SENTINEL-2
GLOBAL REFERENCE IMAGE VALIDATION AND APPLICATION TO MULTI-TEMPORAL PERFORMANCES AND HIGH LATITUDE DIGITAL SURFACE MODEL

A. Gaudel [a], F. Languille [a], J. M. Delvit [a], J. Michel [a], M. Cournet [a], V. Poulain [b], D. Youssefi [c, *]

[a] CNES, 18 Avenue Edouard Belin 31401 Toulouse Cedex 09, France - (angelique.gaudel-vacaresse, florie.languille, jean-marc.delvit, julien.michel, myriam.cournet)@cnes.fr
[b] Thales Services, 290 Allée du lac, 31670 Labège, France - vincent.poulain@thales-services.fr
[c] CS-SI, 5 rue Brindejonc des moulinais, BP 15872, 31506 Toulouse Cedex 04, France - david.youssefi@c-s.fr

Commission VI, WG VI/4

KEY WORDS: Sentinel-2 (S2), Global Reference Image (GRI), multi-temporal registration, Digital Surface Model (DSM)

ABSTRACT:

In the frame of the Copernicus program of the European Commission, Sentinel-2 is a constellation of 2 satellites with a revisit time of 5 days in order to have temporal images stacks and a global coverage over terrestrial surfaces. Satellite 2A was launched in June 2015, and satellite 2B will be launched in March 2017.
In cooperation with the European Space Agency (ESA), the French space agency (CNES) is in charge of the image quality of the project, and so ensures the CAL/VAL commissioning phase during the months following the launch. This cooperation is also extended to routine phase as CNES supports European Space Research Institute (ESRIN) and the Sentinel-2 Mission performance Centre (MPC) for validation in geometric and radiometric image quality aspects, and in Sentinel-2 GRI geolocation performance assessment whose results will be presented in this paper. The GRI is a set of S2A images at 10m resolution covering the whole world with a good and consistent geolocation. This ground reference enables accurate multi-temporal registration of refined Sentinel-2 products.
While not primarily intended for the generation of DSM, Sentinel-2 swaths overlap between orbits would also allow for the generation of a complete DSM of land and ices over 60° of northern latitudes (expected accuracy: few S2 pixels in altimetry). This DSM would benefit from the very frequent revisit times of Sentinel-2, to monitor ice or snow level in area of frequent changes, or to increase measurement accuracy in areas of little changes.

1. INTRODUCTION

As part of the Copernicus program 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 (Drusch, 2012). 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 from −56 ° to 84 ° latitude, a high revisit of five days under the same viewing conditions with the help of two identical Sentinel-2 satellites (called Sentinel-2A launched in June 2015 and Sentinel-2B launched in March 2017), high spatial resolution (10m-20m-60m), and a wide field of view (290 km) for multispectral observations from 13 bands in the visible, near infrared and short wave infrared range of the electromagnetic spectrum.

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 (Gascon, 2016).
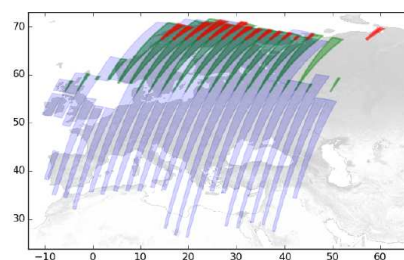

Figure 1. Sentinel-2 swath overlap over Europe.
Blue areas represent overlap between adjacent orbits, while green (resp. red) areas represent overlap between 1 orbit (resp. 2 orbits) apart.

This paper provides a description of one part of the validation activities performed by CNES in the cooperation frame with ESA/ESRIN, focused on Sentinel-2 GRI geolocation performance, and therefore the multi-temporal registration performance ensured by refinement of Sentinel-2 products on this GRI (Languille, 2016).

As an application, this paper also intends to demonstrate the surprising capability of the Sentinel-2 mission for the generation of DSM (Michel, 2017), by taking the advantage of the large swath whose overlaps cover fractions of Earth surface, the area covered by overlaps increasing with the latitude until becoming complete starting at 60° of northern latitude (Figure 1). This way, the principle of stereo-restitution can be applied as deriving the height of a given point on Earth from remote sensing images. It requires at least two images where this point can be seen, with different viewing angles.

2. SENTINEL-2 GLOBAL REFERENCE IMAGE

2.1 Introduction

The Global Reference Image (Dechoz, 2015) can be defined as a set of as cloud free as possible mono-spectral (central B4 red channel, 10m resolution) L1B products, whose geometrical model has been refined through a dedicated process designed by IGN (Institut Géographique National). The area of interest is worldwide, including most of the isolated islands (Figure 2). This GRI is delivered by independent "blocks" for each continent. So that once the GRI is complete the whole archive of L1B products can benefit from its geolocation. GRI can present several layers of images on very cloudy areas (like on equatorial zones or isolated islands). GRI can also contain some holes but without impact on the refining process as S2 acquired segments are very long. 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, and 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 (~2σ).

• The multi-temporal registration performance between refined products is better than 0.3 pixels CE99 (~3σ).

