

*Article*

# 10m Sentinel-2A Cloud-Free Composite – Southern Africa 2016

Fabrizio Ramoino <sup>1,\*</sup>, Florin Tutunaru <sup>2</sup>, Fabrizio Pera <sup>1</sup> and Olivier Arino <sup>3</sup>

<sup>1</sup> SERCO c/o ESA-ESRIN; fabrizio.pera@esa.int

<sup>2</sup> CS-Romania c/o ESA-ESRIN; florin.tutunaru@esa.int

<sup>3</sup> European Space Agency; olivier.arino@esa.int

\* Correspondence: fabrizio.ramoino@esa.int; Tel.: +39-0694180786

**Abstract:** The processing of cloud free geo-referenced imagery is one of the preliminary processing step of any land application. This letter describe the methodology developed to obtain a seamless cloud free composite of Africa for 2016 using Sentinel-2A data at 10 meters resolution freely available from the European Space Agency. The method is based on an hybrid method resulting from the merging of the two most robust time series methods namely the "darkest pixel" and the "maximum NDVI" previously developed with AVHRR time series.

**Keywords:** Sentinel-2; remote sensing; European Space Agency; Copernicus; continental; cloud-free; composite; darkest pixel; maximum NDVI

## 1. Introduction

The study described in this article intends to demonstrate and promote the Sentinel-2A (S2A) capacity to produce a Top Of Atmosphere (TOA) yearly cloud-free composite over Southern Africa up to 8° S at 10 meters resolution, including Blue (490 nm), Green (560 nm), Red (665 nm) and NIR (842 nm) spectral bands.

The S2A mission offers an unprecedented combination of systematic global coverage of land coastal areas, a high revisit under the same viewing conditions, high spatial resolution, and a wide field-of-view (FOV) for multispectral observations from 13 bands in the visible, near infrared and short wave infrared range of the electromagnetic spectrum [1].

As reported in the conclusion of the WorldCover 2017 conference [2], the main objective of Land Cover Community is to have a global high resolution land cover map, and the first step to obtain this is to generate HR cloud-free composites.

The S2A data used in this activity have been acquired from 1st January 2016 to 5th December 2016, the end date does not coincide to the end of the year due to the change in the S2A data format. The input data have been processed with a simple 'two-steps' algorithm based on three-monthly darkest pixel selection to remove clouds and successively yearly composite based on max-NDVI method to further remove cloud shadows. The area of interest (AOI) covers a surface of about 7430000 km<sup>2</sup> and has been processed from 740 S2A L1C granules (110km x 110km, at 10m spatial resolution).

The time series of S2A images has been processed using CalESA system, described in the next paragraph. The yearly composites are based on S2A L1C tiling grid (110km x 110km) and then mosaicked using GDAL utilities and Python routines based on the geo-location information associated to the pixels included in the metadata.

## 2. Materials and Methods

### 2.1. System description

Due to the quantity of S2A data over the AOI, 15 TB, a system for semi-automatic generation of cloud-free composite has been developed.

This approach allows us to use the same resources for data storage and processing.

Processing of the S2A data was done on CalESA system, a cluster composed of 21 server PC, 2 Tower server PC, 1GB network switch, that summarise 120 CPU Core, 580 GB shared Memory and 386 TB storage space. CalESA system comprises the cluster powered by Hadoop, the EO data processing system and the Distributed File System, that stores S2A data. The processing system comprises a processor adapter allowing to integrate Unix or python scripts and SeNtinel Application Platform (SNAP) processor.

The system is based on Hadoop 2.7 and Calvalus with SNAP-core. Logical system hierarchy is as follow: 1 Master PC for process management and HDFS supervision, 20 nodes (slaves) PC for processing and HDFS storage and 2 additional PC for downloading and products ingestion (and products translation for viewer).

Basic platform components:

- Operating system: CentOS 6.8
- Hadoop: 2.7.2
- Java: Oracle java 8
- Calvalus: 2.9 with SNAP 4.0;

The processing workflow is based on Hadoop computational paradigm named Map/Reduce and a formatting step for final product.

#### 1. Map Function

- reads inputs (all available bands)
- re-projects the source pixels to the target grid (same resolution)
- sorts the result pixels into different partitions of the target grid, one per reducer
- and forwards the partitions to the reducers

#### 2. Reduce Function

- sorts the contributions by target grid pixel and time
- applies the temporal aggregator (to all selected bands)
- writes one part of the global result pixel by pixel (all selected bands)

#### 3. Formatting Function

- concatenates the parts and exports the final product (NetCDF)

### 2.2. Algorithm description

#### 2.2.1. Pre-processing

The S2A products acquired before the 5th October 2016 contain more than one S2A L1C granule (110km x 110km), used as basis for the cloud-free composite, therefore the first action was to extract each of them in order to create a stack of data based on S2A L1C tiling grid.

