Article

Copernicus Sentinel-2 Calibration and Products Validation Status

Ferran Gascon [1,*], **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], **Valérie Fernandez** [1]

[1] ESA (European Space Agency), 00044 Frascati, Italy; ferran.gascon@esa.int; enrico.cadau@esa.int; roberto.debonis@serco.com; claudia.isola@esa.int; philippe.martimort@esa.int; valerie.fernandez@esa.int,

[2] CS-SI (Communication & Systèmes - Systèmes d'Information), 31500 Toulouse, France; olivier.thepaut@c-s.fr; bruno.lafrance@c-s.fr; laetitia.pessiot@c-s.fr,

[3] Airbus Defence and Space, 31030 Toulouse, France; mathieu.jung@airbus.com,

[4] Thales Alenia Space, 06150 Cannes, France; benjamin.francesconi@thalesaleniaspace.com,

[5] Telespazio, 31100 Toulouse, France; jerome.louis@telespazio.com,

[6] CNES (Centre National d'Etudes Spatiales), 31400 Toulouse, France; vincent.lonjou@cnes.fr; angelique.gaudel-vacaresse@cnes.fr; florie.languille@cnes.fr; thierry.tremas@cnes.fr,

[7] IGN (Institut Géographique National) Espace, 31520 Ramonville-Saint-Agne, France; stephane.massera@ign.fr

[8] ARGANS, Plymouth PL6 8BX, UK; balhammoud@argans.co.uk; sebastien.clerc@acri-st.fr,

[9] ONERA (Office National d'Etudes et Recherches Aérospatiales), 31055 Toulouse, France; francoise.viallefont@onera.fr

[10] DLR (Deutschen Zentrums für Luft- und Raumfahrt), 12489 Berlin, Germany; bringfried.pflug@dlr.de; jakub.bieniarz@dlr.de,

* Correspondence: ferran.gascon@esa.int; Tel.: +39-06-941-88605

Abstract: As part of the Copernicus programme of the European Union (EU), the European Space Agency (ESA) has developed and is currently operating the Sentinel-2 mission that is acquiring high spatial resolution optical imagery. This paper provides a description of the calibration activities and the current status of the mission products validation activities. Measured performances, from the validation activities, cover both Top-Of-Atmosphere (TOA) and Bottom-Of-Atmosphere (BOA) products. Results presented in this paper 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 Union (EU), the European Space Agency (ESA) has developed and is currently operating the Sentinel-2 mission acquiring high spatial resolution 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 coastal areas, a high revisit of five days under the same viewing conditions, high spatial resolution, and a wide field of view 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. 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 km², 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 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 for both Top-Of-Atmosphere (TOA) and Bottom-Of-Atmosphere (BOA) products (referred respectively as Level-1 and Level-2A and further described later 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 on Figure 1. One focal plane includes the Visible and Near-Infrared (VNIR) bands and the other one the Short-Wave Infrared (SWIR) bands.
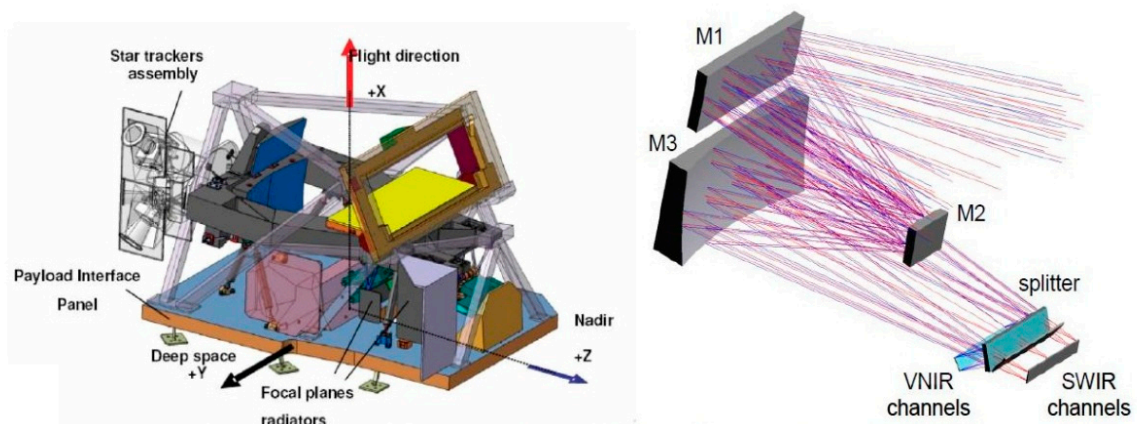

Figure 1. MSI internal configuration. On the left, full instrument view and optical path construction to the SWIR/VNIR (see §2.2) splitter and focal planes.

2.2. Spectral Bands and Resolution

The MSI performs measurements on 13 spectral bands spread over the VNIR and SWIR domains. These spectral channels include:

- 4 bands at 10 m: blue (490 nm), green (560 nm), red (665 nm) and near infrared (842 nm).
- 6 bands at 20 m: 4 narrow bands for vegetation characterization (705 nm, 740 nm, 783 nm and 865 nm) and 2 larger 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 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.
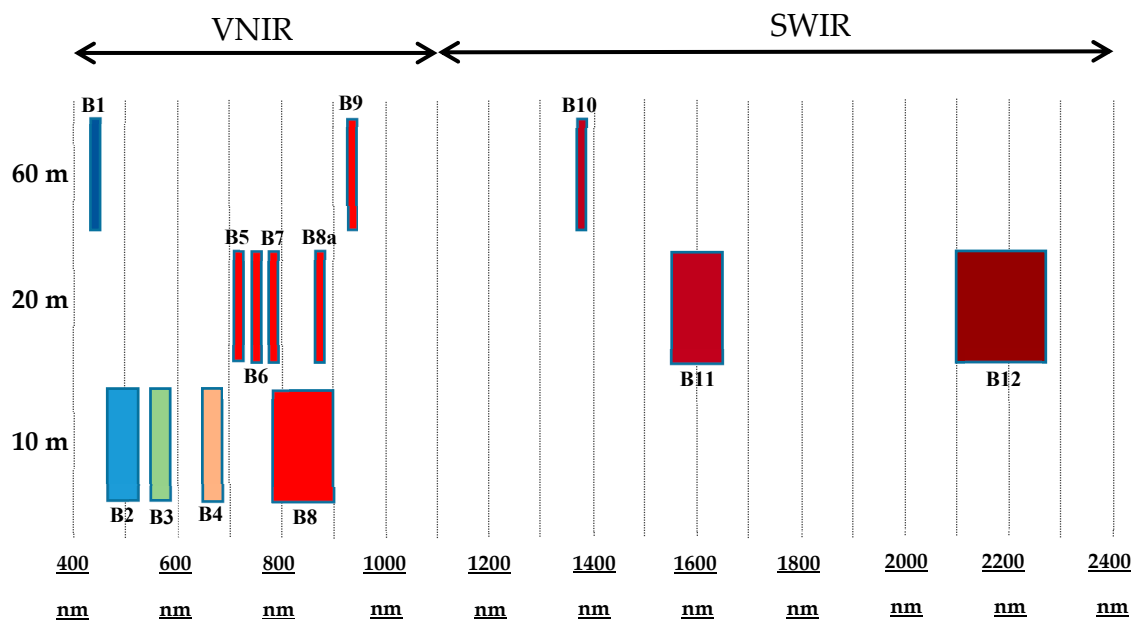

Figure 2. MSI spectral bands vs. spatial resolution with corresponding Full Width at Half Maximum (FWHM)

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 staggered-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 (ACT) at the satellite reference altitude of 786 km.

As illustrated on 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 (ALT) of about 48 km (maximum) and 14 km (minimum). Furthermore, two successive detectors share an overlap area of 2 km ACT.

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 (ALT) displacement of about 17 km between B10 and B12.
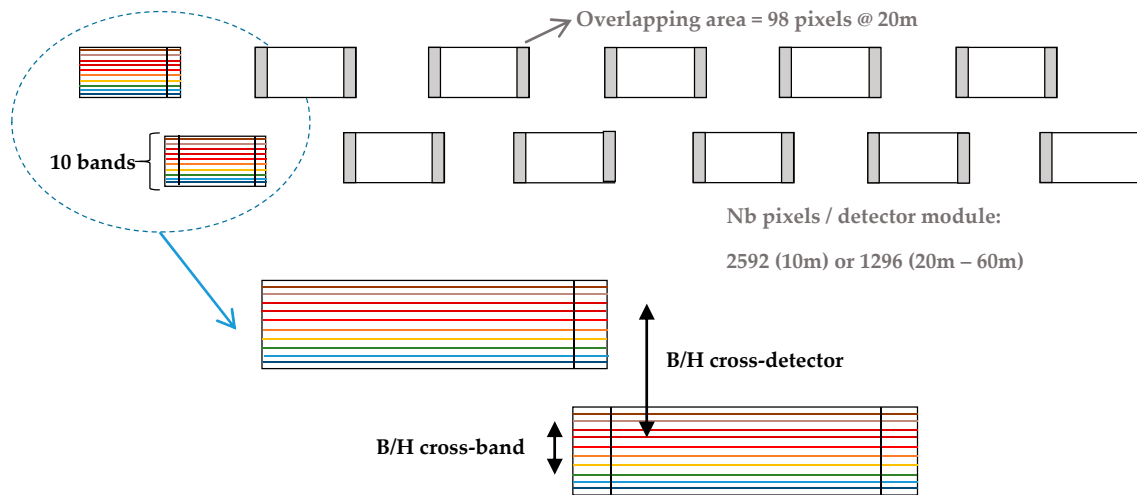
Figure 3. Staggered detector configuration

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 surface/atmosphere 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 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}). Following Figure 4 shows the definition of the LOS.
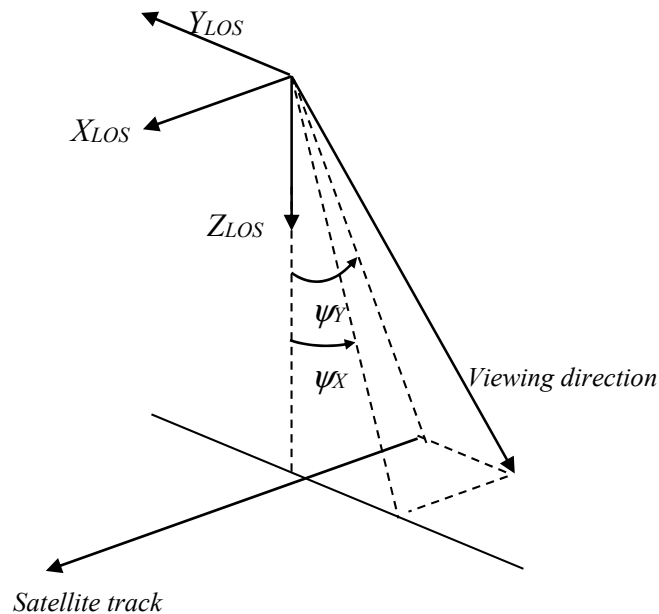

Figure 4. Line of Sight angles definition.

$$\vec{Z}(p) = \frac{1}{\sqrt{1 + \tan(\psi_X(p))^2 + \tan(\psi_Y(p))^2}} \begin{bmatrix} \tan(\psi_Y(p)) \\ -\tan(\psi_X(p)) \\ 1 \end{bmatrix} \tag{1}$$

This expression allows to precisely defining 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 modelling of the earth including a Digital Elevation Model.

Following 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}).
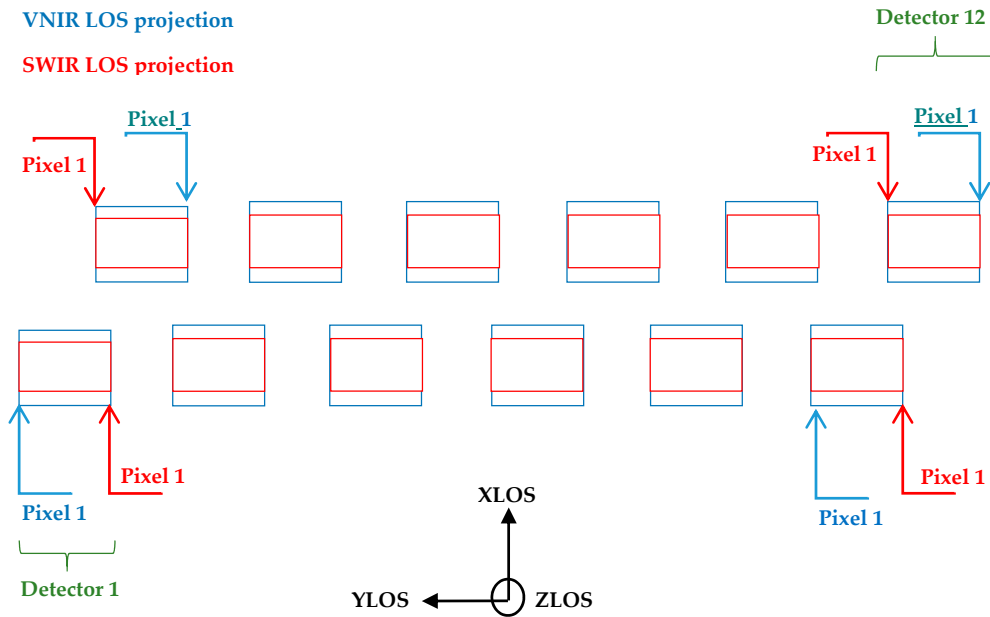
Figure 5. Representation of the pixels LOS projection on ground.

In so far as products are delivered to the used at Level 1C, the MSI pixel number specificities are just given here for information.

2.4. Detector Specificities

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 20m ACT and binning is applied on-ground. The sampling time is also depending on the band. Table 1 synthesizes these instrument specificities.

Table 1. Detector configuration

Band	Resolution (m)	Integration Time (ms)	Number of detector lines	Number of lines after selection	Number of pixels by detector	X-size (across-track) (µm)	Y-size (along-track) (µm)
B1	60	9,396	1	1	1296	15	45
B2	10	1,566	1	1	2592	7,5	7,5
B3	10	1,566	2	2	2592	7,5	7,5
B4	10	1,566	2	2	2592	7,5	7,5
B5	20	3,132	1	1	1296	15	15
B6	20	3,132	1	1	1296	15	15
B7	20	3,132	1	1	1296	15	15
B8	10	1,566	1	1	2592	7,5	7,5
B8a	20	3,132	1	1	1296	15	15
B9	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 on 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.

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.

Figure 6. TDI configuration examples for pixels of SWIR bands. Each column corresponds to one pixel. The lines correspond to available TDI lines. A cross corresponds to the TDI line selected for each pixel.

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.

2.5. On board equalization and compression

Images are compressed on board in order 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 component in charge of these operations is called WICOM.

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 in a discrete wavelet transform (DWT). Before compression, a Non Uniformity Correction (NUC) function is applied on the 12 bits on-board image. This function consists in 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 this table, the compression ratios are pretty low so that the image quality is fully preserved.

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 mechanism is designed to collect the sunlight after reflection by a diffuser, and to prevent a direct viewing of the Sun or contamination during launch and early operation phase (LEOP).

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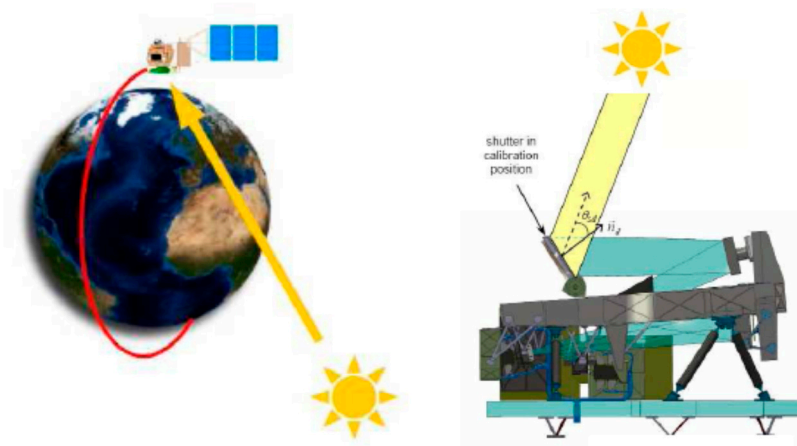

Figure 7. Illustration of the sun diffuser acquisition principle

This CSM is periodically used (typically every month) in order 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.

- 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. Systematic production and dissemination by ESA is planned for the future.

3.2. Level-1 Processing steps

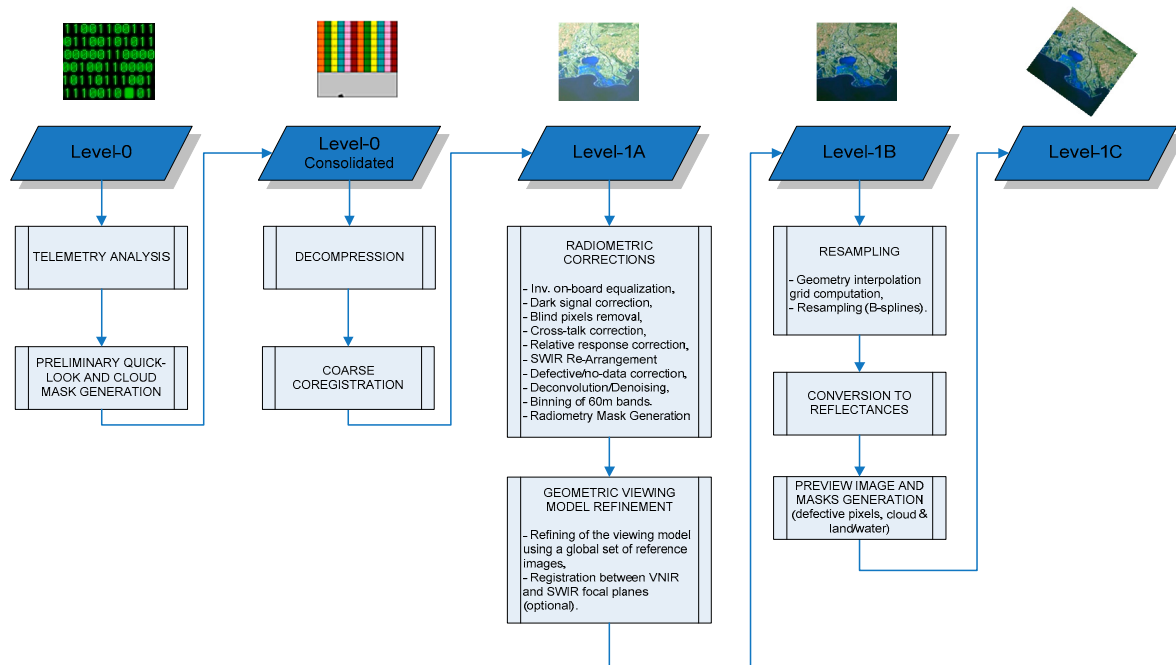The main steps of the processing chain are summarized in Figure 8 below.

Figure 8. MSI processing chain overview

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 in order 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). An optional deconvolution step could be performed in case of poor image quality (currently not used). 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, comes the geometric refining. 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). 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). The pixel value for each spectral band is interpolated using B-splines.

Conversion from radiances to reflectances relies on the Solar Irradiance model of [2].

Finally, quality masks -derived from the L1B masks- as well as opaque and cirrus clouds masks -based on radiometric thresholds- are generated. 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 recalled 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.
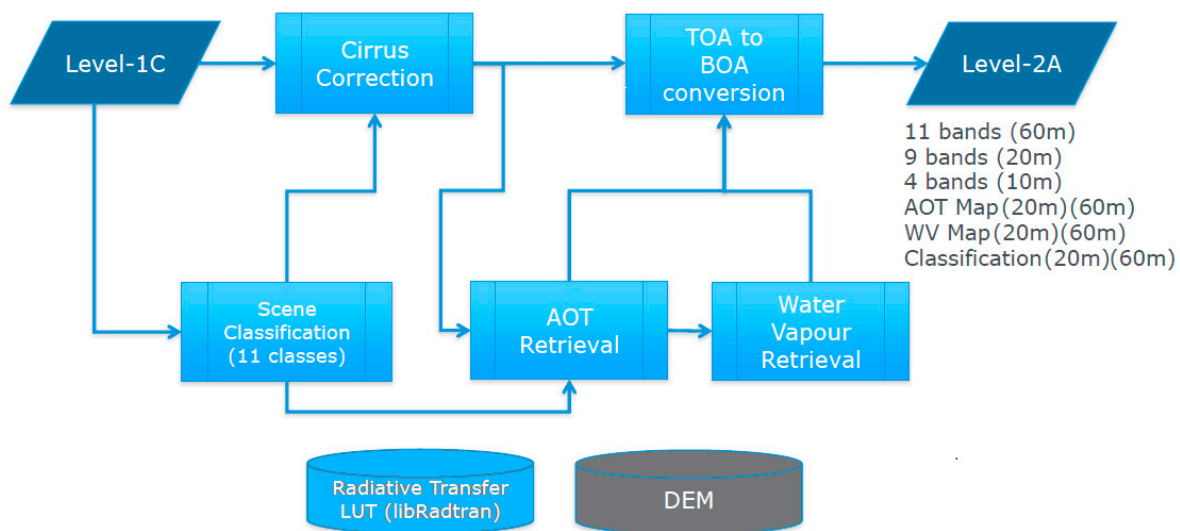

Figure 9. Level-2A processing chain overview

Sen2Cor relies on two main auxiliary data: the Radiative Transfer Look-Up Tables (LUTs) and the Digital Elevation Model (DEM), which one is not embedded in the Sen2Cor software but can be provided by the user in DTED format. The user can also rely on the default Shuttle Radar Topography Mission Digital Elevation Model (SRTM-DEM). Level-2A outputs are a Scene Classification (SCL) image together with Quality Indicators for cloud and snow probabilities, AOT and WV maps and the surface (or BOA) reflectance images which are provided at different spatial resolutions (60 m, 20 m and 10 m).

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 x 110 km² footprint of each tile can be obtained in a Keyhole Mark-up Language (KML) file from

Sentinel-2 website [3]. 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 Datastrip (elementary processing unit) and user product (dissemination unit). Datastrip 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 may include auxiliary data such as meteorological data or processing parameter files.

3.5. Level-2A Products Description

L2A surface reflectance products are generated with Sen2Cor processor whose main purpose is to correct single-date Sentinel-2 L1C products from the effects of the atmosphere. Sen2Cor processor is available as a third-party plugin of the Sentinel-2 Toolbox [4].

L2A products provide:

- Cloud Screening and Pixel Categorization (CSC) map together with Quality Indicators for cloud and snow probabilities,
- Aerosol Optical Thickness (AOT) and Water Vapour (WV) maps
- Surface (or BOA) reflectance images which are provided at different spatial resolutions (60 m, 20 m and 10 m).

The structure of the Level-2A product [5] 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 Pixel Categorization map is available at the root of the IMG_DATA folder at 20 m or 60 m resolution and the cloud and snow probabilities are located in the QI_DATA folder.

The Cloud Screening and Categorization module provides a Pixel Categorization map divided in 11 classes presented in chapter 5. 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 by the APDA (Atmospheric Pre-corrected Differential Absorption) algorithm [6] 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) is estimated using the Dark Dense Vegetation (DDV) pixel method introduced by Kaufmann [7]. The actual Sen2Cor retrieval method is described in [8]. The quantification value to convert Digital Numbers to AOT is equal to 1000.

The surface reflectance values are coded in JPEG2000 with the same quantification value of 10000 as for Level-1C products, i.e. a factor of 1/10000 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 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.

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, evolutions of calibration coefficient do not lead to a change of version as long as the quality of user products is unaffected.

The evolutions of the processing baseline are tracked and justified in the Data Quality Report published on a monthly basis [9]. 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 an incoming light. Its measurement is realized by processing images acquired by night over oceans with the CSM opened, in order 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 acquisition phase). Currently, dark signal acquisitions are performed at least once per 10-day orbit cycle.

Starting from Level-0c products, the images are uncompressed and reversed from the on-board processing in order to retrieve the native values of the measurements. Dark acquisitions last 40

seconds. The average of digital counts among 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-flight. The dark current is particularly stable for VNIR bands: the variation of the measured dark coefficients between two acquisitions is always lower than 1 digit (for extreme variations) and the dark signal noise is about 0.5 DC. Table 3 illustrates this stability by a statistical comparison of the dark coefficients estimated from dark images acquired on 9 May and 8 June 2016. For SWIR bands, the variations of the dark signal can be higher, mainly for the B12 band for which the maximal variations can reach by about 5 digits for a few number of pixel, and the pixel relative variation is noticeable (the standard deviation of variations over pixels is usually about ten times higher for B12 than for VNIR bands). The dark noise for SWIR bands is also slightly higher than for VNIR bands, with mean values by about 0.7 to 1 digit.

