

## CALIBRATION AND VALIDATION PLAN FOR THE L2A PROCESSOR AND PRODUCTS OF THE SENTINEL-2 MISSION

M. Main-Knorn<sup>a,\*</sup>, B. Pflug<sup>a</sup>, V. Debaecker<sup>b</sup>, J. Louis<sup>b</sup>

<sup>a</sup> German Aerospace Center (DLR), Earth Observation Center, Remote Sensing Technology Institute, Photogrammetry and Image Analysis, 12489 Berlin, Germany - (magdalena.main-knorn, bringfried.pflug)@dlr.de

<sup>b</sup> Telespazio France, SSA Business Unit (Satellite Systems & Applications), 31023 Toulouse Cedex 1, France - (vincent.debaecker, jerome.louis)@telespazio.com

**SESSION: Airborne and Innovative Remote Sensing Platforms and Techniques**

**KEY WORDS:** Sentinel-2, Sen2Cor, Atmospheric correction, processor calibration and validation, ground based and in-situ measurements

### ABSTRACT:

The Copernicus programme, is a European initiative for the implementation of information services based on observation data received from Earth Observation (EO) satellites and ground based information. In the frame of this programme, ESA is developing the Sentinel-2 optical imaging mission that will deliver optical data products designed to feed downstream services mainly related to land monitoring, emergency management and security. To ensure the highest quality of service, ESA sets up the Sentinel-2 Mission Performance Centre (MPC) in charge of the overall performance monitoring of the Sentinel-2 mission. TPZ F and DLR have teamed up in order to provide the best added-value support to the MPC for calibration and validation of the Level-2A processor (Sen2Cor) and products. This paper gives an overview over the planned L2A calibration and validation activities.

Level-2A processing is applied to Top-Of-Atmosphere (TOA) Level-1C ortho-image reflectance products. Level-2A main output is the Bottom-Of-Atmosphere (BOA) corrected reflectance product. Additional outputs are an Aerosol Optical Thickness (AOT) map, a Water Vapour (WV) map and a Scene Classification (SC) map with Quality Indicators for cloud and snow probabilities. Level-2A BOA, AOT and WV outputs are calibrated and validated using ground-based data of automatic operating stations and data of in-situ campaigns. Scene classification is validated by the visual inspection of test datasets and cross-sensor comparison, supplemented by meteorological data, if available. Contributions of external in-situ campaigns would enlarge the reference dataset and enable extended validation exercise. Therefore, we are highly interested in and welcome external contributors.

## 1. INTRODUCTION

In the frame of the Copernicus Programme (European initiative for the implementation of information services dealing with environment and security), ESA is developing the space segment with a series of satellite missions, known as Sentinels. Sentinel missions 1 to 3 are designed as a constellation of two satellites to improve revisit and earth coverage requirements. Both radar- and multi-spectral imaging instruments are applied to monitor land, ocean and atmosphere.

### 1.1 Sentinel-2 mission

Sentinel-2 is a polar-orbiting, optical mission for land monitoring and emergency services, designed with an enhanced spectral range and performances compared to previous similar imaging missions as SPOT and Landsat missions. S-2 offers satellite data with systematic global coverage of land surfaces, high spatial and spectral resolution, and wide field of view (290 km) (Table 1).

Multi-spectral instrument aims at measuring reflected radiance in 13 spectral bands spanning from the Visible to the Short Wave Infra-Red spectral range (Table 2). Comparing to the recently launched Landsat 8 (OLI) imaging system, S-2 includes eight similar bands, while has non thermal band.

Sentinel-2	
Imager	multi-spectral
Launch	S-2A: June 12 <sup>th</sup> 2015 S-2B: 18 months after pushbroom
Measurement principle	pushbroom
Swath (in km)	290
Spectral bands	13
Spatial resolution (in m)	10, 20, 60
Altitude (in km)	786.13
Orbit	Sun-synchronous polar LTDN 10h30 AM
Revisit period (in days)	10 days in similar viewing condition for one satellite 5 days for the two satellites (S2A and S2B)

Table 1. Summary of mission specifications. *LTDN* refers to Local Time of Descending Node.

\* Corresponding author.

Band number	Centre $\lambda$ [nm]	Spectral width $\Delta\lambda$ [nm]	Spatial resolution [m]
1	443	20	60
2	490	65	10
3	560	35	10
4	665	30	10
5	705	15	20
6	740	15	20
7	783	20	20
8	842	115	10
8a	865	20	20
9	945	20	60
10	1375	30	60
11	1610	90	20
12	2190	180	20

Table 2. Multi-spectral instrument specifications.

