Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Photogrammetry and Remote Sensing Lab
Prof. Konrad Schindler

Term Paper

# Accuracy and Reliability of Atmospheric Correction for Optical Satellite Images

**Autumn Term 2023**

**Supervised by:**
Prof. Konrad Schindler
Rodrigo Caye Daudt

**Author:**
Tom Lausberg

# Contents

# Acknowledgements

I would like to express my sincere gratitude to Prof. Dr. Konrad Schindler for providing me with the opportunity to undertake this thesis and for the privilege of working in the PRS lab. I extend my heartfelt thanks to Dr. Rodrigo Caye Daudt for his continuous guidance and assistance during the project. His expertise and insights significantly contributed to the development and execution of this research.

# Abstract

The objective of atmospheric correction is to transform raw observations obtained by space-borne earth observation satellites, which measure top-of-atmosphere data, into surface reflectance values on the ground. This thesis aims to assess the accuracy and reliability of atmospheric correction applied to optical satellite images from the Sentinel-2 constellation. The study will focus on the sen2cor algorithm, analyzing its performance under various conditions and with different modified versions. The goal is to comprehensively evaluate the impact of these modifications on the atmospheric correction process.

# Chapter 1

# Introduction

Satellite imagery has evolved into a crucial tool for a large variety of applications, including environmental monitoring, agricultural management, disaster response, urban planning, and climate studies. For instance, it aids in tracking deforestation, monitoring crop health, assessing natural disasters, and studying urban development. As satellite imagery becomes increasingly abundant, ensuring its reliability and consistency becomes vital. One pivotal aspect of post-processing satellite imagery is atmospheric correction. While satellite imagery inherently provides top-of-atmosphere data, most remote sensing applications require bottom-of-atmosphere (BOA) reflectance values. Atmospheric correction algorithms play a pivotal role in this context, aiming to eliminate various atmospheric phenomena such as scattering, absorption, and reflection. Their objective is to provide a robust approximation of BOA reflectance values, thus enhancing the accuracy and reliability of satellite imagery. This paper delves into an exploration of the accuracy and reliability of the sen2cor atmospheric correction algorithm and its impact on Sentinel-2 satellite imagery. Evaluating the effectiveness of such algorithms is crucial for extracting meaningful information from Earth observation data and ensuring its utility in diverse applications. Earth observation imagery from satellites continues to be an indispensable source of data for understanding and addressing various environmental and societal challenges.

## 1.1 Sentinel-2

Sentinel-2 is a constellation of two multi-spectral earth observation satellites which is part of ESA's Copernicus Program. The satellites Sentinel-2A and 2B were both launched into a sun-synchronous polar orbit in 2015 and 2017 respectively. The satellites orbit at a 98.6° inclination and a mean altitude of 786km with a Mean Local Solar Time of 10:30 at the descending node. The orbits are designed to allow the satellites to image all continental land and coastal areas between the 56° southern and 82.8° northern latitudes with a period of 5 days. The imagery is captured in 13 bands in the optical and infrared spectra at 10m, 20m and 60m resolutions [7]. An overview of the bands and their uses can be seen in table 1.1.
The similarities in orbit and imaging instruments allow for the combination of sentinel-2 and landsat data [5]. The Sentinel-2 satellites are design for a lifespan of at least 7.25 years and planned to be replaced by a new generation of satellites from 2024 and 2025 onwards [4].
The Sentinel mission publishes data in three main formats. Level-1B Data contains top-of-atmosphere (TOA) radiances with sensor geometry. It is released in 23km by 25km tiles. The Level-1C product contains TOA reflectance values in cartografic

| Band | Central Wavelength (nm) | Spatial Res. (m) | Description, Use |
|---|---|---|---|
| 2 | 492.4 | 10 | Visible, blue |
| 3 | 559.8 | 10 | Visible, green |
| 4 | 664.6 | 10 | Visible, red |
| 8 | 832.8 | 10 | VNIR, vegetation, shoreline detection |
| 5 | 704.1 | 20 | VNIR, vegetation classification |
| 6 | 740.5 | 20 | VNIR, vegetation classification |
| 7 | 782.8 | 20 | VNIR, vegetation classification |
| 8a | 864.7 | 20 | VNIR, vegetation classification |
| 11 | 1613.7 | 20 | SWIR, soil moisture, snow/cloud differentiation |
| 12 | 2202.4 | 20 | SWIR, soil moisture, snow/cloud differentiation |
| 1 | 442.7 | 60 | Visible, aerosol detection |
| 9 | 945.1 | 60 | VNIR, water vapour detection |
| 10 | 1373.5 | 60 | SWIR, cirrus cloud detection |

***Table 1.1:*** *Sentinel-2 MSI Bands*

geometry. The data is resampled with a constant ground sampling distance of either 10m, 20m or 60m and is released as 100km by 100km tiles in UTM/WGS84 projection. These tiles also contain cloud, cirrus and snow/ice masks. Finally the imagery is compressed with the JPEG2000 standard. The Level-2A product contains bottom-of-atmosphere reflectance values in the same format as the L1C product. It also contains Aerosol Optical Thickness (AOT), Water Vapour (WV) and Scene Classification maps along with the auxiliary data provided with the L1C product. The BOA reflectance and Scene Classification data is computed using the sen2cor atmospheric correction algorithm. In section 1.3 we will go into detail on how this algorithm is structured [7].

## 1.2 Atmospheric Correction Algorithms

The goal an atmospheric correction algorithms are to approximate the radiance at the bottom of the atmosphere with radiance measured by a satellite at the top of atmosphere and the help other auxiliary information. If the radiance at the bottom-of-atmosphere (BOA) is denoted as $I_{BOA}$, the TOA radiance as $I_{TOA}$ and the effect of the atmosphere as

$$\mathcal{F}(I) = \int_{BOA}^{TOA} \frac{dI}{ds} ds \tag{1.1}$$

then the following holds:

$$I_{TOA} = \mathcal{F}(I_{BOA}) \tag{1.2}$$

and hence,

$$I_{BOA} = \mathcal{F}^{-1}(I_{TOA}) \tag{1.3}$$

To be able to accurately calculate the BOA radiance we must thus approximate the effect of the atmosphere in a form which can be efficiently be inverted. To do this, the propagation of visible and infrared light must be modelled. This is done as it travels both from the sun through the atmosphere to the surface and back to the

TOA. As the light propagates it is effected by absorption, emission and scattering. This is described by the radiative transfer equation [2]

$$\frac{dI_\lambda}{ds} = -k_\lambda \rho I_\lambda + k_{\lambda,abs} \rho B_\lambda(T) \tag{1.4}$$

with $I_\lambda$ the intensity of radiation at wavelength $\lambda$, $k_\lambda = k_{\lambda,abs} + k_{\lambda,scat}$ the extinction coefficient comprising of the absorption and scattering components, $\rho$ the density of the gas, $B_\lambda(T)$ the black body radiation at temperature T and $ds$ the thickness of the layer. To simplify this equation, we split the parametriazation into shortwave ($< 4$ microns) and longwave radiation. Shortwave radiative flux can be described with Beer's law

$$\frac{dI_\lambda}{ds} = -k_\lambda \rho I_\lambda. \tag{1.5}$$

The term containing $B_\lambda(T)$ can be omitted, as atmospheric temperatures are too low to emit significant shortwave radiation. The Schwarzschild equation is used to describe longwave radiative flux, as scattering phenomena are negligible for longwave radiation:

$$\frac{dI_\lambda}{ds} = -k_{\lambda,abs} \rho I_\lambda + k_{\lambda,abs} \rho B_\lambda(T) \tag{1.6}$$

## 1.3 Sen2cor

Sen2cor is the atmospheric correction algorithm used for Sentinel-2 products. It was developed by Telespazio VEGA Deutschland GmbH and is available in the Sentinel-2 Toolbox or as a command-line application [10]. It is comprised of five main modules. The Scene Classification, AOT Retrieval and Water Vapour Retrieval create auxillary data needed to perform the Cirrus Correction and TOA→ BOA conversion steps.

