Sentinel-2A and 2B absolute calibration monitoring

Charlotte Revel, Vincent Lonjou, Sébastien Marcq, Camille Desjardins, Bertrand Fougnie, Céline Coppolani-Delle Luche, Nicolas Guilleminot, Anne-Sophie Lacamp, Emmanuel Lourme, Christine Miquel and Xavier Lenot

ABSTRACT
As part of the Copernicus program, Sentinel-2 is the optical imaging mission designed for the operational monitoring of land and coastal areas. It offers a unique combination of global coverage with a wide field of view, a high revisit capability (5 days with two satellites), a high resolution and multi-spectral imagery.

CNES, the French Space Agency, was involved in the commissioning of both Sentinel-2 satellites and is currently working in collaboration with ESA on their long-term monitoring. This paper reviews all the techniques used to ensure an absolute calibration of the 13 spectral bands to better than 5% (the target is 3%) at TOA level. After a brief description of the mission and its related radiometric calibration scheme, we show how standard vicarious calibration methods based on acquisitions over natural targets (oceans, deserts, and Antarctica during winter) are used to check and improve the accuracy of the absolute calibration coefficients. Finally, the verification scheme, exploiting photometer in-situ measurements over the La Crau plain in France and Gobabeb in Namibia, is described. The paper concludes with a summary that includes spectral coherence, agreement between the results obtained with various calibration methods and temporal evolution.

ARTICLE HISTORY
Received 25 July 2017
Revised 17 December 2018
Accepted 19 December 2018

KEYWORDS
Sentinel-2; vicarious calibration; multi-spectral calibration; cross-calibration

Introduction
The Sentinel-2 (S2) mission comprises two satellites: Sentinel-2A, hereafter S2A, launched in 2015, and Sentinel-2B (resp. S2B) launched in March 2017. Each satellite carries a single imaging payload, the so-called MSI after Multi-Spectral Instrument. These constitute the new generation of European satellites dedicated to the global monitoring of land masses. Sentinel-2 offers a unique combination of global coverage with a wide field of view (290 km), a high revisit capability (5 days with two satellites), high resolution (10 m, 20 m and 60 m) and multi-spectral imagery (13 spectral bands in the visible, near-infrared (VNIR) and short wavelength infrared domains (SWIR)). A detailed description of the instruments together with an S2A performance review is available in Gascon et al. (2017).

To maximize its overall use and scientific return, S2 has to be cross-calibrated with past missions (SPOT series, Landsat series, MERIS…) and current missions (Landsat-8, Sentinel-3…) of the same class. It, therefore, has stringent radiometric requirements (Sentinel-2 Mission Requirement Document, 2007), as detailed in Table 1.

There are two focal planes on the MSI: one for the visible spectral bands (<1 μm) and one for the SWIR ones (1–2.5 μm). The SWIR detectors are cooled to 190 K to achieve high radiometric performance. As such, they are susceptible to contamination and consequently their sensitivity changes more rapidly. To properly take this phenomenon into account, the S2 calibration scheme is based on a full-field and full-pupil on-board diffuser mounted on the instrument shutter mechanism. It is used as a calibration device by collecting the solar irradiance after reflection onto the diffuser. This constitutes the nominal method for equalization and absolute calibration. A calibration acquisition sequence is executed on a monthly basis and allows frequent updating of the calibration coefficients. However, diffuser performance in space is known to degrade with time under the effect of the space environment and more particularly the effect of sunlight (Xiong et al., 2007). Consequently, it is essential to continuously monitor the stability of the diffuser with independent calibration methods. In the following sections of this paper, we review all the techniques used to monitor the absolute calibration of Sentinel-2 satellites, and we present the corresponding main results.

These methods have all been published and well validated already (Dinguirard & Slater, 1999). As such they are considered as state-of-the-art by international working groups such as CEOS/IVOS and GSICS for absolute calibration monitoring and cross-calibration.
calibration of operational space sensors (Kaufman and Holben, 1993).

More recent techniques based on supervised vicarious calibration as reported in Brook and Ben-Dor, 2014 and Li et al., 2015 are not considered in this paper because they are not well suited to our work. In particular, they are not suitable for a 300-km-wide swath such as that of Sentinel-2 since the size of the calibration targets is generally of the order of 10 m. Moreover, there is no existing operational system based on this technique, whereas our aim is to achieve continuous monitoring. Although they may be promising, these methods are currently more suitable for small swath instruments doing time-limited acquisition campaigns, such as on aircraft or unmanned aerial vehicles.

Following a brief overview of all calibration methods, the next sections will give a more detailed description. The desert method is based on a comparison of desert images from different sensors (Lachérade, Fougnie, Henry, & Gamet, 2013). This is a sensor cross-calibration method working in the full spectral range of Sentinel-2 between 0.4 µm and 2.5 µm. The geometric matching process makes the method only weakly sensitive to the sites’ BRDF. Since no simultaneity is required, this calibration method is a good way to ensure data continuity between several missions. Another method consists in using the Rayleigh scattering radiance as an absolute reference (Fougnie, Llido, Cross-Colzy, Henry, & Blumstein, 2010). This works on simple and well-known targets such as oceans but only in a limited part of the S2 spectral range (B1-B6) where Rayleigh scattering occurs. Deep Convective Clouds (DCC) are then used for interband calibration in the spectral range B1–B8 (Fougnie & Bach, 2009). This method is inherently insensitive to cloud BRDF, which is quite flat, since only data taken under the same solar and viewing geometries are compared. Finally, a method relying on automated instrumented sites performing in-situ measurements is used for validation purposes (Meygret, Santer, & Berthelot, 2011). It performs well in all Sentinel-2 spectral ranges, except for the atmospheric absorption bands and B12.

