considered as an approximation that works under the assumption that the knowledge of the smile correction and/or spectral calibration is largely correlated across each of the S2 SRF bandpass. This simulation results in a dispersion of TOA spectral radiance values for the S2 band convolution, dependant on the structure in the local TOA spectral radiance spectrum.

Fig. 7 presents the resulting distributions for all the S2 bands with an associated central wavelength and bandwidth knowledge uncertainty of 0.2 nm ($k = 1$) convolved with the TOA spectral radiance 10,000 times. Since the normal distribution is by definition infinite, it was decided to truncate to a maximum of 10 times the standard deviation in order to avoid out-of-range values. The values applied here of central wavelength and bandwidth knowledge...
can be considered as a conservative figure since previous in-flight spectral calibration exercises have proven to keep the spectral knowledge at lower levels, as described in Delwart et al. (2007) for the Medium Resolution Imaging Spectrometer (MERIS) on-board the EnviSat mission. The starting wavelength is set to 410 nm; which corresponds to a zero wavelength position shift in

Fig. 4. The selected interpolation is cubic spline and linear for the TOA spectral radiance reconstruction and linear interpolation for the S2 SRF bands. This is based on the results in Fig. 4 which show the maximum disagreement for the TOA radiance interpolation methods and the limited impact of the S2 SRF interpolation. Note that only the spread of values is of interest here since the spectral

Fig. 7. Distribution of spectral sampling errors for S2 bands with an associated TRUTHS central wavelength and bandwidth knowledge uncertainty of 0.2 nm ($k = 1$).
sampling error was already reported in Fig. 4. Thus, the error distribution has been normalised to the original central wavelength and bandwidth values.

The previous results in Fig. 4 do not directly apply in this case since on the one hand, the resolution of the bands has been modified and, on the other hand, the sensitivity here is based on the wavelength and does not vary the spectral sampling requirement—see Fig. 2—at every shift of the bands central wavelength. Binning step was set to 0.05% for B1, B6, B7, and B8 and 0.01% was the binning for the rest of the bands.

All the results in Table 1 shows a standard deviation below 0.1% with the exception of B1 and B6 which are around 0.2%. That is because the error variations are the most important for this band as reported in Fig. 4. The choice of TOA spectral radiance interpolation method seems to have a negligible effect. The majority of the bands present a symmetrical distribution and thus the reported standard deviation can be reported as uncertainty at $k=1$. However, deriving an uncertainty out of skewed distributions such as B1 or B2 is not directly feasible since this parameter cannot be directly linked to a particular probability distribution (BIPM et al., 2008a).

Recent work in Gorroño et al. (2016a) has also shown the impact of the Sentinel-2 SRF uncertainty using the same TOA radiance input as described in Section 2.1.1. The results show that assuming a spectral response uncertainty of 0.2 nm ($k=1$) for systematic and 0.1 nm ($k=1$) for random spectral calibration knowledge, the dispersion of the data was below 0.3%. Indeed, this maximum dispersion corresponded to B7 for which a clear bimodal distribution was found. This was a consequence of there being different sets of filters used by the Sentinel-2 MSI detectors. The use of the detector footprint mask embedded in the Sentinel-2 LI C products and assigning a specific detector SRF to the results could significantly reduce these numbers and/or alternatively the use of a SRF mean of the S2 as used in Section 2.1.1 would reduce the impact after the data equalisation. In that case, only the spectral residual from the diffuser equalisation would be accounted for.

Table 1

<table>
<thead>
<tr>
<th>Statistics of TRUTHS spectral knowledge</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
<th>B8</th>
<th>B8A</th>
<th>B11</th>
<th>B12</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOA Cubic spline std.</td>
<td>0.21</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.23</td>
<td>0.09</td>
<td>0.10</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>TOA linear interp. std.</td>
<td>0.20</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.23</td>
<td>0.08</td>
<td>0.10</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Fig. 8. (Left) TOA error in estimating the Sentinel-2 MSI equivalent radiance for B1 using TRUTHS as a reference sensor and (centre) the measured TOA radiance as generated by MODTRAN5, resulting measurements of TOA radiance as measured by TRUTHS Earth imager and the Sentinel 2 B1 band and (right) the distribution of spectral sampling errors for S2 B1 band with an associated TRUTHS central wavelength and bandwidth knowledge uncertainty of 0.2 nm ($k=1$) with no spectral binning applied. The legend of the plots are equivalent to Figs. 4, 3, and 7 respectively.

2.1.4. Spectral domain: the impact of spectral binning

As reported in Section 2.1.2, the spectral binning effect can be easily appreciated for B1 and B2. For the cross-calibration events, the binning levels reported in Fig. 2 are not required since the specific application requires the best estimate over a ROI. That is, the accumulation of pixels in the across and along track direction reduces the spatial and temporal uncorrelated component of the pixel noise respectively. Therefore, the SNR requirement for this specific application is comparably lower. Systems like BingO (Binning pattern Generator and Optimiser) described by Dell’Endice et al. (2009) can be used to re-programme the spectral binning pattern of an FPGA card to fit specific application requirements. An example of such type of reprogramming was performed to estimate the spectral instrument response in-flight of MERIS onboard the EnviSat mission (Delwart et al., 2007).

For a mission like TRUTHS, the binning can be adapted to further sample B1 in a specific cross-calibration event. Once the event is finished, the default setup can be restored so that the impact over other applications such as radiation budget and/or memory requirements is minimum. The previous simulations in Fig. 4 have been run without introducing spectral binning and the results for B1 together with the convolved values at the TOA spectral radiance are shown in Fig. 8. In addition, the results for spectral knowledge in Fig. 7 are presented again for B1 with no spectral binning.

The results in Fig. 8 shows that the sampling/resolution largely decreases whereas the spectral knowledge uncertainty slightly
improves with a standard deviation of 0.18%. Increasing the sampling introduces further averaging over the band convolution.

When spectral resolution is reduced, the bands become more sensitive to solar and atmosphere absorptions from strong absorptions such as water-vapour (Green, 1998). However, what the simulation suggests is that this effect, although present at a singular native spectral band, it is largely averaged out when many bands across the S2 bandpass are considered.

2.1.5. Spectral domain: the TOA radiance sampling

The use of a TOA radiance with a sampling of 0.5 nm is justified in Wu et al. (2015) on the basis that this sampling is twice or larger than the main variations in the TOA spectral radiance (i.e. is valid based on Shannon sampling theorem). Here the previous TOA spectral radiance at 0.1 cm⁻¹ has been replaced by the same MODTRAN simulation at 0.5 nm for an empirical verification of the approach and the results of Fig. 4.

By comparing the results between Figs. 4 and 9, the results largely agree in absolute values with a negligible difference. Thus, the use of a 0.5 nm TOA radiance simulation in previous studies as in Wu et al. (2015) is fully justified. The difference arises in the relative shape of the curves due to second order variations that cannot be captured by the 0.5 nm resolution of MODTRAN. For B6, there is a slight decrease on the ringing amplitude whereas for B1 the shape has resulted in a smoothed version of that provided in Fig. 4. Thus, the use of a 0.1 cm⁻¹ MODTRAN spectral resolution is justified in this context since it provides a much more detailed analysis of sampling/resolution error vs. the detector array position and can better help to describe the spectral knowledge as a distribution of errors as already pointed in Section 2.1.1.

