Article

Copernicus Sentinel-2A Calibration and Products Validation Status

**Ferran Gascon [1],*, Catherine Bouzinac [2], Olivier Thépaut [2], Mathieu Jung [3],
Benjamin Francesconi [4], Jérôme Louis [5], Vincent Lonjou [6], Bruno Lafrance [2], Stéphane Massera [7],
Angélique Gaudel-Vacaresse [6], Florie Languille [6], Bahjat Alhammoud [8], Françoise Viallefont [9],
Bringfried Pflug [10], Jakub Bieniarz [10], Sébastien Clerc [8], Laëtitia Pessiot [2], Thierry Trémas [6],
Enrico Cadau [1], Roberto De Bonis [1], Claudia Isola [1], Philippe Martimort [1] and Valérie Fernandez [1]**

[1]   ESA (European Space Agency), Paris 75015, France; enrico.cadau@esa.int (E.C.);
       roberto.debonis@serco.com (R.D.B.); claudia.isola@esa.int (C.I.); philippe.martimort@esa.int (P.M.);
       valerie.fernandez@esa.int (V.F.)
[2]   CS-SI (Communication & Systèmes—Systèmes d'Information), Toulouse 31506, France;
       catherine.bouzinac@c-s.fr (C.B.); olivier.thepaut@c-s.fr (O.T.); bruno.lafrance@c-s.fr (B.L.);
       laetitia.pessiot@c-s.fr (L.P.)
[3]   Airbus Defence and Space, Toulouse 31402, France; mathieu.jung@airbus.com
[4]   Thales Alenia Space, Cannes La Bocca 06156, France; benjamin.francesconi@thalesaleniaspace.com
[5]   Telespazio, Toulouse 31023, France; jerome.louis@telespazio.com
[6]   CNES (Centre National d'Etudes Spatiales), Toulouse 31401, France; vincent.lonjou@cnes.fr (V.L.);
       angelique.gaudel-vacaresse@cnes.fr (A.G.-V.); florie.languille@cnes.fr (F.L.); thierry.tremas@cnes.fr (T.T.)
[7]   IGN (Institut Géographique National) Espace, Ramonville-Saint-Agne 31520, France;
       stephane.massera@ign.fr
[8]   ARGANS, Plymouth PL6 8BX, UK; balhammoud@argans.co.uk (B.A.); sebastien.clerc@acri-st.fr (S.C.)
[9]   ONERA (Office National d'Etudes et Recherches Aérospatiales), Toulouse 31055, France;
       francoise.viallefont@onera.fr
[10]   DLR (Deutschen Zentrums für Luft- und Raumfahrt), Berlin 12489, Germany; bringfried.pflug@dlr.de (B.P.);
       jakub.bieniarz@dlr.de (J.B.)
*   Correspondence: ferran.gascon@esa.int; Tel.: +39-06-941-88605

Abstract: As part of the Copernicus programme of the European Commission (EC), the European Space Agency (ESA) has developed and is currently operating the Sentinel-2 mission that is acquiring high spatial resolution optical imagery. This article provides a description of the calibration activities and the status of the mission products validation activities after one year in orbit. Measured performances, from the validation activities, cover both Top-Of-Atmosphere (TOA) and Bottom-Of-Atmosphere (BOA) products. The presented results show the good quality of the mission products both in terms of radiometry and geometry and provide an overview on next mission steps related to data quality aspects.

Keywords: calibration; validation; optical; instrument; processing; imagery; spatial; operational

1. Introduction

As part of the Copernicus programme of the European Commission (EC), the European Space Agency (ESA) has developed and is currently operating the Sentinel-2 mission acquiring high spatial resolution (10 to 60 m) optical imagery [1]. The Sentinel-2 mission provides enhanced continuity to services monitoring global terrestrial surfaces and coastal waters.

The Sentinel-2 mission offers an unprecedented combination of systematic global coverage of land and coastal areas, a high revisit of five days under the same viewing conditions, high spatial resolution, and a wide field of view (295 km) for multispectral observations from 13 bands in the visible, near infrared and short wave infrared range of the electromagnetic spectrum.

Frequent revisits of five days at the equator require two identical Sentinel-2 satellites (called Sentinel-2A and Sentinel-2B units) and each one carrying a single imaging payload named MSI (Multi-Spectral Instrument). The orbit is Sun-synchronous at 786 km altitude (14 + 3/10 revolutions per day) with a 10:30 A.M. descending node. This local time was selected as the best compromise between minimizing cloud cover and ensuring suitable sun illumination. It is also aligned with other similar satellites, e.g., Landsat. An overview of the MSI imaging payload is provided in the following section.

The Sentinel-2 satellites will systematically acquire observations over land and coastal areas from $-56°$ to $84°$ latitude including islands larger 100 km^2, EU islands, all other islands less than 20 km from the coastline, the whole Mediterranean Sea, all inland water bodies and all closed seas. Over specific calibration sites, for example DOME-C in Antarctica, additional observations will be made. The two satellite units will work on opposite sides of the orbit. Sentinel-2A launch took place in June 2015 and Sentinel-2B is foreseen beginning 2017. Therefore, this paper focuses only on the performances achieved by Sentinel-2A.

The availability of products with good data quality performances (both in terms of radiometry and geometry accuracies) has a paramount importance for many applications. This is indeed a key enabling factor for an easier exploitation of time-series, inter-comparison of measurements from different sensors or detection of changes in the landscape.

Calibration and validation (Cal/Val) corresponds to the process of updating and validating on-board and on-ground configuration parameters and algorithms to ensure that the product data quality requirements are met.

This paper provides a description of the calibration activities one year after Sentinel-2A launch, of the mission products validation activities. Only Sentinel-2A data were available at the manuscript-writing time. Performances derived from the validation activities have been estimated for both Top-Of-Atmosphere (TOA) and Bottom-Of-Atmosphere (BOA) products (referred respectively as Level-1 and Level-2A and further described in this paper).

2. Multi-Spectral Instrument Overview

This section provides a brief overview of Sentinel-2 Multi-Spectral Instrument (MSI). It aims at giving to the reader the basis required to fully understand the measured performances and the Calibration and Validation (Cal/Val) approach.

2.1. MSI Design

The MSI instrument design has been driven by the large swath requirement together with the demanding geometrical and spectral performances of the measurements. It is based on a push-broom concept, featuring a Three-Mirror Anastigmatic (TMA) telescope feeding two focal planes spectrally separated by a dichroic filter, as shown in Figure 1. On the left part of this figure, the yellow part is a Sun diffuser used for radiometric calibration. One focal plane includes the Visible and Near-Infrared (VNIR) bands and the other one the Short-Wave Infrared (SWIR) bands.
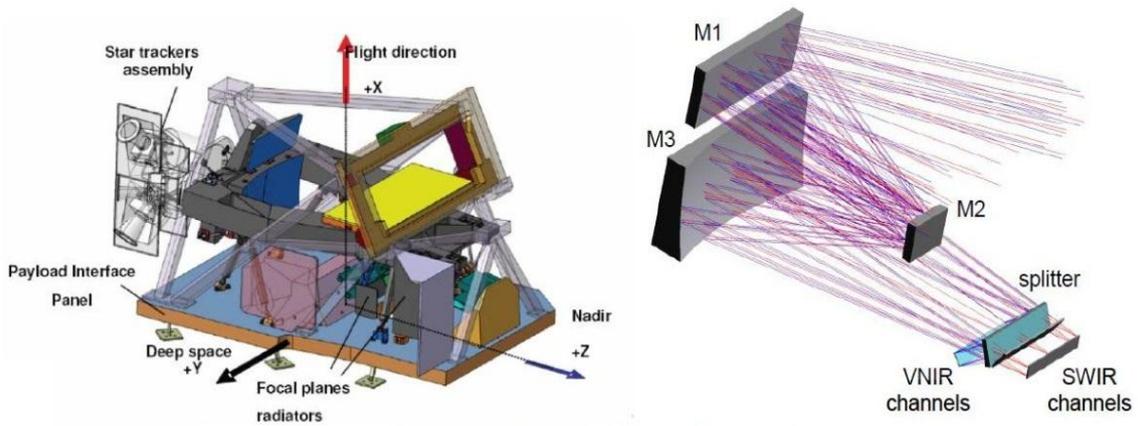
Figure 1. Multi-Spectral Instrument (MSI) internal configuration. **Left**: full instrument view (diffuser panel in yellow, telescope mirrors in dark blue). **Right**: optical path construction to the Short-Wave Infrared (SWIR)/visible to near-infrared (VNIR) (see Section 2.2) splitter and focal planes.

2.2. Spectral Bands and Resolution

The Sentinel-2 MSI is a filter-based push-broom imager. It performs measurements in 13 spectral bands spread over the VNIR and SWIR domains with spatial resolutions ranging from 10 to 60 m. These spectral channels include:

- 4 bands at 10 m spatial resolution: blue (490 nm), green (560 nm), red (665 nm) and near infrared (842 nm).
- 6 bands at 20 m spatial resolution: 4 narrow bands mainly used for vegetation characterization in the red edge (705 nm, 740 nm, 783 nm and 865 nm) and 2 wider SWIR bands (1610 nm and 2190 nm) for applications such as snow/ice/cloud detection or vegetation moisture stress assessment.
- 3 bands at 60 m spatial resolution for applications such as cloud screening and atmospheric corrections (443 nm for aerosols, 945 nm for water vapour and 1375 nm for cirrus detection).

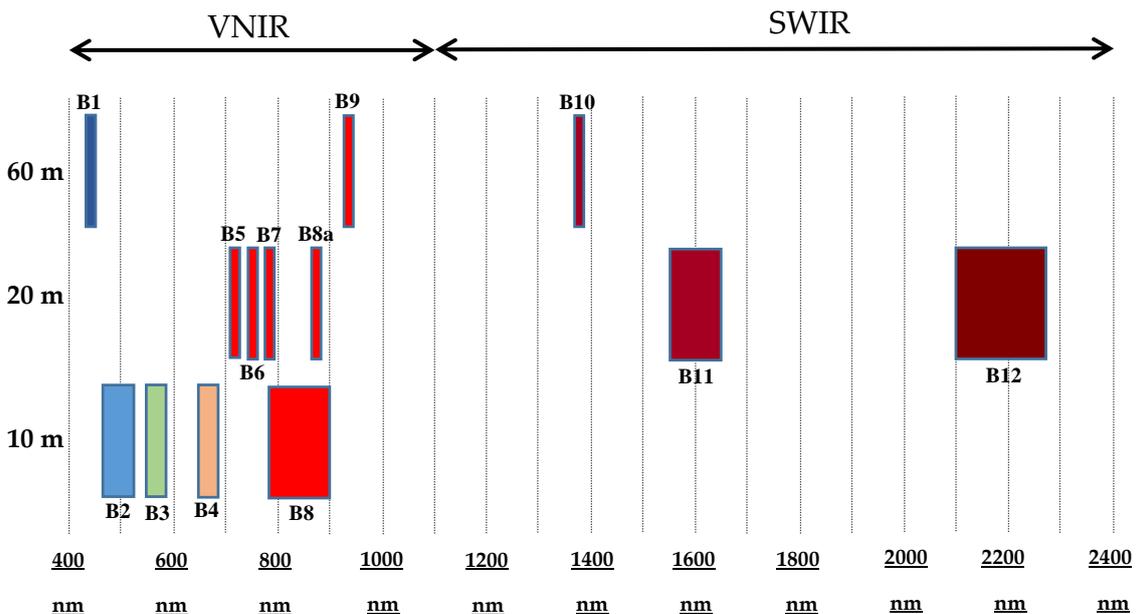The specified spectral band characteristics and resolutions are summarised in Figure 2.

Figure 2. MSI spectral bands vs. spatial resolution with corresponding Full Width at Half Maximum (FWHM).

2.3. Focal Plane Layout—Pixels Line of Sight

Due to different thermal regulation constraints, VNIR and SWIR bands are separated into two distinct focal planes of 12 detector modules each.

The 12 detector modules on each focal plane are stagger-mounted to cover altogether the 20.6° instrument field-of-view resulting in a compound swath width of 295 km on the ground across track at the satellite reference altitude of 786 km.

As illustrated in Figure 3, due to this staggered design of the detectors on the focal planes (and due to the position of the mirrors of the spectral bands between two successive detectors), a parallax angle between the odd and even clusters of detectors is induced on the measurements, resulting in an inter-detector shift along track from 14 km to 48 km at maximum. Furthermore, two successive detectors share an overlap area of 2 km across track.
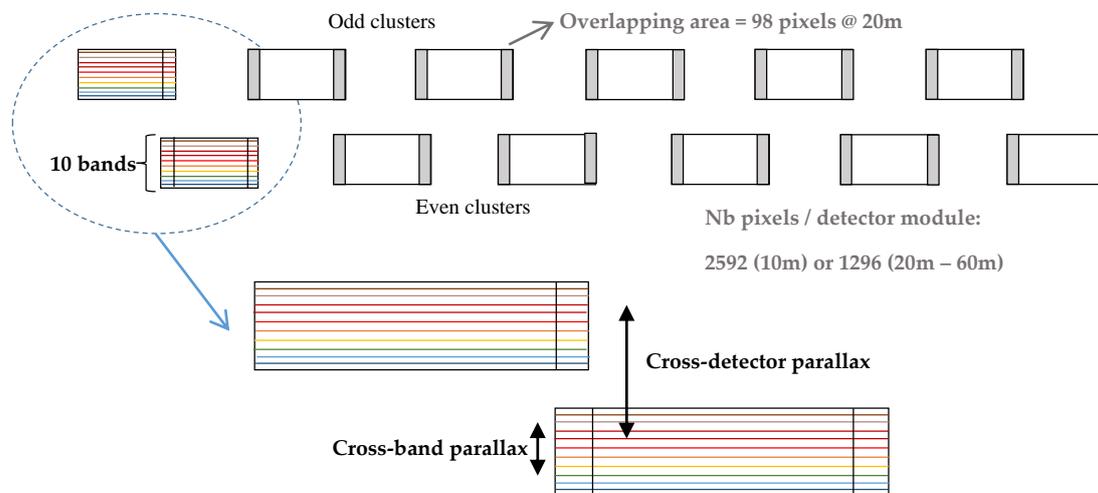
Figure 3. Staggered Detector Configuration.

Likewise, the hardware design of both the VNIR and SWIR detectors induces a relative displacement of each spectral channel sensor within the detector, resulting in an inter-band measurement parallax amounting to a maximum along-track displacement of about 17 km between B10 and B12.

As a consequence of this layout, the landscape is acquired with different viewing angles from one detector to another. This can result in radiometric differences, due to the anisotropy of the atmosphere or surface and as a function of the viewing and sun directions.

For each pixel "p" of each detector module, the Line of Sight (LOS) is given in Equation (1) by a vector $\vec{Z}(p)$ identified by angles $\psi_X(p)$ and $\psi_Y(p)$ expressed in a frame linked to the MSI (X_{LOS}, Y_{LOS}, Z_{LOS}). Figure 4 illustrates the LOS defined by Equation (1).

$$\vec{Z}(p) = \frac{1}{\sqrt{1 + \tan\left(\psi_X(p)\right)^2 + \tan\left(\psi_Y(p)\right)^2}} \begin{bmatrix} \tan(\psi_Y(p)) \\ -\tan(\psi_X(p)) \\ 1 \end{bmatrix} \tag{1}$$
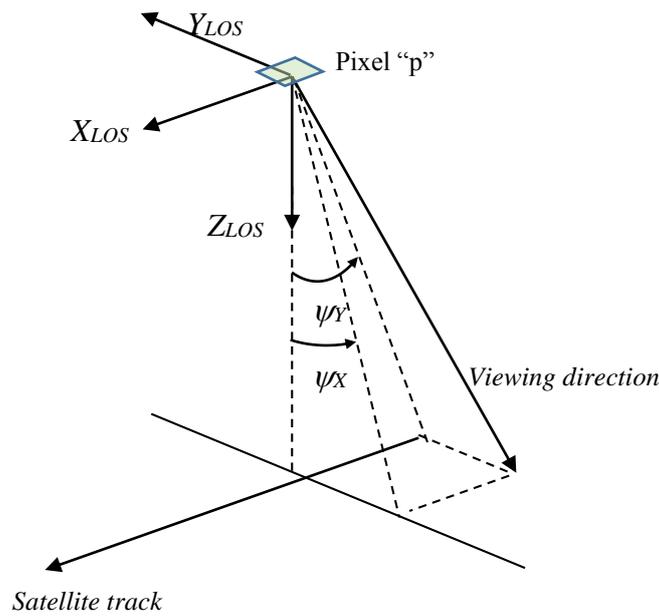
Figure 4. Line of Sight angles definition.

This expression precisely defines the Viewing Angle of each pixel and then the localization of this pixel on ground, when considering the satellite orbital position, the sampling date, and the model of the earth including a Digital Elevation Model. Figure 5 simply presents a flat ground projection of the pixel LOS for the whole focal plane as well as the materialization of the MSI frame (X_{LOS}, Y_{LOS}, Z_{LOS}).
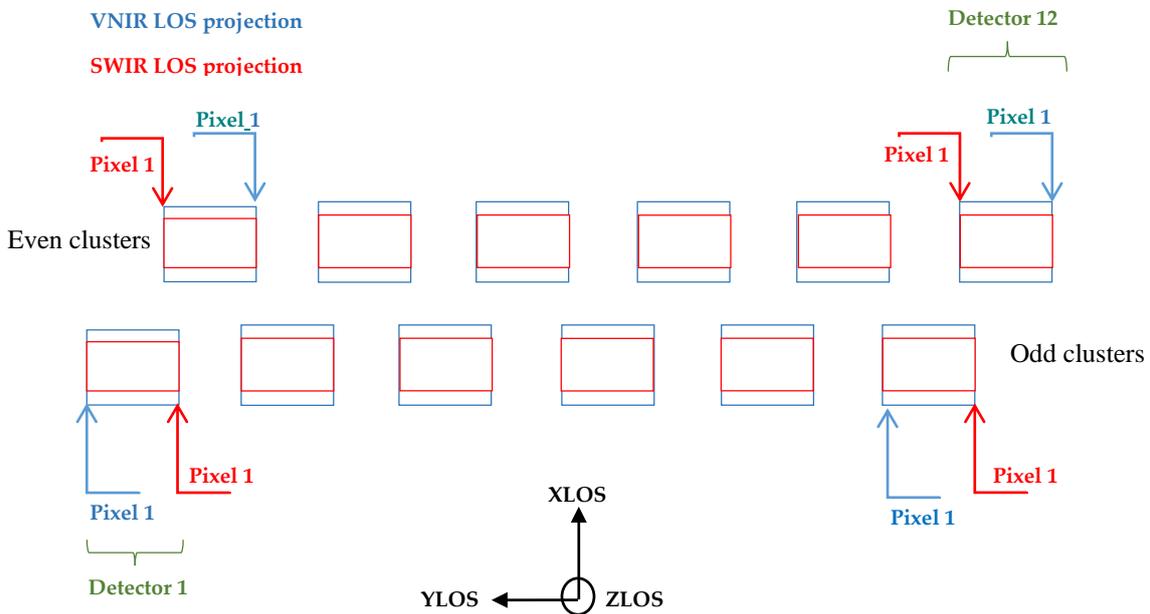

Figure 5. Representation of the pixels Line-Of-Sight (LOS) projection on ground.

2.4. Detector Specificities

The VNIR assembly (built by Astrium-E2V) is based on monolithic Silicon CMOS technology [2] and consists of 10 spectral bands integrated on a single detector. The SWIR detector is based on hybrid HgCdTe-CMOS technology where the Hg1-xCdxTe or Mercury Cadmium Telluride (MCT) is

hybridized to a silicon readout circuit ROIC. The SWIR assembly has 3 spectral bands combined on a single detector respectively [3].

The photosensitive MCT material and the MCT-CMOS technology make the SWIR bands more sensitive to radiation and variations of temperature [4] than the VNIR bands. This may result to a slightly larger dark signal variation or evolution for the SWIR bands comparing to the VNIR ones. To achieve a high radiometric performance, the SWIR detectors are cooled down to 190 K. For each SWIR spectral band, several lines of pixels of $15 \times 15 \ \mu m^2$ have been designed and allow an optimal pixel configuration by the selection of the best responding pixel(s) per column and per band [5].

The number of pixels in each detector module is 1296 or 2592, depending on the spectral band. 60 m bands are acquired with native resolution of 20 m across track and binning is applied on ground. The sampling time is also depending on the band. Table 1 synthesizes these instrument characteristics.

Table 1. Detector Configuration.

Band	Resolution (m)	Integration Time (ms)	Number of Detector Lines	Number of Lines after Selection	Number of Pixels per Detector	X-Size (across-track) (μm)	Y-Size (along-track) (μm)
B01	60	9.396	1	1	1296	15	45
B02	10	1.566	1	1	2592	7.5	7.5
B03	10	1.566	2	2	2592	7.5	7.5
B04	10	1.566	2	2	2592	7.5	7.5
B05	20	3.132	1	1	1296	15	15
B06	20	3.132	1	1	1296	15	15
B07	20	3.132	1	1	1296	15	15
B08	10	1.566	1	1	2592	7.5	7.5
B8A	20	3.132	1	1	1296	15	15
B09	60	9.396	1	1	1296	15	45
B10	60	9.396	3	1	1296	15	15
B11	20	3.132	4	2	1296	15	15
B12	20	3.132	4	2	1296	15	15

As shown in Table 1, SWIR spectral bands are acquired on several detector lines thanks to a Time Delay Integration (TDI) technology. This allows increasing the Signal-to-Noise Ratio (SNR), on the one hand, but on the other hand some reconfiguration is needed, as explained hereafter.

Indeed SWIR bands detectors are more sensitive to noise and to ageing. For this reason, the detector provides more TDI lines than the nominal number used and some rearrangements are possible during the mission lifespan. Figure 6 shows the possible configurations.

SWIR pixel health is monitored all along satellite lifetime and TDI selection rearrangement is performed when required (cf. Section 4.1.4). Any change of the TDI configuration has an impact on the pixel Line-Of-Sight (LOS) that shall also be taken into account by the ground processing system.
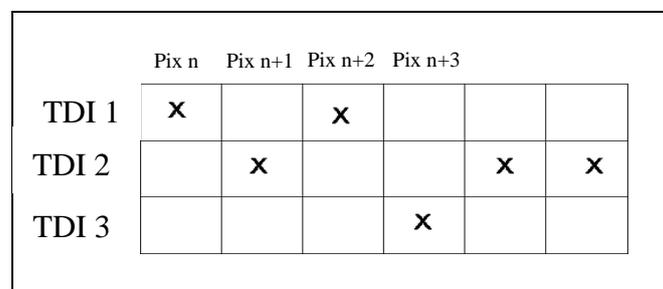

Figure 6. *Cont.*

	Pix n	Pix n+1	Pix n+2			
TDI 1	✗		✗		✗	
TDI 2	✗	✗	✗		✗	✗
TDI 3		✗		✗		✗
TDI 4				✗		

Figure 6. Time Delay Integration (TDI) configuration examples for pixels of SWIR bands. Top figure corresponds to B10 configuration, Bottom figure corresponds to B11 and B12 configuration. Each column corresponds to one pixel. The lines correspond to available TDI lines. A cross corresponds to the TDI line selected for each pixel. For B10, 3 TDI lines are available for only 1 line selected per pixel. Therefore 3 configurations per pixel are possible. The actual configuration may be different from one pixel to the other. For B11 and B12, 4 TDI lines are available, and 2 consecutive lines are used for each pixel. Therefore, 3 configurations per pixel are possible.

2.5. On-Board Equalization and Compression

Images are compressed on board to reduce to volume of data to be downlinked to ground. Before compression, they are roughly equalized to minimize the signal entropy and improve the compression quality at a given rate. The equalization consists in applying a two parts piecewise linear model to each pixel response in order to compensate for the non-uniformity between detectors. The component in charge of these operations is called WICOM (Wavelet Image Compression and Memory).

Detailed description of the WICOM is complex and not interesting in the frame of this paper. Nevertheless, one must be aware of the compression principle as well as the compression rates in order to understand the potential impact on image quality.

WICOM is a high performance image compression module that implements a wavelet image compression algorithm close to JPEG2000 standard consisting of a discrete wavelet transform (DWT). Before compression, a Non Uniformity Correction (NUC) function is applied to the 12 bits on-board image. This function consists of a 2 parts linear function, reversible on ground. Of course, saturated pixels are detected before NUC computation and, for such pixels, the function is not applied.

NUC function as well as compression function can be bypassed, for example in case of calibration images.

The nominal compression ratios and data rates are given in Table 2.

Table 2. Nominal Compression Ratios.

	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
Comp. ratio	2.4	3.33	3.19	3	3.13	3.13	2.86	3	3.13	2.14	2.65	2	2.4
Bits per pixel	5	3.6	3.76	4	3.84	3.84	4.2	4	3.84	5.6	4.52	6	5

As shown in Table 2, the compression ratios are low enough so that the image quality is preserved with little loss.

2.6. On-Board Sun Diffuser

A full-field (or full-pupil) on-board diffuser is used to perform the radiometric calibration and to guarantee a high quality radiometric performance (Figure 7). The diffuser is mounted on the so-called Calibration and Shutter Mechanism (CSM). This CSM is designed to collect the sunlight after reflection by a diffuser, and to prevent a direct view of the Sun or contamination during launch and early operation phase (LEOP).
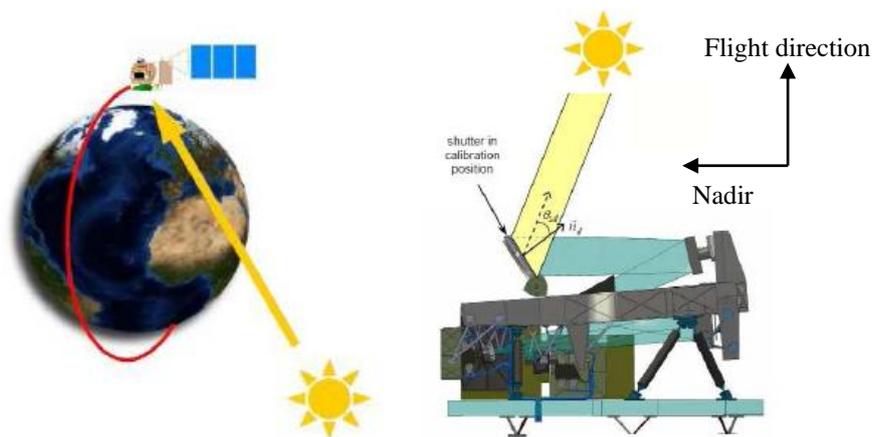
Figure 7. Illustration of the sun diffuser acquisition principle.

The advantage of such on-board device is to provide the instrument with a very uniform and well-known signal, allowing very accurate absolute and relative radiometric calibrations.

This CSM is periodically used (typically every month) to assess the radiometric performance and to compute new calibration parameters (cf. Section 4.1.2).

3. Products Overview

3.1. Processing Levels

Sentinel-2 MSI products comprise the following levels of processing:

- Level-0 (L0) products: Raw instrument data packaged for long-term storage and future reprocessing campaigns. A coarse cloud mask is appended for cataloguing purposes.
- Level-1A (L1A) products: Uncompressed instrument data in sensor geometry and with a coarse registration (i.e., rough pixel alignment between images from different spectral bands and detector modules). No resampling nor radiometric corrections have been applied. These products are used for calibration purposes.
- Level-1B (L1B) products: Top-Of-Atmosphere (TOA) radiances in sensor geometry (same as Level-1A products). Full radiometric corrections have been applied. These products are used for calibration, validation and quality control purposes.
- Level-1C (L1C) products: TOA reflectances in cartographic geometry. These products are publicly disseminated by ESA.
- Level-2A (L2A) products: Bottom-Of-Atmosphere (BOA) reflectances in cartographic geometry (same as Level-1C products). Today, these products can be generated by users with the publicly available Sen2Cor processor. L2A production is now systematic over Europe and dissemination through the Copernicus Open Access Hub [6] started in May 2017.

3.2. Level-1 Processing Steps

The main steps of the processing chain are summarized in Figure 8.
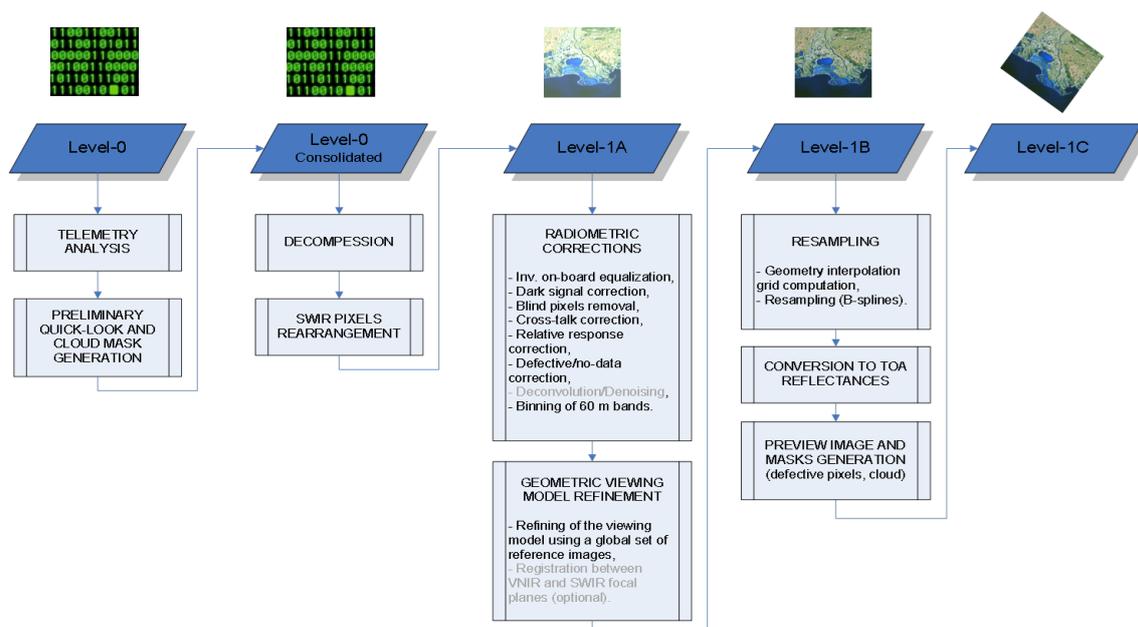
Figure 8. MSI processing chain overview. The gray font indicate optional steps implemented to mitigate risk on instrument performance but currently not activated.

After formatting and decompression, the first processing step for Level-1B is the radiometric correction, which aims at converting instrument counts into physical units (radiances). First, the coarse on-board equalization performed to improve compression levels is inverted to recover raw measurement data. Then, the fine radiometric model using instrument calibration coefficients is applied (including dark signal removal and pixel response non-uniformity correction). A linear correction is applied on SWIR bands to remove the effects of the electronic cross-talk. Blind pixels at the edges of the detector are removed, and defective pixels are interpolated. SWIR pixels are re-aligned consistently with the on-board pixel selection (see Section 2.4). The processing chain implements a contingency deconvolution step to improve image quality. This option has not been activated for Sentinel-2A as the image quality is compliant with pre-launch expectations. Finally, some radiometric quality masks are generated (for saturated and defective pixels). Binning from 20 m to 60 m resolution is applied on the corresponding spectral bands to improve signal to noise ratio.

After application of the radiometric correction, geometric refining is applied. This algorithm aims at improving the geolocation performance and especially the multi-temporal geometric stability. It will be activated after the completion of the Global Reference Image (GRI) currently planned for end 2017. The processing involves a smoothing of the instrument viewing model (attitude and position) using the set of accurately localised ground control points of the GRI for the reference band.

Level-1C processing starts with the reprojection of the images in the cartographic reference frame. The processing is performed for each L1C tile intersecting the instrument swath. Each pixel is projected from the cartographic grid onto the instrument frame, using the viewing model appended to the Level 1B product and the reference Digital Elevation Model (DEM) at 90 m spatial sampling. The pixel value for each spectral band is interpolated using B-splines.

Conversion from radiances to reflectances relies on the Solar Irradiance model of [7].

Finally, quality masks -derived from the L1B masks- as well as opaque and cirrus clouds masks are generated. The opaque cloud mask is based on radiometric thresholds on B01 and B12, while cirrus clouds are detected with band B10. Viewing and solar angles, as well as meteorological data are computed on a coarse grid for each L1C tile and appended as metadata or auxiliary data.

3.3. Level-2A Processing Steps

The main processing steps of the Sen2Cor processor are presented in Figure 9. Level-2A processing is applied to granules of Top-Of-Atmosphere (TOA) Level-1C orthorectified reflectance products. The processing starts with the Cloud Detection and Scene Classification followed by the retrieval of the Aerosol Optical Thickness (AOT) and the Water Vapour (WV) content from the Level-1C image. The final step is the TOA to Bottom-Of-Atmosphere (BOA) conversion. Sen2Cor also includes several options that can be activated like cirrus correction, terrain correction, adjacency correction and empirical Bidirectional Reflectance Distribution Function (BRDF) corrections.
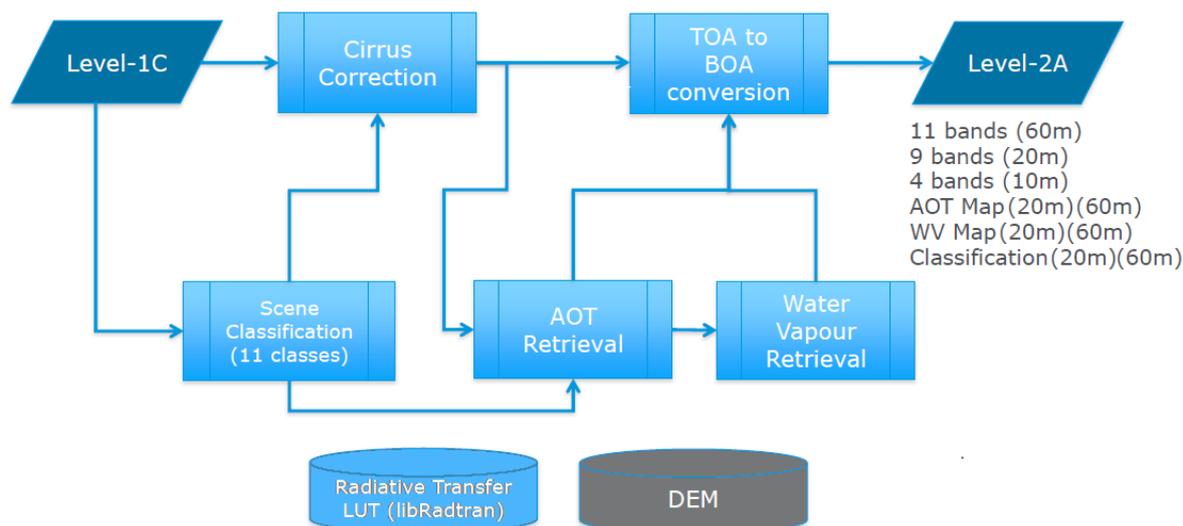

Figure 9. Level-2A processing chain overview.

Sen2Cor relies on two main auxiliary data sets: the Radiative Transfer Look-Up Tables (LUTs) and the Digital Elevation Model (DEM). The latter model is not embedded in the Sen2Cor software but the default DEM (Shuttle Radar Topography Mission Digital Elevation Model at 90 m spatial resolution) is downloaded automatically if requested. Otherwise, the user can add any other DEM in DTED format. Level-2A outputs are the surface (or BOA) reflectance images which are provided at different spatial resolutions (60 m, 20 m and 10 m) together with AOT and WV maps. The internal Scene Classification (SCL) image used for atmospheric correction is also provided together with Quality Indicators for cloud and snow probabilities.

3.4. Level-1C Product Description

L1C products provide TOA normalized reflectances for each spectral band, coded as integers on 15 bits. The physical values range from 1 (minimum reflectance 10-4) to 10000 (reflectance 1), but values higher than 1 can be observed in some cases due to specific angular reflectivity effects. The value 0 is reserved for "No Data".

The computation of the reflectance relies on a solar irradiance model and the cosine of the Sun zenith angle with respect to the reference Earth ellipsoid. Both data are available in the product metadata.

The reflectance values are provided in Universal Transverse Mercator (UTM) projection. The Sentinel-2 Level-1C granules, also called "tiles", are based on the Military Grid Reference System (MGRS) tiles with a 5 km extension on all sides to ensure an overlap with neighbouring tiles. The 110×110 km^2 footprint of each tile can be obtained in a Keyhole Mark-up Language (KML) file from Sentinel-2 internet site [8]. The projection is computed using a global Digital Elevation Model (DEM) with approximately 90 m horizontal resolution and 10 m of elevation uncertainty.