To ensure the highest quality of service for the Sentinel-2 mission, ESA sets up the S-2 Mission Performance Centre (MPC). The MPC is in charge of the overall performance monitoring of the S-2 mission within the S-2 Payload Data Ground Segment (PDGS). It is constituted of a Coordinating Centre (MPC/CC) in charge of the main routine activities, engineering support to the mission, overall service management, and the Expert Support Laboratories (ESLs) providing the scientific expertise according to their area of competency (Figure 1).

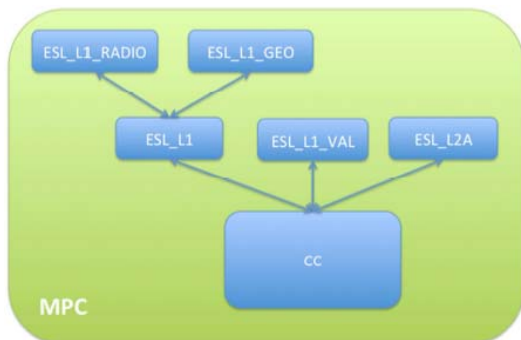


Figure 1. MPC/ESL organization layout © ESA.

The S-2 MPC includes five ESLs responsible for processors calibration and validation of Level-1 and Level-2A Products. In particular, the ESLs are in charge of following activities:

1. Multi-spectral instrument (MSI) configuration
2. L1 and L2A algorithms evolution, verification and maintenance over the mission lifetime
3. Maintaining the processing baseline
4. Calibration of the MSI sensor and L1 processor
5. Radiometric and Geometric validation of L1 products
6. Calibration of the L2A processor (Sen2Cor)
7. Geophysical validation of all the L2A products.

## 2. THE LEVEL-2A PROCESSOR AND PRODUCTS

### 2.1 Sen2Cor Processor and Products

Sen2Cor is the processor for S-2 Level-2A product processing and formatting. The processor performs the tasks of Atmospheric Correction (AC), Cloud Screening and Scene Classification (SC) of Level-1C input data. Level-2A outputs of

Sen2Cor (Figure 2) are the orthorectified Bottom-of-Atmosphere (BOA) corrected reflectance images, Aerosol Optical Thickness (AOT) map, Water Vapour (WV) map, Scene Classification maps, Cloud probabilistic mask and Snow probabilistic mask (Quality Indicators).

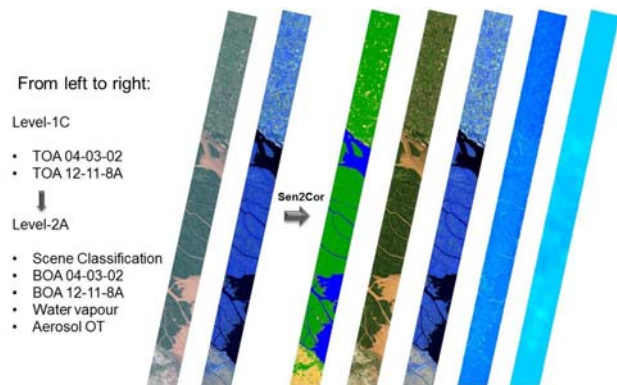


Figure 2. S-2 Level-2A product overview. © Telespazio

The **Scene Classification (SC)** algorithm allows 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: vegetation, soils / deserts, water, snow, shadows and cloud shadows. The algorithm is based on a series of threshold tests that use as input top-of-atmosphere reflectance from the Sentinel-2 spectral bands. In addition, thresholds are applied on band ratios and indexes like the Normalized Difference Vegetation (NDVI) and Snow Index (NDSI). For each of these thresholds tests, a level of confidence is associated. At the end of the processing chain a probabilistic cloud mask quality map and a snow mask quality map is produced. The algorithm uses the reflective properties of scene features to establish the presence or absence of clouds in a scene. Cloud screening is applied to the data in order to retrieve accurate atmospheric and surface parameters, either as input for the further processing steps below or for being valuable input for processing steps of higher levels.

The **Atmospheric Correction (AC)** is performed using a set of Look-Up tables generated via libRadtran (Mayer and Kylling, 2005). Baseline processing is the rural/continental aerosol type. Other Look-Up tables can also be used according to the scene geographic location and climatology. The Atmospheric Correction module is a porting and adaptation of the ATCOR software into Python.