2.1.6. Spectral domain: the impact of the site

The effect described in Section 2.1.2 can vary when the cross-calibration is performed under alternative test-site targets with different spectral properties and atmosphere. The snow simulations in order to be more realistic were done with a sub-arctic summer atmospheric model at 60 degrees latitude and with an associated high SZA of 65 degrees. Aerosol optical thickness (AOT) and water vapour were obtained from AERONET in Greenland. All other sites are done for mid latitude summer, June 22, same solar angles as the simulation in Section 2.1.1 but with AOT and water vapour data taken from La Crau for grassland, Ascension Island for oceanic site.

The errors shown in Fig. 4 have been recalculated for these TOA spectral radiances and for a cubic spline and linear interpolation to reconstruct the TOA radiance and linear interpolation for the Sentinel 2 bands. The results are shown in Fig. 10.

The result shows that the impact of the different modelled TOA spectral radiances has an impact below 0.1% peak-to-peak for the majority of bands. For band B4, the sensitivity to the grass simulation raises the error slightly above 0.1%. For the bands B5 and B6, the error is significantly reduced when using other than a desert simulation. In the B5 case, the error is reduced to a level the below 0.1% at any interpolation at any simulation other than the desert one. In the B6 case, the simulations show a considerable improvement of the error when compared to the desert case. In general, the errors can be bracketed in the range 0.2–0.5% depending on the array position, site and interpolation type.

2.2. Spatial domain

2.2.1. Spatial domain: methodology

In the spatial domain, site non-uniformity, in combination with uncertainty due to misregistration of the instrument scenes, in principle leads to a systematic uncertainty that needs to be accounted for in the total uncertainty budget. The effects of even small differences across the site can lead to a bias when it is used for radiometric calibration; therefore this needs to be carefully assessed and addressed. The results in Chander et al. (2013a) showed that for a 2-pixel spatial knowledge, the potential uncertainty introduced by misregistration was at 0.1% for the VNIR bands and 0.2–0.3% for the SWIR bands of Landsat 7 Enhanced Thematic Mapper Plus (L7 ETM+).

For calibration, multiple sites and multiple observations of the same site will be used and the location knowledge will likely be known to ≲1 km, so reducing significantly the uncertainty in cross calibration due to spatial co-location error. In the multi-temporal case, the TRUTHS and CLARREO orbits are defined with a 90° polar orbit and 61-day ground track repeat cycle at 609 km altitude (Roithmayr et al., 2014). In addition to this, the orbit is asynchronous with a different time overpass over the Equator during the year. Thus, when applying this method to several matchups, the spatial offsets will be largely independent and the effects can be reduced. This same assumption is discussed in Wielicki et al. (2008) where a spatial matching noise below 1% is considered as a threshold to minimise the impact over the temporal aggregation of diverse satellite-to-satellite matches.

This study thus concentrates on the coarser effect of the impact of the spatial non-uniformity in a specific area and imperfect geographic location knowledge and/or lack of co-alignment between the reference sensor and the sensor under calibration. This study considers two arbitrarily chosen ROIs spatially separated within the Libya-4 site and LaCrau site and the resultant systematic uncertainty due to this separation i.e. the variance in TOA reflectance caused by surface non-uniformity. The two sites have been chosen.
as examples of typical calibration sites that represent different levels of uniformity.

In order to model the effect of spatial non-uniformity, a practical approach is presented using real EO data with low relative uncertainty and a sufficiently large swath that covers the area under study. Landsat-8 OLI TOA reflectance factor images have been selected with a large swath of 185 km that allows the selection of multiple ROIs across it (Irons et al., 2012). Landsat-8 OLI pixel-to-pixel non uniformity residual lies between 0.2% and 0.3% for the complete focal plane (Morfitt et al., 2015). This relative uncertainty provides a sufficiently small effect, compared to the expected magnitude of ROI site variability, to allow us to use scenes of Landsat-8 OLI for this evaluation. For a high-radiance scene such as Libya-4 or La Crau, the uncertainty budget is dominated by

![TOA error in estimating the Sentinel-2 MSI equivalent radiance for VNIR bands (above) and SWIR bands (below) due to the TRUTHS sampling bands and preliminary resolution of the detector bands for different modelled sites.](image-url)

**Fig. 10.** TOA error in estimating the Sentinel-2 MSI equivalent radiance for VNIR bands (above) and SWIR bands (below) due to the TRUTHS sampling bands and preliminary resolution of the detector bands for different modelled sites.
highly correlated effects in the spatial and temporal domain, so minimising the weight of uncorrelated spatiotemporal contributions e.g. instrument noise. The effects of such random variations are further reduced by the fact that each ROI covers several pixels, so providing some averaging of signal noise etc.

In addition to Landsat-8 OLI data, the study will also repeat the study using S2 MSI data. This mission also provides a large swath of 295 km. Its pixel-to-pixel non uniformity has been validated by the means of diffuser and natural targets on Earth leading to values well below the specification of 0.2% (Gascon et al., 2016).

The selected bands for the study are B1, B5 and B8 for L8 OLI—443, 865, and 2201 nm central wavelength—and B1, B8 A and B12 for S2 MSI—443, 865, and 2190 nm central wavelength. These bands illustrate boundaries between atmosphere and surface scene composition. At 443 nm the impact from atmospheric effects is significant whereas at 865 nm and ~2200 nm the expected atmosphere transmission is above 80% for both the VNIR and SWIR respectively. In addition, by comparing very similar bands of two different missions, the method can be validated.

The products selected for Libya-4 and La Crau sites were selected based on best temporal coincidence using the CEOS COVE tool (Kessler et al., 2013). The matches found are considered optimum since the overpasses over the same site are delayed by less than 15 min and the cloud conditions are near zero percent for the whole product tile. The selected products are described in Table 2.

In addition to the low cloud cover, each ROI for S2 MSI and L8 OLI has been screened for the possibility of degraded, saturated or invalid pixels. That is possible by reading the quality assessment band in the L1TP L8 OLI products and the different masks integrated as part of the S2 MSI L1C product. The result is that all the pixels in the ROIs processed were valid without exception.

The L1C data from the S2 MSI is directly provided as TOA reflectance factor values scaled by a ‘quantification value’ and have been processed using free software produced by ESA and named Sentinel Application Platform (SNAP). The L1 Digital Number (DN) in the Landsat-8 OLI L1TP product are converted to TOA reflectance, $r_{ij}$, using (1) according to (USGS, 2015):

$$r_{ij} = \frac{M_i \cdot DN(i,j) + A_i}{\cos(SZA(i,j))}$$

where $M_i$ refers to the reflectance multiplicative scaling factor for the band and $A_i$ refers to the reflectance additive scaling factor for the band. Both values can be extracted from the product metadata.

Next, the SZA for the coordinates of each pixel at row $i$ and column $j$ is calculated using the image timestamp and lat/lon position using the PySolar library (Stafford, 2015). The lat/lon at each position is obtained by a coordinate transformation of the Worldwide Reference System (WRS) path/row coordinate system using PROJ.4 libraries.

Once the ROI reflectance factors are obtained, the mean of the pixels is calculated. In an iterative mode, the ROI centre is displaced following a pre-defined grid across a defined area. This process is illustrated in Fig. 11.

The results of this process is a TOA reflectance factor error map for the potential displacements over the calibration site.

The next step is the processing of the error map in order to derive the uncertainty associated with the geolocation knowledge. This process is performed by calculating the standard deviation over a growing area from the centre of the error images. Thus, for each associated geolocation knowledge, an uncertainty $k = 1$ will be associated and a curve of uncertainty vs. geolocation knowledge will be obtained. This curve can be used as a tool for either the refinement of mission design requirements, TRUTHS or CLARREO, and/or the definition of the matching requirements in order to achieve a specified spatial systematic uncertainty.