This absolute GRI geolocation performance has then been assessed to 9,5m CE95 in order to be compliant with requirements on refined Sentinel-2 products for absolute geolocation and mainly for multi-temporal registration.
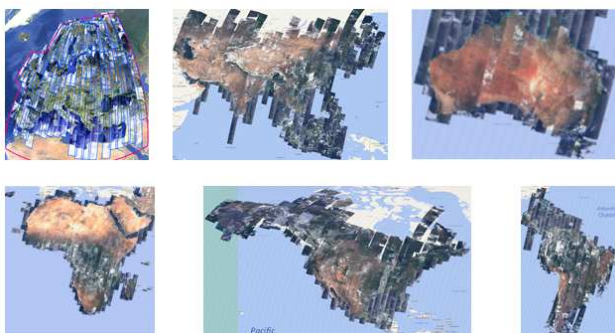

Figure 2. Overview of the Sentinel-2 GRI selection done by IGN, March 2017. European GRI products were selected in the frame of S2A In-Orbit Commissioning Review (IOCR) context, in October 2015, with CNES IQ expertise center (ESA cooperation), whereas other blocks were selected in the frame of S2A routine context with MPC (ESRIN contract).

2.2 Global Reference Image Validation Method

This activity consists 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, the validation of North American block is on-going.

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. The only difference is that the geometric model of the GRI has been refined by IGN.

This assessment is based on detection of Ground Control Points (GCP) by correlation with a database of accurately localised images spread over the world.

These ground control points have to be as accurate as possible, just as for the absolute calibration of viewing reference frames. In that case the quality of the measurement will still depend on the accuracy of the ground control points, their number and their distribution over the scene.

The method used relies on the use of ground control points by estimating the location of these points on the ground, and is illustrated in the Figure 3 below.
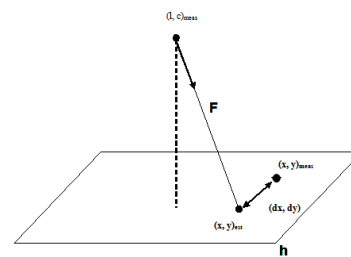

Figure 3. Illustration of the method for estimating location performance in metres.

For a given scene:

• identification of an approximately large number of ground control points in the scene:

(l_i, c_i) $(x_i, y_i, h_i)_{meas}$ where $i=(1, m)$;

• for each measured ground control point i, estimation of the location of this point on the ground $(x_i, y_i)_{est}$ using the location function ;

• for each ground control point i, comparison of $(x_i, y_i)_{est}$ estimated terrain coordinates, with measured terrain coordinates $(x_i, y_i)_{meas}$, which gives the location performance at this ground control point (dx_i, dy_i);

• average location performance of the scene is then obtained by averaging the performance of the all ground control points in this scene.

The estimation of performance is then directly in metres and accounts for all contributing factors.

The most important constraint of this method in view of a global validation is that it requires a lot of GCP as reference images worldwide. Those GCP should be independent from those used for the GRI construction by IGN. Pleiades High Resolution DataBase (BDPHR) including more than 500 images with accuracy less than 5 m is used by CNES for the GRI geolocation performance validation. BDPHR accuracy is estimated at 5m CE95. This performance is measured with the

help of around 15 sites over the World (Figure 4) with 25 GCP for each site very-well geolocalised: planimetric precision between 0,1m and 3m. Few of these points are also used for GRI generation but GRI generation uses a lot of other GCP.

Figure 4. Position of sites used to measure BDPHR geolocation performance.

Pleiades images are then resampled to Sentinel-2 resolution (10m) and degraded with Sentinel-2 MTF (Modulation Transfer Function) to become a reference for geolocation performance assessment of Sentinel-2 GRI. This geolocation performance is estimated by colocation process described just before with the help of a correlation with all BDPHR images intersecting each GRI datastrips, and then spatio-triangulation with homologous points found on all these images. A mean geolocation performance for each GRI datastrips is then computed in meters on ground, and also a global geolocation performance for each GRI block (CE95 of all the mean performances of all datastrips in a same GRI block). At the moment of paper writing, North America block and isolated islands are available but not yet validated by CNES. Even if each GRI products are not necessary in intersection with BDPHR references, it is not worrying as spatio-triangulation ensures a geolocation consistency between neighbor's products.
At the moment of paper writing, North America block and isolated islands are available but not yet validated by CNES.

Figure 5. BDPHR footprints over Europe (in green).

2.3 Europe GRI Validation

Europe GRI block was built during the S2A commissioning phase, L1B segments being selected par IQ geometric team at CNES, and then delivered to IGN for refining, come back to CNES for validation of geolocation performance. As it was the first GRI block, used also to validate the multi-temporal performance of refined S2 products, a lot of BDPHR GCP was provided over Europe to validate the complete process before building other GRI blocks (Figure 5). Figure 6 represents the number of GCP (red bullets) per products as a level of confidence that can be given to the computed geolocation performance of Europe GRI block. Only 5 products over 65 couldn't be validated because of lack of BDPHR references.