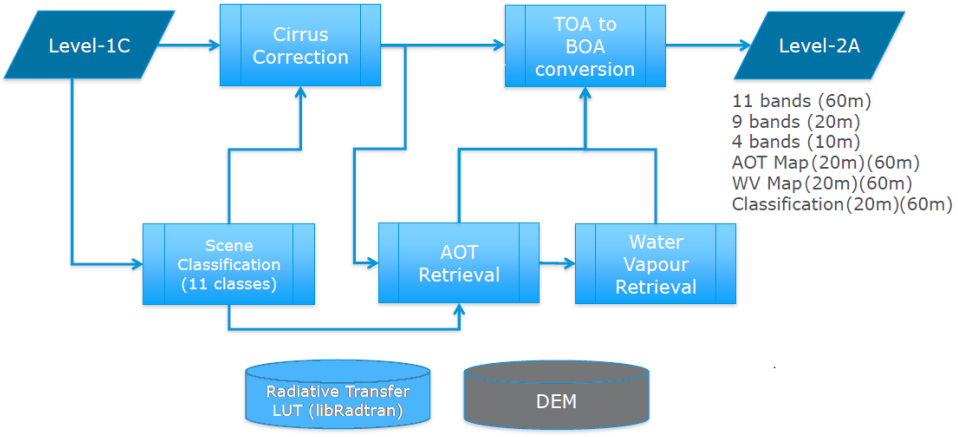


***Figure 1.1:*** *Sen2cor processing chain overview. Source [8]*

It uses the DISORT 8-stream algorithm combined with the correlated k method [11] to discrertize the radiative transfer problem and precalculates the solution for a variaty of atmospheric profiles in a group of look-up tables (LUTs). The LUT

calculations are done with the `libradtran` library [6]. The LUTs are generated with the following variables as seen in the Sen2cor Algorithm Theoretical Basis Document Table 4-I [15]:

| Parameter | Range | Increment / grid points |
|---|---|---|
| solar zenith angle | $0 - 70°$ | $10°$ |
| sensor view angle | $0 - 10°$ | $10°$ |
| relative azimuth angle | $0 - 180°$ | $30°$ ($180°$=backscatter) |
| ground elevation | 0-2.5km | 0.5km |
| visibility | 5-120km | 5,7,10,15,23,40,80,120km |
| water vapour, summer | 0.4-5.5cm | 0.4,1.0,2.0,2.9,4.0,5.0cm |
| water vapour, winter | 0.2 - 1.5cm | 0.2,0.4,0.8,1.1cm |

***Table 1.2:*** *Parameter space for atmospheric correction look-up tables*

Additionally these LUTs are computed for a variety of atmospheric conditions:

- rural / maritime
- atmospheres (mid latitude summer and mid latitude winter)
- 6x ozone concentrations (depending on summer or winter case)
- 6x / 4x water vapour column (depending on summer or winter)

These tables are calculated at a high spectral resolution of $0.6nm$ which is then resampled to the spectral response of the sentinel-2 imaging instruments [15].

## Scene Classification

The scene classification algorithm (SCL) is the first module in the sen2cor algorithm. It uses the spectral data from all bands to segregate the image into 11 classes. These are

| Label | Classification |
|---|---|
| 0 | NO_DATA |
| 1 | SATURATED_OR_DEFECTIVE |
| 2 | CASTED_SHADOWS |
| 3 | CLOUD_SHADOWS |
| 4 | VEGETATION |
| 5 | NOT_VEGETATED |
| 6 | WATER |
| 7 | UNCLASSIFIED |
| 8 | CLOUD_MEDIUM_PROBABILITY |
| 9 | CLOUD_HIGH_PROBABILITY |
| 10 | THIN_CIRRUS |
| 11 | SNOW |

***Table 1.3:*** *List of classes. Source [7]*

The SCL uses a series of threshold tests on different bands and indices to apply this segregation [7]. In a first step, the SCL applies cloud and snow detection. Cloud-free pixel are detected with a threshold in the red spectral band (band 4). It then

uses the Normalised Difference Snow Index (NDSI) to differentiate between clouds and snow. This is possible because clouds absorbe both SWIR and VIR, whereas snow reflects the VIR in band 3 [15].

$$NDSI = \frac{\text{Band } 3 - \text{Band } 11}{\text{Band } 3 + \text{Band } 11} \tag{1.7}$$

A snow confidence mask is then build by applying further thresholds on bands 2,3,8 and 11. To differentiate highly reflective vegetation from snow and clouds, the Normalized Difference Vegetation Index is used.

$$NDVI = \frac{\text{Band } 8 - \text{Band } 4}{\text{Band } 8 + \text{Band } 4} \tag{1.8}$$

Bare soil and bright water pixels are detected using the ratio of bands 2 and 11. To differentiate rocks and sands the ratio of bands 8 and 11 are used. Rocks tend to have a higher reflectance in band 11 than band 8. For clouds, the opposite holds. An optional spacial filtering step is then applied. This corrects the slight error caused by the assumption that all features are located at ground level, which leads to errors for clouds boundaries.
The detection of cirrus cloud relies heavily on the measurements in band 10. This is specifically tuned to measure water vapor absorbtion at $1.38\mu m$. If the reflectance of the pixel is between the clear-sky threshold of 0.012 and thick cloud threshold of 0.035, then the pixel is counted as cirrus cloud.
A simple neural network based on a Kohonen map algorithm is used to calculate radiometrically probable cloud shadows. This is then multiplied with a geometrically probable cloud shadows calculated using a digital elevation model (DEM), the final cloud masks, sun position and cloud height distributions. From sen2cor version 2.10 onwards the `L2A_SceneClass_evolution` class is used by default. This includes improvements to the detection of casted shadows and snow cover based on ESA CCI snow monthly climatology. The final SCL mask is generated using various masks and quality/probability indicators. Sen2cors SCL is not designed to perform a detailed scene classification, rather a simple segregation to allow for efficient atmospheric corrections and further analysis by third parties [10].
A more detailed description of the SCL algorithm which includes quantities analysis of the individual steps and extact numerical values for all thresholds can be found in the Level-2A Algorithm Theoretical Basis Document Section 3.2 [15].