**2.1.1 Aerosol Optical Thickness (AOT):** AOT retrieval provides a measure for the visual transparency of the atmosphere. It is derived using the DDV (Dense Dark Vegetation) algorithm (Kaufman and Sendra, 1988), using the (SWIR) band 12 and correlates its reflectance with bands 4 (red) and 2 (blue). The algorithm requires that the scene contains reference areas of known reflectance behavior, preferably Dark Dense Vegetation (DDV) and/or dark soil and water bodies. The algorithm starts with a user-defined visibility (default: 20 km) as input. If the scene contains no dark vegetation or soil pixels, the surface reflectance threshold of band 12 will be successively iterated in order to include medium brightness reference pixels in the sample. If the scene contains no reference and no water pixels the scene is processed with the start visibility instead. The algorithm delivers an AOT map.

**2.1.2 Water Vapour:** WV retrieval over land is performed with the Atmospheric Pre-corrected Differential Absorption algorithm (Schläpfer et al. 1998) which is applied to the two Sentinel-2 bands B8a, and B9. Band 8a is the reference channel in an atmospheric window region. Band B9 is the measurement channel in the absorption region. The absorption depth is evaluated by calculating the radiance for an atmosphere with no water vapour, assuming that the surface reflectance for the measurement channel is the same as for the reference channel. The absorption depth is then a measure of the water vapour column content.

**2.1.3 Cirrus Correction:** Algorithm uses the Sentinel-2 (cirrus) channel 10. Thin cirrus clouds affect the visible, near- and shortwave infrared spectral regions. They are partially transparent and thus difficult to detect with broad-band multispectral sensors, especially over spatially inhomogeneous land areas. Water vapour, in contrast, dominates in the lower troposphere of 0-5 km altitude. A narrow spectral band in a spectral region of very strong water vapour absorption (Band 10) will thus absorb the ground reflected signal, but will receive the scattered cirrus signal. Cirrus reflectance of band 10 can therefore be correlated with other bands in the VNIR and SWIR region and the cirrus contribution can thus be removed from the radiance signal to obtain a cirrus-corrected scene.

**2.1.4 Topographic correction:** Recommended if more than 5% of the pixels have slopes  $> 8^\circ$ . In mountainous terrain the topography introduces strong brightness variations depending on the orientation of a surface element. The objective of a combined topographic / atmospheric correction is the elimination of topographic effects during the surface reflectance retrieval. An accurate digital elevation model (DEM) of about the same spatial resolution as the pixel size of the instrument and a very accurate ortho-rectification are required to achieve a satisfactory topographic correction. In addition scenes in mountainous regions often exhibit a large variation of terrain slopes, and thus bidirectional brightness variations for a certain surface cover, e.g. meadow or forest. This behaviour cannot adequately be eliminated with the Lambertian assumption and leads to overcorrected reflectance values in faintly illuminated areas. Therefore several BRDF empirical corrections are proposed within Sen2Cor to overcome this problem.

### 3. THE L2A CALIBRATION AND VALIDATION

Since the S-2 launch is scheduled for June 2015, first data can be expected already in July or August 2015. These data have to be very precisely calibrated and validated at all pre-processing levels before dissemination to the users.

Telespazio France and DLR have teamed up in order to provide the best added-value support to the MPC for calibration of the Level-2A processor (Sen2Cor) and geophysical validation of Level-2A products.

#### 3.1 L2A Calibration

Calibration of the L2A processor encompasses two main domains: radiometry, which concerns the AC outputs of Sen2Cor, and classification, which involves Cloud Screening and Classification products. Calibration of radiometry outputs focuses on the Atmospheric Correction parametrization and comprises following activities:

1. Investigation of the processor sensitivity to the configuration parameters (e.g. thresholds). The impact of

the individual parameter variations on the different L2A products is analysed and compared to the in-situ measurements. For AOT and WV calibration, reference data originate from the AERONET network.

2. Qualitative analysis of the impact of the activation/deactivation of yes/no parameter is undertaken (e.g. cirrus correction, BRDF correction)

3. Investigation of Sen2Cor source code itself in order to identify other additional parameters for the processor full calibration.