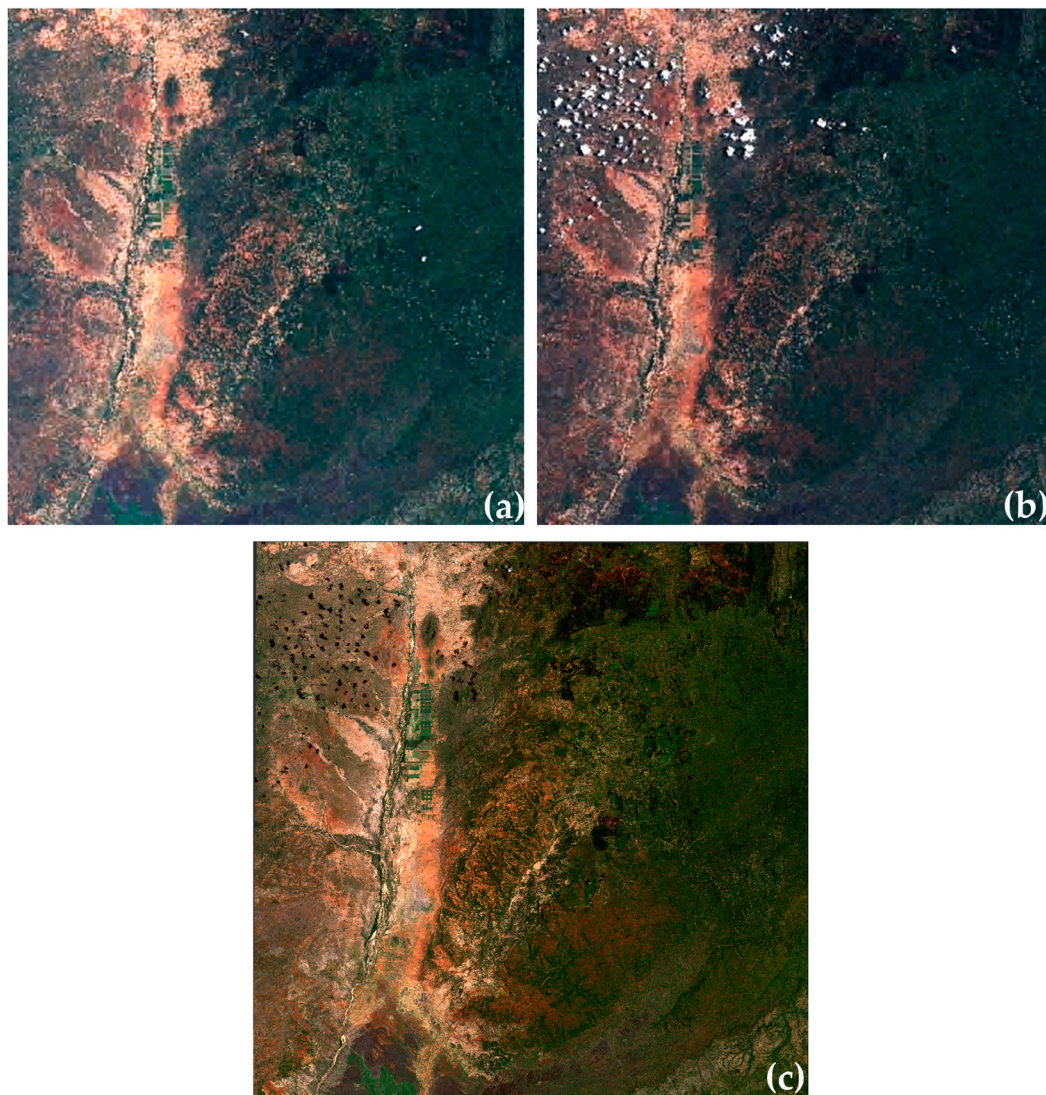
Due to the abundance of S2A data, in order to reduce the processing time, we decided to apply a preliminary screening based on cloud coverage information reported in the metadata (less than 5%), except for the areas with high persistency of clouds where all the available pixels are needed.

The original S2A L1C products are in Universal Transverse Mercator (UTM) coordinate system (CS) based on WGS-84 ellipsoid, this system divides the Earth between 80°S and 84°N latitude into 60 zones, each 6° of longitude in width but it cannot be applied as CRS to map entire Africa. For this reason we re-projected all S2A L1C products in Geographic Latitude and Longitude CS.