Reflectance images are provided with a spatial resolution of 10, 20 or 60 m depending on the spectral bands. Sixty meter images are obtained by spatial binning of measurements sensed at 20 m resolution as recalled in Section 2.4.

Sentinel-2 L1C products take the form of folders gathering geolocalised JPEG2000 images for each spectral band, metadata in xml format and auxiliary data. Images are gathered in folders for each granule, together with quality masks in Geographic Mark-up Language (GML) format and tile metadata. Granule masks include an opaque cloud and cirrus mask, and detector footprint masks for each spectral bands. Viewing angles (zenith/azimuth with respect to North direction) for each spectral band as well as Sun illumination angles are provided on a 5 km grid.

Metadata are provided for the data strip (elementary processing unit) and user product (dissemination unit). data strip metadata provide information on processing configuration and parameters, and reports from automatic quality checks. The user product metadata summarizes information about the instrument spectral and radiometric response and processing parameters.

Finally, Sentinel-2 products include auxiliary data such as meteorological data (systematic distribution) or processing parameter files (specifically added for expertise).

3.5. Level-2A Products Description

Level-2A surface reflectance products are generated with Sen2Cor processor whose main purpose is to correct Sentinel-2 Level-1C products from the effects of the atmosphere. Level-2A processing is applied independently to single granules of TOA Level-1C orthorectified reflectance products. Note that this independent processing of neighbouring granules can lead to visible steps between adjacent granules. Sen2Cor processor is available as a third-party plugin of the Sentinel-2 Toolbox [9].

L2A products provide:

- Surface (or BOA) reflectance images which are provided at different spatial resolutions (60 m, 20 m and 10 m);
- Aerosol Optical Thickness (AOT) and Water Vapour (WV) maps (60 m, 20 m and 10 m);
- Scene Classification (SCL) map used internally as input for atmospheric correction together with Quality Indicators for cloud and snow probabilities (60 m and 20 m).

The structure of the Level-2A product [10] is strictly based on the structure of the Level-1C product. The same tiling geometry and projection are used. The main difference is that the IMG_DATA folder contains three directories: one for each resolution at 60 m, 20 m, and 10 m. The Scene Classification map is available at 20 m or 60 m resolution and the cloud and snow probabilities are located in the QI_DATA folder.

The Scene Classification module provides a Scene Classification map divided in 11 classes presented in chapter 5. This map does not constitute a land cover classification map in a strict sense, its main purpose is to be used internally in Sen2Cor in the atmospheric correction module to distinguish between cloudy pixels, clear pixels and water pixels. Two quality indicators are also provided: a Cloud confidence map and a Snow confidence map with values ranging from 0 to 100%.

The Water Vapour content is retrieved from Level-1C image using a Sentinel-2 adapted APDA (Atmospheric Pre-corrected Differential Absorption) algorithm [11] which uses a ratio between band B8A and band B09. The quantification value to convert Digital Numbers to Water Vapour column in cm is equal to 1000.

The Aerosol Optical Thickness (AOT) at 550 nm is estimated using the Dark Dense Vegetation (DDV) pixel method introduced by [12]. AOT estimation fails when there are no DDV pixels in the image. The fallback solution for that case is to perform the atmospheric correction with a constant AOT which is specified by the start visibility (VIS) set in the configuration file. Default value of start visibility is 40 km which corresponds to an AOT at 550 nm of 0.2 at sea level. The quantification value to convert Digital Numbers to AOT is equal to 1000.

The actual Sen2Cor retrieval method is described in [13]. The method was developed for application over land surface. It includes several options that can be activated like cirrus correction, terrain correction, adjacency correction and empirical Bidirectional Reflectance Distribution Function (BRDF) corrections. Sen2Cor can be also applied over water surface using the AOT estimated over land pixels in the image. However the processor contains no consideration of water surface effects like sun glint.

The surface reflectance values are coded in JPEG2000 with the same quantification value of 10,000 as for Level-1C products, i.e., a factor of 1/10,000 needs to be applied to Level-2A digital numbers (DN) to retrieve physical surface reflectance values.

In the R10m directory of the IMG_DATA folder are located the surface reflectance images with 10 m spatial resolution for the following bands: B02, B03, B04 and B08.

In the R20m directory, the 20 m resampled bands B02, B03 and B04 are provided together with the following surface reflectance bands: the "red edge" bands B05, B06, B07, the NIR band B8A and the two SWIR bands B11 and B12. Note that the band B08 is not provided resampled at 20 m because the band B8A at 20 m native resolution provides a narrower spectral width more suitable for vegetation monitoring applications.

In the R60m directory, the band B01 and the band B09 are provided at their 60 m native resolution together with all other bands at the exception of bands B10 and B08. Band B10 is not provided in surface reflectance as it does not provide information on the surface and band B08 is replaced by band B8A which provides a narrower spectral width.

3.6. Processing Baseline Evolutions

The public archive of Sentinel-2 product was opened with processing baseline 02.00. The processing baseline evolves to take into account corrections of errors in the processing chain and introductions of new product features or processing steps (e.g., introduction of the geometric refinement).

On the other hand, evolution of calibration coefficient does not lead to a change of version. The list of auxiliary files used for processing can be traced in the data-strip metadata of the products.

The evolutions of the processing baseline are tracked and justified in the Data Quality Report published on a monthly basis [14]. This document is available from the Sentinel-2 on-line technical guide.

4. Level-1 Calibration and Validation Status

Calibration and Validation (Cal/Val) correspond to the process of updating and validating on-board and on-ground configuration parameters and algorithms to ensure that the product data quality requirements are met. These activities are performed within the frame of the Mission Performance Centre (MPC). These operations are achieved through products processing and analyses at different processing levels.

This section provides an overview of the Cal/Val activities that have been performed on Sentinel-2A unit.

4.1. Radiometry Calibration Activities

The radiometric calibration activities allow determination of Ground Image Processing Parameters (GIPPs) of the radiometric calibration model, which aims at converting the electrical signal measured by the instrument, transformed in digital counts, into the physical incoming radiance enlightening the sensor. The nominal calibrations are based on the exploitation of the on-board sun diffuser images (relative gains calibration, absolute radiometric calibration) or images acquired over ocean at night (for dark signal calibration). Some other calibrations are foreseen only in case of contingency (crosstalk, refocusing).

At the moment, the radiometric calibration is very close to what was expected before launch and the discrepancies observed are weak and under control. In particular, SWIR focal plane is subject to predictable electronic crosstalk which is corrected by a ground post-processing.

4.1.1. Dark Signal Calibration

The dark signal is the signal delivered by the sensor when it is not illuminated by any light. Its measurement is realized by processing images acquired by night over ocean to avoid light coming from human activities. Such acquisitions were performed several times per week up to mid of May 2016 (during the commissioning phase and the beginning of the operational phase). Currently, dark signal acquisitions are performed at least once per 10-day orbit cycle.

Starting from Level-0 products, the images are uncompressed and reversed from the on-board processing to retrieve the native values of the measurements. Dark acquisitions last 40 s. The average of digital counts over acquisition lines provides the dark signal coefficients for each pixel of each detector for each band. The standard deviation of the dark signal over the column gives an estimate of the dark signal noise, expressed in digital counts (DC) or Least Significant Bit (LSB).

The MPC monitors the time evolution of the dark signal. No significant dark signal variation has been noticed since the sensor is in orbit. The dark current is particularly stable for VNIR bands: the variation of the measured dark coefficients between two acquisitions is always smaller than 1 digit (for extreme variations) and the dark signal noise is about 0.5 DC. Table 3 shows the statistics of the differences between dark images obtained on 9 May and 8 June 2016. For SWIR bands, the variations of the dark signal can be larger, mainly for the B12 band for which the maximal variations can reach about 5 digital counts for a few pixels, and the pixel relative variation is noticeable (the standard deviation of variations over pixels is usually about ten times larger for B12 than for VNIR bands). The dark noise for SWIR bands is also slightly larger than for VNIR bands, with mean values of about 0.7 to 1 digital count. SWIR detectors are known to be more sensitive to variations than VNIR ones due to a difference in detector technology (see Section 2.4).

Table 3. On the left, statistics on dark signal variations between the dark estimate on June 8, 2016 and the operational dark coefficients deduced from an acquisition on May 9, 2016 (i.e., difference of the mean dark signal measured for the two acquisitions). On the right, statistics on the dark noise for estimate on June 8, 2016 (standard deviation of the dark signal measured by pixel). Measurements are expressed in digital counts (LSB).

Dark Signal Variation Relative to the Operational GIPPs on 9 May 2016					Dark Noise (RMS in LSB)				
08/06/2016	Min	Mean	Max	Std	08/06/2016	Min	Mean	Max	Std
B01	−0.1	0.0	0.1	0.02	B01	0.33	0.51	0.78	0.05
B02	−0.1	0.0	0.1	0.02	B02	0.37	0.51	2.41	0.06
B03	−0.1	0.0	0.2	0.02	B03	0.21	0.42	0.75	0.08
B04	−0.1	0.0	0.3	0.02	B04	0.22	0.43	0.66	0.08
B05	−0.1	0.0	0.2	0.02	B05	0.37	0.53	0.75	0.05
B06	−0.2	0.0	0.2	0.02	B06	0.36	0.52	0.77	0.05
B07	−0.1	0.0	0.2	0.03	B07	0.35	0.53	0.83	0.05
B08	−0.1	0.0	0.3	0.02	B08	0.36	0.52	0.86	0.05
B8A	−0.1	0.0	0.6	0.03	B8A	0.29	0.53	0.77	0.05
B09	−0.2	0.0	0.7	0.03	B09	0.29	0.50	0.75	0.06
B10	−1.0	0.0	1.1	0.12	B10	0.91	1.06	1.37	0.03
B11	−0.9	0.0	0.5	0.08	B11	0.54	0.67	0.82	0.03
B12	−5.0	0.0	5.5	0.34	B12	0.55	0.73	2.02	0.05

Figures 10 and 11, respectively for the B01 and B12 bands, give a global overview of the dark signal versus the pixel number, its variation between the two dates, and its noise (the successive detectors are agglomerated for these illustrations). They illustrate both the very good time stability of

the dark signal and the weak inter-pixel variations for VNIR bands, while the inter-pixel variations appear larger for SWIR bands (mainly for B12 as plotted on Figure 11).
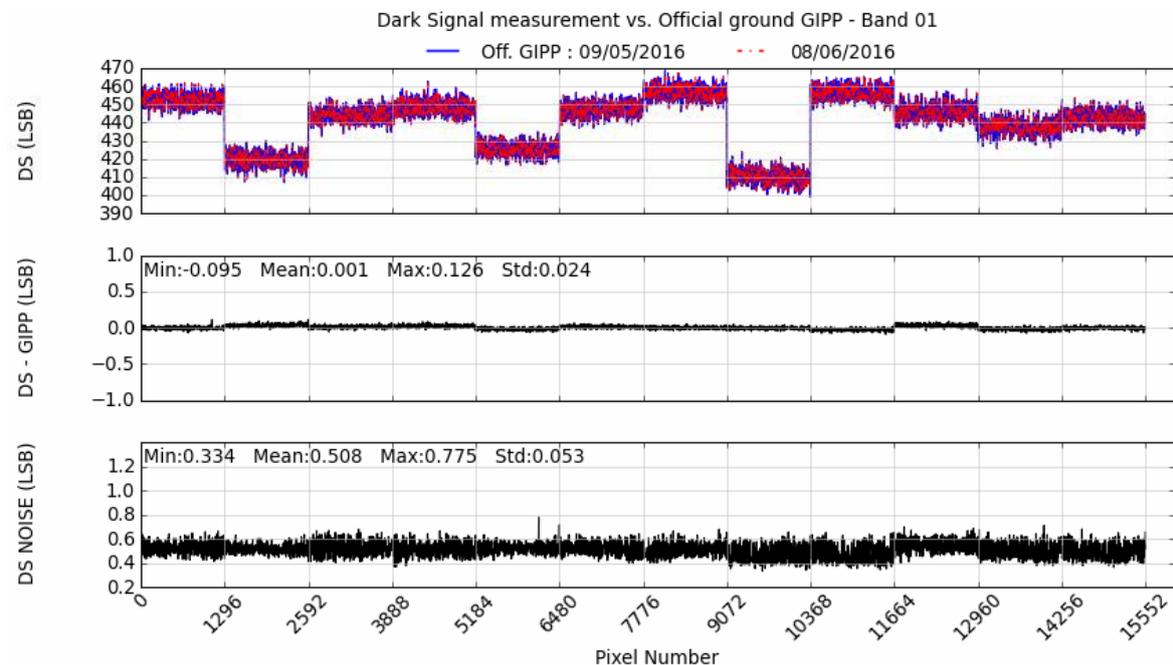

Figure 10. Variation of the dark signal measured on 8 June 2016, for the B01 band, with respect to the coefficients calculated from dark acquisition on 9 May 2016. The first plot shows the dark signal level (in digital counts, LSB) as a function of the pixel number, the second plot shows its variation between the two dates, the third plot shows its noise (in LSB).
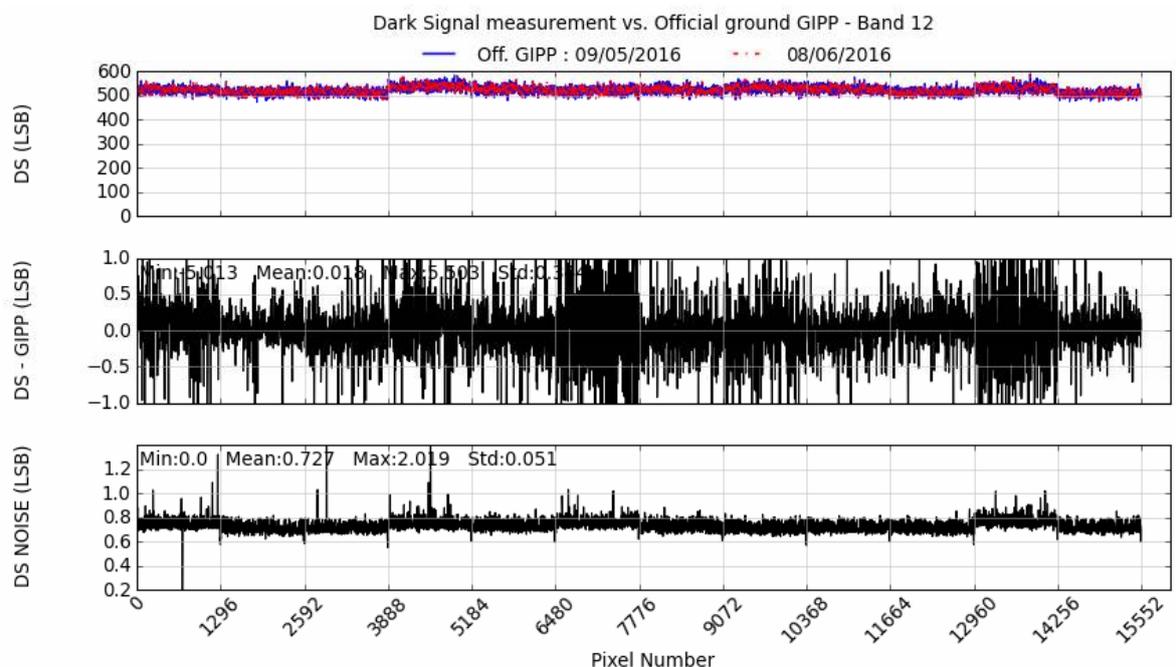
Figure 11. Same as Figure 10 but for the B12 band.

4.1.2. Absolute Radiometric Calibration

The Sun-diffuser calibration aims to assess the absolute calibration coefficients and the relative gain coefficients (or equalisation coefficients). The method relies on the comparison of the Sun-diffuser acquisition measurement to a simulation of the reflected radiance.

The simulation takes into account the solar irradiance corresponding to each acquisition line ℓ, i.e., the mean solar irradiance (E_{sun}) corrected for the Earth to Sun distance (d_{sun}) and for the incidence angle (θ_{sun}) of the solar beam incoming on the Sun-diffuser for each line (as there is a slight time variation depending on the acquisition line). The model of the reflexion of the solar light on the Sun-diffuser considers the non-uniformity and bidirectional reflectivity of the diffuser from a pre-launch characterisation of its BRDF. This one gives for each pixel the diffuser reflectance as a function of the incident direction of the solar light (solar zenith angle θ_{sun} and solar azimuth angle φ_{sun}), accounting for a constant viewing direction for a given pixel p. This leads to a simulation of the radiance illuminating the MSI sensor which is given, for each band b and each detector d, by:

$$Z_{simu}(b,d,\ell,p) = \frac{E_{sun}(b) \times cos\,\theta_{sun}(\ell)}{d_{sun}^2} \times BRDF_{dif}(b,d,\theta_{sun}(\ell),\varphi_{sun}(\ell),p) \qquad (2)$$

The solar irradiance used as reference for Sentinel-2 comes from [7]. This is the CEOS recommended solar irradiance spectrum for use in Earth Observation applications. The Earth to Sun distance is calculated by Orekit flight dynamics library [15].

The Level-0 Sun-diffuser acquisition is corrected from the dark signal, then is equalised with the current calibration parameters (absolute and relative gain coefficients) to provide the Level-1B equalised measurement, $Z_{meas}(b,d,\ell,p)$, by line and by pixel for each detector and band.

For each band b, the absolute calibration coefficient, $A(b)$, is then calculated by the average value of the ratio between measurement and simulation, for each pixel of each detector, over the acquisition lines:

$$A(b) = \frac{1}{N_d.N_\ell.N_p} \times \sum_{d,\ell,p} \frac{Z_{meas}(b,d,\ell,p)}{Z_{simu}(b,d,\ell,p)} \qquad (3)$$

where N_d is the total number of detectors (12), N_p is the number of pixels by detector, and N_ℓ is the number of acquisition lines selected for the processing (5100 lines for a 10 m-resolution band).

The non-valid pixels are not taken into account in the calculation of the average value. SWIR bands are corrected beforehand from the cross-talk effect. A stray-light correction is also performed, by applying a correction factor on simulations (1.007 whatever the band) taking into account a uniform illumination of the sensor.

Note that, for a given spectral band, the absolute calibration coefficient is related to the mean sensitivy of the radiometric response over all the detectors. The inter-detector variation of the sensor sensitivy is defined by the relative gain coefficients.

The monitoring of the absolute calibration coefficients response is an important output of the calibration activity as their time variation impacts directly the global level of the measured radiance. The evolution of the absolute calibration coefficients, as it is included in the processing baseline 02.04, is illustrated on Figure 12 for VNIR and SWIR bands.

The variation of the absolute calibration coefficients for VNIR bands is below 0.85% from the 06 July 2015 (reference date) until July 2016. Except for the B01 band, a decrease of the sensitivity is visible since the first Sun-diffuser acquisition to March 2016, by about −0.5% to −0.8% depending on the VNIR spectral band. A part of this decrease is due to the calculation of the Earth to Sun distance which was slightly improved by using Orekit in April 2016. Since April 2016, absolute calibration coefficients have remained stable by about 0.1% for VNIR bands.
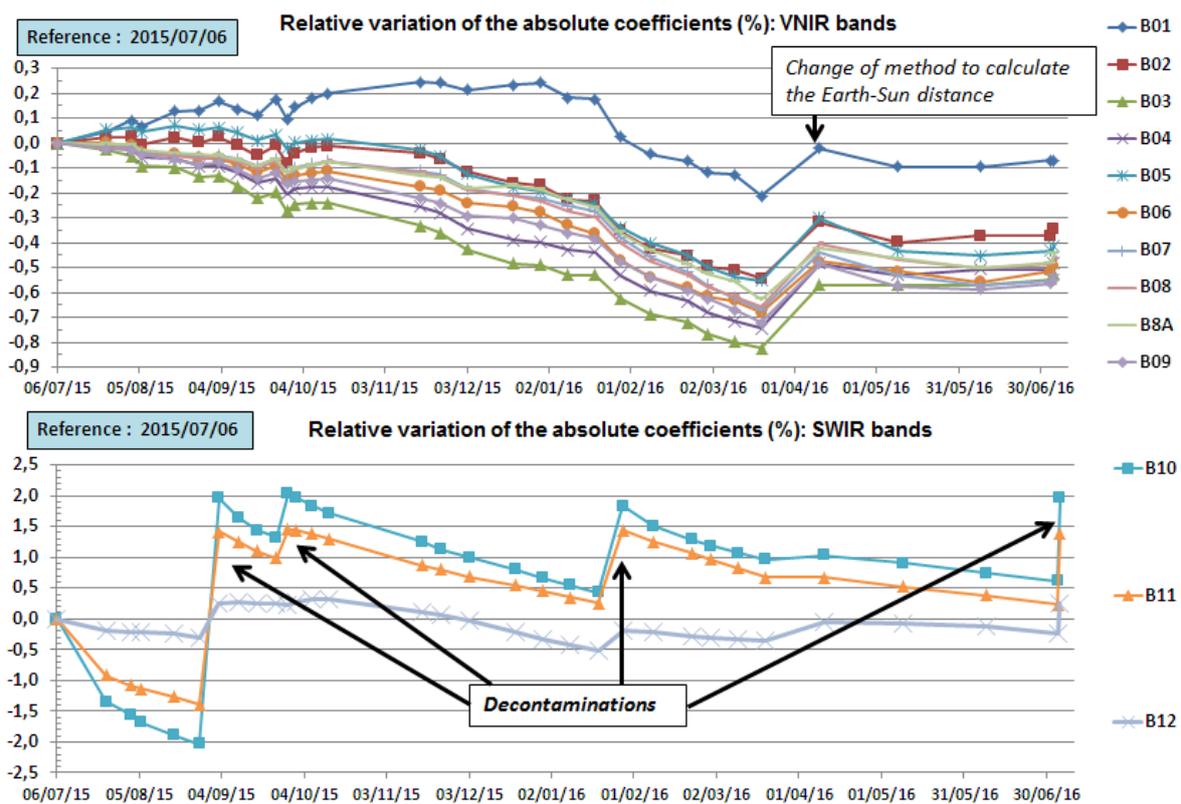
Figure 12. Time variation of the absolute calibration coefficients (for the processing baseline version 02.04), normalized to the coefficients estimated from the first Sun-diffuser acquisition performed on 06 July 2015.

The variation of the absolute calibration coefficients for SWIR bands is larger, mainly for the B10 and B11 bands. As mentioned in Section 2, these bands are respectively centred on wavelengths 1375 nm (30 nm width) and 1610 nm (90 nm width). The B10 band was designed to cover a strong water vapour absorption band and is especially sensitive to this gas. During the sensor manufacture, a tiny amount of water vapour was trapped in the optical system. A slight condensation effect occurs and impacts the sensitivity of this band (as well for B11 and B12, even if they are less sensitive to this absorption). This contamination effect was expected before launch and can be reversed by a decontamination operation consisting in increasing the temperature of the sensor during a short time: while its nominal temperature is around 200 K, the SWIR focal plane is heated to around 300 K during 90 min (the MSI is unavailable for nominal acquisitions during around 15.5 h, i.e., 3 h for heating up, 1.5 h with regulated decontamination temperature and 11 h to recover the operational temperature). The resulting effect is clearly noticeable in Figure 12. The trend of decrease of the absolute calibration for the B10 band is about −2.5% per six months (about −2% and −1% respectively for the B11 and B12 bands). This is completely compensated after decontamination (in the range of uncertainty due to the calculation of the Earth to Sun distance before the use of the more accurate computation by Orekit).

4.1.3. Relative Gains Calibration

Taking into account the dark signal variations, the measurements $Y_{meas}(b, d, \ell, p)$ in digital counts are equalised by pre-defined functions $\gamma(b, d, p, Y)$ which were estimated pixel by pixel in pre-launch characterisations and are updated in orbit from Sun-diffuser acquisitions. These functions deal both with the non-linearity of the sensor response and with the inter-pixel differences of response. They are defined by a cubic model for VNIR bands and by a bi-linear model for SWIR bands. A piece-wise linear function has been shown to provide a better fit to non-linearity measurements for the SWIR

detectors with respect to a bi-linear model. Here is detailed the mathematical approach for the cubic model, a similar approach being applied for the bi-linear model.

The relation between the equalised measurements $Z_{meas}(b, d, \ell, p)$ and measurements corrected for the dark signal, $Y_{meas}(b, d, \ell, p)$, in digital counts, for VNIR bands is:

$$Z_{meas}(b, d, \ell, p) = \gamma(b, d, p, Y_{meas}(b, d, \ell, p)) = \sum_{n=1}^{3} G_n(b, d, p) \times [Y_{meas}(b, d, \ell, p)]^n \tag{4}$$

In orbit, the relative gains can be adjusted only for the solar radiance level. Again, the method is based on the simulation of the solar radiance illuminating the sensor, as defined by Equation (2). The average value is considered over lines, giving a simulated radiance by pixel:

$$< Z_{simu} >(b, d, p) = \frac{1}{N_\ell} \times \sum_{\ell=1}^{N_\ell} Z_{simu}(b, d, \ell, p) \tag{5}$$

where N_ℓ is the number of selected lines for the processing (5100 lines for a 10 m-resolution band).

The relation between the simulated radiance by pixel $< Z_{simu} >(b, d, p)$ and the corresponding simulation of measurement corrected for the dark signal $< Y_{simu} >(b, d, p)$ is given by:

$$\gamma(b, d, p, < Y_{simu} > (b, d, p)) - A(b) \times < Z_{simu} >(b, d, p) = 0 \tag{6}$$

The first positive real root of the polynomial function gives $< Y_{simu} >(b, d, p)$. This value is then compared to the measurement (average over lines too):

$$R_a(b, d, p) = \frac{< Y_{simu} > (b, d, p)}{< Y_{meas} > (b, d, p)} \tag{7}$$

One can deduce the update of coefficients G_n in Equation (4), for equalisation functions $\gamma(b, d, p, Y)$ of VNIR bands, by:

$$G_n^{updated}(b, d, p) = G_n^{current}(b, d, p) \times [R_a(b, d, p)]^n \tag{8}$$

The temporal evolution of the equalisation coefficients is monitored by the MPC. During the operational phase of Sentinel-2A, the sun-diffuser acquisitions are performed on a monthly basis. After each sun-diffuser acquisition, a new set of relative gains calibration parameters is generated. Table 4 shows typical results of the R_a estimate. The time variation of relative gains is weak for VNIR bands: typically maximal variations do not exceed 0.2 to 0.4% between two assessments. Variations are higher for SWIR bands with a change of gain coefficients up to 3% for some pixels and with an inter-pixel variation more pronounced for SWIR bands than for VNIR ones. For instance, Figure 13 illustrates the change of equalisation coefficients for the B10 band as a function of the pixel number. The maximum variations appear on the figure. They induce stripes along track on the images as explained in Section 4.3.1. The update of the equalisation coefficients eliminates these artefacts.

Table 4. Statistics on the R_a factor for relative gains variation between the sun-diffuser acquisition on 08 June 2016 and the previous update of GIPPs from the acquisition on 09 May 2016, without cross-talk correction. Minimal, mean and maximal values of R_a and its standard deviation over all pixels are given for each spectral band.

2016-06-08	Min	Mean	Max	STD
B01	0.999	1.000	1.001	0.0004
B02	0.999	1.000	1.001	0.0003
B03	0.999	1.000	1.001	0.0004
B04	0.999	1.000	1.001	0.0004
B05	0.999	1.000	1.002	0.0006
B06	0.999	1.000	1.001	0.0004
B07	0.999	1.000	1.001	0.0005
B08	0.999	1.000	1.002	0.0005
B8A	0.999	1.000	1.002	0.0005
B09	0.999	1.000	1.002	0.0005
B10	0.989	1.000	1.028	0.0014
B11	0.992	1.000	1.027	0.0012
B12	0.994	1.000	1.006	0.0007
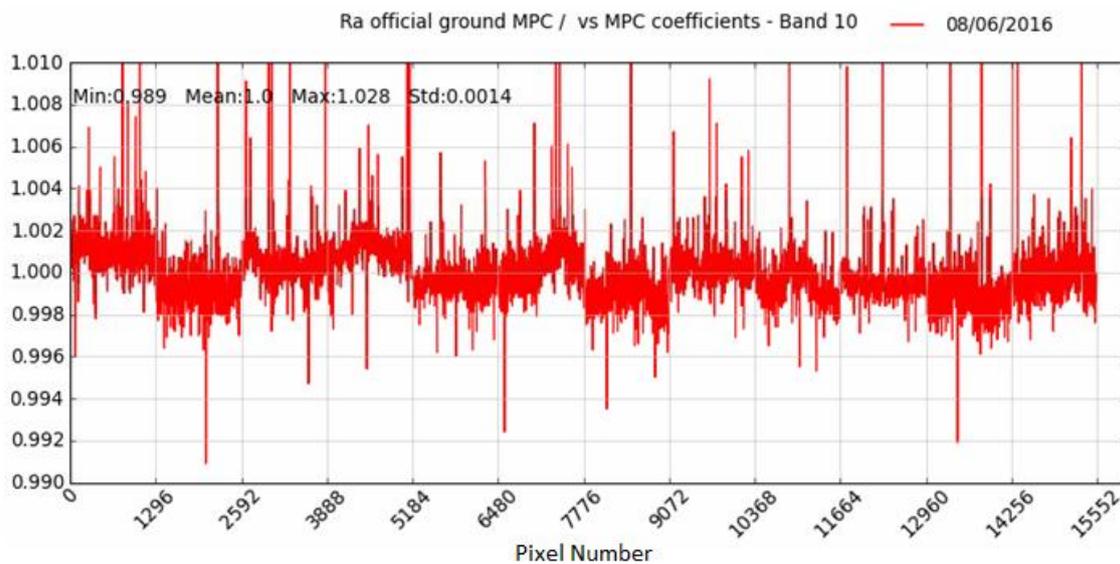
Figure 13. R_a factor as a function of the pixel number (over the 12 detectors) for the update of gain functions measured on 8 June 2016, for the B10 band, with respect to the previous operational coefficients calculated from the sun-diffuser acquisition on 9 May 2016.

4.1.4. SWIR Detectors Re-Arrangement Parameters Generation

As presented in Section 2.4, individual SWIR pixels TDI configuration is likely to be modified, depending on the radiometric performance of individual pixels.

The initial configuration mode has been determined before launch during the calibration of the instrument with the aim of achieving no defective pixel and the highest SNR. During the flight, the performance may degrade and a new configuration may be decided and uploaded to the satellite. The relative sensitivity of each pixel varies with the detector physical lines and pixels. There is no dependence on the line number of the detector matrix. It is correct to assume that all detectors are virtually single-line detectors -only one value per pixel will remain- as long as the calibration coefficients are recomputed every time the TDI configuration or the line selection changes.

Before the launch, each pixel of each TDI line has been tested on ground and affected to one of the following categories:

- 1: valid pixel with SNR compliant to specification,
- 2: valid pixel with SNR below specification,
- 3: non-valid saturated pixel,
- 4: non-valid blind pixel,
- 5: non-valid with SNR below specification.

Then, a rearrangement rule has been elaborated to classify the configurations relying on these indicators. The rules aim at maximising the SNR.

When a rearrangement is required, the decision is taken with respect to a radiometric performance threshold and the best possible remaining configuration is adopted. A memory of past configurations and associated performances is obviously necessary to be maintained.

As reselection is done on board the satellite, the impact on the mission is important. The new selection has to be uploaded and immediately followed by a sun-diffuser calibration as well as a dark signal acquisition to calibrate properly the new equalisation model of the pixel and validate the radiometric performance increase. Moreover, relevant GIPPs have also to be updated, particularly to take into account the new LOS of the pixel.

If none of the TDI configuration is satisfying, the pixel might be declared invalid and interpolated by ground processing.

4.1.5. Crosstalk Correction Calibration

An electrical crosstalk phenomenon has been characterised on ground in SWIR bands. As shown on Figure 14, when a band is illuminated (only a part of B12 on the figure) the two other bands are affected by a negative signal perfectly proportional to the signal in the illuminated band. The impact on a real landscape is a decrease of the actual signal in the affected bands.
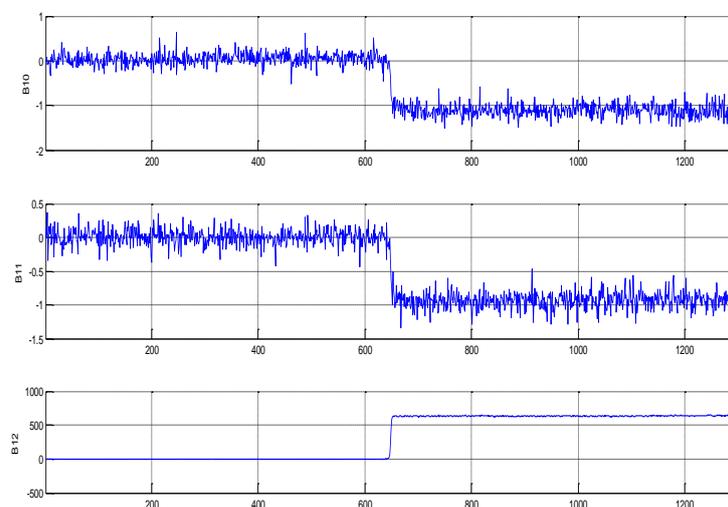

Figure 14. Characterisation of the electrical crosstalk: example of test result from ground measurement at Focal Plane Assembly (FPA) level (digital counts vs. pixel number).

Electrical crosstalk affects the 3 SWIR bands and does not evolve along time. This crosstalk slightly varies from one detector to the other. Averaged attenuations in dB are given in Table 5.

The figures from Table 5 expressed in percentage represent −0.45% for crosstalk from B11 to B10 and −0.2% from B12 to B10 whatever the radiance level. Note again that the figures are negative, which means that the crosstalk is subtracted from original signal.

Table 5. Averaged SWIR electrical crosstalk.

		Measurement on Band (dB)		
		B10	**B11**	**B12**
	B10	0	−83	−60
Illuminated Band	**B11**	−47	0	−76
	B12	−54	−53	0

Moreover, ghost images induced by crosstalk are shifted by constant number of pixels along track because of focal plane layout (see Section 2.3). For B10, for example, this number of pixel is exactly 66.6 from crosstalk from B12.

An example of such a crosstalk on a real landscape is given in Figure 15. On this figure, dynamic has been extremely stretched. Indeed, the order of magnitude of the effect is inferior to 2 or 3 DC.
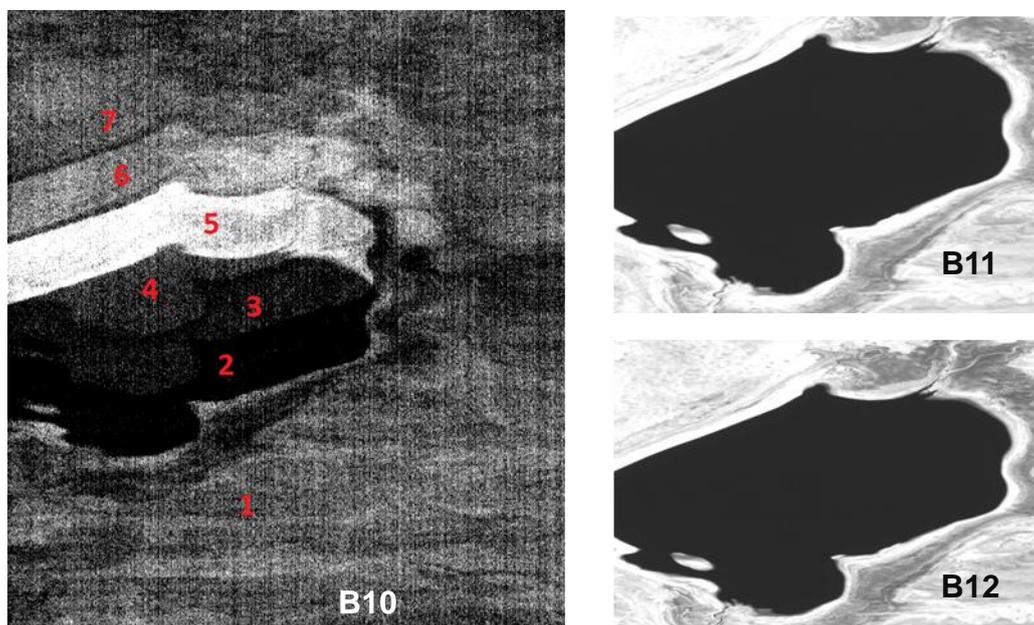
Figure 15. Left image shows an example of crosstalk affecting B10 (contrast strongly enhanced). **Right** represent the corresponding acquisitions in B11 and B12.

