S2check: a tool for quality control of sentinel-2 products

Sébastien Clerc, Christophe Bévy, Jan Jackson, Theodora Papadopoulou, and Alejandro Garcia-Soto

Department of Operational Services, ACRI-ST, Sophia Antipolis, France; ARGANS, Plymouth, UK; Ground Segment Systems Department, DEIMOS, Tres Cantos, Spain

ABSTRACT
With more than 4 TB of data acquired every day, the Sentinel 2 mission poses new challenges for quality control (QC). To support the QC operations of Sentinel 2, a dedicated tool has been developed for improved efficiency and traceability of analyses. This tool aims at helping visual inspection by human operators through automatic product retrieval, generation of quicklook images and production of quality indicators.

The article describes the product sampling strategy used to select products suitable for quality control checks. Then we present the methodology used to generate quicklooks, compute relative and multi-spectral co-registration performance. The final step is the production of end-to-end service performance indicators.

Lessons learnt from the first year of operation of this tool are presented regarding detected product anomalies and measured performance trends.

ARTICLE HISTORY
Received 30 July 2017
Revised 3 June 2018
Accepted 4 June 2018

KEYWORDS
Quality control; Sentinel-2; co-registration

Introduction: quality control activities for sentinel-2

The Copernicus program is introducing a revolution in Earth Observation by providing massive amounts of free and open data. The Sentinel-2 mission in particular acquires multi-spectral images at medium resolution with a revisit time of 5–10 days worldwide (Gascon et al., 2017). Two satellites are operating on the same orbit with a descending node at 10:30 AM local time. Images are acquired at 10 m resolution in the visible (blue, green and red bands B02, B03 and B04 respectively) and the near infra-red (B08). Narrow spectral bands at 20 m resolution in the red-edge domain (bands B05, B06, B07 and B8A) are especially useful to monitor vegetation, while SWIR bands (B11 at 1.6 µm and B12 at 2.1 µm) support a better characterization of surfaces (in particular for cloud vs ice discrimination). Additional bands at 60 m resolution are more focused on the characterization of the atmosphere: B01 for coastal aerosols, B09 for water vapour, and B10 for cirrus. More details can be found in the aforementioned reference article.

Performing quality control inspections on the 4 TB of data produced by the mission every day is a challenge which requires some specific solutions. Moreover, timeliness performance is an important goal of the Copernicus programme, which limits the possibility for a-priori quality checks before dissemination.

The initial quality control strategy relied on two approaches:

- An automatic On-Line Quality Control performing systematic elementary checks (number and size of files in the products, threshold and ranges on metadata values, analysis of satellite telemetry,…). These checks are performed before dissemination of products.
- Human inspections on sample products using visualization and analysis tools. These checks are performed after dissemination of products.

However the automatic On-Line Quality Control fails to detect most anomalies, while the human inspections are limited to one or two products per day and can therefore miss non-systematic anomalies. In the first months of the missions, the need for more thorough semi-automatic checks on sample products was identified. The S2check tool was developed with the objective to perform tests requiring a few minutes of processing on a dozen of products per day, or approximately one product per orbital period. As most anomalies affect all products of a given orbit, this sampling rate ensures a high probability of detection.

Automatic procedures for quality control are relatively standard in the domain of satellite data processing, see, for example (Han & Jochum, 2016, July). The originality of our approach is the combination of automatic analytics and visualization tools to support
human visual inspection. A specific random sampling strategy has been designed to cope with the amount of data while ensuring a high probability of detection of anomalies.

A semi-automatic quality control tool

More precisely, the overhead for human inspection can be reduced by automatizing some of the tasks:

- Selection of sample products suitable for quality control and download of selected products
- Visualization of the images for all spectral bands
- Assessment of geometric relative and multi-spectral co-registration performance
- Assessment of radiometric stability

Human visual inspection is still needed to detect some anomalies and to clear some false alarms. However the inspection time is now much faster because there is no need to select and download a product, and then to open each spectral band in a viewer. The time is decreased to 1 min, versus 1 h previously. Thanks to the tool it becomes now possible to reach the objective of inspecting one product per orbit approximately.

An e-mail containing quicklooks of the different spectral bands as well as performance metrics is sent to the operators every 2 h. The mail provides some product metadata (sensing date, orbit number, platform reference…) and a geographic context (position on a world map as well as the name of the location at the centre of the image) in order to help image analysis. An additional benefit of the S2check tool comes from the recording of all inspections in a database. This improves traceability and robustness of the quality control process, and allows to compute history of performance metrics.