### Atmospheric Correction

To perform the atmospheric correction of the Level-1C images, atmospheric optical thickness (AOT) and Water Vapour (WV) maps must first be generated. AOT gives a measure for visual transparency of the atmosphere. It is estimated using the Dense Dark Vegetation (DDV) Algorithm. This uses band 12 (SWIR), band 2 (blue) and band 4 (red) which are compared to reference values found in dense dark vegetation or soil. If the image contains less than 1% suitable reference pixels, it can fall back to auxillary Copernicus Atmosphere Monitoring Service (CAMS) AOT data. If no CAMS data is supplied, a default starting value is taken from the Ground Image Processing Parameter (GIPP) File. The water vapour is approximated using the Atmospheric Pre-corrected Differential Absorption (ADPA) algorithm. This uses band 8a to get a baseline measurement, which is then compared to the reflectance of band 9. The values of bands 8a and 9 are assumed to be the same in the case of no WV [7].
Before the TOA to BOA conversion is applied, the effect of cirrus clouds can be removed using the cirrus cloud mask produced in SCL section. The atmospheric

correction is the performed pixelwise for each of the 12 bands. A digital elevation model can be used to model lambert's relfactance law for areas that are not flat.

# Chapter 2

# Code

## 2.1 Modification of the Sen2cor Algorithm

Sen2cor is structured as a Python-based software package which is used in the generation of Level-2A (L2A) products in the Copernicus Data Space Ecosystem. It is also available as a standalone package here [3]. Sen2cor is packaged as a standalone Python environment, complete with essential dependencies such as GDAL, OSGEO and other required Python libraries. The core codebase is encapsulated in the `lib/python2.7/site-packages/sen2cor` (linux) or `Lib/site-packages/sen2cor` (windows) directories. Within these directories, one can find essential components like Look-Up Tables (LUTs) in the `lib_S2A/B` directories, definitions outlining file structures for the various processing baselines and auxiliary data such as global snow maps. Notably, Sen2cor follows an open-source model for most components, with the exception of the Atmospheric correction module, which is available as precompiled `.pyd` or `.so` dynamically linked libraries.
The original sen2cor codebase was adapted to enable analysis and control of the Sen2cor module. Modifications were made to facilitate debugging and enhance comprehension of the Atmospheric Correction (AC) module. Specifically, alterations were introduced to the L2A_Tables class, a component crucial to the AC module for reading and writing data. By adjusting existing functions in the L2A_Tables class , it was possible to inject code which is executed by the AC module. These functions provide basic information about which tables are used by the AC and how it changes the data. Attempts were also made to decompile .so files, albeit yielding inconclusive information. Modifications were incorporated into the Scene Classification (SCL) module, enabling manual manipulation of data. This functionality allows for the assessment of the impact of different components of the algorithm on final results. All these modifications are controlled by environment variables, ensuring simple and consistent communication communication between the Python environment conducting the analysis (Python 3.10+) and the Sen2cor Python 2.7 environment. The modified Sen2cor code, along with associated analysis software, is available on GitHub via the provided link. [9]

### 2.1.1 Copernicus Data Space Ecosystem Migration

At the end of October 2023 ESA transitioned the dissemination of Sentinel-2 data from the "Copernicus Open Access Hub" to the "Copernicus Data Space Ecosystem". This new system is an open standard, facilitating integration with diverse data sources and leveraging increased public cloud resources. Because of this transition, legacy tools like the `sentinelsat` package [1] have become obsolete.

To address this, new tools were developed to efficiently access the upgraded service. Principally, a simple client was created to interact with the Copernicus Data Space OData API. This client not only manages authentication and tokens but also streamlines the processes of product search and batch downloads.

## 2.2 Sen2cor Data Management and Analysis Tools

To gain insights into the functionality of the sen2cor algorithm, a series of experiments were conducted, involving systematic modifications to sen2cor across various test locations. Given the multitude of potential permutations arising from these variables, the development of a dedicated package became necessary to manage the resulting dataset effectively. In response to this necessity, the `l2a_analysis.py` script defines three main classes. The `L2A_Band` and `L2A_Band_Stack` classes serve as dataclasses, encapsulating information about the location and metadata of individual bands or sets of bands. This structuring enables systematic data access and manages memory usage, a critical consideration when dealing with expansive datasets. The central component, the `L2A_Analysis` class, assumes a pivotal role within the module. Its functionalities include initializing data locations, orchestrating combinations of sen2cor modifications and test locations, executing scene classification and atmospheric correction (AC) for each permutation, organising to the newly processed L2A data for subsequent analysis, and preserving metadata for each run in a designated directory to ensure possible retrieval at a later time. The report directory created by the L2A_Analysis class has the following structure:

```
REPORT_DIR
├── location_1
│   ├── modification_1
│   │   └── example.SAFE
│   ├── modification_2
│   └── ...
├── location_2
│   └── ...
└── data_info.json
```

The main report directory contains a set of locations and a `data_info.json` file. The json file contains all of the metadata from the creation of the L2A products. Each location also contains a set of modifications. Each modification directory contains the `.SAFE` file for that location and modification.

### Visualisation Tools

For the visualization and analysis of various versions of L2A data, a set of utility functions has been established within `helpers.py`. These functions perform computations on the reflectance of individual bands, providing statistical metrics such as mean, standard deviation, minimum, and maximum values.
Additionally, these utility functions enable the generation of visual representations, such as true-color and false-color RGB images. Notably, key functions within this set allow for plotting of histograms illustrating the differences between the modified sen2cor algorithm and a reference version. This feature is crucial for assessing the impact of algorithmic adjustments.
Moreover, the utility functions extend to the creation of comparison matrices, offering a systematic means to evaluate disparities between various modifications. These functionalities collectively enhance the capacity to visually and quantitatively assess the outcomes of different sen2cor algorithm versions, providing valuable insights into their performance characteristics.

## 2.3 Building Detection Algorithm

To evaluate the differences in modified versions of the sen2cor algorithm, an illustrative example involves employing a straightforward building detection algorithm on the SpaceNet 7 (SN7) [12] dataset. The dataset consists of 100 locations across the globe with each containing 4km x 4km images and hand-labeled buildings. Detailed information on the SN7 dataset is available in Section 3.3. The `copernicus_dataset.ipynb` notebook serves to construct the dataset on which the building detection algorithm is executed.
The process begins by creating a directory for each location within the SN7 dataset. Subsequently, the first set of building labels for each location is selected and added to the respective directory. The CRS and extent of each label set is then extracted to calculate the footprint for a specific location.
Utilizing the Copernicus Data Space Ecosystem client, as detailed in Section 2.1.1, L1C products are found and downloaded for each location. Locations lacking cloudless L1C images for the designated time period are excluded.
Next, a list of modifications to be applied to the sen2cor algorithm is defined. The `L2A_Analysis` class is then utilized to execute each run of sen2cor for every location. Following this, the `create_dataset` function is executed. This function samples images from various locations and a specific modification and incorporates the corresponding SN7 building labels. The resultant dataset can be exported as a .csv file, offering a structured representation of the imagery and associated building labels for model training.

# Chapter 3

# Analysis

## 3.1 Modifcations

Modifications to the sen2cor algorithm fall into three primary categories. Firstly, there are Lookup Table (LUT) modifications, in which lookup tables utilized for atmospheric correction. A comprehensive discussion on these LUTs can be found in Section 1.3.
The second category involves various flags that can be set for the sen2cor algorithm. Notably, the Water Vapor and Cirrus Correction flags are included, which toggle these corrections on or off. Additionally, the Bidirectional Reflectance Distribution Function (BRDF) correction is an optional adjustment addressing reflection off sloped terrain. It encompasses different modes where $G$ is the correction factor, $\beta_i$ the local solar zenith angle, $\beta_T$ the : threshold for surface reflectance and $g$ the lower bound of the correction factor (default 0.22) [11]:

$$0 : \text{(no correction)} \tag{3.1}$$

$$1 : G = \frac{cos(\beta_i)}{cos(\beta_T)} \leq g \tag{3.2}$$

$$2 : G = \sqrt{\frac{cos(\beta_i)}{cos(\beta_T)}} \leq g \tag{3.3}$$

$$\tag{3.4}$$

The third category comprises modifications related to scene classification. This involves the use of a flag to override the Scene Classification (SCL) mask generated during the scene classification process and set it to a singular class. This modification allows for an assessment of the impact of SCL on the Atmospheric Correction (AC).