Figure 10 and Figure 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 higher for SWIR bands (mainly for B12 as plotted on Figure 11).

Table 3. On the left, statistics on dark signal variations between the dark estimate on 8 June 2016 and the operational dark coefficients deduced from an acquisition on 9 May 2016. On the right, statistics on the dark noise for estimate on 8 June 2016. 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
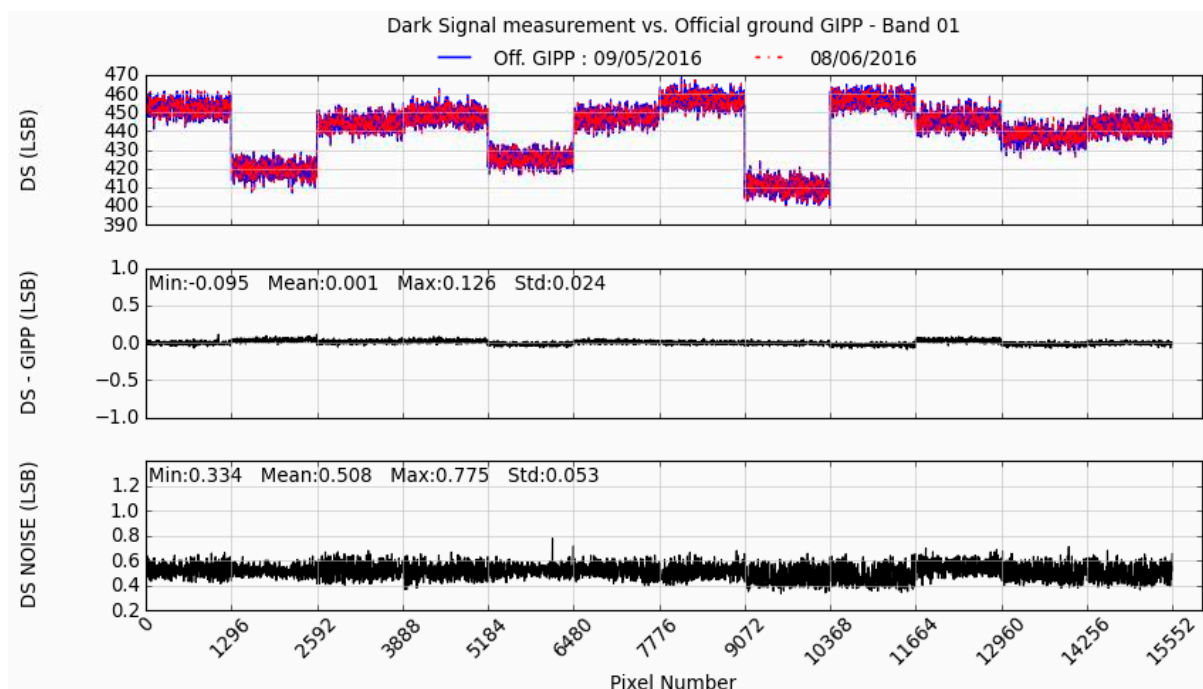
Figure 10. Variation of the dark signal measured on 8 June 2016, for the B01 band, with respect to the coefficients calculated from dark acquisition on 9 May 2016. The first plot shows the dark signal level (in digital counts, LSB) as a function of the pixel number, the second plot shows its variation between the two dates, the third plot shows its noise (in LSB).
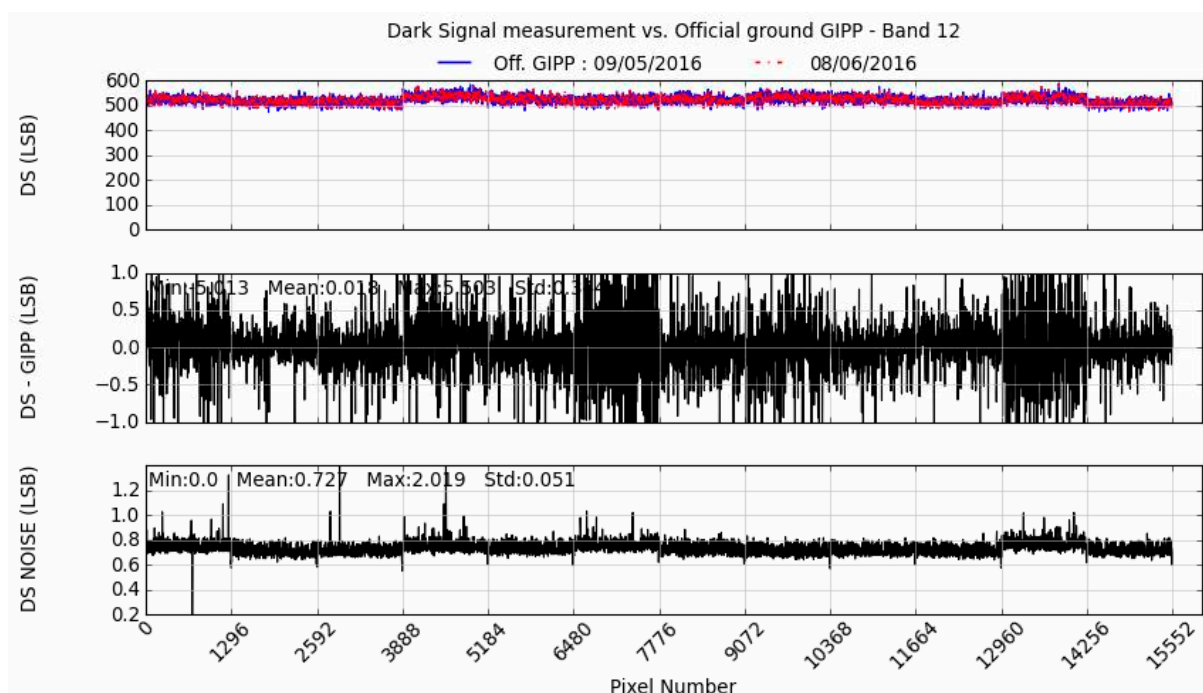

Figure 11. Same legend as Figure 10 but for the B12 band.

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, 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 modelling of the reflexion of the solar light on the sun-diffuser considers the non-uniformity and bidirectional reflectivity of the diffuser from a pre-flight 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. This leads to a simulation of the radiance illuminating the MSI sensor which is given, for each band *b* and each detectord *d*, by:

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

The solar irradiance used as reference for Sentinel-2 comes from [2]. 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 [10].

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

The absolution calibration coefficient, for each band *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:

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

The non-valid pixels are not taken into account in the calculation of the average value. SWIR bands 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.

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.
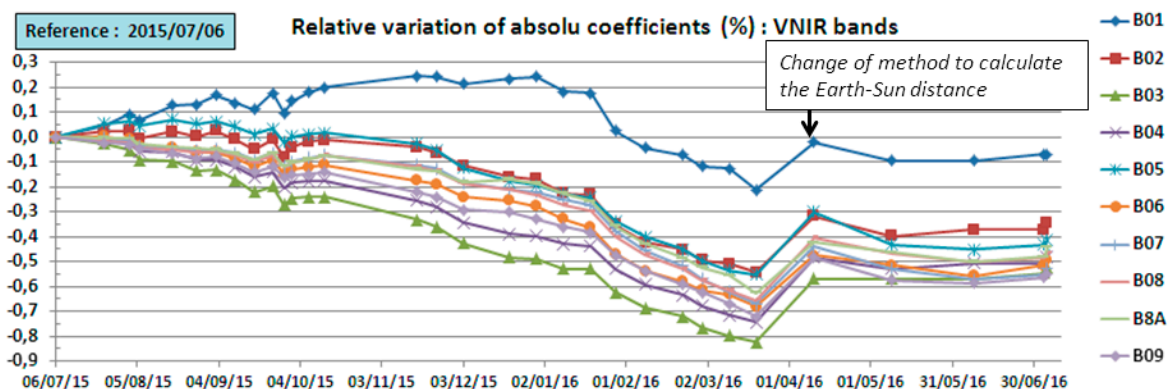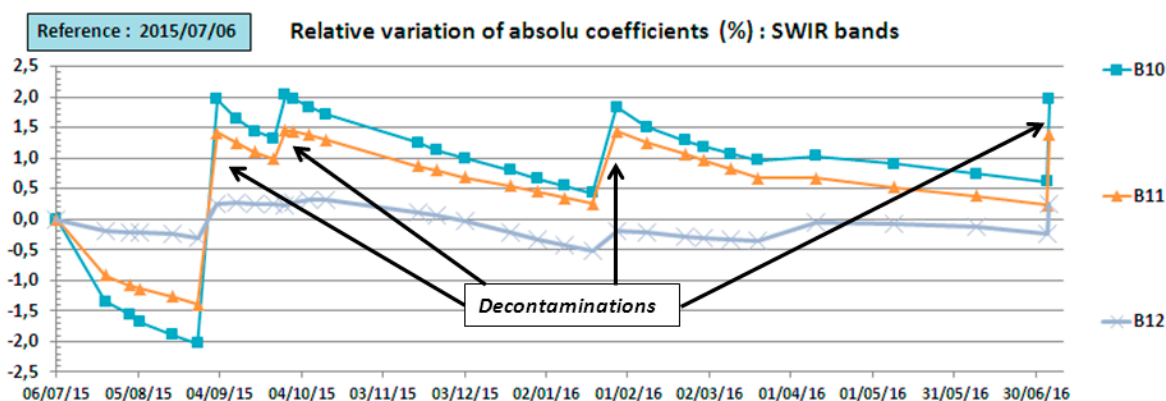
Figure 12. Time variation of the official absolute calibration coefficients (for the processing baseline version 02.04), normalized by the coefficients estimated from the first sun-diffuser acquisition performed on 6 July 2015.

The variation of the absolute calibration coefficients for VNIR bands is below 0.8 % from the 6 July 2015 (reference date) until July 2016. Except for the B01 band, we notice a decrease of the sensitivity 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. It is foreseen to reprocess all the time serie of sun-diffuser acquisitions in order to correct this slight seasonal effect during a future data reprocessing. Since April 2016, absolute calibration coefficients have remained stable by about 0.1 % for VNIR bands.

The variation of the absolute calibration coefficients for SWIR bands is higher, mainly for the B10 and B11 bands. As mentioned in section 0, 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 manufacturing, a tiny amount of water vapour was trapped in the optics 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 is 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 minutes (however, the MSI is unavailable for nominal acquisitions during around 15.5 hours, i.e. 3 hours for heating up, 1.5 hour with regulated decontamination temperature and 11 hours to recover the operational temperature). The resulting effect is clearly noticeable on 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). It is completely compensated after decontamination.

4.1.3. Relative Gains Calibration

The measurements are equalised by pre-defined functions $\gamma(b, d, p, Y)$ which were estimated pixel by pixel on pre-flight characterisations and are updated in-flight from sun-diffuser acquisitions. These functions deal both with the non-linearity of the sensor response and with the inter-pixels differences of response. They are defined by a cubic model for VNIR bands and by a

bi-linear model for SWIR bands. We focus here on 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 :

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

In-flight, 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 relation (2). We consider the average value over lines, giving a simulated radiance by pixel:

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

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:

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

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):

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

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

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

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 illustrated in Figure 24. 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 8 June 2016 and the previous update of GIPPs from the acquisition on 9 May 2016. Minimal, mean and maximal values of R_a and its standard deviation over all pixels are given for each spectral band.

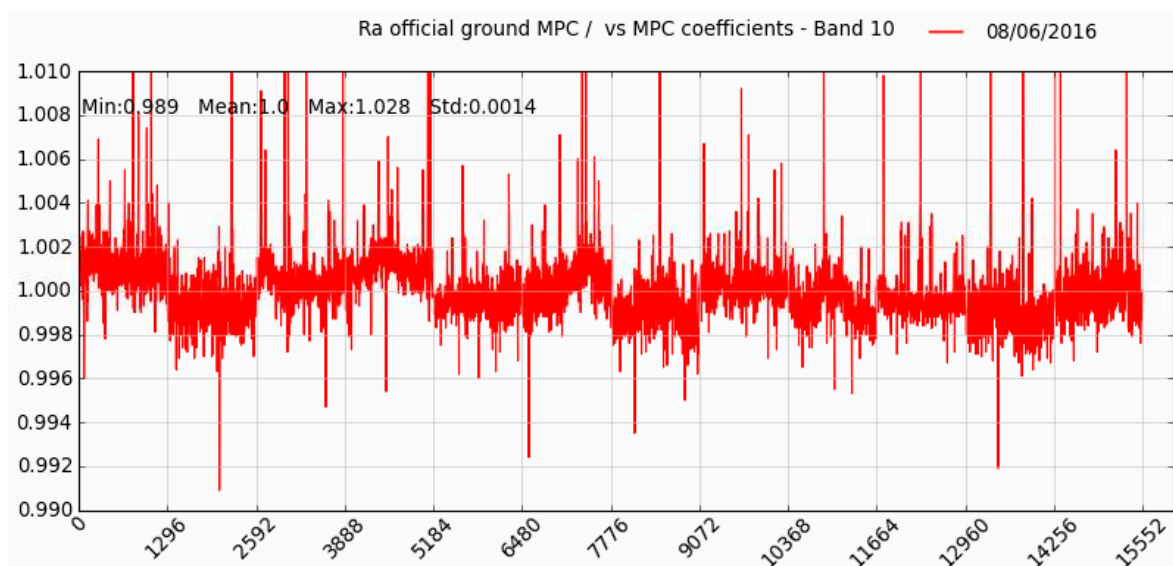| 08/06/2016	Min	Mean	Max	Std
B01	0,999	1,000	1,001	0,0004
B02	0,999	1,000	1,001	0,0003
B03	0,999	1,000	1,001	0,0004
B04	0,999	1,000	1,001	0,0004
B05	0,999	1,000	1,002	0,0006
B06	0,999	1,000	1,001	0,0004
B07	0,999	1,000	1,001	0,0005
B08	0,999	1,000	1,002	0,0005
B8A	0,999	1,000	1,002	0,0005
B09	0,999	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 for the updating 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.

Prior to the launch, each pixel of each TDI line has been tested on ground and affected with 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 this 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 in order 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.
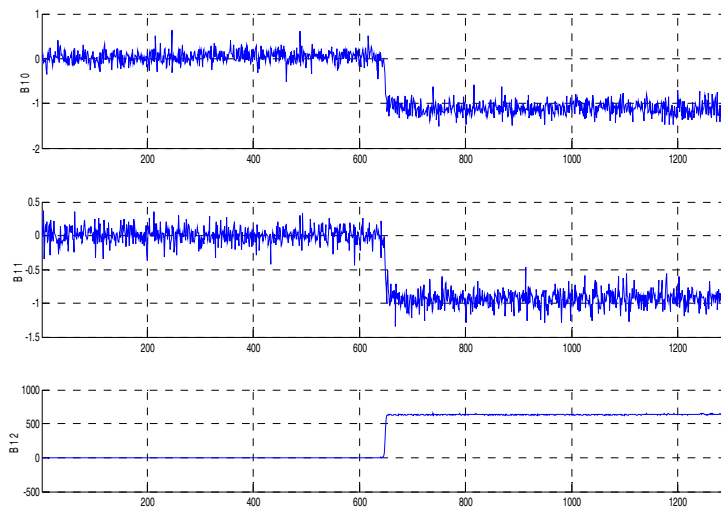

Figure 14. Characterisation of the electrical crosstalk**:** example of test result from ground measurement at Focal Plane Assembly (FPA) level

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

Table 5. Average SWIR electrical crosstalk

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

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.

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.

Example of such a crosstalk on a real landscape is given 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.
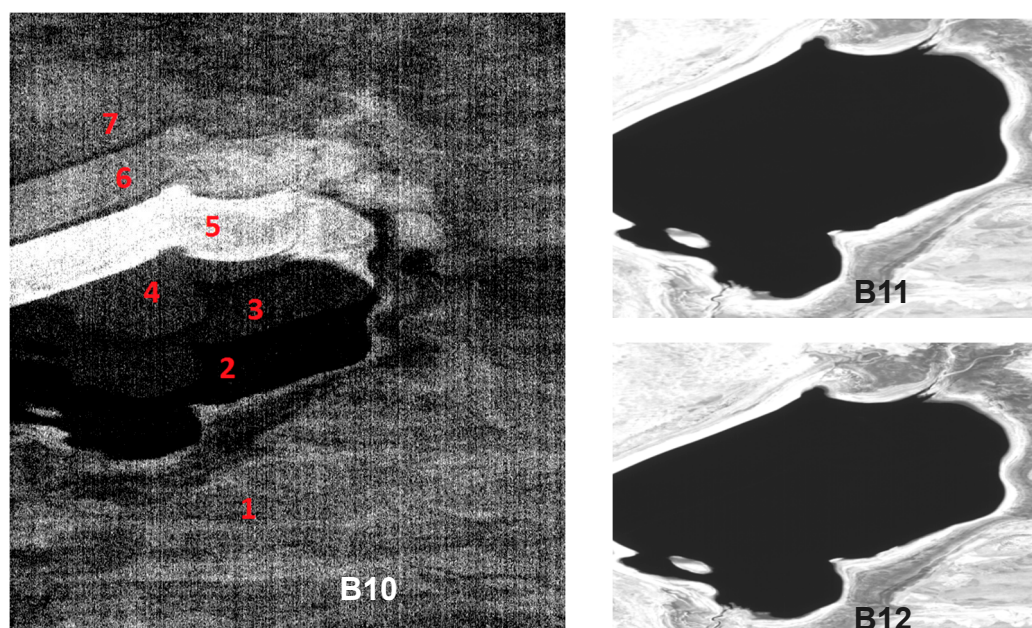

Figure 15. Left image shows an **e**xample of crosstalk affecting B10. Right represent the corresponding acquisitions in B11 and B12.

On 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.
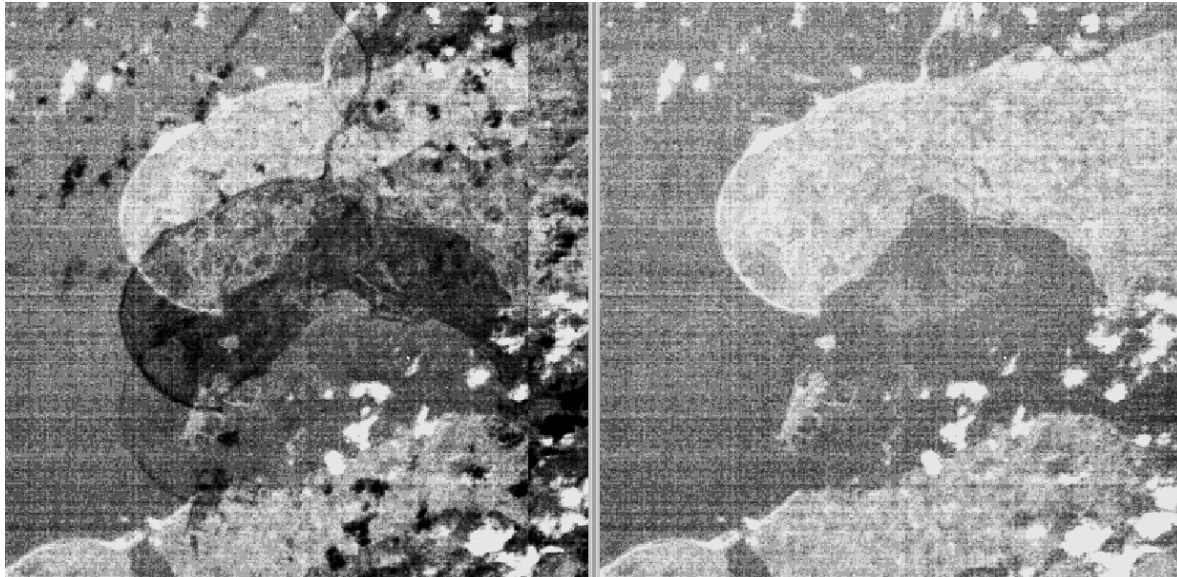

Figure 16. Example of crosstalk correction (before/after)

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 (M. Jung)

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 MTF assessment). This degradation is not expected, even considering the ageing of the instrument.

Nevertheless, if required, the MSI would 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, in order 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 monospectral (B4, red channel) L1B products, whose geometrical model has been refined through a dedicated process designed by IGN. 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

- 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, a low cloud coverage does not mean that the part of the image we are interested in is really cloud free. Conversely, products with a 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. Quicklook images at 320 m 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.
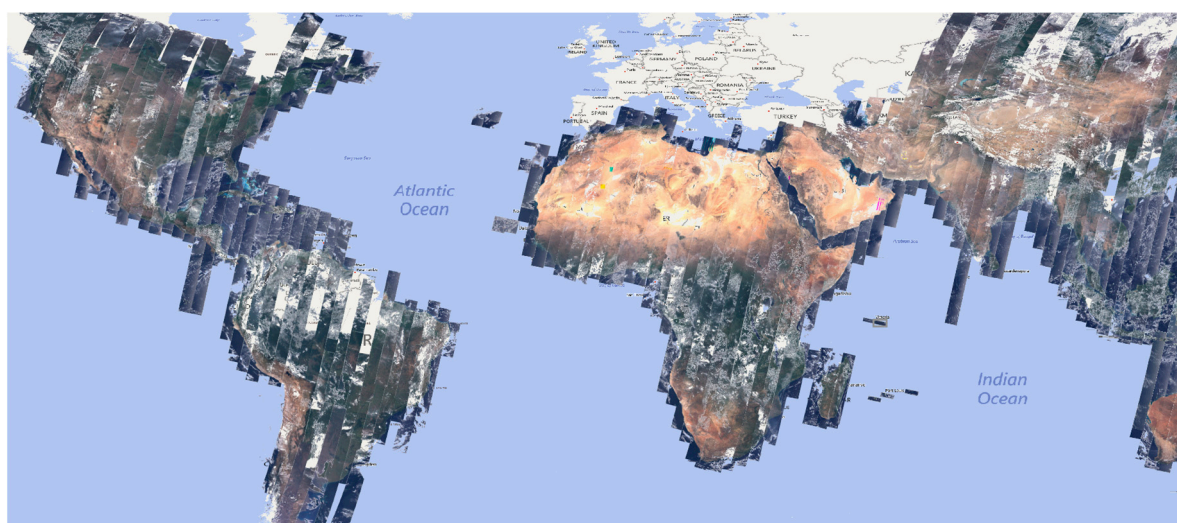

Figure 17. Overview of the Sentinel-2 GRI selection, July 2016 (European GRI products are not present on this map, since they were produced in the In-Orbit Commissioning Review (IOCR) context, in October 2015)

The GRI will not be made of a 100 % cloud free products: we just need to find enough GCPs (Ground Control Points) in the GRI to refine other L1B products. Since segments are very long, some cloudy parts can be accepted by the registration method. Moreover equatorial areas are cloudy most of the time, and there are very few chances to get a cloud free image over there.

Figure 18 below 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 the clouds, and no satisfying images could be found yet. From a Sentinel-2 point of view, it may not be a real issue since there are enough cloud free data and long image strips to retrieve a satisfying density of GCPs. This

also illustrates that the GRI will not be an aesthetical view of the world, or a complete database to use as is.
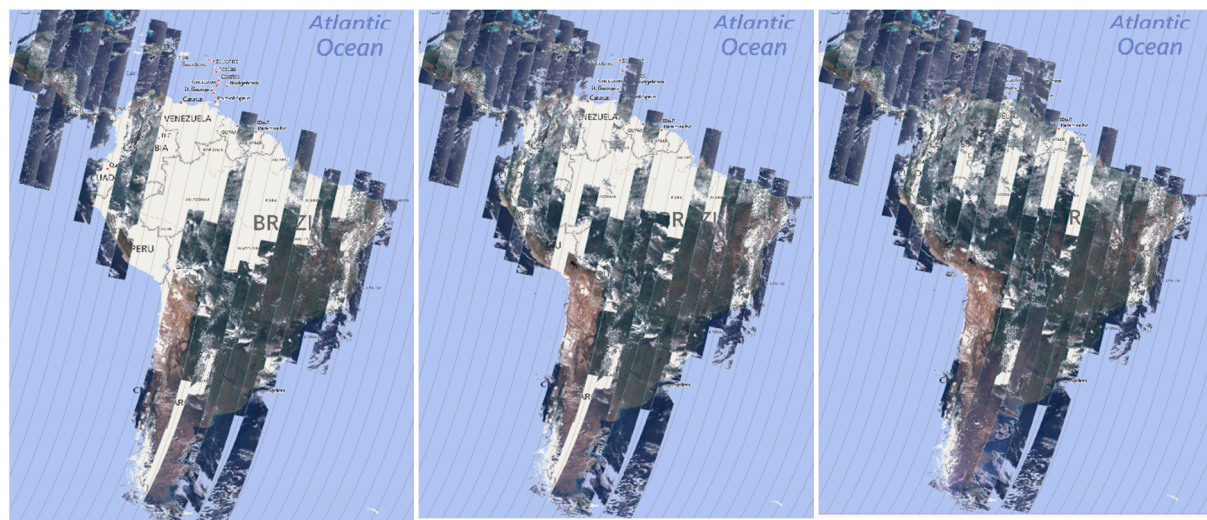

Figure 18. Monthly evolution of the image selection of GRI (Apr to Jul 2016)

The complete GRI is expected to comprise around 1,000 products. This includes numerous products over isolated islands.