S2check methodology

Processing chain overview

Figure 1 presents a synthetic view of the S2check process. The S2check tool is implemented as a ‘cron’ job which starts a new inspection every 2 h. This periodicity allows inspecting approximately one product per orbit for one satellite. Note that the inspections are performed after dissemination. S2check is implemented as a set of python scripts, using common python libraries (requests, numpy and matplotlib).

S2check uses the Sentinel Data Hub API (Application Programming Interface, see Martin & Knowelden, 2015) to select a tile recently ingested in the archive which is suitable for inspection, then a previous acquisition of the same tile to be used as a reference.

S2checks generate quicklooks of the image for a fast human visual inspection, then some automatic checks are performed:

- Relative registration performance between the current and reference image for the reference spectral band
- Multi-spectral registration performance between two spectral bands of the current product
- Average radiometric gain and offset between the current and reference products for R,G,B bands.

The performances are compared to reference thresholds to help in the identification of potential anomalies.

Once all the analyses have been completed, an e-mail including the quicklooks and a performance summary table is sent to the service operators. Analyses are also recorded in a database, and can be accessed through a web page. The web page also provides standardize plots of the mission performance history.

Figure 1. Overview of the S2check process.
**Sampling strategy**

The first step is to obtain from the Sentinel Data Hub archive the list of all products ingested since the last inspection, using the archive API. Some filters are then applied to select products which are suitable for quality control:

- Filtering on cloud cover percentage (<20% currently)
- Filtering on data coverage (less than >50% of 'No Data' pixels): this filter uses the size of the product as a proxy for data coverage.
- Filtering on latitude to avoid snow or ice covered tiles
- Filtering with a land/water mask to avoid tiles located completely over the sea

The first product found matching these criteria is selected for inspection. This ensures in practice a random sampling in the sense that the probability of choosing a particular sense depends only on the local cloud statistics. S2check also retrieves a reference acquisition of the same scene from the archive for comparison. This is done by requesting the list of previous acquisitions of the same tile. The current period used to select the reference product is from now – 8 months to now – 1 month. The list is ordered by increasing age so that more recent products are used whenever possible. This minimizes the impact of image changes due to seasonal effects (on vegetation and snow cover in particular). However, too recent products are avoided. The one-month lag allows to detect small systematic changes through short-term statistics (linear regression on performance indicators). Filters are again applied at this step to eliminate tiles not suitable for analysis (filter on cloud and data coverage).

Geographic maps of inspected products are produced to check for potential sampling biases. The current map reflects the expected effect of cloud coverage and shows no other systematic bias, see Figure 2.

**Quicklook generation**

For a fast visual inspection, quicklooks for all spectral bands (at 400 pixels) and an RGB quicklook (1000 pixels, see Figure 8) are generated. This resolution is sufficient to identify major image anomalies and remains compatible with a conventional e-mail size. The main issue for this step is to deal with the high variability of the reflectance, especially for SWIR bands. Saturation levels are fixed to 0.5 for band B01, 0.1 for band 10 (cirrus band), and 0.3 for other bands. Automatic contrast adjustment is not used in order to be able to detect radiometric anomalies. For example, a processing anomaly in fall 2016 led to generate products with an incorrect quantification value of 1000 (instead of 10,000). The quicklooks generated by S2check were extremely dark, which led to a quick identification of the anomaly. If a contrast adjustment had been used, the images would have looked normal.

For many images on the other hand, the contrast is not strong enough to support visual inspection. This is true in particular for deserts (high radiometry for SWIR bands) and snow covered areas (high radiometry for VISNIR bands).

**Co-registration performance**

An important added value of the S2check to is to provide automatic measurement of the relative and multi-spectral co-registration performance. The co-
registration assessment is based on the optical flow method (Lucas-Kanade algorithm, Lucas and Kanade (1981). This method is fast but limited to small shift vectors (less than 2 pixels typically). If the shift is larger the method will fail, which can indicate that the image is affected by an anomaly.

The method uses a large number of small imagettes randomly extracted from the 2 images (old and new), denoted I₁ and I₂. First a contrast adjustment is performed:

\[ I_2 \rightarrow \alpha I_2 + \beta, \]  (1)

Where the contrast adjustment gains and offsets \( \alpha \) and \( \beta \) are computed from:

\[
\begin{align*}
\alpha &= \text{std}(I_1)/\text{std}(I_2) \\
\beta &= \langle I_1 \rangle - \alpha \langle I_2 \rangle
\end{align*}
\]  (2)

The contrast adjustment allows to get rid of radiometric variations due to atmospheric effects, but also to seasonal effects (vegetation change, snow) in some cases.