In the final section, particular attention is paid to the cross-calibration between S2A and S2B. Finally, it is important to note that, due to the respective launch dates of S2A and S2B, namely 2015 and 2017, the sizes of the datasets used in this work are quite disproportionate. Hence, while S2A results are considered reliable, those from S2B are still preliminary.

All the figures presented in this paper are relative absolute calibration coefficients, also called calibration coefficients, \( \Delta A_{k_{\text{method}}} \):

\[
\Delta A_{k_{\text{method}}} = \frac{A_{k_{\text{method}}}}{A_{k_{\text{official}}}}
\]  

(1)

where \( k \) is the spectral band, \( A_{k_{\text{method}}} \) the best absolute calibration coefficient determined by any calibration method and \( A_{k_{\text{official}}} \) the official absolute calibration based on on-board diffuser data analysis. \( \Delta A_{k} \) is equivalent to:

\[
\Delta A_{k} = \frac{\rho_{\text{measured-S2}}}{\rho_{\text{reference}}}
\]

(2)

where \( \rho_{\text{measured-S2}} \) is the reflectance measured by S2 and \( \rho_{\text{reference}} \) a reference reflectance provided by another sensor or simulation depending on the calibration method.

Cross-calibration over desert sites

Description of the method

Calibration over deserts provides cross-calibration between two sensors (the reference sensor and the sensor to be calibrated). Assuming the temporal stability of the desert sites, also called pseudo-invariant calibration sites (PICS), and a good characterisation of the atmosphere using exogenous data to avoid errors due to non-simultaneous acquisitions, the inter-calibration process is performed by comparing the TOA reflectance of the same scene observed by two different sensors (Bhatt et al., Cabot, Hagolle, & Henry, 2000, Chander et al., 2013, Henry, Dinguirard, & Bidilis, 1993, Lachérade et al., 2013, Smith & Cox, 2013, Sterckx, Livens, & Adriaensen, 2013). To deal with directional effects, the scene images can only be compared if they are acquired with similar viewing and solar directions. An alternative approach based on BRDF model fitting using MERIS data has been developed by Bouvet (2014). Since it is very difficult to acquire data from two sensors simultaneously under identical geometrical conditions, the requirement for geometrical matching is relaxed somewhat. The results we present later were obtained assuming that the viewing angle and the solar direction were constrained as follow:

\[
|\theta_{S}^{S2} - \theta_{S}^{REF}| < 5^\circ
\]

(3)

\[
|\theta_{V}^{S2} - \theta_{V}^{REF}| < 5^\circ
\]

(4)
where $\theta_S^{S2}$ (resp. $\theta_S^{SW}$) is the solar zenith angle, $\theta_V^{S2}$ (resp. $\theta_V^{SW}$) the ground-level zenith viewing angle, and $\phi_S^{S2} - \phi_V^{S2}$ (resp. $\phi_S^{SW} - \phi_V^{SW}$) the relative azimuth angle of Sentinel-2 (resp. reference sensor). These conditions can be more restrictive as in Lachérade et al. (2013) but these values are a good compromise between precision and the number of matches.

After the selection of the match between the reference and to-be-calibrated data, the measurements made by the reference sensor are corrected for atmospheric effects and used to fit a spectral reflectance model which is then included in the spectral response of the sensor to be calibrated. This surface reflectance is finally transferred to the top of atmosphere to be used for comparison. The radiative transfer is based on atmospheric conditions derived from a mix of meteorological data (H2O, O3, surface pressure) and climatological data for aerosol optical thickness. The radiative transfer is performed by using the SMAC model (Rahman & Dedieu, 1994) based on 6SV code (Vermote et al., 2006).

Using Meteosat data (Cosnefroy, Leroy, & Briottet, 1996), 19 deserts over North Africa and the Middle East were selected for their spatial uniformity and temporal stability. Their locations are depicted in Figure 1. Despite these sites are not perfectly Lambertian, the directional variation of their reflectances was monitored using different sensors. In particular, the choice of the constraints (1), (2) and (3) is justified. All the collected data were stored in a database. When possible, in this database, the desert reflectances were extracted from the data over two areas:

- Standard site: the desert sites were defined by a ± 0.45° square in latitude and longitude.

- Small site: the desert sites were defined by a ± 0.1° square in latitude and longitude. All the small sites were located inside the standard site and usually correspond to a more uniform area. The small sites were chosen to be radiometrically consistent with the large sites to within 1 to 2%; this consistency is still true today.

There are approximately 48,000 non-cloudy PICS acquisitions from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard AQUA and 22,000 from the MEEdium Resolution Imaging Spectrometer (MERIS) onboard ENVISAT since their respective launches. Theoretically, the swath of S2 (300 km) would allow standard sites to be imaged, but only a few of them would be acquired in a single acquisition (two orbits needed). Hence, the analysis presented here is only based on small desert sites except for the cross-calibration between S2A and Sentinel 3 where small sites were calibrated with standard sites.

The pixel selection takes advantage of the cloud mask provided at the level 1C. There is no restriction on the geometry of the acquisition as Sentinel-2 is limited to nadir viewing in nominal mode.