In Figure 15:

- Region 1: the signal is the signal measured in B10 far from the lake from which is subtracted 0.45% of the signal measured in B11 (around 5 DC) and 0.2% of the signal measured in B12 (around 2 DC) also far from the lake. B10 own signal is a landscape residual that is seen because of the lack of cirrus.
- Region 2: B10 is on the lake. B10 own radiometry is around 0 DC but B11 and B12 are not yet on the lake and crosstalk induces negatives values at focal plane level on B10. At ground level those negative values are truncated to zero.
- Region 3: B10 is still on the lake. B11 is now on the lake but B12 not yet. Thus the radiometry is superior to the one observed in region 2 because B11 crosstalk is null (because B11 radiometry is null).
- Region 4: B10, B11 and B12 are on the lake. Crosstalk is negligible. On this region we observe the real response of B10 on the lake.
- Region 5: B10 is no more on the lake but B11 and B12 are still. On this region we observe the real response of B10 on the lake border.

- Region 6: B10 and B11 are no more on the lake. B12 is still on the lake.
- Region 7: idem region 1.

The signal level measured is fully consistent with the crosstalk measured on ground. The 3 main ghosts of the lake are exactly shifted by 67 pixels one from the other.

Thus, this kind of crosstalk is fully predictable and corrigible thanks to a simple linear combination of SWIR bands. This processing had been anticipated on ground and implemented in the image processing chain. It is now systematically applied. Figure 16 shows an example of such a correction.
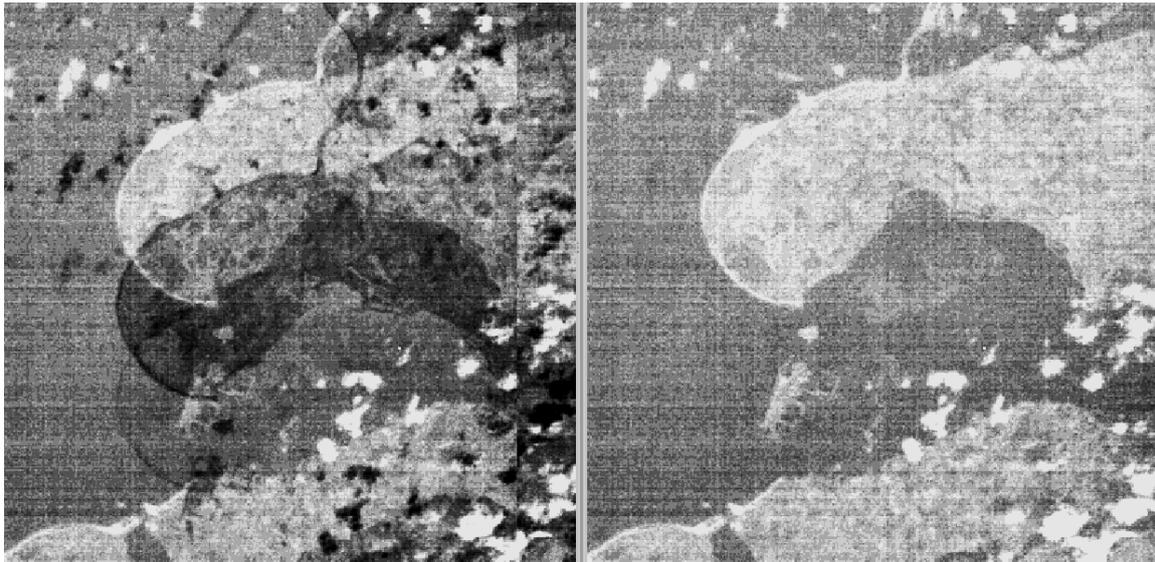
Figure 16. Example of crosstalk correction on band 10 (before/after).

Finally, it is noted that also sun-diffuser acquisitions are affected by crosstalk, and therefore are also corrected before being processed for calibration purposes.

4.1.6. MSI Refocusing

MSI refocusing is only foreseen as a contingency activity in case of Modulation Transfer Function (MTF) degradation (see Section 4.3.7 for more details on MTF assessment). This degradation is not expected, even considering the ageing of the instrument.

Nevertheless, if required, the MSI can allow some focal adjustment. Indeed, M3 mirror (cf. Figure 1) temperature has a direct impact on the instrument focus. It is nominally regulated at 20 °C, temperature for which the focus has been optimised on ground. When modifying this temperature, the consequence is a focus variation in the range of 5.5 μm/°C in the focal plane that has been pre-validated on ground.

4.2. Geometric Calibration Activities

Geometric calibration activities allow the determination of all the GIPPs of the geometric calibration model which aims at better ensuring geometry compliance to the requirements for Sentinel-2 images (orientation of the viewing frames, lines of sight of the detectors of the different focal planes). These parameters have been estimated before launch and the purpose of the geometric calibration activities is to take into account any update of these parameter values that might occur. Moreover, to meet the multi-temporal registration and absolute geo-location requirements, a Global (to be understood as worldwide) Reference Image called GRI is generated and used for automatic extraction of Ground Control Points (GCP) for systematic refinement of the geometric model within the Image Processing Facility (IPF).

4.2.1. Global Reference Image Generation

4.2.1.1. Definition and Goals of the Global Reference Image

The Global Reference Image (GRI) can be defined as a set of as cloud free as possible mono-spectral (B4, red channel) L1B products, whose geometrical model has been refined through a dedicated process designed by IGN (French Geographic Institute). The area of interest is worldwide, including most of the isolated islands, so that once the GRI is complete the whole archive of L1B products can benefit from its geolocation. The GRI will be actually used as a ground control reference: homologous points will be found between the GRI and the L1B product to be refined within the processing chain, then a geometric correction will be estimated.

Consequently, the GRI must have an absolute geolocation performance which allows respecting the following two specifications:

- The geolocation of L1C products refined with the GRI is better than 12.5 m CE95 (95 percentile of the circular error).
- The multi-temporal registration performance between refined products is better than 0.3 pixels.

4.2.1.2. Selection and Constitution

Since the GRI has a worldwide extent, the image selection work is a quite long step. It is done continent by continent. The main constraint is to find cloud free products. Even if the metadata files contain some information about the cloud coverage, it is not sufficient in our case. Indeed, L1B products may be very long (up to 8000 km): consequently, low cloud coverage does not mean that the part of the image we are interested in is really cloud free. Conversely, products with high cloud coverage may be partially useful when building up the GRI. Finally, almost none of the products are totally cloud free, which means that only parts of full products shall be selected to produce good quality mosaics.

The selection of products remains a fully manual task. Quick-look images at 320m resolution are used, their quality being sufficient to select or reject products upon a cloud criterion. End of July 2016 almost 500 products had been selected throughout the world, as illustrated on Figure 17.
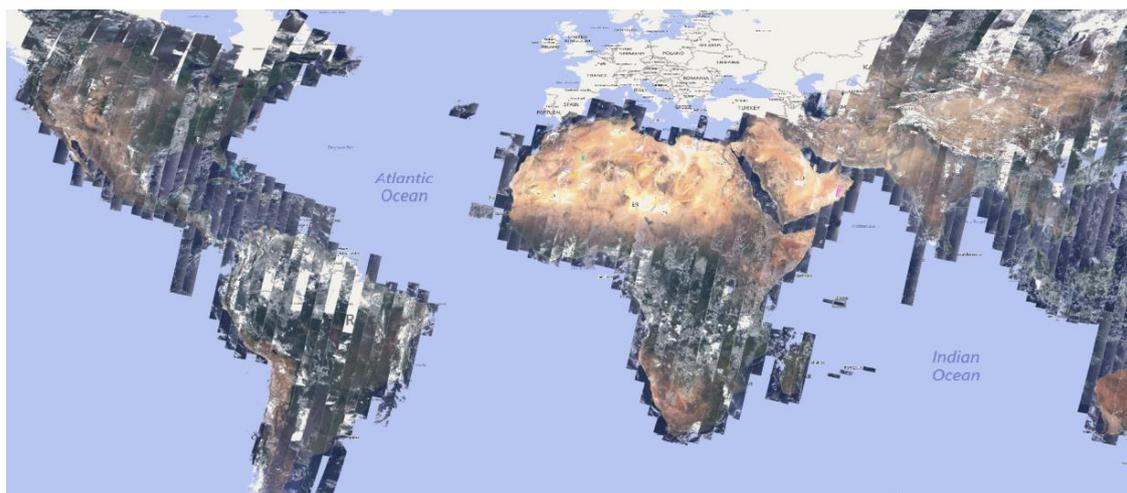

Figure 17. Overview of the Sentinel-2 GRI selection, July 2016 (European GRI products are not present on this map, since they were produced in the In-Orbit Commissioning Review (IOCR) context, in October 2015).

Equatorial areas are cloudy most of the time, and there are very few chances to get a cloud-free image over these regions. So, the GRI is not made of 100 % cloud-free products, but it is only necessary

to find enough GCPs in the GRI to refine other L1B products. Since segments are very long, some cloudy parts can be accepted by the registration method.

Figure 18 shows the monthly evolution of the selection over South America between end of April and end of July 2016: the Amazonian rainforest is always under clouds, and no satisfying images could be found yet. From a Sentinel-2 point of view, it may not be a real problem since there are enough cloud-free images and long data strips to retrieve a satisfying density of GCPs. This also illustrates that the GRI is not an aesthetical view of the world.
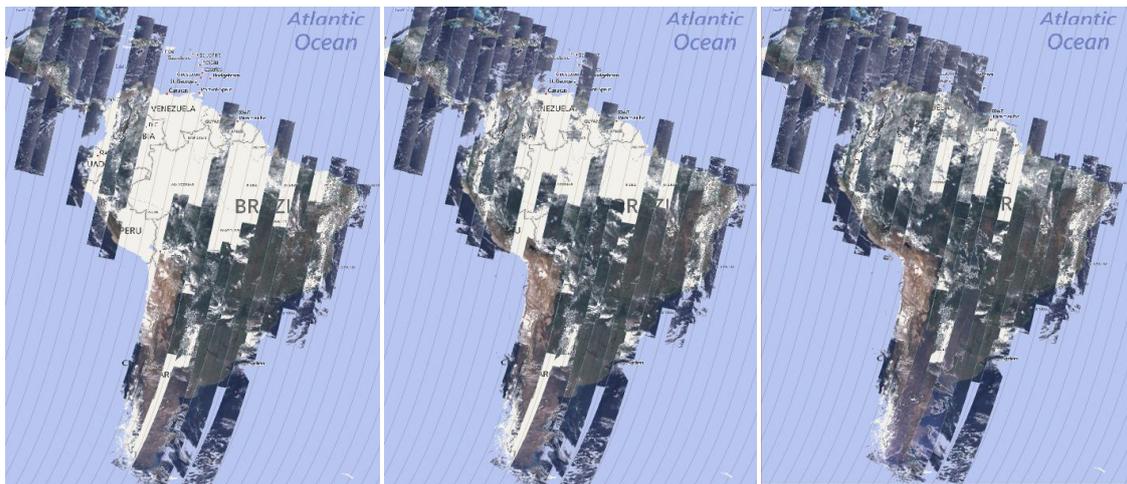
Figure 18. Monthly evolution of the image selection for the GRI (April to July 2016).

The complete GRI is expected to comprise around 1000 products, including products over isolated islands. The completion is scheduled for mid 2017.

4.2.1.3. Methods of Refinement

Basically, the method used to refine L1B products is the block spatio-triangulation: it provides refined values of the attitude of the satellite platform by using observations from ground and images (GCPs and tie points). The full coverage of L1B products used for the GRI will be split into continental or multi continental blocks in order to achieve the calculation process. Tie points are extracted using SURF (Speeded Up Robust Features): these features are more robust, less sensitive to shadow effects, and provide in general better results than the usual points of interest (e.g., Harris corners) used in a classical correlation process. These homologous points are very numerous and located everywhere there is an overlap: along the same orbit when available and between two (or more in high latitudes) neighbouring orbits.

GCPs can be selected using various methods and sources of GCPs databases. Several sources can be used:

- Identification on ground calibration sites or ITRF GPS network: the ITRF (International Terrestrial Reference Frame) network is worldwide. When possible, GCPs from these sites are used.
- Identification in an image of reference: several sources can be used as a reference because they are accurate enough for GRI need. This is the case of aerial BD Ortho ® over France made by IGN, AGRI (Australian Geographic Reference Image) made with ALOS imagery over Australia, aerial imagery over the USA made by the USGS, etc. In this use case, homologous points of interest are extracted between the reference and the GRI to be built. This extraction is done automatically or manually when the automatic process fails.

In the specific case of small and/or isolated islands, GCPs may not be available at all. Thanks to the excellent initial geolocation performance of Sentinel-2 (around 10–12 m), it is reasonable to

only compute a relative spatio-triangulation (using only tie points) of several Sentinel-2 overlapping products. Indeed, experiences show that the geolocation performance increases along with the number of overlapping images. Thanks to this method without any GCP, the specifications are respected.

Using these observations (tie points and GCPs), it is possible to refine the chosen set of geometrical parameters (attitudes, position): it is the spatio-triangulation step itself, i.e., the resolution of the least square system. At this step, outlier observations are removed automatically. As the initial geolocation of the satellite is very good, only an order 0 or 1 (according the situation) correction on yaw, pitch and roll is estimated. Most of the time, the values of adjustment are weak, corresponding to a few meters on ground.

4.2.1.4. Internal Controls

As producer of the GRI, IGN is not part of the validation team. However, some internal checks are performed before the GRI is delivered to the validation team. This step is achieved just after the refining task, using a derivative of the Reference 3D® product. This derivative, called BDAmer, is a worldwide set of GCPs, directly extracted from the whole Spot 5 HRS archive, that has been spatio-triangulated before being turned into the Reference 3D® product (a global orthoimage layer). This database has a very high density, which helps to provide trustful statistical results in IGN checking step.

The GCPs of the BDAmer database were automatically extracted and fully qualified. One strong advantage of this product is that every GCP is seen in stereoscopy (thanks to Spot 5 HRS instrument), sometimes with a higher multiplicity, and a long time interval. Consequently, a quality index can be computed from these parameters, as well as other indicators such as geometrical accuracy, multiplicity, durability, etc. All available 200×200 pixels Spot 5 HRS cropped and centred images are associated to each GCP.

As derived from Reference 3D®, the BDAmer database is accurate enough to be considered as a reference for Sentinel-2 images, although the estimated accuracy may locally vary. Since the multiplicity of each BDAmer point is higher than 2, each check point is correlated at least twice to limit the risk of false correlation. It also helps to increase the confidence in the result.

These checking operations can be considered as fully independent since the GCPs used to refine the models are not extracted from this product (or other derivative). To collect homologous points between Sentinel-2 and BDAmer, the method is identical to the extraction of GCPs. Finally, the locations of the homologous points are compared, and statistical results can be extracted. The main advantage of this method is that checks can be done anywhere in the Sentinel-2 products, and not only on local areas.

In the case of Europe during the In-Orbit Commissioning Review (IOCR), more than 5000 check points were used to compute a significant statistical result. Finally, the accuracy of the GRI over Europe was estimated at 8 m CE95 after IGN checking operations. CNES team made additional checks with Pleiades imagery in many local areas, as confirmed by Figure 19.
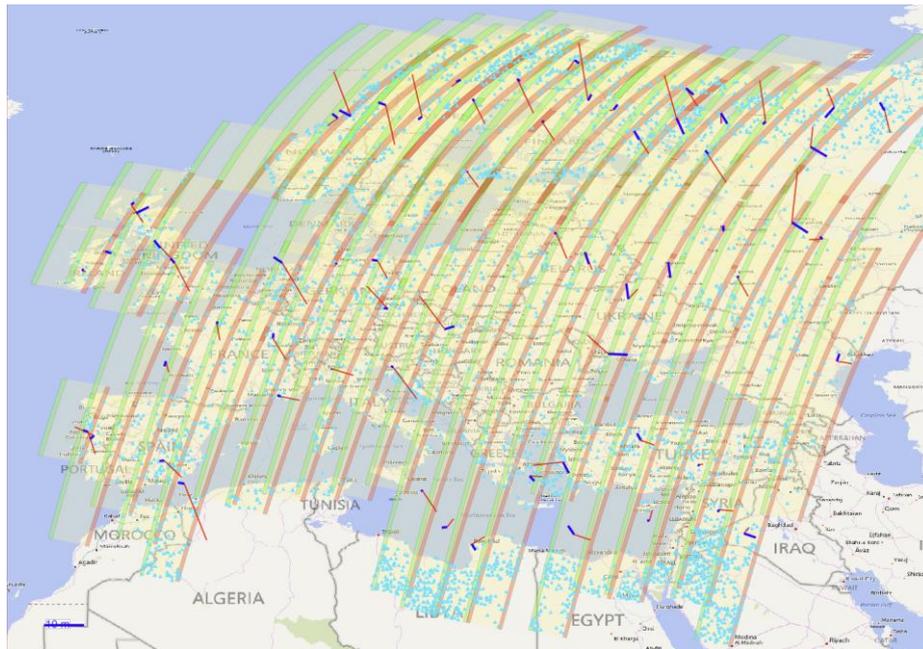
Figure 19. Checks from Reference 3D® /BDAmer (turquoise markers) data over Europe. Red and blue lines represent error vector of the product before and after refinement.

4.2.1.5. The Example of Australia

This paragraph illustrates the method described above, through the example of the processing of 33 products selected for the GRI over Australia, as shown in Figure 20. GRI processing over Australia is relatively straightforward for several reasons: the climate is helpful to find cloud-free images and the territory is very compact. The final selection covers almost the whole territory without any cloud. There are a couple of holes left, but it is not considered as blocking the use of the GRI in the Sentinel-2 context. Indeed, acquisitions are several thousands of kilometres long, so that enough GCPs can be found almost everywhere in the image.
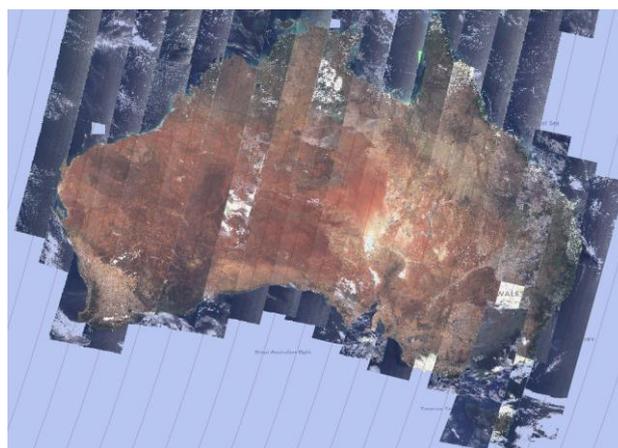

Figure 20. Selection of 33 products over Australia for the GRI.

The extraction of the GCPs for this block was facilitated by the availability of the AGRI database. At first, candidate points were extracted from the orthorectified AGRI database, using SURF methods (see Figure 21). Then these points were resampled at the resolution of Sentinel-2, and a correlation process was applied to find the homologous point in the Sentinel-2 image (orthorectified). About 800

points were found using this method, leading to a homogeneous repartition shown in Figure 22. Most of the points of interest found by the process are dark or light items of the landscape.
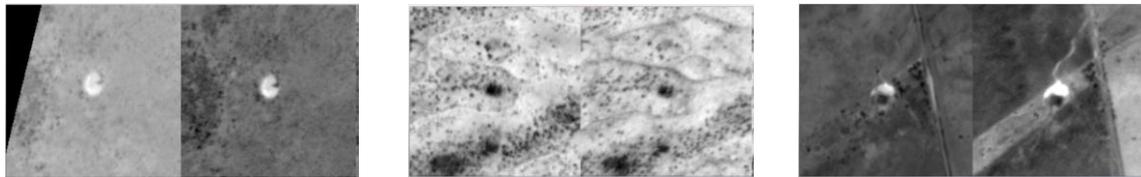

Figure 21. Three couples (3 examples over Australia) of homologous points between Sentinel-2 (left) and Australian Geographic Reference Image (AGRI) (right). Each time, the correlation between both images is made at 10m Ground Sample Distance from ground (orthorectified) images.
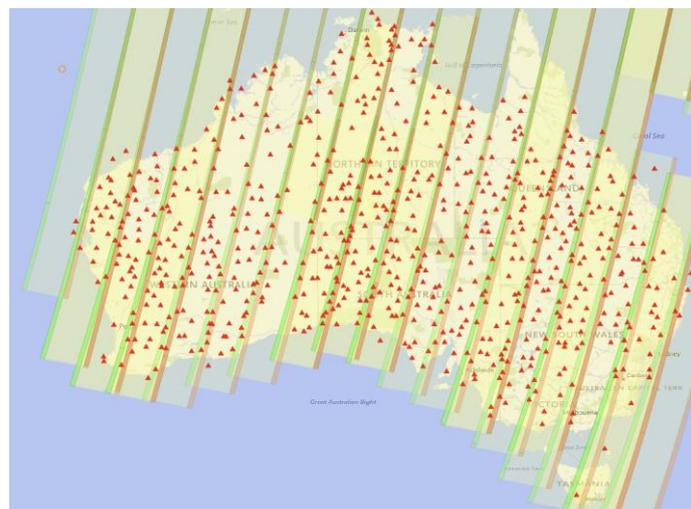

Figure 22. Repartition of the Ground Control Points (GCPs) extracted from AGRI.

The identified parameters to refine are an order 1 on yaw, pitch and roll and an order 0 on magnification. If acquisitions are very long, the order 1 is very constrained, to avoid a "leverage effect" along track. In this case, there are seven unknowns by product to determine, with a high redundancy in the observations, since there are 800 GCPs, and about 35,000 tie points.

After refining, the accuracy is about 9.50 m CE95 over Australia. This value was estimated thanks to the checking step based on BDAmer. This has still to be confirmed by the validation team, and then integrated in the system, that will be able to deliver refined products to the final users.

4.2.2. Absolute and Relative Calibration of the Focal Plane

4.2.2.1. Methods

The focal plane calibration consists in re-estimating the lines of sight of the detectors from the various retinas, i.e., estimating distortion and possible discontinuities between the different arrays of the focal plane to improve internal image consistency. Such accurate measurements are not possible on ground before launch.

In flight, an absolute method is used: it is based on the correlation between an image acquired by one of the retina to be mapped and an absolute reference image (orthorectified database, the so called BDortho, done by IGN over France). B4 has been chosen as the reference retina for Sentinel-2.

Using absolute reference images guarantees planimetric and altimetric accuracy consistency with requirements because they are geometric references close to ground truth. Moreover, they cover the full swath so that calibration of all detectors can be done.

The absolute reference images are rectified into the geometry of the sensor focal plane to be calibrated (B4 spectral band). Two sets of images are then obtained: the image from the B4 spectral band to be calibrated, which represents the physical reality of the system, and the reference image, rectified in the geometry of the estimated sensor. These rectified images are then correlated. The aim here is to measure the deviation parallel and perpendicular to the track (ALT and ACT respectively).

These raw image measurements resulting from correlation are reduced in the focal plane to angular units, e.g., radians, taking into account the aperture angle for each detector: this varies along track as a function of the roll angle.

Deviations (in the form of polynomials or other functions) are modelled per detector or per array. This modelling is then added to the current model of lines of sight, in such a way that minimises the row and column deviations between the absolute reference and the image of the sensor being calibrated.

The relative method for focal plane calibration uses the absolute method, but takes a well-calibrated existing band as a reference and uses an appropriate DEM (Digital Elevation Model). This method is used to calibrate SWIR focal plane from VNIR focal plane and more generally to calibrate the different multispectral bands to ensure consistency between channels LOS. Calibration parameters are thus estimated by correlating the various bands with each other. But the correlation reaches its limits when the spectral bands are very different. A study has been carried out before the commissioning phase to determine the best pairs for correlation. The following pairs are used for calibration:

- B4 spectral band is used as a reference to calibrate the focal plane of the B2, B3, B5 and B8 spectral bands;
- B5 spectral band is used as a reference to calibrate the focal plane of the B1, B6, B7, B11 and B12 spectral bands;
- B8 spectral band is used as a reference to calibrate the focal plane of the B8a and B9 spectral bands;
- B2 spectral band is used as a reference to calibrate the focal plane of the B10 spectral band.

Figure 23 details all the couples of bands used to perform the multi-spectral registration. The black-links refer to the couples used for the calibration. The blue-links are additional couples used for final checking of the performances.
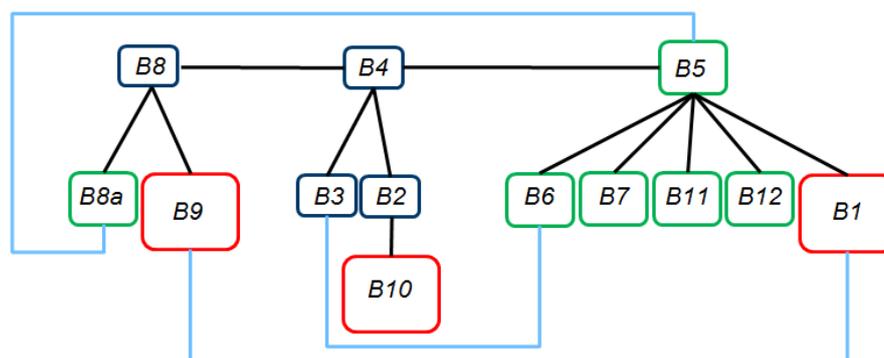

Figure 23. Couples of bands used to perform the multi-spectral registration. Blue boxes: 10 m bands; green boxes: 20 m bands; red boxes: 60 m bands.

4.2.2.2. Results

A calibration has been performed to correct the geometric model computed on ground. After correction of the geometric model, the residuals errors were less than 0.2 pixels.

To evaluate any evolution of the focal plane after this calibration, another method was used by correlation (whatever the site) between two products from the same orbit. This is the same method but:

- products are located anywhere around the world so cloud-free acquisitions can be chosen;

- correlation is computed between two Sentinel-2 products: a reference image is not needed as the correlation noise is low.

This method is a relative method to determine any evolution of the absolute calibration. There has been no evolution of the focal plane since the beginning of October 2015.

Concerning the relative calibration, a particular case has been observed for band B10. At first, the images used for calibration of B10, with B01 as reference, were cloudy areas with cumulus that are visible in both bands. But the results were too noisy, because there is 0.45 s of time lag between acquisitions by the two bands and the movement of clouds during this time lag can be up to 1 pixel of 60 m. For clouds at 10 km of altitude, parallax error between these 2 bands is 42.5 m. In addition, the cloud speed at this altitude can be around 200 km/h, which means 50 m of error. So the images finally used for calibration are over very high mountains, without clouds. In that case, the integrated water vapour content is sufficiently low to allow a good detection in B10. Therefore, B10 is calibrated on ground not at cloud altitude

Generally speaking, the LOS is well calibrated on ground with an in-orbit correction of less than 2 pixels for VNIR bands and less than 15 pixels for SWIR bands. Only few gaps have been corrected (except on SWIR bands where the focal plane is not the same as the VNIR one, so the gaps were larger). There are 3 main contributors to the registration error:

- Static LOS calibration residuals;
- Dynamic vibrations residuals that represent the main contributor because of on-board oscillations;
- Correlation noise and outliers which depend on landscapes and spectral couples.

The performances are very good and homogeneous on every couple. The registration residual (positive distance) is better than 0.3 pixels at 3σ, in all cases except on band B10 where it reaches 0.6 pixels at 3σ. This is principally due to the very noisy measurements obtained on this particular band, so this is not a bias between bands but the measured gap for each pixel with a lot of noise. Moreover, as the LOS is measured for this band on high mountains, there is an effect from the DEM: products from different repeat orbits must be used to average biases coming from DEM.

Viewing directions are very stable. From the beginning of September 2015, no evolution has been observed. It is important to note that if viewing frame biases have a temporal evolution, there is an evolution on viewing directions. A correction of these biases should correct the focal plane residuals.

4.2.3. Absolute Calibration of the Viewing Frames

4.2.3.1. Methods

Absolute calibration of the viewing frames or calibration of the absolute alignment biases consists in determining the absolute orientation of these frames. This calibration is achieved by refining the geometric models on a large set of scenes, using Ground Control Points (GCP) or well-located reference images. Scenes are acquired on various sites (geographical sites which are perfectly known geometrically, also called GIQ sites, as reference for Geometric Image Quality purposes) distributed all over the world; the aim is to cover the maximum number of longitude and latitude ranges to:

- Determine any possible change in alignment biases as a function of criteria such as latitude (analysis of possible thermo-elastic effects),
- Avoid being too dependent on weather conditions.

For all the scenes, the geometric models are then refined by space-triangulation to determine an average biases set per scene (pitch, roll and yaw biases). An analysis of the alignment biases estimated for all scenes and all GIQ sites then allows to:

- Determine a mean biases set to update GIPP,

- Observe a possible evolution of biases according to such criteria as latitude, date, etc.

For Sentinel-2, this method is used for the calibration of the VNIR viewing frame. Sentinel-2 images are acquired on the various GIQ sites. The B04 spectral band of these Sentinel-2 images is systematically correlated with the ground control points or the well-located reference images of the GIQ sites to refine the geometric model of the VNIR frame.

4.2.3.2. Results

The output of alignment biases calibration of a viewing frame is the mean biases set in pitch, roll, and yaw, and the associated GIPP to be delivered (GIPP_SPAMOD, standing for spacecraft model). Those biases are corrected by on-ground processing parameters, although in case of an error larger than 2 km just after launch, an update of the on-board model would have been performed. A new ground spacecraft model GIPP would then be associated to these new on-board biases. But it has not been necessary.

Just after launch, measured alignment biases were around 2.5 km. After on-board Stars Trackers calibration, on 3 July 2015, biases decreased to around 700 m. Then, to reach the specification (under 20 m for non-refined products, and under 12.5 m for refined product with GRI), calibration of viewing frames is managed via Image Quality Geometric Model. As those biases are evolving over time, several sets of alignment biases have to be applied depending on a validity date indicated in the GIPP spacecraft model. From the launch until July 2016, Table 6 shows the values of pitch, roll and yaw angles to be applied.

Table 6. Viewing frames alignment biases.

Applicability Date	GIPP DATATI	GIPP SPAMOD		
		Roll	Pitch	Yaw
03/07/2015	−2 ms	−81.0 µrad	953.0 µrad	0
01/09/2015	−2 ms	−77.8 µrad	949.4 µrad	22 µrad
15/11/2015	−2 ms	−77.8 µrad	946.0 µrad	47 µrad
01/02/2016	−2 ms	−78.7 µrad	952.4 µrad	62 µrad

Alignment biases in pitch and roll are very stable. The yaw angle is slowly drifting and is fully monitored.

4.3. Radiometry Validation Activities

Radiometric validation activities aim at assessing all radiometric performances related to image quality requirements. The validation activities are performed regularly but with various time scales from weekly basic checks to yearly in-depth analyses, including long-term statistical and trend analysis. In case of reported performance degradation, specific calibration activities can be triggered.

4.3.1. Equalisation Validation

The goal of the equalisation validation is to verify that all pixels in a given band have the same response to a uniform radiance level. This is directly linked to the quality of the dark signal calibration and relative gains calibration (see Section 4.1.1 to Section 4.1.3) and to the linearity model characterization. In case of an observed degraded performance, a dedicated calibration activity and possible investigations can be triggered. Figure 24 illustrates the effect of equalisation on the image quality.
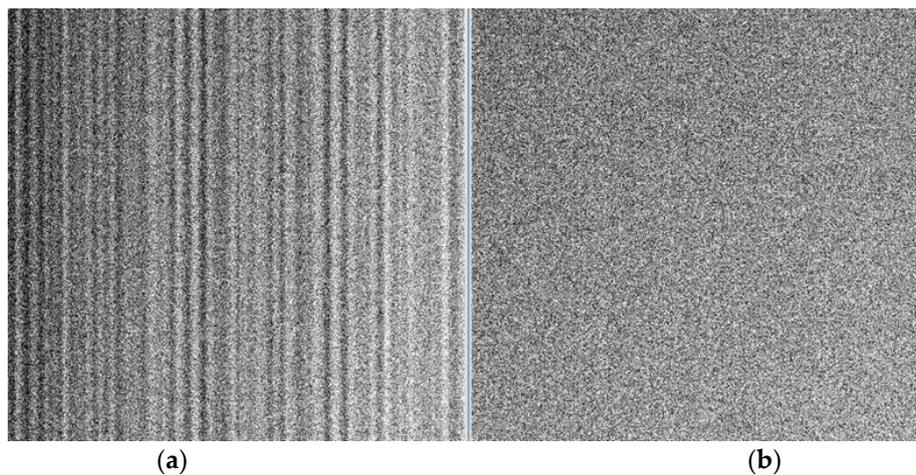
(**a**) (**b**)

Figure 24. Illustration of image equalisation impact: (**a**) Non-equalised diffuser image; (**b**) Equalised diffuser image.

4.3.1.1. Methods

The equalisation validation is based on the analysis of images over radiometrically uniform scenes that can be:

- On-board Sun-diffuser acquisitions: in this case the true radiance is known;
- "Uniform" natural targets on Earth (vicarious method): deserts, ice, etc.

The method using acquisitions over natural sites is always more pessimistic because the natural targets are never perfectly uniform and thus the scene non-uniformity contributes to the measurement error. The non-uniformity at detector transitions due to BRDF impact (the viewing angle being different for adjacent detectors) makes this method relevant only inside each detector. Moreover as it is not possible to find a uniform landscape for the full field of view, this method shall be performed combining a set of sub-swaths. This makes the operations quite demanding.

For these reasons the nominal method for the computation of the equalisation criteria is the one using diffuser data. The vicarious method on uniform scenes is used for cross-validation of the parameters provided by the diffuser and correction if necessary.

The instrument response non-uniformity is assessed through the Fixed Pattern Noise (FPN) and Maximum Equalisation Noise (MEN) which quantify local non-uniformities in the responses of physical pixels across the swath. Although the general principle, shown in Figure 25, is applicable to both sun-diffuser image and to uniform scenes, the computation methods are slightly different. Indeed in the case of diffuser image, one can use the sun-diffuser characterization to know the true radiance of the image.
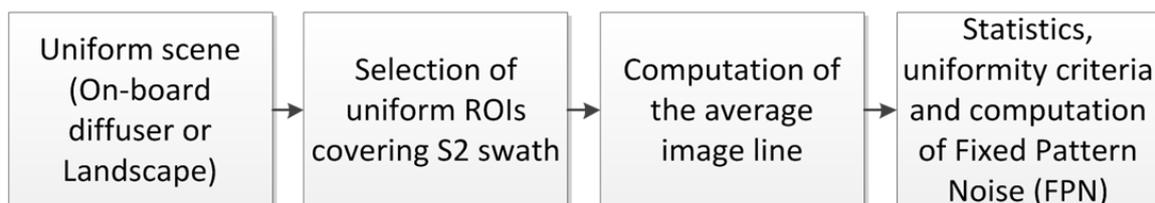

Figure 25. General principle for equalisation assessment.

For Sun-diffuser L1B images, FPN and MEN are computed as follows:
For each band,

(1) For each detector, correct each pixel of the sun-diffuser BRDF (to avoid local fluctuations due to the spatial response of the sun-diffuser) and of solar angle: Compute the average line of the ratio between observed radiance and sun-diffuser simulated radiance (see Equation (3) in Section 4.1.2):

$$R(b,d,p) = \frac{1}{N_\ell} \sum_{\ell=1}^{N_\ell} Z_{meas}(b,d,\ell,p) / Z_{simu}(b,d,\ell,p) \tag{9}$$

In practice the average is computed over Nl = 5000 lines (at 10 m resolution).

(2) Build a full-swath line concatenating the previous average line but keeping only pixels that will be seen in the end-user product: $R(b,p')$

(3) Using a sliding window with a step of 1 pixel, compute the mean and standard deviations over a section of 100 pixels and derive the (normalised) FPN and MEN:

- $FPN(b,p') = STD(R)/MEAN(R)$, where MEAN and STD are respectively mean and standard-deviation over pixels in the sliding window centred at p.
- $MEN(b) = MAX(STD(R))$ over all pixels across-track

Note: All defective pixels are ignored in the computation.
For uniform L1B images, FPN and MEN are computed as follows:
For each Region of Interest in the image (included in one detector) and for each band,

(1) Compute the mean line of the image.
(2) Compute FPN and MEN performing step (3) of the method using sun-diffuser acquisitions.

4.3.1.2. Results on Sun-diffuser Acquisitions

This method is performed at the same time that the relative gains calibration using sun-diffuser acquisitions. The FPN and MEN are estimated on the sun-diffuser acquisition converted to L1B product using the current relative gains and dark offsets. It allows assessing the status of the equalisation just before the new calibration (worst case).

For instance, Table 7 gives a statistical overview of the values of FPN measured for each spectral band on the sun-diffuser acquisition on 4 July 2016 by applying the equalization coefficients estimated from the previous sun-diffuser acquisition on 8 June. About 5000 lines (N_ℓ) at 10 m resolution are used to compute the average line in Equation (9). For all the VNIR bands, the maximal value of FPN is clearly below the specified limit (0.2% for all bands, except for B09 for which the specification is 0.3%) typically by an order of magnitude. For SWIR bands, it is not the case for the B10 and B11 bands: for these bands, the maximal value of FPN can be higher than the specified limit (0.3% for B10, 0.2% for B11), but the FPN values exceed the limit for a limited number of pixels as the values of FPN for the quantiles 98% are significantly below the specified acceptable limits, even for these bands. The update of equalization coefficients resets the FPN to zero, and ensures that the requirements will be met until the next calibration.