Figure 7 presents the mean geolocation performance obtained for each Europe GRI products. The global mean performance reached on the whole block is 8m CE95, which is compliant with the performance required. Red products bad geolocation performance is explained because products are cloudy and not linked to other products on the GRI. Red and white products are to be produced again to complete the Europe GRI block.
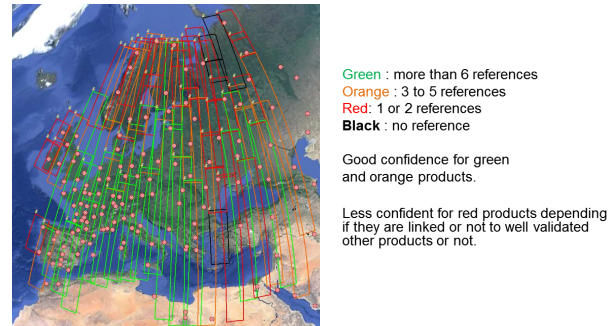

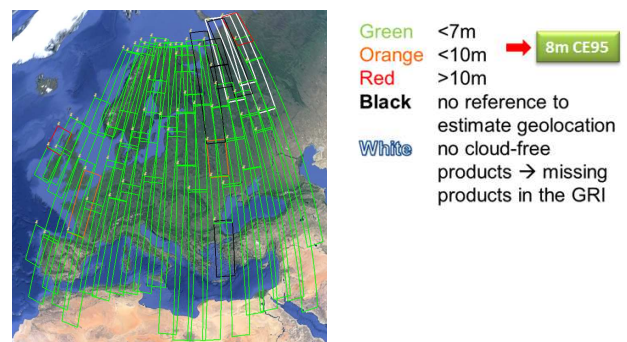Figure 6. BDPHR number per European GRI products.

Figure 7. Europe GRI mean geolocation performance.

2.4 Australia GRI Validation

Figure 8 represents the validation level of confidence versus the number of GCP per products. As there is less references than for European block, more products cannot be validated (black ones) or with a less confident validation level (red ones). 29 datastrips over 60 have been checked. Figure 9 illustrates the mean geolocation performance per products, the global mean geolocation performance for the whole Australian block being assessed to 8.3m CE95 (Table 1). Two defective products have to be produced again on orbits 30 and 131).

Average mean differences	5.1m
Standard deviation of the mean differences	1.8m
Minimal gap (product best localized)	1.7m
Max Deviation (product worst localized)	9.4m
CE68 of estimated values at a lower location	6.0m
CE95 of estimated values at a lower location	**8.3m**

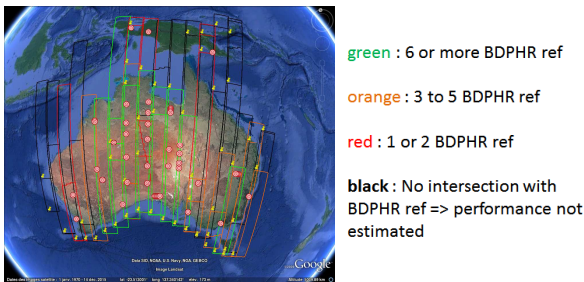Table 1. Geolocation performances statistics on Australian GRI.

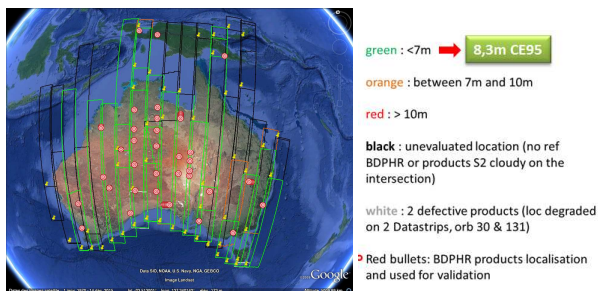Figure 8. BDPHR number per Australian GRI products.

Figure 9. Austrialia GRI mean geolocation performance.

2.5 North Africa and Middle East GRI Validation

Same statistics are presented in this section. The global mean geolocation performance for this block is estimated at 9.1m CE95. 58 datastrips over 86 have been checked. Only one product is estimated at a bad performance of 15m, but this product has only one BDPHR reference so the result is not reliable.
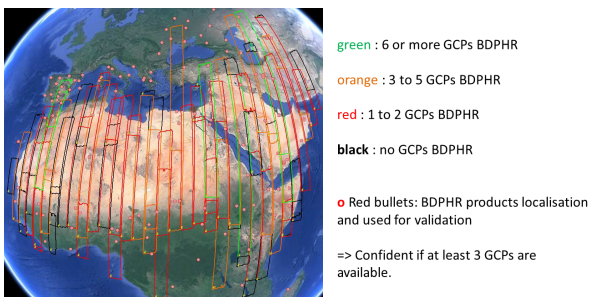

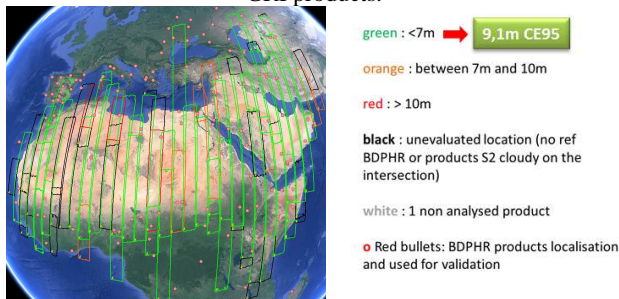Figure 10. BDPHR number per North Africa and Middle East GRI products.