4.2.1.3. Methods of refining

Basically, the method used to refine L1B products is the block spatiotriangulation: it provides refined values of the attitude of the satellite by using observations from ground and images (GCPs and tie points). To cover the whole world, the calculation process will separately concern continental or multi continental blocks such as Europe, Africa, Asia, Australia, North and South America, and also most of isolated islands.

Tie points will be extracted using SURF (Speeded Up Robust Features) algorithms: SURF method main interest lays in the nature of the detected points, which are more robust, less sensitive to shadow effects, and provide in general better results than the usual points of interest used in a classical correlation process (e.g. Harris points). These homologous points are very numerous and located everywhere there is an overlap: along the same orbit when available and between 2 (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 network is worldwide. When possible, GCPs located on 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 use case of little 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 spatiotriangulation (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 will be respected.

Using these observations (tie points and GCPs), it is possible to refine the chosen set of geometrical parameters (attitudes, position): it is the spatiotriangulation step itself, i.e. the resolution of the least square system. At this step, outlier observations are removed automatically. One more time, the initial geolocation of the satellite is very good. Consequently, 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.

4.2.1.1. Internal controls

Even though it is not the task of IGN to make the official control of its own production, it seems essential to deliver the GRI to the validation team after it has been pre-checked in house. 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 our 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. Finally, every point has a quality index, computed from several parameters, such as geometrical accuracy, multiplicity, durability, and other criteria related to the interest point itself (SURF). All available 200 x 200 pixels Spot 5 HRS cropped and centred images are associated to each GCP.

Finally, 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 will be 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 5,000 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 our checking operations. CNES team, that made other checks from Pleiades imagery in many local areas, confirmed this figure.
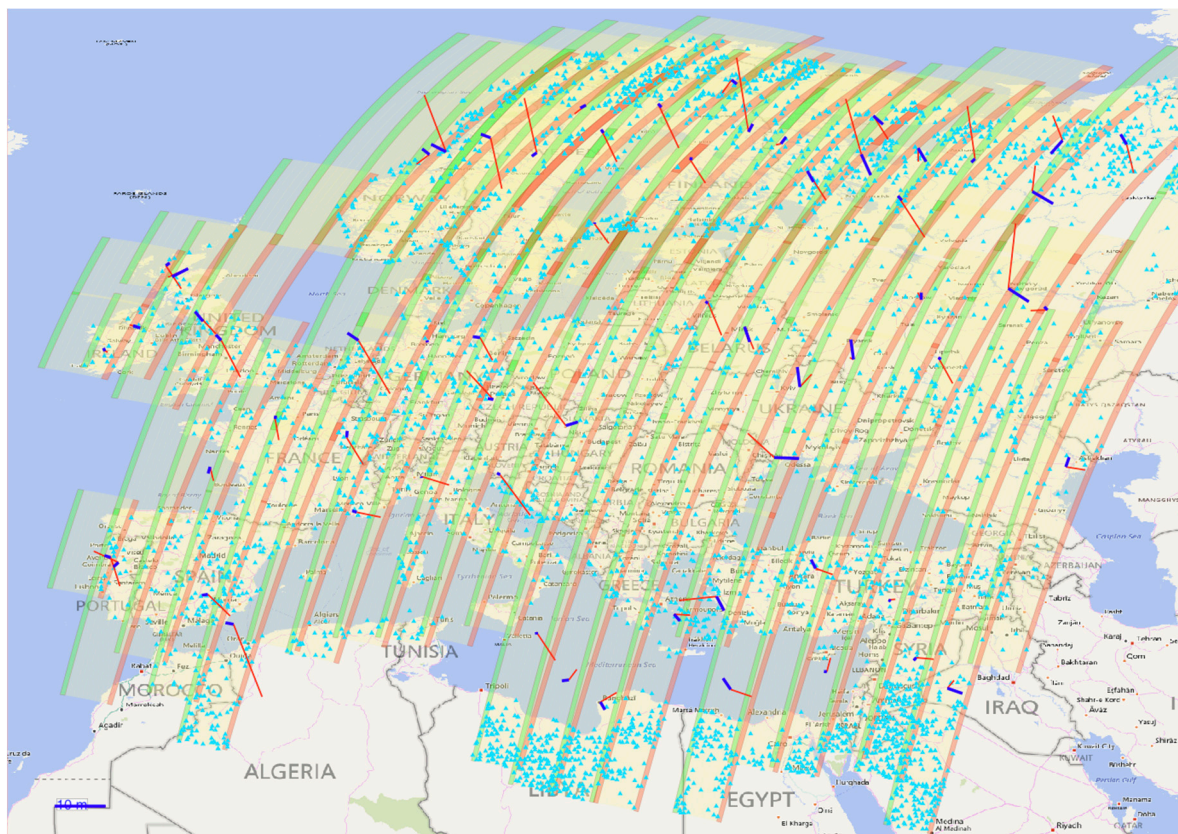
Figure 19. Checks from Reference 3D® / BDAmer (blue dots) data over Europe. Red and blue arrows: accuracy of the product before and after refining.

4.2.1.2. 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. Australia is a quite easy territory to process for many 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 to use 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.
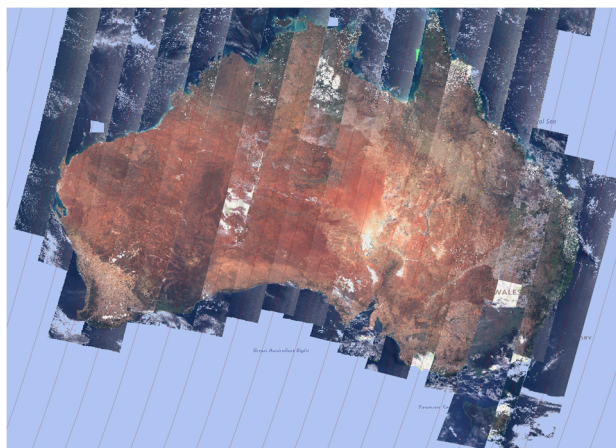

Figure 20. Selection of 33 products over Australia for the GRI

The extraction of the GCPs for this block was quite easy, due to the existence of the AGRI database. At first, "interesting points" were extracted from the orthorectified AGRI database, using SURF methods. 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. Most of the points of interest found by the process are dark or light spots of the landscape.
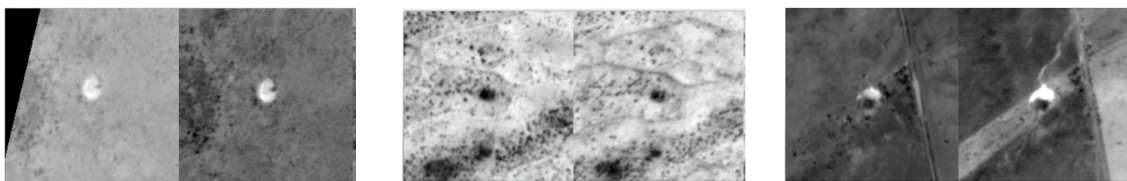

Figure 21. Three couples of homologous points between Sentinel-2 (left) and AGRI (right). Each time, the correlation between the 2 images is made at 10m GSD from ground (orthorectified) images.
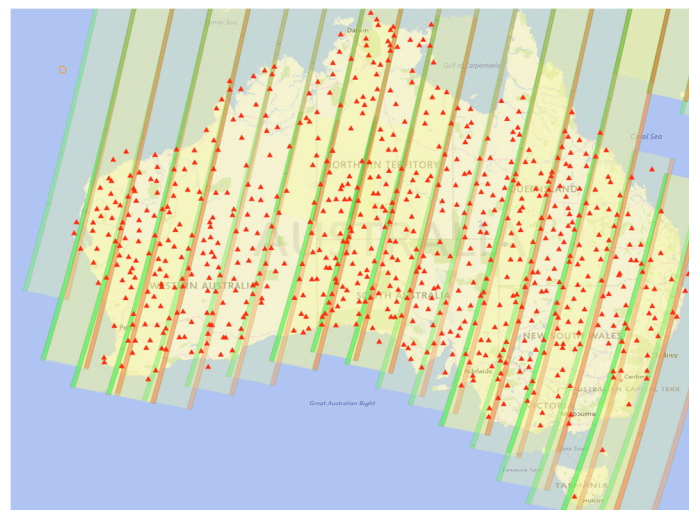

Figure 22. Repartition of the GCPs extracted from AGRI.

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, we have 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 public.

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 retina 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 retina 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.
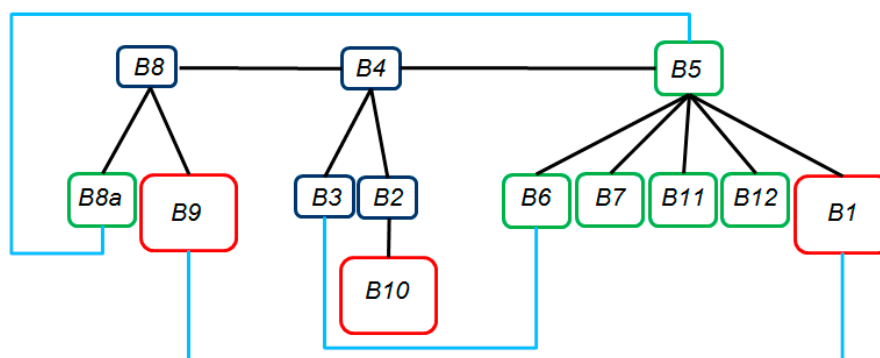

Figure 23. Couples of bands used to perform the multi-spectral registration. Blue boxes: 10m bands; green boxes: 20m bands; red boxes: 60m bands.

4.2.2.2. Results

A calibration has been done in order 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 around the world so cloud free acquisitions can be chosen;
- sites are well-correlated: no need to use reference images, but use of two Sentinel-2 products (no noise in the correlation).

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

Concerning relative calibration, particular case has been observed for 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 seconds between acquisitions by the two bands because the movement of clouds during this timeslot could be up to 1 pixel at 60m. 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.

Generally speaking, the LOS was very well calibrated on ground with an in-flight 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 higher). There are 3 main contributors:

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

The performances are very good and very homogeneous on every couple. The registration residual (positive distance) is better than 0.3 pixels @ 3 , in all cases except on band B10 where it reaches 0.6 pixels @ 3 . This is principally due to the very noisy measurements obtained on this particular band, so this is not the gap between bands but the measured gap with a lot of noise. Moreover as we measure LOS on this band on high mountains there is an effect of the DEM, so we have to use products from different orbits to average, which is not so straight forward.

Viewing directions are very stable. From beginning of September 2015 no evolution has been observed. It is important to notice that if viewing frames biases have a temporal evolution, there will be 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 in order 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 thus 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, and so on…

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 B4 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, in order 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). It was planned that if the pointing accuracy was more than 2km, biases could be given to be put on board, and a new spacecraft model GIPP would then be associated to these new on-board biases. But actually, it has not been necessary.

Just after launch, measured alignment biases were around 2.5 km. After Stars Trackers (SST) calibration on-board, on 3 July 2015, biases decreased at 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. Since 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	-2ms	-81 µrad	953 µrad	0
01/09/2015	-2ms	-77.8 µrad	949.4 µrad	22 µrad
15/11/2015	-2ms	-77.8 µrad	946 µrad	47 µrad
01/02/2016	-2ms	-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-frequency from weekly basic checks to yearly in-depth analyses, including long-term statistical and trend analysis. In case of spotted performance degradation, specific calibration activities could 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 sections 4.1.1 to 4.1.3) and to the linearity model characterization. In case of a degraded performance observed a dedicated calibration activity and possible investigations can be triggered. Figure 24 illustrates the effect of equalisation on the image quality.
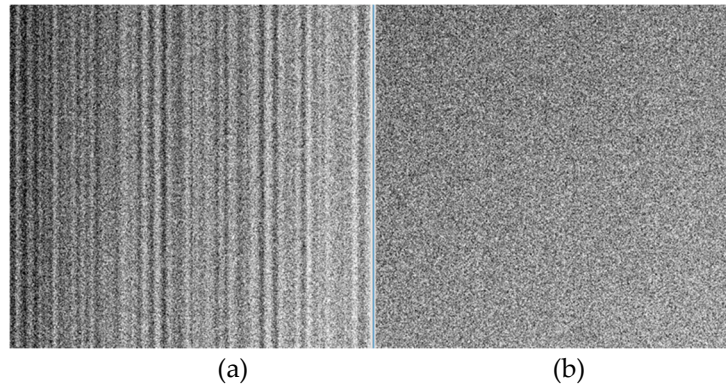
(a)   (b)

Figure 24. Illustration of image equalisation impact: (a) Not-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 only for cross-validation of the parameters provided by the diffuser in order to potentially update them 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, see Figure 25, is applicable to both sun-diffuser image and to uniform scenes, the computations 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.
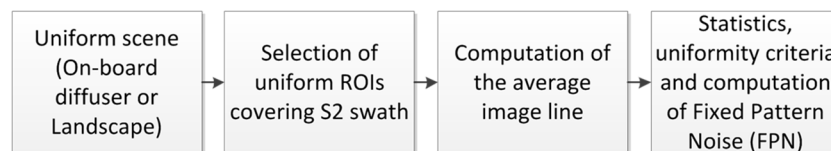

Figure 25. General principle for equalisation assessment.

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 (in order 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):

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

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.
   - $MEN(b) = MAX\big(STD(R)\big)$ 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. 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 allows meeting the specification again.

Table 7: Statistics on the FPN (in %) estimated on the sun-diffuser acquisition on 4 July 2016 by applying the operational coefficients EQOG calculated from the SUN acquisition on 8 June 2016. Values are given for the minimal, average and maximal values, and for the quantile at 98 %. The specified maximum acceptable values are reminded.

Band	Spec	Min	Avg	Q. 98%	Max
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

4.3.1.3. Results on Uniform Scenes

We present hereafter 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 be white and intrinsically uniform. However, we are looking at very small effects, so the zone selection has to be done very 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.
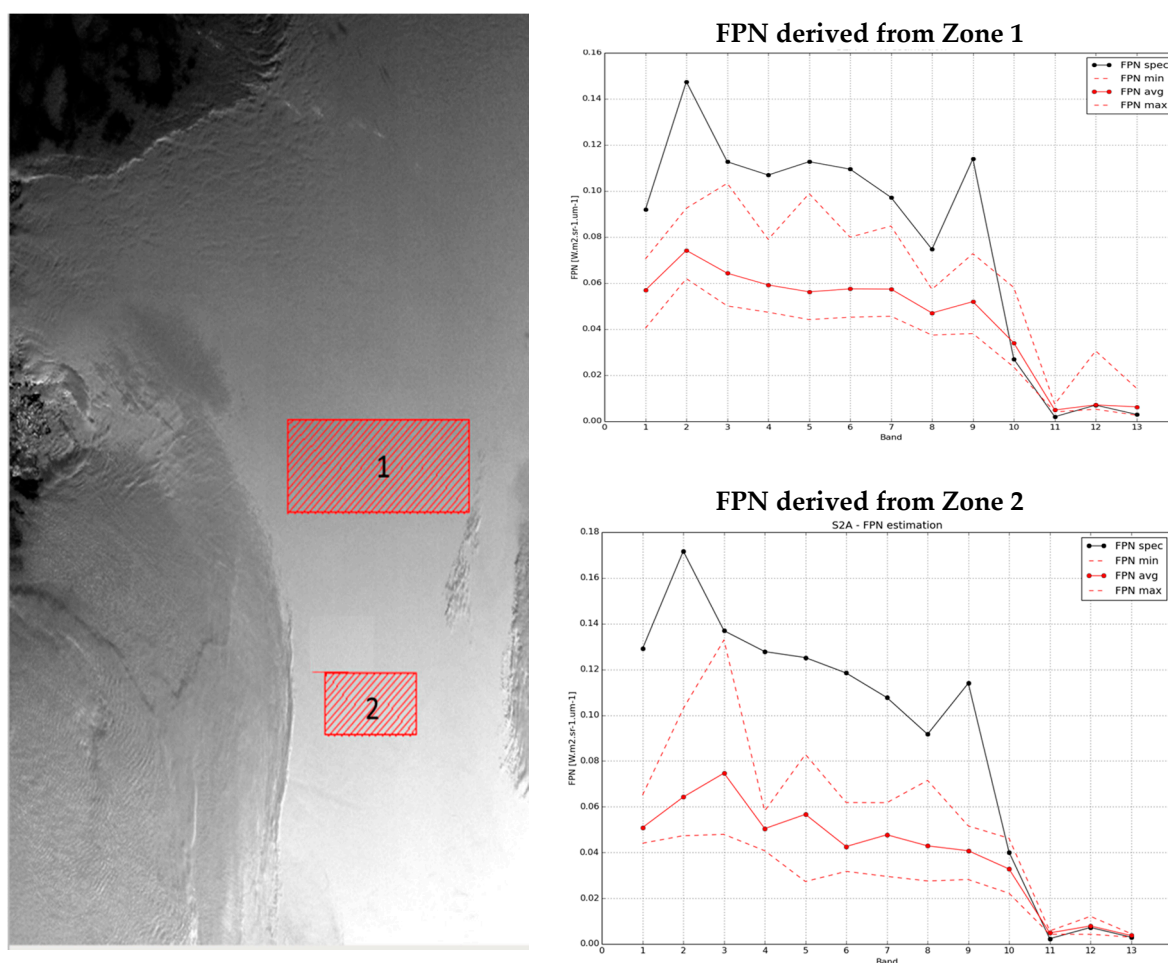

Figure 26. Left hand side: Sentinel-2 Greenland image date from 2015/09/04 together with the 2 uniform zones that are used to estimate the equalization quality. Right hand side: the FPN that are derived independently from the 2 zones.

In the VNIR range, the FPN is almost systematically lower than the specification. 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 mean PFN is close to the specification while the maximum FPN is always higher. 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 will rely on the on board solar diffuser. This calibration has been validated using vicarious methods. As sensor life will increase, the vicarious calibration will aim 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 k, the ratio between the TOA observed reflectance and the computed equivalent TOA reflectance gives the absolute calibration coefficient A_k. Except for band B10, the vicarious calibration will rely on three methods, the Rayleigh scattering method, the Deep Convective Clouds (DCC) method and the ground-based reflectance method.

4.3.2.1. Methods

4.3.2.1.1. 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 [11]. In ideal conditions -stable oceanic region, with low concentration of phytoplankton and sediment, and far from land to ensure a purely maritime aerosol model [12] 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 [13]. 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 [14; 15] 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.

In order to ensure a proper computation of the vicarious coefficients over Rayleigh scattering, one needs to satisfy the following conditions: 1) A 0 % cloud coverage at ROI; 2) A low wind speed is required; typically <5 m/s; 3) A small content of aerosol must be ensured, assuming the Rayleigh corrected normalized radiance at 865 nm directly related to aerosol amount [14]; 4) A very stringent threshold at 865 nm of 0.002 also avoids using further data screening for sun glint

We use a marine model following [16] which 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 [17].

In order to ensure a 0 % cloud coverage over the ROI, the target dataset automatically cloud screened using DIMITRI (Database for Imaging Multi-spectral Instruments and Tools for Radiometric Inter-comparison) package [18]. In order to ensure a low concentration of phytoplankton and sediment, and pure marine aerosol, six Cal/Val sites were selected following [19] (Figure 27 and Table 8).

In order to discard any contaminated pixel by sun glint, we use a threshold of 0.002-0.008 on the Rayleigh corrected normalized radiance at 865 nm as well we apply a threshold of ±15° around the specular direction..

After the selection of L1C-tiles that are within the area of interest (AOI), the radiometric and geometric measurements of the 13 bands are re-sampled into B01-grid (60 m spatial resolution). This option has been justified by the big size of S2A-products with full resolution (10 m). In order to minimise the cloud contamination and the processed products size, sub-ROIs surface of 0.1 ° x 0.1 ° up to 0.3 ° x 0.3 ° latitudes 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.

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	-10.00	+10.00	60.00	90.00
Railroad Valley	38.49	38.50	-115.68	-115.69

4.3.2.1.2. Inter-band calibration over Deep Convective Clouds

This method is usually used for lower resolution sensors (at least hectometric) [20], 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 an absolute calibration coefficient with respect to another band. It is useful in order 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 B4 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 x 700 km²) has been chosen over the Maldives during the IOC (Figure 27 and Table 8).
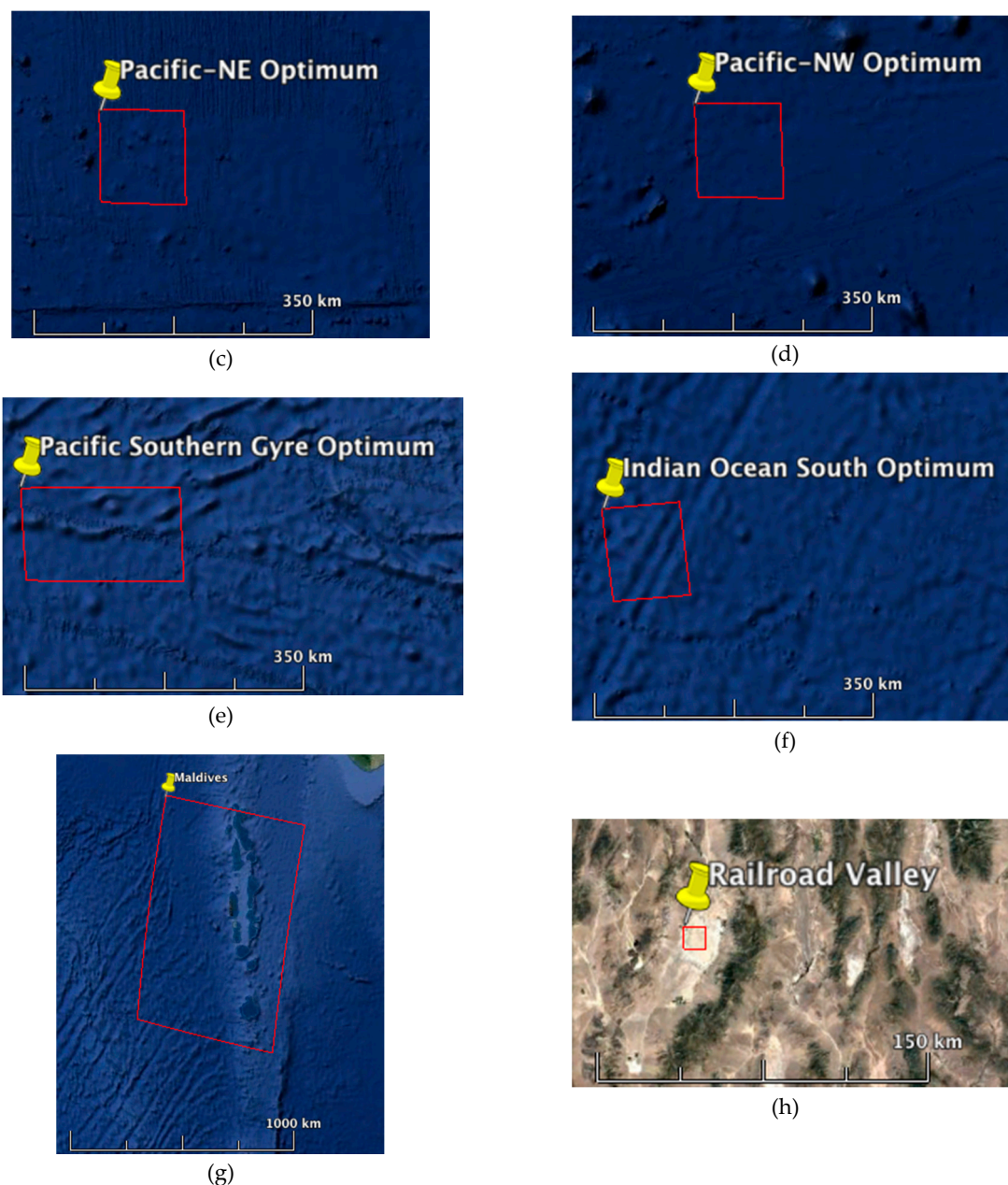

(a)

(b)

Figure 27. Cal/Val test sites over (a-f) Rayleigh scattering, (g) DCC, and (h) Ground-based chosen for Sentinel-2A commissioning activities.