As already mentioned, all comparisons between the reference sensor and the to-be-calibrated sensor were made on the basis of TOA reflectances. The following formulation is used to compute the TOA signal for a given spectral band:

$$p_{TOA}(\lambda, \theta_S, \theta_V, \phi) = t_g(\lambda, \theta_S, \theta_V) \cdot \left( p_A(\lambda, \theta_S, \theta_V, \phi) + p_{surface}(\lambda, \theta_S, \theta_V, \phi) \right) \cdot T(\lambda, \theta_S, \theta_V)$$

where $\theta_S$, $\theta_V$, and $\phi$ are, respectively, the solar and viewing zenith angles and relative azimuth angle, $t_g$ is the gaseous transmittance, $p_A$ is the molecular and
aerosol contribution including coupling term, T is the total atmospheric transmission for aerosol and molecules, \( \rho_{\text{surface}} \) is the surface reflectance over the desert site, and \( S_A \) is the atmospheric albedo.

The surface reflectance \( \rho_{\text{surface}} \) is derived from the TOA reflectance measured by the reference sensor \( \rho_{\text{RS,reference}} \):

\[
\rho_{\text{RS,reference}} = \frac{\rho_{\text{RS}} - \rho_A \cdot t_g}{T \cdot t_g - \rho_A \cdot t_g + S_A \cdot \rho_{\text{RS}}} \tag{7}
\]

where \( t_g \), \( S_A \), \( \rho_A \) and T are defined as previously (for clarity the dependence of each term on \( \theta_5 \), \( \theta_V \) and \( \phi \) was omitted). The surface reflectance \( \rho_{\text{surface}} \) can be assessed by a combination of the transmission t of the spectral band \( \lambda_{S2} \) and a surface reflectance spectrum \( \rho_{\text{spectral}} \) according to:

\[
\rho_{\text{surface}}(\lambda_{S2}) = \int_{\lambda_{S2}} \rho_{\text{spectral}}(\lambda) \cdot t(\lambda) \, d\lambda
\]

where the surface spectrum is obtained through a spectral interpolation of \( \rho_{\text{RS,reference}} \) using a spline function.

The atmospheric TOA reflectance, originating from molecular diffusion, aerosol diffusion, and coupling terms, was computed using the 6S radiative transfer code (Vermote, Tanré, Deuzé, Herman, & Morcrette, 1997). A desert aerosol model, with configurable aerosol optical thickness fixed thanks a climatology, was assumed. Ozone content, water vapour and surface pressure were taken from meteorological data coming from the National Centers for Environmental Prediction (NCEP). Gaseous transmittances were computed for atmospheric corrections using the SMAC transmittance model (Rahman & Dedieu, 1994).

S2 was cross-calibrated with respect to different sensors (MERIS, MODIS, LANDSAT8, SPOT5, its alter ego S2) with the following objectives:

- To get absolute information for S2, taking these sensors as a reference
- To monitor the change in instrument sensitivity and assess the diffuser stability

**Results**

Table 3 and Figure 2 show the results obtained by performing cross-calibration between S2A and LANDSAT8, MERIS, MODIS, SPOT-5 and Sentinel-3 OLCI. S2B results are presented in Table 4 and Figure 3. SPOT-5 and OLCI were not compared to S2B because the corresponding results are not considered sufficiently significant in this early S2B mission phase.

The error bars shown in Figures 2 and 3 are the standard deviations associated with the determination of each absolute calibration, assuming that there is no bias in the process, using the best method to our knowledge. However, there are some flaws in the method, mostly depending on the spectral interpolation. Hence, results presented in bold on the figure correspond to where the spectral bands are well matched between S2 and the other sensor, and there is little spectral interpolation error. The choice of these spectral bands is described in the section on spectral band analysis (Section Description of the method).

Generally speaking, for spectral bands which correspond to little spectral interpolation, almost all the calibration coefficients are within the 3% goal specification. For all the bands, even those for which we knew the method might perform less well, many of the coefficients lie within the 3% but all are within the 5% (threshold specification) of the diffuser calibration.

The cross-calibration between S2A and MERIS seems to out-perform the other cross-calibrations, with small errors bars and calibration coefficients.

**Table 2.** S2 spectral bands which can be cross-calibrated over desert sites with different sensors. X indicates that S2 has spectral bands close to the other sensor. In that case, the result can be used for absolute calibration. X* indicates that there are spectral discrepancies between the two sensors. The analysis is then limited to temporal monitoring and validation of dataset consistency. An empty cell indicates that the reference sensor does not fit at all for calibration.

| Reference sensor | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B8A | B11 | B12 |
|------------------|----|----|----|----|----|----|----|----|-----|-----|-----|
| MODIS            | X  | X* | X  | X  | X* | X  | X  | X  | X   |     |     |
| MERIS            | X  | X  | X  | X  | X* | X* | X  | X  |     |     |     |
| LANDSAT8        | X  | X* | X  | X  | X* | X* | X  | X  | X   |     |     |
| SPOT5           | X* | X* | X  | X* | X* | X  | X* | X  |     |     |     |
| Alter ego       | X  | X  | X  | X  | X  | X  | X  | X  |     |     |     |
| SENTINEL-2      | X  | X  | X  | X  | X  | X  | X  | X  |     |     |     |
close to 1. These good results were expected since MERIS and S2 sensors have close spectral bands (see Figure 4). The S2A/SPOT-5 cross-calibration is quite poor since the two sensors have very different spectral bands.

The cross-calibration between S2A and Sentinel 3 is an example of cross-calibration that mixes small and standard sites. A bias of about 2% between the two sensors can be observed. The bias also appears in the case where OLCI is the reference Sensor and S2A the to-be-calibrated one. However, these results have to be treated with caution since errors can be due to calibration with non-identical sites and also to the small number of matches (only 600).

In the particular case of the S2A-S2B cross-calibration, the number of matches is still quite high thanks to the S2 mission’s fixed observing geometry. Hence, the results can be assumed to be reliable. A small bias of about 1–2% between S2A and S2B in Figure 3 can be observed. This point was adopted for the study and discussed in detail in the Synthesis section.