#### 2.2.2. Darkest pixel method

Usually, to generate cloud-free maps complex algorithms for clouds detection and removal are applied, such as Automated Cloud-Cover Assessment Algorithm (ACCA) [3] or F-mask [4].

To simplify the processing, we decided to use a simpler method that chooses the darkest value for the 10 meters bands (VIS and NIR) for each pixel of the three-monthly stack of S2A images, as shown in Figure 1.



**Figure 1.** (a) S2A image acquired on 04/02/2016; (b) S2A image acquired on 14/02/2016; (c) 1<sup>st</sup> Quarter 2016 composite based on darkest pixel selection. © Contains modified Copernicus Sentinel data (2016) / Processed by ESA.

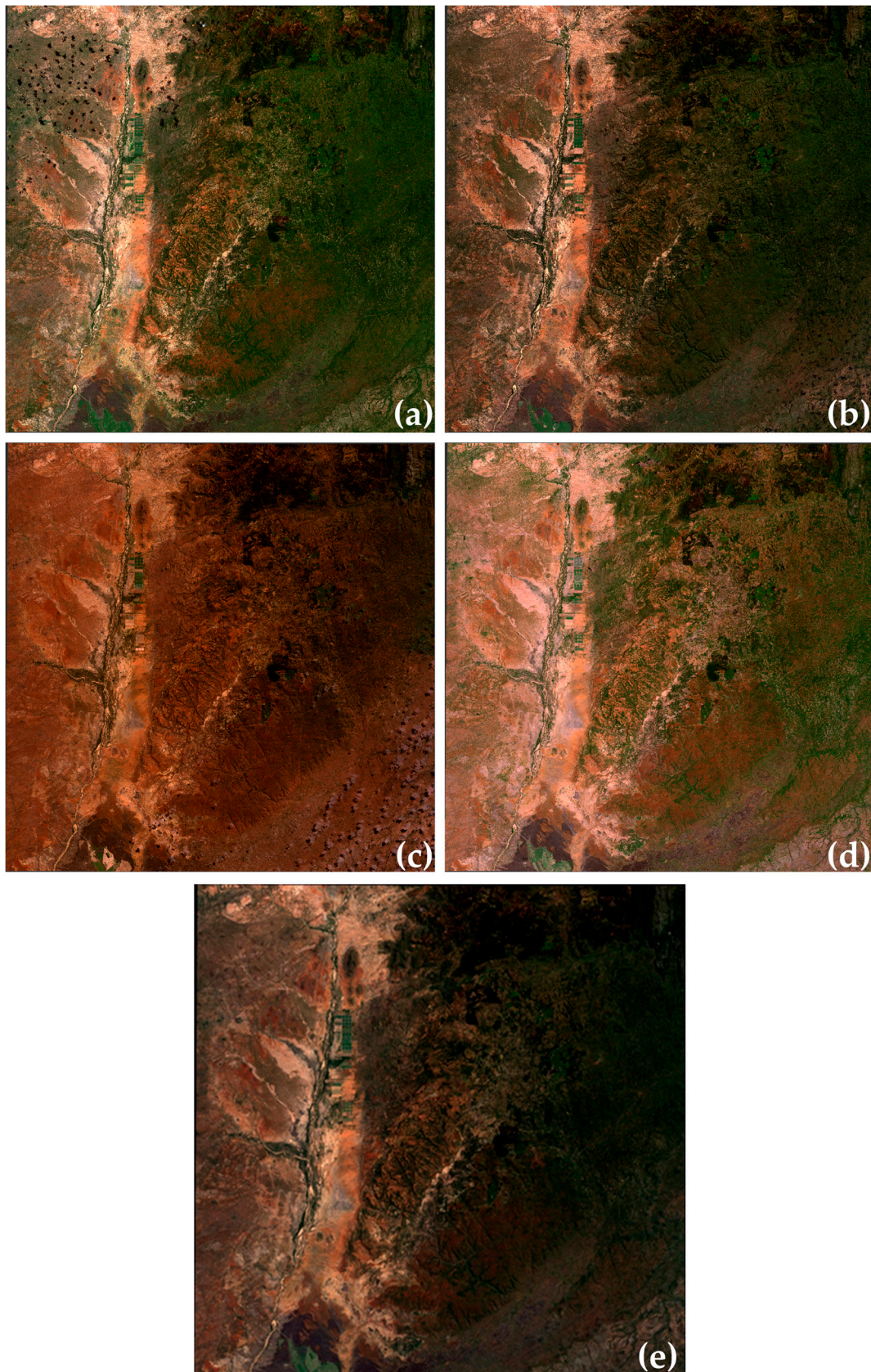
As described by Saunders [5], in the histogram of visible reflectance the land and sea peaks are well defined at the 'dark' end with cloudy reflectance producing a broad higher reflectance tail.