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

$$L_{DCC}(\lambda, \theta_v, \theta_s, \Delta\phi) = \left(L_{cloud}(\tau_{cloud}, \tau_{aer}, \tau_{mol}) + L_{Ray} + L_{aer}\right) \times T_g(\lambda, \theta_s, \theta_v) \qquad (10)$$

Where:

L_{cloud} is the radiance at top of the cloud, which depends on the geometrical conditions, the cloud particle type and at a very little extent of 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 all contributors are 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 TOA radiance over DCC (average case).

Band	Cloud	Molecular signal	aerosols	gaseous transmission
		Contribution to the signal, except absorption		
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%

4.3.2.1.3. Ground-based reflectance measurements

The ground-reflectance-based approach is used over the Radiometric Calibration Test Site (RadCaTS; Figure 27) to perform the absolute radiometric vicarious calibration since two decades and recently as operationally for Landsat-8 [21; 22; 23]. In this study we use the in-situ 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-site (RRV) but only one overpass coincides with the S2A time-series from the baseline 02.01 used here. We use as reference the TOA normalized reflectance reconstructed by the University of Arizona team using ground and atmospheric radiometric measurements. Note that, the dataset has been specifically generated for the spectral bands of S2A using hyperspectral reference profiles of the site.

4.3.2.2. Results and analysis

4.3.2.2.1. Rayleigh scattering results

The assessment of a given sensor calibration consists on comparing the observed TOA reflectance (*Rho_obs*) by the satellite sensor to the simulated TOA reflectance (*Rho_sim*) by the radiative transfer model. The resulting ratio Ak, defined as $\varrho_{obs}/\varrho_{sim}$ provides a measure of the quality of the calibration of the instrument.

Rayleigh method in DIMITRI is performed over the valid S2A-L1C products (11 acquisitions) over the commissioning period June 23rd 2015-April 1st 2016. The results seem sufficiently reliable to provide a first estimate of the absolute vicarious calibration coefficients in the visible spectral range (Table 10). The results are consistent overall 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 [24] obtained over shorter period, which attests the temporal stability of the sensor measurement. The estimated uncertainty associated to each site is found less than ±5% and it drops up to less than ±2% overall the test-sites. This attests the accuracy of the method in compliance with the mission requirement.
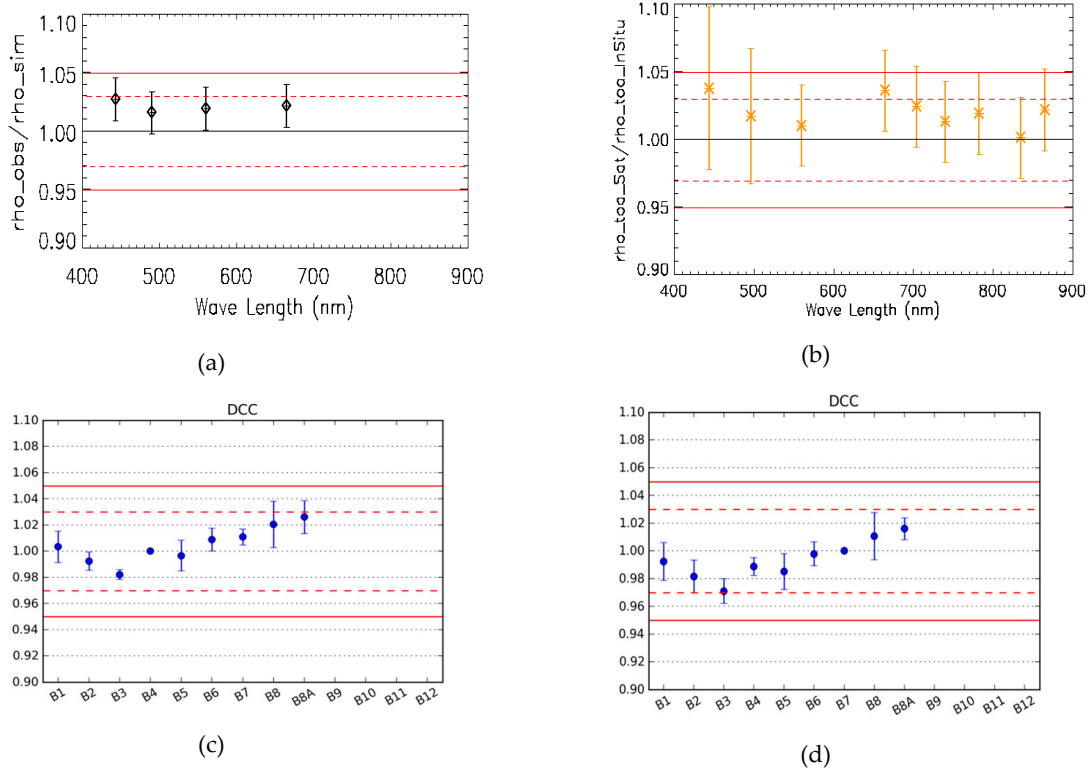
(a)

(b)

(c)

(d)

Figure 28. Calibration results over (a) Rayleigh Scattering, (b) Ground-based measurements, (c) and (d) Deep Convective Clouds using Sentinel-2 images calibrated with the PDGS-absolute calibration coefficients for (a and b) and the sun-diffuser for (c and d). The dashed (resp. solid) red lines represent the 3% goal specification (resp. 5% threshold specification). The reference band is B4 and B7 for (c) and (d) respectively. Error bars indicate the estimated uncertainty for Rayleigh and In-situ measurements and the Stddev for the DCC method.

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

S2A/MSI Bands	Wave length (nm)	Rayleigh		DCC		In-Situ	
		Vic. Cal. Coeff.	Uncert. (%)	Vic. Cal. Coeff.	Stddev. (%)	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

4.3.2.2.2. Deep convective clouds method results

This method was used for the first time for a high-resolution satellite with Sentinel-2A. A lot of experimental work has been achieved to transpose the method from low to high resolution, 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-2, 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, with so few available products, the results presented in and Table 10 show spectacular consistency, with all bands from VNIR range within 3 % of each other, which is compliant with the goal specification. To make sure there was no artefact of the method, a different reference band has been chosen (B4 or B7): the results show the same interband pattern so they are consistent.

Note that B3 may be off the other bands because this band 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.

B8 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 it is not surprising they present relative different behaviours.

The method is also less limited by geometric conditions than the calibration method over molecular scattering, so it allows deriving calibration results within the whole field of view of the instrument. The distribution of valid measurements presented in Figure 29 show an increased number of points in the centre of the FOV, but is only due to the cloud distribution within the images. More acquisitions would even the number of measurements per detector and confirm the behaviour of the calibration coefficients within the FOV. Figure 30 show a slight relative decrease on the East side of the images for B1 (~1-2%) and B2 (<1%). It this tendency were confirmed, it would help to improve the consistency of the Rayleigh method results for these bands with other methods, since the distribution of valid measurements over the oceans is biased on the West side of the FOV. This method shows great results: the 3% goal inter-band specification is confirmed with only a few products.
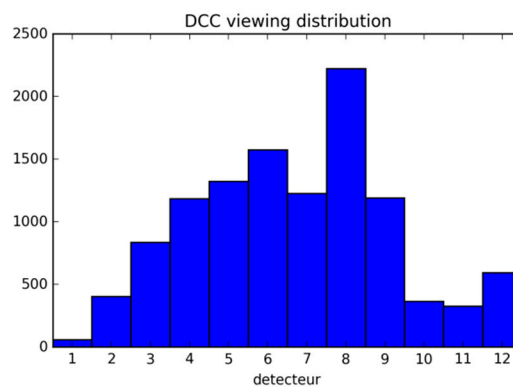

Figure 29. Distribution of valid measurements over DCC sites as a function of detector number.
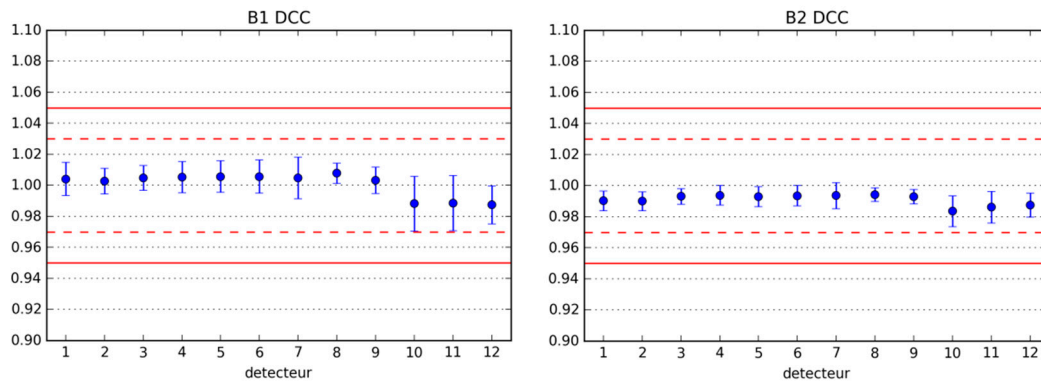
Figure 30. B1 (left) and B2 (right) inter-band calibration coefficients with respect to B4 as a reference band as a function of the detector number, based on Sentinel-2 images calibrated using the diffuser.

4.3.2.2.3. Ground-based reflectance measurements results

In this analysis we used as reference TOA normalized reflectance reconstructed by the University of Arizona team using ground and atmospheric 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_k) as 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 UoA team is of the order of 3 % for most bands, and up to 6 % for B1 and B2 due to a higher impact of atmospheric effects [22; 23]. 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. We aim to identify relative biases in the radiometric calibration and, and thereby, to correct for biases if/where necessary. Moreover, the long-term trends in S2A/MSI sensor measurements are assessed in order to monitor the sun diffuser aging along the mission life. We use pseudo-invariant calibration sites (PICS) methodology following [26], and implemented in DIMITRI (dimitri.argans.co.uk) by ARGANS [24]. PICS method allows observation of the biases and trends in any given sensor and provides a reliable assessment of its radiometric performance.

4.3.3.1. Pseudo-Invariant Calibration Site Methodology

The pseudo-invariant calibration sites (PICS) as deserts have been exploited since decades for multi-temporal monitoring and cross-calibration in the solar spectral range. PICS algorithm aims to simulate the TOA reflectance in the visible to near-infrared (NIR) spectral range over pre-defined desert (e.g. CEOS PICS sites [25]). 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 [26]). 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 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 included. Then it has been validated

over the whole archive of MERIS 3rd reprocessing (2002-2012) [26]. The method allows performing multi-temporal analysis, as well comparison of multiple sensors on the same site over the same or different observation periods.

4.3.3.2. Sites Selection

[27] 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 x 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). DIMITRI uses the six CEOS-PICS sites (Table 11) for the PICS calibration. It is worth to note that due to the high resolution of MSI L1C products, and for practical reasons, we subsampled these sites to reduce the site-size into 20 km x 20 km and 30 km x 30 km. To keep the properties of the sites, the small ones are selected as a subset of the standard one 28. Figure 31 displays RGBs quick-looks of the six desert test-sites used in DIMITRI for this study.

Note that the surface BRDF is computed considering the full size of the test-site. Thus the impact of the test-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.

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

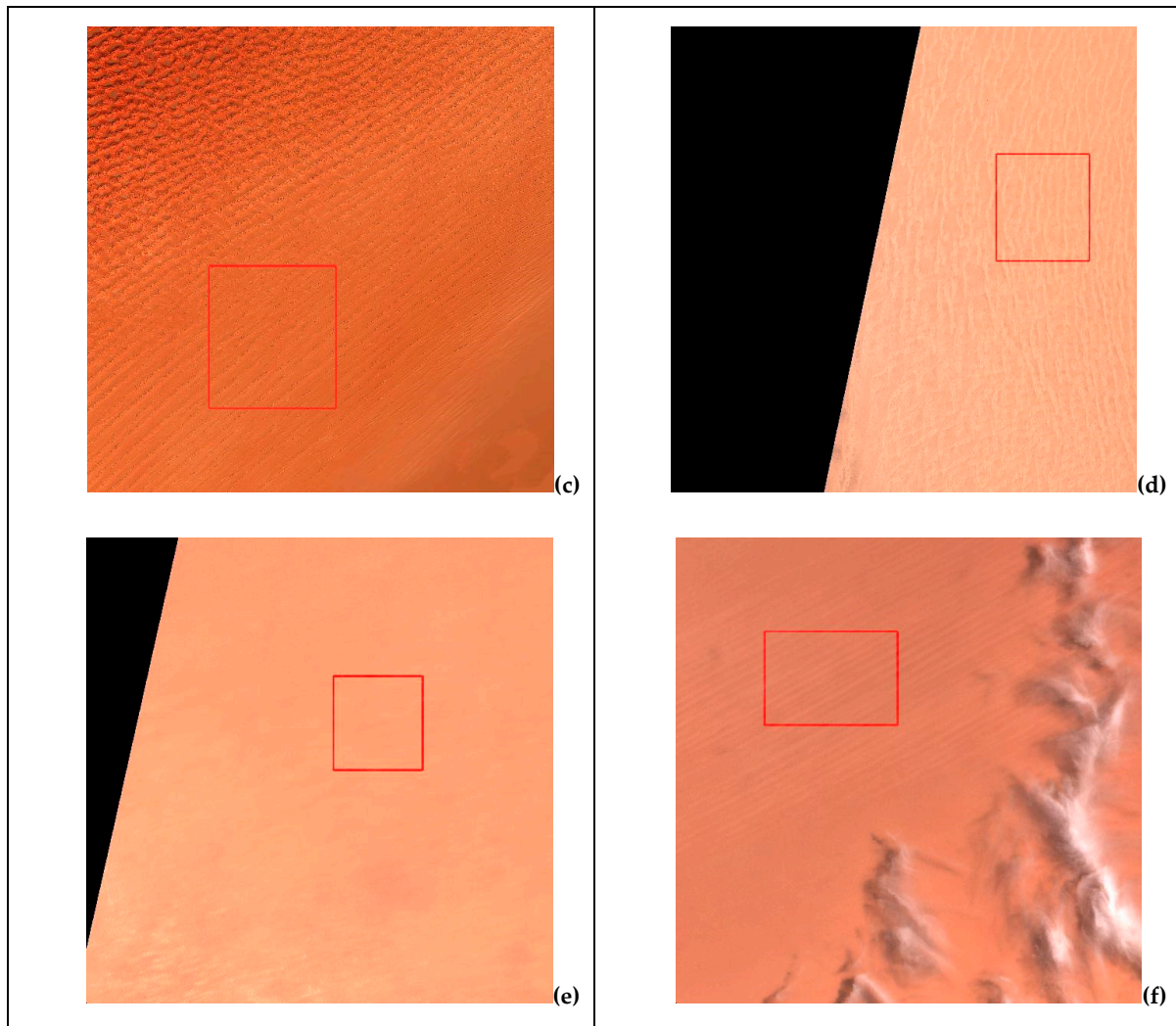| N°	Name	Latitude (°)		Longitude (°)	
		min	max	min	max
003b	Algeria 3	29.82	30.82	7.16	8.16
005b	Algeria 5	30.52	31.52	1.73	2.73
031b	Libya 1	23.92	24.92	12.85	13.85
034b	Libya 4	28.05	29.05	22.89	23.89
038b	Mauritania 1	-9.8	-8.8	18.8	19.9
039b	Mauritania 2	-9.28	-8.28	20.35	21.35
031a	Railroad Valley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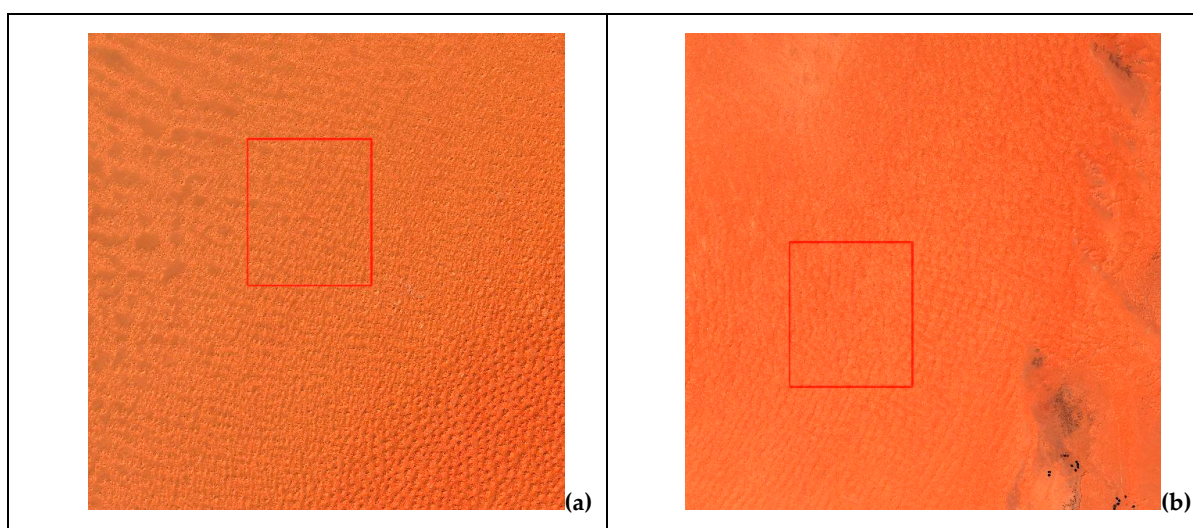
Figure 31. RGB Quick-Looks from S2A/MSI over (a) Algeria3, (b) Algeria5, (c) Libya1, (d) Libya4, (e) Mauritania1, and (f) Mauritania2 sites. Red squares indicate the ROI.

4.3.3.3. Results and Analysis

As mentioned above, this method considers the VNIR spectral range only. Moreover, the gain coefficients are, in fact, relative calibration coefficients as we have considered MERIS 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_k, 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 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 waves (e.g. B01) show more scattered ratios than the long waves ones. This can be easily seen on the stddev on the bottom-left of each plot. This effect has been observed by [28] 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 clearly be seen over these time-series where we find trend values within/less-than 1% per year within the mission requirements. While the uncertainty of the method is estimated to be of about 5% (except the absorption bands), the standard deviations over each site (Figure 32) represents less than 2% 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 [24], a clear improvement on the trends values due to the longer time-series of S2A/MSI acquisition. This confirms the statistical nature of the used PICS methodology, thus the longer the time-series the

better the estimation of trends, and the more acquisition we have the more reliable the estimation we get.
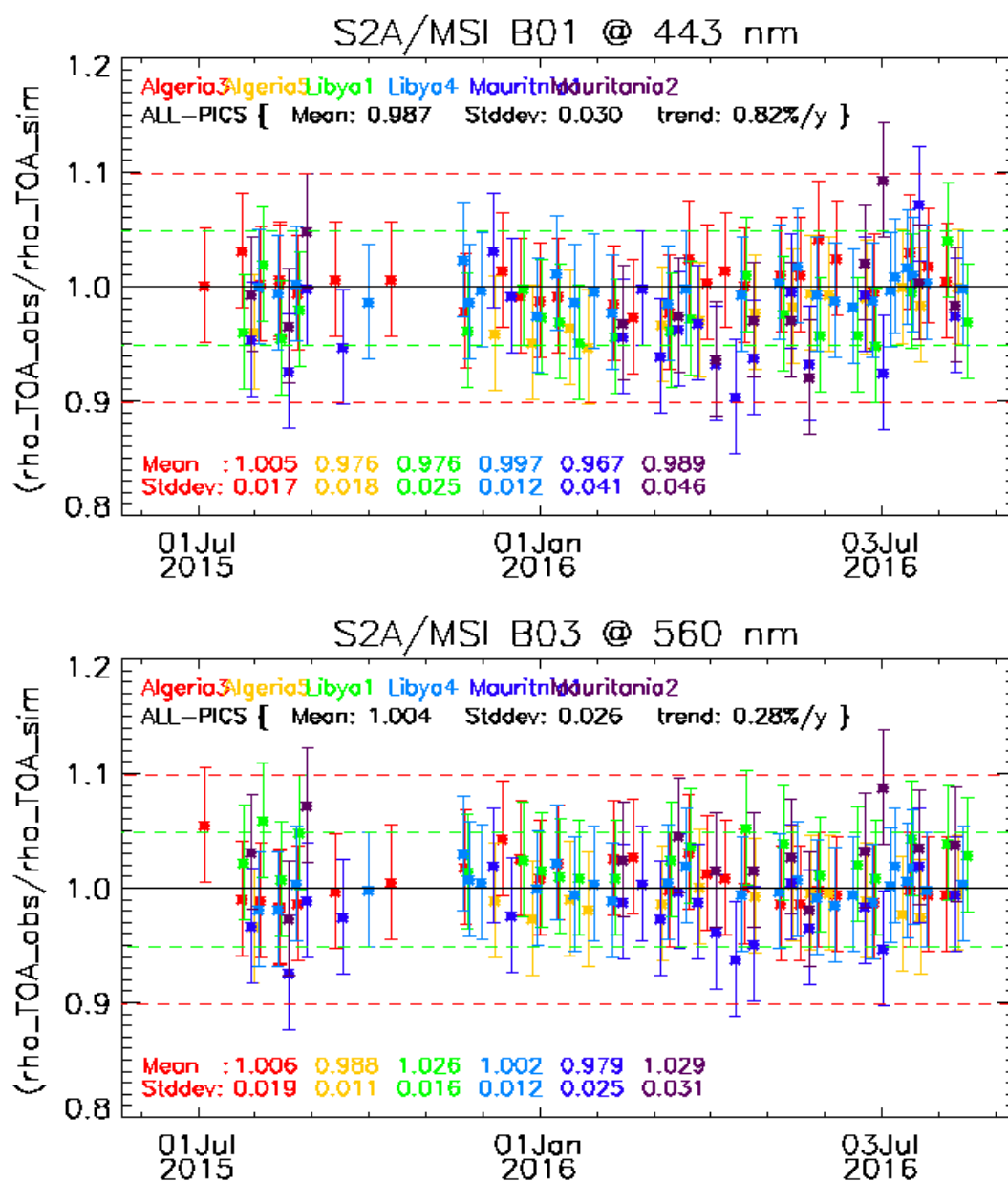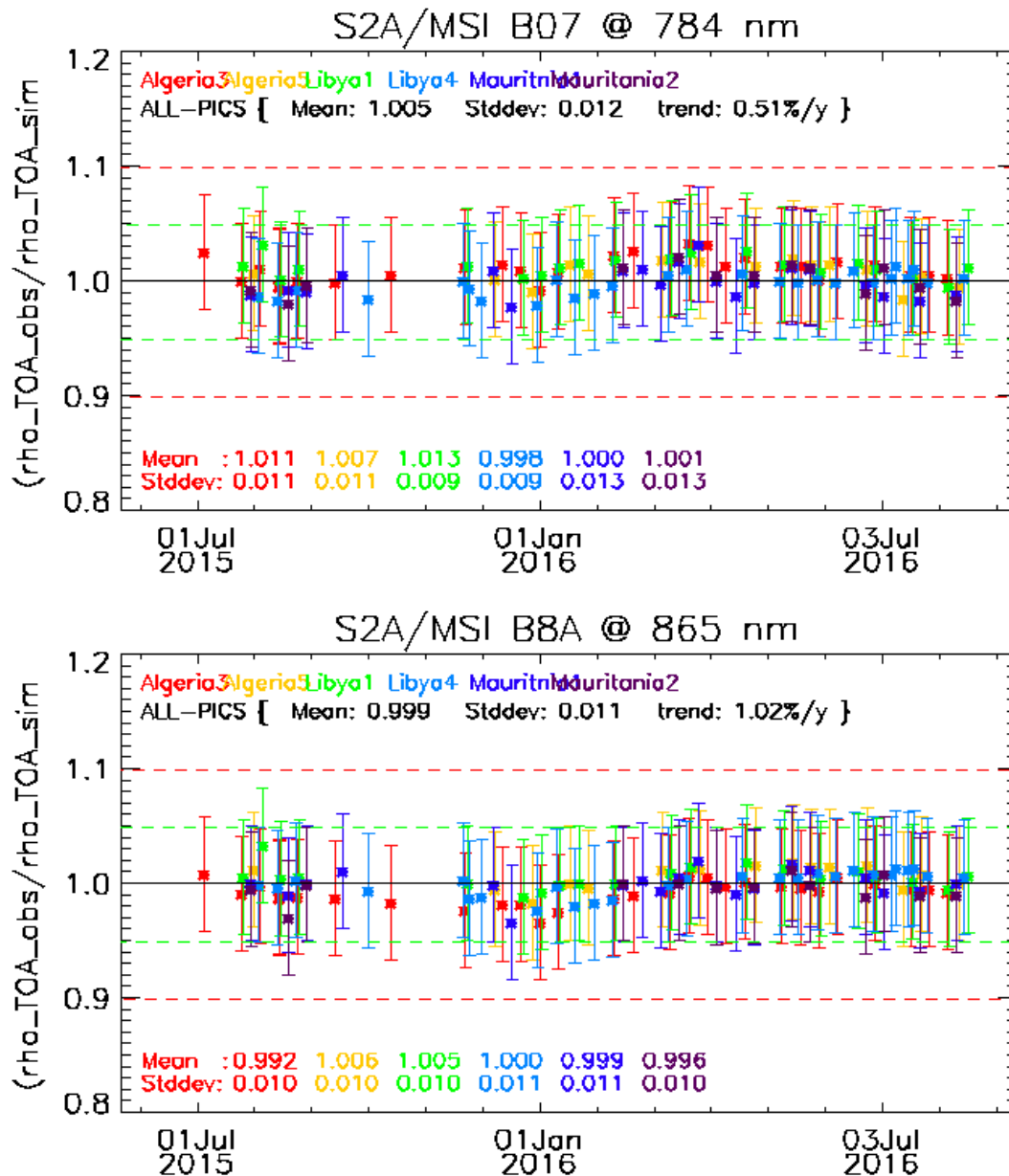
Figure 32. Time series of TOA reflectance ratio R_k from S2A/MSI over (red) Algeria3, (yellow) Algeria5, (green) Libya1, (light-blue) Lybia4, (dark-blue) Mauritania1 and (purple) Mauritania2 for (from top to bottom) B01, B03, B07 and B8A over the commissioning period. Error bars indicate the estimated uncertainty for the PICS method.