In general terms, the dispersion of the results could be related to the spatial non-uniformity or directional effects of the desert sites. Indeed, a study showed that the sites are not equally homogeneous (Cosnefroy et al., 1996). In this study, they were grouped into five classes (very homogeneous, homogeneous, acceptable, heterogeneous, and inhomogeneous). When using a subset containing only the most uniform sites, the results are significantly improved, with lower dispersion around the mean.

To verify the temporal stability of S2A, a temporal study was performed on the same data set. Figure 4 shows the change in the calibration coefficient for four bands (B3, B4, B8A and B11) as a function of time. Slow seasonal variations can be observed, in particular by studying the cross-calibration between S2A and MERIS. These variations are interpreted as the result of climatological approximations, small site reflectance variations and potentially directional effects. Figure 4 also shows large error bars for the cross-calibration between S2A and MODIS. This high standard deviation is due to an outlier measurement during the associated 30 day period. It illustrates the need for a large number of acquisitions to obtain reliable statistics. The temporal stability of S2B is not reported here because the corresponding dataset is considered too small to be meaningful.

### Calibration using Rayleigh scattering

#### Description of the method

The top-of-atmosphere signal observed over Deep Ocean at short wavelengths is mainly due to atmospheric molecular scattering. This scattering, also known as Rayleigh scattering, can be predicted accurately and computed with a radiative transfer code. The other contributions to the TOA signal are aerosol scattering, back-scattering by the water body, diffuse reflection from whitecaps, specular (or Fresnel) reflection from the surface and gaseous absorption. Satellite acquisitions over such oceanic targets are selected so as to minimize these contributions.

This method is used to calibrate B1, B2, B3 and B4 (and eventually B5 and B6) where molecular scattering represents the bulk of the signal.

The calibration consists in simulating the TOA normalized reflectance seen by the sensor and comparing it to the measurement (Fougnie et al., 2010, Frouin et al. 2014).

The marine contribution represents 10 to 15% of the TOA signal for the blue bands and is consequently an important source of error in the TOA signal. A climatological study based on the analysis of 1 year of SeaWiFS data, selected six oceanic sites with good spatial homogeneity and moderate seasonal effects (Fougnie et al., 2010). These sites are located in the North and South Pacific, the North and South Atlantic, and in the Indian ocean (Figure 5).

Exogenous data are necessary to accurately compute the TOA signal, such as surface pressure, surface wind speed, or total ozone.
The TOA signal (reflectance) is computed according to the following formula:

\[
\rho_{\text{TOA}}(s, v, \varphi) = \frac{\rho_A(s, v, \varphi)}{\rho_{\text{w}}(s, v, \varphi)} + \frac{\rho_{\text{w}}(s, v, \varphi) T(s, v) [1 - S_A \rho_{\text{w}}(s, v, \varphi)]}{C_8 + \rho_{\text{w}}(s, v, \varphi) T(s, v) = 1/C_0 S_A \rho_{\text{w}}(s, v, \varphi) / C_2 / C_3}
\]

where \( \theta_s \), \( \theta_v \) and \( \varphi \) are the solar and viewing zenith angles and relative azimuth angles, respectively, \( \rho_{\text{g}} \) is the total gaseous transmittance, \( \rho_A \) is the molecular and aerosol contribution including coupling terms and specular reflection from the wave-covered surface, \( \rho_{\text{w}} \) is the remote-sensing reflectance, \( T \) is the total atmospheric transmission for aerosols and molecules, and \( S_A \) is the atmospheric albedo. Note that \( \rho_{\text{g}} \) depends on the amount of absorbers (essentially ozone), \( \rho_A \) on aerosol optical thickness, surface pressure and wind speed, \( T \) on surface pressure and wind speed, and \( S_A \) on aerosol optical thickness. \( \rho_{\text{w}} \) is defined as the ratio of water-leaving radiance to downwelling irradiance just above the surface. It varies with both water constituents and angular geometry as described in Lee et al. 2011.

The atmospheric functions \( \rho_A \), \( T \), and \( S_A \) are computed using an accurate radiative transfer model such as shown in Figure 2. S2A cross-calibration over desert sites with respect to reference sensors. The dashed (resp. solid) red lines represent the 3% goal specification (resp. 5% threshold specification). The bold blue points/lines are those where the method theoretically should perform the best based on spectral analysis.
as the successive order of scattering code (Lenoble et al., 2007). The molecular scattering contribution is accurately computed knowing the surface pressure and the molecular optical thickness corresponding to the spectral band being considered. The background aerosol contribution is computed knowing its estimated optical thickness of 865 nm (band B8a) and extrapolated for the spectral band being considered using a Maritime aerosol model.

The marine contribution is estimated over the pre-defined oceanic sites through the climatological study. Typical sea remote-sensing reflectances for these sites are 0.033 at 443 nm, 0.020 at 490 nm, 0.0049 at 555 nm and 0.0007 at 670 nm, and are close to values derived through a bio-optical model using a surface pigment concentration of 0.07 m/m³. A spectral interpolation is performed for S2 when the spectral band of interest is not exactly the same as one of the SeaWiFS spectral bands (Feldman, 2003, Hooker & Mcclain, 2000) for which the climatological values are available. In addition, a bidirectional correction is added as an option in order to take into account particular differences in the viewing and solar geometries of the pixel to be calibrated and Table 4. S2B cross-calibration results over desert sites for different sensors. The bold values are those associated with the bands for which the method theoretically should perform the best based on spectral band analysis (cf. Table 2).