## 3.2 Results

### 3.2.1 Locations

#### Algeria

The modifications to the sen2cor algorithm were exemplified through a focused analysis on three key locations, with the initial focus on Algeria. Algeria was chosen due to its characteristic dry atmosphere, providing an ideal environment for assessing atmospheric correction algorithms. The simplicity of the scene classification in this region, primarily characterized by bare soil, facilitated a clear evaluation of the algorithmic modifications. The image is located at UTM tile `T30RXT`.

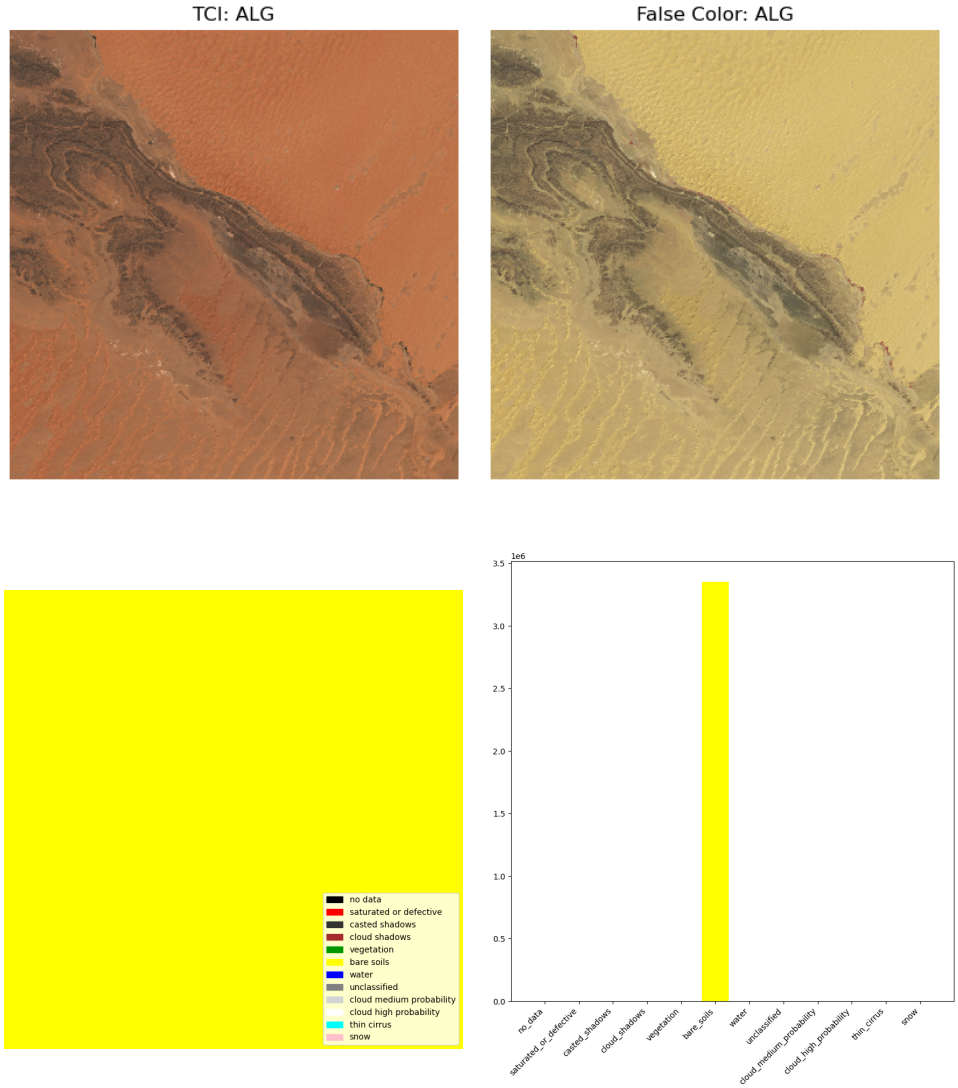


***Figure 3.1:*** *Algeria, Top left: TCI, Top Right: False Color with B8a, B4, B3, Bottom Left: SCL Mask, Bottom Right: distribution of SCL classes*

### Switzerland

The second area chosen for examining modifications to the sen2cor algorithm is situated in the Swiss Alps. This region was selected for its diverse and challenging topography, featuring a mix of snow-covered areas, water bodies, and varied vegetation. The Swiss Alps present a contrasting environment to the initial Algerian location, introducing complexities related to different surface types. Moreover, the Swiss Alps are characterized by a mid latitude atmosphere with the presence of cirrus clouds. This atmospheric condition adds an additional layer of complexity for atmospheric correction algorithms. By focusing on this region, the analysis aims to assess the adaptability of sen2cor modifications to diverse terrains and atmospheric conditions, providing valuable insights into the algorithm's performance in more complex geographical settings. The image is located at UTM tile `T32TMS`.