Calibration of Cloud Screening and Classification algorithm is based on an empirical approach and encompasses full calibration of threshold parameters. The calibration procedure consists of:

1. Run of Sen2Cor processor using default thresholds to produce SC products
2. Manual inspection of pixel classification results to superimpose the scene classification map and L1C spectral bands. 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 cases: ("easy", "medium", "difficult"). These cases correspond to a tuning of the SC thresholds that improves the Scene Classification: without having a negative impact on the other classes ("Easy"), having a limited negative impact on other classes ("Medium"), or improving results for one class but having a negative impact on other classes ("Difficult")
4. The proper activity of fine tuning consists in manually slightly varying the SC thresholds to improve the Scene Classification based on these three different cases (easy, medium, and difficult)
5. Run of Sen2Cor processor using a new set of updated SC parameters
6. A quantitative comparison exercise 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.

A calibration dataset of Sentinel-2 Level-1C products is constituted, covering different land cover types (e.g. snow, rocks, desert, urban, vegetation, grass, forest, cropland, vineyard, irrigated crops, rivers, lakes, sand, coastal area, wetlands, ocean water) and different atmospheric conditions (e.g. cloud cover, aerosol optical thickness, water vapour content). In addition, the calibration dataset is worldwide including different latitudes in order to cover various solar angles and seasons (Figure 3).

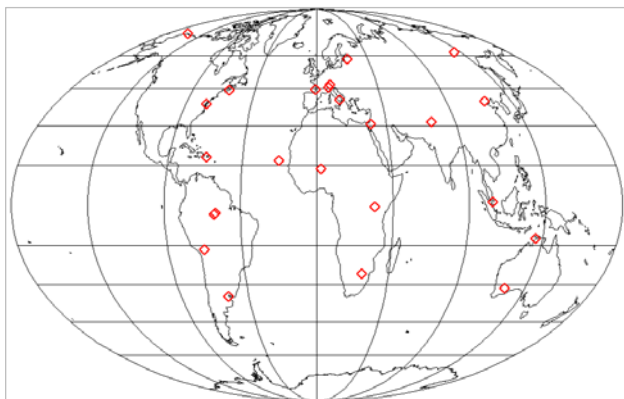


Figure 3. Geographical distribution of the 24 selected Aeronet test sites for the L2A Calibration (core + back-up sites).

### 3.2 L2A Validation

Validation activities consist of validation of radiometry and classification outputs of Sen2Cor processor. Geophysical validation of L2A products encompasses:

**3.2.1 Validation of AOT and WV:** This task is performed by comparing Sen2Cor outputs with in-situ data provided by collocated S-2A overpasses and sunphotometer measurements at the ground for a set of test sites representative of main surface and atmosphere types. Validation comprises following activities:

1. Acquisition, quality assurance and archiving in-situ sunphotometer measurements from selected Aeronet sites representative for all latitude regions, all continents, different land cover types and different topography (Figure 4).
2. Microtops measurements on test site Potsdam/Berlin performed in a time series from ~2 h before to ~2 h after the Sentinel-2 overpass, if the meteorological conditions are acceptable. Short time series with a time step of 15 minutes allows recognition of trends and outliers (quality assurance). Even at the stable land surface it is required for each time step to perform groups of at least 10 scans with each instrument for the final data analysis.
3. Processing of Microtops measurements using a coupled analysis of sunphotometer and ozonometer measurements (Pflug, 2013). First, ozonometer data are used for computation of vertical column ozone content [cmSTP]. Actual cmSTP is used to calculate vertical column AOT-spectra. AOT spectra allow computation of the vertical column Ångström exponent  $\alpha$ , which contains information about aerosol particle size respectively aerosol type. Spectral dependency of AOT given with the AOT-spectra is also used to compute vertical column water vapour content [cm precipitable water column] and to interpolate the AOT at 550 nm.
4. Direct comparison of Sen2Cor outputs with ground-truth from in-situ data and statistical analysis.

More basic meteorological information like humidity of air, wind speed, wind direction and cloud cover can assist interpretation of results as additional information characterizing aerosol properties and the quality of the satellite image.

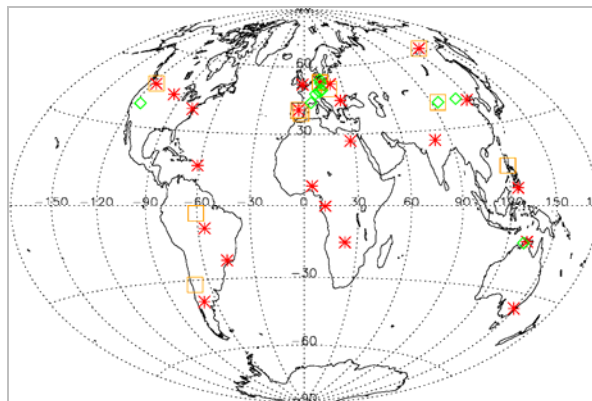


Figure 4. Geographical distribution of the selected core+back-up test sites for the L2A Validation (Red asterisks: Test sites with sunphotometers for AOT, WV and BOA validation; Green diamonds: Test sites for in-situ surface reflectance measurements; Orange squares: Test sites for SC validation).