### 2.2.2. Spatial domain: results for La Crau calibration site

The site of La Crau has been traditionally used for the calibration of sensors as SPOT (Santer et al., 1992). The calibration site is defined as 400 × 400 m² area centred at 43.556°N 4.858°E, in a 60 km² flat area composed of pebbles and sparse low vegetation.

The region studied for this example corresponds to ±0.002° off-centre in latitude and longitude. This corresponds to approximately 0.32 × 0.44 km² rectangular spacing from the centre of the site. The TOA reflectance factor at each 400 × 400 m² pixel ROI and band considered are shown in Fig. 12.

TOA reflectance factors for both missions show similar results. The S2 B1 shows a less defined pattern due to its coarser spatial resolution of 60 m. The ROI has been selected as 400 × 400 m² which can be achieved for S2 B8A and S2 B12. However for the S2 B1 and L8 bands, this distance cannot be delimited within an exact number of pixels and the final spatial dimensions are 360 × 360 m² and 390 × 390 m² respectively. This point together with the geolocation uncertainty for each sensor and band results in a displacement between bands that can be visually estimated at around 1 pixel.

Fig. 13 shows the spatial error as a consequence of the ROI displacement ±0.002° off-centre in latitude and longitude.

The spatial error between both missions largely agree in terms of gradient and image shape. These variations, however, seem slightly shifted between the missions with the spatial variations slightly shifted towards positive values for the S2 MSI with respect to the L8 OLI. This consistency between the two missions for all three bands suggest that the small delay between overpasses and, especially, the different angular configuration is likely to be the cause of these variations.

Fig. 14 presents the standard deviation of the spatial error pixels as growing from the centre of Fig. 13.

The results presented here show a strong consistency between the two missions which provides confidence to the results. There is a clear linear dependency of the spatial uncertainty vs. the spatial knowledge of the ROI centre. The linear dependency shows a small slope variation between S2B8A and L8 B5 and S2 B12 and L8 B7. The results for S2 B1 follow a linear dependency in a more irregular way that can be attributed to the coarser spatial resolution. These results can be easily applied to determine the spatial knowledge requirements for TRUTHS/CLARREO missions. If we were to account for a more realistic scenario with e.g. Spatial uncertainty 10 times lower than the maximum displacement shown here,—

### Table 2

<table>
<thead>
<tr>
<th>Site</th>
<th>Sensor</th>
<th>Product ID</th>
<th>Timestamp centre</th>
<th>Cloud [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Libya-4</td>
<td>L8 OLI</td>
<td>LC08_L1TP_181040_20170513_20170513_01_RT</td>
<td>2017-05-13T08:54:34Z</td>
<td>0</td>
</tr>
<tr>
<td>Libya-4</td>
<td>S2 MSI</td>
<td>S2A_MSIL1C_20170513T090021_N0205_R007_T348GS_20170513T090803.SAFE</td>
<td>2017-05-13T09:08:03Z</td>
<td>0</td>
</tr>
<tr>
<td>La Crau</td>
<td>L8 OLI</td>
<td>LC08_L1TP_190603_20170420_20170501_01_T1</td>
<td>2017-04-20T10:23:17Z</td>
<td>0.08</td>
</tr>
<tr>
<td>La Crau</td>
<td>S2 MSI</td>
<td>S2A_MSIL1C_20170420T103021_N0204_R108_T311FJ_20170420T134554.SAFE</td>
<td>2017-04-20T10:34:54Z</td>
<td>0.5321</td>
</tr>
</tbody>
</table>
Fig. 11. Methodology process for the assessment of spatial variations.

Fig. 12. TOA reflectance factor at the 400 × 400 m² at the LaCrau site for the considered L8 OLI and S2 MSI bands.
i.e. approximately $32 \times 44 \text{ m}^2$—The uncertainty would be 10 times that of the maximum displayed. This means approximately 0.12% for B1, 0.27% for B5 and 0.5% for B12. This value would represent the systematic uncertainty introduced over one overpass and would be largely uncorrelated in between matches and will tend to reduce with increasing number of overpasses and match-ups.

2.2.3. Spatial domain: results for PICS sites

The ROI reference centre position is selected to be at the centre of the Libya-4 site as defined in Lacherade et al. (2013)—i.e. 28.55° N, 23.39° E—and with a size of 20 km × 20 km. The region studied corresponds to ±0.05° off-centre in latitude and longitude. This corresponds to approximately $10 \times 10 \text{ km}^2$ rectangular spacing from the Libya-4 centre. The centre of the ROI has been moved in 15 equidistant points in each of the directions. The reason for using an ROI size of 20 km × 20 km is that based on the results of Govaerts (2015), this is a sufficiently large area at which the dune effect can be reasonably integrated. A smaller area would introduce a larger dispersion due to the dune effect whereas a larger area would require a bigger swath and eventually would introduce low-frequency spatial variations. The TOA reflectance factor at each $20 \times 20 \text{ km}^2$ pixel ROI and example band are shown in Fig. 15.

The values of TOA reflectance factor are very similar for both missions and no visual shift in geolocation can be identified. The size of the ROI used is sufficiently large so that the impact of a small miss-registration between missions and/or number of pixels in the ROI does not have any impact.

Fig. 16 shows the spatial error as a consequence of the ROI displacement ±0.05° off-centre in latitude and longitude. The image error variations and levels are very similar for S2 B1 and L8 B1 and S2 B8A and L8 B5. Indeed the results for S2 bands are slightly larger than those of L8 bands in similar manner to the results for La Crau in Fig. 13. Both images for S2 B1 and L8 B1 present an irregular pattern typical of real scene variations. However this irregular pattern cannot be found in S2 B8A and L8 B5. The pattern of errors suggests in both a dependency of TOA reflectance factors with directionality of the sun illumination. The results suggest that at these bands the errors that have been measured could be the consequence of the viewing angular variations of the

![Fig. 13. TOA reflectance factor error map for the LaCrau site and the considered L8 OLI and S2 MSI bands.](image)

![Fig. 14. Spatial uncertainty vs. spatial offset for the LaCrau site and the considered L8 OLI and S2 MSI bands.](image)
push-broom sensor. The difference of results for L8 B7 and S2 B12 cannot be explained by a difference on angular configuration or scene change between overpasses.

Fig. 17 presents the standard deviation of the spatial error pixels as growing from the centre of Fig. 16.

The results show again an almost linear dependency of the spatial TOA uncertainty with the spatial positioning knowledge.

Following the same logic as in Section 2.2.2, the impact over a realistic geolocation knowledge of a TRUTHS/CLARREO like mission would be well below 0.1%.

The results agree very well for both missions except for the S2 B12 and L8 B7. In this case the S2 B12 and L8 B7 results are largely variable and confirm the disagreement between the missions seen in Fig. 16.
Fig. 18 presents the error distribution for L8 B7 and S2 B12 in Fig. 16.

On the one hand, the results for L8 B7 present a map of error with an irregular map with a normal or similar distribution pattern. On the other hand, the results for S2 B12 show a North-South variation following the orbit overpass with a highly skewed distribution. Thus, these variations cannot be caused by the sensor angular variations or natural variability. Indeed it suggests an ACT sensor effect.

Several effects can be suggested for these variations although none of them can be fully verified here. A crosstalk effect in the SWIR bands has been detected and accordingly corrected for in the S2 L1C products. In addition, the SWIR bands undergo regular decontamination due to deposition of moisture on the top of the detectors (Gascon et al., 2016).

The results obtained for Libya-4 are lower than those obtained in Chander et al. (2013a). One reason for this disagreement could be the potential pixel non-uniformity of the L7 ETM. This should have limited impact due to the whiskbroom design of the ETM instrument. However, the main difference between the two studies arises from the selection of the ROI over Libya-4. The selection of a small ROI of 3 \times 3 km\textsuperscript{2} as used by Chander et al. (2013a) introduces a larger dispersion due to the dune effect (Govaerts, 2015).