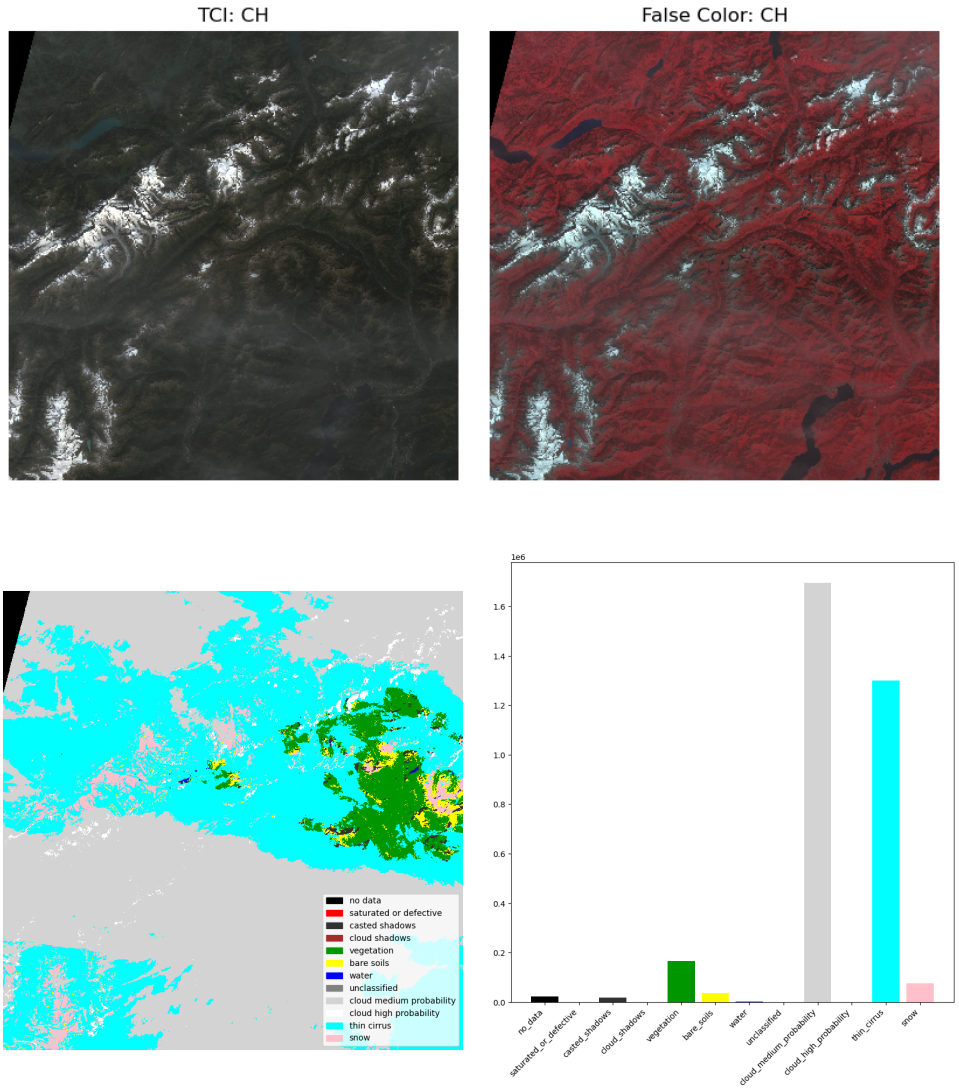


***Figure 3.2:*** *Switzerland, Top left: TCI, Top Right: False Color with B8a, B4, B3, Bottom Left: SCL Mask, Bottom Right: distribution of SCL classes*

This image proves much harder for the scene classification to effectively segment

and highlight a major shortfall of the SCL algorithm. This is because large portions of the image contain trace amounts of low-altitude and cirrus clouds. Because these are segmented in the same mask as the classes which describe the ground conditions (i.e snow, vegetation, bare soils), much of the information is obscured by the cloud classes. Where the image is deemed cloudless, the SCL algorithm can discern the different terrain classes with some accuracy.

**Brazil**

The final location selected for evaluating modifications to the sen2cor algorithm is situated in the Brazilian Amazon rainforest. This region stands out for its highly moist atmosphere and dense vegetation cover, offering a stark contrast to the previous locations. The Amazon rainforest environment introduces specific challenges related to abundant vegetation, with patches of bare soil interspersed throughout. Additionally, the presence of low-altitude clouds in the Brazilian Amazon further complicates atmospheric correction procedures. This region serves as a critical test case for sen2cor modifications in environments with high humidity, extensive vegetation, and prevalent low-altitude clouds. The image is located at UTM tile `T21MTS`.

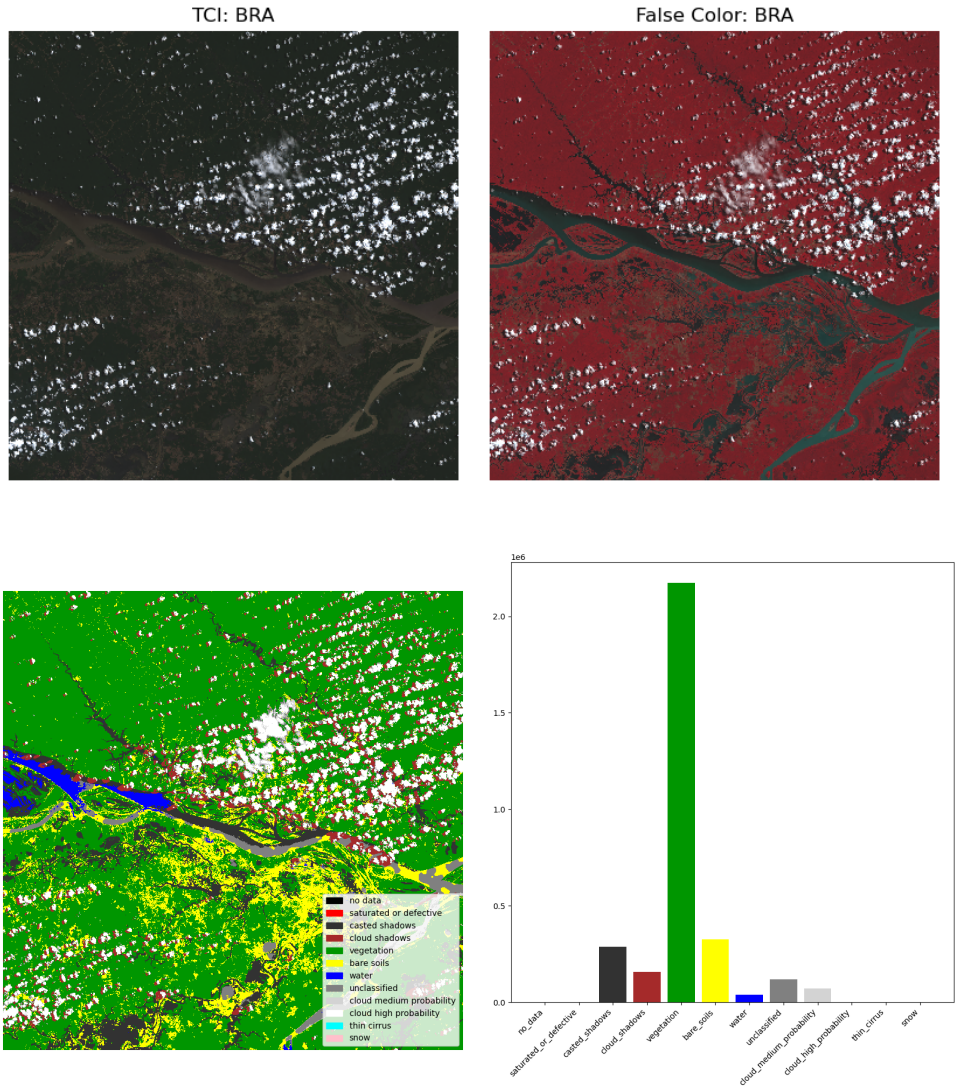


***Figure 3.3:*** *Brazil, Top left: TCI, Top Right: False Color with B8a, B4, B3, Bottom Left: SCL Mask, Bottom Right: distribution of SCL classes*

The image above provides insights into the strengths and limitations of the SCL algorithm in a distinct scenario. Noteworthy strengths of the algorithm include its effectiveness in:

- Distinguishing between vegetation and bare soil,
- Identifying cloud shadows and low-altitude clouds,
- Recognizing certain areas of the Amazon River.

However, the algorithm exhibits limitations in:

- Incomplete classification of bare soil,
- Inaccuracies in water classification, occasionally misinterpreting it,
- Occasional confusion between water and bare soil due to sediment presence,
- Misclassification of dark areas of water as cast shadows.

### 3.2.2 Flags

#### Water Vapour Correction

Figure 3.4 shows the effect of the water vapour correction on the L2A product in the three locations described above. The values plotted in the histogram are the pixel-wise difference between the L2A product without the WV correction and with. Hence, positive values denote pixel made brighter by WV correction.