**3.2.2 Validation of BOA reflectance:** Validation focuses on comparing Sen2Cor outputs with surface reflectance reference data. These reference data are computed with a radiation transport model using AOT and WV information provided by collocated sunphotometer measurements as input. This validation makes use of the same test sites and in-situ data sets used for validation of AOT and WV products (Figure 4). The selected test sites are representative of main surface and atmosphere types. Specifically, the validation procedure consists of:

1. Computation of surface reflectance reference data. For a subset of 10 km x 10 km around the sunphotometer location, AOT, aerosol model and column water vapour derived from sunphotometer measurements are used in a radiative transfer model libRadtran to perform atmospheric correction of S2 L1C TOA reflectance. Only pixels that were not flagged out by the L1 processor and not labelled as cloudy are considered for a pixel-by-pixel comparison of Sen2Cor BOA reflectance product and surface reflection reference computed (Ju et al, 2012, Vermote and Saleous, 2008).
2. Provision of surface reflection in-situ data for BOA reflectance product validation from permanently operating stations and ad-hoc campaigns. This activity encompasses acquisition, quality assurance and archiving of in-situ surface reflectance measurements. So far only one permanently operating station confirmed data provision for L2A-validation. DLR organizes one ad-hoc campaign in Germany per year, with the DLR-Hyspex sensor concomitant to Sentinel-2 overpasses, complemented by sunphotometer and surface reflectance measurements on the ground. Provision of BOA reference data in addition relies on other campaigns coordinated by international parties e.g. ESTEC, EUFAR. Low altitude aircraft flights are considered to minimize atmospheric effects on measurements/data. Then accurate surface reflectance data with high spectral and spatial resolution can be computed from the data takes. Hyperspectral data are convolved to match the S2 spectral channels. High spatial resolution enables upscaling of hyperspectral data to pixel resolution and pixel location of S2 data.

Several additional test sites of ad-hoc experiments planned by other parties for 2015 were identified. Investigators invited DLR to participate or confirmed providing their data. Search for more ad-hoc campaigns and test sites and attempts to get confirmation for data provision of more equipped, permanently operating sites is a running process.

**3.2.3 Cloud Screening & Classification Validation:** The main objectives include verification of the cloud screening/masking and classification accuracy to limit L2A products and final products uncertainties. The Scene Classification Validation is limited due to the lack of “ground truth” data sets. Therefore, validation of classification map relies on the visual inspection, supplemented by the meteorological data, if available. Validation is performed using a set of test sites (Figure 4) representative for a large range of land cover types and weather condition (e.g. sun position, elevation, cloud cover). The Cloud Screening and Scene Classification Validation involves following steps:

1. Selection of a reference database based on a stratified random sampling, in order to guarantee statistical consistency (validity), and avoid exclusion of spatially limited classes from the validation.
2. Visual inspection of the samples using ENVI © software. In case of difficulties in class identification as for example distinction of dark water and shadow pixel, or between cloud and snow pixel, additional tools like TimeSync © software (Cohen et al., 2010) and open data archive Google Earth are involved. Visual inspection is provided for following classes: no data, saturated or defective pixels, dark pixels and clouds shadows, vegetation, bare soils, water, cloud probabilities (low, medium, high), thin cirrus, and snow. Besides of existing 11 class labels, one additional label for non-classified samples is introduced. The Cloud Screening Validation is performed based on the visual inspection of original images and on-ground clouds observations, available from the meteorological stations. Cloud Confidence Quality Indicator (QI) ranges from 0 (clear sky) to 100 (clouds). Using confidence thresholding, probability of clouds is validated based on 3 classes: cloud low probability (confidence below 35%), cloud medium probability (confidence between 35% and 65%), and cloud high probability (confidence above 65%). Additionally, a Snow Confidence Quality Indicator QI is used as a support for snow class validation.
3. Establishment of reference database and comparison with the classification results.
4. Calculation of an error matrix and accuracy statistics.