The three-monthly intermediate composites replace cloudy pixels with cloudless ones but also cloud shadows pixels are included.

### 2.2.3. Maximum NDVI method

The Normalised Difference Vegetation Index (NDVI) measures the spectral response of the vegetation and it is defined as difference between near-infrared and red reflectance divided by the sum of the two  $(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$ . As reported by Holben in the Maximum-Value Composite (MVC) procedure [6], any attenuation of the spectral signals will reduce the NDVI value from that of an unattenuated signal. Therefore, the best possible pixel value for a particular location is achieved by choosing the highest pixel value from multi-temporal data (Figure 2).





**Figure 2.** (a) Darkest pixel composite for 1st Quarter 2016; (b) Darkest pixel composite for 2nd Quarter 2016; (c) Darkest pixel composite for 3rd Quarter 2016; (d) Darkest pixel composite for 4th Quarter 2016; (e) Yearly composite based on Max-NDVI. © Contains modified Copernicus Sentinel data (2016) / Processed by ESA.

#### 2.2.4. Combining method

Comparing the output of the two methods described above, we observed that each method taken individually is insufficient, and for this reason we combined them in the most efficient manner considering the processing time and final results.

STEP-1 - To remove the clouds, for each quarter (January-March, April-June, July-September, October-December) darkest pixel selection from the stack of S2A L1C TOA images identified as input.

STEP-2 - From the quarterly composites choose the best pixel applying the “max-NDVI” method replacing the pixels affected by cloud shadow with clear ones, if present.

#### 2.2.5. Post-processing

To visualise the final product we developed a series of python scripts that combine different GDAL utilities described by R. Simmons [7], applying histogram stretching to the RGB bands, input data compression and generating an overview at different zoom levels of each S2A L1C TOA tile.

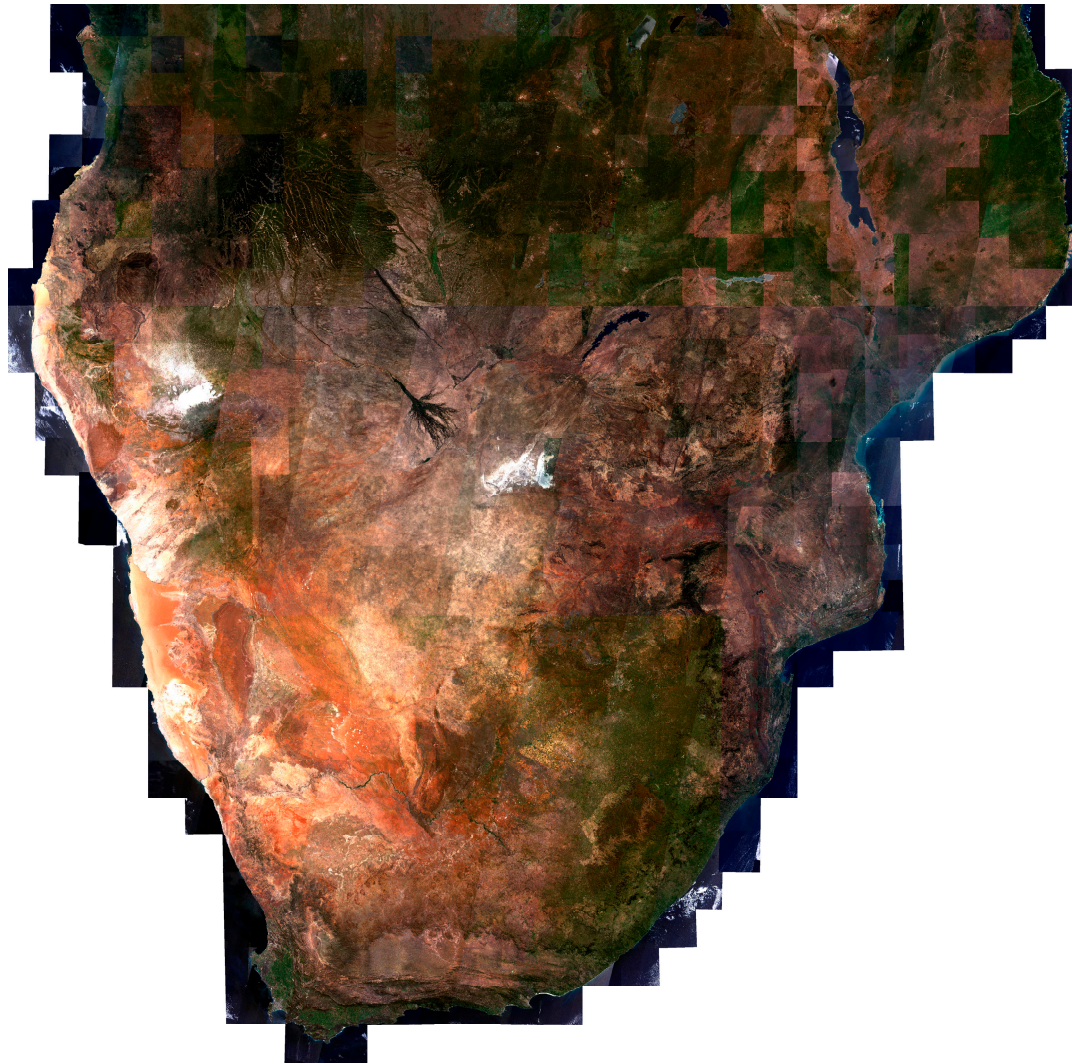
The output of the post-processing is the RGB mosaic at 10m spatial resolution available at <http://2016africacomposite10m.esrin.esa.int/>.