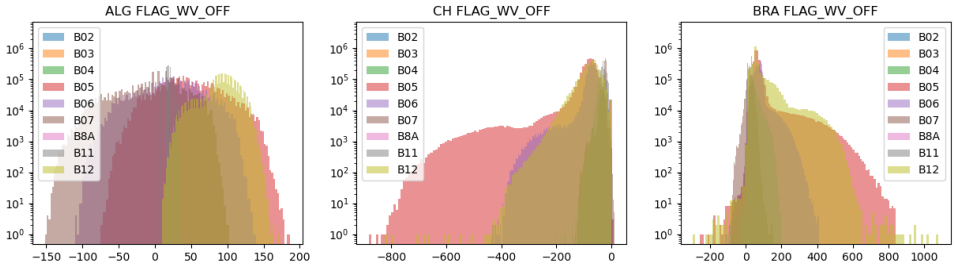


***Figure 3.4:*** *Water vapour correction*

The impact of Water Vapor (WV) correction is evidently correlated with the moisture content in the respective locations. While the y-axis remains consistent across all three plots, it is crucial to note the varying x-axis scales, indicating differences in the specific effects observed. Interestingly, the WV correction tends to result in a brighter representation for the Brazilian Amazon (BRA) location but a darker representation for the Swiss Alps (CH) location. Moreover, it is noteworthy that the effect of WV correction is more pronounced in the Infrared (IR) bands compared to the visible ones. This observation underscores the algorithm's sensitivity to moisture content in the atmosphere, particularly in the IR spectrum

#### Cirrus Correction

Figure 3.5 shows the effect of the cirrus cloud correction on the L2A product in the three locations described above. Contrary to the example above, the values plotted show the effect of adding the cirrus correction rather than removing it. This is because the default for the WV correction is to be turned on. For the cirrus correction, the opposite holds.

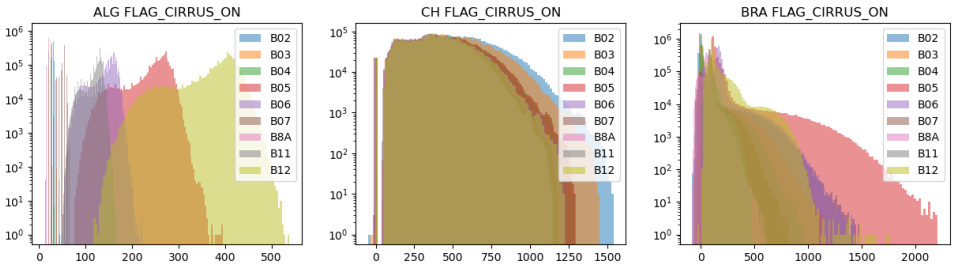


***Figure 3.5:*** *Cirrus correction*

| | std | mean | max | min |
|---|---|---|---|---|
| ALG | 17.42 | 115.26 | 170.89 | 41.00 |
| CH | 205.03 | 421.79 | 1,278.33 | -16.22 |
| BRA | 94.32 | 88.07 | 1,226.33 | -48.22 |

***Table 3.1:*** *Cirrus correction: difference in reflectance averaged over all bands*

The impact of Cirrus Correction is observed to vary across locations and bands. In the Algerian (ALG) example, the effect is mainly noticeable in the infrared bands. Conversely, in the Swiss Alps, there is a substantial effect on visible bands, indicating that cirrus correction has a more pronounced influence on these bands in this specific location. In the Brazilian Amazon, the cirrus correction notably affects Band 5. A consistent trend across all locations is the darkening of images with cirrus correction. This outcome aligns with expectations, as cirrus clouds tend to reflect more light than the underlying ground, and the correction aims to mitigate this reflective influence.

#### BRDF Correction & Default Visibility

For the BRDF Correction and the default visibility flags, are no differences between the reference L2A product and the modified versions. The default visibility has no effect on the atmospheric correction as the algorithm is always able to detect DDV pixels for AOT estimation. In the investigation of various test locations, no scenarios were found in which the algorithm fell back default visibility values. The BRDF correction has no effect on the default version as no digital elevation model is supplied.

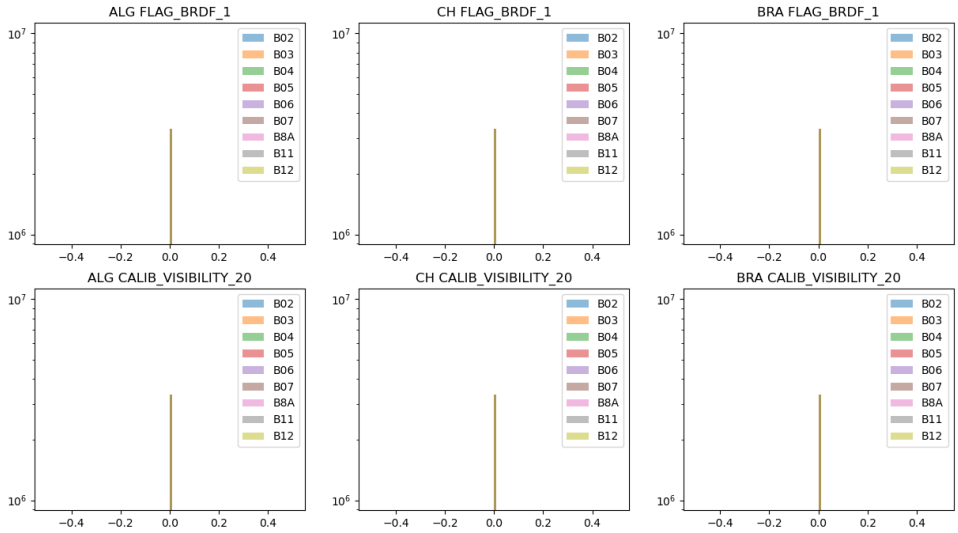


***Figure 3.6:*** *BRDF correction & default visibility flags*

### 3.2.3 Look-Up Tables

In the various setups seen in figure 3.7, the default Lookup Table (LUT) parameters were utilized, with only one variable being changed. For example, in the LUT_AEROSOL_MARITIME plots, the mid-latitude summer configuration and an ozone concentration of 331 were applied. It is important to note that the standard deviation observed in the images generated with LUTs is smaller compared to those

with Cirrus and Water Vapor (WV) corrections, as highlighted in Table 3.2.

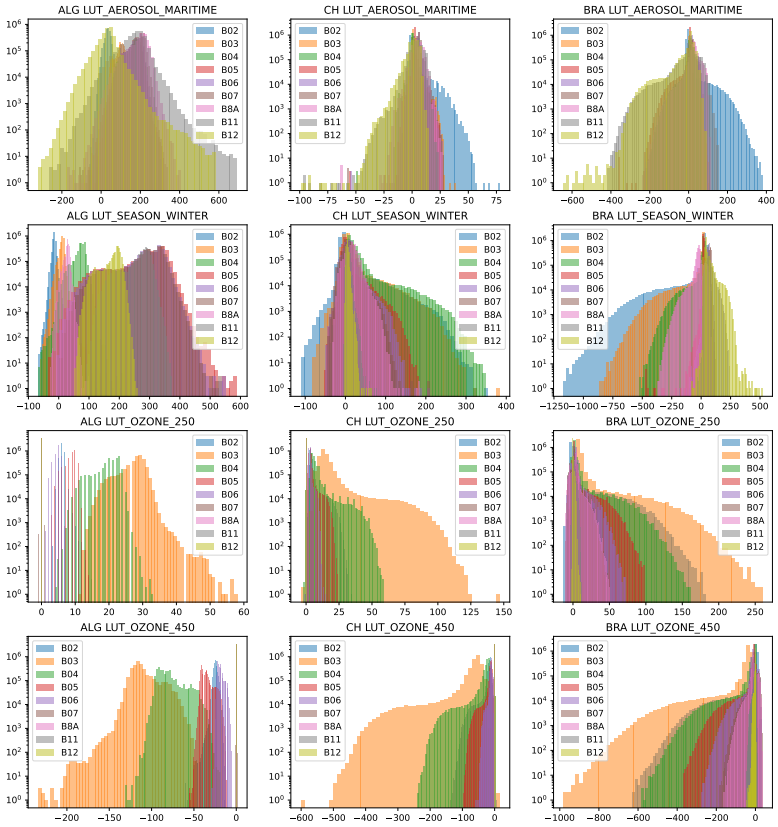


***Figure 3.7:*** *Look-Up Tables*

| | ALG | CH | BRA |
|---|---|---|---|
| LUT_AEROSOL_MARITIME | 36.86 | 3.16 | 27.83 |
| LUT_SEASON_WINTER | 35.25 | 16.53 | 39.91 |
| LUT_OZONE_250 | 1.12 | 2.59 | 8.52 |
| LUT_OZONE_450 | 4.36 | 10.43 | 31.07 |