2.3. Temporal domain

2.3.1. Temporal domain: methodology

This section describes the impact on the TOA reflectance due to changes in the sun angle (azimuth and zenith) for a specified time-span after the overpass of the reference satellite (TRUTHS/CLARREO) and the resulting residual uncertainty of the correction between the reference and target satellites.

Recent work in McCorkel et al. (2013) studied the effect of temporal mismatch between MODIS vs. Hyperion matches. The latter instrument was measuring in an orbit 40 min preceding the MODIS one until mid-2005. The orbit of Hyperion was changed from mid-2005 resulting in a rare cross-calibration between the two missions. This unusual situation triggered the possibility to compare the impact of the temporal overpass differences between coincident overpasses—within 30–40 min—and non-coincident overpass—within 30 days separation—over the Railroad Valley calibration site. The results showed that although the dispersion of the data significantly increased, the bias between the two cases was between 1–2%. To a large extent, BRDF and temporal mismatches were largely averaged out even for such a large timespan difference.

The orbit choice of CLARREO is set as a polar 90 degrees asynchronous orbit (Roithmayr et al., 2014). This type of orbit permits the sparse sampling of brightness temperature over the diurnal cycle and subsequently improves the sampling error (Kirk-Davidoff et al., 2005). The climate benchmark of missions like TRUTHS or CLARREO is largely improved by this type of orbit since it assures full diurnal cycle sampling for spectral fingerprints as well as full reference inter-calibration sampling over all climate regimes and all satellite orbit thermal conditions (Wielicki et al., 2013).

In the work presented in Roithmayr et al. (2014), the selection of cross-calibration matchups was set to a global scale within a
and Wielicki et al. (2013), it is conceivable to assume that the temporal noise was found to be at the 1% level and with sufficient samples the noise reduces to <0.3% (Wielicki et al., 2008). Extending the concept of sparse sampling studied in Kirk-Davidoff et al. (2005) and Wielicki et al. (2013), it is conceivable to assume that the temporal systematic errors are largely uncorrelated and converge to a low bias not only when a global scale is taken into account but also when a more restricted area is considered.

The approach developed here seeks to consider a complementary scenario where the surface and atmospheric conditions are considered stable in time but the inter-calibration matchups are limited to the specific locations that fulfill these conditions. Thus, the sites considered here are PICS and more specifically, the Libya-4 site due to the large amount of prior work available and its representativeness of the PICS sites. Due to the better temporal stability of these types of sites, the considered delay between overpasses can be increased and consequently the opportunities increased. Here a delay of 30 min will be considered as representative of the upper limit considered for the SNO cross-calibration over PICS. Furthermore, the accurate observation of missions like TRUTHS and CLARREO in conjunction with a pointing capability offer the possibility of an improved modelling of the surface BRDF models of PICS sites. The achievable uncertainty of a temporal correction using this approach will also be considered here.

2.3.2. Temporal domain: atmospheric variation and radiative transfer code impact

In this section, the TOA reflectance factor variation over a 30 min timespan will be evaluated as a consequence of the atmospheric solar reflected radiance due to solar angle variations. In addition, the study will also consider any discrepancy between different atmospheric correction algorithms.

Solar angles at 30 s intervals are calculated and used as inputs to the MODTRANv5 model for a parameterisation as described in Section 2.2; the starting time of the simulation is taken as a typical Landsat-8 OLI product reference timestamp for a Libya-4 overpass, specifically 8:56:32 local time for days of the year 173 and 355. The study has been undertaken at three wavelengths—443, 865, and 2201 nm—that represent the central wavelengths of the Landsat-8 OLI bands with a high atmospheric sensitivity (B1) and with lower atmospheric sensitivity in the VNIR (B5) and SWIR (B7). The radiance is further normalised by the cosine as shown in (3) so that a reflectance factor difference can also be calculated. Due to the large amount of time and MODTRAN runs—400 simulations covering the VNIR and SWIR spectrum at 1 nm spectral resolution—the simulations have been setup in parallel using the JASMIN facilities (Lawrence et al., 2013).

In addition to using MODTRAN, we have also carried out a similar analysis using the 6SV1 (second simulation of a satellite signal in the solar spectrum, vector, and version 1) radiative transfer code (Vermote et al., 1997) and interfaced using the Py6S library (Wilson, 2013). By comparing them, it is possible to assess the effect of any radiative transfer (RT) code biases. In particular, the work described in Kotchenova et al. (2008) or Kotchenova et al. (2006) pointed out important differences in the aerosol and molecular scattering between these two RT codes. The parameterisation follows that described in Section 2.2 for MODTRAN but using the sand model in 6SV1.

Fig. 19 shows the reflectance factor temporal differences using both RT codes—6SV1 and MODTRAN—at the studied wavelengths for the year day 173 and 355. These two days represent the most extreme SZA angle conditions (summer and winter solstice, respectively).

The results show two anomalies at 865 nm and 2201 nm. These are produced by the MODTRANv5 simulations and a zoom to the radiance trend showed that at that point, the trend was slightly changing. That is, the assumption is that there is a software interpolation discontinuity and/or an ill-conditioned solution. This is further amplified when small errors are calculated and the SZA round off at 2 decimal digits introduces a small noise.

The graph shows the importance of the atmosphere at 443 nm due to the strong impact of scattering at this wavelength. For winter periods, the atmospheric radiance variation becomes dominant whereas in the summer period it is of the same order as the cosine effect. Similarly, the result at 865 nm shows how these two variations are largely compensated. Indeed, here the changes are so small that the different atmosphere and surface balance between the two RT codes and setup is clearly shown. Where 6S shows a minimum error increase in winter, MODTRAN does it for summer. Finally, the TOA reflectance error at 2201 nm shows an important change as a consequence of the dominant cosine effect and an almost negligible atmospheric effect.

The discrepancy between the two radiative codes can be seen at 443 nm and 865 nm. No discrepancy can be seen at 2201 nm as a result of the low atmospheric impact. The work in Kotchenova et al. (2008) described differences between the radiative codes in the order of several percent. However, the results here demonstrate these differences are limited to approximately 0.2% at 443 nm and below 0.1% at 865 nm. This is an expected result since the constant biases are cancelled out and only the differences in temporal effect between the two radiative codes is relevant.

2.3.3. Temporal domain: atmospheric knowledge

In addition to the temporal discrepancies between radiative codes, it is important to understand the potential impact of the atmospheric knowledge as it contributes to the temporal effect. In order to study its impact, the TOA radiance calculation has been repeated nearly 1000 times with varying inputs of AOT and water vapour. The AOT and water vapour values are random samples.
from a normal distribution determined from a mean and standard deviation as specified in Mishra et al. (2014a) as representative of Libya 4. These are 0.0858 and 0.0486 for mean and standard deviation of AOT and 2.85 and 0.7 for the water vapour mean and standard deviation. Just a very small percentage of AOT samples out of the normal distribution were negative. These samples were set to 0 for the simulations.

The resulting residual in the sun angular correction can be understood as the repetition of the error trend shown in Fig. 19 for each simulation. This provided approximately 1000 potential error curves which will increasingly vary with length of time. The distribution of corresponding reflectance errors versus timespan is studied here. Fig. 20 shows the results for the reflectance error dispersions for a 30 min timespan and Table 3 contains the statistical parameters for the dispersions respectively.

The results indicate the low impact that the radiative transfer code inputs have on the residual uncertainty after the sun angular correction.