The temporal average of the ratios for each band over all the four test-sites is summarized on 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 B05, which shows high ratio (>5% error and >11% uncertainty), this is due to the methodology limitation to simulate accurately the TOA-reflectance for bands with significant O2 and water vapour absorption. However, the results illustrate the good quality of S2A/MSI calibration. The results compare well with those obtained by CNES over the same PICS and the same commissioning period, but using independent desert-calibration methodology [29].
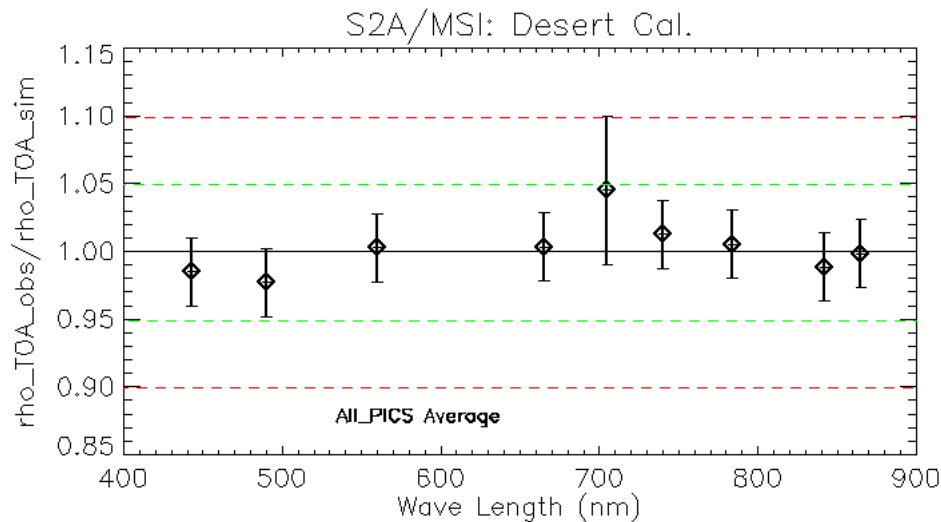
Figure 33. Temporal average of the ratio of observed TOA-reflectance to simulated one from S2A/MSI over the four-PICS as a function of wavelength. Error bars indicate the estimated uncertainty for the PICS method.

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, LANDSAT-8/OLI in our case. Our 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. We use the so-called Sensor-to-sensor inter-calibration methodology implemented in DIMITRI by ARGANS Ltd. In addition we use the PICS methodology 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) over the similar bands over the commissioning period.

4.3.4.1. Methods

4.3.4.1.1. 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 [30]. The day offset and strictness of angular matching between observations, as well as the allowable cloud cover percentage values can be user defined. The strictness of the angular match between two concomitant sensors observation, 1 and 2, is controlled through the Angular Matching Criteria (AMC) parameter which is defined as:

$$AMC = \sqrt{\left([\theta_{s1} - \theta_{s2}]^2 + [\theta_{v1} - \theta_{v2}]^2 + \frac{1}{4}[|RAA_1| - |RAA_2|]^2\right)} \qquad (11)$$

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 that one of the reference sensor is computed. The comparison is done over three desert-sites, Algeria3, Libya4 and Railroad Valley (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 and azimuth and zenith (SAA, SZA), the viewing azimuth and zenith angles (VAA, VZA). These parameters are averaged over all pixels included in the test-site. In order to maximize the number of the match-ups (doublets), we select all acquisitions that satisfy the condition AMC of <15, <45 and <55 for Algeria3, Libya4 and Railroad Valley sites respectively and time-lag (day-offset) between both sensors of 11 days. The spectral response of the seven reference bands used for the Intercomparison is depicted on Figure 34. Except bands B02 and B04, which display a shift of about 10 nm, all the considered bands show a rather comparable spectral response, 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.
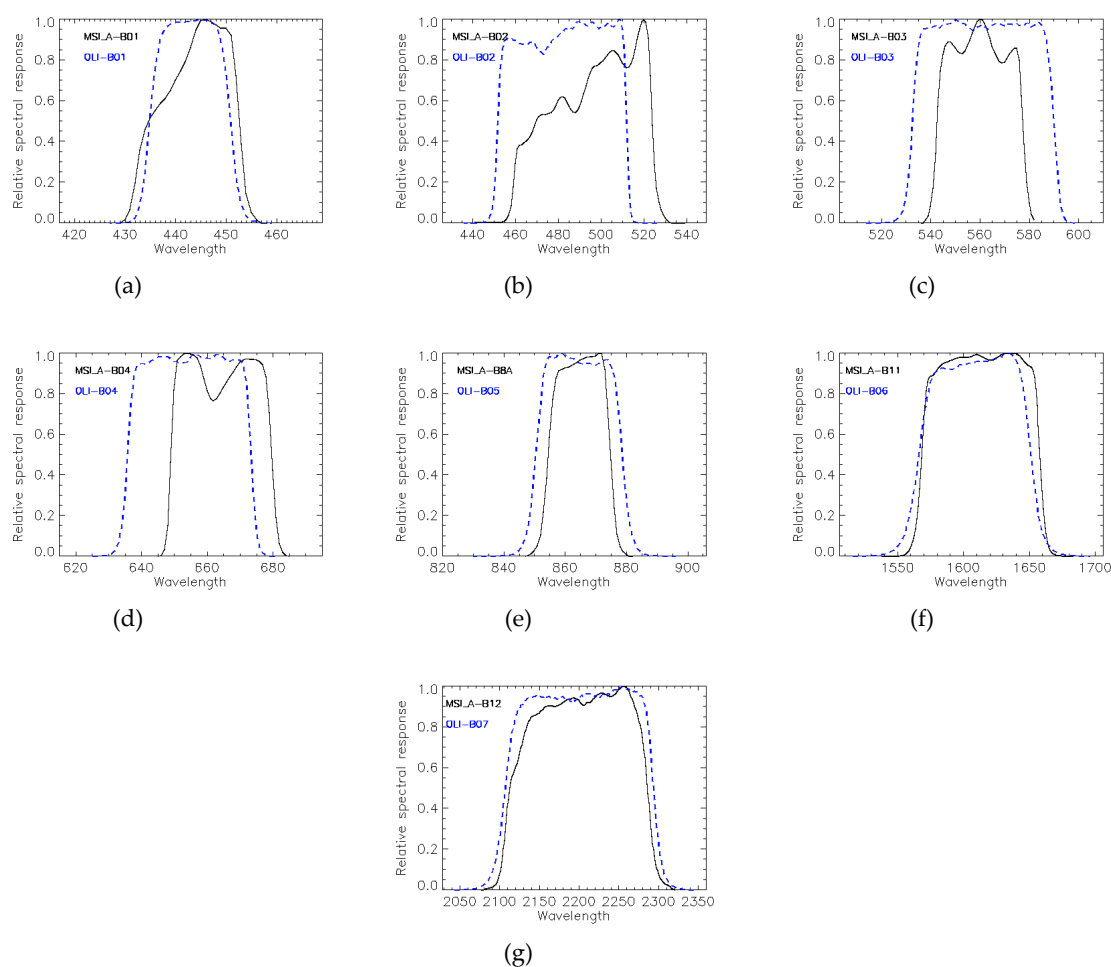

Figure 34. Relative spectral response of (solid-black) S2A/MSI and (dashed-blue) LANDSAT-8/OLI for bands (a-g) 443, 490, 560, 665, 865, 1610 and 2190 nm respectively.

4.3.4.1.2. Pseudo-Invariant Calibration Site Methodology

In addition to the Sensor-to-sensor inter-calibration method, we use PICS method described above (section 4.3.3.1). 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 into the visible to near-infrared (NIR) 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.

4.3.4.2. Results and Analysis

4.3.4.2.1. Sensor-to-Sensor Inter-Calibration Results

By applying the previously described geometrical and temporal criteria to identify doublets of remote sensing reflectance between S2A/MSI and LANDSAT8/OLI we obtain 12, 4 and 5 doublets over 40, 36 and 15 acquisitions for each site Algeria3, Libya4 and RRV respectively. Figure 35 displays the ratios between the observed TOA reflectance from the sensor under calibration (*Rho_cal*) to the TOA reflectance from the reference sensor (*Rho_ref*). In general, the comparison shows remarkable consistency over all the test sites over all the spectral range. The maximum discrepancy over the average gain coefficients is observed over Bands B02 and B04. This discrepancy could be related to the shift in the RSR as shown in Figure 4.3.3.1. Libya4 results show higher bias overall the spectral range, which could be explained by the low 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 [29] when comparing similar bands from MODIS-A and MERIS over about 20 PICS sites. The estimated uncertainty overall the considered doublets is about 3% for bands B01 and B02, while we find uncertainties less than 2.5% for the other bands (B03, B04, B8A, B11 and B12). Our uncertainty estimation is in good agreement with the estimated uncertainty of about 3% by [30] for MODISA, AATSR and MERIS Intercomparison. In fact the uncertainty difference between the VIS band s and SWIR bands is most likely related to the decrease of the surface reflectance and the increase of the atmospheric reflectance in the VIS domain, reaching 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 (REF-X, REF-Y). It is worth to note that there is several factors contribute to the discrepancy between both sensors: 1) the day-offset of 11 days which induce 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, which induces difference on TOA reflectance of about 1% [30].
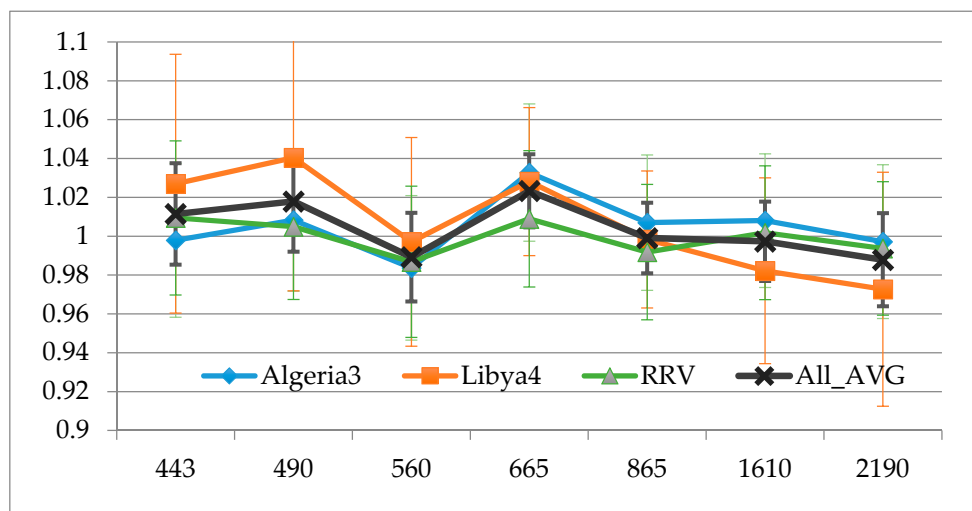

Figure 35. TOA reflectance ratio defined as S2A-MSI/LS8-OLI measurements derived from extracted doublets over (blue) Algeria3, (orange) Libya4, (green) RRV and (black) the average over the three test-sites as a function of the wavelength. Error bars indicate the estimated uncertainty.

4.3.4.2.2. 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 only the matchups and it considers the VNIR spectral range only. Thus the gain coefficients are, in fact, relative calibration coefficients as we have considered MERIS 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_k, defined as observed/simulated (ρ_{obs}/ρ_{sim}) reflectance is then computed over the six desert-sites. Figure 36 shows time series of R_k 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 result Rk. In particular, the concomitant acquisitions show similar response of both sensors. Nevertheless, OLI ratios exhibit slightly higher variability in the blue and red bands than MSI (about 0.2 % difference on the standard deviation), the NIR band of OLI show lower variability than MSI bands of about 0.06 % on the standard deviation (1σ). The green band of OLI shows slightly higher ratios than that MSI ones. While the bands L05/B8A show a very good agreement in term of temporal evolution and radiometric gain, which is within 2-3%. Overall the similar bands from both sensors show slight difference in the standard deviation less than 0.3%. In spite of the difference in the acquisition time and the difference of the spectral response of S2A/MSI to LANDSAT/OLI, a striking agreement between both sensors in term of radiometric measurements and image qualities.
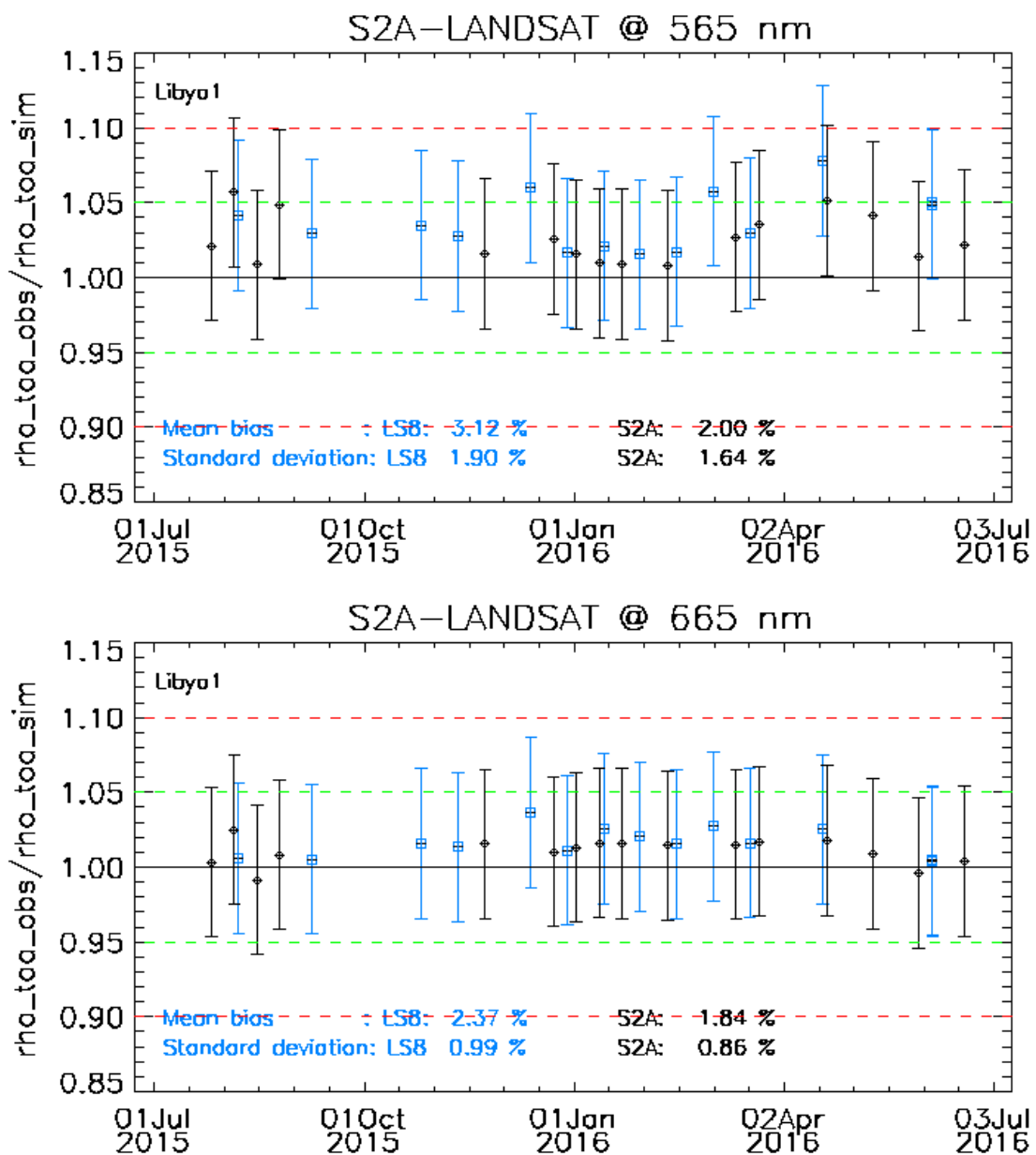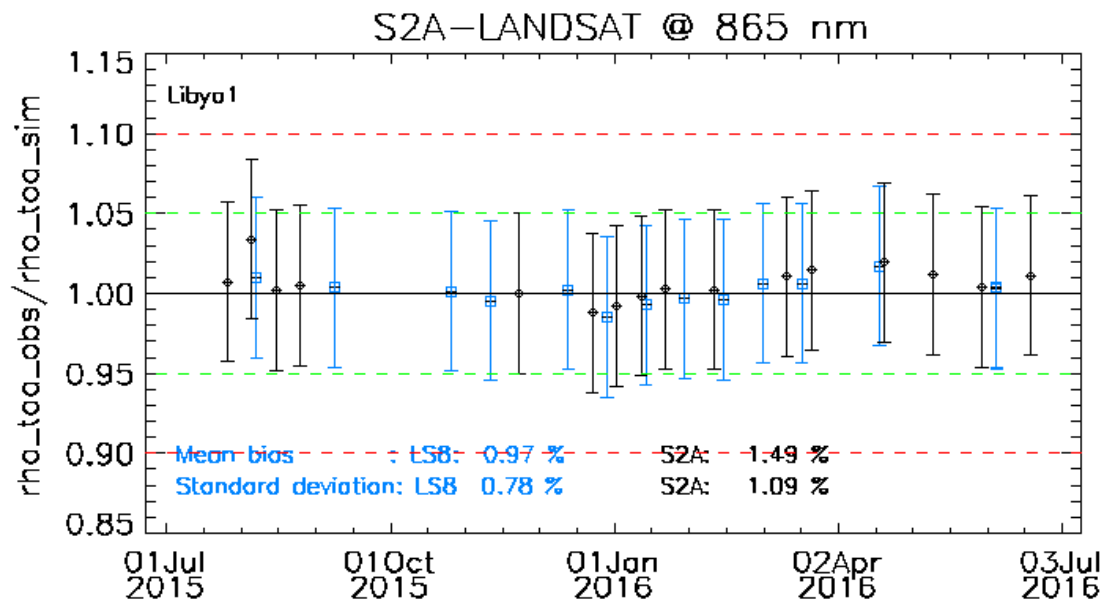
Figure 36. Time series of TOA reflectance ratio R_k from (black) S2A/MSI and (blue) LS8/OLI over Lybia1 for (from top to bottom) B03, B04 and B8A from S2A over the commissioning period. Error bars indicate the estimated uncertainty for the PICS method.

Figure 37 illustrates the cross-calibration results over Algeria3, Libya1 and Libya4 desert sites. Except slight bias of 1-2 % in bands B01 over Algeria3 and B02 aver Libya4, the results are consistent over the VNIR spectral range, and no significant discrepancy can be observed. The six PICS show almost the same results (not shown), where the whole ratio values are within ±5 %.
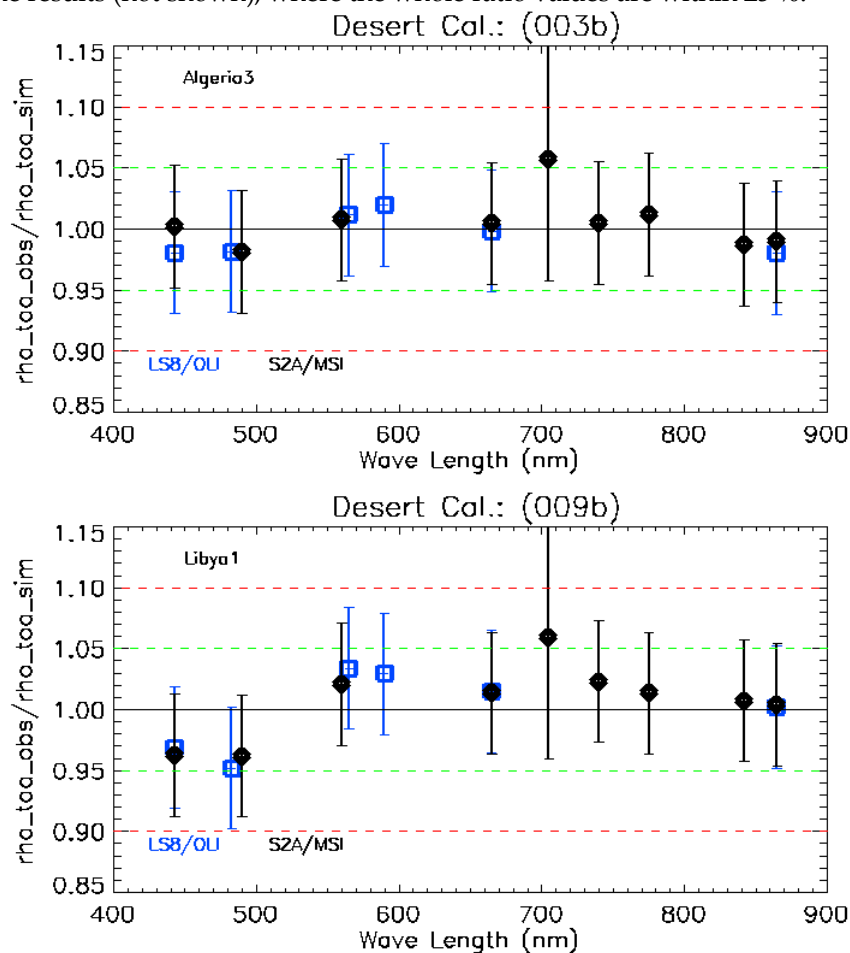
Figure 37. Ratio of observed TOA-reflectance to simulated one for each sensor (black) S2A/MSI and (blue) LANDSAT-8/OLI over Libya1 site as a function of wavelength. Error bars indicate the estimated uncertainty for the PICS method.
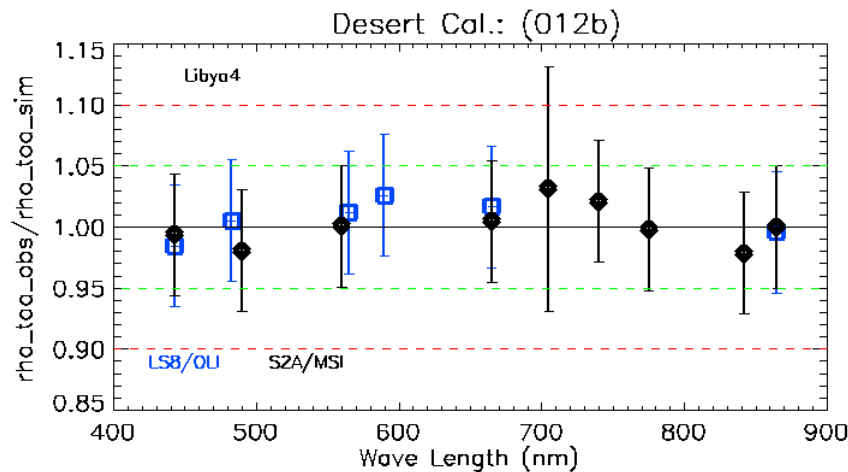
4.3.5. Inter-band Relative Radiometric Uncertainty Validation

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) 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 will 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 spectral bands k and l 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 envisaged over snowy areas like Antarctica and Greenland. As the snow spectrum varies according to the kind of snow (thin or thick), the inter-band validation nominally relies on deserts sites. It is performed twice a year.

4.3.5.1. Method

It relies on spectral BRDF of desert sand samples measured in 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 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, a priori climatologic data and the spectral BRDF available, feed a radiative transfer code which provides expected TOA reflectances for each spectral band.

The reflectance ratios of pairs of spectral bands are computed for the both cases and compared.