***Table 3.2:*** *Look-Up Tables: standard deviation*

### 3.2.4 Scene Classification

To investigate the impact of scene classification, a manual adjustment was made by setting the Scene Classification (SCL) mask to a singular value. The thin-cirrus, medium-probability clouds and high-probability cloud classes revealed similar effects, with both casted and cloud shadows also being treated similarly. Across

various locations, the effects of these modifications were largely consistent. However, the Algerian location exhibited intriguing effects. Given that the unmodified SCL mask categorized the entire area as bare soil, the manual modification to a NOT_VEGETATED did not show any discernible difference. Interestingly, figure 3.8 shows the algorithm treated snow and vegetated earth the same as bare soil in the atmospheric correction process.

| | ALG | CH | BRA |
|---|---|---|---|
| class_NO_DATA | 165.32 | 161.36 | 344.79 |
| class_CLOUD_SHADOWS | 42.09 | 100.72 | 155.90 |
| class_VEGETATION | 0.00 | 125.82 | 130.44 |
| class_NOT_VEGETATED | 0.00 | 90.01 | 143.26 |
| class_WATER | 42.09 | 100.72 | 155.90 |
| class_CLOUD_HIGH_PROBABILITY | 92.86 | 115.14 | 233.71 |
| class_SNOW | 0.00 | 90.01 | 120.30 |

***Table 3.3:*** *Scene classification: standard deviation*

## 3.3 Building Detection Test Case

To examine the impact of modifications, a straightforward building detection algorithm was developed. Employing a cross-validated grid search from the scikit-learn [13] package, optimal parameters for a support vector classifier (SVC) were identified. The best-performing classifier utilized a radial basis function kernel with a regularization parameter $C$ set to 1000. The gamma value was configured as 'scale', calculated as the inverse of the number of features multiplied by the variance of the training data. Datasets were crafted for each set of modified sen2cor products and the SVC with the specified parameters was trained on each dataset. The resulting accuarcy, precision, recall, f1-score and mean-squared error for each modification are presented in Table 3.4.

| **Modification** | **Acc.** | **Prec.** | **Recall** | **F1-score** | **MSE** |
|---|---|---|---|---|---|
| reference | 0.74 | 0.79 | 0.74 | 0.75 | 0.26 |
| class_VEGETATION | 0.73 | 0.79 | 0.73 | 0.74 | 0.27 |
| class_NO_DATA | 0.72 | 0.77 | 0.72 | 0.73 | 0.28 |
| class_CLOUD_HIGH_PROBABILITY | 0.74 | 0.79 | 0.74 | 0.75 | 0.26 |
| FLAG_WV_OFF | 0.68 | 0.76 | 0.68 | 0.70 | 0.32 |
| FLAG_CIRRUS_ON | 0.71 | 0.76 | 0.71 | 0.72 | 0.29 |
| LUT_AEROSOL_MARITIME | 0.72 | 0.78 | 0.72 | 0.73 | 0.28 |
| LUT_OZONE_250 | 0.72 | 0.78 | 0.72 | 0.73 | 0.28 |
| LUT_OZONE_450 | 0.74 | 0.79 | 0.74 | 0.75 | 0.26 |
| LUT_SEASON_WINTER | 0.73 | 0.78 | 0.73 | 0.74 | 0.28 |

***Table 3.4:*** *SVC Metrics*

The metrics indicate that, in this case, the modifications had no significant impact on the performance of the building detection algorithm. Unsurprisingly, the reference sen2cor configuration consistently outperformed the modified versions across all evaluated metrics. Notably, turning off the water vapor correction had the most substantial effect on the results. However, the overall error is too high to draw clear and definitive conclusions from these observations.