2.3.4. Temporal dimension: atmospheric variation

Thus far, the analysis has been conducted assuming that the atmospheric parameters have remained constant over the 30 min timespan being studied. Similar to the description of Section 2.2, the potential atmospheric variations in this timespan are difficult to predict, although likely to be small for the types of site chosen unless an unusual weather event occurs. However, the approach taken here is to predict the worst case uncertainty in the correction and limit the potential minimum and maximum uncertainty in a temporal correction.

Fig. 21 shows the results for the TOA radiance dispersion at a point in time as a consequence of AOT and water vapour variations—in this case 30 min—and Table 4 contains the statistical parameters for the dispersion respectively.

The results indicate the potential dispersion of the TOA reflectance factor—the time is constant and the dispersion is indifferent of radianc or reflectance factor—are expected to be below the 1% level for 443 nm and 865 nm. For the SWIR region at 2201 nm the relative uncertainty increases due to the weak signal measured. Table 4 indicates standard deviation values around 0.5% for the simulation at 443 nm. This is produced by the large impact of aerosol variations in this region. Nonetheless, the simulation here has used any potential variation in AOT and water vapour that could occur throughout the year. The variations in a 30 min timespan are expected to be much smaller and the impact will be several times lower. In fact this uncertainty could be used as a worst case to account for aerosol variation at any time of the year and thus an uncertainty for any time gap. In addition, for a PICS site characterisation using multiple TRUTHS overpasses, the variations of the atmosphere can be assumed random to a large extent. Since each acquisition represents a different solar angular geometry, there is

![Fig. 20. Dispersion of reflectance factor errors at 30 min at 443 nm and yearday 173 (a), 865 nm and yearday 173 (b), 2201 nm and yearday 173 (c), 443 nm and yearday 355 (d), 865 nm and yearday 355 (e), and 2201 nm and yearday 355 (f).](image)

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Statistical parameters for Fig. 20.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day of Year 173</td>
<td>Day of Year 355</td>
</tr>
<tr>
<td>443 nm</td>
<td>865 nm</td>
</tr>
<tr>
<td>Mean</td>
<td>–0.0303</td>
</tr>
<tr>
<td>Median</td>
<td>–0.0290</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.01</td>
</tr>
<tr>
<td>68.27% coverage interval</td>
<td>[−0.04, −0.02]</td>
</tr>
<tr>
<td>Skewness</td>
<td>−1.2034</td>
</tr>
<tr>
<td>Excess Kurtosis</td>
<td>2.1070</td>
</tr>
</tbody>
</table>
no direct method to assess the random improvement over time. However, in a fully uncorrelated atmospheric situation and assuming a maximum improvement, just 10 cloudless overpasses, would reduce the uncertainty levels by a factor of over 3. This would set the uncertainty levels below the 0.5% level or close to it, without placing any limitation on time gap between sensor and test overpasses.

2.3.5. Temporal domain: surface impact

The earlier sections have evaluated the effects of the radiative code, temporal atmospheric change and atmospheric knowledge in the temporal correction. There is, however, a final factor to consider related to the knowledge of the surface reflectance. In this case, the surface reflectance model from Bouvet (2014) has been taken as a reference. This surface reflectance model is the result of an inversion of MERIS observations over the 2006–2009 period that tunes the four parameters of a Rahman-Pinty-Verstraete (RPV) model. The aerosol optical depth has been extracted from the mean year value described in Mishra et al. (2014a). The selected aerosol profile is 'continental'. The discussion in Bouvet (2014) suggested this model not because of its better representation of the aerosol profile in Libya but because of its apparent better performance when combined with the RPV model. The characterisation of the surface angular variation uses this model (Rahman et al., 1993). The model provides the reflectance, $\rho$, defined by four parameters ($\rho_0$, $k$, $\Theta$ and $\rho_c$) for the viewing and illumination conditions (SZA $\equiv \delta_s$, VZA $\equiv \delta_v$ and RAA $\equiv \Delta \phi$) as follows:

$$\rho(\delta_s, \delta_v, \Delta \phi, \rho_0, k, \Theta, \rho_c) = \rho_0 M_1(\delta_s, \delta_v, k) F_{\text{inc}}(g, \Theta) H(\rho_c, G)$$  \hspace{1cm} (4)

Where each one of the terms is defined as:

$$M_1(\delta_s, \delta_v, k) = \frac{\cos^{k-1} \delta_s \cos^{k-1} \delta_v}{(\cos \delta_s + \cos \delta_v)^{1-k}}$$ \hspace{1cm} (5)

$$F_{\text{inc}}(g, \Theta) = \frac{1 - \Theta^2}{(1 + 2\Theta \cos g)^{3/2}}$$ \hspace{1cm} (6)

$$H(\rho_c, G) = 1 + \frac{1 - \rho_c}{T + G}$$ \hspace{1cm} (7)

$$\cos g = \cos(\delta_s) \cos(\delta_v) + \sin(\delta_s) \sin(\delta_v) \cos(\Delta \phi)$$ \hspace{1cm} (8)

$$G = (\tan^2(\delta_s) \tan^2(\delta_v) - 2 \tan(\delta_s) \tan(\delta_v) \cos(\Delta \phi))^{1/2}$$ \hspace{1cm} (9)

The terms described in (5)–(9) represent different features of the reflectance function (Rahman et al., 1993). The amplitude component is set by $\rho_0$ and then modified by the term $M_1$ which defines the overall shape of the angular field using the parameter $k$. $F_{\text{inc}}$ is a Henyey-Greenstein function that provides the balance...
between forward and backward scattering and is described through the parameter $\Theta$ and $g$ (described in Eq. (9)). $H$ describes the hotspot effect through the parameter $\rho_c$.

Values for $k$, $\Theta$, and $\rho_c$ for the Libya-4 site have been extracted from the results obtained in Bouvet (2014) for a surface bidirectional reflectance distribution function (BRDF) model. The values have been derived from data pertaining to the whole Libya-4 ROI site and therefore describe the BRDF of large scale structures at the site such as the dunes.

The work in Bouvet (2014) also discussed the limitations of a model that cannot be traceable in-flight and discussed the possibility of providing an absolute traceable standard by using observations of missions like TRUTHS or CLARREO. The overpasses of these missions can be used in the same manner as MERIS. In addition, these two missions incorporate in their design a gimbal mechanism that can further provide different angular observations over the same site. Thus, in this section the surface reflectance model from Bouvet (2014) is tested to understand the potential surface reflectance variation and uncertainty residual correction in an overpass of TRUTHS or CLARREO after 30 min of a Landsat 8 OLI overpass over Libya-4.

Here, the simulation has provided a similar approach to the one described in Fig. 20 and Table 3. The change in the surface reflectance over 30 min has been repeated 10,000 times. The RPV parameters from Bouvet (2014) have been described as normal distributions with a 5% standard deviation for which a sample out of a normal distribution is extracted for each parameter at each iteration. That is, the uncertainty of the four parameters in the RPV model have been assumed as uncorrelated. As mentioned earlier, significant upgrade in the performance of these models require accurate reference measurements of sites and surfaces. Missions like TRUTHS not only propose accurate measurement but also point to capabilities that can further tune the model as indicated in Bouvet (2014). Thus, the level of uncertainty in the surface reflectance model used for this simulation can be taken as a worst case with the expectation that the knowledge of the parameters would be much lower—i.e. the parameter $\rho_0$ that represents the albedo in (5). The resulting distributions for nadir viewing and summer and winter overpass are provided in Fig. 22 and their main statistics in Table 5. Only mean and standard deviation are reported due to the near-normal distribution shape (Gorroño et al., 2016b).