### 3.3 Time schedule

Sentinel-2 L2A Calibration and Validation have begun in November 2014 with the MPC Set-Up and Deployment Phase (Phase 1) focusing on the definition and description of the ESL L2A Service Design, Communication Plan and the Operational Baseline, as well as preparation and specification of L2A Calibration and Validation activities and plans.

After the S2A launch (scheduled for June 2015), the 9 months MPC Commissioning Phase (Phase 2) follows, subdivided into a 3 months period (Phase E1) ending by the first satellite In-Orbit Commissioning Review, and a 6 months MPC ramp-up phase (Phase E2) preceding the PDGS Routine Operation Phase (Phase 3) (Figure 5).

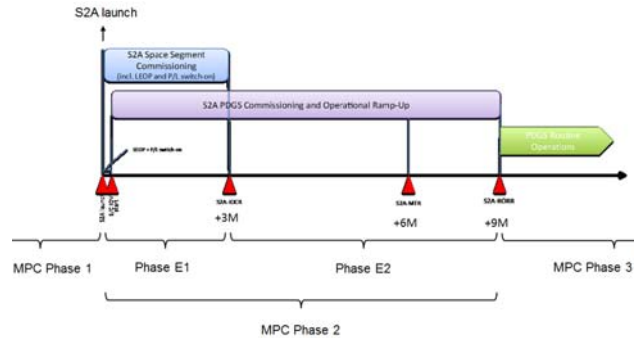


Figure 5. PDGS/MPC Time schedule © ESA

**3.3.1 Phase E1:** Collection and archiving of in-situ data for BOA, AOT and WV product calibration and validation start as soon as MPC/CC informs that Sentinel-2 acquires data over the defined test sites. Processing of the corresponding Sentinel-2 images begins with the release of initial calibrated Level-1C data.

First statistical analysis relies on Aeronet level 1.5 data due to the delay in access to Level 2 Aeronet data from few months to more than 1 year after the date of measurements. Level 1.5 data are automatically cloud cleared but may not have final calibration applied yet. These data are not Quality Assured. The preliminary calibration of the Sen2Cor Cloud Screening and Classification algorithm is performed using the initial calibration dataset. The Cloud Screening and Scene Classification are validated as soon as initial calibrated L2A products are released.

**3.3.2 Phase E2:** Similarly to activities in Phase E1, the objectives are acquisition, quality assurance and archiving of Aeronet data, Microtops and surface reflectance measurements required to perform the Sen2Cor processor and products calibration and validation. Processing of the corresponding satellite data will be continued respectively Sentinel-2 data will be reprocessed with a revised radiometric calibration. As far as available, the Aeronet Level 1.5 data will be replaced by Level 2.0 data. Level 2.0 Aeronet data are pre and post field calibrated, automatically cloud cleared and manually inspected. The full calibration and validation of the Sen2Cor Cloud Screening and Classification algorithm is performed using the calibration and validation datasets enlarged by new acquisitions, considering different solar/viewing angles and various atmospheric conditions.

## 4. INVITATION TO COLLABORATION

In the frame of AC validation, three types of BOA reflectance references are considered:

1. equipped sites with systematic on-ground measurements (e.g. RADCALNET, CEOS and Aeronet)
2. ad-hoc campaigns (own and coordinated by other parties)
3. Landsat BOA products (cross-comparison exercise).

Since all CEOS test sites represent very bright targets (e.g. desert, salt lake, ice), it is unreasonable to include many of them into the L2A-validation. The influence of the atmosphere is smaller for bright targets than for darker surfaces, due to the higher part of directly reflected radiation. Thus, concentrating the validation to bright surfaces would lead to an underestimation of the atmospheric correction uncertainty.

Additionally, applications using atmospherically corrected Sentinel-2 data mainly concern darker targets like vegetation. Therefore, a high quality validation of BOA reflectance product has to be based on dominance of darker surfaces. Until now this is the most critical part of the Sen2Cor BOA validation activities.

To ensure high representativeness and quality of S2 L2A BOA validation, provision of in-situ surface reflectance data over various (including darker) surfaces, both from equipped sites as well as from international cooperation with laboratories or other entities has to be strived.

Since ad-hoc campaigns can be performed at flexibly chosen sites (e.g. vegetation sites, urban sites, or coastal regions), they give good opportunity to fill the gaps in the coverage of surface types within the collection of in-situ data.