Figure 11. North Africa and Middle East GRI mean geolocation performance.

Average mean differences	5.9m
Standard deviation of the mean differences	2.0m
Minimal gap (product best localized)	2.5m
Max Deviation (products worst localized)	15.1m
CE68 of estimated values at a lower location	6.2m
CE95 of estimated values at a lower location	**9.1m**

Table 2. Geolocation performances statistics on the North Africa and Middle East GRI block.

2.6 South Africa GRI Validation

For this block, 44 datastrips are intersecting BDPHR references over 91 and then can be validated with this method. The global mean geolocation performance for this block is estimated at 7.5m CE95. Only one product is estimated at a bad performance of more than 11m, but this product has only one BDPHR reference and a difficult correlation, so the result is not very reliable.

Average mean differences	5.2m
Standard deviation of the mean differences	1.7m
Minimal gap (product best localized)	1.3m
Max Deviation (product worst localized)	11.1m
CE68 of estimated values at a lower location	7.5m
CE95 of estimated values at a lower location	**7.5m**

Table 3. Geolocation performances statistics on the South Africa GRI block.

Figure 12. BDPHR number per South Africa GRI products.
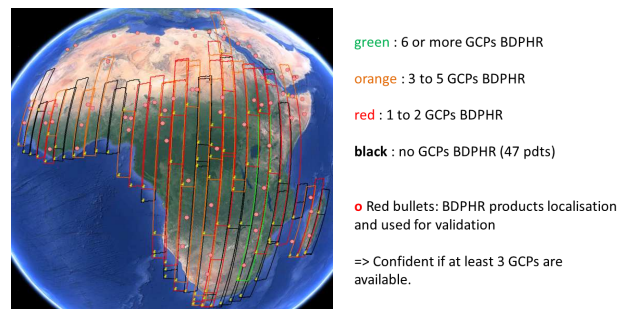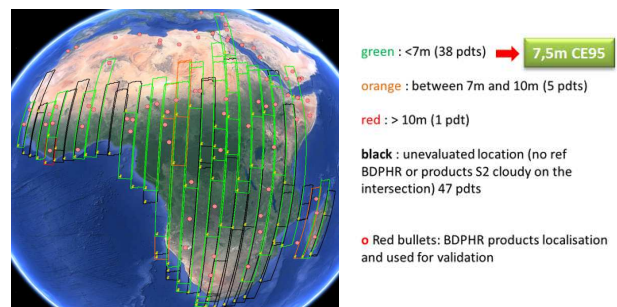

Figure 13. South Africa GRI mean geolocation performance.

2.7 Asia GRI Validation

Asia GRI is a very big block, and it was difficult to find good BDPHR GCP in some regions of this block and its Islands. So only 74 products over the 163 are intersecting at least one BDPHR reference. The global mean geolocation performance for this block is estimated at 9.4 m CE95.

Average mean differences	5.6m
Standard deviation of the mean differences	1.5m
Minimal gap (product best localized)	2.4m
Max Deviation (product worst localized)	11.2m
CE68 of estimated values at a lower location	8.6m
CE95 of estimated values at a lower location	**9.4m**

Table 4. Geolocation performances statistics on the Asia GRI.

green : 6 or more GCPs BDPHR

orange : 3 to 5 GCPs BDPHR

red : 1 to 2 GCPs BDPHR

black : no GCPs BDPHR (47 pdts)

o Red bullets: BDPHR products localisation and used for validation

=> Confident if at least 3 GCPs are available.
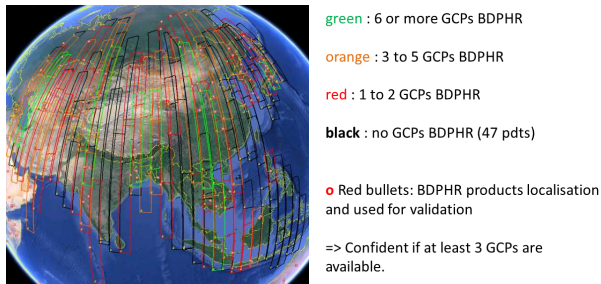
Figure 14. BDPHR number per Asia GRI products.

green : <7m (64 pdts)   9,4m CE95

orange : between 7m and 10m (10 pdts)

red : > 10m (1 pdt: 11,23m/2BDPHR)

black : unevaluated location (no ref BDPHR or products S2 cloudy on the intersection) 88 pdts

Figure 15. Asia GRI mean geolocation performance.

Two products with a quite high CE95 performance that cannot be well validated either by CNES or IGN. The South East part of the GRI (mainly over Philippines and Indonesia) is hard to check (by both CNES and IGN) because of a lack of references. Refined parameters analysis reported a datatake with a high value on yaw. But the geolocation performance of this product cannot be evaluated as it doesn't intersect any references.