4.3.5.2. Results

Due to problems in the tool extracting reflectances from L1C products, there are no consolidated results for the moment. Nevertheless, preliminary results showing expected and computed spectral reflectance for one Algeria-4 acquisition are presented on Figure 38.
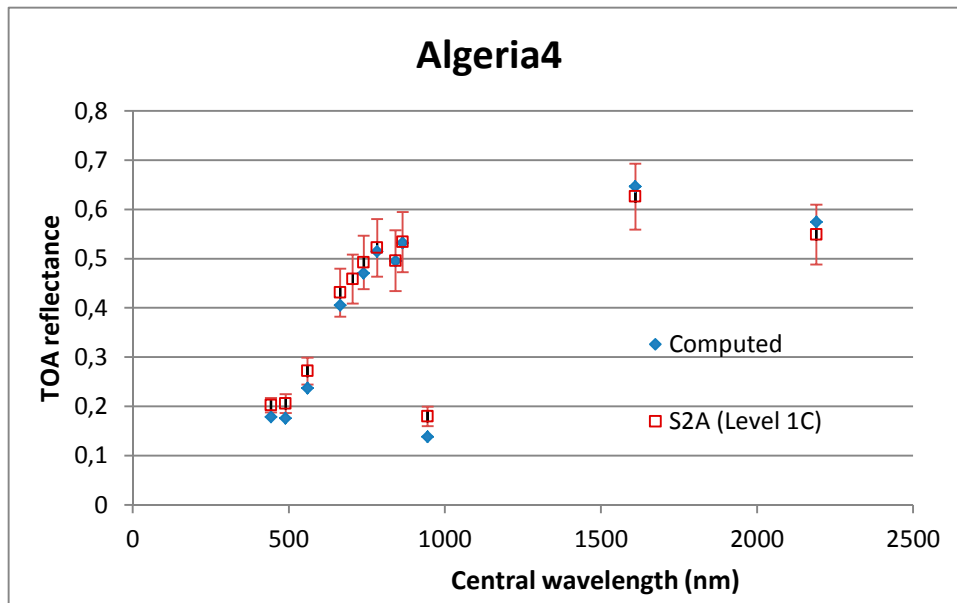
Figure 38. Extracted and computed reflectances for one acquisition over Algeria4

4.3.6. SNR 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:

$$SNR = \frac{<L>}{\sigma_L} \tag{12}$$

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:

$$\sigma_L(p,d,b) = \sqrt{\alpha_L(p,d,b)^2 + \beta_L(p,d,b)L(p,d,b)} \tag{13}$$

Where

- p,d,b are respectively indices for pixel, detector, band
- σ_L is the radiometric noise standard-deviation (in W/m²/sr/μm)
- L is the radiance (in W/m²/sr/μm)
- α_L and β_L are the noise model parameters, both in in W/m²/sr/μm.
    - α_L represents the dark noise standard-deviation (at radiance 0)
    - β_L.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 (13) 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 (14) below:

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

The method for assessing the 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 01/2016) as shown in Figure 39 and Figure 40.

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 39. SNR requirements at reference radiance L_{ref} and current measured values
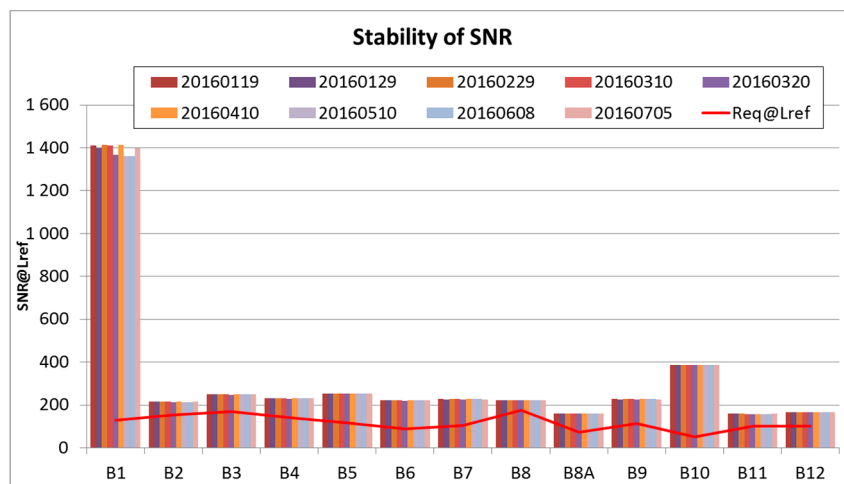

Figure 40. Average SNR at reference radiance L_{ref} measurements (per band) since January 2016.

4.3.7. MTF Validation

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

Let us recall the sensor model for spatial resolution:

$$Image(x,y) = Landscape(x,y)*h(x,y).Dirac\ Comb(x,y) \tag{15}$$

x varies along the line of the Image,
y varies along a columns of the Image,
h is the impulse response in other words the Point Spread Function (PSF),
The *Dirac_Comb* corresponds to the sampling by the detectors in the focal plane.
A Fourier transform enables to reach the frequency domain:

$$ImageSpectrum(f_x, f_y) = LandscapeSpectrum(f_x, f_y).H(f_x, f_y)*\ Dirac\ Comb(f_x, f_y) \tag{16}$$

fx is the spatial frequency for the line (or across track) direction,
fy is the spatial frequency for the column (or along track) direction,
H is the Transfer Function.
The MTF is the modulus of the transfer function H.
The convolution by the *Dirac_Comb* 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.

4.3.7.1.1. Reference image method

Unlike the 2 other 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. Finally, a least square algorithm is used to compare images and determine the most probable MTF.

4.3.7.1.2. 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 bridge observed needs 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:

$$MTF = \frac{DFT(Profile_{measured})}{DFT(Profile_{theoretical})} \qquad (17)$$

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).

4.3.7.1.3. 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.

In order to manage the replication, through oversampling, there should be an inclination of edge relative to the line or row of the image. Thanks to this inclination a 1D oversampled edge response is created thanks to the 2D undersampled image.

Thanks to this oversampling, in the direction perpendicular to the edge, equation (16) becomes:

$$ImageSpectrum(f) = Heaviside(f).H(f) \qquad (18)$$

Which leads to:

$$MTF(f) = |H(f)| = \left| \frac{ImageSpectrum(f)}{Heaviside(f)} \right| \qquad (19)$$

4.3.7.2. Results

For the edge method, MTF was deduced from an image from Maricopa field and from an image of Ross ice Shelve 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 Figure 41.
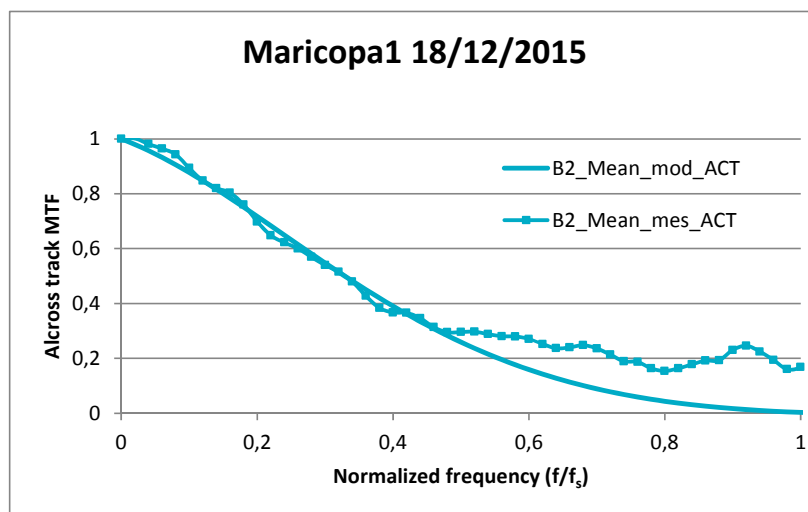

Figure 41. MTF curve for B2 band for the across track direction

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 Figure 42 and Figure 43.
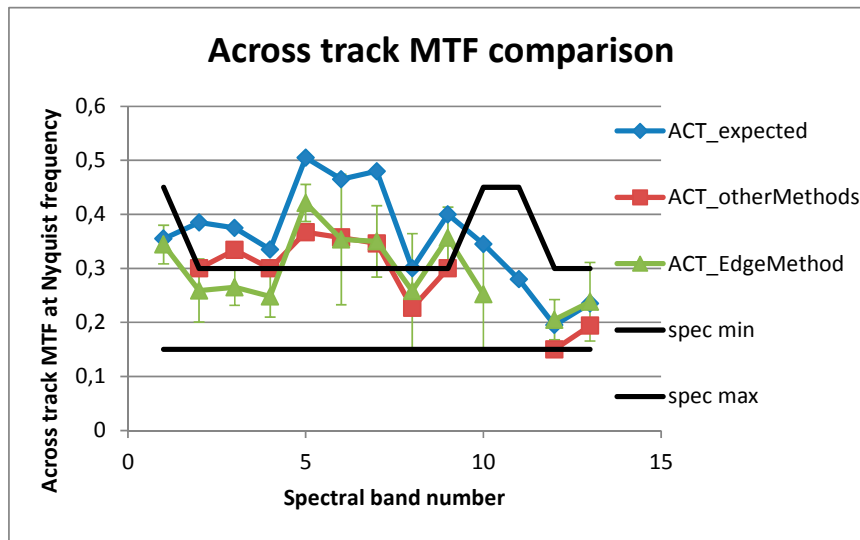
Figure 42. MTF results for the across track direction
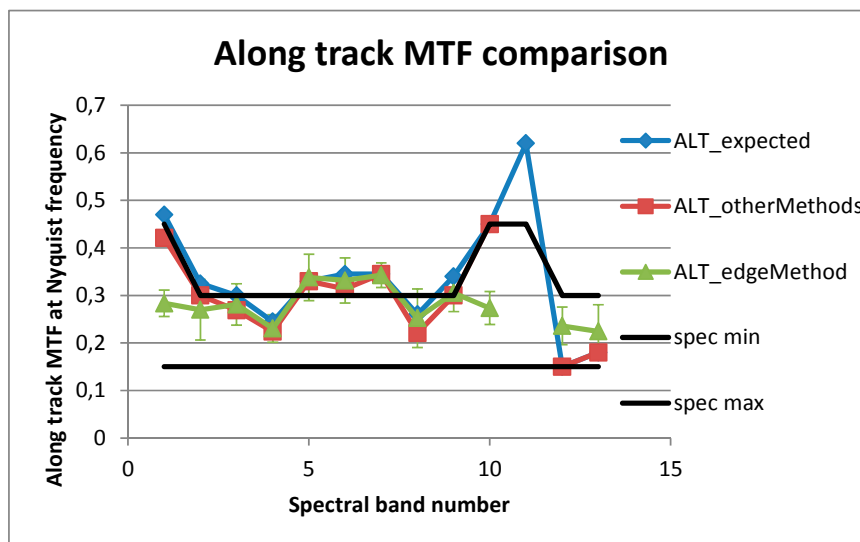
Figure 43. MTF results for the along track direction

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

For the sensor estate of December 2015, we can consider that the requirements were met, marginally in some cases, for the along track direction and MTF was above requirement for B5, B6, B7 and B8A for the across track direction.

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-frequency 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 time of writing, 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 in Sentinel 2 images by correlation with a database of accurately localised images spread over the world. The general principle is shown in Figure 44 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 the two images in the same geometry: Either S2 images are 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. @95.45% confidence level).

The processing is performed using Sentinel 2 band B4 as reference band. The 10m bands obviously provide the better correlation accuracy.
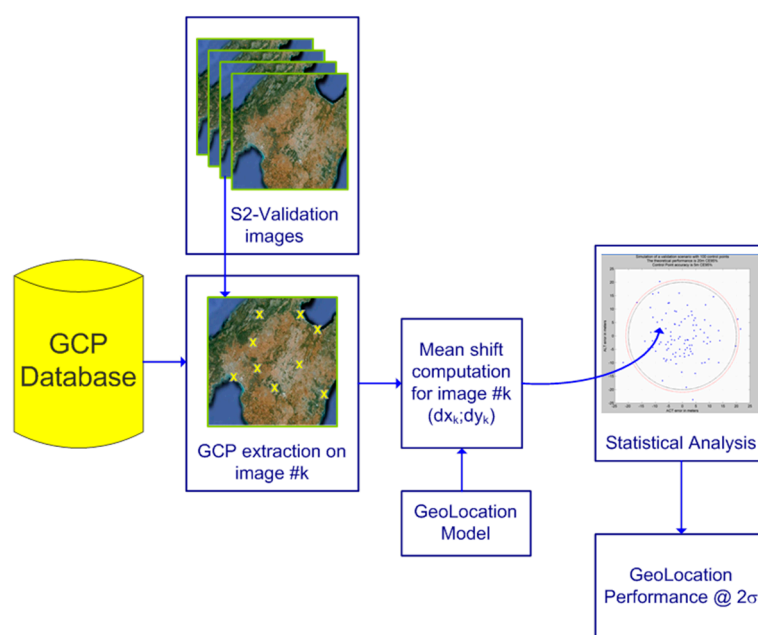
Figure 44. General principle of geolocation uncertainty validation using ground control points. Mean shift computation is obtained by correlation technique.

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 45.
- Pleiades HR database including more than 500 images with an accuracy less than 5 m is used by CNES.
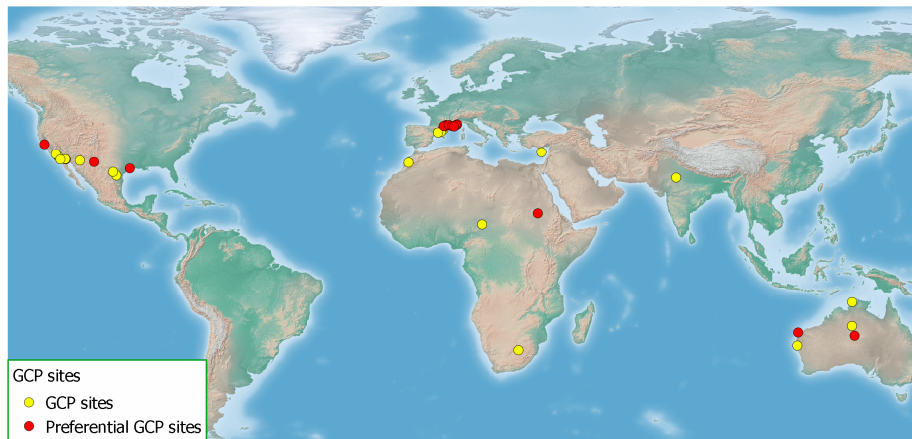
Figure 45. Map of TAS GCP Database sites (Red rings are GCP sites currently available and used in baseline for S2 validation activities).

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 shift 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 GCP image (in cartographic projection). The estimation of performance is then directly in meters 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 no longer estimated in metres on the ground, but in the focal plane.

In this case, we use the inverse location function: from the GCP ground coordinates we estimate the image coordinates by means of the inverse geometric model and then compare them with measured image coordinates. Hence we obtain 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.

Performance metrics: the performance indicator to be compared to the requirement is computed as follws:

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.

4.4.1.1.1. Results for Non-Refined Products

Geolocation performance on non-refined products represents the absolute location performance as provided by the system, and which 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 product baseline 2.04 suffered from a yaw bias correction anomaly, particularly visible at the edge of the swath, leading to a performance of about 14.5m CE@95.5% confidence Level on L1C products.

Since product baseline 2.04 the performances are very good on both types of products L1B and L1C, and considerably under the limit fixed by specifications. Following table Table 12 and Figure 46 and Figure 47 present the statistics obtained on system geolocation performances for L1 non-refined products.

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
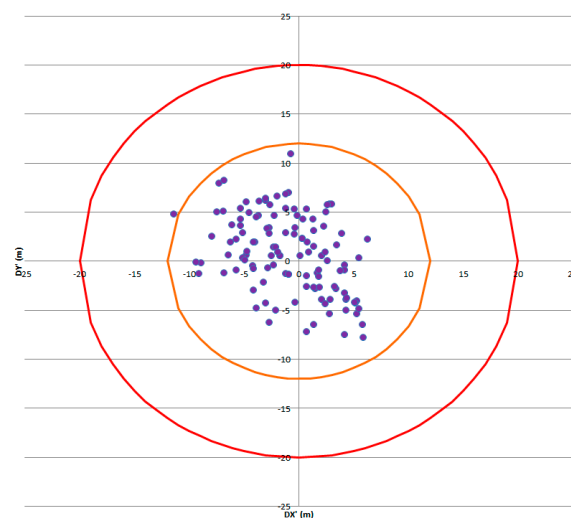
Figure 46. System Geolocation Performances in meters, (B4 Band), L1B non refined products.

(a)   (b)
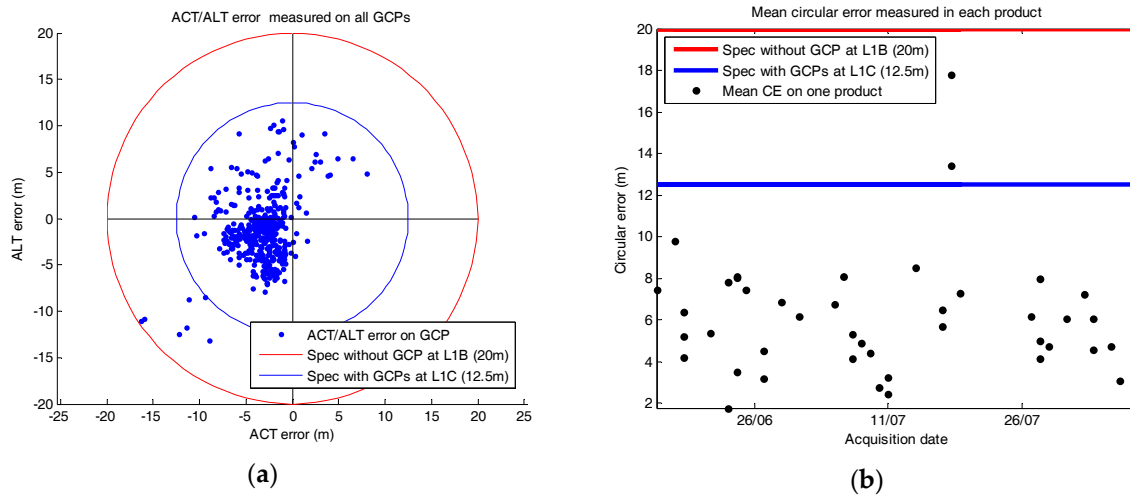
Figure 47. System Geolocation Performances in meters for L1C non refined products. Reference band is B04. (a) geolocation error (in metre) along-track and across-track measured on each GCPs. (b) This figure shows the mean geolocation error in each product processed in function of the acquisition dates of the product. The outliers seen on the 2 figures correspond to products from orbit 5601 and are currently investigated.

The 95.45 % conf. level ("2 σ") system geolocation performance value obtained on analysed products is less than 10 m for non-refined L1B and L1C products.

It is compliant with the specification of 20m at 2 σ.

4.4.1.1.2. Results for Refined Products

The objective is to characterize the geolocation performances of a L1C product whose geometric model has been refined by on-line 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 48.

Table 13. Statistics on geolocation for refined L1C products

System Geolocation performance (Circular error)	L1C refined
95.45 % conf. level	8 m
Mean value	4.5 m
Max value	11 m

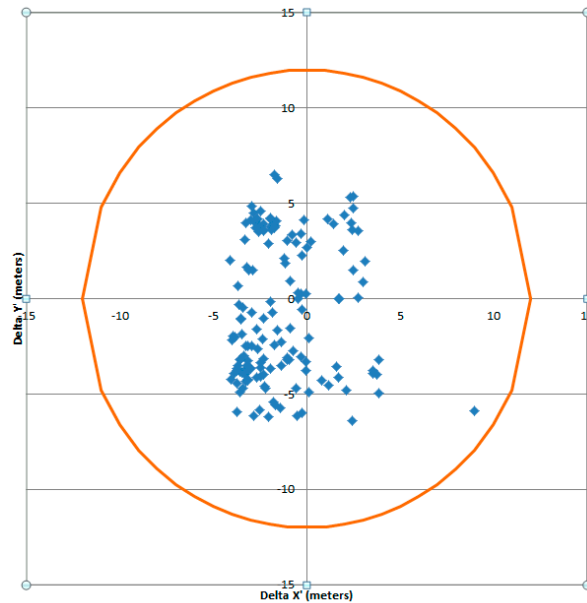**Figure 48.** Product Geolocation Performances in meters, (B4 Band), L1C refined tiles.

The 95.45 % conf. level ("2σ") product geolocation performance value obtained on analysed products is less than 8m 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 of images 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
- Product level, on L1B products with focal plane registration processing enabled. In that case the registration performance within each focal plane is the same as system level performance.

At the time of writing the focal plane registration being stable and excellent, the registration processing 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 follows (summarized in Figure 49):

For any band couple and detector in one L1B product:

1) Resample the image with the finest resolution in the geometry of the image with the coarsest resolution, using the geometric model and an accurate interpolation method (B-spline).
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 processing is applied to cloud-free products acquired on several areas at different latitudes and with various kinds of landscapes in to order 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 bad correlation points based on correlation quality criteria

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

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.
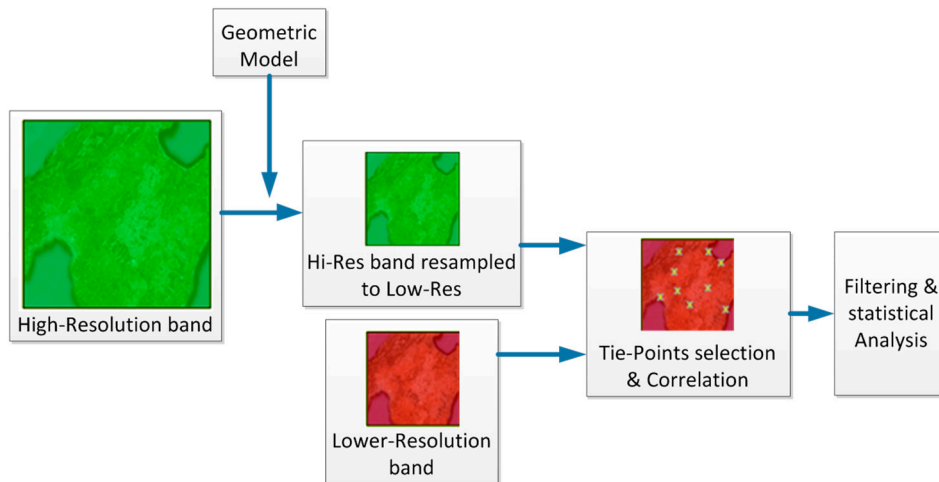
Figure 49. High-level principle of the multi-spectral registration uncertainty assessment

When the registration processing 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 and within the requirement of 0.3 pixels of the band with the coarser resolution, for any correlated band couple.

Figure 50 shows an example of measured error clouds in one detector for two band couples. The product baseline of the product used is 2.04.

Due to the very good and stable multi-spectral registration performance the VIS/SWIR registration processing has not been activated so far.
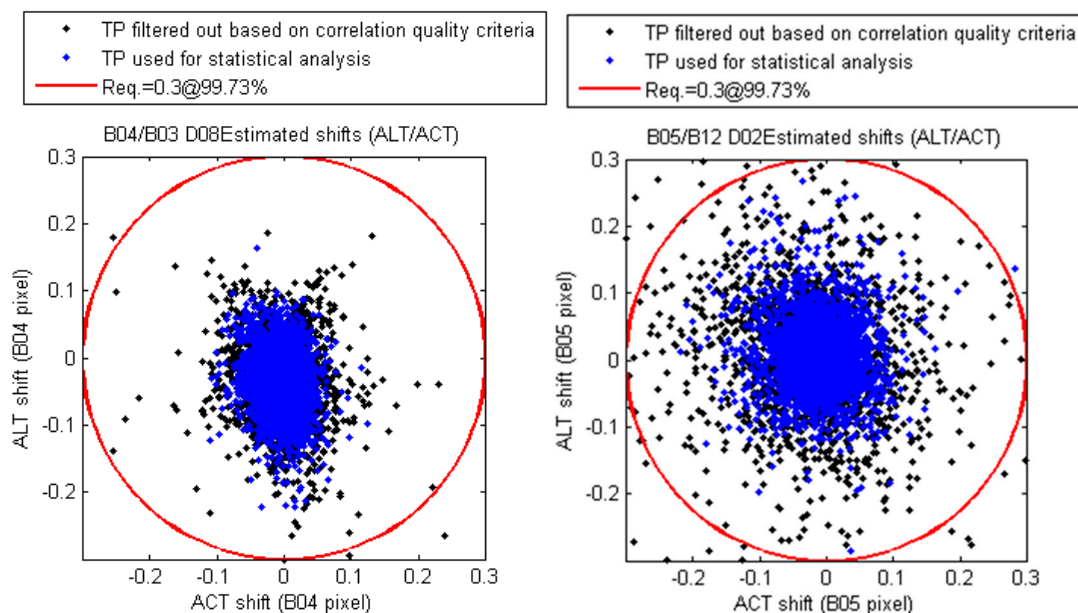
Figure 50. Example of global mis-registration shifts measured on 'Central Australia' for B04/B03-D08 and B05/B12-D02. Black points are those rejected based on correlation quality criteria (step 4)).