After contrast adjustment, the second image \( I_2 \) is equal to the first image \( I_1 \) shifted by a small displacement vector \( u \):

\[ I_2(x) = I_1(x + u) \]  (3)

If \( u \) is small enough one can linearize:

\[ I_2 \approx I_1 + \nabla I_1 \cdot u, \]  (4)

or, (with second-order accuracy):

\[ I_2 - (\nabla I_2 \cdot u)/2 \approx I_1 + (\nabla I_1 \cdot u)/2. \]  (5)

Image gradients are computed using finite differences.

The Lucas-Kanade method consists in evaluating \( u \) from this method using a least-square method:

\[ \min_u \|(\nabla I_1 + \nabla I_2)/2 \cdot u - (I_2 - I_1)\|. \]  (6)

For each imagette, a shift vector \( u \) can be computed. To obtain meaningful results, a filtering must be done to select only good features:

- Filtering on the value of the smallest eigenvalue of the matrix of the least square problem \((\nabla I_1 + \nabla I_2)/2\). This filter removes imagettes without enough texture
- Filtering on the quality of the match between the two imagettes: first image \( I_1 \) and second image after shift correction \( \bar{I}_2 = I_2 - (\nabla I_1 + \nabla I_2)/2 \cdot u \). Pearson’s correlation coefficient is used to select good matches (homologous features).

The correlation filter removes the main sources of variability of the image which could affect the shift estimation (cloud and cloud shadows, vegetation change, white caps on shoreline…), as illustrated in Figure 3.

The co-registration workflow starts with the computation of the relative co-registration between the current and reference images is computed. B04 is used as reference spectral band. A total of 5000 imagettes are extracted randomly in the image, but only a small fraction is retained after filtering for low texture imagettes and poor matches.

The multi-spectral co-registration between bands B02 and B04 is then computed for the current product. The computation uses the ‘good’ features retained at the previous step. This ensures that co-registration measurements will not be affected by clouds or sun-glint effects.

Co-registration performance indicators and compared to threshold values determined from statistics (3 sigma values). When the thresholds are exceeded, a warning message is produced to draw the attention of operators and trigger more detailed analyses. In most cases the products are correct and the warning message can be explained by another reason. On the other hand, we verified a posteriori that warning messages are issued on all products affected with a previously identified anomaly (no missed detection).

Table 1 below lists the warning messages and their possible origin.

To further help interpretation of the temporal co-registration assessment, a composite RGB image is created by overlaying a feature from the current

---

Figure 3. Imagette selection algorithm. Red boxes: imagettes rejected by the correlation criteria. Gray boxes: imagettes rejected by the texture content criteria. Green boxes: selected homologous imagettes. The algorithm is able to retrieve homologous features even when the scenes look visually dissimilar.
image over a background from the reference image (see Figure 4). The composite image can in particular quickly reveal the presence of snow on the reference image, or of strong changes of in the vegetation cover. The composite image sometimes reveals an anomaly on the reference product: spectral mis-registration (Figure 4 last example), dark product, etc.

### Radiometric stability
More recently, a radiometric stability check has been introduced in the S2check processing. This check computes the average radiometric gains and offsets between the reference and current image, considering all homologous features identified during the relative co-registration step. The computation is done for bands B02, B03 and B04. Gains and offset between two top-of-atmosphere images can be relatively large as a result of atmospheric effects, but the general trend should remain constant. A systematic discrepancy for one spectral band can indicate an anomaly on the radiometry of this band.

### End-to-end performance metrics
All inspections are recorded in a database, which allows to monitor trends on the mission performances. A synthetic timeline of multi-spectral and relative co-registration performances is generated (see Figure 5). On this plot, the size of the circle is proportional to the standard deviation of shift measurements among the various features. A large standard deviation can signal an anomaly affecting a part of the image. Red stars indicate failed inspections (no matching points have been found between the reference and current image). This is generally the result of high cloud or snow coverage, low land or data coverage, or a combination of these factors.

The inspection database can also be used to produce end-to-end mission performance plots for availability (Figure 6) and timeliness (Figure 7). The timeliness plot details the contribution of the space segment, ground segment and user dissemination segment. The availability plot reports outages of the Copernicus data hub (red flags) and ingestion or production failure (orange flags). Yellow flags correspond to the case where no reference product has been found for comparison. This happens generally when sampling a very cloudy area where few cloud-free products are available. Note that the S2check service itself may have outages (internet access failure for instance) in which case no flag is produced.