2.8 South America GRI Validation

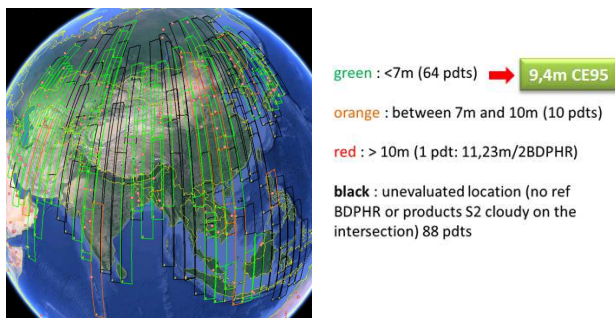| Average mean differences	5.3m
Standard deviation of the mean differences	1.4m
Minimal gap (product best localized)	2.6m
Max Deviation (product worst localized)	9.7m
CE68 of estimated values at a lower location	7.1m
CE95 of estimated values at a lower location	**7.5m**

Table 5. Geolocation performances statistics on the South America GRI.

50 products of 141 can be validated with the BDPHR references. The global mean geolocation performance for this block is estimated at 7.5 m CE. It can be noticed that only 19 among 50 datastrips intersect 3 or more BDPHR references.
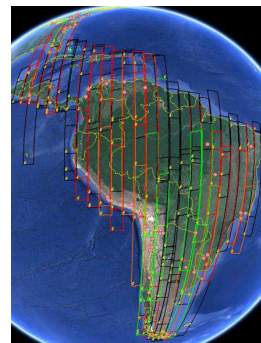

green : 6 or more GCPs BDPHR (9 pdts)

orange : 3 to 5 GCPs BDPHR (10 pdts)

red : 1 to 2 GCPs BDPHR (31 pdts)

black : no GCPs BDPHR (91 pdts)

o Red bullets: BDPHR products localisation and used for validation

=> Confident if at least 3 GCPs are available.

Figure 16. BDPHR number per South America GRI products.

green : <7m (46 pdts)   7,5m CE95

orange : between 7m and 10m (4 pdts)

red : > 10m (0 pdt)

black : unevaluated location (no ref BDPHR or products S2 cloudy on the intersection) 91 pdts
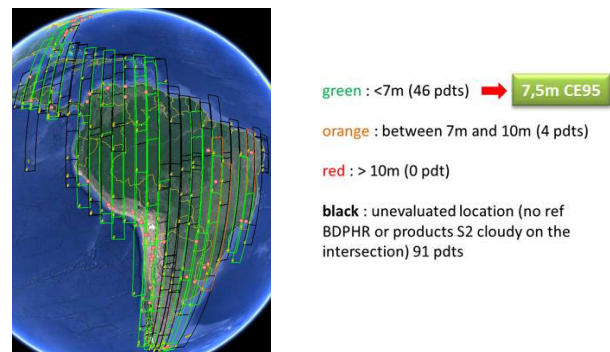
Figure 17. South america GRI mean geolocation performance.

2.9 Conclusion on GRI Validation

On the GRI blocks validated, the geolocation performance is consistent with the requirement of 9.5m CE95. North America block is still to be validated, as well as isolated islands even if it will be difficult on islands because of lack of PHR references. But this is not worrying as spatio-triangulation ensures a geolocation consistency between neighbor's products.

3. MULTI-TEMPORAL REGISTRATION

3.1 Introduction

The GRI permits to fulfil the requirement on multi-temporal registration (Figure 18). Indeed, as Sentinel-2 has a high revisit (5 days with 2 satellites) under the same viewing conditions, images registration from different dates over the same orbit can be performed by systematically refining images over a reference. This reference is then the GRI. This multi-temporal registration has been checked on ortho-rectified and refined L1C tiles. These performances are measured by correlation, without resampling, since it is done in a given projection.

These performances have to be checked only in products acquired from the same orbit in the cycle. Products acquired on the same site from neighbour orbits, therefore with different viewing angles, cannot meet this requirements, since the system DEM error is higher than 0.3 SSD CE95 : assuming a DEM with 16m of elevation error CE95, 16m*tan(10°)*2=5.6m of error. Nevertheless, these performances have been also estimated on a set of adjacent orbits for user's information.

As the number of products was not very high at the time of S2A commissioning, the assessment of this registration is not statistical and so not fully relevant. It will be better estimated with adding S2B images. It is also difficult to work on a large number of products because of computing time and good weather conditions (cloudy free) or seasonal effects (snow,

cultures…) on the same area during several months. B4 and B11 bands are the reference bands for this analysis.