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

4.4.3. Multi-Temporal 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 51:

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 MGRS tile, the older L1C tile is taken as reference and all new cloud-free acquisitions of the same tile is 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 bad correlation points 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.
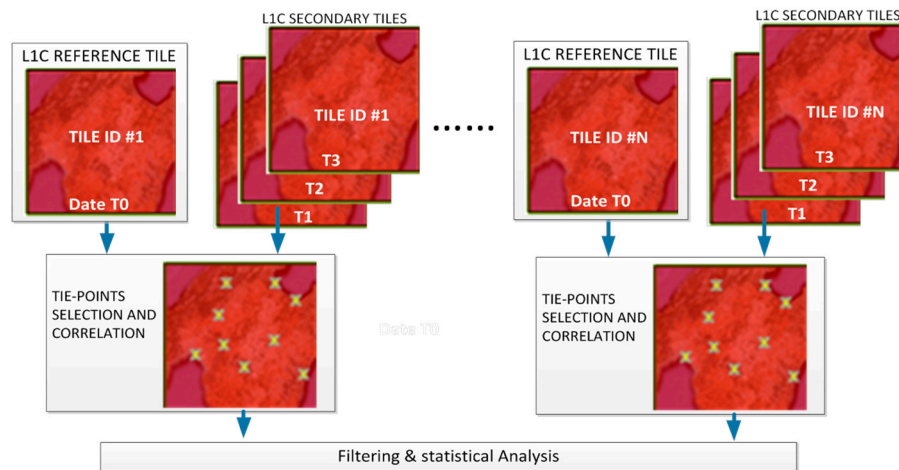

Figure 51. High-level principle of the multi-temporal registration uncertainty assessment

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 terrain, relief, surface reflectance, and image content,
- When the two tiles correlated have been acquired from two adjacent orbits (i.e. cover 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 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 (Spatial Sample Distance) at 2σ. The elevation accuracy of the PlanetDEM90 used in operational processing being 16m @2σ, the error obtained is about *16m*tan(10°)*2=5.6m=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 01/06/2016. 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. 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. Figure 53 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 processing over GRI is not activated.

The sites selected for this analysis are shown in Figure 52. They include the geolocation validation sites and additional sites specific for multi-temporal registration assessment. All sites include at least one MGRS tile.
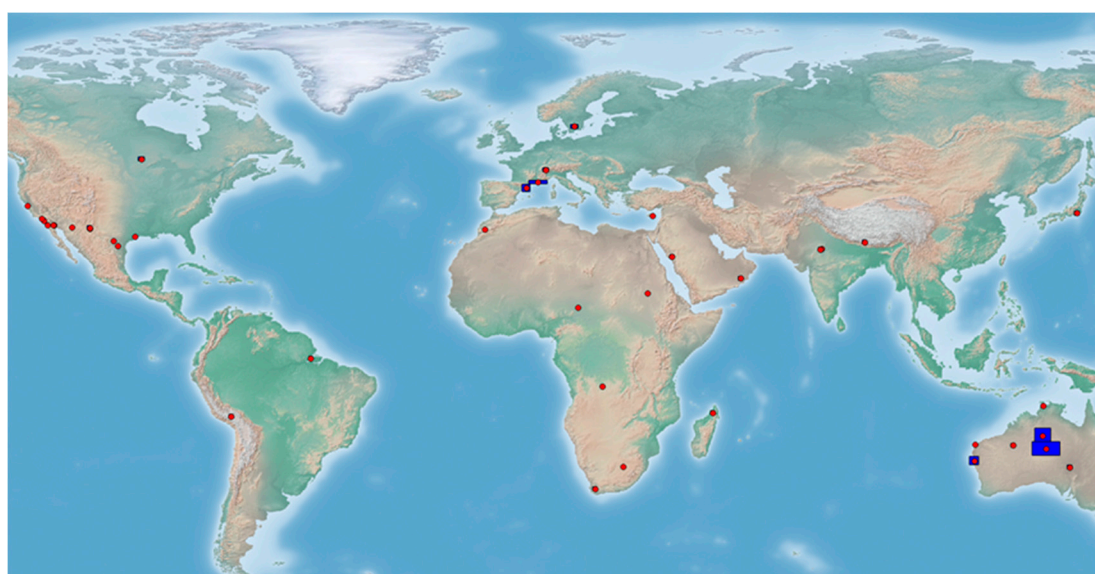

Figure 52. Sites used for multi-temporal registration uncertainty validation on non-refined products. They include the geolocation validation sites and others specific for multi-temporal registration assessment.

Figure 53. 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.

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

Multi-temporal registration performance (PB 2.04) (Circular error)	L1C non refined
95.45% conf. level	1 pixel@10m
Mean value	0.36 pixel@10m
Max value	1.5 pixel@10m

4.4.3.3. Results for Refined Products over European GRI (A. Gaudel)

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 has been 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). B4 and B11 bands are the reference bands for this analysis.

The 95.45 % confidence level multi-temporal registration performance value obtained on analysed L1C products (mean value per tile) is 0.22 pixel for B4 band (with max mean value per tile of 0.36 pixel) and 0.17 pixel for B11 band (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 high, 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 took in 9 older products with the same relative orbit. The reference and secondary tiles were acquired between July and September 2015.
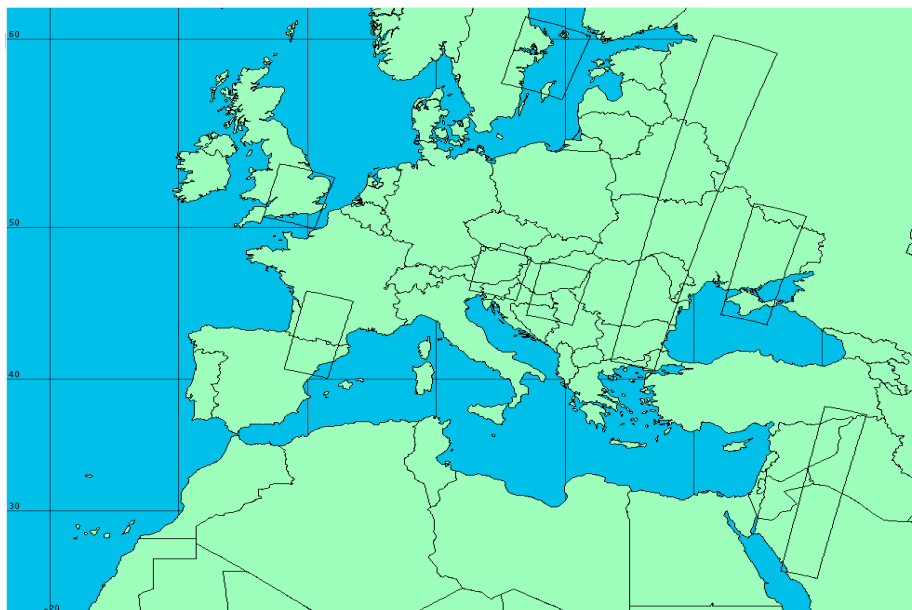
Figure 54. Sites used for multi-temporal registration uncertainty validation on refined products.

- 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)
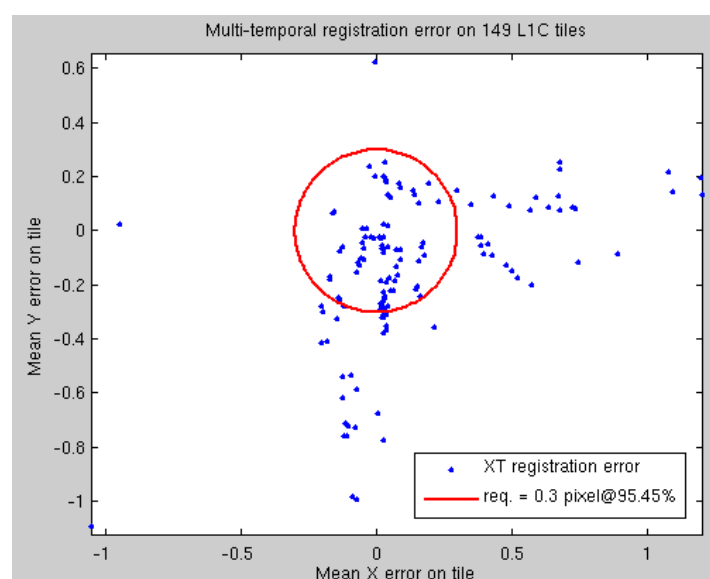- Orbit 64: 2 dates, 27 tiles (Ukraine)
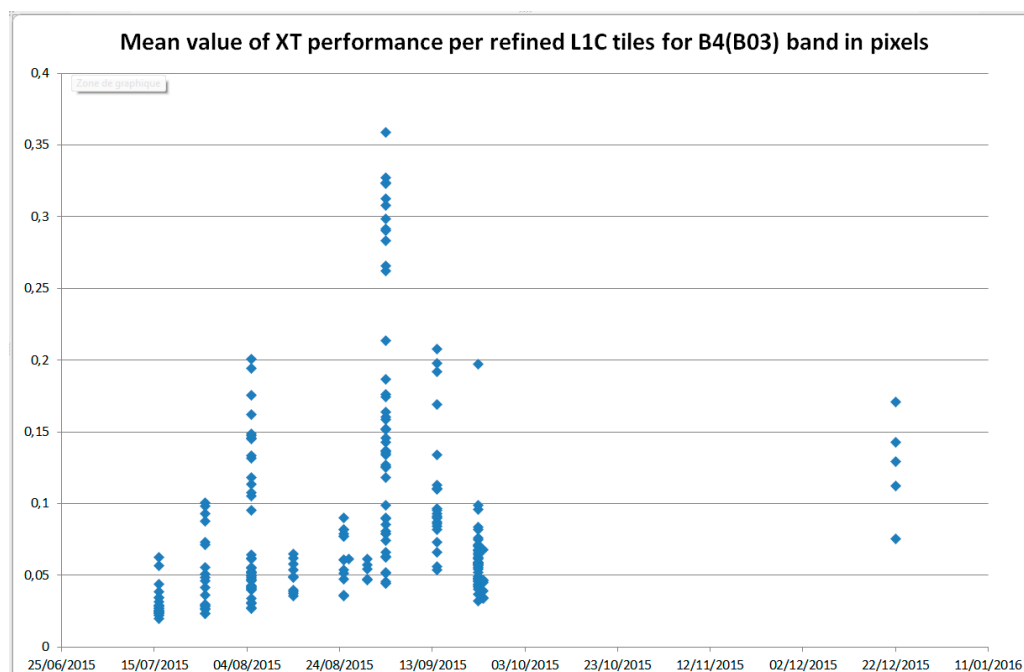
Figure 55. Multi-temporal registration uncertainty for B4 band. Each point represents the average shift measure in one secondary tile with respect to the reference tile.
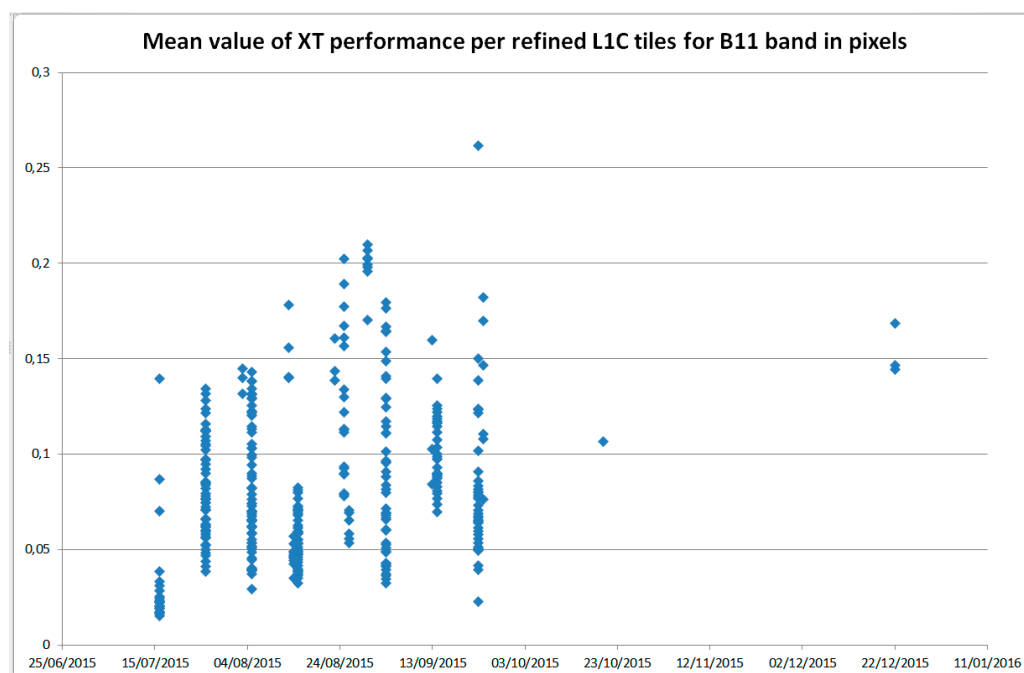
Figure 56. Multi-temporal registration uncertainty for B11 band. Each point represents the average shift measure in one secondary tile with respect to the reference tile.

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

Note that at the time of writing 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 on-going, 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 / 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 at construction).

4.4.4.2. Results for European Block of Global Reference Image

The cover percentage of each orbit of the final GRI over Europe is given by Figure 57.
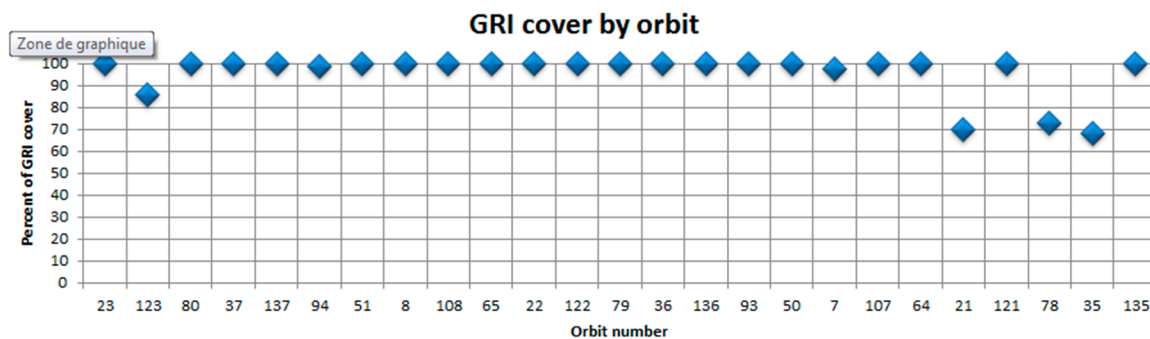
Figure 57. Cover percentage of Europe GRI versus relative orbit number.

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 is 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 (starting from November 2015).

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

The location precision of the final GRI was validated using Pleiades database (references with a location precision better than 5 m). Figure 58 represents the footprint 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.
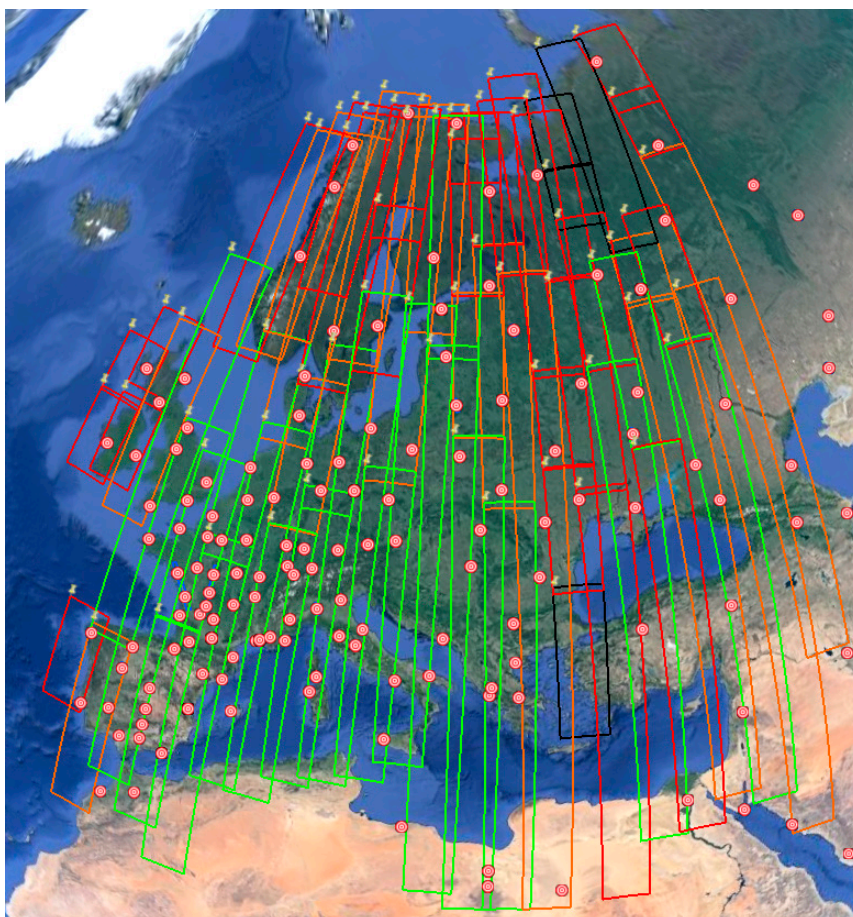
Figure 58. European GRI product footprints according to Pleiades references.

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

Figure 59 represents location precision of each European GRI product after refining, with 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 references to estimate geolocation
- White: missing products in the final GRI
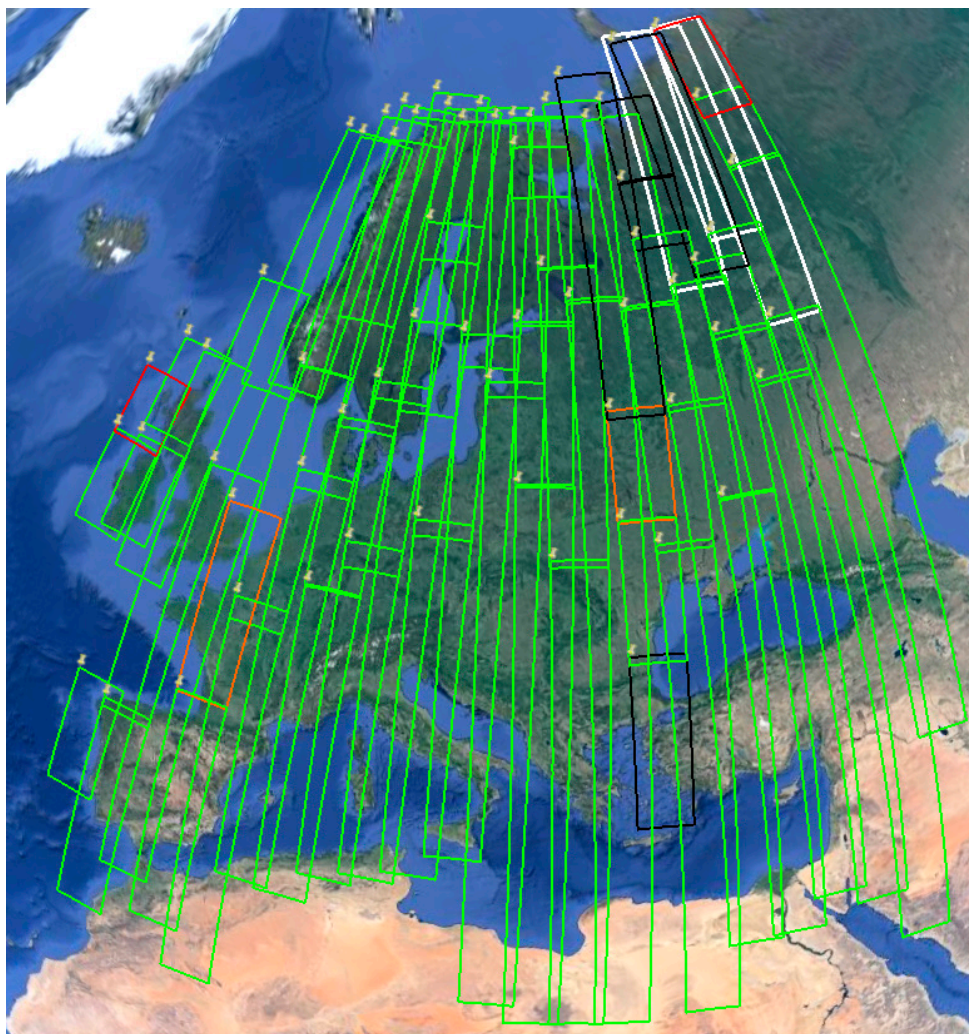
Figure 59. European GRI product geolocation performance.

Red products geolocation performance is explained because products are cloudy and not linked to other products on the GRI. So they were not linked to the Europe block in refining, and so their geolocation is not homogeneous to other products.

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

The global performance of European GRI is assessed to be 8m@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 Categorization (CSC) outputs of Sen2Cor: a Scene categorization which assigns to each pixel a category (Vegetation, Soil, Water, 2 types of Clouds, Thin Cirrus, Snow, etc.), a Cloud probabilistic mask and a Snow probabilistic mask.

Figure 60 gives an example of the outputs of Sen2Cor for one selected test site over Easton-MDE (USA). 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. SC correctly

classified water, vegetation and soil pixels. The WV and AOT-differences between the four granules representing the product are negligible. Figure 61 provides the legends and colour tables associated to the Level-2A outputs.
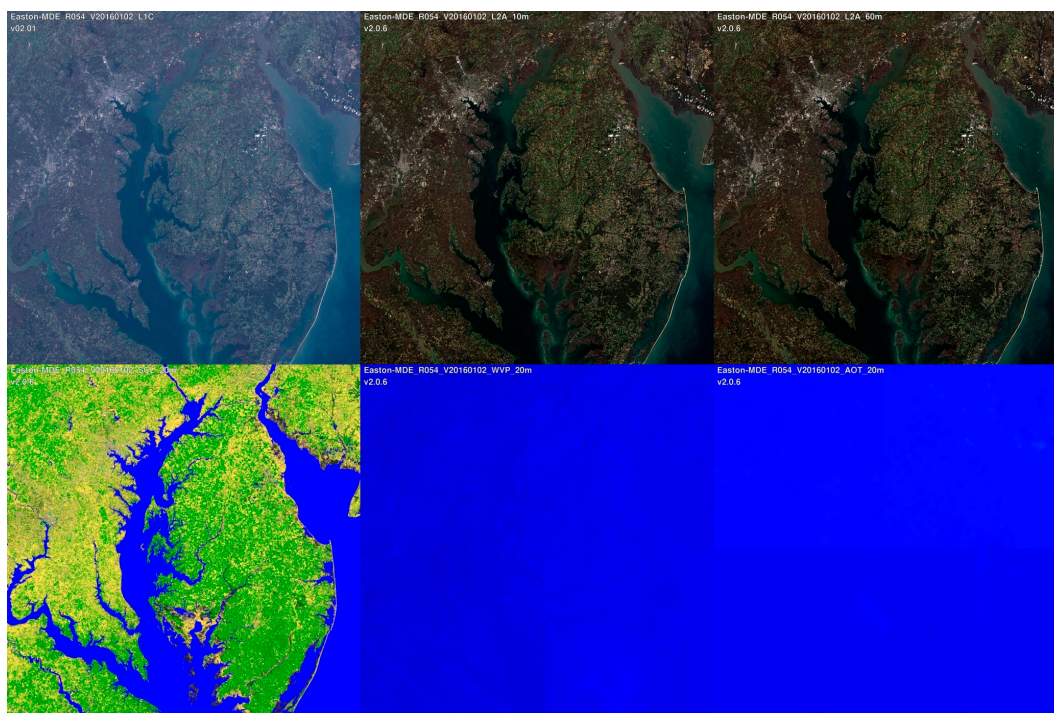
Figure 60. Level-2A product example. Site Easton-MDE (USA), acquired on 2nd of January 2016, characterized by MidlatitudeN, Flat terrain, Forest, Croplands, Water, Urban. (Top) RGB compositions (B04, B03, B02) for L1C TOA, L2A SR at 10 m, L2A SR at 60 m; (Bottom) Scene Classification, L2A Water Vapour at 20 m, L2A AOT at 20 m.
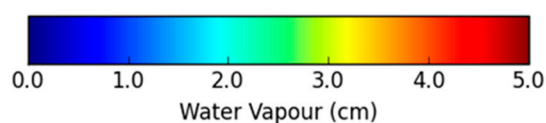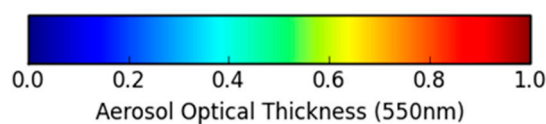
Figure 61. (**a**) Cloud Screening and Categorization (CSC) pixel legend; (**b**) Total Water Vapour Column colour table; (**c**) Aerosol Optical Thickness at 550 nm colour table.

