

ACIX I: CEOS-WGCV ATMOSPHERIC CORRECTION INTER-COMPARISON EXERCISE (SUMMARY)

Exercise coordinated by: Georgia Doxani (ESA/Serco), Ferran Gascon (ESA), Eric Vermote (NASA), Jean-Claude Roger (University of Maryland/NASA).

Motivation and Objectives

Landsat 8 and Sentinel-2 provide geometrically and radiometrically high resolution spectral measurements of land and water surfaces in a timely dense manner, and the nowadays available computing possibilities ease processing and analysis of the corresponding large data volumes. Accessing the full information of the data requires a proper pre-processing. CEOS includes in its definition of Analysis Ready Data for optical Earth Observation data specifically the atmospheric correction (AC), i.e. “Adjustments for atmospheric effects (absorption and scattering) due to water vapor, ozone, molecular scattering, and aerosols” as well as Quality Meta Data including clouds and cloud shadow, as minimum requirements [1]. Over the past years several organisations worldwide have started to generate, or plan to generate in the short term, surface reflectance products at global scale for Landsat-8 and Sentinel-2 missions. On behalf of CEOS, NASA and ESA have organised ACIX, an intercomparison exercise on AC processors to point out strengths and weaknesses as well as commonalities and discrepancies of various AC processors. The exercise has been designed to lead to a better understanding of the different uncertainty components and to the improvement of the AC processors’ performance in order to suggest and define ways for their further improvement. This first exercise, ACIX-I, focused on the AC over land surfaces. A follow-on activity, ACIX-II, will include a cloud masking intercomparison exercise (CMIX) that will consider AC over both land and water surfaces.

Exercise Overview

ACIX-I was conducted between June 2016 and November 2017. It was organised in 4 steps:

- 1) Agreement on the analysis protocol (concept, test sites & products, evaluation metrics); this was prepared by the organisers, reviewed by the participants and finalized during a workshop
- 2) Running of the AC on the agreed set of L8 and S2 test products (participants)
- 3) Evaluation of the quality of the AC according to the agreed metrics (organisers)
- 4) Analysis of results, drafting conclusions and recommendations (common, workshop)

The exercise was free and open, and eventually 12 organisations participated (see Annex 1). The exercise was conducted in very close cooperation between the organisers and the participants. The outcome of the exercise has been published in a peer reviewed paper [2]. All detailed results are available from the CEOS CalVal Portal [3]

Protocol

The protocol was designed to include typical experimental cases over diverse land cover types and atmospheric conditions. A total of 19 globally distributed sites were selected based on the locations of the international Aerosol Robotic Network (AERONET), including also 5 coastal and inland water sites to examine the performance of the AC over water. A detailed procedure to select match-ups, to filter and extract macro-pixels, and to calculate statistical metrics was defined: the analysis was performed on three parameters - Aerosol Optical Thickness (AOT), Water Vapour (WV), and Surface Reflectance (SR) - and on image subsets of 9 km x 9 km, centered on the location of the AERONET Sunphotometer stations of each site. The quality masks originating from each algorithm were either blended altogether or in combinations, and only the pixels commonly flagged as ‘good quality’ were considered in the

analysis. Since there were no other global networks mature enough or with similar global representation, the AERONET in-situ measurements of AOT and WV were considered to be the ground truth in ACIX. As AERONET does not provide SR measurements, the protocol includes (a) an analysis of the differences among the SRs from the different ACs and (b) a comparison of the AC results with a reference SR. This was calculated using the 6SV [4] Radiative Transfer (RT) model, using the AERONET AOT and WV values. The authors acknowledge that the choice of 6SV as the RT model for the computation of the reference SR could constitute a moderate, but not negligible advantage for the AC codes that use this same model in their RT simulations. For each comparison the statistical metrics accuracy, precision and uncertainty were used.