| Sensor     | B1 Mean | Standard deviation | B2 Mean | Standard deviation | B3 Mean | Standard deviation | B4 Mean | Standard deviation |
|------------|---------|--------------------|---------|--------------------|---------|--------------------|---------|--------------------|
| SENTINEL-2B| 0.98    | 3%                 | 0.98    | 3%                 | 0.98    | 3%                 | 0.98    | 3%                 |
| LANDSAT8   | 0.96    | 2%                 | 0.99    | 3%                 | 0.99    | 3%                 | 0.99    | 3%                 |
| MERIS      | 0.96    | 2%                 | 0.99    | 3%                 | 0.96    | 2%                 | 0.99    | 3%                 |
| MODIS      | 0.98    | 3%                 | 1.00    | 3%                 | 0.98    | 3%                 | 0.98    | 3%                 |
| SENTINEL-2A| 0.98    | 4%                 | 0.99    | 4%                 | 1.02    | 3%                 | 0.99    | 3%                 |
| B5         | 0.97    | 2%                 | 0.98    | 3%                 | 0.97    | 2%                 | 0.99    | 3%                 |
| B6         | 0.98    | 3%                 | 1.00    | 3%                 | 0.98    | 2%                 | 0.98    | 3%                 |
| B7         | 0.98    | 2%                 | 0.99    | 3%                 | 0.97    | 2%                 | 0.97    | 3%                 |
| B8         | 0.99    | 1%                 | 1.00    | 3%                 | 1.01    | 2%                 | 0.99    | 3%                 |
| B8A        | 1.00    | 1%                 | 1.00    | 3%                 | 0.98    | 2%                 | 1.00    | 3%                 |
| B11        | 0.97    | 2%                 | 0.95    | 3%                 | 1.00    | 4%                 | 1.00    | 3%                 |
| B12        | 0.99    | 1%                 | -       | -                  | 0.98    | 2%                 | 0.99    | 3%                 |

Figure 3. S2B cross-calibration over desert sites with S2A as reference sensor. The dashed (resp. solid) red lines represent the 3% goal specification (resp. 5% threshold specification). The bold blue points/lines are those where the method theoretically should perform the best based on spectral analysis.
the angular conditions of the climatological values derived from SeaWiFS.

The main gaseous absorption contributors are water vapour (mainly around 565 and 865 nm), ozone (mainly around 490, 565, and 670 nm), oxygen and nitrogen dioxide (mainly around 443 and 490 nm). The correction is made according to the SMAC model using exponential variation with air mass and gaseous partial volumes. Table 5 summarizes the mean contribution to the TOA signal.

This is the only "absolute" calibration method. Some external data are used for the calculation (SeaWiFs for surface reflectance climatology, TOMS for O₃,), but these contributions are of second order compared with the physical calculation of molecular scattering as visible in Table 5.

**Results**

Due to S2 near-nadir viewing conditions, the contribution from the surface is quite significant, mostly in the blue bands, as compared with other sensors that were calibrated by CNES (POLDERs, VEGETATION, SPOT etc.). Less strict data selection parameters have, therefore, been used in order to maximize the number of valid measurements.

Results are given in Table 6 and Figure 6. All spectral bands show an absolute calibration coefficient within 3% of the diffuser. The absolute calibration coefficients for B4, B5 and B6 are even closer since they confirm the values derived from the diffuser to within 1%. Note that it is at these wavelengths that the method performs best as stated in the

![Figure 4. Change of S2A calibration coefficients since launch every 30 days determined using different reference sensors.](image)

![Figure 5. The six oligotrophic zones used for calibration over molecular scattering.](image)

| Spectral band | Average relative contributions to TOA signal | gas | Average measured reflectance |
|---------------|---------------------------------------------|-----|-----------------------------|
|               | Rayleigh | aerosols | marine | Transmission | Absorption |                   |                  |
| B1            | 80%      | 2%        | 18%     | 0.998        | 0.2%       | 0.147             |                  |
| B2            | 83%      | 3%        | 14%     | 0.983        | 1.7%       | 0.103             |                  |
| B3            | 89%      | 6%        | 5%      | 0.918        | 8.2%       | 0.055             |                  |
| B4            | 87%      | 11%       | 2%      | 0.969        | 3.1%       | 0.028             |                  |
previous sub-section, so the confidence in these results is high. The impact of the marine contribution is visible on B1 and B2; the results are less satisfactory, although they still meet the specification.

Figure 7 depicts the distribution of the valid measurement for S2A. That for S2B is similar but not reported. The left-hand histogram is a function of the acquired sites. The sites that are closest to the equator (Pacific North-West, Atlantic South and Atlantic North) are viewed only by the first detectors. This is due to the restrictive condition added to avoid the effects of sun glint on the data. It explains the shape of the Rayleigh distribution in the histogram on the right-hand side of Figure 7.

This variation in the number of acquisitions by detectors could lead to errors. Figure 8 shows the calibration coefficient for B4, respectively, as a function of the geographical site (left) and as a function of the detector number. This figure leads to two conclusions: due to the variation of the sea surface reflectance and other parameters (see Table 5) a small variation appears in the calibration results between the sites, then the calibration coefficients are not stable in the field of view; this could be due to the variation in the number of valid measurements for each detector.