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 in order to cover various solar angles and seasons. The 24 Level-2A calibration sites were selected above an active AERONET [31] site. Their geographical locations are shown on Figure 62. Their characteristics are further defined in [32].
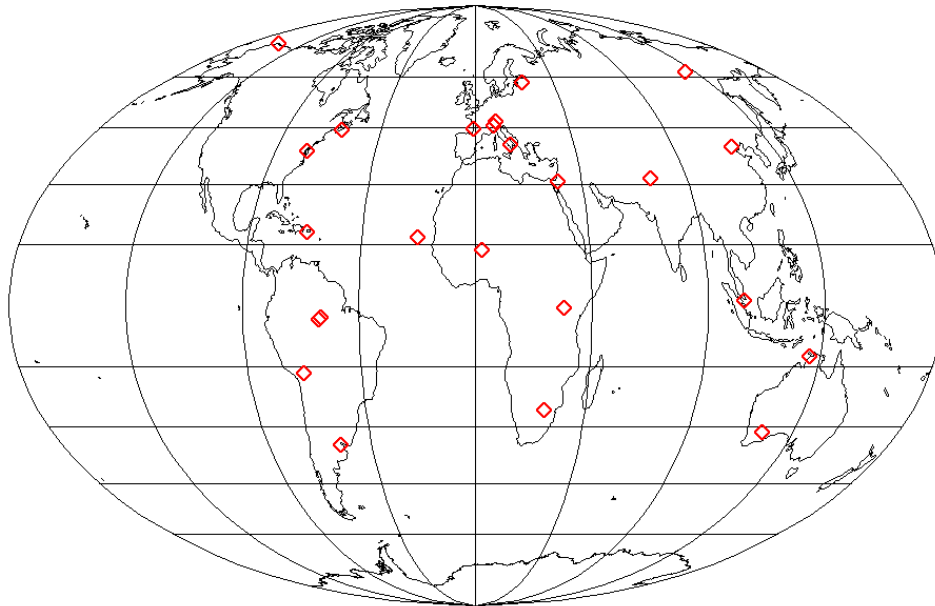
Figure 62. Geographical distribution of the 24 selected AERONET test sites for the Level-2A Calibration

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_ATM_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.

- 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.

5.1.2.1. Results and Outlook

The main result to date of this calibration activity is the Sen2Cor version 2.2.1 issued on 4th of May 2016. This version includes an updated internal resampling method that preserves the geolocation information when spectral band are resampled to another resolution. This update has improved the quality of the WV map retrieved with Atmospheric Pre-corrected Differential Absorption (APDA) algorithm [6] 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.

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 & Categorization Calibration

The objective of this task is to perform a regular check and calibration of thresholds parameters concerning the Cloud Screening and Scene Categorization (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 follow:

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 x 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 using ENVI © software to superimpose 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 the improvements to be performed are listed in three classes: ("easy", "medium", "difficult"). "Easy" corresponding to a tuning of the SC thresholds that improves the Scene Classification without having a negative impact on the other classes. (Independent thresholds) "Medium" corresponding to a tuning of the SC thresholds that improves the Scene Classification having a limited negative impact on other classes. (Relatively independent thresholds) "Difficult" corresponding to a tuning of the SC thresholds that improves the Scene Classification for one class but could have a negative impact on other classes. (interdependent thresholds between classes)

4. The proper activity of fine tuning consists in manually slightly varying the SC thresholds to improve the Scene Classification based on the three different cases (easy, medium, difficult) listed above.

5. When the fine tuning is finished, a new processing baseline is produced with a new set of updated SC parameters delivered in an updated Sen2Cor Calibration file "L2A_CAL_AC_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 x 100 km) a new scene classification map and 2 new quality indicators maps (Could confidence and Snow confidence).

7. A quantitative comparison exercise is performed between the results of the updated baseline and the results of the current baseline to assess the impact of the algorithm. i.e., the absolute and relative variation of the number of pixels per class, the class origin for the new classified pixels.

Based on this quantitative comparison exercise, the decision is taken to end the calibration procedure or to start another fine tuning (step 4) if it seems possible to improve the SC parameters calibration (improvement that doesn't require a further evolution of the SC algorithm).

5.1.3.2. Results and Outlook

The main result of this calibration activity is the Sen2Cor version 2.2.1 issued on 4th of May 2016. This version includes an updated set of thresholds for Cloud Screening and scene Categorization (CSC) which improve the overall categorization accuracy.

In addition to the thresholds tuning, several evolutions have been introduced in the CSC algorithm that are listed hereafter:

*   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 63 (a) gives an example of the cloud shadow class appearance in the Scene Classification map.

(a)   (b)
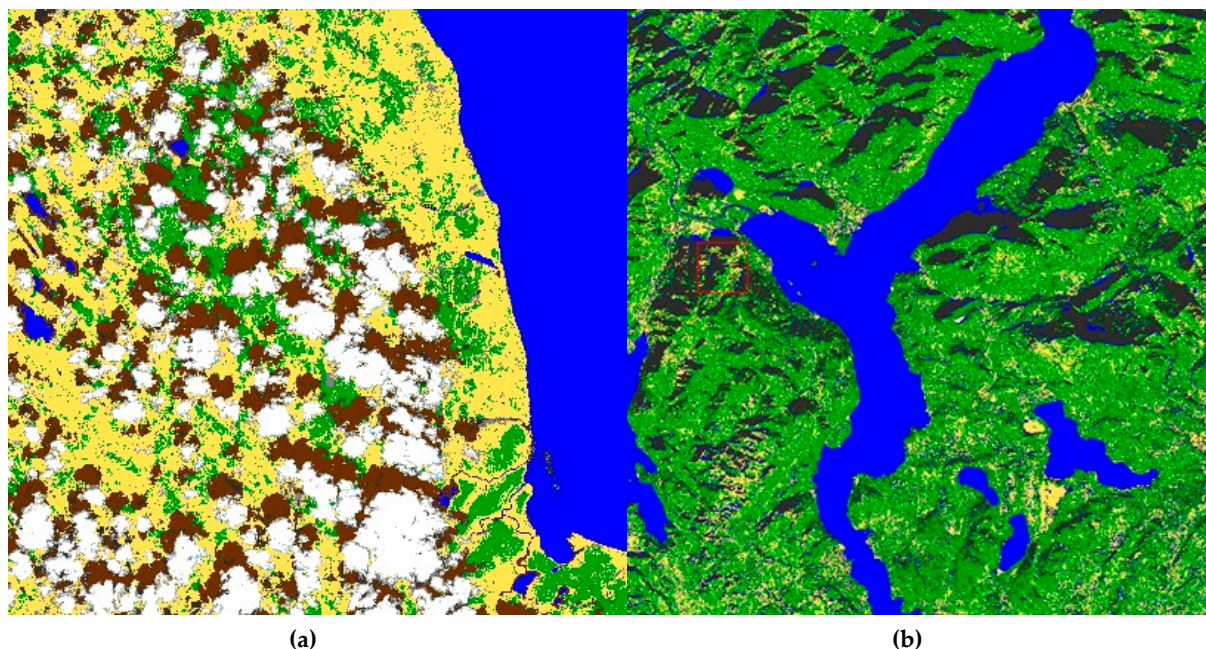
Figure 63. (**a**) Cloud shadows classification; (**b**) Topographic shadows classification

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 63 (b) 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 classification validation of the CSC-product and radiometric validation of WV, AOT- and SR-products. Validation data sets differ for CSC validation and for radiometric validation. CSC-validation is performed on full granules with area of 100x100 km2 whereas radiometric validation works on area of only 9x9 km2 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 Classification Validation

The scene classification algorithm is designed to detect clouds, snow and cloud shadows and to generate a classification map, which consists of 4 different classes for clouds (including cirrus), together with six different classifications for shadows, cloud shadows, vegetation, soils / deserts, 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 CSC validation is limited due to the lack of "ground truth" data sets. The applied validation method performs stratified random sampling followed by establishment of 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

Classification ground-truth images were created and compared to corresponding pixels of the sen2cor classification images for five scenes at different locations. One example for data analysis is shown in Figure 64

TOA-RGB   **Scene classification**   **Ground-truth classification image**   **Corresponding pixels from scene classification image**

Figure 64. Example images for Cloud Screening & Classification Validation analysis, Site Potsdam (Germany), acquired on 22th of April 2016, characterized by Boreal region, flat terrain

The image contains some clouds, cloud shadows, vegetated, soil 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, bare soils, water and high probability clouds and cloud shadow classes have been correctly classified for a very large portion of validation data. The resulting classification precision is 0.98 for vegetation, 0.96 for bare soils, 0.93 for water, 0.88 for dark area & cloud shadows and 0.85 for high probability clouds. Classification precision reaches only 0.71 for thin cirrus and 0.30 for class medium probability clouds. Precision is a value indicating the probability that the pixel was classified correctly. Average precision for the example image presented is 0.64. Other measures for characterization of classification performance are recall and overall accuracy. Recall indicates the probability that a pixel belongs to the class classified by Sen2Cor and overall accuracy indicates the relative number of correctly classified pixels relative to the total number of pixels. Recall and overall accuracy for the example presented are 0.56 and 84%.

Table 15. Example of confusion matrix for classification validation

Sen2Cor class	Reference class									
	(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
(2+3)	Dark area, cloud shadows	0	**6785**	880	0	0	5	0	0	0	0
(4)	Vegetation	0	31	**15398**	52	0	46	0	0	198	0
(5)	Bare soils	0	441	406	**38063**	0	177	34	460	0	0
(6)	Water	0	3223	3	0	**42351**	110	0	0	89	0
(7)	Cloud low probability	0	1	154	354	0	**205**	163	212	33	0
(8)	Cloud medium probability	0	21	75	599	0	517	**823**	347	327	0
(9)	Cloud high probability	0	0	0	24	0	14	533	**4197**	148	0
(10)	Thin cirrus	0	3234	6444	2205	0	934	7010	187	**48308**	0
(11)	Snow	0	0	0	0	0	0	0	0	0	**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%. Mean precision is 0.58±0.08 and mean recall is 0.55±0.07. Water (if present in the image) and high probability clouds are classified with similar high precision for all examples investigated. Classification precision for classes vegetation and bare soils vary from scene to scene due to the different type of landscape contained. Class bare soils is a class including a relative broad set of possible objects. It contains not only soil but urban objects, roads or railroads. Also classification of dark area pixels & cloud shadows vary due to the different terrain 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 with 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 [31] sunphotometer measurements. For WV validation, the average Sen2Cor output over all soil and vegetation pixels within 9km x 9km 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. Results for processing with 60 m, 20 m or 10 m spatial resolution give consistent results. AOT validation results are shown in Figure 65. 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 around 0.30 when there are not sufficient DDV-pixels in the image. Limiting the statistical analysis to samples with sufficient 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 0.05. This

result is in accordance with experience from earlier validation studies for Landsat and RapidEye sensors. Present validation results for Sen2Cor are in-line with validation results for other atmospheric correction algorithms found in the literature [33].
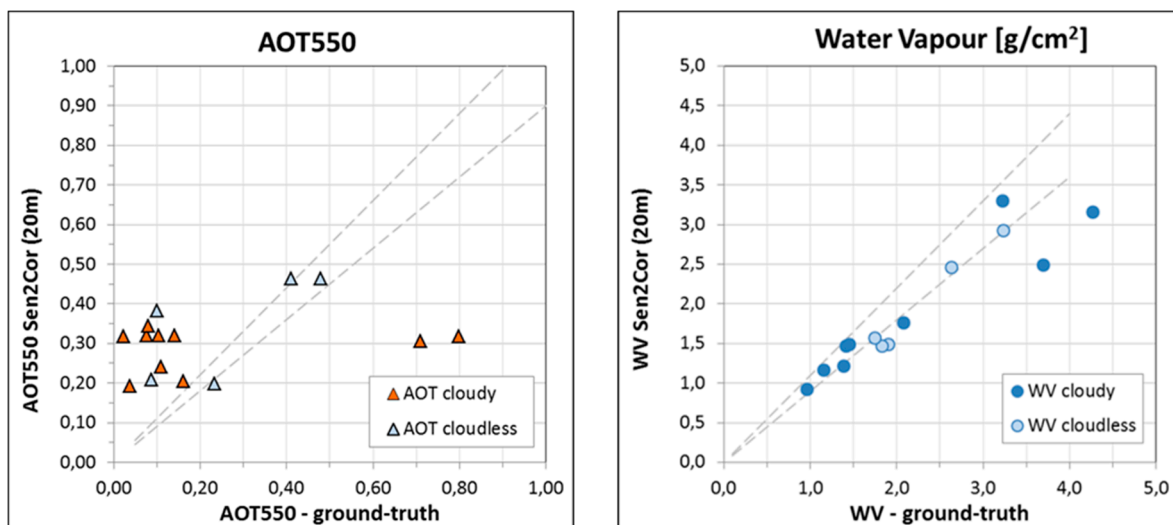

Figure 65. AOT- and WV Validation, Direct comparison of Sen2Cor with ground reference from AERONET. Sen2Cor is represented by average over 9 km x 9 km area with AERONET sunphotometer in the centre. Filled symbols represent samples with cloudiness > 5% and not filled with cloudiness < 5 %.

Water vapour retrieval gives better results than AOT (Figure 65). 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 with maximum difference 0.42.

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 Level-2A processor Sen2Cor. The validation makes use of the same test sites and in-situ data sets 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 x 9 km subset around the sunphotometer location. Reference data are computed by a second run of Sen2Cor using AOT 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

Results for processing 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 sufficient DDV-pixels in the image (Figure 66). Example spectra were extracted at different locations in the reference image and in the image to validate. Example spectra for dark and bright soil, for forest and for different other vegetated locations show the expected spectral dependency and agree within the target accuracy of 5 % relative for BOA-reflectance. The SR differences up to 0.04 between Sen2Cor processing and reference leads to a Normalized Density Vegetation Index (NDVI) uncertainty up to 0.06.
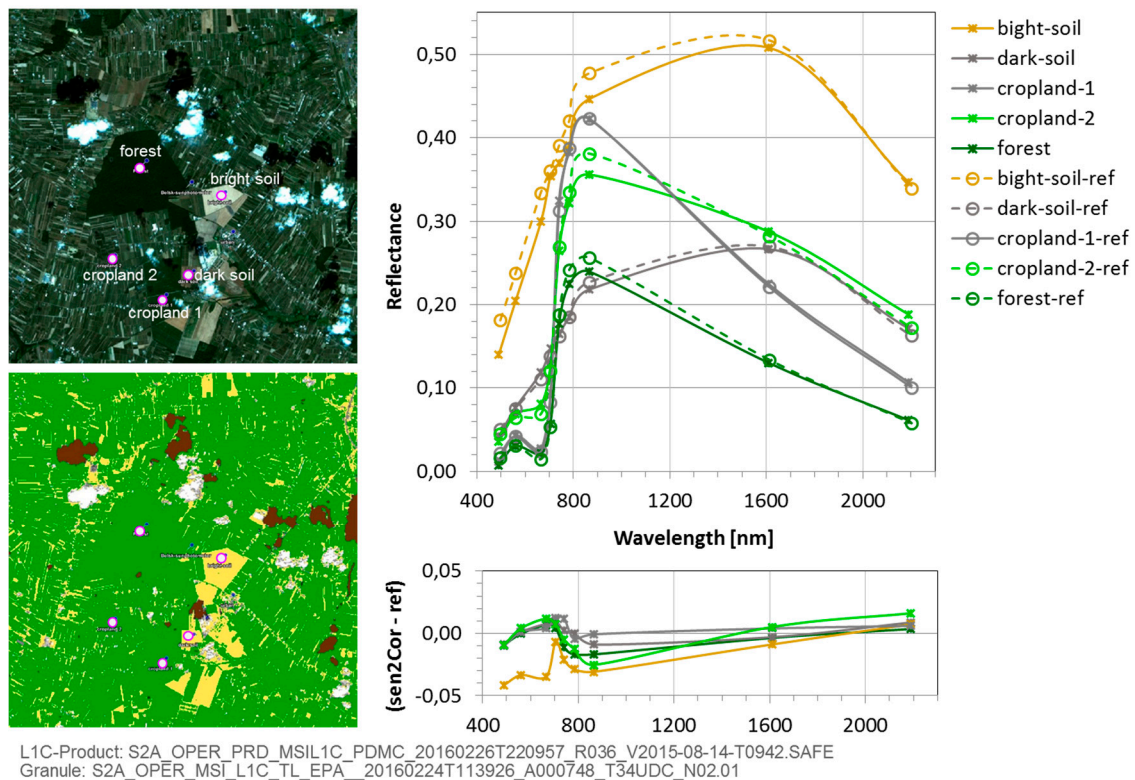
L1C-Product: S2A_OPER_PRD_MSIL1C_PDMC_20160226T220957_R036_V2015-08-14-T0942.SAFE
Granule: S2A_OPER_MSI_L1C_TL_EPA__20160224T113926_A000748_T34UDC_N02.01

Figure 66. Validation example for BOA-reflectance, Site Belsk (Poland), acquired on 14th of August 2015, characterized by flat terrain. Left: BOA-RGB and scene classification image. Right: Example spectra and relative difference to reference for soil and vegetated pixels at indicated locations.

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.

Finally, the Sentinel-2 timeseries need to meet the high availability, reliability and timeliness performance requirements in order 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.

7. References

1. M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, P. Bargellini, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services", Remote Sensing of Environment, Volume 120, 15 May 2012, Pages 25–36.

2. Thuillier, G., M. Hersé, D. Labs, T. Foujols, W. Peetermans, D. Gillotay, P.C. Simon, and H. Mandel, "The solar spectral irradiance from 200 to 2400 nm measured by the SOLSPEC spectrometer from the ATLAS and EUREKA missions", Solar Physics, 214: 1–22, 2003.

3. Footprint of all Level-1C tiles in a Keyhole Mark-up Language (KML) file from Sentinel-2 website https://sentinels.copernicus.eu/documents/247904/1955685/S2A_OPER_GIP_TILPAR_MPC__20151209T09 5117_V20150622T000000_21000101T000000_B00.kml

4. Sentinel-2 Toolbox: Available online: http://step.esa.int/main/toolboxes/sentinel-2-toolbox (accessed on 07 09 2016).

5. Louis, J. (2016). Sentinel 2 MSI - Level 2A Product Definition. Issue 4.4. 2016-08-12. Available online: https://sentinel.esa.int/documents/247904/1848117/Sentinel-2-Level-2A-Product-Definition-Document.pdf (accessed on 07 09 2016).

6. Schläpfer, D., Borel, C.C., Keller, J., and Itten, K.I. (1998). Atmospheric precorrected differential absorption technique to retrieve columnar water vapour, Remote Sens. Environ., Vol. 65, Pages 353-366. doi:10.1016/S0034-4257(98)00044-3 Schläpfer, D., Borel, C.C., Keller, J., and Itten, K.I. (1998).

7. Kaufman, Y. J., and Sendra, C. (1988). Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery, Int. J. Remote Sensing, Vol. 9, Pages 1357-1381. doi: 10.1080/01431168808954942

8. Richter, R., Louis, J., Müller-Wilm U. (2012). Algorithm Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document.S2PAD-ATBD-0001, Issue 2.0.

9. Data Quality Report published on a monthly basis: https://sentinel.esa.int/web/sentinel/data-product-quality-reports

10. Maisonobe, L., Pommier-Maurussane, V., "Orekit: an Open-source Library for Operational Flight Dynamics Applications", in 4th ICATT, May 2010.

11. Gordon, H. R., and M. Wang. 1994. Retrieval of Water-Leaving Radiance and Aerosol Optical Thickness Over the Oceans with SeaWiFS: A Preliminary Algorithm. Applied Optics 33: 443–452.

12. Morel, A. 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 water), Journal of Geophysical Research, 93, 10,749-10,768.

13. Morel, A., and L. Prieur. 1977. Analysis of Variations in Ocean Color. Limnology and Oceanography 22 (4): 709–722.

14. Hagolle et al. 1999. Results of POLDER in-flight Calibration, IEEE Transactions on Geoscience and Remote Sensing, May 1999, Volume 37, Number 03.

15. Vermote, E., R. Santer, P.Y. Deschamps and M. Herman. 1992. In-flight Calibration of Large Field-of-View Sensors at Short Wavelengths using Rayleigh Scattering, Int. Journal of Remote Sensing, 13, No 18, 1992.

16. Morel, A. and Maritorena, S. 2001. Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical research, 106, 7763-7780.

17. Morel, A., and B. Gentili. 1996. Diffuse reflectance of oceanic waters. 3. Implication of bidirectionality for the remote-sensing problem. Applied Optics, 35, 4850-4862.

18. DIMITRI: Database for Imaging Multi-spectral Instruments and Tools for Radiometric Inter-comparison: https://dimitri.argans.co.uk

19. Fougnie, B., Patrice Henry, P., Morel, A., Antoine, D., Montagner, F. 2002. Identification and characterization of stable homogenous oceanic zones: climatology and impact on in-flight calibration of space sensors over rayleigh scattering. Ocean Optics XVI, Santa Fe, NM.

20. Fougnie, B. and Bach, R. 2009. Monitoring of Radiometric Sensitivity Changes of Space Sensors Using Deep Convective Clouds: Operational Application to PARASOL. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 47, NO. 3, DOI: 10.1109/TGRS.2008.2005634.

21. Thome, K.J., Helder, D.L., Aaron, D., and Dewald, J.D. 2004. Landsat-5 TM and Landsat-7 ETM+ absolute radiometric calibration using the reflectance-based method. IEEE Transactions on Geoscience and Remote Sensing, Vol. 42, pp. 2777–2785. doi:10.1109/TGRS.2004.839085.

22. Czapla-Myers, J., McCorkel, J., Anderson, N., Thome, K. Biggar, S., Helder, D., Aaron, D., Leigh, L., & Mishra, N. 2015. The Ground-Based Absolute Radiometric Calibration of Landsat 8 OLI, 626; Remote Sens. doi:10.3390/rs70100600.

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

24. Alhammoud, B., Bouvet, M., Jackson, J., Arias, M., Thepaut, O., Lafrance, B., Gascon, F., Cadau, E., Berthelot, B., Francesconi, B., 2016, On the vicarious calibration methodologies in DIMITRI: Applications on Sentinel-2 and Landsat-8 products and comparison with in-situ measurements, proceeding ESA Living Planet Symposium, 9-13 May 2016, Prague, Czech Republic, ESA Special Publication SP-740 (CD-ROM), 2016.

25. CEOS-PICS sites: http://calvalportal.ceos.org

26. Bouvet, M. 2014. Radiometric comparison of multispectral imagers over a pseudo-invariant calibration site using a reference radiometric model, Remote Sensing of Environment 140 (2014).

27. Cosnefroy, H., Leroy, M. and Briottet M. (1996). Selection and characterisation of saharan and Arabian desert sites for the calibration of optical satellite sensors, Remote Sens. Environ. Vol. 58, No. 1., pp. 2713-2715.

28. IOCCG (2013). In-flight calibration of Satellite ocean-colour sensors, Frouin, R. (eds.), report of the IOCCG No 14, Dartmouth, Canada, 2013.

29. Lacherade, S., Fougnie, B., Henry, P. and Gamet, P. 2013. Cross calibration over desert sites: description, methodology, and operational implementation, IEEE transaction on geosciences and remote sensing, Vol.51, No.3.

30. Bouvet, M. 2006. Intercomparison of imaging spectrometer over the Salar de Uyuni (Bolivia), Proceedings of the 2006 MERIS AATSR Validation Team Workshop.

31. 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.; Lavenu, F.; Jankowiak, I.; Smirnov, A). AERONET—A federated instrument network and data archive for aerosol characterization. Remote Sens. Env. 1998, 66, 1–16, DOI 10.1016/S0034-4257(98)00031-5. Available online: http://www.sciencedirect.com/science/article/pii/S0034425798000315 (accessed on 07 09 2016).

32. S2MPC - Calibration and Validation Plan for the MPC Routine Operation Phase, issue 05, 2016-04-11.

33. Breon, F.M., Vermeulen, A. Descloitres, J. An evaluation of satellite aerosol products against sunphotometer measurements, Remote Sens. Env. 2011, 115, 3102-3111, DOI 10.2016/j.rse 2011.06.017. Available online: http://www.sciencedirect.com/science/article/pii/S0034425711002410 (accessed on 07 09 2016).