4.3.1.3. Results on Uniform Scenes

Here are presented the results obtained on a Greenland snow image. Comparable results have been obtained on other similar images. The first step is to select one or more uniform areas to be able to compute the FPN. At first glance, that seems easy since snowy landscapes are known to have a high and quite spectrally flat radiance in VNIR and to be spatially uniform. However, very small effects are investigated, so the zone selection has to be performed carefully, avoiding variation in snow properties, and any patterns caused by wind or elevation variation. As Sentinel-2 has a very wide field of view, only subparts of the swath are considered. Once this is done, it is possible to compute, for each zone, statistical quantities on FPN as depicted in Figure 26.

Table 7. Statistics on the Fixed Pattern Noise (FPN) (in %) estimated on the sun-diffuser acquisition of 04 July 2016 by applying the operational equalisation coefficients calculated from the sun-diffuser acquisition of 08 June 2016. Values are given for the minimal, average and maximal values, and for the quantile at 98 %. The specified maximum acceptable (requirement) values are recalled.

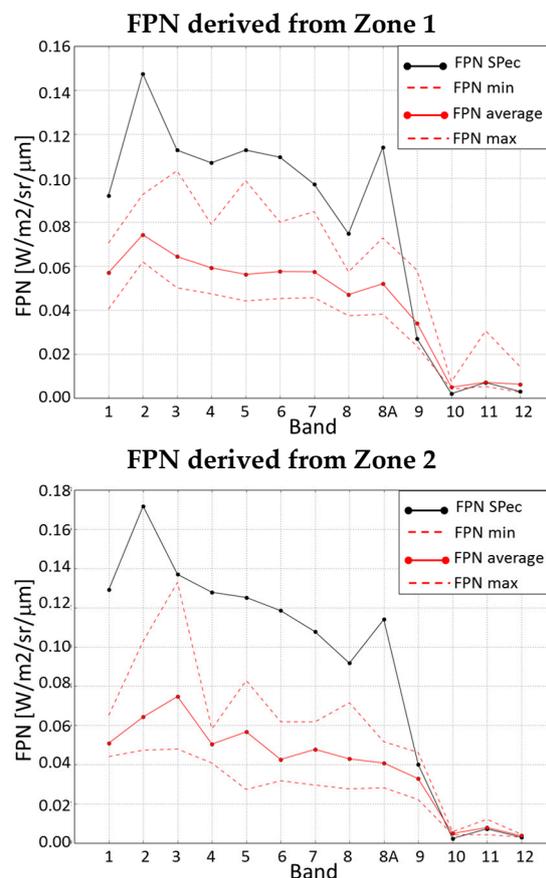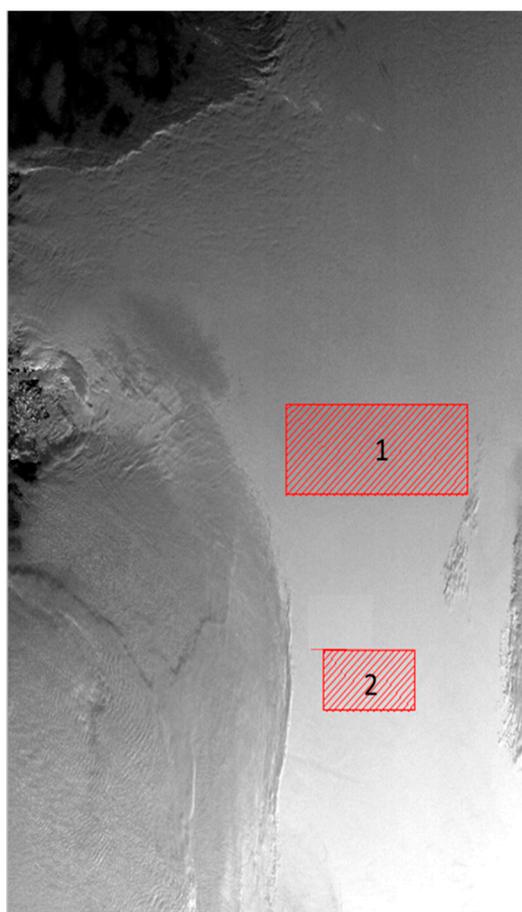| Band	Req.	Min FPN	Avg FPN	Q. 98% FPN	Max FPN
B01	0.2	0.00	0.01	0.02	0.03
B02	0.2	0.01	0.01	0.02	0.02
B03	0.2	0.00	0.01	0.01	0.03
B04	0.2	0.01	0.01	0.01	0.03
B05	0.2	0.01	0.01	0.02	0.04
B06	0.2	0.01	0.01	0.01	0.04
B07	0.2	0.01	0.01	0.02	0.05
B08	0.2	0.01	0.01	0.02	0.03
B8A	0.2	0.01	0.01	0.02	0.03
B09	0.3	0.01	0.01	0.02	0.06
B10	0.3	0.00	0.11	0.24	0.95
B11	0.2	0.03	0.07	0.17	0.35
B12	0.2	0.00	0.03	0.07	0.10

Figure 26. Left: Sentinel-2 Greenland image from 2015/09/04 together with the 2 uniform zones that are used to estimate the equalization quality. **Right**: FPN derived independently from the 2 zones.

In the VNIR range, the FPN is almost systematically lower than the specification (0.3% of input radiance for B09 and B10; 0.2% for other bands). Even in the worst case, the equalization quality meets

the requirement. This is a very strong confirmation of the quality of the non-uniformity coefficients derived from the diffuser. FPN derived on diffuser images are of course even better because it is a purely uniform target. But it is conservative to claim that the VNIR equalization quality is very good since the FPN derived, and that are certainly tainted by landscape residuals, are within specification.

However, for any band of the SWIR domain, the FPN in this image is higher than the requirement. Looking at the selected zones for those specific spectral bands, it is clear that they are less uniform than VNIR ones and that this non uniformity is caused by the landscape itself. Indeed, the reflectance is very sensitive to the snow properties (snow/ice ratio, dust ratio, etc.) in that spectral range. Consequently, we consider that SWIR results obtained here are not relevant. The use of other landscapes, like desert or ocean, is on-going to remove this restriction.

4.3.2. Absolute Radiometry Vicarious Validation

The absolute radiometric calibration relies on the on board solar diffuser. This calibration has been validated using vicarious methods. As sensor life increases, the vicarious calibration aims at distinguishing between the sensor aging and the diffuser aging. Vicarious calibration consists in equivalent TOA radiance computation for a known surface reflectance and a known atmosphere. For the spectral band b, the ratio between the TOA observed reflectance and the computed equivalent TOA reflectance gives the absolute calibration coefficient A_b. Except for band B10, the vicarious calibration relies on three methods, the Rayleigh scattering method, the Deep Convective Clouds (DCC) method and the ground-based reflectance method.

4.3.2.1. Methods

Rayleigh Scattering Over Ocean Surface

The TOA signal measured by the satellite sensor over ocean targets is mainly due to the scattering by atmospheric components in the visible spectral range [16]. In ideal conditions—stable oceanic region, with low concentration of phytoplankton and sediment, and far from land to ensure a purely maritime aerosol model [17]—the molecular scattering (so-called Rayleigh scattering) constitutes about 90% of the TOA signal. The assumption holds for Case 1 waters with low chlorophyll concentration and where phytoplankton is the only optically significant water column contributor [18]. Thus, Rayleigh scattering can accurately be calculated based on the surface pressure and viewing angles. In this study the Rayleigh scattering is based on the methodology of [19,20] using open ocean observations, to simulate the molecular scattering (Rayleigh) in the visible and comparing against the observed TOA reflectance to derive a calibration gain coefficient.

To ensure a proper computation of the vicarious coefficients over Rayleigh scattering, the following conditions need to be satisfied: (1) 0% cloud coverage on ROI; (2) low wind speed, typically less than 5 m/s; (3) low content of aerosol. Following [19], the Rayleigh-Corrected Normalized Radiance (RCNR) at 865 nm is used for screening aerosol content. The RCNR is defined as:

$$RCNR = \left(\rho_{TOA} - \rho_{Rayleigh} \right) \cos \left(SZA \right), \tag{10}$$

where ρ_{TOA} is the TOA reflectance, $\rho_{Rayleigh}$ is the Rayleigh reflectance, and SZA is the Sun-Zenith Angle. This coefficient is dimensionless. Images are processed only if the RCNR is lower than 0.008. This threshold avoids using further data screening for sun glint too.

A marine model following [21] provides an estimate of irradiance reflectance at null depth from 350 to 700 nm, as a function of chlorophyll concentration and sun zenith angle. The conversion from irradiance reflectance to marine reflectance above sea surface (ρw) follows [22].

To ensure a 0% cloud coverage over the region of interest (ROI), the target dataset is automatically cloud-screened using DIMITRI (Database for Imaging Multi-spectral Instruments and Tools for Radiometric Inter-comparison) package (https://dimitri.argans.co.uk). To ensure a low concentration

of phytoplankton and sediment, and pure marine aerosol, six Cal/Val sites were selected following [23] (Figure 27 and Table 8).
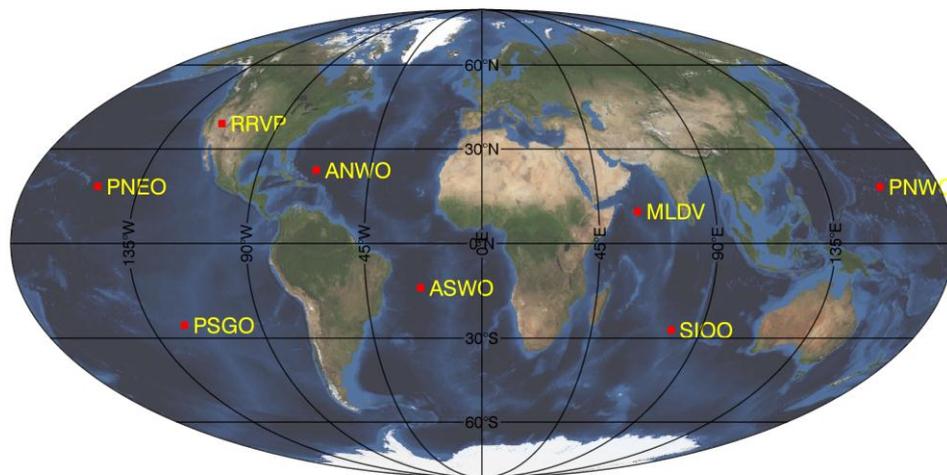

Figure 27. Cal/Val test sites location for Rayleigh scattering, Deep Convective Clouds (DCC), and ground-based methods chosen for Sentinel-2A commissioning activities. ANWO: Atlantic NW-Optimum, ASWO: Atlantic SW-Optimum, PNEO: Pacific NE-Optimum, PNWO: Pacific NW-Optimum, PSGO: Pacific Southern-Gyre-Optimum, SIOO: South-Indian Ocean-Optimum, MLDV: Maldives and RRVP: Rail-Road Valley Playa. (world image is taken from http://visibleearth.nasa.gov).

Table 8. Definition of the absolute radiometry vicarious validation sites used in this analysis.

Name	Latitude (°)		Longitude (°)	
	Min	**Max**	**Min**	**Max**
Atlantic-SW-Optimum	−14.5	−13.5	−24.5	−23.5
Atlantic-NW-Optimum	22.5	23.5	−67.5	−66.5
Pacific-NE-Optimum	17.5	18.5	−152.5	−151.5
Pacific-NW-Optimum	17.5	18.5	156.5	157.5
Pacific-Southern-Gyre-Optimum	−26.5	−25.5	−121.5	−119.5
Southern-Indian-Ocean-Optimum	−27.5	−26.5	77.8	78.5
Maldives	−10.0	10.0	60.0	90.0
Rail-Road Valley Playa	38.495	38.505	−115.685	−115.695

After the selection of L1C tiles that are within the ROI, the radiometric and geometric measurements of the 13 bands are re-sampled into B01 grid (60 m spatial resolution). This option has been motivated by the big size of S2 products with full resolution (10 m). To minimise the cloud contamination and the processed product size, sub-ROIs surface of 0.1 × 0.1 up to 0.3 longitude degree × 0.3 latitude degree are chosen. Then the TOA reflectance, the sun and viewing angles, cloud-mask and auxiliary variables are stored for each pixel. Quick-looks for each tile are generated.

Inter-Band Calibration over Deep Convective Clouds

The inter-band calibration method is usually used for lower resolution sensors (at least hectometric) [24], and has been experimented as a validation one for Sentinel-2 to demonstrate its potential for higher resolution sensors. It is an interband calibration method, which means it only gives relative measurements of calibration coefficients accuracy with respect to a reference band. It is useful to verify interband calibration or to be used in the synthesis of all calibration methods. Any spectral band in the VNIR range can be used, but usually, the red spectral band is used (670 nm or B04 for Sentinel-2) since the Rayleigh calibration is supposed 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, like over the Maldives or in the Gulf of Guinea. Since this method has been experimental for Sentinel-2A and there were operational constraints to have additional images, only a small zone (600 × 700 km^2) has been chosen over the Maldives during the In-Orbit Commissioning Phase (Figure 27 and Table 8).

Cumulonimbus clouds or Deep Convective Clouds (DCC) in this area are high and dense, and strongly diffuse the incoming solar radiance with a spectrally flat and lambertian reflectance. Moreover, they reach a very high altitude (8–12 km), and there is almost no atmospheric perturbation of the signal over the cloud. The measured signal at satellite level can be written as:

$$L_{DCC}(\lambda,\ \theta_v,\ \theta_s,\ \Delta\phi) = \left(L_{cloud}(\tau_{cloud},\ \tau_{aer},\ \tau_{mol}) + L_{Ray} + L_{aer}\ \right)\ \times T_g(\lambda,\ \theta_s,\ \theta_v) \tag{11}$$

where:

L_{cloud} is the radiance at top of the cloud, which depends on the geometrical conditions, the cloud particle type and, less significantly, on what is under the cloud (aerosols, tropospheric molecular signal, etc.). The clouds are so thick that the surface has zero influence on the result.

L_{Ray} is the residual molecular signal over the cloud.

L_{aer} is the residual aerosol signal over the cloud.

T_g is the gaseous transmission above the cloud.

The different radiances are computed in pre-calculated Look-Up Tables (LUTs), and the typical relative importance of each contribution is given in Table 9. The main difficulty of this method is the characterization of DCC and in particular the components of the top layer usually constituted 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.

Table 9. Estimation of the relative importance of all contributors to the Top-Of-Atmosphere (TOA) radiance over DCC (average case).

	Contribution to the Signal, Except Absorption			
Band (nm)	Cloud	Molecular Signal	Aerosols	Gaseous Transmission
443	95.7%	4.3%	0.0%	0.1%
490	97.2%	2.8%	0.0%	1.2%
560	98.4%	1.6%	0.0%	5.6%
665	99.2%	0.8%	0.0%	2.4%
775	99.5%	0.5%	0.0%	4.9%
865	99.7%	0.3%	0.0%	0.0%

Ground-Based Reflectance Measurements

The ground-reflectance-based approach has been used over the Radiometric Calibration Test Site (RadCaTS; Figure 27) to perform the absolute radiometric vicarious calibration for two decades [25] and recently operationally for Landsat-8 [26–28]. In this study are used the ground-based measurements provided by NASA (Landsat Cal/Val Team) via ESA as part of the ESA-NASA agreement. Five cloud-free S2A overpasses were obtained over the Railroad Valley Playa site (RRVP) but only one overpass coincides with the S2A time-series from the processing baseline 02.01 used here. The chosen reference is the TOA normalized reflectance reconstructed by the University of Arizona team using ground and atmosphere radiometric measurements. Note that, the dataset has been specifically generated for the spectral bands of S2A using hyperspectral reference profiles of the site.

4.3.2.5. Results and Analysis

Rayleigh Scattering Results

The assessment of a given sensor calibration consists in comparing the observed TOA reflectance (ρ_{obs}) by the satellite sensor to the simulated TOA reflectance (ρ_{sim}) by the radiative transfer model. The resulting ratio A_b, defined as ρ_{obs}/ρ_{sim}, provides a measure of the quality of the calibration of the instrument.

The Rayleigh method in DIMITRI is performed over the valid S2A-L1C products (11 acquisitions) over the commissioning period June 23 2015-April 1 2016. The results seem sufficiently reliable to provide a first estimate of the absolute vicarious calibration coefficients in the visible spectral range (Figure 28 and Table 10). The results are consistent over all test sites (not shown) and in good agreement with independent results obtained by CNES using the on-board diffuser calibration over the commissioning period of S2A (Sentinel-2 Image Quality Team, 2016). In addition, these results are in good agreement with the results of [29] obtained over shorter period, which attests on the temporal stability of the sensor measurement. The estimated measurement uncertainty associated to each site is less than ±5% and it is less than ±2% when taking into account all the test sites. This suggests that the accuracy of the method is compliant with the mission requirement.
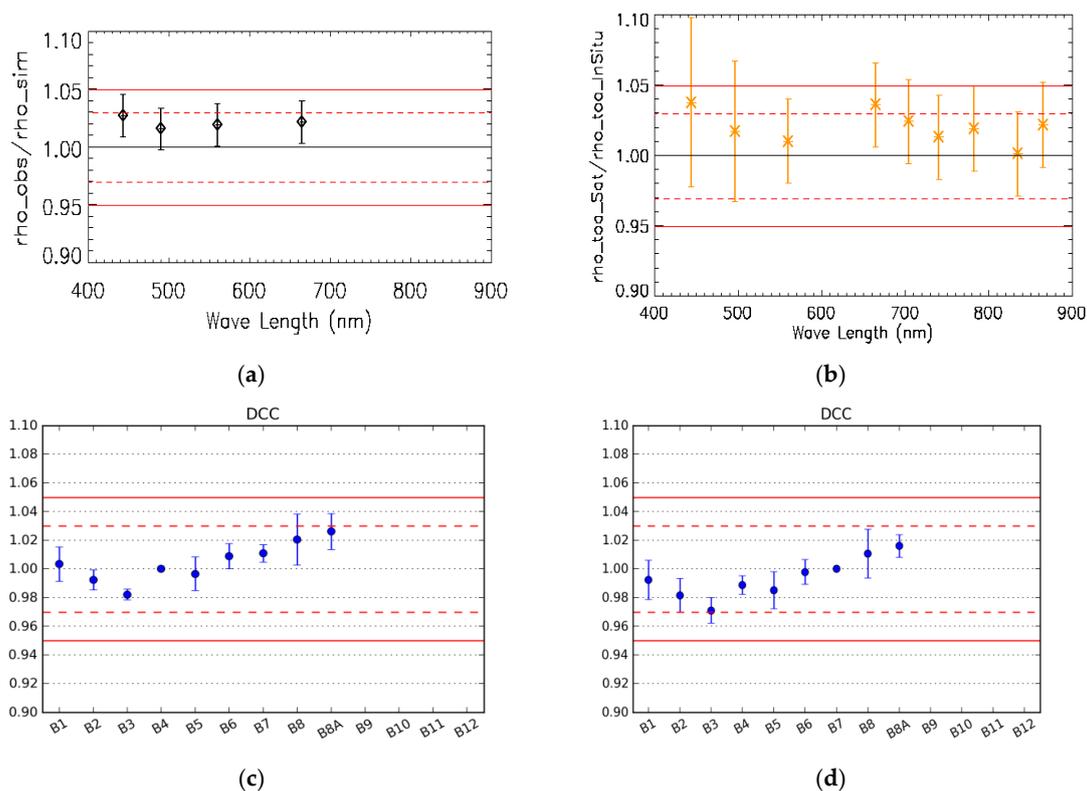

Figure 28. Calibration results over (**a**) Rayleigh Scattering; (**b**) ground-based measurements; (**c**,**d**) Deep Convective Clouds. The dashed (resp. solid) red lines represent the 3% goal specification (resp. 5% threshold specification). The reference band is B04 and B07 for (**c**,**d**) respectively. Error bars indicate the estimated uncertainty for Rayleigh and ground-based measurements and the standard deviation for the DCC method.

Table 10. Vicarious calibration coefficients estimated from Rayleigh and ground-based reflectance methodology and the associated uncertainty.

S2A/MSI Bands	Wave Length (nm)	Rayleigh		DCC		In-Situ	
		Vic. Cal. Coeff.	Uncert. (%)	Vic. Cal. Coeff.	Std. Dev. (%)	Vic. Cal. Coeff.	Uncert. (%)
B01	443	1.028	1.8	1.000	1	1.048	6
B02	490	1.024	1.8	0.990	1	1.028	5
B03	560	1.023	1.8	0.980	<1	1.020	3
B04	665	1.021	1.8	–	–	1.046	3
B05	705	NA	NA	1.000	1	1.034	3
B06	740	NA	NA	1.010	1	1.023	3
B07	783	NA	NA	1.010	1	1.029	3
B08	842	NA	NA	1.020	2	1.011	3
B8A	865	NA	NA	1.030	1	1.031	3
B09	945	NA	NA	NA	NA	0.994	3
B10	1375	NA	NA	NA	NA	NA	3
B11	1610	NA	NA	NA	NA	1.053	3
B12	2190	NA	NA	NA	NA	1.091	3

Deep Convective Clouds Method Results

This method was used for the first time for a high-resolution satellite with Sentinel-2A. Transposing the method from low to high resolution has required extensive tests, especially about the extraction of valid products. Only 22 images were considered for this analysis, since the acquisitions over the Maldives only started at the end of August 2015 and over a limited area. Out of these 22 total images that were acquired by Sentinel-2A, 7 presented cloud features, but only one image accounted for 68% of the total valid areas. The results presented here might therefore contain some natural variability, and some more cloudy products could help confirm them. However, despite so few available products, the results presented in Table 10 are consistent, with all VNIR bands within 3% of the reference band. This is compliant with the mission requirements. To make sure there is no artefact in the method, a different reference band has been chosen (B04 or B07): the results show the same inter-band pattern so they are consistent.

Note that B03 may be off the other bands because it is very sensitive to the ozone content which is derived from meteorological exogenous data, so having only one major DCC product may create a bias there.

Bands B08 and B8A are also in the limit region of the method because the reflectance of the cloud starts to become sensitive to the cloud particle type, so their relatively different behaviour is not surprising.

The method is also less limited by geometric conditions than the calibration method over molecular scattering, so results can be derived for the whole field of view of the instrument. The distribution of valid measurements presented in Figure 29 shows an increased number of points in the centre of the FOV. This is only due to the cloud distribution within the selected images. More acquisitions would increase the number of measurements per detector and consolidate the behaviour of the calibration coefficients within the FOV. Considering that Detector 1 corresponds to West edge of the swath, Figure 30 shows a slight relative decrease on the right (East) side of the images for B01 (~1–2%) and B02 (<1%). If this tendency were confirmed, it would help improve the consistency of the Rayleigh method with other methods, since the distribution of valid measurements over the oceans is biased on the left (West) side of the FOV. In conclusion, the 3% goal inter-band specification is confirmed with only a few products, which confirms the power of this method.
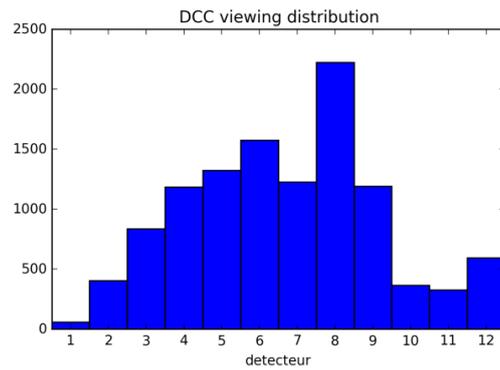
Figure 29. Distribution of valid measurements over DCC sites as a function of detector number. The detector-1 corresponds to the West edge of the swath.
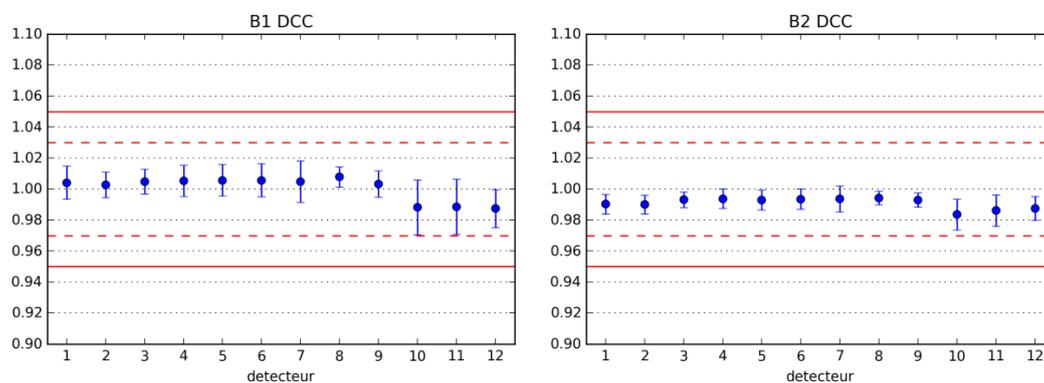

Figure 30. B01 (**left**) and B02 (**right**) inter-band calibration coefficients with respect to B04 as a reference band as a function of the detector number, based on Sentinel-2 images calibrated using the diffuser.

Ground-Based Reflectance Measurements Results

In this analysis, the reference is the TOA normalized reflectance reconstructed by the University of Arizona team using ground and atmosphere radiometric measurements. It is worth to note that the dataset have been specifically generated for the spectral bands of S2A/MSI using hyperspectral reference profiles of the site. Table 10 displays the vicarious calibration coefficients (A_b) estimated for S2A/MSI. Almost all the VNIR bands show gain less than 3%, except B01 and B04 where the gain is about 5%. Both SWIR bands display gain coefficient higher than 5% (not shown). The uncertainty of the method estimated by the University of Arizona team is of the order of 3% for most bands, and up to 6% for B01 and B02 due to a higher impact of atmospheric effects [27,28]. The results from the three methods show high consistency and good agreement.

4.3.3. Multi-Temporal Relative Radiometry Vicarious Validation

This section reports the assessment of S2A/MSI sensor performance in term of radiometry measurements, which aims to identify relative biases in the radiometric calibration and, thereby, to correct for biases if/where necessary. Moreover, the long-term trends in S2A/MSI sensor measurements are assessed to monitor the sun diffuser aging along the mission life. The method uses pseudo-invariant calibration sites (PICS) following [30], and is implemented in DIMITRI (dimitri.argans.co.uk) [29].

4.3.3.1. Pseudo-Invariant Calibration Site Methodology

Desertic PICS have been exploited since decades for multi-temporal monitoring and cross-calibration in the solar spectral range [31]. PICS algorithm aims to simulate the TOA reflectance in the visible to near-infrared (VNIR) spectral range over pre-defined desert sites (e.g., CEOS PICS

sites [32]). The first step of this method consists on building a reference reflectance model for the selected site, and the model calibration using TOA measurements from a reference sensor (MERIS in [29]). TOA measurements are propagated to the surface using an inverse Radiative Transfer simulation. A database of BOA measurements with various acquisition geometries (Sun and viewing zenith angles and relative azimuth angle) is built. The database is used to fit a four-parameter BRDF model (for each spectral band). A hyperspectral model is constructed using spectral interpolation.

The second step consists on the simulation of the observed TOA measurements (e.g., S2A/MSI) using the reference BRDF model and considering the observation geometry. Ozone and WV content is retrieved from the European Centre for Medium-range for Weather Forecasts (ECMWF), while a constant aerosol optical thickness is assumed. The resulting hyperspectral signal is then convolved with the spectral response of the instrument to produce the simulated TOA reflectance of the sensor under test. The model is 'calibrated' over each PICS site using 4 years of MERIS observations between 2006 and 2009 inclusive. Then it has been validated over the whole archive of MERIS 3rd reprocessing campaign (2002–2012) [29]. The method allows the performance of multi-temporal analysis and comparison of multiple sensors on the same site.

4.3.3.2. Sites Selection

Reference [33] identified 20 desert sites in Africa and Arabia, which are characterized as homogeneous and suitable for vicarious calibration. The authors selected site-size of 100 km × 100 km with relative spatial heterogeneity of about 3%. Six of these sites were recognized by the Infrared and Visible Optical Sensors (IVOS) working group from the Committee on the Earth Observation Satellite (CEOS). The six CEOS-PICS sites (Table 11) are used for the PICS calibration. It is worth to note that due to the high spatial resolution of MSI L1C products, and for practical reasons, these sites are subsampled to reduce the size into 20 km × 20 km and 30 km × 30 km (Figure 31). To keep the properties of the sites, the small ones are selected as a subset of the standard one [34]. Figure 31 displays RGB quick-looks of the six desert test sites used for this study.

Table 11. Definition of the desert sites used for this analysis.

N°	Name	Latitude (°)		Longitude (°)	
		Min	Max	Min	Max
003b	Algeria 3	29.82	30.82	7.16	8.16
005b	Algeria 5	30.52	31.52	1.73	2.73
031b	Libya 1	23.92	24.92	12.85	13.85
034b	Libya 4	28.05	29.05	22.89	23.89
038b	Mauritania 1	−9.8	−8.8	18.8	19.9
039b	Mauritania 2	−9.28	−8.28	20.35	21.35
031a	Railroad Valley Playa	38.495	38.505	−115.685	−115.695
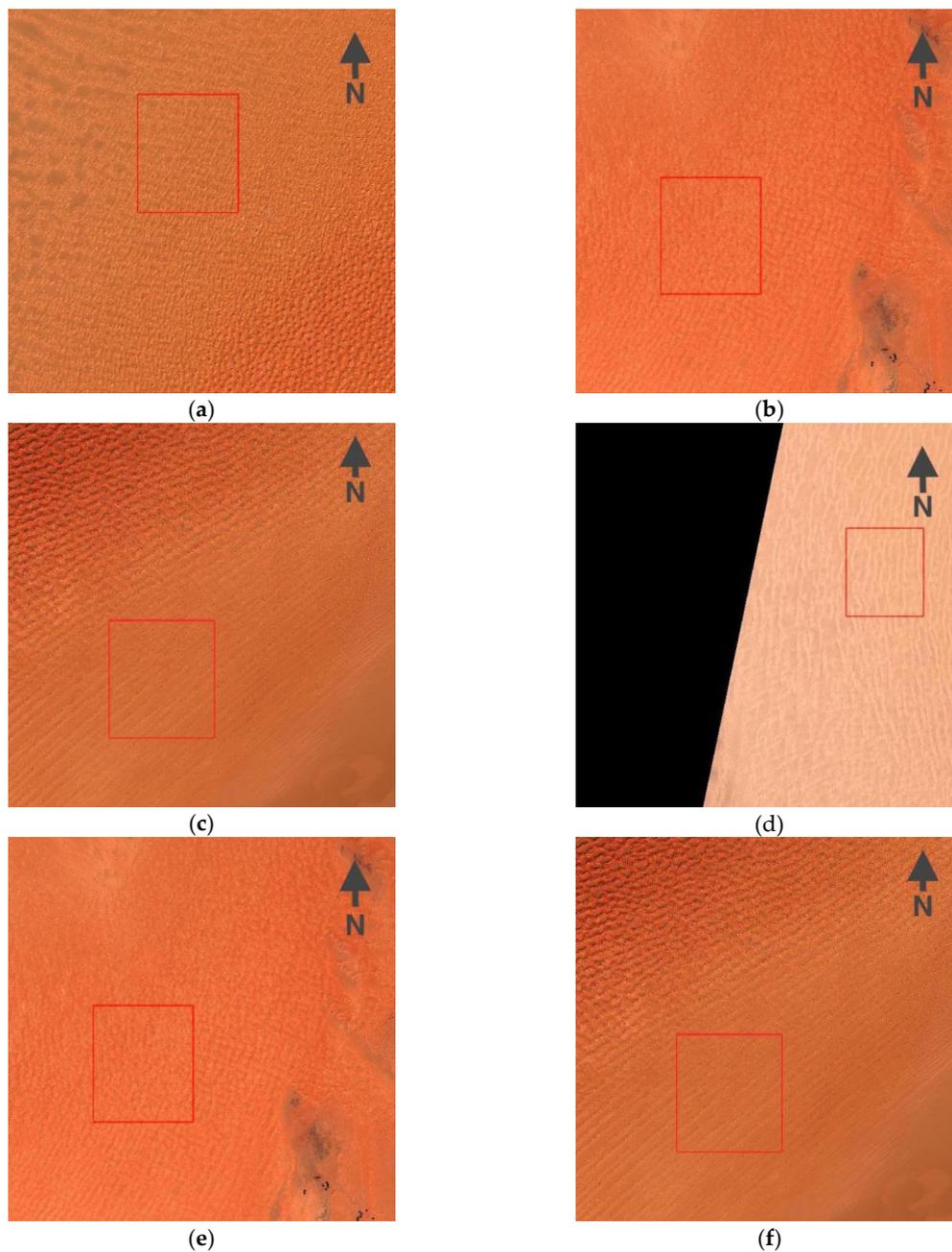
Figure 31. RGB Quick-Looks from S2A/MSI over (**a**) Algeria3, (**b**) Algeria5, (**c**) Libya1, (**d**) Libya4, (**e**) Mauritania1, and (**f**) Mauritania2 sites. Red squares indicate the region of interest (ROI).

Note that the surface BRDF is computed considering the full size of the test site. Thus the impact of the site size on the BRDF model has been computed over the six test sites and correction is then applied when needed over the VNIR spectral range.

4.3.3.3. Results and Analysis

As mentioned above, this method considers the VNIR spectral range only [29,30]. Moreover, the gain coefficients are, in fact, relative calibration coefficients as MERIS is considered as a reference to compute the surface BRDF of the test site. After the ingestion of the whole datasets from S2A/MSI, the ratio R_b, defined as observed/simulated (ρ_{obs}/ρ_{sim}) reflectance, is then computed over the six desert sites (Table 11). Figure 32 shows time series of R_k from S2A/MSI bands B01, B03, B07 and B8A over

the six PICS sites over the period June 2015—July 2016. The gap in the time-series is mainly due to the reprocessing campaign (on-going), consequently, the computed trends are clearly impacted over the first half-year of the data acquisition. In general, all the ratios are close to 1 and within ±5%, which attests a clear consistency between the six PICS. One may observe that short wave lengths (e.g., B01) show more scattered ratios than the long ones. This can be easily seen on the standard deviation on the top of each plot. This effect has been observed by [29] and could be attributed to the low surface reflectance and the high contribution of atmospheric signal. However, a stable temporal evolution of the sensor can be seen over these time-series where trend values are less than 1 ± 0.1% per year, within the mission requirements. Any reliable trends of the sensor would need more than two years of acquisitions (J. Barsi, personal comm.). While the uncertainty of the method is estimated to be about 5% (except for the absorption bands B05), the standard deviations over all sites (Figure 32) represent less than 3% of the ratios, attesting the good accuracy of the sensor calibration. Even if these ratios are in good agreement with the previous results over shorter period [30], the trend values are clearer now due to the longer time-series of S2A/MSI acquisition. This confirms the statistical nature of the used PICS methodology.