Correlation and refining parameters have been tuned to comply with features of Sentinel-2 products (cloud free, with a lot of water, cloudy, very cloudy and desert). That has defined the geometrical model refining using tie points between Sentinel-2A images and GRI, and space triangulation.
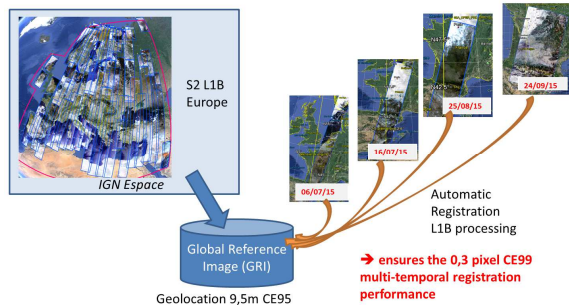

Figure 18. Refining and multi-temporal process with GRI.

3.2 Difference between refined and non-refined products

In order to evaluate the improvement of multi-temporal registration when using GRI, a comparison on some products was done. The same products are used, once non-refined and once refined to compare the performance. This study was done on 2 long products on 2 orbits, for 6 different dates and for the same relative orbit in the cycle. Results obtained are presented in Table 6, they are fully demonstrative.

B4 band	Multi-temporal registration performance (pixels)	
	Non-refined products	Refined products
Maximum value of mean values of all tiles	1.47	0.21
CE95 value of mean values of all tiles	1.40	0.13

Table 6. Statistics on multi-temporal performances for L1C products: comparison for refined and non-refined products.

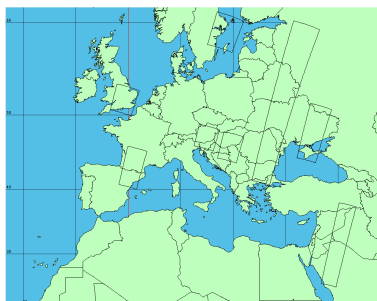### 3.3 Refined products on same relative orbit

Figure 19. Footprints of refined products analyzed for multi-temporal registration assessments.

The map in Figure 19 presents the areas covered by the refined L1C products used to check the multi-temporal registration for product performance. Products are distributed over Europe in the GRI zone. The requirements are verified on products from the same orbit. The study has been performed on 8 different orbits: Orbit 51-6 dates (France), Orbit 50-6 dates (East Europe), Orbit 35-4 dates (Middle East), Orbit 137-2 dates (UK), Orbit 122-2 dates (Austria), Orbit 36-3 dates (Hungary), Orbit 22-4 dates (Sweden) and Orbit 64-2 dates (Ukraine).

In order to have the most relevant results, tiles with a number of valid points for correlation (between a reference tile and the measured tile) under 60% of the maximal number of points of a half-tile, have been filtered. This way, for the B4 band, products over UK (orbit 137) and Sweden (orbit 22) don't participate to the statistics. So, a total of 242 tiles are taken into account for B4 band, and a total of 377 tiles for B11 band. Statistics on all tiles are detailed in Table 7 and the CE95 value on all tiles is considered as the multi-temporal performance achieved.

Multi-temporal performance in pixels	B4	B11
Mean value of mean values of all tiles	0,085	0,081
Min value of mean values of all tiles	0,020	0,015
Max value of mean values of all tiles	0,360	0,260
CE95 value of mean values of all tiles	0,220	0,170

Table 7. Statistics on multi-temporal performances for refined L1C products on the same relative orbit.

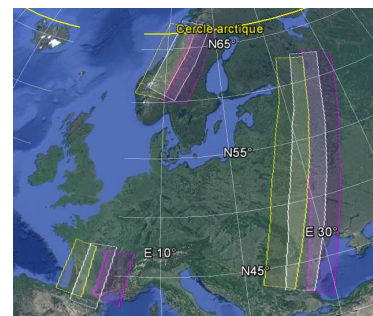### 3.4 Difference between same and adjacent relative orbits

Figure 20. Footprints of products analyzed for multi-temporal analysis on adjacent orbits.

Multi-temporal performances in pixels On B4 band	Same orbit	Adjacent orbits	
	Refined	refined	non refined
Mean value of mean values of all tiles	0,085	0,29	0,47
Max value of mean values of all tiles	0,36	0,80	1,26
CE95 value of mean values of all tiles	0,22	0,57	0,92

Table 8. Statistics on multi-temporal performances for L1C products on same or adjacent relative orbits.

For user's information, an estimation of the multi-temporal registration performance between adjacent orbits has been performed. Table 8 compare the results obtained with the performance achieved on the same orbit. The analysis has been done on 3 different sites over Europe, for refined (400 tiles) and non-refined (230 tiles) L1C products (Figure 20). Of course, this performance depends on the relief and DSM accuracy as already explained in introduction at section 3.1.

4. HIGH LATITUDE DIGITAL SURFACE MODEL

One of the key aspects of the Sentinel-2 mission is its high revisit rate, which allows creating cloud-free time-series. In the case of Digital Surface Models (DSM) generation, this ability can be used for different purposes:
- Increasing completeness by reducing areas masked by clouds,
- Increasing accuracy by averaging measures in areas of little elevation changes,
- Monitoring terrain changes such as ice and snow levels by creating DSM time-series.

4.1 Required product level and characteristics

The stereo-restitution process has two main steps: first, pixels are automatically matched between images using some disparity estimation method, and then the lines of sight generated by the matched pixels are triangulated. Both steps influence what product should be used and how they should be selected.