Other options have been considered to explain the difference in consistency between the Rayleigh calibration coefficients and the diffuser results. The difference in spectral response in the field of view could lead to a bias in the absolute calibration coefficients due to the preferential distribution of valid points in the Western part of the field of view (FOV, Figure 7, right-hand side). Figure 9 proves that this is not the

|        | SENTINEL-2A |        | SENTINEL-2B |
|--------|-------------|--------|-------------|
|        | Mean | Standard deviation | Mean | Standard deviation |
| B1     | 1.02 | 1.8% | 1.02 | 0.02 |
| B2     | 1.02 | 1.2% | 1.02 | 0.02 |
| B3     | 0.99 | 1.1% | 0.99 | 0.01 |
| B4     | 0.99 | 1.3% | 1.00 | 0.01 |
| B5     | 1.00 | 1.0% | 1.01 | 0.01 |
| B6     | 1.00 | 1.1% | 1.02 | 0.01 |

Figure 6. Absolute calibration coefficients using the absolute calibration method over molecular scattering, using Sentinel-2A images calibrated using the diffuser. The dashed (resp. solid) red lines represent the 3% goal specification (resp. 5% threshold specification).

Figure 7. Distribution of S2A valid measurements over Ocean sites as a function of detector number and the geographic site (left) and as only a function of detector number (right). The western part of the swath (detector 1) is less filtered because it is less sensitive to sun glint thanks to the local acquisition time of 10h30.
case, and although the calibration coefficients derived from detectors 8 to 12 (Eastside) are lower than the others, they are not significant because there are few measurements there and there is good consistency between all values to within 1%. These inconsistencies could also come from other instrument effects like residual error in the aerosol retrieval band or non-linearity, in particular the blue bands where all instrument defects are amplified. The last option is the locations of the acquisitions. While the selected sites are more or less oligotrophic (see Figure 7), the measurements cover only a part of the south east of the Pacific site, which means that some errors can be due to the observed scene.

Hence, the Rayleigh scattering calibration method appears to be a good calibration method for bands B3 to B6. For the B1 and B2 bands, however, particular attention has to be paid to the effects of sea remote-sensing reflectance and instrumental artefacts.

**Calibration over Deep Convective Clouds**

**Description of the method**

This method is usually used for lower resolution sensors where pixel sizes are in the order of a few hundred metres (Fougnie & Bach, 2009), and has been experimentally validated for Sentinel-2 to demonstrate its potential for higher resolution sensors (Lamquin, Bruniquel, & Gascon, 2017). The results obtained during S2A in-orbit commissioning (IOC) have been very promising. This method is an inter-band calibration method, which means it only gives an absolute calibration coefficient with respect to the Rayleigh scattering band.
to another band. It is useful in order to verify inter-band accuracy specifications or in the synthesis of all calibration methods. Any spectral band in the VNIR range can be used, but the red spectral band is usually used (670 nm or B4 for Sentinel-2B) since the Rayleigh calibration is assumed to be very efficient at this wavelength and can provide an absolute calibration value for this band.

The favourable imaging zones for this method are warm ocean sites in tropical latitudes where cumulonimbus clouds develop, such as over the Maldives or in the Gulf of Guinea. Since this method was experimental for S2A and there were operational constraints preventing additional images, only a small zone (600 x 700 km²) was chosen over the Maldives during the IOC (Figure 10). This site was preserved in the S2 acquisition plan and is now used for DDC calibration.

Cumulonimbus clouds or Deep Convective Clouds (DCC) in this area are high and dense. They strongly diffuse the incoming solar radiance with a practically white and Lambertian spectral response. Moreover, they reach a very high altitude (8–12 km), and there is almost no atmospheric perturbation of the signal over the cloud.

The different radiances are computed in pre-calculated Look-Up Tables (LUTs). The main difficulty of this method is the characterization of DCC and in particular the components of the top layer that usually consists of ice particles. The reflectance of the cloud becomes sensitive to the type of particle for wavelengths greater than 850 nm, which is why this method is only used in the VNIR range.

Results

The results presented in Table 7 and Figure 11 show very good consistency, with all bands from the VNIR range being within 2% of each other; this complies with the inter-band calibration specification of 3%.

Note that B3 may differ from the other bands because this band is very sensitive to the ozone content, which is derived from meteorological exogenous data; hence having only one major DCC product may create a bias there.

The calibration results of B8 and B8A, which are at the limits of the method, are surprisingly good. Indeed, for these bands, the reflectance of the cloud starts to become sensitive to the type of cloud particle, so poorer results were expected. It may be advantageous to study an extension of this method to other bands.

The method is also less limited by geometric conditions than the calibration method based on molecular scattering, so calibration results can be derived within the whole field of view of the instrument. The distribution of valid measurements presented in Figure 12 shows an increased number of points in the central and eastern part of the FOV, but is only due to the distribution of cloud within the images. More acquisitions would increase the number of measurements per detector and confirm the behaviour of the calibration coefficients within the FOV. Figure 13 shows the variation of the calibration coefficients over the FOV.

In conclusion, this method shows impressive results: the 3% inter-band specification goal is confirmed for bands from B1 to B8A, both for S2A and S2B.

Figure 10. Maldives DCC site chosen for S2 interband calibration.
Calibration based on instrumented sites

Description of the method

CNES has developed and installed on the La Crau site (FRANCE) an automatic ground-based station named ROSAS (Robotic Station for Atmosphere and Surface). See Meygret et al. (2011) for a detailed description of the system. ROSAS continuously provides atmosphere and ground characterization allowing the calibration of any high-resolution sensor which passes over the site. This station comprises a CIMEL photometer mounted on top of a 10 m pole.

Every non-cloudy day, the photometer automatically and sequentially performs direct sun irradiance, sky radiance (almucantar and principal plane) and ground radiance measurements. Data are transmitted to CNES and processed. The surface reflectance in the viewing direction of Sentinel-2 acquisition is estimated using ROSAS acquisitions for the whole day and the TOA reflectance is simulated with a radiative transfer code (6S) with the atmospheric parameters measured simultaneously with the satellite overpass.