### Experience after 1 year of operation
The S2check tool started operation on 5 July 2016. Since then a number of anomalies have been identified thanks to this tool. They are listed in Table 2. Depending on the situation, the following remediation actions are implemented:

- In case or recoverable anomaly, a modification of the processing chain is implemented. The correction is verified by reprocessing the anomalous product. Once verified, the new processing chain is transferred to operation for Near-Real...
Figure 5. Multi-spectral (top) and multi-temporal co-registration (bottom) error: norm of 2D error vector in pixel, as a function of the orbit number. Circle size is proportional to the standard deviation of measurements over the product. Red stars indicate failure of the analytics.

Figure 6. Availability history, request status versus date. Blue: product retrieved normally. Red: Data Hub unavailability. Orange: no new product ingested in the last 2 hours. Yellow: no product matching selection criteria found. Unreported status corresponds to s2check maintenance down-time.

Figure 7. Timeliness history. Blue: age of the product at downlink. Khaki: contribution of processing at Payload Data Ground Segment. Orange: dissemination time.
Time production and anomalous products are reprocessed.
- If the anomaly cannot be recovered, affected products are removed from the archive and added to a global list of ‘black orbits’.

In all cases extensive investigations have been needed to identify or characterize the anomaly and find the root cause. The tool does not provide a direct indication of the origin of the anomaly. This is inevitable for a system which aims at detecting new anomalies and not just a re-occurrence of previously characterized ones.

One can note that the detection of the first and fourth anomalies relied on the automatic performance analytics, while the second and third could only be identified by visual inspection. This justifies a posteriori the choice of combining these two approaches.

Two detected anomalies (second and fourth) were affecting the ‘reference’ product used for comparisons instead of the ‘current’ product under test. While the tool was primarily designed to detect anomalies on the near real time production, the fact that several anomalies were found on older products can be explained in retrospect:

- The number of anomalies during the commissioning and ramp-up phase largely exceeds those of the current operational production; so the probability to find an anomaly on older products is higher
- The older anomalous products were generated before the deployment of the S2check tool; the fact that the anomalies found with S2check were not found previously show the improved efficiency of the tool with respect to the previous ‘manual’ inspection process.

In its current version, the tool is not able to determine which of the current or reference products are anomalous. Operators must check both products thanks to a third source if needed.

The end-to-end timeliness plots also provided useful warning on anomalies affecting product generation and dissemination. However specific investigations are needed to identify the origin of a download failure. For instance, a missing file from a product could be due to an ingestion error in the Data Hub or defect in the original product.

The percentage of products flagged as potentially defective by the co-registration analytic is relatively high (40%). This occurs mostly when the number of suitable homologous features is too low, which happens on scenes with a high proportion of water, snow/ices surface, undetected clouds or vegetation (when the reference and current images are in different seasons). Although these cases could be filtered thanks to specific masks (water, snow, vegetation and improved cloud masks), we believe that it is safer to rely on human
analysis at this stage. An additional filtering could indeed unwittingly hide some image defects. Human analysis is therefore still needed to understand the origin of quality warnings. On the other hand we have not observed any failed detection of a co-registration anomaly.

**Conclusion and perspectives**

S2check has proved an efficient tool to support quality control analysis. The concept of automatic random sampling for quality control inspection seems a valid approach that could be implemented in future Payload Data Ground Segments. It complements the simple quality checks applied systematically on the production as well as detailed investigations that can be performed by human operators.

The key ingredients of the method are as follows:

- Use a powerful API to download products suitable for inspection
- Robust and fast performance analytics
- Efficient reporting strategy (summary plots, e-mail and web page)
- Recording of inspections in a database for trend detection and long-term performance monitoring

The main drawback of the present approach is that quality checks are performed after dissemination in order not to impact the timeliness performance. To solve this problem, a new service is being developed to inform users in a convenient way about anomalous product. Planned evolutions for the S2check tool concern the elaboration of additional analytics and performance reports, in particular with the aim of detecting small degradations of the image quality in the time series.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**References**

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., & Louis, Fernández, V. (2017). Copernicus sentinel-2a calibration and products validation status. *Remote Sensing*, 9, 584–664.

Han, W., & Jochum, M. (2016, July). *Near real-time data quality monitoring and control*, Geoscience and Remote Sensing Symposium (IGARSS), Beijing. doi: 10.1109/IGARSS.2016.7729045

Lucas, B.D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of imaging understanding workshop* (pp. 121–130). Pittsburgh: DARPA.

Martin, J., & Knowelden, R. (2015, April). *Copernicus space component esa data access overview*. Paper presented at the Copernicus Space Component Data Access Workshop, Frascati, http://old.esaconferencebureau.com/custom/15M34/Ppt%20Copernicus%20only%20PDF/01%20Day-1/Session%201/03%20CSCDA_ESA_infrastructure%20overview_1.pdf