The former step implies that images should have been acquired in a short range of time, so as to minimize landscape changes due to seasonal changes for instance. The images should be as similar as possible so as to get the best performances from the matching algorithm. Also, cloud coverage should be kept low, even if it is always possible to perform matching in cloud-free areas of the images.

The latter implies that images are in the acquisition geometry (i.e. not projected to ground), which means that the DSM can only be generated from L1B products (sensor geometry). Those products can be generated by CNES in the frame of its cooperation with ESA on the Sentinel-2 mission. Also, since the triangulation process is very sensitive to attitude angular error, sensor modelling of those images should be refined using the GRI dataset. Each L1B product is pre-processed so as to stitch all sub-swath (detectors) into a single image and to compensate residual oscillations. The corrected, stitched perfect sensor products can then be fed into the DSM generation tool.

4.2 Data selection and coverage assessment

In order to analyse coverage opportunities and automatically select qualified data for a given area, we developed a small tool to parse the entire Sentinel-2 archive and return the product set given an area to produce, the required period of the year, the expiration time of data and the maximum cloud cover.

As shown in figure 21(a-b), if we consider the entire European area, this tool allows us to determine for a given date what fraction of the area we can actually produce given the current available archive.

For instance, for a DSM in august with one month expiration time and cloud coverage lower than 10 %, we can see that around half of the theoretical area can be produced with available data. For the same area in spring, this fraction falls down to only 20 %.

In the frame of this work, we focused on a 4 squared degrees area over Sweden, for which reference data were available and fraction of DSM that can be produced given current archive is almost 100 %. Figure 21(c-d) shows the coverage available for both dates in this area. For the summer date, the tool selected 7 products, whereas only 6 products where selected for the spring date. Each set of products comes from 4 different orbits.

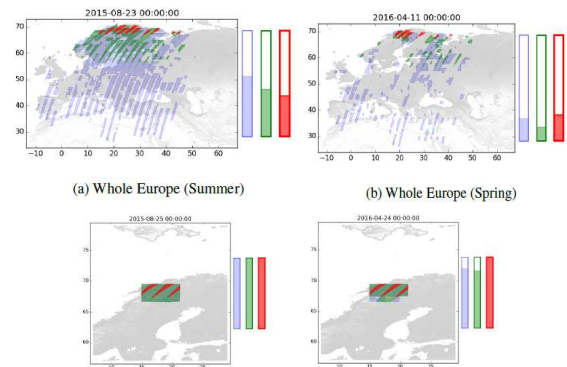(a) Whole Europe (Summer)   (b) Whole Europe (Spring)

Figure 21. Area that can actually be produced given current archive for two dates (one in summer and one in spring), with a time frame of one month and cloud cover less than 10 %, over the whole GRI area, and the Sweden test site. Blue areas represent overlap between adjacent orbits, while green (resp. red) areas represent overlap between 1 orbit (resp. 2 orbits). Side bars represent the fraction of producible areas for each orbits spacing.

4.3 Expected accuracy

It is noteworthy that the baseline ratio (i.e. stereoscopic angle between two different orbits) will vary depending on orbits pairs and latitude. These variations are important because they relate directly to the accuracy that can be expected. Provided that pixel size is 10 meters and that most disparity estimation methods cannot reach accuracy better than 0.1 pixels, a baseline ratio of 0.1 will lead to DSM vertical accuracy of only 10 meters. Increasing this accuracy to 1 meter requires both to increase the disparity measurement accuracy and to increase the baseline ratio. The most visible effect of varying baseline ratio is quantization or noise dependant on disparity estimation algorithm. As many combinations of swath and orbits are needed to produce a given area, the quantization and accuracy will vary across the spatial extent of the generated DSM. Nevertheless, accuracy can be estimated for each DSM value by analysing orbits and swath that were combined to derive it.

4.4 DSM generation

DSM generation has been carried out with the open-source tool S2P developed by CMLA and CNES (De Franchis, 2014 and Michel, 2015). This tool uses local stereo-rectification to increase matching performance and is built with a modular architecture that allows to change the matching (or disparity estimation) algorithm. In this work, matching has been carried out by the in-house correlation tool Medicis (Cournet, 2016), which seems the most appropriate given the observed landscape and allows efficiently filtering bad matches over shadows, cloud or watering areas. Some modifications were made by CNES to the S2P software. The major one concerns the use of the viewing directions intersection method (Delvit, 2006), instead of the height maps fusion method. Other modifications were made to accommodate some corner cases and the specificity of this work. Also notes that S2P generates 3D point clouds and that a raster DSM can then be derived from those point clouds.
S2P uses a reference image and processes N secondary images. Reference image is tiled and for each tile, stereo-rectification and disparity estimation is performed against each of the secondary image. Heights are then estimated either by fusing estimation from each pair or by triangulating lines of sights for

all images (for which a match is found) at once. The latter has been used in the frame of this study. In order to obtain the full coverage of the target area, each of the selected images is alternatively used as the reference image, and the final raster DSM uses all generated point clouds.