Table 7. Interband calibration results over Deep Convective Clouds using B4 as the reference band.

|             | SENTINEL-2A |             | SENTINEL-2B |
|-------------|-------------|-------------|-------------|
|             | Mean        | Standard deviation | Mean        | Standard deviation |
| B1          | 1.00        | 0.5%        | 1.00        | 0.4%        |
| B2          | 0.99        | 0.2%        | 0.99        | 0.4%        |
| B3          | 0.99        | 0.2%        | 0.99        | 0.2%        |
| B4          | -           | -           | -           | -           |
| B5          | 1.00        | 0.2%        | 0.99        | 0.3%        |
| B6          | 1.00        | 0.2%        | 0.99        | 0.3%        |
| B7          | 1.01        | 0.5%        | 1.00        | 0.5%        |
| B8          | 1.00        | 0.6%        | 0.99        | 0.7%        |
| B8A         | 1.00        | 0.7%        | 1.00        | 0.9%        |

Figure 11. Interband calibration results over Deep Convective Clouds using B4 as the reference band.

Figure 12. S2B distribution of valid measurements over DCC sites as a function of detector number.
More recently, ESA and CNES have set up a system equivalent to ROSAS in Gobabeb, Namibia, presented in Marcq et al. (2018). Use is made of both sites in this work. Other sites from the RadCalNet (Automated Radiative Calibration Network) project can be used for such calibration such as Railroad Valley (USA), Czapla-Myers, McCorkel, Anderson, and Biggar (2017), or Baotou (Chine), Wang et al. (2017).

The S2 images are pre-processed in order to extract the non-cloudy reflectance over the area characterized by a photometer. This reflectance is then compared to the simulated one, thereby providing an absolute calibration coefficient.

Results

Results obtained from the La Crau instrumented site are presented in Figure 14. They are not yet satisfactory. S2A results show large error bars that are attributed to seasonal variations and systematic effects not present in S2B since its associated dataset is smaller. Indeed, there are 14 usable S2B acquisitions, but only six of them present good quality corresponding ground data, and no single one has more than the six complete ground cycles that are usually required to perform a satisfactory absolute calibration. The results presented here are for information only and are not taken into account in the summary. Even so, all results are still on average within the 5% threshold specification.

Regular acquisitions of the La Crau area will be performed during the operational phase, so this result will be updated in the future with a larger dataset.

A calibration was also performed with data from the new CNES/ESA instrumental site, Gobabeb in Namibia. The results are shown in Figure 15 for S2A and S2B, and are very promising. The Gobabeb site is more homogeneous than La Crau and weather conditions are generally better in Namibia than in France. This might explain why the Gobabeb calibration seems superior to that of La Crau.

Synthesis

The S2A-S2B cross-calibration over desert sites indicates a bias of about 1–2% between the two satellites. One way to confirm this trend is to compute so-called double ratios. This consists simply in dividing

Figure 13. B1 (left) and B2 (right) inter-band calibration coefficients as a function of the detector number, with B4 as a reference band.

Figure 14. ROSAS calibration results for La Crau acquisitions.
the calibration coefficients obtained by both sensors for all absolute calibration methods. The result of this ratio is by construction free of potential method biases because they are roughly the same for S2A and S2B. Note that the DCC method is not considered in this work as an absolute calibration method but only as an interband calibration method and is therefore not used here. The S2A-S2B cross-calibration by double ratio is reported in Figure 16. This clearly confirms the ~1–2% S2A-S2B bias in VNIR that was established using the desert method and also shows an overall good consistency for all our methods.

An overview of all vicarious calibration results for S2B is presented in Figure 17. There is good consistency between them and with respect to the diffuser. All calibration results obtained with the best effective methods (bold square) are within the 3% goal specification. Even for the methods for which good results were not expected (some spectral bands over deserts, La Crau), the calibration coefficients obtained are still within the 5% threshold. This confirms the usefulness of vicarious methods in addition to calibration with a diffuser to confirm the uncertainty budget associated with these diffuser results, and also to see how it varies with time.

**Conclusion**

This paper reports describes the work done by CNES to monitor the Sentinel-2 mission absolute calibration and cross-calibration with other sensors. A set of independent methods are used, each one having its own features and limitations. However, the combination of those methods provides a comprehensive view of the Sentinel-2 calibration accuracy.

First of all, most of the bands are cross-calibrated with other reference sensors in the field to better than 3%. In the other cases, calibration accuracy is always better than 5%.

No change overtime has been observed for both VNIR and SWIR focal planes. The latter is subject to contamination, so any change in its sensitivity is well compensated thanks to monthly diffuser calibrations.

From the beginning of 2018, there have been indications of a cross-calibration offset of the order of 1 to 2% between S2B and S2A or MERIS in the VNIR range. This result was derived using about 6 months worth of S2B data and as such, has to be used with caution and balanced with an updated analysis based on a more significant dataset.
The authors would like to thank the reviewers for their time and valuable comments, which helped considerably to improve the quality of the manuscript.

Disclosure statement
No potential conflict of interest was reported by the authors.