Results

Not all processors were able to process all input data. In total four AC processors, i.e. ATCOR, FORCE, iCOR and LaSRC, were applied to most of the Landsat 8 (L8) datasets. Eight processors, i.e. CorA, FORCE, iCOR, LaSRC, MACCS, S2-AC2020, GFZ-AC and Sen2Cor, provided results for Sentinel-2 (S2) datasets over most of the sites. The analysis was performed per sensor and parameter. The main results are:

- L8/AOT: With the exception of specific cases, all AC retrieved the aerosol in a similar and satisfactory manner, with a mean difference of 0.1 in AOD at 550nm. Among all AC processors, iCOR has the lowest root mean square (RMS) value, i.e. scatter around the 1:1 line. Algorithms that derive the aerosol over dense dark vegetation (DDV) show problems in arid areas, as expected. Problems also occur when the surface is water or snow.
- L8/SR: The analysis was performed on accuracy, precision and uncertainty (APU), when the theoretical SR was used as reference. The APU was compared with a theoretical specification, in which the APU follows a linear line, derived from instrument specification ($0.005+0.05 \times p$). See Annex 2 for an example. It was found that the uncertainty does not exceed the specification for most of the ACs and sites. In particular, ATCOR and LaSRC perform best with low APU scores.
- S2/AOT: As with L8, the arid sites were found to be challenging for all the AC processors, basically because of the absence of DDV pixels. One algorithm, GFZ-AC, used a non-image-based AOT retrieval method, but extracted it from ECMWF forecast data. In this case, the agreement with AERONET reference is not as good, presumably due to the large grid cell size. Overall, the AC processors produced good AOT results, with a mean deviation in the order of 0.13. LaSRC achieved the best overall agreement between the AOT estimates and the reference AERONET measurements, as indicated by the lowest RMS value.
- S2/WV: Overall, the processors succeeded in more accurate retrievals of WV than AOT values. Most of the points fall close to 1:1 line, apart from a few exceptions over arid and equatorial forest sites. In particular, a very good agreement is observed for MACCS estimates, although in this case the results provided were limited to specific sites (the algorithm requires a time series of imagery as input).
- S2/SR: The analysis shows more diverse results for the different ACs compared to L8 case. This result indicates that the development of ACs for S2 is still in a development phase and improvements can be expected in the future. Overall, FORCE, LaSRC, MACCS and Sen2Cor managed to estimate the SRs quite well over all the sites. However, when a time series of imagery is required, as in the case of MACCS, a significantly lower number of match-ups is retrieved, but this is – at least partly – caused by the fact that S2 started providing images every 10 days only in April 2016. Spectral differences in the quality of the retrievals exist also between the different ACs that should be further analysed and understood. As previously mentioned, the reference BOA reflectance were computed by 6SV RT code that is the same as the one used in some of the AC codes. This can provide a modest but non-negligible advantage to these AC codes, e.g. LaSRC.

A comprehensive compilation of the main results can be found in [2]. All detailed results are available online [3].

Overall, all ACs provided good retrieval with generally a quite comparable quality but differences in detail. It should be noted that larger differences might have been observed, if the comparison had been performed at higher resolution where disturbances such as cloud shadows occur or adjacency effects become more important.

Conclusions and Recommendations

The availability of high quality optical Earth observation data from Landsat 8 and Sentinel-2, their accessibility, and the disposability of powerful computational resources, enable the generation of specifically cloud masked and atmospherically corrected data. ACIX is a very timely activity to provide a state-of-the-art overview and assessment of multiple algorithms, knowing that they are rapidly evolving. ACIX is steered by NASA and ESA as organisers on behalf of CEOS Working Group on Calibration and Validation (WGCV), but the work was performed by the participants. ACIX is free and open to everyone. Altogether, this is a basis for a fair and objective intercomparison, with a wide and representative coverage of AC algorithms. ACIX aims at a scientific understanding of strengths and weaknesses, and resulted in an open access, peer reviewed publication. This publication is considered by the user community to be a key resource for future algorithm improvements and guidance for work going forward.