Finally, it can be assumed that there are worldwide many institutions performing ad-hoc campaigns for validation of higher-level Sentinel-2 products. These institutions are highly invited to cooperate with ESL L2A team by giving potential of free access to their data suitable for L2A validation.

## REFERENCES

Cohen, W.B., Yang, Z., & Kennedy, R., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync - Tools for calibration and validation. *Remote Sens. Environ.*, 114, 2911-2924.

Ju J, Roy D, Vermote E, Masek J, Kovalskyy V., 2012. Continental-scale validation of MODIS-based and LEDAPS ETM+ Atmospheric correction methods. *Remote Sens. Environ.*, 122 (SI):175-184.

Kaufman, Y. and Sendra, C., 1988. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *International Journal of Remote Sensing*, Volume 9, Issue 8, 1357-1381.

Mayer, B. and Kylling, A., 2005. Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use. *Atmos. Chem. Phys.*, 5, 1855-1877, 2005

Pflug, B., 2013. Ground based measurements of aerosol properties using Microtops instruments. *AIP Conf. Proc.* 1531, 588-591; doi: 10.1063/1.4804838.

Schläpfer, D. et al. (1998). Atmospheric precorrected differential absorption technique to retrieve columnar water vapour. *Remote Sens. Environ.*, 65, 353-366.

Vermote, E. and Saleous, N., 2008. LEDAPS surface reflectance product description. *Version 2.0, January 2007*. Online available: [http://ledaps.nascom.nasa.gov/docs/pdf/SR\\_productdescript\\_dec06.pdf](http://ledaps.nascom.nasa.gov/docs/pdf/SR_productdescript_dec06.pdf)

## APPENDIX

The selection of test-sites for Sentinel-2 L2A Calibration and Validation is described in this appendix. The nomenclature for latitude region (zone) was specified according to the one's from Nasa Goddard for the Landsat global dataset used for ACCA validation ( i.e. Polar > 60°, Boreal > 45°N, MidlatitudeN > 30°N, SubtropicalN >15°N, 15°N < Tropical < 15S and equivalent for southern hemisphere latitudes). Vegetation land cover stands for tundra, prairie, steppe, scrubland or semi-arid. Topography was assigned to flat if elevation varies over the 10km×10km area by less than 100m, and above 100m to rough.

Site name (country)	Zone	Relief	Elev [m asl]	Land cover
<i>Core sites</i>				
Arcachon (France)	Midlat N	flat	11	forest, water, vineyard
ARM_Darwin (Australia)	Tropical	flat	30	wetlands, urban
Banizoumbou (Niger)	Tropical	flat	250	desert, cropland
CEILAP-BA (Argentina)	Midlat S	flat	10	urban, water, vegetation
Easton_MDE (US)	Midlat N	flat	15	forest, cropland, water
Halifax (Canada)	Midlat N	flat	65	urban, vegetation, lakes
Ispra (Italy)	Boreal	rough	235	vegetation, lake, snow
La Paz (Bolivia)	Subtropic S	rough	3439	forest, snow, agriculture
La Parguera (Puerto Rico)	Subtropic N	rough	12	vegetation, coastal
Lahore (Pakistan)	Midlat N	flat	270	urban, irrigated crops
Lake_Lefroy (Australia)	Midlat S	flat	300	forest, sand, rocks
Manaus_EMBR APA (Brazil)	Tropical	flat	115	Forest
SEDE_BOKER (Israel)	Midlat N	rough	480	desert,
Singapore (Singapore)	Tropical	flat	30	urban, vegetation, water
Toravere (Estonia)	Boreal	flat	70	forest, lakes, croplands
Yakutsk (Russia)	Polar N	flat	118	forest, river, snow
<i>Back-up sites</i>				
ARM_Manacapuru (Brazil)	Tropical	flat	115	semi-arid, forest
Barrow (US)	Polar N	flat	0	lakes, snow, grass
Beijing (China)	Midlat N	rough	92	urban, mountains
Capo_Verde (Capo Verde)	Subtropic N	flat	60	desert, ocean
Davos (Switzerland)	Boreal	rough	1596	forest, snow, agriculture
ICIPE_Mbita (Kenya)	Tropical	rough	1125	forest, lake
IMAA_Potenza (Italy)	Midlat N	rough	820	forest, cropland
Pretoria_CSIR-DPSS (South Africa)	Subtropic S	rough	1449	urban, semi-arid

Table 3. Selected core and back-up Aeronet test-sites for L2A Calibration.