The temporal average of the ratios for each band over all the four test sites is summarized in Figure 33. Again, the ratio values are close to one over all the VNIR spectral range, and the biases are found to be better than ±3% except for B05, which shows high ratio (about 5% error and 11% uncertainty); this is due to the methodology limitation to simulate accurately the TOA-reflectance for bands with significant oxygen and water vapour absorption. However, these results illustrate the good quality of S2A/MSI calibration in agreement with [35]. The results compare well with those obtained by the French Space Studies Centre (CNES) over the same PICS and the same commissioning period, but using independent desert-calibration methodology [36].
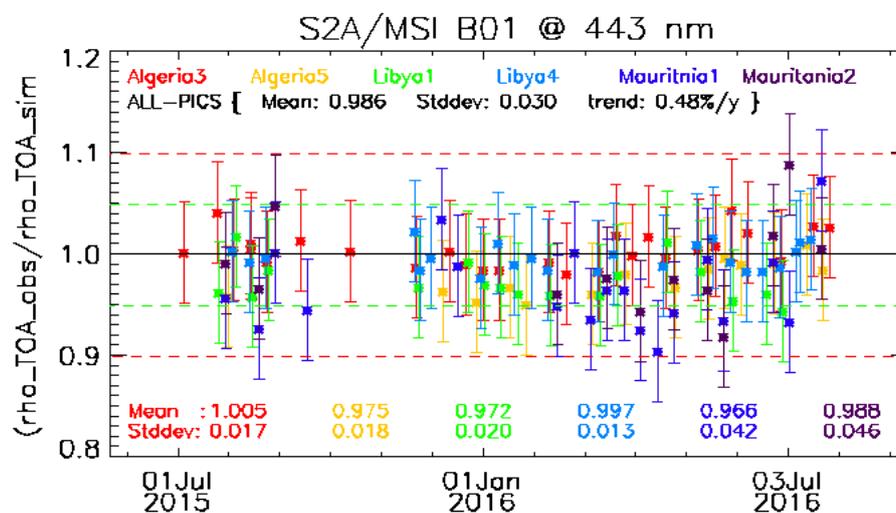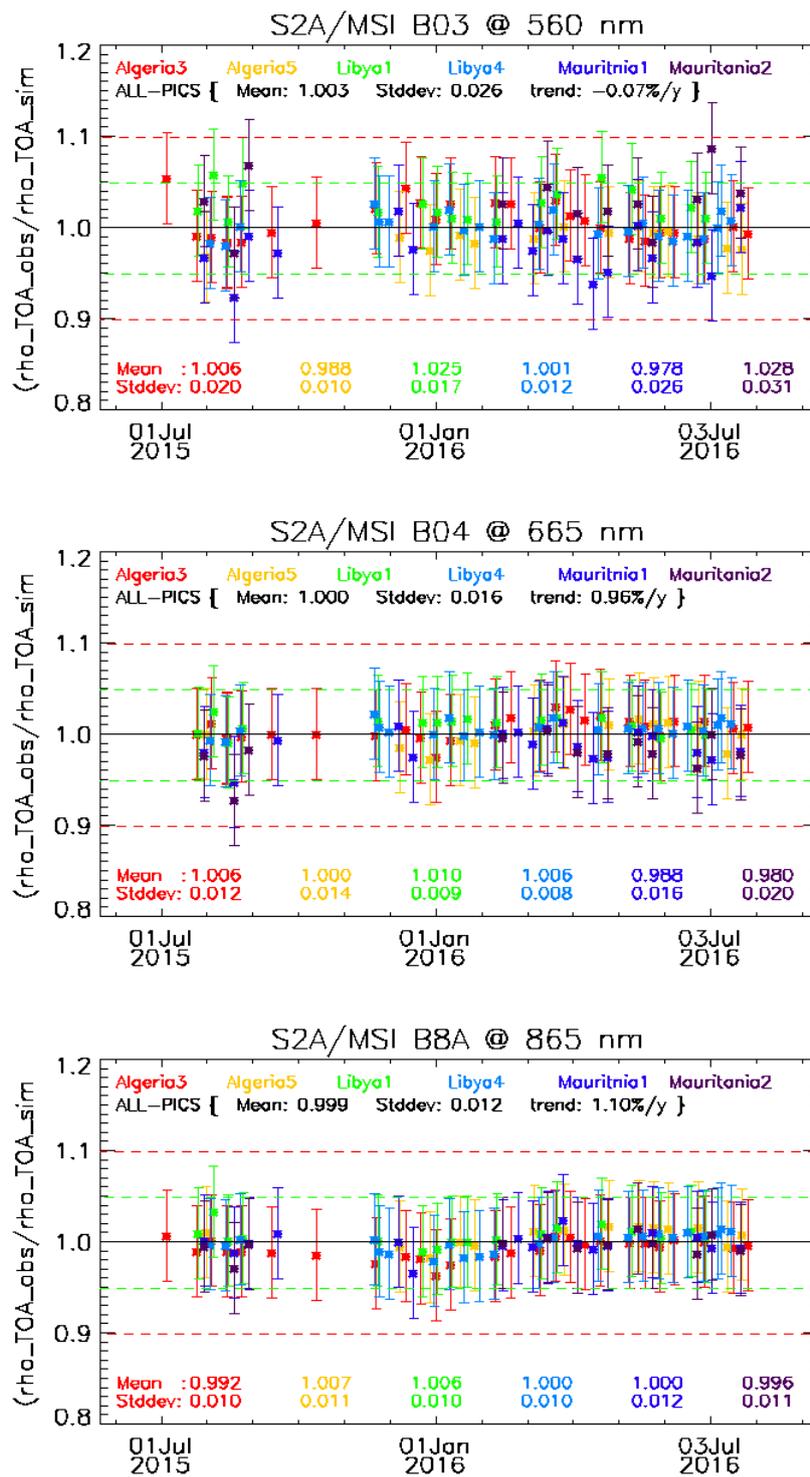

Figure 32. *Cont.*

Figure 32. Time series of TOA reflectance ratio R_b from S2A/MSI over (red) Algeria3, (yellow) Algeria5, (green) Libya1, (light-blue) Lybia4, (dark-blue) Mauritania1 and (purple) Mauritania2 for (from top to bottom) B01, B03, B04 and B8A over the commissioning period. The green (resp. red) dashed lines represent the 5% (resp. 10%) threshold specification. Error bars indicate the estimated uncertainty for the pseudo-invariant calibration sites (PICS) method.
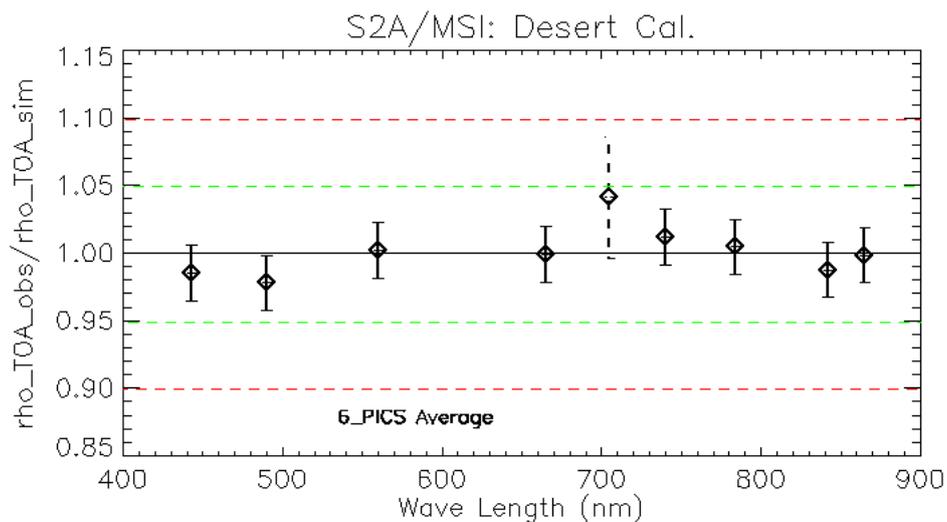
Figure 33. Temporal average of the ratio of observed TOA reflectance to simulated one from S2A/MSI over the four-PICS as a function of wavelength. B05 result is shown with black dashed line due to significant gaseous absorption. The green (resp. red) dashed lines represent the 5% (resp. 10%) threshold specification. Error bars indicate the estimated uncertainty for the PICS method.

4.3.4. Absolute Radiometry Cross-Mission Inter-Comparison

The goal of this section is to assess the performance of S2A/MSI versus a reference sensor, the Operational Land Imager (OLI) on board LANDSAT-8. The aim is to identify radiometric calibration differences between those two sensors by comparing their observations made over the same natural target, at TOA, at the same time, and under the same geometrical configuration. The so-called sensor-to-sensor inter-calibration methodology implemented in DIMITRI is used for this task. In addition, the PICS methodology is used to monitor the performance and the temporal evolution of the radiometry measurements of both sensors and then to weight S2A/MSI to LANDSAT8/OLI (as reference) for similar spectral bands over the commissioning period.

4.3.4.1. Methods

Sensor-to-Sensor Inter-Calibration

The methodology is based on the identification of directly comparable near-simultaneous TOA reflectance measurements from two space sensors made under similar observational geometries. The methodology does not require neither atmospheric nor BRDF correction to radiometrically compare two sensors. It assumes the TOA reflectance angular distribution (1) obeys the principle of reciprocity and (2) is symmetrical with respect to the principal plane [37]. Temporal and angular match-up criteria as well as the allowable cloud cover cover can be defined by the user. The strictness of the angular match between two concomitant sensors observations, 1 and 2, is controlled through the Angular Matching Criteria (AMC) parameter which is defined as:

$$AMC = \sqrt{\left([\theta_{s1} - \theta_{s2}]^2 + [\theta_{v1} - \theta_{v2}]^2 + \frac{1}{4}[|RAA_1| - |RAA_2||]^2\right)} \tag{12}$$

where: θ_s and θ_v are the solar and viewing zenith angles respectively, and RAA the Relative Azimuth Angle for sensors 1 and 2. Comparisons are done between similar spectral bands between two sensors. Then the ratio of TOA reflectance from the sensor under test to the one from the reference sensor is computed. The comparison is done over three desert sites, Algeria3, Libya4 and Railroad Valley Playa (Table 11).

After the ingestion of L1C and L1T from MSI and OLI respectively into DIMITRI, the dataset is automatically cloud-screened and the following parameters are archived: the reflectances in all spectral bands, the Sun azimuth and zenith angles (SAA, SZA) and the viewing azimuth and zenith angles (VAA, VZA). These parameters are averaged over all pixels included in the considered site. To maximize the number of the match-ups (so-called doublets), are selected all acquisitions that satisfy the condition AMC <15, <45 and <55 for Algeria3, Libya4 and Railroad Valley Playa sites respectively and maximum time lag (time window) between sensing times of 11 days. The spectral response of the seven reference bands used for the intercomparison is depicted in Figure 34. Except bands B02 and B04, which display a shift of about 10 nm, all the considered bands show a rather comparable relative spectral response (RSR), particularly in term of band centre. Hence no correction of the shift is considered in this analysis; nevertheless, its impact has been computed and added to the uncertainty analysis following [38].
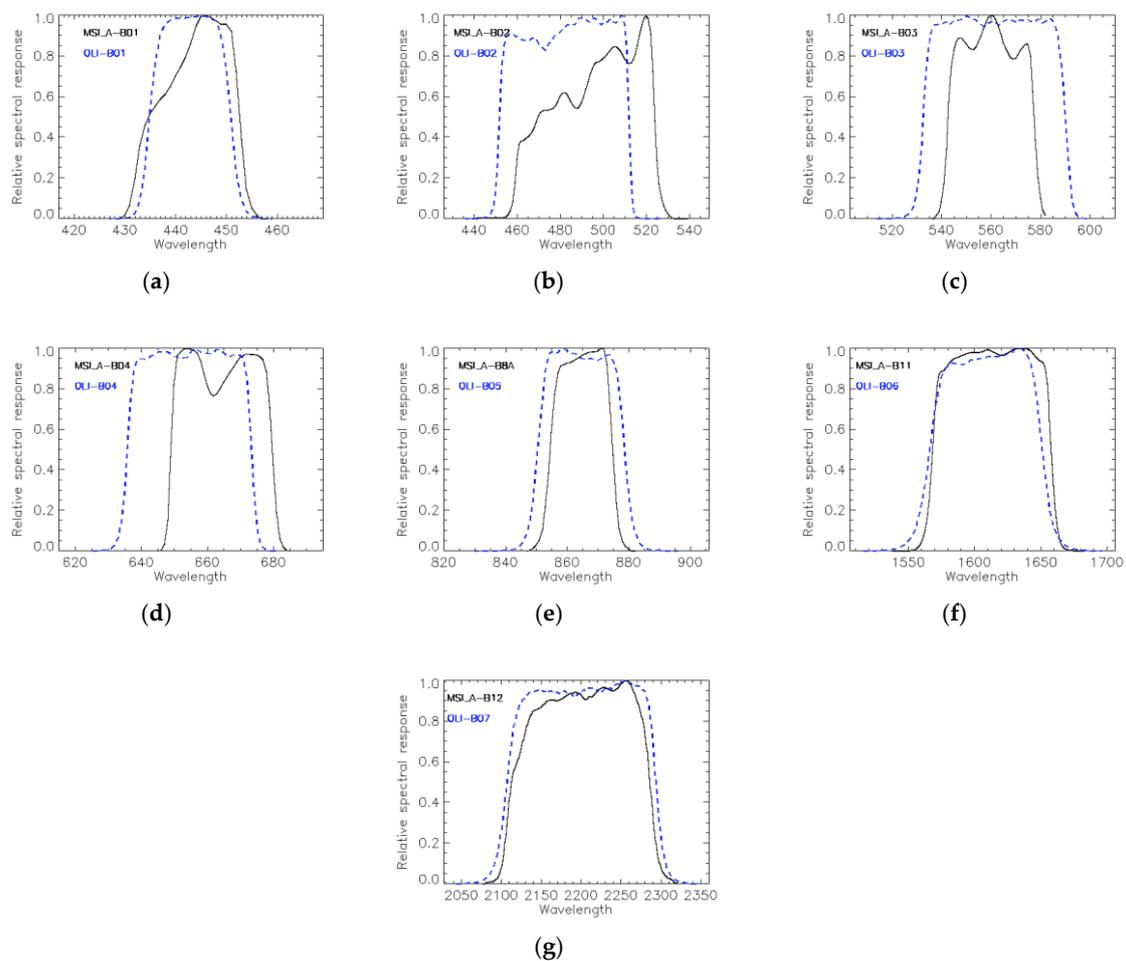

Figure 34. Relative spectral response of (solid-black) S2A/MSI and (dashed-blue) LANDSAT-8/OLI for bands (**a**–**g**) 443, 490, 560, 665, 865, 1610 and 2190 nm respectively.

Pseudo-Invariant Calibration Site Methodology

In addition to the sensor-to-sensor inter-calibration method, the PICS method described above (Section 4.3.3.1) is used. As mentioned above, in the cross-calibration of two sensors, simultaneous acquisitions are required in order to minimize the geometric effect. Therefore, the use of PICS method allows a wider applicability of such comparison through a bidirectional and spectral characterization of the surface reflectance of the site. Nevertheless, the PICS method is limited to the visible to

near-infrared (VNIR) spectral range. It is worth to keep in mind that the gain coefficients obtained from PICS method are relative when the previous method gives absolute ones.

4.3.4.4. Results and Analysis

Sensor-to-Sensor Inter-Calibration Results

By applying the previously described geometrical and temporal criteria to identify doublets of reflectance from S2A/MSI and LANDSAT8/OLI, 12, 4 and 5 doublets have been selected over 40, 36 and 15 acquisitions for Algeria3, Libya4 and RRVP sites respectively. Figure 35 displays the ratios between the observed TOA reflectance from the sensor under calibration (ρ_{cal}) and the TOA reflectance from the reference sensor (ρ_{ref}). In general, the comparison shows remarkable consistency over all the test sites for all the spectral ranges. The maximum discrepancy over the average gain coefficients is observed for bands B02 and B04. This discrepancy could be related to the shift in the RSR as shown in Figure 34. Libya4 results show larger bias over all the spectral ranges, which could be explained by the small number of matchups. However, the averaged ratios show a good agreement between both sensors with gain coefficients better that 2.5%. These results are found to be of the same order of magnitude as [36] when comparing similar bands from MODIS-A and MERIS over about 20 PICS sites. The estimated uncertainty over all the considered doublets is about 3% for bands B01 and B02, while uncertainties are found less than 2.5% for the other bands (B03, B04, B8A, B11 and B12). The uncertainty estimation is in good agreement with the estimated uncertainty of about 3% by [37] for MODIS-A, AATSR and MERIS intercomparison. In fact, the uncertainty difference between the visible bands and the SWIR bands is most likely related to the decrease of the surface reflectance and the increase of the atmospheric reflectance in the visible domain, as the atmospheric reflectance can reach about 50% of the TOA reflectance at 443 nm band. These results support the good quality of S2A/MSI radiometric calibration knowing the high quality of the radiometry calibration of LANDSAT8/OLI [27,28,39]. It is worthwhile noting that there are several factors contributing to the discrepancy between both sensors: (1) the maximum time lag of 11 days induces day-to-day variability; (2) the assumption of symmetry of the TOA BRDF with respect to the principal plane is valid as far as there is no azimuth-dependent structure on the surface; (3) the difference in the RSR of both sensors induces difference on TOA reflectance of about 1%, see [37].
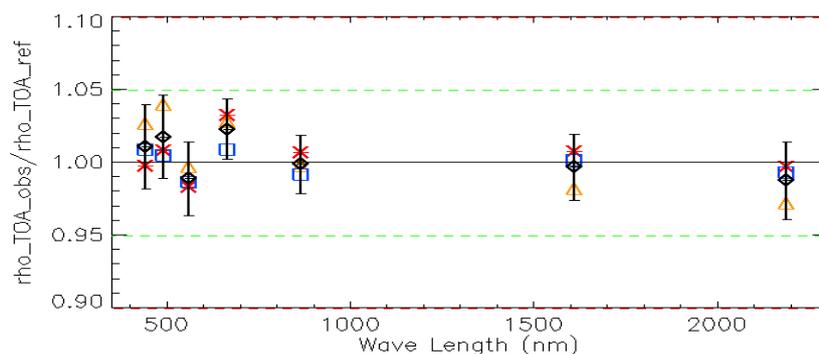

Figure 35. TOA reflectance ratio defined as S2A-MSI/LS8-OLI measurements derived from extracted doublets over (red asterisk) Algeria3, (orange triangle) Libya4, (blue square) Railroad Valley Playa site (RRVP) and (black diamond) the average over the three test sites as a function of the wavelength. The green (resp. red) dashed lines represent the 5% (resp. 10%) threshold specification. Error bars indicate the estimated uncertainty.

PICS Methodology Results

As mentioned above, this method considers all time-series of both sensors over a predefined period. In other words, it is not restricted to the matchups only but it considers the VNIR spectral range only. Thus the gain coefficients are, in fact, relative calibration coefficients as MERIS is considered as a reference to compute the surface BRDF of the test site. After the ingestion of the whole datasets from S2A/MSI and LANDSAT-8/OLI, the ratio R_b, defined as observed/simulated (ρ_{obs}/ρ_{sim}) reflectance is then computed over the six desert sites. Figure 36 shows time series of R_b from both S2A/MSI bands (B03, B04 and B8A) and LANDSAT-8/OLI bands (L03, L04 and L05) over Libya-1 site.

Both sensors display nearly the same temporal evolution, scatter and variability of the resulting R_b. In particular, the concomitant acquisitions show similar responses of both sensors (Figure 36). The results show good consistency with the sensor-to-sensor method and the conclusions of [40]. OLI ratios exhibit slightly higher variability in the blue and red bands than MSI (standard deviation is larger by 0.2%), while the NIR band of OLI shows lower variability than MSI band B8A (standard deviation is smaller by 0.3%). In spite of the acquisition-time difference and the RSR differences of both sensors, the results exhibit a good correspondence of the TOA reflectance between both sensors over the VNIR and SWIR spectral ranges in agreement with [40].

Figure 37 displays the cross-calibration results over Algeria3, Libya1, Libya4 and Mauritania2 desert sites. Algeria5 and Mauritania1 sites have been excluded from this comparison due to the short time series of acquisitions, which leads to less reliable results. Except a slight bias of 2 or 3% in bands B01 over Algeria3 and Mauritania2 and B02 over Libya4, the results are consistent over the VNIR spectral range, and no significant discrepancy can be observed. Except for B05 from MSI, the whole ratio values are within ±5%. These results demonstrate the complementarity between MSI and OLI, and show that MSI products can provide data continuity for OLI in agreement with [40].
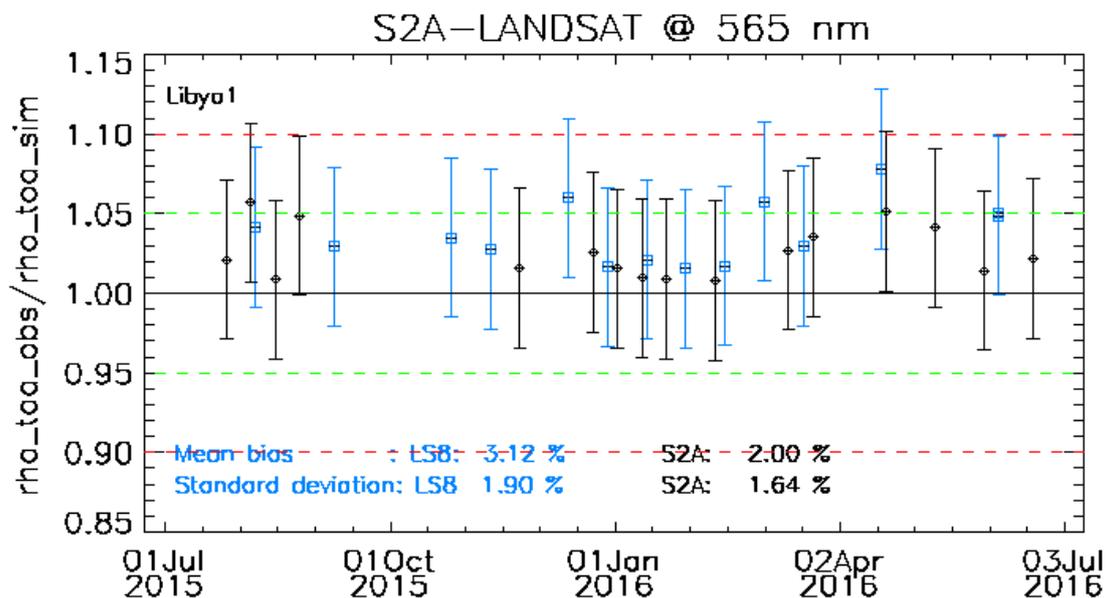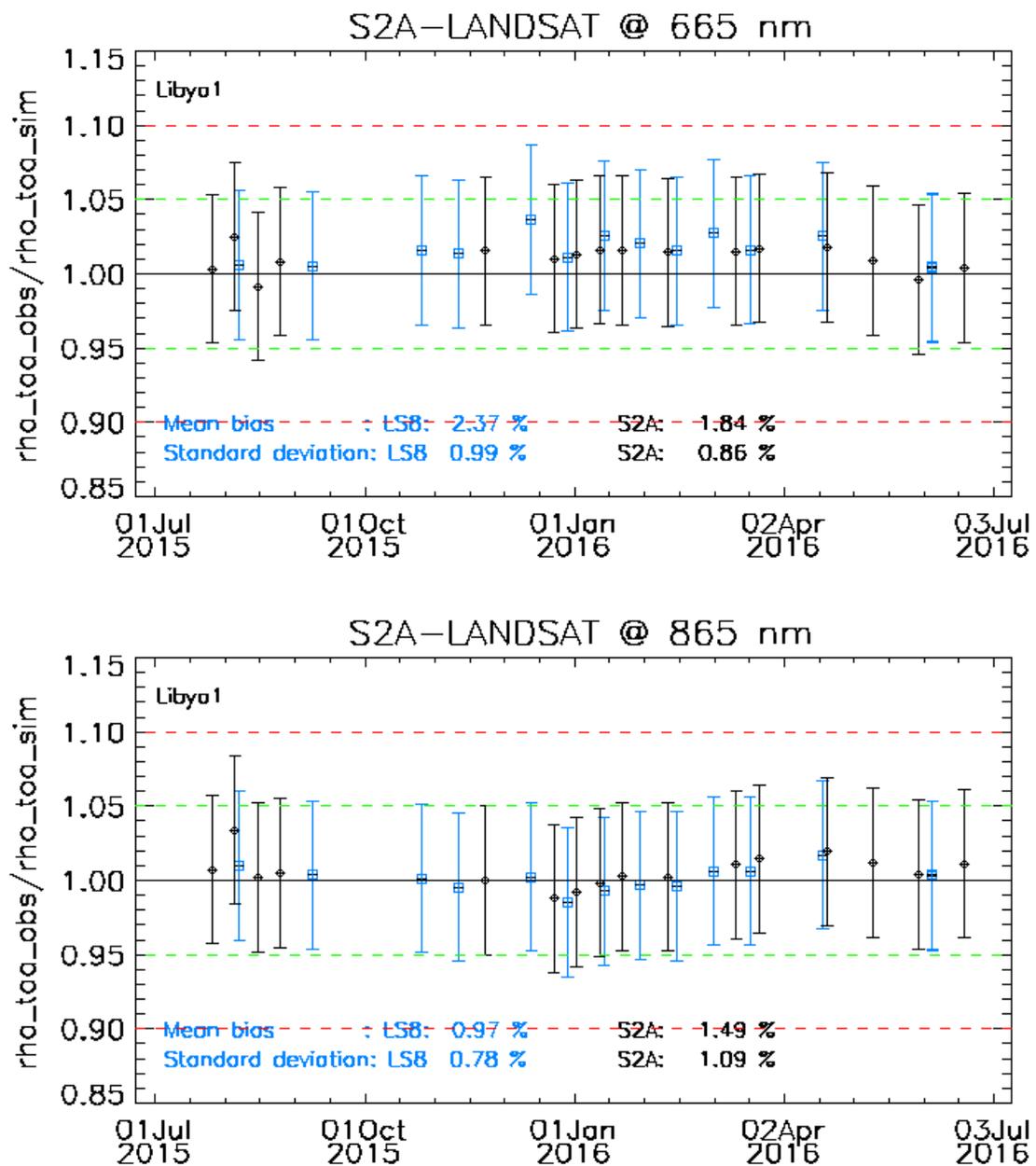

Figure 36. *Cont.*

Figure 36. Time series of TOA reflectance ratio R_b from (black) S2A/MSI and (blue) LS8/OLI over Lybia1 for (from top to bottom) B03, B04 and B8A from S2A over the commissioning period. The green (resp. red) dashed lines represent the 5% (resp. 10%) threshold specification. Error bars indicate the estimated uncertainty for the PICS method.
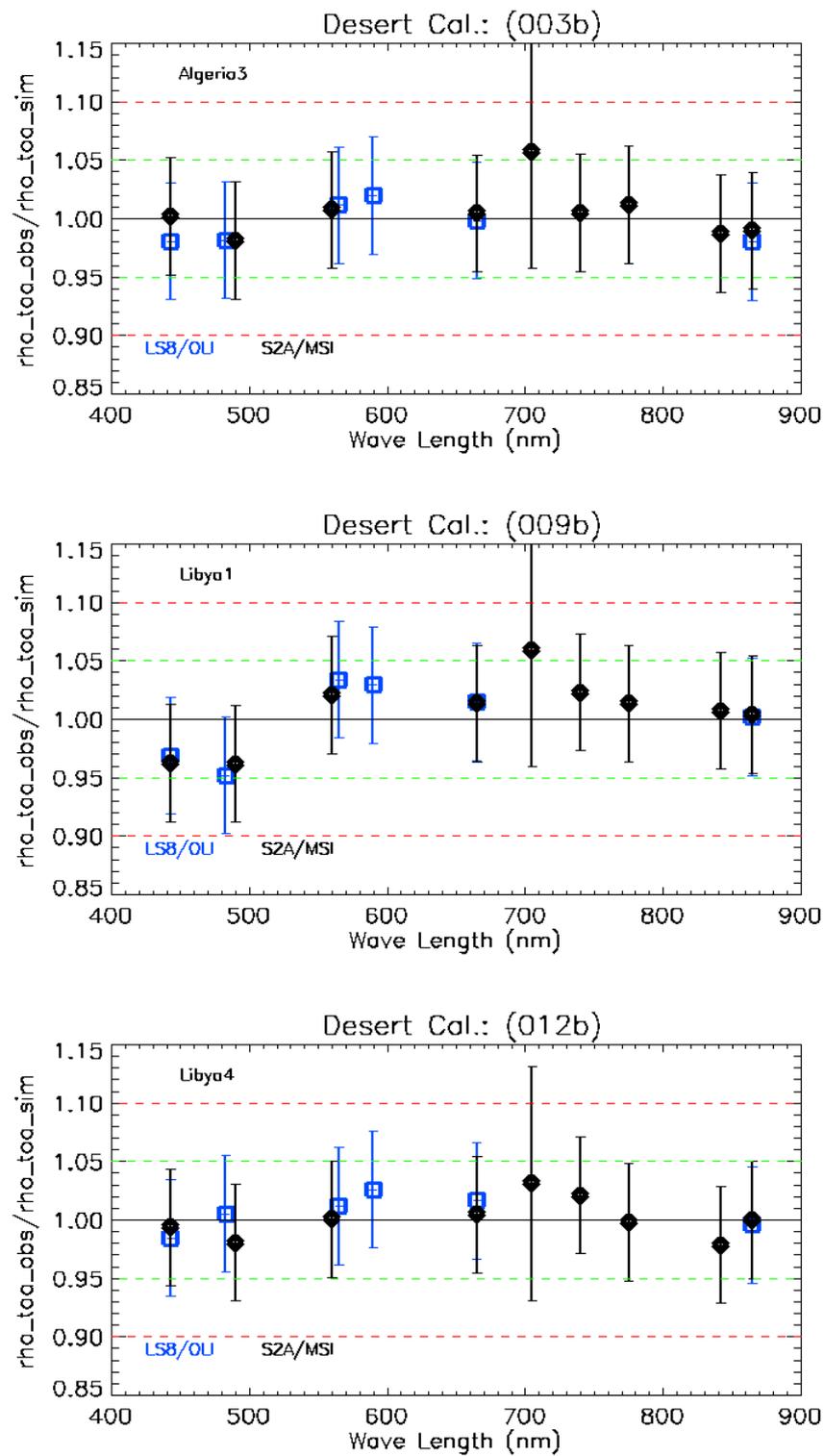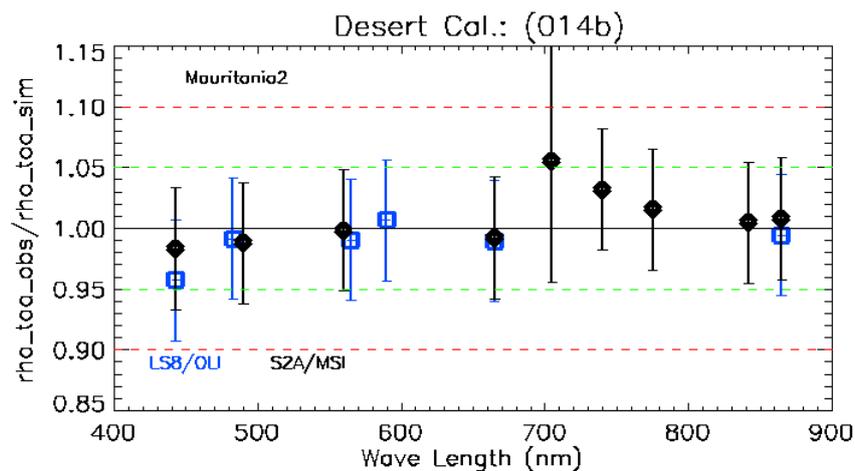
Figure 37. *Cont.*

Figure 37. Ratio of observed TOA reflectance to simulated one for each sensor (black) S2A/MSI and (blue) LANDSAT-8/Operational Land Imager (OLI) over Algeria3, Libya1, Libya4, and Mauritania2 sites as a function of wavelength. The green (resp. red) dashed lines represent the 5% (resp. 10%) threshold specification. Error bars indicate the estimated uncertainty for the PICS method.

4.3.5. Inter-band Relative Radiometric Uncertainty Validation

The sensor calibration provides absolute calibration coefficients, one per spectral band, which enables to convert the digital count at the output of the sensor into TOA radiance. For each spectral band, the TOA reflectance of the Level 1C products is directly proportional to the TOA radiance. An error on the absolute calibration coefficients may have various effects depending on its type. The Inter-band Relative Radiometric Uncertainty Validation (shortly, inter-band validation) is devoted to assess the impact of the calibration coefficients assessment between spectral bands. For a reflectance ratio, any common multiplicative error on the calibration coefficients does not affect the performance. On the contrary, some errors on individual spectral bands may "cumulate" when the ratio is used.

Thus, the inter-band validation compares the reflectance ratio between two different spectral bands computed from a L1C product to the expected TOA reflectance ratio. The key point of such assessment is the knowledge of the spectral shape of the observed landscape.

The inter-band validation was initially envisaged over snowy areas like Antarctica and Greenland. But the snow spectrum varies according to the characteristics of the snow (thickness, wet or dry, size of the snow crystals, etc.), so the inter-band validation now relies on desert sites. It will be performed twice a year.

4.3.5.1. Method and Outlook

The method relies on spectral BRDF of desert sand samples measured in a laboratory. The available samples correspond to the sites named Algeria-3, Algeria-4 and Negev.

For each acquisition over these sites, a tool extracts from the L1C product, for each spectral band, the mean value of the TOA reflectance, the solar angles and the viewing angles.

The site coordinates, the date and the viewing angles, climatology data and the available spectral BRDF feed a radiative transfer code, such as 6S [41] or MODTRAN [42], which provides the expected TOA reflectance for each spectral band.

The reflectance ratios of pairs of spectral bands are computed for both cases and compared. The implementation of the method is in progress and results are expected in the coming months.

4.3.6. Signal-to-Noise Validation

The objective of this activity is to measure the temporal noise (or column noise) standard-deviation and Signal to Noise Ratio (SNR) of the instrument pixels in order to compare it to the data quality

baseline and identify potential instrument performance degradations. The method used for this assessment relies on on-board sun-diffuser device and dark product acquisitions.

4.3.6.1. Method

The SNR is a function of the mean radiance of the landscape, generally expressed as:

$$SNR = \frac{<L>}{\sigma_L} \tag{13}$$

where *<L>* is the mean of a set of radiances over a uniform landscape (the signal) and σ_L is the standard-deviation of this set of radiances (i.e., the noise).

SNR depends on the radiance level. It is usually lower for low values of radiance (dark landscape) because the relative influence of the noise is larger. For large radiances, the SNR increases as the relative influence of the noise decreases. Therefore, the SNR should be known at different radiance levels.

For that purpose a 2-parameter instrument noise model (per pixel) is established:

$$\sigma_L(p,d,b) = \sqrt{\alpha_L(p,d,b)^2 + \beta_L(p,d,b)L(p,d,b)} \tag{14}$$

where

- σ_L is the radiometric noise standard-deviation (in W/m^2/sr/μm).
- L is the radiance (in W/m^2/sr/μm).
- α_L and β_L are the noise model parameters, both in in W/m^2/sr/μm.

  ○ α_L represents the dark noise standard-deviation (at radiance 0).
  ○ $\beta_L \cdot L$ represents the variance of the shot noise (due to the particle nature of light) at radiance L.

The two noise model parameters (per pixel/band/detector) are estimated from one dark acquisition and one sun-diffuser acquisition (L0 products processed up to L1B):

- α_L is directly determined from measurement of noise standard-deviation on a dark image (L ≈ 0).
- β_L is determined from measurement of noise standard-deviation in a sun-diffuser image and by inversion of Equation (14) using the knowledge (from diffuser model) of diffuser radiance $L_{diff}(p,d,b)$ and current estimation of α_L.

Then the SNR at any radiance *L* for each instrument pixel is given by Equation (15) below:

$$SNR(p,d,b) = \frac{L}{\sqrt{\alpha_L(p,d,b)^2 + \beta_L(p,d,b).L}} \tag{15}$$

The method for assessing the temporal (or column) noise standard deviation on the two L1B products is as follows:

- On dark image:

  ○ $\sigma_L(p,d,b)$ on dark image is simply the standard-deviation of the radiance levels in one column.
- On sun-diffuser image:

  ○ $\sigma_L(p,d,b)$ on sun-diffuser image is the standard-deviation of the radiance levels in one column corrected from the sun-diffuser non-uniformity (BRDF and solar angles effects) known from sun-diffuser simulated radiance (see Equation (2) in paragraph 4.1.2).

4.3.6.2. Results

S2A SNR is well inside the requirements (>20% margin) for all bands and appears to be very stable over the period analysed (since January 2016) as shown in Figures 38 and 39.

Spectral Band	Lref	SNR SPEC @ Lref	SNR@Lref
B1	129	129	1361
B2	128	154	214
B3	128	168	249
B4	108	142	230
B5	74.5	117	253
B6	68	89	220
B7	67	105	227
B8	103	174	221
B8A	52.5	72	161
B9	9	114	227
B10	6	50	387
B11	4	100	158
B12	1.5	100	166

Figure 38. Sentinel-2 reference radiance L_{ref} (W/m^2/sr/µm), Signal-to-Noise Ratio (SNR) requirement at reference radiance and current SNR measurement on sun-diffuser at reference radiance.
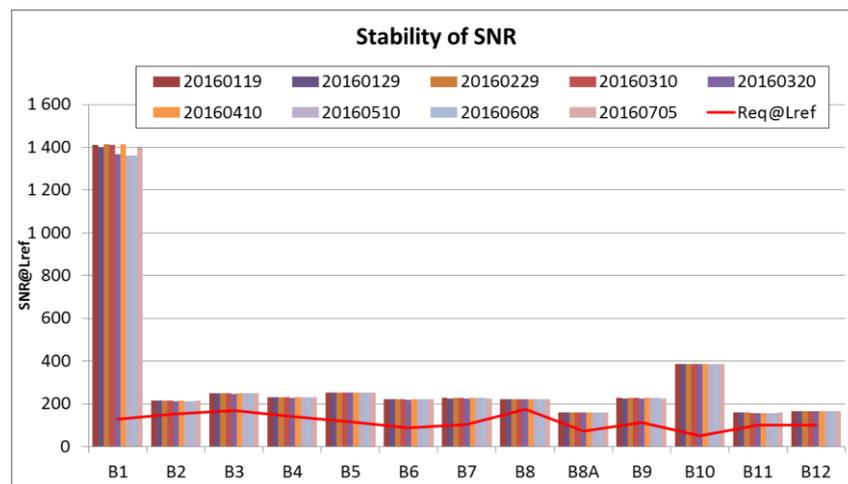

Figure 39. Average SNR at reference radiance L_{ref} measurements (per band) since January 2016.