More specifically, ACIX has proven to be successful in a) quantifying the discrepancies of AOT and WV products compared to AERONET measurements, and b) identifying the similarities among the AC algorithms by analysing and presenting all the results in the same manner. The objective to address the strengths and weaknesses of each over diverse land cover types and atmospheric conditions was achieved to some extent but could be improved. Different factors, e.g. time constraints, tuning of the processors, processor's limitations, etc., constraint the application of some AC algorithms on certain L-8 and/or S-2 datasets. Due to this variance of the submitted results, it was not feasible to draw common conclusions among all the algorithms, but fortunately these cases were not the majority.

ACIX started only six months after the beginning of the availability of S-2 Level-1 data to all users. The research community was still inexperienced at that phase, and time and effort were needed to adapt the processors to the new data requirements. Hence, the discrepancies observed in ACIX inter-comparison results have assisted the developers to learn about the performance and also identify the flaws in their algorithms. As a matter of fact, the participants have already modified and improved their processors.

Although this situation is specific to the timing of Sentinel-2, it is a general fact that scientific data processing algorithms are continuously evolving. The success of ACIX, the lessons learnt, imply automatically that its results will be outdated sooner or later as algorithm developers respond. Thus it is recommended to repeat the exercise, at least once due to the specific case of Sentinel-2, but even better at regular intervals of 2 – 3 years. The continuation ACIX-II is already under preparation.

Based on the experience of ACIX-I, the inter-comparison strategy shall be refined to:

- address the cloud masking explicitly, guided by a dedicated protocol;
- include additional sites for wider coverage of atmospheric and surface conditions; emphasis shall be on coverage of different land cover classes and aerosol types;
- include a time series of test data, at least a one-year period from L-8, S-2A and S-2B;
- include other sources of reference data, e.g. RadCalNet;
- analysis of higher spatial resolution (pixel scale) in order to allow testing the adjacency effect;
- assess the time consistency of time series;
- assess the sensitivity of ACs to instrument noise;
- study effects of undetected clouds or shadows;

- strive that all the participants will apply their processors over all sites, in order to gain an overall assessment of their inter-performance.

References

[1] <http://ceos.org/ard/>

[2] Doxani, G.; Vermote, E.; Roger, J.-C.; Gascon, F.; Adriaensen, S.; Frantz, D.; Hagolle, O.; Hollstein, A.; Kirches, G.; Li, F.; Louis, J.; Mangin, A.; Pahlevan, N.; Pflug, B.; Vanhellefont, Q. "Atmospheric Correction Inter-Comparison Exercise." *Remote Sens.* **2018**,*10*,352. doi:[10.3390/rs10020352](https://doi.org/10.3390/rs10020352)