Site name (country)	Zone	Relief	Elev [m asl]	Land cover
<i>Core sites</i>				
Yakutsk (Russia)	Polar	rough	92	vegetation, water, snow
Berlin (Germany)	Boreal	flat	36	urban, cropland, forest, water, snow
Potsdam (Germany)	Boreal	flat	47	cropland, forest, urban, water, snow
Rimrock (US)	Boreal	rough	824	cropland, vegetation
Madrid (Spain)	Midlat N	rough	660	urban, vegetation, forest,
GSFC (US)	Midlat N	Flat	64	urban, forest, cropland, water
El Farafra (Egypt)	Subtropic N	flat	92	desert, urban, cropland, forest
Kanpur (India)	Subtropic N	flat	135	cropland, urban, forest, water
Korondal (Phillipines)	Tropical	rough	70	urban, forest, cropland
Alta Floresta (Brazil)	Tropical	flat	277	cropland, urban, forest
Mongu (Zambia)	subtropic S	flat	1040	vegetation, urban
Itajuba (Brazil)	subtropic S	rough	851	vegetation, urban, forest, water
Canberra (Australia)	Midlat S	rough	595	urban, vegetation, water,
Neuquen (Argentina)	Midlat S	rough	271	cropland, urban, water
<i>Back-up sites</i>				
Belsk (Poland)	Boreal	flat	176	cropland, forest, snow
Wytham_Woods (UK)	Boreal	flat	155	cropland, forest, water
Bucharest (Romania)	Midlat N	flat	77	urban, cropland, forest
Sioux Falls (US)	Midlat N	flat	500	vegetation, cropland
XiangHe (China)	Midlat N	flat	36	cropland, urban, forest, water
Guadeloup (Guadeloupe)	Subtropic N	flat	0	forest, urban, cropland, water
Ilorin (Nigeria)	Tropical	flat	381	vegetation
Lope Gabon (Gabon)	Tropical	rough	280	forest
Jabiru (Australia)	Tropical	flat	30	forest, urban, vegetation, water

Table 4. Selected core and back-up Aeronet test-sites for AOT and WV L2A products validation.

Site name (country)	Zone	Relief	Elev [m asl]	Land cover
<i>Instrumented, automatic operating stations (core sites)</i>				
Litchfield Park (Australia)	Tropical	rough	221	vegetation, forest
<i>Instrumented, automatic operating stations (back-up sites)</i>				
Baotou (China)	Midlat N	rough	1307	desert, cropland
Dunhuang (China)	Midlat N	flat	1067	desert, urban
La_Crau (France)	Midlat N	flat	18	vegetation
Railroad valley (US)	Midlat N	flat	1435	desert
<i>Test-sites for ad-hoc campaigns</i>				
Demmin (Germany)	Boreal	flat	9	cropland, forest, urban, water
Potsdam (Germany)	Boreal	flat	47	cropland, forest, urban, water
Bavarian Forest (Germany)	Boreal	rough	982	forest, urban
Kranzberger Forest (Germany)	Boreal	rough	502	cropland, forest, urban
Laegeren (Switzerland)	Boreal	rough	735	forest
Litchfield Park (Australia)	Tropical	rough	221	vegetation, forest

Table 5. Presently identified potential test-sites for surface reflectance measurements.

Site name (country)	Zone	Relief							
			dark features/shadow	vegetation/forest	bare soil/desert	water (dark&bright)	clouds	snow/ice	
<i>Core sites</i>									
Yakutsk (Russia)	Polar	rough	x	x	x	x	x	x	
Potsdam/Berlin (Germany)	Boreal	flat	x	x	x	x	x	x	
Rimrock (US)	Boreal	rough		x	x	x	x	x	
Madrid (Spain)	Midlat N	rough	x	x	x	x	x		
Manila Observatory (Phillipines)	Tropical	rough	x	x	x	x	x		
<i>Back-up sites</i>									
Dunhuang LZU (China)	Midlat N	flat	x	x	x	x	x		
Barrax (Spain)	Midlat N	rough		x	x	x			
Manacapuru/Manaus (Brazil)	Tropical	flat	x	x	x	x	x		
CASLEO (Argentina)	Midlat S	rough	x	x	x	x		x	
Babia_Gora (Poland)	Boreal	rough	x	x		x	x	x	
Demmin (Germany)	Boreal	flat		x	x	x	x	x	

Table 6. Selected core and back-up test-sites for Cloud Screening and Scene Classification Validation.