4.3.7. Modulation Transfer Function Validation

Modulation Transfer Function (MTF) measurement aims at quantifying sensor spatial resolution. MTF may be affected by launch vibrations, transition from air to vacuum and thermal state. This leads to check the MTF value in orbit. Depending on the origin, an MTF loss can be compensated thanks to the refocusing mechanism. During the first phase of the post-launch calibration, two methods have been used for MTF assessment: the mean square methods and the linear object method. For routine period, the well-known edge method is used. The MTF measurement is an activity performed once a year.

4.3.7.1. Methods

From [43], the sensor model for spatial resolution is:

$$Image(x,y) = Landscape(x,y)*h(x,y).DiracComb(x,y) \qquad (16)$$

x varies along the line of the image,

y varies along the row of the image,

h is the impulse response in other words the Point Spread Function (PSF),

The *DiracComb* corresponds to the sampling by the detectors in the focal plane,

* is the convolution product,

. is the multiplication.

A Fourier transform enables to reach the frequency domain:

$$ImageSpectrum(f_x, f_y) = LandscapeSpectrum(f_x, f_y).H(f_x, f_y)* DiracComb(f_x, f_y) \qquad (17)$$

fx is the spatial frequency for the line (or across-track) direction,

fy is the spatial frequency for the row (or along-track) direction,

H is the Transfer Function.

The MTF is the modulus of the transfer function *H*.

The convolution by the *DiracComb* in the Fourier domain corresponds to a replication of the pattern *LandscapeSpectrum(fx, fy).H(fx, fy)*.

MTF measurement methods aim at:

- Choosing a landscape so that the term *LandscapeSpectrum* can be known,
- Managing the replication in order to avoid aliasing effect that corresponds to mixing, for some frequency range, of the pattern with its replica.

Reference Image Method

Unlike the other two methods described hereafter, the "reference image method" is used to provide MTF measurement in all directions and in many parts of the field of view. Given a Sentinel-2 image target, it uses an acquisition of another instrument on the same area, but at higher resolution. This reference image is resampled many times at S2 resolution, applying a different MTF each time and searching for the one that gives best resemblance with S2 product.

Five Pleiades images over Toulouse, Albuquerque, Las Vegas, Los Angeles and Dallas were selected as high resolution references. This method requires good geometrical superimposable reference and target images. The reference product is thus resampled in S2 level-1B geometry as a pre-processing step. Finally, a least square algorithm is used to compare images and determine the most probable MTF.

Bridge Method

This method uses images of bridges for MTF verification. It is inspired by method on slanted edge targets. As such, it uses an object which radiometric response is easy to model and whose response after transmission through the system is compared to its theoretical initial form.

The observed bridge has to be straight, of uniform radiometry (both bridge and background) and its orientation slightly tilted with respect to X or Y axis. This inclination is used to build an oversampled profile of the bridge acquired by the instrument, which is the profile of the real bridge after application of the system MTF. This real bridge is itself simply modelled as a rectangular function. The MTF of the system is then obtained as the ratio between the Discrete Fourier Transforms (DFT) of the oversampled profile and of this rectangular function:

$$MTF = \frac{DFT(Profile_{measured})}{DFT(Profile_{theoretical})} \qquad (18)$$

Three bridges were selected thanks to their orientations with respect to Sentinel-2 swath and to their widths: King Fahd Causeway (Saudi Arabia), Albermarle Bay Bridges (USA), Rikers Island Bridge (New York).

Edge Method

For the edge method, the landscape is chosen so that it can stand for a Heaviside function. This can be done thanks a large mosaic of fields.

To manage the replication through oversampling an inclination of the edge relative to the line or row of the image is needed. A 1D oversampled edge response is created from the 2D under-sampled image.

Thanks to this oversampling in the direction perpendicular to the edge Equation (17) becomes:

$$ImageSpectrum(f) = Heaviside(f).H(f) \tag{19}$$

which leads to:

$$MTF(f) = |H(f)| = \left| \frac{ImageSpectrum(f)}{Heaviside(f)} \right| \tag{20}$$

4.3.7.5. Results

For the edge method, MTF was deduced from an image from Maricopa field and from an image of Ross Ice Shelf in Antarctica. Due to the noise mainly due to the non-uniformity or the sides of the natural edges, a model was fitted between 0.1 fs and 0.5 fs, fs being the sampling frequency and 0.5 fs being the Nyquist frequency. An example of MTF curves is given in Figure 40.
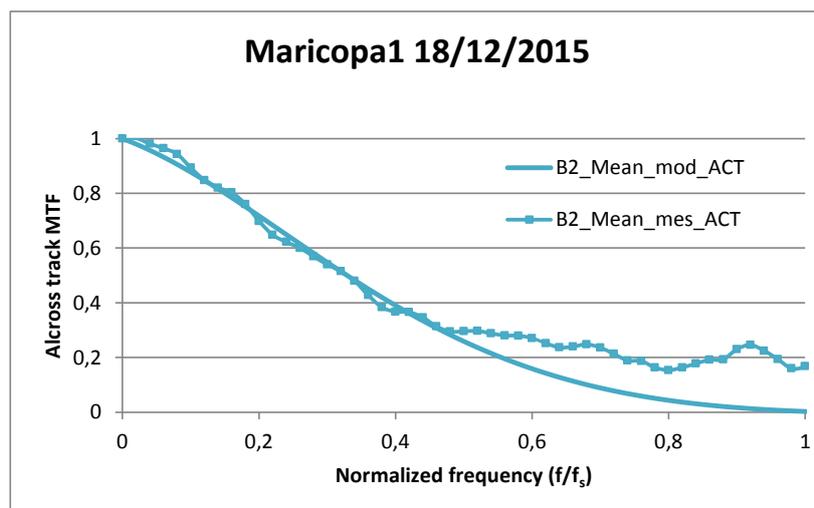

Figure 40. Modulation Transfer Function (MTF) curve for B02 band (centred at 490 nm) for the across-track direction.

The MTF values at Nyquist frequency obtained with the various methods are mixed and compared to the requirements and the expectations deduced from ground measurements in Figures 41 and 42.

Most validation measurements are in agreement. Globally, the across-track values are lower than those expected from ground measurements.

In December 2015, ESA accepted a waiver on the MTF maximum value requirement since the impact on image quality was considered acceptable. Therefore, the measured across-track MTF performance for B05 (705 nm), B06 (740 nm), B07 (783 nm) and B8A (865 nm) is not a concern. As there is no MTF below the minimum value specified, the spatial resolution reduced to MTF [44] is as expected or even better.
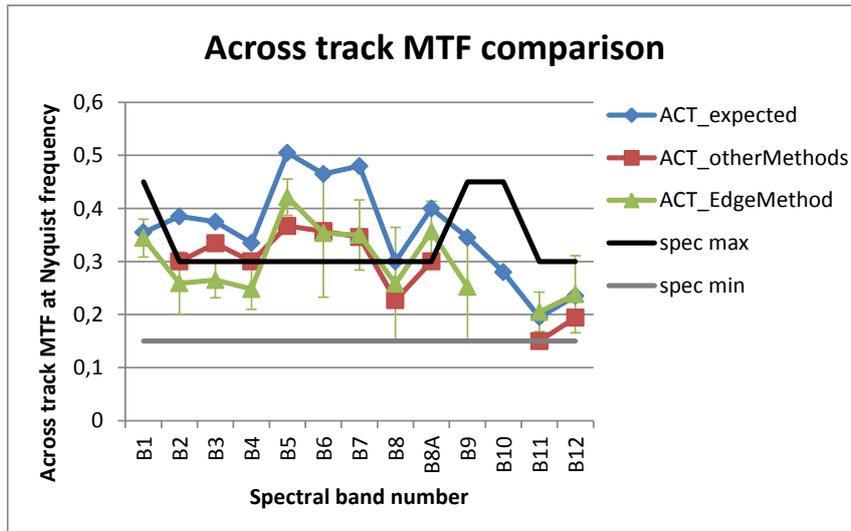
Figure 41. MTF results for the across-track direction, and MTF requirements (min value, spec min and max value, spec max).
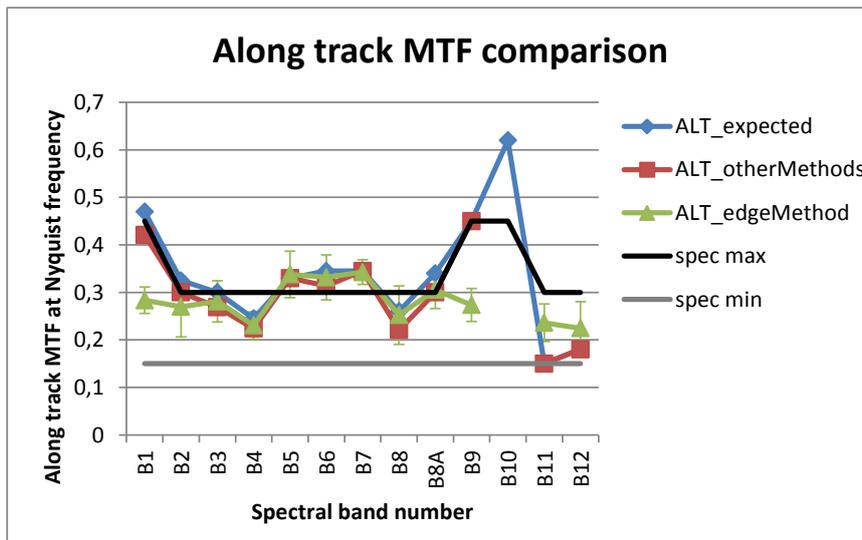

Figure 42. MTF results for the along-track direction, and MTF requirements (min value, spec min and max value, spec max).

4.4. Geometry Validation Activities

Geometric validation activities aims at assessing all geometric performances related to image quality requirements. The validation activities are performed regularly but with various time frequencies from weekly basic checks to yearly in-depth analyses, including long-term statistical and trend analysis. In case of observed performance degradation, specific calibration activities could be triggered.

4.4.1. Geolocation Uncertainty Validation

Geolocation performance is assessed both at level 1B and at level 1C with or without geometric refinement using the GRI. All results are presented in the following paragraphs. Geometric refinement being not activated in the operational processing chain at the manuscript-writing time, the results related to refined products are obtained off-line by CNES using the Ground Processor Prototype

(GPP) and should be representative of the operational products performance when refinement will be activated. In practice the geolocation performances at L1C should be similar to those at L1B level, except the small error introduced by the resampling.

4.4.1.1. General Methods and Data

Geolocation assessment is based on detection of Ground Control Points (GCP) in Sentinel 2 images by correlation with a database of accurately localised images spread over the world. The general principle is shown in Figure 43 and is composed of the following steps:

(1) Find approximate GCP location in Sentinel 2 image using product geolocation metadata (L1B physical geolocation model or L1C georeferencing metadata).
(2) Resample both images in the same geometry: Either S2 image is resampled to GCP image or GCP image is resampled to Sentinel 2 image. This is based on L1B physical geolocation model or L1C georeferencing metadata.
(3) Assess the GCP geolocation error (shift between the GCP image and Sentinel-2 image) by correlation technique.
(4) Repeat the previous steps for several GCPs in one product and several products.
(5) Filter out bad correlation points based on correlation quality criteria
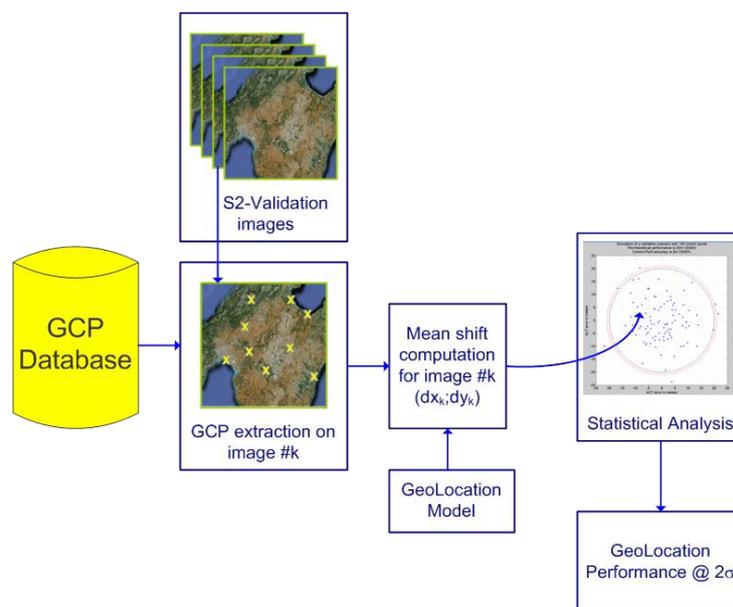(6) Compute statistics to assess performance (e.g., at 95.45% confidence level).

Figure 43. General principle of geolocation uncertainty validation using ground control points. Mean shift computation is obtained by correlation technique.

The process is performed using Sentinel-2 band B04 as reference band. The 10m bands obviously provide a better correlation accuracy.

Two different GCPs database are used for the various geolocation validations:

- A dedicated high-resolution ortho-images database with geolocation accuracy better than 5m is used by Thales Alenia Space in the frame of MPC, see world distribution in Figure 44.
- Pleiades HR database including more than 500 images with accuracy better than 5 m is used by CNES.
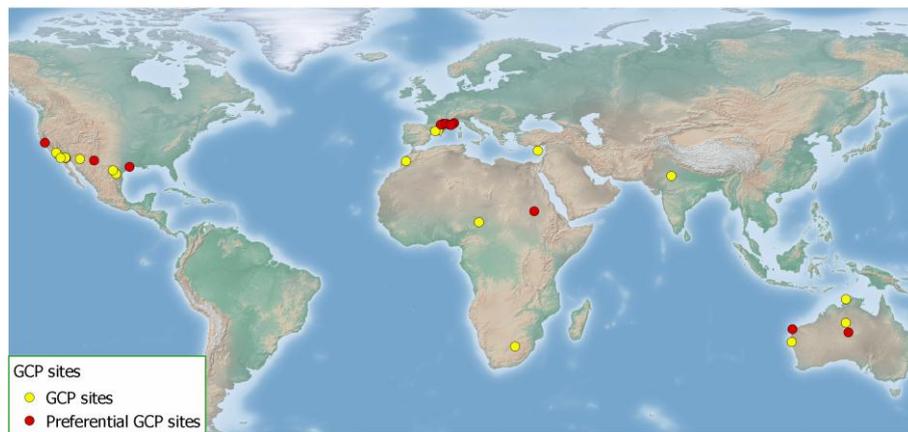
Figure 44. Map of Thale Alenia Space GCP database sites (red dots are GCP sites currently available and used as baseline for S2 validation activities).

These ground control points have to be as accurate as possible and on flat areas to limit parallax effects, shadows, etc., that would impact measurement accuracy. The quality of the measurement also depends on:

- the accuracy of matching algorithm, given the difference in acquisition date, sensors characteristics and acquisition conditions between Sentinel-2 images and the reference image,
- the number of GCPs and their distribution over the scene.

For the assessment of geolocation on refined products a special care shall be taken to use GCPs independent from the ones used for the GRI construction.

The general performance assessment principle is declined into several methods depending on the product level processed:

Validation on L1C products (refined or not):

- In step (2) GCPs ortho-images (in any cartographic projection) are resampled onto Sentinel-2 L1C tile cartographic projection.
- The error shifts are thus observed in fraction of L1C pixels directly convertible into meters or X/Y UTM coordinates. The shifts can be converted into across-track and along-track directions using knowledge of satellite acquisition direction on ground.

Validation on L1B products (refined or not):

Method 1: In step (2) Sentinel-2 L1B image is resampled on the GCPs image (in cartographic projection). The estimation of performance is then directly in metres and accounts for all contributing factors.

Method 2: In step (2) the GCP image is resampled on Sentinel-2 L1B image (in sensor geometry). Thus the geolocation performance is not estimated in metres on the ground, but in the focal plane.

In this case, the inverse location function is used: from the GCP ground coordinates, the image coordinates are estimated by means of the inverse geometric model and then compared with measured image coordinates. This provides the location performance along rows (in pixels), also known as column location performance, and the location performance along columns (in pixels), also known as row location performance. The benefit of this method lies in obtaining the performance in the focal plane and therefore in better understanding the physical phenomena. Hence pitch, roll or yaw bias, magnification or even drift in pitch or roll, can be shown.

The performance indicator to be compared to the requirement is computed as follows:

(1)　For each processed S2 product, compute the mean circular error on all GCPs inside the product.

(2)   Compute the quantile at 95.45% of the mean circular errors over all the products.

Sentinel-2 acquisition needs are segments of L1B or L1C (refined or not refined), depending on the product level analysed. They shall be distributed over the world, including various latitudes and various seasons, the accuracy of the image geolocation depending on these two parameters.

The geolocation performance assessment is by essence a long-term activity requiring at least one year of product acquisitions.

Results for Non-Refined Products

Geolocation performance on non-refined products represents the absolute location performance as provided by the system, and depends uniquely on the calibrated imaging parameters without any ground-processing improvement based on external information (such as Ground Control Points—GCP—for example).

Performances before production baseline 2.04 suffered from a yaw bias correction anomaly, particularly visible at the edge of the swath, leading to a performance of about 14.5 m CE at 95.5% confidence level on L1C products.

Since production baseline 2.04, the performances are very good on both types of products L1B and L1C, and considerably under the limit fixed by specifications of 20 m CE at "2σ". Table 12, Figures 45 and 46 present the statistics obtained on system geolocation performances for L1 non-refined products.

The 95.45% confidence level ("2σ") system geolocation performance value obtained on analysed products is 10 m for non-refined L1B and L1C products. It is compliant with the specification of 20 m at 2σ.

Table 12. Statistics on geolocation for non-refined L1B and L1C products.

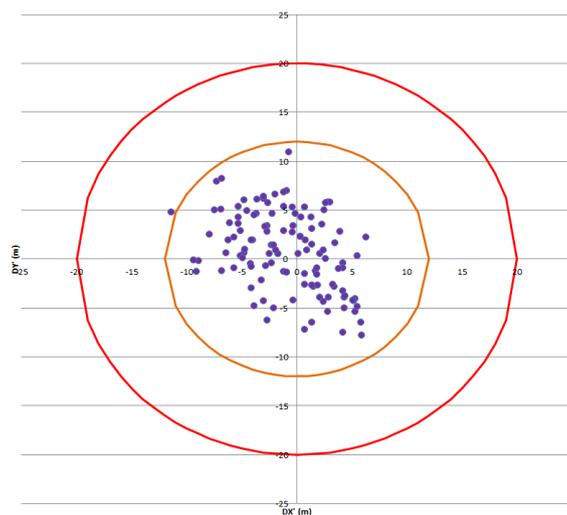| System Geolocation Performance (Circular Error)	L1B Non Refined	L1C Non Refined
95.45% conf. level	10 m	10 m
Mean value	5 m	5 m
Requirement (for non-refined products)	20 m	20 m

Figure 45. System Geolocation Performances in metres (Band 04) for L1B non-refined products. Red (resp. orange) circle is the requirement without (res. with) geometric refinement on GCPs.
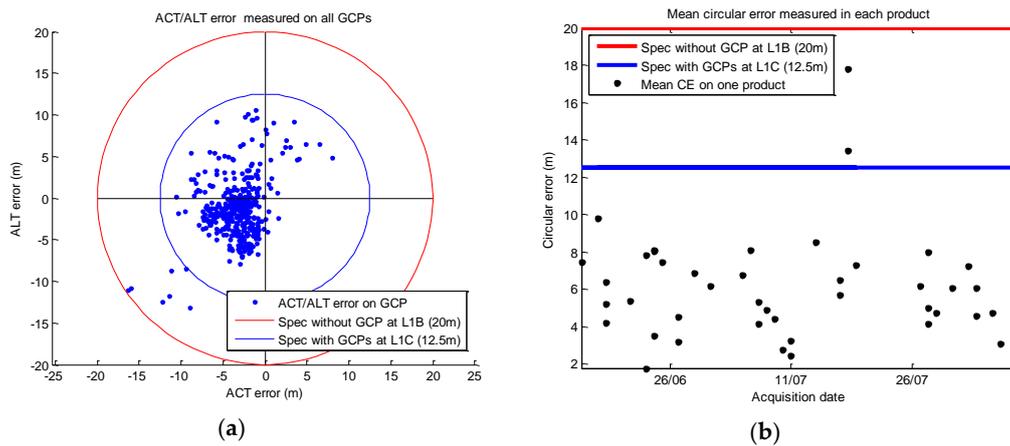
Figure 46. System Geolocation Performances in metres for L1C non-refined products. Reference band is B04. (**a**) Geolocation error (in metre) along-track and across-track measured on each GCP; (**b**) Mean geolocation error in each product processed in function of the acquisition dates of the product. The two outliers seen on both figures correspond to products from orbit 5601 and are due to contingencies (star-tracker outage).

Results for Refined Products

The objective is to characterize the geolocation performances of a L1C product whose geometric model has been refined by matching with a well-geolocalised Global Reference Image (GRI).

The statistical results obtained on refined L1C product geolocation performances are summarised in Table 13 and represented in Figure 47.

Table 13. Statistics on geolocation for refined L1C products.

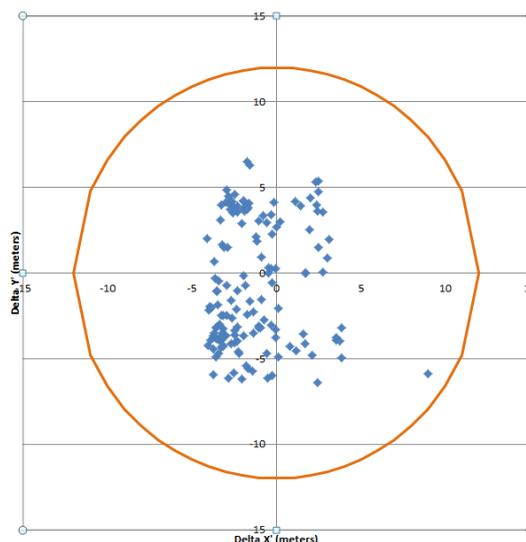| System Geolocation Performance (Circular Error)	L1C Refined Product
95.45% conf. level	8 m
Mean value	4.5 m
Max value	11 m

Figure 47. Product Geolocation Performances in metres (Band 04) for L1C refined tiles. Orange circle is the requirement with geometric refinement on GCPs.

The 95.45% confidence level ("2σ") product geolocation performance value obtained on analysed products is 8 m for L1C refined products. It is compliant with the specification of 12.5 m at 2σ.

4.4.2. Multi-spectral Registration Uncertainty Validation

The goal of the multispectral registration assessment is to verify that the spatial registration of spectral bands is within the specifications, with a special focus on the registration performance between the focal planes.

Multi-spectral registration assessment can be performed at:

- System level, on L1B products without any improvement of VIS/SWIR focal planes registration by processing task.
- Product level, on L1B products with focal plane registration processing between VNIR and SWIR focal planes enabled. In that case the registration within each focal plane is not refined and the performance is the same as system level performance.

At the manuscript-writing time, the focal plane registration being stable and excellent, the registration processing task has not been activated.

4.4.2.1. Methods

The nominal method to assess multi-spectral registration performance, within each focal plane and between VIS and SWIR focal planes, is based on correlation of spectral band couples (Figure 48):
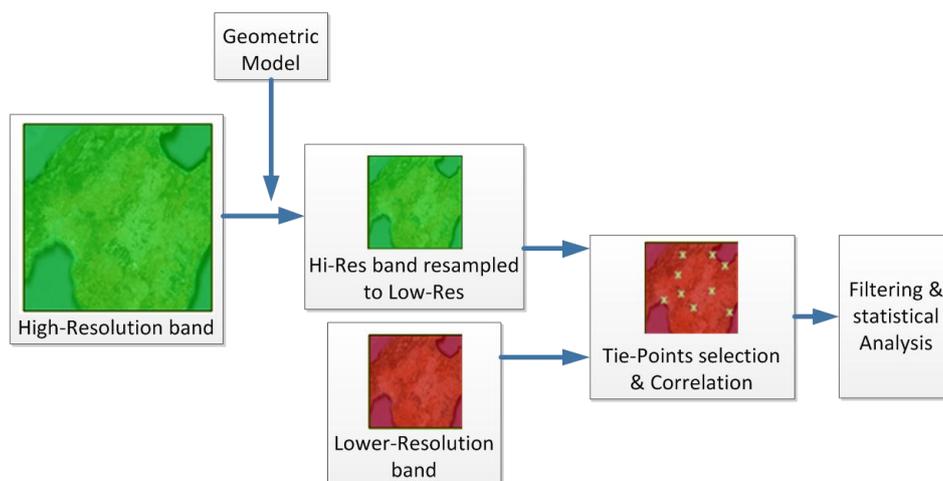

Figure 48. High-level principle of the multi-spectral registration uncertainty assessment.

For any band couple and detector in one L1B product:

(1) Resample the image with the finest resolution (secondary image) in the geometry of the image with the coarsest resolution (reference image). For that purpose a resampling grid establishing the correspondence between pixels of the reference band and their position in the secondary band is computed using the line of sights physical model. Then a B-spline interpolation method is used to interpolate the radiometric values of the secondary band at locations defined by the resampling grid.

(2) Select a list of well-matching tie-points in the images where to perform the correlations (avoiding clouds, water, etc) and extract small image chips around tie-points in both images.

(3) This process is applied to cloud-free products acquired on several areas at different latitudes and with various kinds of landscapes to estimate registration uncertainty (along-track and across-track shifts) at each tie point by correlation of the reference and secondary image chips.

(4) Filter out outliers with poor correlation based on correlation quality criteria

(5) Compute the registration uncertainty parameter: 99.73% quantile among shifts measured on all tie-points. This is compared to the requirement.

(6) Compare results against requirements.

Several products on various areas are used to decrease the influence of landscape radiometric content on the correlation measurements. Indeed, on some landscapes, the correlation quality between certain band couples can be poor due to very different spectral responses. In particular band B10 is especially difficult to correlate with other bands, since it is radiometrically very different from other bands.

The spectral band couples used to perform multi-spectral registration assessment are shown in Figure 23. They have been chosen in order to optimize the correlation quality.

When the registration processing task is enabled, the system-level registration uncertainty between VIS and SWIR focal planes can also be assessed by a complementary analysis of the registration processing outputs (registration residuals).

4.4.2.2. Results

The system-level multi-spectral registration performance is very good since the first calibrations during commissioning phase are within the requirement of 0.3 pixels of the band with the coarser resolution, for any correlated band couple.

Figure 49 shows an example of measured error clouds in one detector for two band couples. The production baseline of the product used is version 2.04.
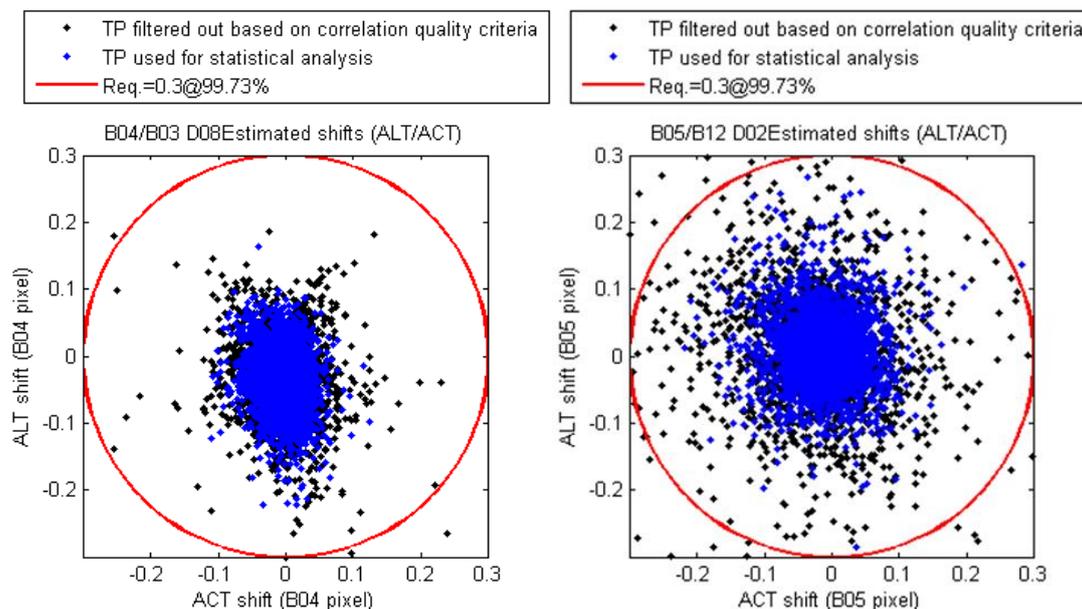

Figure 49.   Example of global mis-registration shifts measured on 'Central Australia' for B04/B03-Detector 08 and B05/B12-Detector 02. Black points are tie points rejected based on correlation quality criteria (step 4). Units are pixels of the coarser band in the couple.

Due to the very good and stable multi-spectral registration performance, the VIS/SWIR registration processing task has not been activated so far.

The 99.73% confidence level ("3σ") multi-spectral registration performance (without registration processing task) obtained on analysed products is always less than 0.3 Spatial Sampling Distance (SSD) of the coarser band correlated.

4.4.3. Multi-Temporal (XT) Registration Uncertainty Validation

The objective of the multi-temporal registration uncertainty validation is to assess the performance of co-registration between L1C tiles acquired on the same site at different dates (superimposable time-series). This relative measurement is important to appreciate stability of the instrument, to distinguish year-to-year and season-to-season variations due to global change from technical issues that may arise during operations.

The achievement of the multi-temporal registration performance requirement nominally relies on the geometric model refinement processing based on Global Reference Image (GRI). This processing is not yet activated in the operational chain, thus in this section we present:

- The current multi-temporal registration performance for non-refined product which is representative of the current situation.
- A preliminary assessment of the performance with refinement activated, based on products processed with the ground-processor prototype and the available European part of the GRI.

4.4.3.1. Method

The assessment method, for both refined and non-refined products is based on correlation between L1C tiles as follows and illustrated in Figure 50:

(1) Select two cloud-free L1C tiles acquired on the same area on Earth at different dates. Select also the reference band to be processed. In practice, for a given tile, the older L1C image is taken as reference and all new cloud-free acquisitions of the same tile are compared to the reference tile.
(2) Select a list of well-matching tie-points in the images where to perform the correlations (avoiding clouds, water, etc.) and extract small image chips around tie-points in both images.
(3) Estimate registration uncertainty (shifts along X and Y axes) at each tie point by correlation of the reference and secondary image chips.
(4) Filter out outliers with poor correlation based on correlation quality criteria.
(5) Compute the registration uncertainty metrics at tile level: Average circular error over all non-rejected tie-points of the tile.
(6) Repeat the previous steps (1) to (5) for as many tile couples as possible.
(7) Compute the global registration uncertainty metrics, to be compared to the requirement: 95.45% quantile among the average shifts measured on each L1C tile couples.
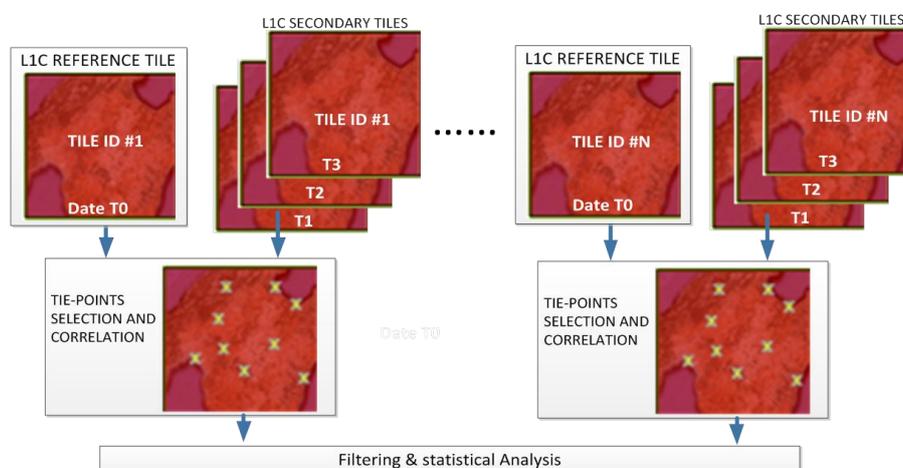

Figure 50. High-level principle of the multi-temporal registration uncertainty assessment.

Various factors influence the multi-temporal matching accuracy, some of which are listed here after:

- The quality of the images,
- The seasonal scene variations and meteorological/atmospheric properties (water vapour and aerosols should be limited in the processed scenes),
- The properties of the scene, relief, surface reflectance, and image content,
- When the two tiles correlated have been acquired from two adjacent orbits (i.e. covering the overlapping areas), the viewing angles are different. In areas with relief this causes a parallax between the two images that can degrade the correlation accuracy.

These factors imply that the multi-temporal performance assessment by correlation is always pessimistic.

It is here highlighted that the performances presented in the following paragraphs are assessed only in products acquired from the same relative orbit in the repeat cycle. Indeed, products acquired on the same site from adjacent orbits, therefore with different viewing angles, cannot meet the stringent requirement, since the system error due to operational DEM accuracy is higher than 0.3 SSD at 2σ. The elevation accuracy of the PlanetDEM90 used in operational processing being 16 m @2σ, the error obtained is about *16 m * tan(10°) * 2 = 5.6 m = 0.56* SSD (for 10m bands) of co-registration error.

4.4.3.2. Results for Non-Refined Products

The assessment of multi-temporal registration uncertainty on non-refined L1C products has been performed on operational products, using the band B04 as reference band for correlation.

Performances for product baseline 2.00 to 2.03 suffered from a yaw bias correction anomaly, particularly visible at the edge of the swath, leading to a performance of about 1.83 SSD CE@95.5% confidence level on L1C products. The analysis was based on 933 L1C tiles included in 184 cloud-free products correlated against reference tiles took in 55 older products with the same relative orbit. The reference and secondary tiles were acquired between mid-December 2015 and the first of June 2016. They cover 195 MGRS tiles.

Multi-temporal registration performance in product baseline 2.04 is improved and has been assessed to 1 SSD CE@95.5% confidence level on L1C products as synthesized in Table 14. The analysis was based on 149 L1C tiles included in 46 cloud-free products correlated against reference tiles took in 19 older products with the same relative orbit. The reference and secondary tiles were acquired between 15/06/2016 and 10/08/2016. They cover 55 MGRS tiles. The sites selected for this analysis are shown in Figure 51. They include the geolocation validation sites and additional sites specific for multi-temporal registration assessment. Each site includes at least one MGRS tile. Figure 52 shows the registration error in each correlation tile with respect to the reference tile from which the 95.5% confidence level performance has been estimated. The performance is still not compliant with the requirement but this is expected because the refinement process with GRI is not activated.

Table 14. Multi-temporal registration performance for non-refined L1C products.