### Table 5

<table>
<thead>
<tr>
<th></th>
<th>Day of Year 173</th>
<th>Day of Year 355</th>
</tr>
</thead>
<tbody>
<tr>
<td>443 nm</td>
<td>865 nm</td>
<td>443 nm</td>
</tr>
<tr>
<td>Mean</td>
<td>2.3337</td>
<td>-0.4745</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>0.22</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Fig. 22. Results for surface reflectance error dispersion at 30 min at 443 nm and yearday 173 (a), 443 nm and yearday 355 (b), 865 nm and yearday 173 (c), and 865 nm and yearday 355 (d).
The results at 443 nm result in an uncertainty that oscillates between 0.2–0.3% at any time in the year whereas the results at 865 nm range between 0.3–0.4% at any time of the year. If we consider the combined effect of the atmospheric impact, then the radiative code differences must also be accounted for at the bands dominated by the atmospheric scattering (see Fig. 20). Considering an equal weight on the TOA reflectance at 443 nm, the same levels of uncertainty at 0.2–0.3% can be kept. At 865 nm the atmospheric contribution is much lower and the uncertainty in the surface reflectance variation can be considered as dominant.

The SZA variations in the summer case start at 21.2 degrees and end with 14.8 degrees. For the winter case, the SZA ranges from 56.1 down to 53.8 degrees. These results can be easily compared with the TOA reflectance factor dependency over Libya 4 performed by Mishra et al. (2014b). Applying these SZA variations to the empirical SZA curve using TOA reflectance factor images from MODIS band 2 (841.9 nm) reveals a predicted variation of 0.88% and 0.33% respectively for the summer and winter case.

These results are consistent with the ones obtained in Fig. 19 for the summer and winter cases at 865 nm respectively. The predicted atmospheric variation at 865 nm in Fig. 19 shows a small but present increase of the TOA reflectance factor at 0.1–0.2% either in the winter or summer case. If these values are considered together with the reported variations of the surface at 865 nm in Fig. 22, the agreement with the predicted results in Mishra et al. (2014b) gets very close in the summer case and just above 0.5% in the winter case.

The work in Mishra et al. (2014b) obtained different TOA reflectance linear fitting curves for the rest of MODIS bands and their slopes were fitted in an exponential model. At 443 nm the predicted slope is very close to zero. Assuming that the surface represents around half of the total signal at the TOA in the spectral region around 443 nm, the impact of the surface reflectance variation reported in Fig. 22 can be set at around −0.2% for winter and +1.16%. When combining with the results of atmospheric variation at 443 nm in Fig. 19, the global variations are close to zero and agrees with the reported slope close to zero by Mishra et al. (2014b).

To sum up, for a temporal delay of 30 min between a target sensor and a reference sensor the expected TOA reflectance factor variations will be at the 1%-level without further corrections. The asynchronous orbit of a mission like TRUTHS and CLARREO will be translated in an inconsistent delay pattern. That is, if these errors are largely uncorrelated, for just nine matchups over PICS, the expected temporal uncertainty will be reduced below 0.33%. Furthermore, if the temporal correction over PICS is carried out the MCM approach and radiative transfer code comparison have shown that it is possible to correct the temporal bias with an uncertainty residual in the 0.2–0.3% range.

3. Discussion

This paper studies the uncertainty contributions in the spectral, spatial and temporal domain of sensors of the form used in Sentinel-2 MSI or L8 OLI when cross-calibrated using TRUTHS as a reference over difference cross-calibration sites. The results obtained support previous work in Chander et al. (2013a) and conclude that the uncertainty from the main sources related to the radiometric properties of the site and the spectral matching of the sensors generally falls well below the 0.5% (k = 1) level. At this level, the uncertainty in the reference sensor becomes the main contribution in the cross-calibration uncertainty budget. In metrological terms, it means that any effort directed towards an improvement of the calibration transfer methodology will have a limited impact in terms of the overall uncertainty budget. However, the use of PICS (and other sites with similar properties e.g. snowfields) for in-flight inter-calibration using a high accuracy sensor like TRUTHS/CLARREO presents a major opportunity to provide in-flight calibration upgrade to EO optical missions, leading to a significantly reduced uncertainty budget.

The use of specific examples, with real data has facilitated the use and description of a more rigorous methodology. In particular, this paper has clarified the use of error and uncertainty and has discussed the implications for uncertainty improvement through use of multiple acquisitions. This type of approach is specifically relevant for an instrument like TRUTHS in asynchronous orbit with potentially different time and spatial matchings over time. An asynchronous orbit like the one proposed in the TRUTHS or CLARREO missions means that the overpasses will be slightly different in time over the year (Roithmayr et al., 2014). Compared to sun-synchronous missions, this represents an advantage for inter-calibration since the delay between sensors is expected to be largely uncorrelated. However, this temporal and multi-site improvement has not been extensively studied at this point. Where necessary, boundary conditions showing a best and worst case have been provided. The next steps of this study will look more carefully at the temporal correlation effects and the uncertainty improvement over several overpasses and use of multiple sites.

The effort in this study has resulted in a set of tools and methodologies that are under continuous evolution and will ultimately be useful for the operational exploration of missions like TRUTHS or CLARREO. However, at this point in time, the results are already beneficial as a feedback to the TRUTHS mission team and helpful for refinements to the mission design.

3.1. Discussion: spectral domain

Section 2.1 studies the spectral response effect. The method is similar to the one applied in Wu et al. (2015) for the CLARREO mission. However, in this case, the data used is based on a preliminary design of the TRUTHS sensor and includes the further effect of spectral binning. For spectrally flat sites such as Libya-4 the error due to the spectral response effect for a cross-calibration of Sentinel-2 with a TRUTHS-like reference sensor is small, with values below 0.1% for most bands. For specific bands in regions with significant spectral features within the band—e.g. Sentinel-2 B1 and/or a smaller bandwidth e.g. Sentinel-2 B5 and B6—the error has been found to rise to around the 0.5% level. The results of CLARREO use a larger spectral resolution of about 8 nm for 4 nm spectral sampling and Gaussian shape. Although different in design and spectral simulation of the native spectrometer, both examples show the low error introduced by most of the reference sensor bands. For the cases were the error increases significantly above 0.1%, the comparison cannot be made due to the absence of similar bands between S2 bands studied here and the MODIS bands studied in Wu et al. (2015).

The starting wavelength of the calculation represents the alignment of the detector to these optical requirements. It has been found to be significant for certain bands especially for those like S2 B1 that utilise spectral binning. This latter effect has been also found to introduce discontinuities in the results due to the linear interpolation of the binning requirements (see Fig. 2). Due to this dependency on the alignment, a study of the rms error to minimise the impact for all the bands in the focal plane has been carried out. Since the TRUTHS design introduces separate focal planes for the ultraviolet (UV), VNIR and, SWIR regions, the rms has been calculated separately for each focal plane.

The TOA reconstruction method and S2 band SRF interpolation were evaluated with the linear, PCHIP and cubic spline interpolation. The interpolation methods showed insignificant differences for the S2 SRF interpolation and found a level of improvement up
to 0.2% for the most affected bands due its greater flexibility in capturing the relevant spectral variations. Further iterations of the tool will include more detailed spectral information at a pixel level, if available, and further discuss the methods to reconstruct the TOA spectral radiance. Furthermore, a separate study of spectral effect of the SWIR bands should be undertaken and effects like filter contamination discussed.