In order to enhance DSM quality and remove outliers, two post-processing steps are applied. The first step occurs right after disparity has been estimated. Validity flags from the Medicis tool are used to filter out bad matches corresponding to shadows, clouds or water. Left-right consistency is also used to this end. The second post-processing step occurs once point clouds have been generated. A statistical filter is applied to filter outliers based on the mean distance to the K nearest neighbours of each point. The PCL open-source software has been used to this end (Radu Bogdan Rusu, 2011).

For each image used as reference image, a set of points clouds is generated. Those 3D points clouds can then be fused to increase completeness of the final DSM. Because of clouds, shadows or water areas, it might be interesting to use several images from the same orbit for this purpose. Fusion is done at point cloud level, i.e. before projecting the cloud to the raster grid.

4.5 Sentinel-2 DSM test areas

Figure 22 shows on the left side the summer and spring DSM generated over the Sweden test area, with the same colour scale. One can note that the spring DSM has more missing data due to the higher cloud cover and the lower number of images that could be selected for the DSM generation. Nevertheless, one can also note that some river areas that are missing in the summer DSM are actually completely reconstructed in spring. This is probably because those rivers are free water in summer but ice and snow in spring.

To demonstrate our almost fully automated process, we also generated a second test site over the Svalbard Island (Figure 22 on right side), with images from 5 different orbits. The only manual step in our workflow is the ordering of selected L1B products. Other steps, including product selection, pre-processing, DSM generation and post processing are fully automated.

(a) Summer DSM   (b) Spring DSM

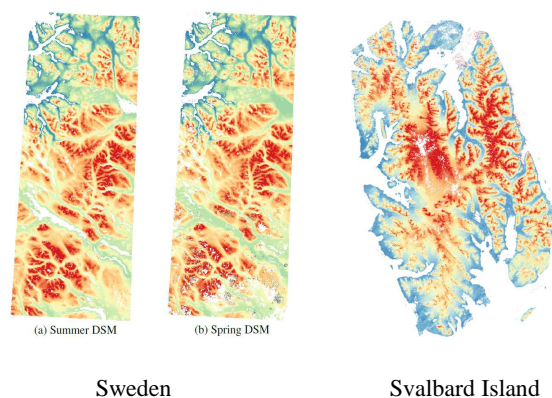Sweden   Svalbard Island

Figure 22. DSM generated over the test areas in Sweden (2 left) and Svalbard Island (right). Values range from 0m (blue) to 1500m (left red) and 1000m (right red). DSM Resolution is 10 meters.

5. CONCLUSION

The main objective of the Sentinel-2 MSI mission is to provide stable time series of images at medium-high spatial resolution.

After nearly two years in orbit, Sentinel-2 time series are consistent and comparable with other missions such as Landsat-8 OLI. The geometric accuracy is very satisfactory, the typical multi-temporal co-registration between two acquisitions is better than one pixel. CNES has shown that the full performance (better than 0.3 pixel CE95 confidence level) can be reached with the activation of the geometric refinement based on the GRI, as the geolocation performance of this GRI is estimated better than 9.5m CE95 all over the world.

In this work, we then demonstrated that the Sentinel-2 data can be used for DSM generation across tracks. For northern latitudes between 60° and 84°, the overlap between swaths yields a complete coverage of Earth, and some interest site can also be available for lower latitudes. With respect to other elevation sources for those latitudes, Sentinel-2 DSM would benefit from the mission high revisit rate, allowing producing several DSMs per year in order to increase accuracy or monitor rapidly changing surfaces such as ice and snow. In this work, we focused on a test area of 4 squared degrees but it could be easily scaled to the entire GRI area over whole lands between 60° and 84° of northern latitudes. Accuracy assessment is still in progress and will be presented at the conference.

REFERENCES

Cournet, M., 2016. 2D sub-pixel disparity measurement using qpec/medicis. In:*International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 41.

Dechoz, C., 2015. Sentinel-2 Global Reference Image. In: *SPIE Remote Sensing, International Society for Optics and Photonics*, pp. 96430A–96430A.

De Franchis, C., 2014. An automatic and modular stereo pipeline for pushbroom images. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, no. 3, pp. 49.

Delvit J.M., 2006. The geometric supersite of Salon de Provence. In: *ISPRS Congress Paris.*

Drusch, M., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. In: *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.

Gascon, F., 2016. Copernicus Sentinel-2 Calibration and Products Validation Status.
Preprint doi:10.20944/preprints201610.0078.v1.

Languille, F., 2016. Sentinel-2A image quality commissioning phase final results: geometric calibration and performances. In: *SPIE Remote Sensing. International Society for Optics and Photonics*, pp. 100000F–100000F.

Michel, J., 2015. S2p: a new open-source stereo pipeline for satellite images. In: *Free and Open Source Software for Geospatial (Europe).*

Michel, J., 2017. High latitude multi-temporal DSM from Sentinel-2 data: a proof of concept. IN: *IGARSS.*

Radu Bogdan Rusu, 2011. 3D is here: Point cloud library (pcl). In: *Robotics and Automation (ICRA).*