Multi-Temporal Registration Performance (PB 2.04) (Circular Error)	Non-Refined L1C Results
95.45% conf. level	1 pixel at 10 m
Mean value	0.36 pixel at 10 m
Max value	1.5 pixel at 10 m

Figure 51. Sites used for multi-temporal registration uncertainty validation on non-refined products (red points). They include the geolocation validation sites (in blue) and other sites specific for multi-temporal registration assessment.
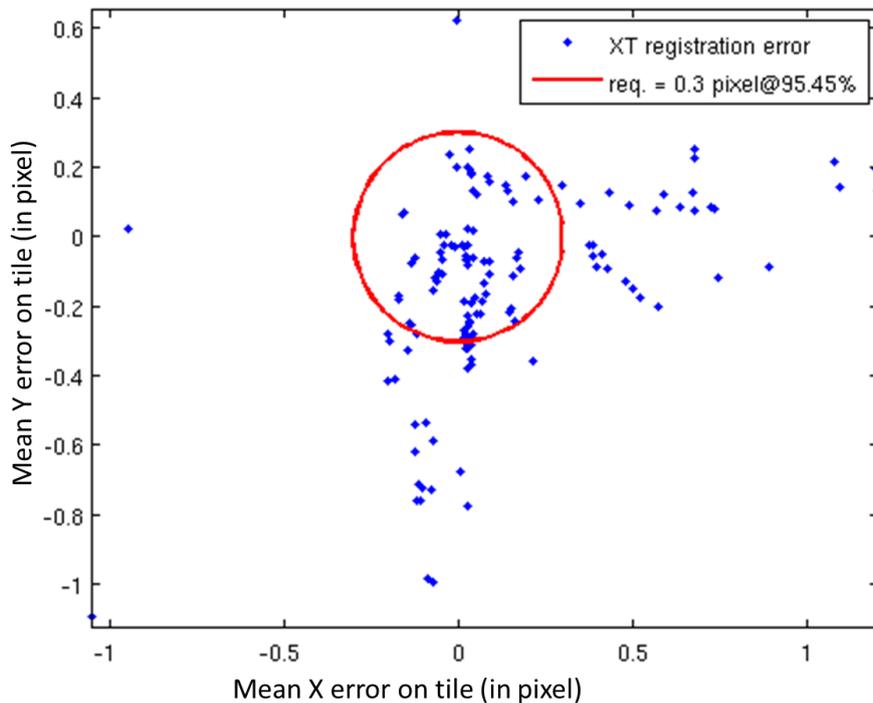

Figure 52. Multi-temporal registration uncertainty. Each point represents the average X shift and Y shift measure in one secondary tile with respect to the reference tile.

4.4.3.3. Results for Refined Products over European GRI

The objective is to characterise the multi-temporal registration of the L1C tiles, after geometric processing refining Sentinel-2 products over European GRI. This preliminary analysis was performed by CNES during the Sentinel-2 commissioning phase, with the Ground Processor Prototype.

Correlation and refining processing parameters have been tuned to comply with features of Sentinel-2 products (cloud-free, with a lot of water, cloudy, very cloudy and desert). B04 and B11 bands

are the reference bands for this analysis, considering that they give respectively the best and the worst performances (shown by studies achieved before launch).

The 95.45% confidence level multi-temporal registration performance value obtained from 505 analysed L1C products (see data strips in Figure 53) is 0.22pixel for B04 band (see Figure 54, with max mean value per tile of 0.36 pixel) and 0.17 pixel for B11 band (see Figure 55, with max mean value per tile of 0.26 pixel). However these figures must be consolidated in the future because the number of products processed is not very large, leading to possible statistical estimation errors. Indeed, the analysis was based on 505 L1C tiles included in 29 cloud-free products dispatched on 8 different relative orbits and correlated against reference tiles in 9 older products with the same relative orbit. The reference and secondary tiles were acquired between July and September 2015.

- Orbit 51: 6 dates, 312 tiles (France).
- Orbit 50: 6 dates, 76 tiles (East Europe).
- Orbit 35: 4 dates, 48 tiles (Middle East).
- Orbit 137: 2 dates, 12 tiles (UK).
- Orbit 122: 2 dates, 10 tiles (Austria).
- Orbit 36: 3 dates, 12 tiles (Hungary).
- Orbit 22: 4 dates, 8 tiles (Sweden).
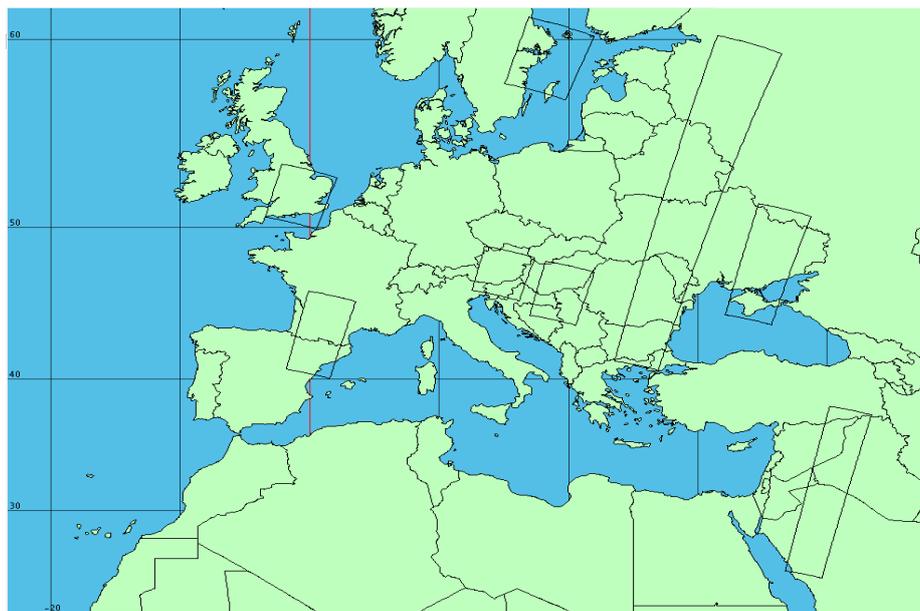- Orbit 64: 2 dates, 27 tiles (Ukraine).

Figure 53. Areas used for multi-temporal registration uncertainty validation on refined products.
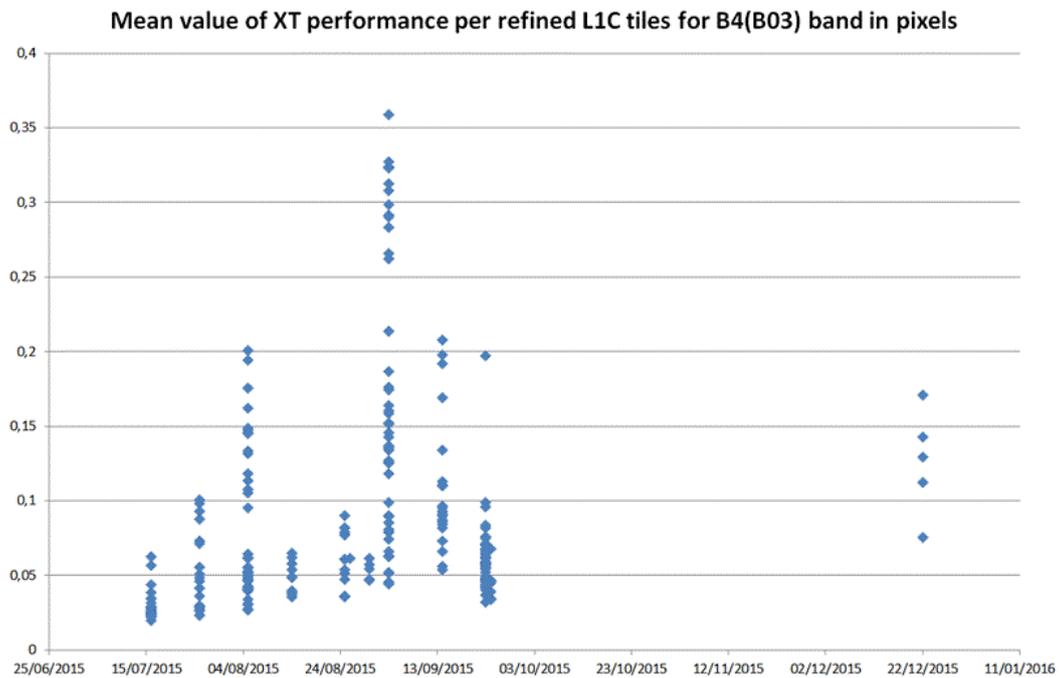
Figure 54. Multi-temporal (XT) registration uncertainty for B04 band. Each point represents the average shift measured in one secondary tile with respect to the reference tile.
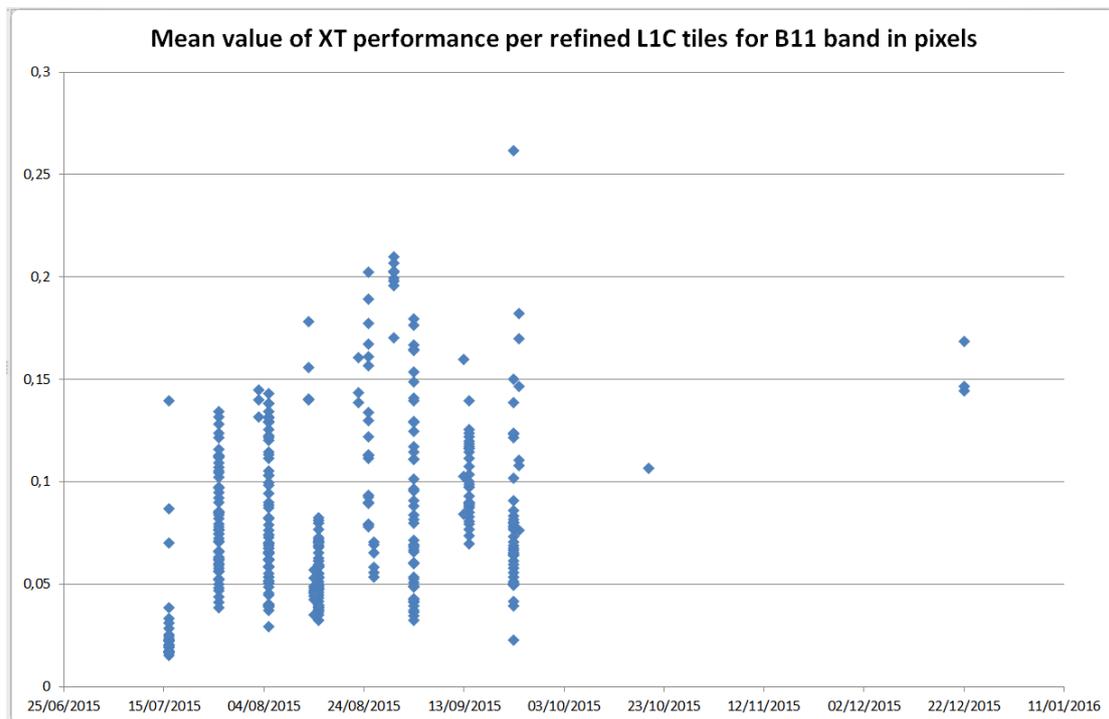

Figure 55. Multi-temporal (XT) registration uncertainty for B11 band. Each point represents the average shift measured in one secondary tile with respect to the reference tile.

The requirement at 95.45% confidence level of 0.3 pixel multi-temporal registration performance is achieved for refined products on GRI and acquired from the same relative orbit in the cycle.

4.4.4. Global Reference Images Validation

This activity consists in validating the completeness (coverage) and geolocation performance of the Global Reference Image that will be used in the operational ground processor to refine the geometric model of images by image matching and spatio-triangulation techniques.

Note that at the manuscript-writing time, only the European block of the GRI has been completely validated by CNES in the frame of the commissioning phase. This is presented in paragraph 4.4.4.2. The validation of Australian block is in progress, while the validation of the other blocks will start when they will be available.

4.4.4.1. Methods

GRI being a collection of L1B products, the nominal method for assessing its geolocation performance is exactly the same as the one used for geolocation validation on L1B products (see paragraph 4.4.1.1). The only difference is that the geometric model of the GRI has been refined.

The most important constraint of this method in view of a global validation is that it requires a lot of GCPs or reference images worldwide, independent from those used for the GRI construction. Specific procurement is under consideration in order to be able to assess geolocation performance in areas where the GRI geometric accuracy is potentially weak (e.g., far from the GCPs used during GRI construction).

4.4.4.2. Results for European Block of Global Reference Image

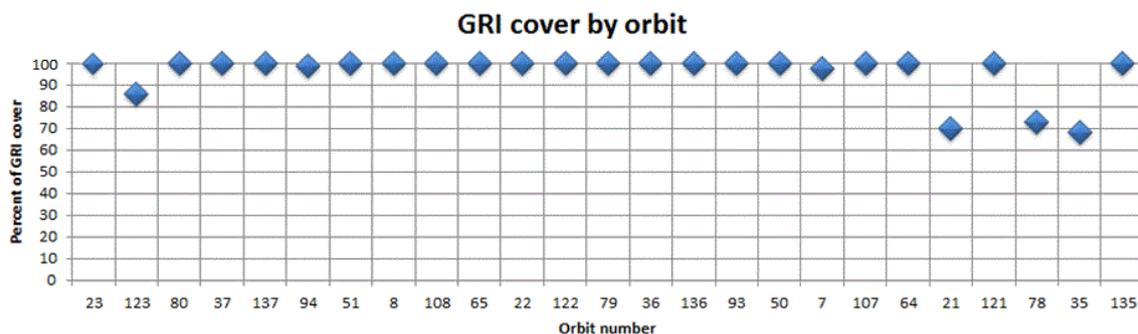The coverage percentage of each orbit of the final GRI over Europe is given by Figure 56.

Figure 56. Coverage percentage of Europe GRI versus relative orbit number.

Final GRI over Europe covers 95% of the Europe area. Missing images are located on:

- Orbit 021: 1240 km missing on the North.
- Orbit 035: 1600 km missing on the North.
- Orbit 078: 1100 km missing on the North.

Moreover, 66 km of image are missing on orbit 123, due to the acquisition of the 29/09/2015 split in two segments. Those parts of orbits should be filled later to complete the GRI over Europe, when cloud-free products will be available.

The geometrical model of the 65 European GRI products was refined by space triangulation using tie points between images and GCPs.

The location precision of the final GRI was validated using Pleiades database (references with a location precision better than 5 m). Figure 57 presents the area of each European GRI product according to the number of Pleiades references available, with the following legend of colours:

- Green: more than 6 Pleiades references crossing. The computed location precision will be very reliable.

- Orange: 3 to 5 Pleiades references crossing
- Red: 1 or 2 Pleiades references crossing
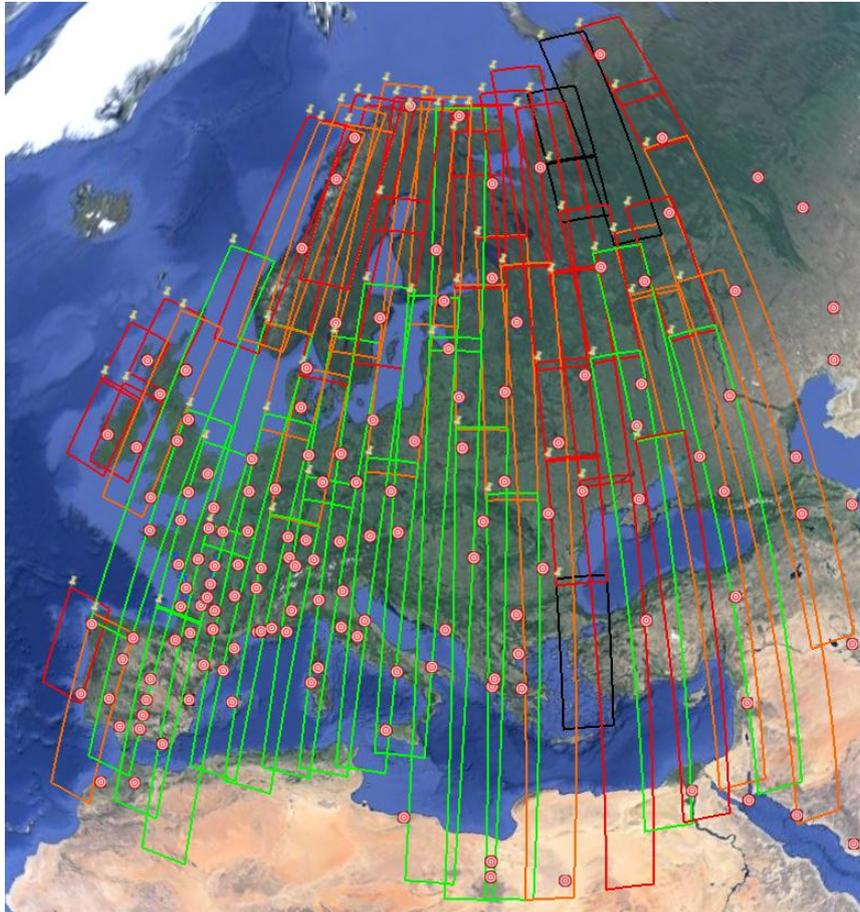- Black: no Pleiades references crossing. The location performance cannot be computed for these products.

Figure 57. European GRI product strips according to Pleiades references.

If all products were crossing other products (that is to say if one block was available for the whole Europe), it would not be important to validate geolocation on each segment separately. When the refinement is performed with all segments together, the geolocation is homogeneous on the block.

Figure 58 represents location precision of each European GRI product after refinement, with the following legend of colours:

- Green: geolocation better than 7 m.
- Orange: geolocation between 7 m and 10 m.
- Red: geolocation worse than 10 m.
- Black: no reference to estimate geolocation.
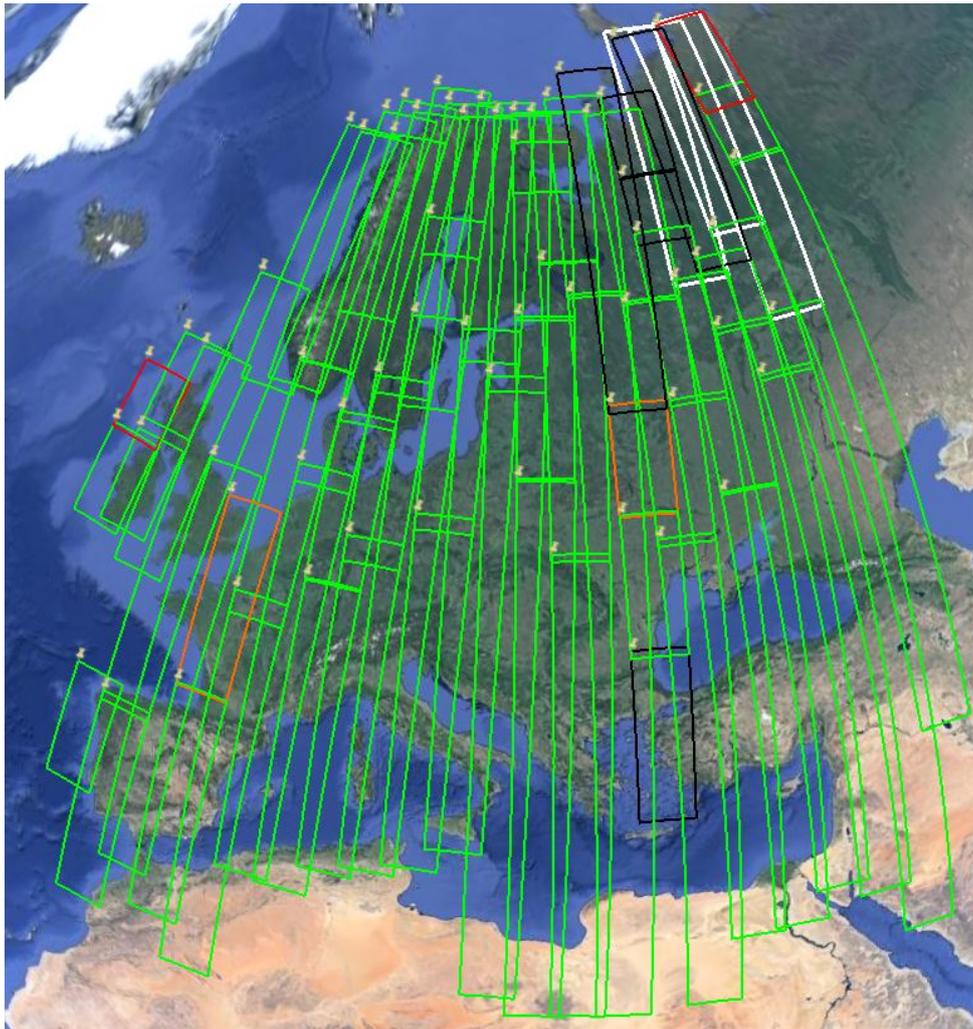- White: missing products in the final GRI.

Figure 58. European GRI product geolocation performance.

Red products with bad geolocation performance are cloudy and not linked to other products on the GRI. So they were not linked to the Europe block in refinement, and therefore their geolocation is not homogeneous to other products.

Consequently, it is recommended to add new products on red and white strips when available to complete the European GRI.

The global performance of the European GRI is assessed to be 8m at 95.45% confidence level.

5. Level-2A Calibration and Validation Status

Level-2A products are generated by the Atmospheric Correction (AC) processor Sen2Cor, which generates BOA reflectance, i.e., Surface Reflectance (SR), from a single-date Level-1C TOA product. Sen2Cor outputs can be classified in two main domains: (1) Radiometry, which concerns the Atmospheric Correction outputs of Sen2Cor: Surface reflectance products (SR), Aerosol Optical Thickness (AOT) and Water Vapour (WV) maps and (2) Cloud Screening and Scene Classification (SCL) outputs of Sen2Cor: a scene classification map which assigns a class to each pixel (Vegetation, Not-vegetated, Water, 2 types of Clouds, Thin Cirrus, Snow, etc.), a Cloud probabilistic mask and a Snow probabilistic mask. The scene classification map does not constitute a land cover classification map in a strict sense. Its main purpose is to be used internally in Sen2Cor in the atmospheric correction module to distinguish between cloudy pixels, clear pixels and water pixels.

Figure 59 with associated legends and colour tables gives an example of the outputs of Sen2Cor for one selected test site over Easton-MDE (USA) processed with default configuration. The Level-2A surface reflectance images are much clearer than the uncorrected Level-1C image showing no visible difference between the 2 different spatial resolutions. SCL correctly classified water, vegetation and not-vegetated pixels. The Sentinel-2 product example is composed of 4 granules, for which Sen2Cor retrieves WV and AOT maps in an independent manner (tile by tile processing). The WV and AOT values retrieved are very close and in this example no border effect at the boundaries is visible for WV, AOT and RGB composites of surface reflectance. The WV and AOT images appear to be quite homogeneous because the colour scale used for their visualisation is large with respect to the internal variations within the image.
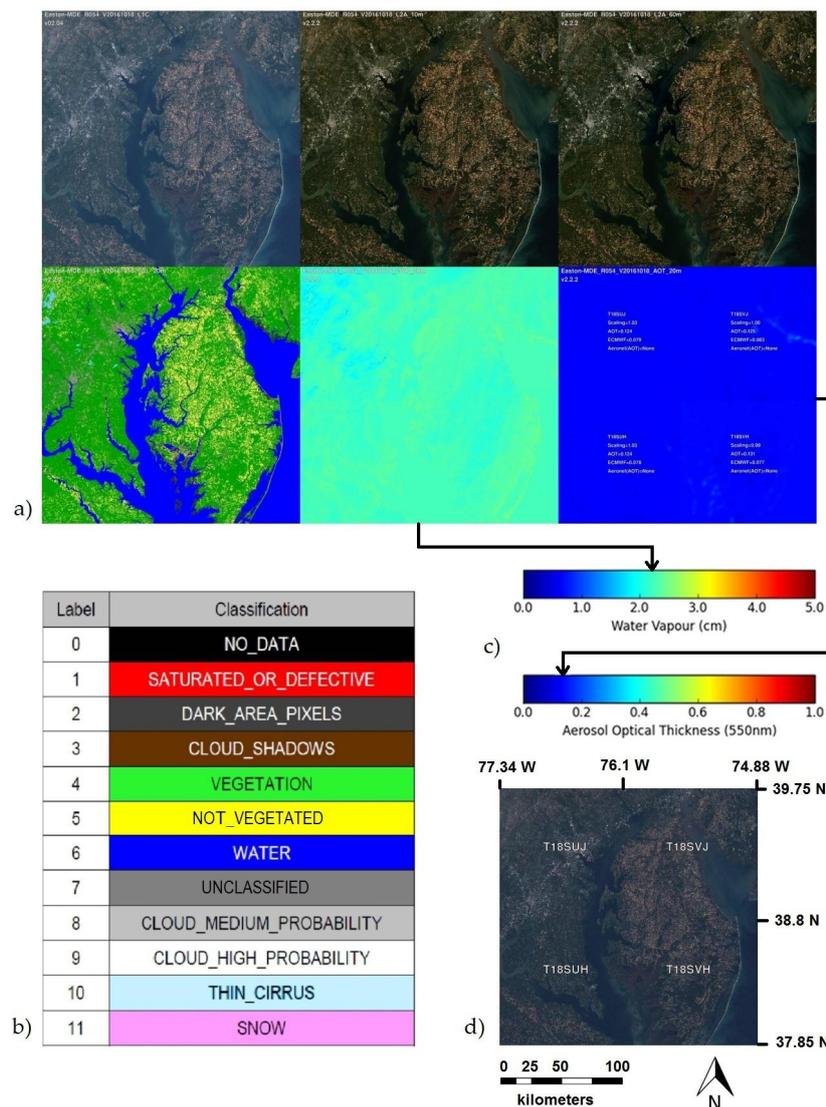
Figure 59. (**a**) Level-2A product example. Aeronet site Easton-Maryland-Department-Environment (USA), acquired on October 18, 2016, characterized by MidlatitudeN, Flat terrain, Forest, Croplands, Water, Urban. (Top) RGB compositions (B04, B03, B02) for L1C TOA, L2A Surface Reflectance (SR) at 10 m, L2A SR at 60 m; (Bottom) Scene Classification, L2A Water Vapour at 20 m, L2A Aerosol Optical Thickness (AOT) at 20 m. (**b**) Cloud Screening and Scene Classification (SCL) pixel legend; (**c**) Colour tables for Total Water Vapour Column and Aerosol Optical Thickness at 550 nm; (**d**) Scale, north arrow and coordinates of the four product granules.

5.1. Level-2A Calibration Activities

5.1.1. Level-2A Calibration Dataset

A calibration dataset of Sentinel-2 Level-1C products is constituted and regularly updated, covering different land cover types (e.g., snow, rocks, desert, urban, vegetation, grass, forest, cropland, vineyard, irrigated crops, rivers, lakes, sand, costal area, wetlands, ocean water) and different atmospheric conditions (e.g., cloud cover, aerosol optical thickness, water vapour content). The calibration dataset is worldwide including different latitudes to cover various solar angles and seasons. The 24 Level-2A calibration sites were selected above an active AERONET [45] site. Their geographical locations are shown on Figure 60. Their characteristics are further defined in [46].
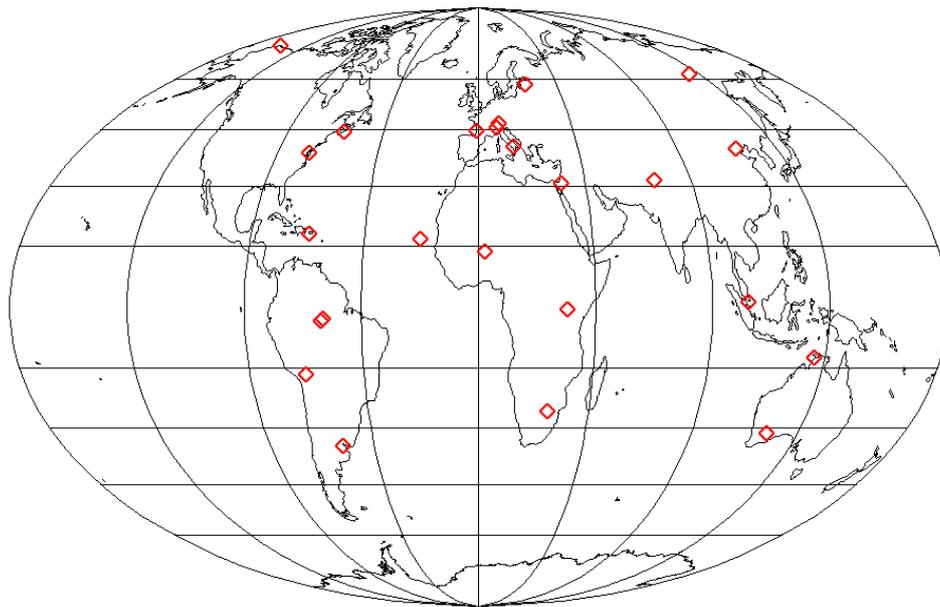

Figure 60. Geographical distribution of the 24 selected AERONET test sites for the Level-2A Calibration.

5.1.2. Atmospheric Correction Parameterization

The objective of this task is to perform a regular check and calibration of AC parameters concerning the Atmospheric Correction processor using the Level-2A calibration dataset covering different land cover types, atmospheric conditions, solar and viewing conditions. Updated calibration parameters are delivered under the form of an updated configuration file of Sen2Cor processor.

5.1.2.1. Method

The following main activities take place during this calibration task:

- The sensitivity of the Sen2Cor processor to the parameters stored in the processor configuration files "L2A_CAL_AC_GIPP.xml" and "L2A_GIPP.xml" is investigated by performing several runs of Sen2Cor with modification of the (single) calibration parameter of interest: Min Dark Dense Vegetation (DDV) area, SWIR reflectance lower threshold, DDV 1.6 μm reflectance threshold, SWIR2.2μm-red reflectance ratio, red-blue reflectance ratio, cut-off for AOT-iterations: max percentage of negative reflectance vegetation pixels (B4), cut off for AOT-iterations: max percentage of negative reflectance water pixels (B8), Aerosol type ratio threshold, topographic correction threshold, slope threshold, water vapour map box size, cirrus correction threshold, BRDF lower bound; The impact of these individual parameter variations on the different Level-2A products (AOT, WV and BOA reflectance) is analysed and compared to the in-situ data of AOT

and Water Vapour measurements. For "continuously varying" calibration parameter, a best value is retained to be the default configuration in the L2A_CAL_AC_GIPP.xml and L2A_GIPP.xml files. This activity will be performed once the internal atmospheric calibration parameters have been moved to the L2A_CAL_AC_GIPP.xml configuration file.

- Concerning yes/no calibration parameters (e.g., cirrus correction, BRDF correction), a qualitative analysis of the impact of the activation/deactivation of each parameter is undertaken. (These choices usually depend on particular kind of user applications or scene landscape). The individual impact of each parameter is assessed on a variety of landscapes and weather conditions. The outputs of this calibration activity are expected to provide (1) advices to the Sen2Cor user, based on these previous assessments, when Sen2Cor is used within the Sentinel-2 Toolbox and (2) a default configuration (more generic) to cover the needs of systematic Level-2A production.

5.1.2.2. Results and Outlook

The main result to date of this calibration activity is the Sen2Cor version 2.2.1 issued on May 4, 2016. This version includes an updated internal resampling method that preserves the geolocation information when spectral bands are resampled to another resolution. This update has improved the quality of the WV map retrieved with Atmospheric Pre-corrected Differential Absorption (APDA) algorithm [11] which for Sen2Cor relies on spectral bands B8A and band B09 that have different native resolutions, respectively 20 m and 60 m.

A thorough investigation has been performed on Sen2Cor handling of two different DEMs, mostly SRTM v4.1 in GeoTIFF format, the default DEM used in Sen2Cor and Planet DEM as a user-provided DEM in DTED format used in the Level-1C processing chain. Several issues have been identified, regarding the SRTM geolocation itself, the DTED processing and the handling of DEM NoData values. For each of these issues a recommendation has been formulated and a solution proposed for implementation in the next release of Sen2Cor, version 2.3.

A number of algorithm evolutions are foreseen for the atmospheric correction module of Sen2Cor: improvements of cirrus correction and AOT retrieval as well as the implementation of an AOT fallback estimation based on ECMWF aerosol information when AOT retrieval fails in case of missing DDV pixels.

5.1.3. Cloud Screening and Scene Classification Calibration

The objective of this task is to perform a regular check and calibration of thresholds parameters concerning the Cloud Screening and Scene Classification (SC) module using the Level-2A calibration dataset covering different land cover types, atmospheric conditions, solar and viewing conditions. Updated calibration parameters are delivered under the form of an updated configuration file of Sen2Cor processor.

5.1.3.1. Method

The procedure for the calibration of the Cloud Screening and Scene Classification algorithm is performed as follows:

1. Sen2Cor processor (with the option for scene classification only) runs on the Level-2A Calibration dataset using all thresholds by default from the current Level-2A Processing Baseline, producing for each S2 scene (100 km × 100 km) a scene classification map and 2 quality indicators maps (Cloud confidence and Snow confidence). These outputs are stored with the objective to be used later in the calibration procedure as reference.
2. For each class of the Scene Classification Map, the pixel classification is verified by manual inspection by superimposing the scene classification map and Level-1C spectral bands. (At this stage the work performed by DLR on Level-2A Product Validation is also used as input when available). The outputs of this step are a performance assessment of the classification for each

class reporting on: over-detection, under-detection, misclassification (indicating the wrong class assigned) and how the edges/boundaries between classes are handled by the processor.

3. Based on this first assessment, a list of potential areas of improvements is identified with their corresponding thresholds. Some thresholds are more critical than others, bearing in mind that the principal objective of the scene classification algorithm is to distinguish between cloudy and clear pixels. Some of the most difficult optimal thresholds to find are the thresholds linked to thin clouds (e.g., thin cloud over desert areas) because in some cases their tuning will lead to over detection (some clear urban landscape being classified as medium probability clouds) and in other cases under detection (missed thin clouds). On the other hand, some other thresholds have less importance for the atmospheric correction algorithm e.g., some NDVI thresholds have an influence on the number of pixels classified as vegetation versus the number of pixel classified as not-vegetated.

4. The proper tuning activity consists in manually slightly varying the SC thresholds to improve the Scene Classification.

5. When the tuning activity is finished, a new processing baseline is provided with a new set of updated SC parameters delivered in an updated Sen2Cor calibration file "L2A_CAL_SC_GIPP.xml".

6. Sen2Cor processor (with the option for scene classification only) runs again on the Level-2A Calibration dataset using the updated Level-2A Processing Baseline, producing for each S2 scene (100 km × 100 km) a new scene classification map and 2 new quality indicators maps (Cloud confidence and Snow confidence).

7. A comparison exercise is performed between the SC results with the updated thresholds and the SC results with the standard baseline to assess the impact of the changes. Through visual assessment, the pixel class transitions are evaluated on a large range of land cover types, atmospheric conditions, solar and viewing conditions.

Based on this comparison exercise, if it seems possible again to improve the SC parameters through thresholds tuning process, the decision is taken to start another tuning (step 4) calibration. However in other cases the tuning activity is not sufficient and the improvement of the scene classification can only be achieved by a further evolution of the SC algorithm, e.g., including an additional test on Sentinel-2 band or including DEM auxiliary information.

5.1.3.2. Results and Outlook

The main result of this calibration activity is the Sen2Cor version 2.2.1 issued on May 4, 2016. This version includes an updated set of thresholds for Cloud Screening and Scene Classification (SCL) which improves the overall classification accuracy.

In addition to the thresholds tuning process, several evolutions have been introduced in the SCL algorithm. They are:

- Cloud shadow detection evolution.
- Topographic shadow evolutions.
- Snow and water classification improvements.
- Cirrus detection improvements.

The implementation of the cloud shadow algorithm has been reviewed to allow a proper computation of the cloud shadows for all configurations of solar angles and viewing angles in northern and southern hemispheres. In addition, the dark features identification used as input for the cloud shadow computation has been calibrated to improve the resulting cloud shadow mask. Figure 61a gives an example of the cloud shadow class appearance in the Scene Classification map.
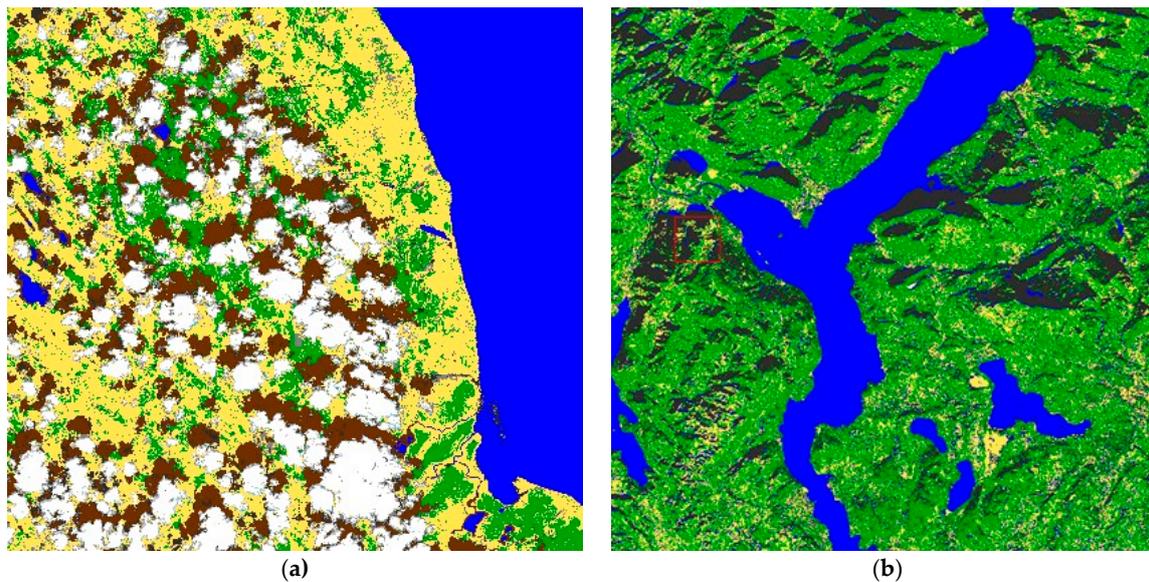
(**a**)   (**b**)

Figure 61. (**a**) Cloud shadows classification; (**b**) Topographic shadows classification.