In addition, a simplified uncertainty propagation of the spectral response knowledge is introduced in Section 2.1.3, where the central wavelength and bandwidth distributions of the bands are fully propagated to show the radiometric impact in the cross-calibration application. The uncertainty levels obtained are below or at 0.1% for all bands with the exception of B1 and B6 for which the lower bandwidth and high TOA spectral radiance variability compared with other bands increases the error up to 0.3% in some cases. The simulation here included only the systematic effects (the normal distribution shift is applied to all detector pixel bands). Further study should provide a better description of non-linear spectral knowledge contributions in the spectrometer. The results of TRUTHS’ spectral knowledge complement the ones studied in Gorroño et al. (2016a) for the impact of Sentinel-2 spectral knowledge, where similar levels of uncertainty were reported. These preliminary results also pointed to a small impact from the spectral degradation of the S2 bands. Other studies as (Chander et al., 2013a) have reported an impact of the spectral degradation at the level of 2% for some ETM + bands. Although a similar desert scene was used, this is considered here a pessimistic value based on the consideration of spectral filter shift up to 5 nm in both edges, where similar levels of uncertainty were reported. These preliminary results also pointed to a small impact from the spectral degradation of the S2 bands. Other studies as (Chander et al., 2013a) have reported an impact of the spectral degradation at the level of 2% for some ETM + bands. Although a similar desert scene was used, this is considered here a pessimistic value based on the consideration of spectral filter shift up to 5 nm in both directions. The work in (Wu et al., 2015) proposed a 0.5 nm change based on the changes observed in-flight as derived from MODIS in-flight spectral monitoring (Xiong et al., 2006).

Section 2.1.4 has studied the potential improvement of the spectral sampling/resolution error if the spectral binning were not applied for the spectral region around S2 B1. The results in Fig. 9 have shown that further sampling of the region would significantly reduce the sampling/resolution error to a level below 0.1% and also lead to an improvement in the spectral knowledge uncertainty. Recalling the results presented in Green (1998) it can be seen how the sensitivity error is largely symmetrical. That suggests that when decreasing the spectral resolution, the sensitivity increases as does the number of samples across the bands which tend to favourably balance the sensitivity error in this simulation. This approach would only increase a little the memory requirements of the mission. Systems like the one developed in Dell’Endice et al. (2009) are suggested to be included in a mission like TRUTHS/CLARREO so that the change of spectral binning pattern in-flight can be applied to the specific application and provide further flexibility in the cross-calibration with other sensors.

The use of a MODTRAN simulation at 0.1 cm\(^{-1}\) is justified in order to derive meaningful results of the spectral knowledge impact of the TRUTHS sensor. Small spectral variations in this case are well-captured by the fine MODTRAN simulation. Section 2.1.5 has compared the results in sampling/resolution error when using a MODTRAN spectral radiance at 0.5 nm resolution. The comparison of results in Fig. 9 for the B1 and B6 has shown the validity of using such a resolution as in Wu et al. (2015). However, when the results are intended to provide an evolution of the error with the array positioning, it shows a more accurate description of the error evolution when using a narrower TOA spectral radiance.

Section 2.1.6 has studied the variability of the spectral sampling/resolution error for different types of sites. The results in Fig. 10 reveal that the 0.1% is largely maintained for the S2 bands other than B1, B5, and B6. For B5 and B6. The desert simulation has shown to be the worst scenario with all other scene types showing improved values.

3.2. Discussion: spatial domain

Section 2.2 follows a pragmatic approach similar to the one proposed in Chander et al. (2013a) to study the effect of spatial non-uniformity produced by spatial offsets. Using real EO data with low pixel-to-pixel uncertainty and displacements of the ROI—approximately ±10 km and ±0.4 km in latitude and longitude for Libya-4 and La Crau sites respectively—it has been possible to generate a map of TOA reflectance factor error from the site centre. The association of a position knowledge with a distribution of errors in the image have been processed to generate a site curve that links the TOA reflectance factor uncertainty with spatial positioning knowledge. This curve has been found to be highly linear for the studied cases and can be used as a direct input for the definition of cross-calibration requirements.

The method has been evaluated for both the S2 MSI and L8 OLI sensors in near-coincident cloud-free overpasses. The bands selected for the study are B1, B5 and B8 for L8 OLI—443, 865, and 2201 nm central wavelength—and B1, B8A and B12 for S2 MSI—443, 865, and 2190 nm central wavelength—which share an almost coincident SRF shape and positioning. Thus, the results can be cross-validated and provide a reliable result for different parts of the VNIR and SWIR region.

An approximate error of 0.12% for B1, 0.27% for B5 and 0.5% for B12 is calculated for a single overpass over La Crau site. The results for Libya-4 show values below 0.1% for all the studied bands. These values are less than the ones provided by Chander et al. (2013a) most likely due to the impact of dune dispersion as a consequence of a much smaller ROI over the site.

This method uses images with a low relative uncertainty (see earlier discussion) however, neither the solar nor the viewing angles are constant within the study area. This means that the variations across the selected area are the result of TOA reflectance changes combined—either in a constructive or destructive manner—with angular changes. These angular variations cannot be expected to be caused by the displacement over La Crau since the displacement of just 400 m represent a very small angular variation; however they could have a larger impact in the studied case of Libya-4. The TOA reflectance factor error map in Fig. 16 suggest that for S2 B8A and L8 B5, these variations could be attributed to viewing angular variations linked to solar illumination direction. This conclusion is subject to further analyses to understand the effect of such variations. The angular information introduced by both the L8 L1TP and S2 L1C products can be ingested in a model of the site that can estimate these variations.

The agreement between the results for L8 OLI and S2 MSI is excellent with the exception of the L8 B7 and S2 B12 over Libya-4. In this case, the study of the map error in Fig. 16 and their distribution in Fig. 18 suggest that the S2 B12 might have an instrument derived variation. The comparison of the results with the ones described in Chander et al. (2013a), has found that the derived uncertainty is significantly lower. However, this difference could be largely to the consideration of an uncertainty rather than an error and the use of a significantly larger ROI that minimises the dune dispersion.

The association of a standard deviation as a proxy of an uncertainty \( k = 1 \) is based in the fact that it represents a 68.27% of the probability error distribution. The simulations suggest that might slightly vary from the normal distribution —see Fig. 18— and further analysis might be considered to provide an impact of this effect (BIPM et al., 2008a).

3.3. Discussion: temporal domain

Section 2.3 studies the effect of angular changes with time and/or any potential loss of knowledge of angle as a function of time.
Table 6

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Resultant uncertainty on Sentinel-2 TOA reflectance/%Single overpass</th>
<th>Resultant uncertainty on Sentinel-2 TOA reflectance/%Mean of multiple overpasses</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral resolution/accuracy of TRUTHS</td>
<td>0.1</td>
<td>0.1</td>
<td>See above</td>
</tr>
<tr>
<td>Non-spatial co-alignment of TRUTHS with Sensor under test</td>
<td>0.2</td>
<td>0.2</td>
<td>See above</td>
</tr>
<tr>
<td>Error due to 30 min difference in overpass times: solar view angle</td>
<td>0.1</td>
<td>0.1</td>
<td>See above</td>
</tr>
<tr>
<td>Error due to 30 min difference in overpass times: sunposition</td>
<td>0.1</td>
<td>0.1</td>
<td>See above</td>
</tr>
<tr>
<td>Error due to 30 min difference in overpass times: surface BRDF</td>
<td>0.1</td>
<td>0.1</td>
<td>See above</td>
</tr>
<tr>
<td>Error due to lack of knowledge of surface BRDF for 30 min period</td>
<td>0.1</td>
<td>0.1</td>
<td>See above</td>
</tr>
<tr>
<td>Error due to atmospheric path difference between sensor tests</td>
<td>0.2</td>
<td>0.2</td>
<td>See above</td>
</tr>
<tr>
<td>Total uncertainty due to cross-calibration and process for anticipated level of knowledge and conditions (uncertainty sources considered marked with asterisk)</td>
<td>0.3–0.7</td>
<td>0.3–0.7</td>
<td></td>
</tr>
</tbody>
</table>