[3] <http://calvalportal.ceos.org/projects/acix>

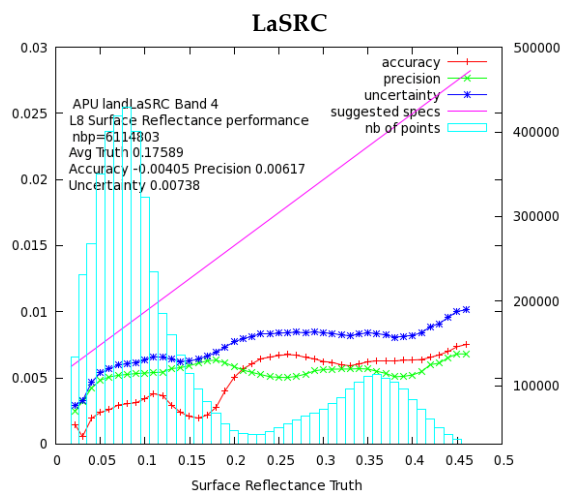
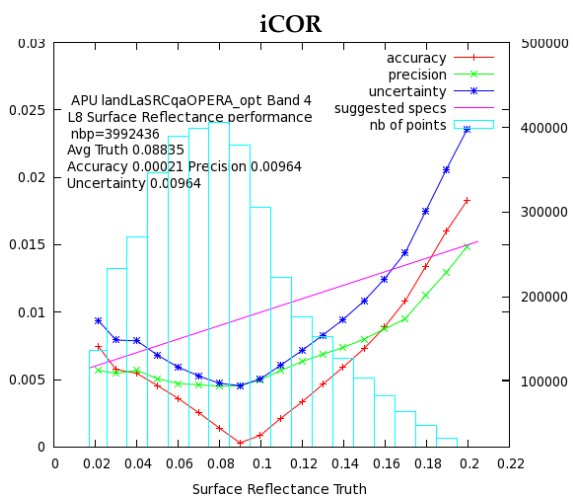
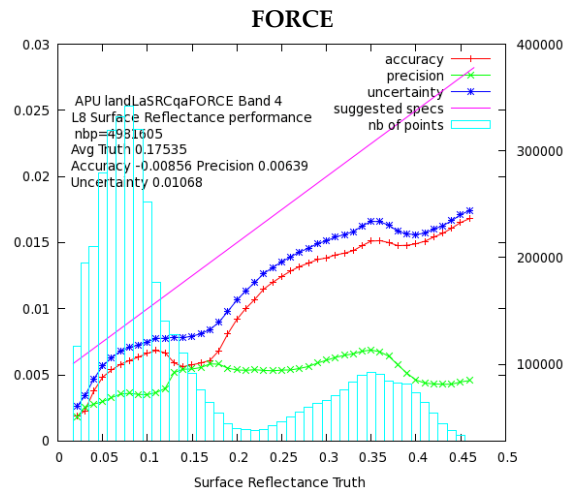
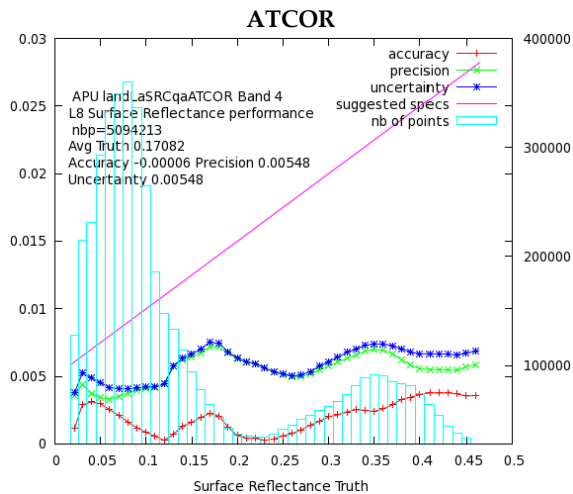
[4] <http://6s.ltdri.org/>

Annex 1: List of ACIX participants

AC Processor	Participants	Affiliation
ACOLITE	Quinten Vanhellefont	Royal Belgian Institute for Natural Sciences [Belgium]
ATCOR/ S2-AC2020	Bringfried Pflug, Rolf Richter, Aliaksei Makarau	DLR - German Aerospace Center [Germany]
CorA	Grit Kirches, Carsten Brockmann	Brockmann Consult GmbH [Germany]
FORCE	David Frantz	Trier University [Germany]
iCOR	Stefan Adriaensen	VITO [Belgium]
GA-PABT	Fuqin Li	Geoscience Australia [Australia]
LAC	Antoine Mangin	ACRI [France]
LaSRC	Eric Vermote	GSFC NASA [USA]
MACCS	Olivier Hagolle	CNES [France]
GFZ-AC	André Hollstein	GFZ German Research Centre for Geosciences [Germany]
SeaDAS	Nima Pahlevan	GSFC NASA [USA]
Sen2Cor v2.2.2	Jerome Louis ¹ , Bringfried Pflug ²	¹ Telespazio France [France] ² DLR - German Aerospace Center [Germany]

Annex 2: Surface Reflectance Results

A) Landsat-8



The accuracy (red line), precision (green line), and uncertainty (blue line) as computed in bins (blue bars) for OLI Band 4 (Red). The total number of pixels (nbp) used in the computations is given also in the plot. The magenta line represents the theoretical SR reference for Landsat SR ($0.005+0.05 \times p$).