In version 2.2.1 the topographic shadows can now be identified by using an illumination map derived from the solar position and a DEM. Additionally the DEM slope information is used to filter out water pixels detected on steep slopes. Figure 61b illustrates this new feature.

The issue of snow and water detection in the clouds has been partially solved by adjusting the calibration parameters and adding threshold conditions on TOA reflectance. There are still areas of improvement for this algorithm and several ideas of investigation using auxiliary data (DEM altitude, Snow Climatology, Water Bodies) have been proposed for evolutions.

5.2. Level-2A Validation Activities

Validation includes an accuracy assessment of the SCL product and radiometric validation of WV, AOT and SR products. Validation data sets differ for SCL validation and for radiometric validation. SCL validation is performed on full granules with area of 110 km × 110 km whereas radiometric validation works on area of only 9 km × 9 km around sun photometer location. Several AERONET sunphotometer sites have been selected for radiometric validation. Radiometric validation relies also on fiducial reference data gathered on ground during ad-hoc campaigns. However, analysis of data from ad-hoc campaigns is no part of the present paper. Only first validation results for a limited number of test cases and test sites will be presented in the following sections.

5.2.1. Cloud Screening and Scene Classification Validation

The scene classification algorithm is designed to detect clouds, snow and cloud shadows and to generate a classification map, which consists of 3 different classes for clouds (including cirrus), together with six different classifications for shadows, cloud shadows, vegetation, not-vegetated, water and snow, at 60 m and 20 m resolution. The goal of this section is to assess the quality of cloud screening and classification provided by Sen2Cor.

5.2.1.1. Method

The SCL validation is limited due to the lack of "ground truth" data sets. The applied validation method performs stratified random sampling followed by establishment of a reference database by visual inspection of the samples, and finally establishment of an error matrix allowing calculation of accuracy statistics.

Each scene classification product from selected test sites contains more than 30 million pixels. As the validation is performed manually in a pixel based manner, a representative subset of these data has to be selected. Stratified random sampling guarantees statistical consistency (validity) and avoids exclusion of spatially limited classes from the validation.

Visual, pixel-based manual inspection of the samples is supported by RGB band composition image (Bands 4,3,2), Colour Infra-Red composition image to highlight vegetation features (bands 8A,4,3) or snow (bands 12, 11, 8A), shape of the spectral curve, snow confidence Quality Image and cloud confidence Quality Image. After the Scene Classification results are compared against the reference database, an error matrix (confusion matrix) is calculated. Overall, user's and producer's accuracies are then computed as validation measures.

5.2.1.2. Results and Analysis

Ground-truth classification images were created and compared to corresponding pixels of the Sen2Cor classification images for four scenes at four different locations. One example of data analysis is shown in Figure 62.
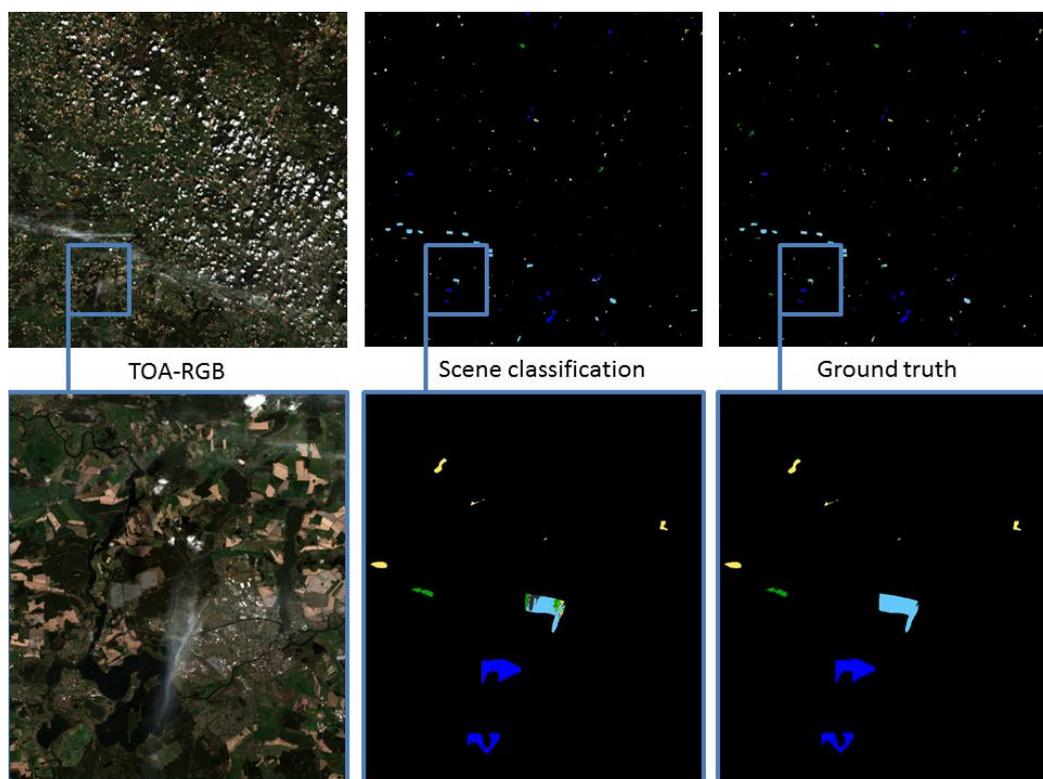

Figure 62. Example images for Cloud Screening and Classification Validation analysis, Site Potsdam (Germany), acquired on April 22, 2016, characterized by temperate climate zone and flat terrain. Whole granule at the top and zoomed area at the bottom.

The image contains clouds, cloud shadows, vegetation, not-vegetated and water pixels. Table 15 contains the corresponding confusion matrix. The diagonal in bold shows the number of correctly classified pixels in the validation set. In this example, vegetation, not-vegetated, water, high probability clouds and cloud shadow classes have been correctly classified for a very large portion of validation data. The resulting user's accuracy (UA) ranges from 18.3% for class unclassified to 97.9% for class "vegetation". UA is a value indicating the probability that the pixel was classified correctly. Other measures for characterization of classification performance are producer's accuracy (PA), overall accuracy (OA) and Cohen's Kappa coefficient. PA indicates the probability that a pixel belongs to the

class classified by Sen2Cor. PA per class is reported in Table 15. OA indicates the relative number of correctly classified pixels relative to the total number of pixels whereas Cohen's Kappa indicates the inter-rater agreement. OA and Cohen's Kappa, for the example presented, are respectively 84.0% and 0.796.

Table 15. Example of confusion matrix for classification validation. The classes are numbered like in Figure 60. Note that classes 2 and 3 are merged.

Sen2Cor Class		Ground-Truth Class										UA (%)
		(1)	(2+3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	
(1)	Saturated or defective	**0**	0	0	5	0	0	0	9	7	16	0
(2+3)	Dark area, cloud shadows	0	**6785**	880	0	0	5	0	0	0	0	88.5
(4)	Vegetation	0	31	**15,398**	52	0	46	0	0	198	0	97.9
(5)	Not-vegetated	0	441	406	**38,063**	0	177	34	460	0	0	96.2
(6)	Water	0	3223	3	0	**42,351**	110	0	0	89	0	92.5
(7)	Unclassified	0	1	154	354	0	**205**	163	212	33	0	18.3
(8)	Cloud medium probability	0	21	75	599	0	517	**823**	347	327	0	30.4
(9)	Cloud high probability	0	0	0	24	0	14	533	**4197**	148	0	85.4
(10)	Thin cirrus	0	3234	6444	2205	0	934	7010	187	**48,308**	0	70.7
(11)	Snow	0	0	0	0	0	0	0	0	0	**0**	0
	Producer Accuracy (%)	0	49.4	65.9	92.2	100.0	10.2	9.6	77.5	98.4	0	

The analysis was performed for four more example scenes located in Madrid (Spain), Manila (Phillipines), Tatra Mountains (Poland) and Potsdam (Germany). The overall accuracy of all scenes processed is relatively stable with mean of 80.8% and standard deviation 3.4%. This is a satisfactory result in comparison with those of other studies. Crop and tree species classifications on single data Sentinel-2 acquisitions had been studied with a supervised Random Forest classifier and a classical pixel-based classification [47]. The achieved overall accuracies ranged between 65% and 76%.

Water (if present in the image) and high probability clouds are classified with similar UA for all examples investigated. UA for vegetation and not-vegetated classes vary from scene to scene due to the different types of included landscape. Not-vegetated class includes a relative broad set of possible objects. It contains not only soil but urban objects, roads or railroads. Also, classification of dark areas and cloud shadows varies due to the different terrains and cloud types at different altitudes in the atmosphere.

5.2.2. Validation of AOT and WV Products

The objective of this study is to validate AOT and WV products estimated by Level-2A processing chain in Sen2Cor. AOT and WV are the key parameters for atmospheric correction and have a large influence on the accuracy of BOA-SR product.

5.2.2.1. Method

The principle of validation of AOT and WV products is a direct comparison of Sen2Cor outputs with ground truth from AERONET [45] sunphotometer measurements. For WV validation, the average Sen2Cor output over all soil and vegetation pixels within 9 km × 9 km area around sunphotometer location is computed. AOT validation includes also water pixels for the same area.

5.2.2.2. Results and Analysis

The Level-1C products were processed to Level-2A using the default Sen2Cor configuration (e.g., cirrus correction deactivated) and with SRTM DEM activated. Processing results with 60 m, 20 m or 10 m spatial resolution give consistent results. AOT validation results are shown in Figure 63 Note that only a small number of products was included into the analysis until now. Sen2Cor produced some very good AOT estimation samples, however also some with large differences between Sen2Cor and ground truth. Most of bad samples suffer from lack of DDV-pixels, which are necessary for the AOT-retrieval algorithm implemented in Sen2Cor. AOT is set to a default value of 0.20 when there

are not sufficient DDV pixels in the image. Limiting the statistical analysis to samples with enough DDV pixels in the image and with cloudiness less than 5% gives a mean difference between Sen2Cor and ground truth of 0.03 ± 0.02 with maximum difference of 0.05. This result is in agreement with experience from earlier validation studies for Landsat and RapidEye sensors [48]. Present validation results for Sen2Cor are aligned with validation results for other atmospheric correction algorithms found in the literature [49].
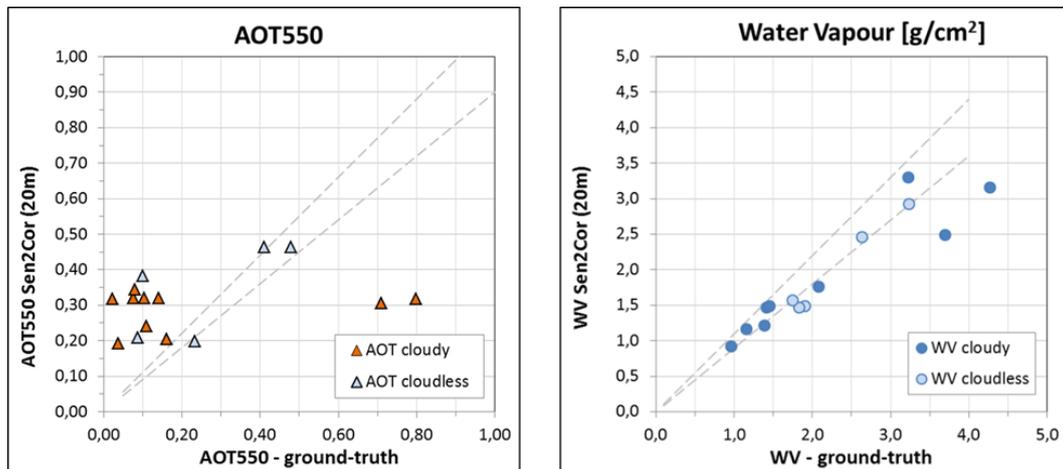

Figure 63. AOT and Water Vapour (WV) Validation. Direct comparison of Sen2Cor with ground reference from AERONET. Sen2Cor is represented by average over 9 km × 9 km area with AERONET sunphotometer in the centre. Filled symbols represent samples with cloudiness >5% and not filled with cloudiness <5%.

Water vapour retrieval gives better results than AOT (Figure 63). Retrieval accuracy is less influenced by cloud cover and missing DDV pixels. Limiting the statistical analysis to samples with cloudiness less than 5% gives a mean difference between Sen2Cor and ground truth of (0.29 ± 0.11) g/cm^2 with maximum difference of 0.42 g/cm^2.

5.2.3. Validation of BOA Reflectance Products

The objective of this investigation is to estimate the uncertainty of BOA SR product resulting from processing Sentinel-2 data using Sen2Cor Level-2A processor. The validation uses the same test sites and in-situ data sets as the ones used for validation of AOT and WV products.

5.2.3.1. Method

The validation is performed comparing Sen2Cor outputs with surface reflectance reference data for a 9 km × 9 km subset around the sunphotometer location. Reference data are computed by a second run of Sen2Cor using AOT550 information provided by collocated sunphotometer measurements as input. The resulting "sunphotometer-corrected" surface reflectance data are considered to provide the surface reflectance "truth", since the greatest uncertainty in atmospheric correction comes from the aerosol characterization.

5.2.3.2. Results and Analysis

Processing results with 60 m, 20 m or 10 m spatial resolution give consistent results. Example for BOA-SR validation at test site Belsk (Poland) is typical for locations with enough DDV pixels in the image (Figure 64). Example spectra were extracted at different locations in the reference image and in the image to validate. Example spectra for dark and bright soils, for forest and for different other vegetated locations show the expected spectral dependency and agree within the target accuracy of

relative 5% for BOA-reflectance. The SR differences up to 0.04 between Sen2Cor results and reference
lead to a Normalized Density Vegetation Index (NDVI) uncertainty up to 0.06.
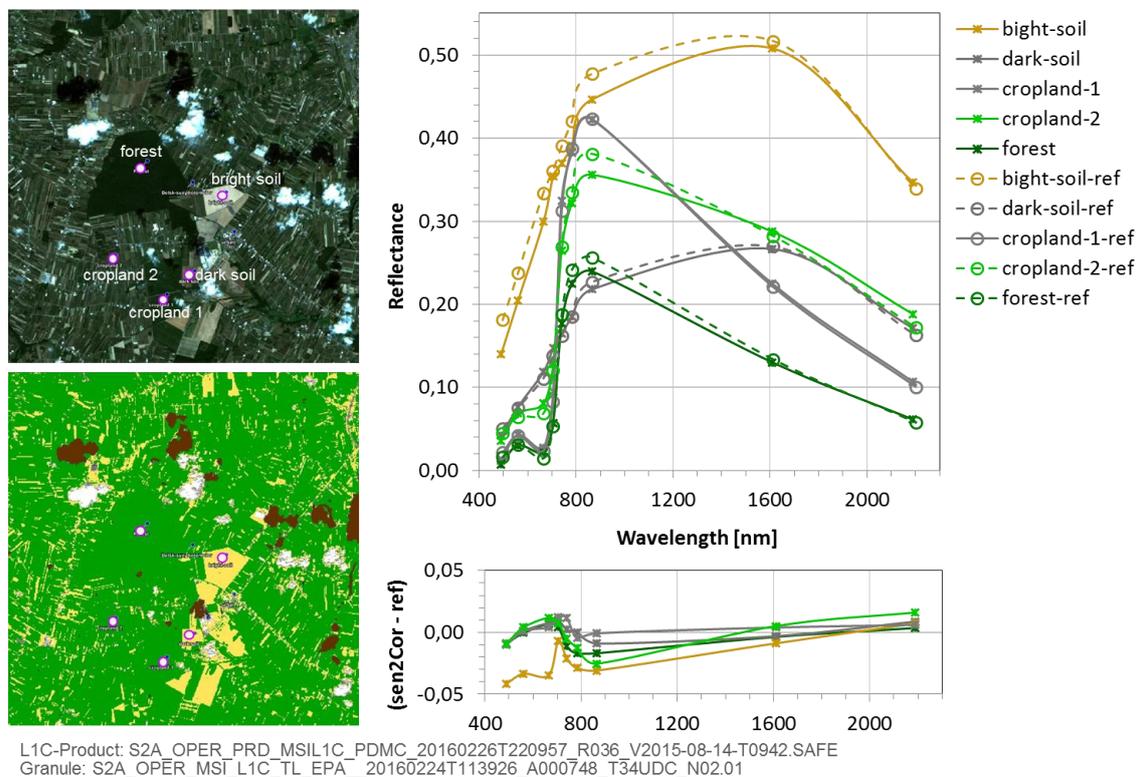

L1C-Product: S2A_OPER_PRD_MSIL1C_PDMC_20160226T220957_R036_V2015-08-14-T0942.SAFE
Granule: S2A_OPER_MSI_L1C_TL_EPA__20160224T113926_A000748_T34UDC_N02.01

Figure 64. Validation example for Bottom-Of-Atmosphere (BOA) reflectance, Site Belsk (Poland),
acquired on 14th of August 2015, characterized by flat terrain. (**Left**): BOA-RGB and scene
classification image of the 9×9 km^2 area around AERONET sunphotometer (longitude 20.79 E,
latitude 51.84 N). (**Right**): Example spectra and relative difference to reference for soil and vegetated
pixels at indicated locations.

These results are consistent with other investigations. Sentinel-2 Level-2A data processed with
Sen2Cor were compared to atmospherically-corrected Landsat-8 data achieving an overall Root
Mean Square Error (RMSE) of 0.031 over all six bands used for comparison [50]. RMSE ranges from
0.023 for the green to 0.043 for the near-infrared band. The pixel-based comparison was conducted
on acquisitions of Sentinel-2 and Landsat-8 on the same day over six test sites in Europe. The
atmospherically-corrected Landsat Surface Reflectance Climate Data Record (CDR) was used for the
comparison, which is in very good agreement with MODIS data and AERONET measurements [51].
The same study also found very good agreement between Sentinel-2 surface reflectance data and
reflectance measured on ground during two field campaigns. Sentinel-2 Level-2A product has also
potential for application over water surface [52], even if Sen2Cor was developed as atmospheric
correction over land surface. Results of processing Sentinel-2 data with three different atmospheric
corrections relative to spectra measured on a lake have been compared: Sen2Cor, ACOLITE and MIP
(Modular Inversion and Processing System). Both ACOLITE and MIP are algorithms developed for
application over water bodies. Comparison at seven test points showed that all 3 algorithms give
comparable results. On average over all test points MIP performed a bit better than Sen2Cor and
ACOLITE considering both the shape and the intensity of resulting spectra. RMSE between Sen2Cor
surface reflectance spectra and in-situ measurements range from 0.002 to 0.005. Another study could
not rely on reflectance measurements on ground in parallel to Sentinel-2 overpasses [53]. They found
that shape and magnitude of spectra over lakes generated with Sen2Cor are consistent with field

measurements from previous years. The agreement is better in case of more turbid water giving higher reflectance. This may be due to sun-glint effects not considered in Sen2Cor. Sun glint has a larger influence on lower signals coming from the water. Another result is that water parameter retrievals on basis of band ratio algorithms performed better with TOA-data than with BOA-data. Band ratio algorithms are very sensitive to small errors in the signals. They relate signals from the NIR bands to the green band which is much more affected by the atmosphere than the NIR. Consequently, atmospheric correction uncertainty has a larger effect on the green band leading to error magnification on the band ratio. Bio-optical modelling is less sensitive to atmospheric correction uncertainty [52].

6. Conclusions and Perspectives

The main objective of the Sentinel-2 MSI mission is to provide stable time series of images at medium-high spatial resolution. After one year in orbit, this objective is mainly achieved. Sentinel-2 time series are now consistent and comparable with other missions such as Landsat-8 OLI.

This paper provided a description of the calibration activities and the current status, one year after Sentinel-2A launch, of the mission products validation activities. Measured performances, derived from the validation activities, have been estimated and presented for the different mission product levels.

Results obtained show the very good performances of the mission products both in terms of radiometry and geometry. Thanks to a robust in-flight calibration strategy, the radiometry is both accurate (<5% absolute uncertainty) and stable (<1%/year variation estimated). Cancelling seasonal effects on diffuser acquisition is the key to this performance: this involves an accurate model of the Sun-Earth distance and of the diffuser bi-directional Reflectance Function. Further progress on the latter point should lead to improved pixel response stability (i.e., Fixed Pattern Noise) in the near future.

The geometric accuracy is also very satisfactory, especially after the correction of the LOS yaw bias at the beginning of June 2016. The typical multi-temporal co-registration between two acquisitions is now better than one pixel. The full performance (better than 0.3 pixel at 2σ confidence level) will however be reached later, after the activation of the geometric refinement based on the GRI. At this point, the accuracy of the DEM becomes the main limiting factor for the geometric performance.

Level-2A Sen2Cor processor provides Cloud Detection and Scene Classification with an overall accuracy of (81 ± 3)% for the products investigated. Uncertainty of retrieval of the Aerosol Optical Thickness is 0.03 ± 0.02 if the image contains Dense Dark Vegetation pixels. If there are no Dense Dark Vegetation pixels in the image, then Aerosol Optical Thickness estimation fails. A processor evolution is being developed using aerosol information from ECMWF. Water Vapour content is retrieved from the Level-1C image with uncertainty of (0.29 ± 0.11) g/cm^2. Top-of-Atmosphere to Bottom-Of-Atmosphere conversion gives uncertainties in surface reflectance up to 0.04 in products with Dense Dark Vegetation pixels.

Finally, the Sentinel-2 time series need to meet the high availability, reliability and timeliness performance requirements to support the Copernicus services. These aspects are particularly challenging for the Sentinel-2 mission given the amount of data produced by the satellite. After one year or ramp-up, the processing baseline (currently 02.04) is essentially stabilized with few anomalous products. New challenges will be faced in the coming months with the arrival of Sentinel-2B and the start of the production with geometric refinement. Operational procedures and monitoring tools are being optimized in order to meet this challenge, taking into account lessons learnt during the ramp-up phase. Innovative solutions could again improve the level of service of Sentinel-2: for instance a machine-to-machine Application Programming Interface (API) could provide real-time information about product anomalies or unavailabilities.

Author Contributions: Jerome Louis and Bringfried Pflug designed and prepared the L2A-part of the paper. Jerome Louis concentrated on Sen2Cor calibration and Bringfried Pflug on L2A validation. Jakub Bieniarz worked on scene classification validation. Ferran Gascon is the ESA technical officer of the MPC activity and the Sentinel-2 data quality manager. Catherine Bouzinac is the coordinator of the MPC Calibration and Validation activities, replacing Olivier Thepaut. Sebastien Clerc is the technical manager of the MPC. Stephane Massera is responsible for the GRI generation. Mathieu Jung is the leader of the L1 calibration activities. Bruno Lafrance is responsible for the radiometric calibration analyses. Benjamin Francesconi is the leader of the L1 validation activities. Francoise Viallefont is responsible for the MTF validation. Bahjat Alhammoud is responsible for the vicarious validation. Laetitia Pessiot is the service manager of the MPC. Vincent Lonjou is the CNES radiometry expert for Sentinel-2. Angélique Gaudel-Vacaresse and Florie Languille are the CNES geometry experts for Sentinel-2. Thierry Tremas is the CNES leader for Sentinel-2 image quality. Enrico Cadau is the ESA Sentinel-2 Mission Management Support Engineer. Roberto De Bonis is the ESA Sentinel-2 Data Processing Support Engineer. Valérie Fernandez is the ESA Sentinel-2 Mission Engineering and Payload Manager. Philippe Martimort was, up to 2015, the Sentinel-2 ESA Mission Engineering and Payload Manager and is currently with ESA Future Missions Division. Claudia Isola is the ESA Sentinel-2 Cal/Val and Instrument Performance Engineer.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Drusch, M.; del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

2. Martin-Gonthier, P.; Magnan, P.; Corbiere, F.; Rolando, S.; Saint-Pe, O.; Breart de Boisanger, M.; Larnaudie, F. CMOS detector for space applications: from R & D to operational program with large volume foundry. *Proc. SPIE* **2010**, *7826*. [CrossRef]

3. Chorvalli, V.; Cazaubiel, V.; Bursch, S.; Welsch, M.; Sontag, H.; Martimort, P.; del Bello, U.; Sy, O.; Laberinti, P.; Spoto, F. Design and development of the Sentinel-2 Multi Spectral Instrument and satellite system. *Sens. Syst. Next-Gener. Satell. XIV* **2010**, *7826*. [CrossRef]

4. Dariel, A.; Chorier, P.; Leroy, C.; Maltère, A.; Bourrillon, V.; Terrier, B.; Molina, M.; Martino, F. Development of a SWIR multi-spectral detector for GMES/Sentinel-2. *Proc. SPIE* **2009**, *7474*. [CrossRef]

5. Chorvalli, V.; Espuche, S.; Delbru, F.; Haas, C.; Astrium, G.; Martimort, P.; Fernandez, V.; Kirchner, V. The multispectral instrument of the Sentinel 2 EM program results. Available online: http://esaconferencebureau.com/custom/icso/2012/papers/FP_ICSO-023.pdf (accessed on 22 May 2017).

6. Copernicus Open Access Hub. Available online: https://scihub.copernicus.eu/ (accessed on 22 May 2017).

7. Thuillier, G.; Hersé, M.; Labs, D.; Foujols, T.; Peetermans, W.; Gillotay, D.; Simon, P.C.; Mandel, H. The solar spectral irradiance from 200 to 2400 nm measured by the SOLSPEC spectrometer from the ATLAS and EUREKA missions. *Sol. Phys.* **2003**, *214*, 1–22. [CrossRef]

8. Keyhole Mark-up Language (KML). Available online: https://sentinels.copernicus.eu/documents/247904/1955685/S2A_OPER_GIP_TILPAR_MPC__20151209T095117_V20150622T000000_21000101T000000_B00.kml (accessed on 22 May 2017).

9. Sentinel-2 Toolbox. Available online: http://step.esa.int/main/toolboxes/sentinel-2-toolbox (accessed on 6 December 2016).

10. Louis, J. Sentinel 2 MSI–Level 2A Product Definition. Available online: https://sentinel.esa.int/documents/247904/1848117/Sentinel-2-Level-2A-Product-Definition-Document.pdf (accessed on 7 September 2016).

11. Schläpfer, D.; Borel, C.C.; Keller, J.; Itten, K.I. Atmospheric precorrected differential absorption technique to retrieve columnar water vapour. *Remote Sens. Environ.* **1998**, *65*, 353–366. [CrossRef]

12. Kaufman, Y.J.; Sendra, C. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *Int. J. Remote Sens.* **1988**, *9*, 1357–1381. [CrossRef]

13. Richter, R.; Louis, J.; Müller-Wilm, U. *Sentinel-2 MSI—Level 2A Products Algorithm Theoretical Basis Document*; S2PAD-ATBD-0001, Issue 2.0; Telespazio VEGA Deutschland GmbH: Darmstadt, Germany, 2012.

14. Data Quality Report. Available online: https://sentinel.esa.int/web/sentinel/data-product-quality-reports (accessed on 22 May 2017).

15. Maisonobe, L.; Pommier-Maurussane, V. Orekit: An Open-source Library for Operational Flight Dynamics Applications. In Proceedings of the 4th ICATT, Madrid, Spain, 3–6 May 2010.

16. Gordon, H.R.; Wang, M. Retrieval of Water-Leaving Radiance and Aerosol Optical Thickness over the Oceans with SeaWiFS: A Preliminary Algorithm. *Appl. Opt.* **1994**, *33*, 443–452. [CrossRef] [PubMed]

17. Morel, A. Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 water). *J. Geophys. Res.* **1988**, *93*, 10749–10768. [CrossRef]

18. Morel, A.; Prieur, L. Analysis of Variations in Ocean Color. *Limnol. Oceanogr.* **1977**, *22*, 709–722. [CrossRef]

19. Hagolle, O.; Goloub, P.; Deschamps, P.Y.; Cosnefroy, H.; Briottet, X.; Bailleul, T.; Nicolas, J.-M.; Parol, F.; Lafrance, B.; Herman, M. Results of POLDER in-flight Calibration. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1550–1566. [CrossRef]

20. Vermote, E.; Santer, R.; Deschamps, P.Y.; Herman, M. In-flight Calibration of Large Field-of-View Sensors at Short Wavelengths using Rayleigh Scattering. *Int. J. Remote Sens.* **1992**, *13*, 3409–3429. [CrossRef]

21. Morel, A.; Maritorena, S. Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.* **2001**, *106*, 7763–7780. [CrossRef]

22. Morel, A.; Gentili, B. Diffuse reflectance of oceanic waters. III. Implication of bidirectionality for the remote-sensing problem. *Appl. Opt.* **1996**, *35*, 4850–4862. [CrossRef] [PubMed]

23. Fougnie, B.; Patrice Henry, P.; Morel, A.; Antoine, D.; Montagner, F. Identification and characterization of stable homogenous oceanic zones: Climatology and impact on in-flight calibration of space sensors over rayleigh scattering. In Proceedings of Ocean Optics XVI, Santa Fe, NM, USA, 18–22 November 2002.

24. Fougnie, B.; Bach, R. Monitoring of Radiometric Sensitivity Changes of Space Sensors Using Deep Convective Clouds: Operational Application to PARASOL. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 851–861. [CrossRef]

25. Slater, P.N.; Biggar, S.F.; Holm, R.G.; Jackson, R.D.; Mao, Y.; Moran, M.S.; Palmer, J.M.; Yuan, B. Reflectance- and radiance-based methods for the in-flight absolute calibration of multispectral sensors. *Remote Sens. Environ.* **1987**, *22*, 11–37. [CrossRef]

26. Thome, K.J.; Helder, D.L.; Aaron, D.; Dewald, J.D. Landsat-5 TM and Landsat-7 ETM+ absolute radiometric calibration using the reflectance-based method. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2777–2785. [CrossRef]

27. Czapla-Myers, J.; McCorkel, J.; Anderson, N.; Thome, K.; Biggar, S.; Helder, D.; Aaron, D.; Leigh, L.; Mishra, N. The Ground-Based Absolute Radiometric Calibration of Landsat 8. *Remote Sens.* **2015**, *7*, 600–626. [CrossRef]

28. Markham, B.; Barsi, J.; Kvaran, G.; Ong, L.; Kaita, E.; Biggar, S.; Czapla-Myers, J.; Mishra, N.; Helder, D. Landsat-8 operational land imager radiometric calibration and stability. *Remote Sens.* **2014**, *6*, 12275–12308. [CrossRef]

29. Bouvet, M. Radiometric comparison of multispectral imagers over a pseudo-invariant calibration site using a reference radiometric model. *Remote Sens. Environ.* **2014**, *140*, 141–154. [CrossRef]

30. Alhammoud, B.; Bouvet, M.; Jackson, J.; Arias, M.; Thepaut, O.; Lafrance, B.; Gascon, F.; Cadau, E.; Berthelot, B.; Francesconi, B. On the vicarious calibration methodologies in DIMITRI: Applications on Sentinel-2 and Landsat-8 products and comparison with in-situ measurements. In Proceeding of the ESA Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016.

31. Holben, B.N.; Kaufman, Y.J.; Kendall, J.D. NOAA-11 AVHRR visible and near IR in-flight calibration. *Int. J. Remote Sens.* **1990**, *11*, 1511–1519. [CrossRef]

32. CEOS-PICS. Available online: http://calvalportal.ceos.org (accessed on 22 May 2017).

33. Cosnefroy, H.; Leroy, M.; Briottet, M. Selection and characterisation of saharan and Arabian desert sites for the calibration of optical satellite sensors. *Remote Sens. Environ.* **1996**, *58*, 2713–2715. [CrossRef]

34. International Ocean-Colour Coordinating Group (IOCCG). *In-Flight Calibration of Satellite Ocean-Colour Sensors*; Frouin, R., Ed.; IOCCG: Dartmouth, Canada, 2013.

35. Kääb, A.; Winsvold, S.H.; Altena, B.; Nuth, C.; Nagler, T.; Wuite, J. Glacier Remote Sensing Using Sentinel-2. Part I: Radiometric and Geometric Performance, and Application to Ice Velocity. *Remote Sens.* **2016**, *8*, 598. [CrossRef]

36. Lacherade, S.; Fougnie, B.; Henry, P.; Gamet, P. Cross calibration over desert sites: Description, methodology, and operational implementation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1098–1113. [CrossRef]

37. Bouvet, M. Intercomparison of imaging spectrometer over the Salar de Uyuni (Bolivia). In Proceedings of the 2006 MERIS AATSR Validation Team Workshop, Frascati, Italy, 20–24 March 2006.

38. Adriaensen, S.; Barker, K.; Bourg, L.; Bouvet, M.; Fougnie, B.; Govaerts, Y.; Henry, P.; Kent, C.; Smith, D.; Sterckx, S. CEOS IVOS Working Group 4: Intercomparison of vicarious calibration methodologies and radiometric comparison methodologies over pseudo-invariant calibration sites. Available online: http://calvalportal.ceos.org/ceos-wgcv/ivos/wg4/final-report (accessed on 22 May 2017).

39. Barsi, J.A.; Schott, J.R.; Hook, S.J.; Raqueno, N.G.; Markham, B.L.; Radocinski, R.G. Landsat-8 Thermal Infrared Sensor (TIRS) Vicarious Radiometric Calibration. *Remote Sens.* **2014**, *6*, 11607–11626. [CrossRef]

40. Van der Werff, H.; van der Meer, F. Sentinel-2A MSI and Landsat 8 OLI Provide Data Continuity for Geological Remote Sensing. *Remote Sens.* **2016**, *8*, 883. [CrossRef]

41. Vermote, E.F.; Tanré, D.; Deuzé, J.L.; Herman, M.; Morcrette, J.-J. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An Overview. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 675–686. [CrossRef]

42. Berk, A.; Anderson, G.P.; Acharya, P.K.; Bernstein, L.S.; Muratov, L.; Lee, J.; Fox, M.; Adler-Golden, S.M.; Chetwynd, J.H.; Hoke, M.L.; et al. MODTRAN™ 5, A Reformulated Atmospheric Band Model with Auxiliary Species and Practical Multiple Scattering Options: Update. *Algorithms Technol. Multispectr. Hyperspectr. Ultraspectr. Imag.* **2005**, *5655*. [CrossRef]

43. Viallefont-Robinet, F.; Léger, D. Improvement of the edge method for on-orbit MTF measurement. *Opt. Express* **2010**, *18*, 3531–3545. [CrossRef] [PubMed]

44. Fourest, S.; Briottet, X.; Lier, P.; Valorge, C. *Satellite Imagery From Acquisition Principle to Processing of Optical Images for Observing the Earth*; CEPADUES Editions: Toulouse, France, 2012.

45. Holben, B.N.; Eck, T.F.; Slutker, I.; Tanré, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [CrossRef]

46. *S2MPC—Calibration and Validation Plan for the MPC Routine Operation Phase*; S2-PDGS-MPC-CCVP, issue 12; Telespazio VEGA Deutschland GmbH: Darmstadt, Germany, 2017.

47. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]

48. Pflug, B.; Main-Knorn, M.; Makarau, A.; Richter, R. Validation of aerosol estimation in atmospheric correction algorithm ATCOR. In Proceedings of the 36th International Symposium on Remote Sensing of Environment (ISRSE), Berlin, Germany, 11–15 May 2015. [CrossRef]

49. Breon, F.M.; Vermeulen, A.; Descloitres, J. An evaluation of satellite aerosol products against sunphotometer measurements. *Remote Sens. Environ.* **2011**, *115*, 3102–3111. [CrossRef]

50. Vuolo, F.; Zółtak, M.; Pipitone, C.; Zappa, L.; Wenng, H.; Immitzer, M.; Weiss, M.; Baret, F.; Atzberger, C. Data Service Platform for Sentinel-2 Surface Reflectance and Value-Added Products: System Use and Examples. *Remote Sens.* **2016**, *8*, 938. [CrossRef]

51. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary Analysis of the Performance of the Landsat 8/OLI Land Surface Reflectance Product. *Remote Sens. Environ.* **2015**, *185*, 46–56. [CrossRef]

52. Dörnhöfer, K.; Göritz, A.; Gege, P.; Pflug, B.; Oppelt, N. Water Constituents and Water Depth Retrieval from Sentinel-2A—A First Evaluation in an Oligotrophic Lake. *Remote Sens.* **2016**. [CrossRef]

53. Toming, K.; Kutser, T.; Laas, A.; Sepp, M.; Paavel, B.; Nõges, T. First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sens.* **2016**. [CrossRef]