To display and interact with the map on the web page we used Leaflet an open-source JavaScript library.

### 3. Results

The final result shown in Figure 3, obtained applying our simple processing chain, is comparable to the ones generated using more complicated algorithms.





**Figure 3.** S2A cloud-free yearly composite - Southern Africa up to 8° S. © Contains modified Copernicus Sentinel data (2016) / Processed by ESA.

#### 4. Discussion

The HR cloud-free composites are of crucial importance for the generation of HR land cover maps that represent key point to support monitoring of changes to vegetation within the growing season, to improve the efficiency of management of agricultural resources and to avoid emerging threats to the future of agriculture (e.g. land degradation, scarcity of land and water resources, biodiversity loss and climate change).

The near-future activities in the frame of the study reported in this article will be focused to generate the cloud-free composite over entire Africa for 2016 at 10 meters spatial.

#### 5. Conclusions

The strong points of the 'two steps' method described in this study are the easy implementation, the quick computation time and the possibility to export and to adapt it easily on cloud processing system expanding the AOI and the time series.

**Acknowledgments:** The S2A Level 1C TOA data have been kindly provided by Eric Monjoux and Adriana Grazia Castriotta from ESA Payload Data Ground Segment (PDGS).

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

## References

1. Gascon, F.; Thépaut, O.; Jung, M.; Francesconi, B.; Louis, J.; Lonjou, V.; Lafrance, B.; Massera, S.; Gaudel-Vacaresse, A.; Languille, F.; B. Alhammoud, B.; Viallefont, F.; Bieniarz, J.; Pflug, B.; Clerc, S.; Pessiot, L.; Trémas, T.; Cadau, E.; De Bonis, R.; Isola, C.; Martimort, P.; Fernandez, V. Copernicus Sentinel-2 Calibration and Products Validation Status. *Preprints*, **19 October 2016**, DOI: 10.20944/preprints201610.0078.v1.
2. WorldCover 2017 Conference. Available online: <http://worldcover2017.esa.int/> (accessed on 31 May 2017).
3. Irish, R.R.; Barker, J.L.; Goward, S.N.; Arvidson, T. Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm. *Photogrammetric Engineering & Remote Sensing*, **2006**, Volume 72, (pp. 1179-1188), DOI: 10.14358/PERS.72.10.1179.
4. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *RSE*, **2015**, Volume 159, (pp. 269-277), DOI: 10.1016/j.rse.2014.12.014.
5. Saunders, R.W.; Kriebel, K.T. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *IJRS*, **1998**, Volume 9, (pp. 123-150), DOI: 10.1080/01431168808954841.
6. Holben, B.N. Characteristics of Maximum Value Composite images from temporal AVHRR data. *JRS*, **1986**, Volume 7, (pp. 1417-1434), DOI: 10.1080/01431168608948945.
7. Simmons, R. A Gentle Introduction to GDAL Part 4: Working with Satellite Data, **2017**. Online article available at: <https://medium.com/@robsimmon/a-gentle-introduction-to-gdal-part-4-working-with-satellite-data-d3835b5e2971> (accessed on 31 May 2017).