References
Bhatt, R., Doelling, D.R., Wu, A., Xiong, X., Scarino, B.R., Haney, C.O., & Gopalan, A. Initial stability assessment of S-NPP VIIRS reflective solar band calibration using invariant desert and deep convective cloud targets", remote sensing. Bouvet, M. (2014). Radiometric comparison of multispectral imagers over a pseudo-invariant calibration site using a reference radiometric model. Remote Sensing of Environment, 140, 141–154. doi:10.1016/j.rse.2013.08.039
Czapla-Myers, J., McCorkel, J., Anderson, N., & Biggar, S. (2017). Earth-observing satellite intercomparison using the Radiometric Calibration Test Site at Railroad Valley. Journal of Applied Remote Sensing, 12, 1. doi:10.1117/1.JRS.12.012004

Dinguiard, M., & Slater, P.N. (1999). Calibration of Space-multipelspectral imaging sensors: A review. Remote Sensing of Environment, 68, 194–205. doi:10.1016/S0034-4257(98)00111-4

Feldman, G.C. (2003). SeaWiFS project – detailed description. Retrieved from: http://oceancolor.gsfc.nasa.gov/SeaWiFS/BACKGROUND/SEAWIFS_970_BROCHURE.html

Fougne, B., & Bach, R. (2009, March). Monitoring of radiometric sensitivity changes of space sensors using deep convective clouds: Operational application to PARASOL. IEEE Transactions On Geoscience And Remote Sensing, 47(3), doi:10.1109/TGRS.2008.2005634

Fougne, B., Lildo, J., Gross-Colzy, L., Henry, P., & Blumstein, D. (2010). Climatology of oceanic zones suitable for in-flight calibration of space sensors. Proceedings Earth Observing Systems XV, SPIE.

Frouin, R., Sei, A., Hauss, B., & Pratt, P. (2014). Operational in-flight calibration of S-NPP VIIRS in the visible using Rayleigh scattering. Proceedings of Earth Observing Systems XIX, SPIE.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., & Fernandez, V. (2017). Copernicus Sentinel-2A calibration and products validation status. 9, 584.

Henry, P., Dinguiard, M., & Bidilis, M. (1993). SPOT multitemporal calibration over stable desert areas. Proceedings of SPIE International Symposium Aerospace Remote Sensor (pp. 67–76), Orlando.

Hooker, S.B., & Mcclain, C. (2000). The calibration and validation of SeaWiFS data. Progress in oceanography. Progress in Oceanography, 45, 427–465. doi:10.1016/S0079-6611(00)00012-4

Kaufman, Y.J., & Holben, B.N. (1993). Calibration of the AVHRR visible and near-IR bands by atmospheric scattering, ocean glint and desert reflection. International Journal of Remote Sensing, 14, 21-52

Lachérade, S., Fougnie, B., Henry, P., & Gamet, P. (2013). Cross calibration over desert sites: Description, methodology and operational implementation. IEEE Transactions On Geoscience And Remote Sensing, 51(3), doi:10.1109/TGRS.2012.2227061

Lamquin, N., Bruniquel, V., & Gascon, F. (2017). Sentinel-2 L1C radiometric validation using deep convective clouds observations. European Journal of Remote Sensing, 51(1), 11–27. doi:10.1080/22797254.2017.1395713

Lee, Z.P., Du, K., Voss, K.J., Zibordi, G., Lubac, B., Arnone, R., & Weidemann, A. (2011). An inherent-optical-property-centered approach to correct the angular effects in water-leaving radiance. Applied Optics, 50(19). doi:10.1364/AO.50.003155

Lenoble, J., Herman, M., Deuzé, J.L., Lafrance, B., Santer, R., & Tans, D. (2007). A successive order of scattering code for solving the vector equation of transfer in the earth’s atmosphere with aerosols. Journal of Quantitative Spectroscopy and Radiative Transfer, 107(3), 479–507. doi:10.1016/j.jmsrt.2007.03.010

Marcq, S., Meygret, A., Bouvet, M., Fox, N., Greenwell, C., Scott, B., … Damiri, B. (2018). New RadCalNet site at Gobabeb, Namibia: Installation of the instrumentation and first satellite calibration results. Submitted to IGARSS 2018, Valencia.

Meygret, A., Santer, R., & Berthelot, B. (2011). ROSAS a Robotic Station for Atmosphere and Surface characterization dedicated to on-orbit calibration. Proceeding Of SPIE, (pp. 8153).

Rahman, H., & Dedieu, G. (1994). SMAC: A simplified method for the atmospheric correction of satellite measurements in the solar spectrum. Remote Sensing, 15(1), 123–143. doi:10.1016/0146-6811(94)90001-6

Sentinel-2 Mission Requirement Document, (2007), issue 2 revision 0, EOP-SM/1163/MR-dr

Smith, D.L., & Cox, C.V. (2013). (A)ATS solar channel on-orbit radiometric calibration. IEEE Transactions On Geoscience and Remote Sensing, 51, 3. doi:10.1109/TGRS.2012.2230333

Sterckx, S., Livers, S., & Adriaensen, S. (2013). Rayleigh, deep convective clouds, and cross-sensor desert vicarious calibration for the validation for the PROBA-V. IEEE, Transactions on Geoscience and Remote Sensing, 51, 3. doi:10.1109/TGRS.2012.2236682

Vermote, E., Tanré, D., Deuzé, J.L., Herman, M., Morcrette, J.J., & Kotchenova, S.Y. (2006). 6S User Guide Version 3, s.l.: S.n.

Vermote, E.F., Tanré, D., Deuzé, J.L., Herman, M., & Morcrette, J.J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview. IEEE Transactions On Geoscience and Remote Sensing, 35(3), 675–686. doi:10.1109/36.581987

Wang, N., Li, C., Ma, L., Liu, Y., Meng, F., Zhao, Y., … Wang, D. (2017). Ground-based automated radiometric calibration system in Bantou site, China. Proceeding of SPIE, Image and Signal Processing for Remote Sensing XXIII.

Xiong, X., Xie, X., Angal, A., Choi, J., Sun, J., & Barnes, W. L. (2007). Characterization of MODIS solar diffuser on orbit degradation. Proceedings of SPIE, (pp. 6677).