Calculating change in TOA reflectance that would occur over a 30 min period due to angular variation allows an estimation of potential error (which should be corrected) due to time delay between overpasses of a satellite under test e.g. Landsat-8 OLI and TRUTHS. Two radiative transfer codes were selected—MODTRAN5 and 6SV1—to assess any differences that might occur due to time lapse between overpasses as a consequence of the radiative code used. The result has shown a 0.2% difference between the simulated corrections for spectral regions dominated by the atmospheric scattering i.e. shorter wavelengths. Of additional interest here was also whether this correction and associated differences could be optimised through improved parameterisation. For this, the dominant parameters, aerosol and water vapour, have been modelled as distributions making use of the values presented in Mishra et al. (2014a) and propagated to the TOA radiance/reflectance factors. The results have shown a minimum impact in the correction factor with levels below 0.1%. Nonetheless, this assumes that the aerosol and water vapour knowledge are perfectly constant during the 30 min of the simulation. In considering the overall uncertainty due to knowledge of atmospheric parameters, an analysis of TOA reflectance distribution was performed using the full range of observed atmospheric variations over a 1 year period. The results showed a resultant maximum uncertainty below 1% for B1 and B5 and between 1% and 2% for B7. However, these values represent the uncertainty which would occur without any real correction for atmosphere (assuming worst case annual variations) and most importantly the relatively slow temporal change in atmosphere conditions. For a simultaneous nadir overpass (SNO) cross-calibration, as would be envisaged, the atmospheric variation is likely to very small and thus have little contribution to the uncertainty. However, if we consider a characterisation of a PICS site the temporal variations of the atmosphere can be considered random providing a large uncertainty improvement over several overpasses even without any knowledge of the atmosphere. It should be noted that as a hyperspectral imager, TRUTHS will be able to make some atmospheric retrievals at the time of overpass and thus correct its own observations. The temporal module could be improved by further varying other factors such as the ozone concentration and temperature and by analysing real observations of atmospheric short-term variations. In addition, the aerosol uncertainty distribution should be further improved. A more refined model should look for a distribution of aerosols that only considers positive values and provides an expected distribution of values as e.g. a log-normal distribution.

The effect of the knowledge in the surface reflectance angular correction has been also studied Section 2.3.5. The method employed is similar to that described for the atmospheric variation. Here we assume that over a period of 30 min, the surface reflectance is invariant; for a PICS site this is probably true for a much longer period, except under extreme conditions such as sand storms. The RPV model in Bouvet (2014) has been modified to introduce a 5% uncertainty on each BRDF parameter. This is a worst-case assumption that does not consider the optimisation of the model that would be possible from the TRUTHS observations. The results at 443 nm show an uncertainty that oscillates between 0.2–0.3% at any time in the year whereas the results at 865 nm range between 0.3–0.4%. If we consider the combined effect of the atmosphere, then in the worst case, without applying corrections, the uncertainty due to temporal knowledge is well below the level of 0.5%. These variations due to solar angular change have been found to be consistent with the empirical results in Mishra et al. (2014b) for Libya-4 site.
parameterisation as the ozone content, temperature or pressure. Finally, the surface reflectance correction uncertainty should be upgraded by introduced the impact of the correlation between the different RFV model parameters.

Finally, it is important to mention the benefit of an asynchronous orbit of the reference sensor in terms of temporal effects in cross-calibration. Time delays between the sensor under test and a cross-calibration using TRUTHS or CLARREO will tend to zero as the number of match-ups increase, due to randomness and in turn reducing overall uncertainties. That is whether the systematic uncertainty is produced by the delay between overpasses or in the correction knowledge as studied in Section 2.3, the accumulation of them over different overpasses will tend to reduce the impact.

3.4. Discussion: uncertainty budget

Effects due to viewing angle have not been discussed in this paper but since the reference sensor is considered to be agile it can be aligned to match that of the sensor under test. CLARREO and TRUTHS are designed to have a polarisation sensitivity of less than 0.5% (k = 2) below 1000 nm, and less than 0.75% (k = 2) above 1000 nm (Wielicki et al., 2013). Even though the sensitivity is low, the degree of polarisation might be certainly high for certain spectral regions, sites and angular configuration. Recent work for CLARREO has shown that desert areas present a degree of polarisation at the 10%-level for longer wavelengths but that can raise up to 50% at the shorter wavelengths (Sun et al., 2015). In order to account for the polarisation effect in the cross-calibration, a set of Degree Polarisation Models (DPMs) have been derived as a methodology to account for them has been proposed in (Lukashin et al., 2013).

Table 6 provides a summary of the sources of uncertainty and their relative importance for a range of cross-comparison scenarios using the characteristics of TRUTHS as a reference sensor and Sentinel-2 as the sensor to be calibrated. In this table, the polarisation error and viewing angle effect have not been considered.

4. Conclusion

This paper presents a rigorous approach to evaluate the sources and quantification of uncertainty in post launch Level 1 radiometric gain obtainable from sensor-to-sensor cross-calibration. The novel approach analyses the derived probability distributions for the three main error domains: spectral; spatial; and temporal. The analysis for the considered sites show that a worst case cross-calibration uncertainty (at k = 1) below or at 0.5% can be achieved for a single match-up for each of the three domains for the majority of the overpasses and satellite-to-satellite matching conditions. These values indicate that missions like the proposed TRUTHS or CLARREO with an SI-traceable accuracy of below 0.3% (k = 2) below 1000 nm, and less than 1000 nm (Wielicki et al., 2013). Even though the sensitivity is low, the degree of polarisation might be certainly high for certain spectral regions, sites and angular configuration. Recent work for CLARREO has shown that desert areas present a degree of polarisation at the 10%-level for longer wavelengths but that can raise up to 50% at the shorter wavelengths (Sun et al., 2015). In order to account for the polarisation effect in the cross-calibration, a set of Degree Polarisation Models (DPMs) have been derived as a methodology to account for them has been proposed in (Lukashin et al., 2013).

Table 6 provides a summary of the sources of uncertainty and their relative importance for a range of cross-comparison scenarios using the characteristics of TRUTHS as a reference sensor and Sentinel-2 as the sensor to be calibrated. In this table, the polarisation error and viewing angle effect have not been considered.

Acknowledgements

This work was supported by the National Measurement System of the UK government’s Department for Business, Energy and Industrial Strategy, the UK space Agency’s Centre for Earth Observation Instrumentation and the European Metrology Research Programme (EMRP). The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union. Thanks to the European Space Agency (ESA), National Aeronautics and Space Administration (NASA) and United States Geological Survey (USGS) for the free access to data and inputs used in this tool. We also thank CEDA for the access to the JASMIN supercomputer as well as personal thanks to Dan Lobb ( Surrey Satellite Technology Ltd, SSTL) for the access to the preliminary spectral design of the TRUTHS spectrometer, Paul Kessler (AMA, Inc.) for his support in using the COVE tool, Dr. Marc Bouvet for the access to the surface reflectance model of Libya-4, and Tracy Scanlon and Teresa Goodman for their helpful comments on this paper. Thanks also to the anonymous reviewers for their useful comments.

References


for atmospheric correction and aerosol retrieval: intercomparison study. Appl.
Opt. 47, 2215–2226.


Stafford, B., 2015. PySolar Documentation, 0.7 ed.