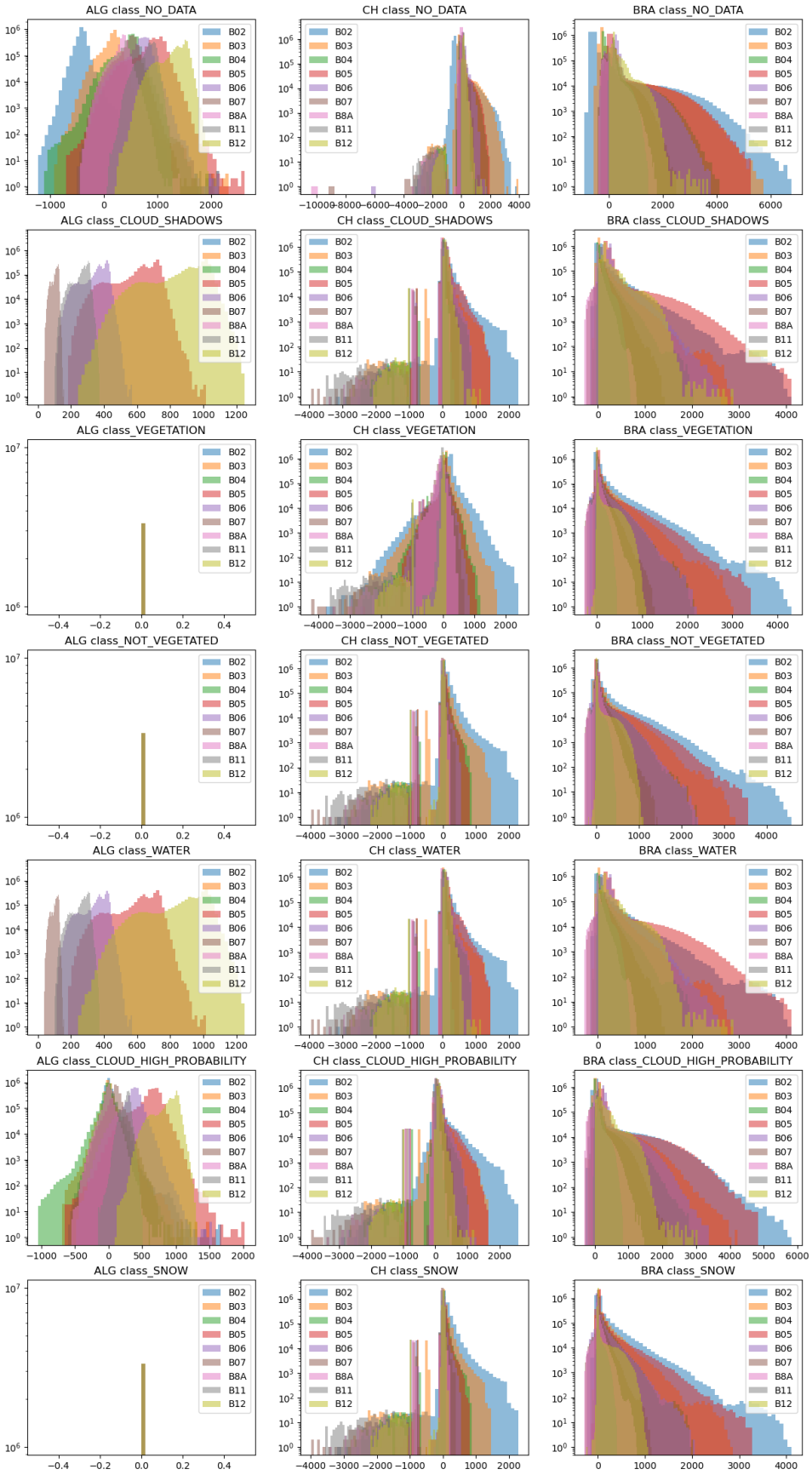


***Figure 3.8:*** *Scene Classification*

# Chapter 4

# Conclusion

Sen2cor is an algorithm designed to derive bottom of atmosphere estimates and auxiliary data from Sentinel-2 imagery by implementing scene classification and atmospheric correction algorithms. Its robustness has been rigorously tested and validated by the European Space Agency [16].
The conversion of BOA values poses a fundamentally challenging problem, as obtaining large-scale reference values is not practically feasible. Maintaining consistency in the L2A product, both spatially and temporally, is of paramount importance. Furthermore, compatibility with other Earth Observation (EO) constellations, such as Landsat, necessitates a degree of conservatism in sen2cor's approach.
While sen2cor employs threshold-based methods for the scene classification algorithm, alternative approaches have been proposed in the literature, as evidenced by [14]. These approaches offer a noteworthy improvement using convolutional neural networks, showing a Micro-F1 score enhancement from 0.59 to 0.84. Whilst these improvements could have a large impact on the atmospheric correction of misclassified data, such methods have not been integrated into sen2cor. The reason for this could be the simplicity of the threshold-based SC algorithm currently employed.
In a case study, a simple building detection algorithm based on the labels of the space net 7 dataset [12] was used to compare the L2A products created by various modified sen2cor algorithms. Unfortunately, the accuracy of this building detection algorithm was not high enough to make conclusive statements about the quality of the L2A input products. This may by due to the lower resolution of the sentinel-2 imagery, which is not able to resolve of building in the dataset or due to the simplicity of the model employed in the building detection algorithm.
Due to the closed-source nature of the atmospheric correction code in sen2cor, the rapid testing of modifications to this core component proved challenging. Consequently, the focus shifted to elements of the algorithm outside the AC core or those exposed by flags. Notably, modifications related to cirrus and water vapor correction showed the most substantial impact on the L2A product. Manually choosing lookup tables used by the AC also had a measurable impact on the L2A product. This approach likely improves the quality of local images but comes with the downside that L2A correction is less consistent over large geographic areas.
Additionally, the scene classification emerged as a significant factor influencing AC. Given its open-source nature, scene classification proved more accessible for modification and improvement. Furthermore, the scene classification algorithm is easier calibrate and validate, as obtaining ground truth data for scene classification is comparatively more feasible.

# Bibliography


[1] Sentinelsat. URL: pypi.org/project/sentinelsat/.

[2] Chapter 7 - application of radiative transfer principles to remote sensing. In K.N. Liou, editor, *An Introduction to Atmospheric Radiation*, volume 84 of *International Geophysics*, pages 348–441. Academic Press, 2002. URL: https://www.sciencedirect.com/science/article/pii/S0074614202800225, doi:10.1016/S0074-6142(02)80022-5.

[3] European Space Agency. Sen2cor 2.11. URL: step.esa.int/main/snap-supported-plugins/sen2cor/sen2cor-v2-11/.

[4] European Space Agency. Gearing up for third sentinel-2 satellite, Accessed: 12.01.2024. URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Gearing_up_for_third_Sentinel-2_satellite.

[5] Martin Claverie, Junchang Ju, Jeffrey G. Masek, Jennifer L. Dungan, Eric F. Vermote, Jean-Claude Roger, Sergii V. Skakun, and Christopher Justice. The harmonized landsat and sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219:145–161, 2018. URL: https://www.sciencedirect.com/science/article/pii/S0034425718304139, doi:10.1016/j.rse.2018.09.002.

[6] C. Emde, R. Buras-Schnell, A. Kylling, B. Mayer, J. Gasteiger, U. Hamann, J. Kylling, B. Richter, C. Pause, T. Dowling, and L. Bugliaro. The libradtran software package for radiative transfer calculations (version 2.0.1). *Geoscientific Model Development*, 9(5):1647–1672, 2016. URL: https://gmd.copernicus.org/articles/9/1647/2016/, doi:10.5194/gmd-9-1647-2016.

[7] European Space Agency. *Sentinel-2 User Handbook*, 2015. URL: https://sentinel.esa.int/documents/247904/685211/sentinel-2_user_handbook.

[8] Ferran Gascon, Catherine Bouzinac, Olivier Thepaut, Mathieu Jung, Benjamin Francesconi, Jerome Louis, Vincent Lonjou, Bruno Lafrance, Stéphane Massera, Angélique Gaudel-Vacaresse, Florie Languille, Bahjat Alhammoud, Françoise Viallefont, Bringfried Pflug, Jakub Bieniarz, Sebastien Clerc, Laëtitia Pessiot, Thierry Trémas, Enrico Cadau, and Valérie Fernandez. Copernicus sentinel-2a calibration and products validation status. *Remote Sensing*, 8, 06 2017. doi:10.3390/rs9060584.

[9] Tom Lausberg. sen2cor analysis, 2023. URL: github.com/tomlausberg/sen2cor_analysis.

[10] Magdalena Main-Knorn, Bringfried Pflug, Jerome Louis, Vincent Debaecker, Uwe Müller-Wilm, and Ferran Gascon. Sen2Cor for Sentinel-2. In Lorenzo

Bruzzone, editor, *Image and Signal Processing for Remote Sensing XXIII*, volume 10427, page 1042704. International Society for Optics and Photonics, SPIE, 2017. doi:10.1117/12.2278218.

[11] Uwe Müller-Wilm. *[S2-PDGS-MPC-L2A-IODD] S2 Level 2A Input Output Data Definition 2016*. European Space Agency, 2016.

[12] SpaceNet on Amazon Web Services (AWS). "datasets." the spacenet catalog, Last modified October 1st, 2018. Accessed on January 21st, 2024. URL: https://spacenet.ai/datasets/.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[14] Kashyap Raiyani, Teresa Gonçalves, Luís Rato, Pedro Salgueiro, and José R. Marques da Silva. Sentinel-2 image scene classification: A comparison between sen2cor and a machine learning approach. *Remote Sensing*, 13(2), 2021. URL: https://www.mdpi.com/2072-4292/13/2/300, doi:10.3390/rs13020300.

[15] J. Louis & L2A team. *[S2-PDGS-MPC-ATBD-L2A] Sentinel-2 Level-2A ATBD 2021*. European Space Agency, 2021.

[16] S2 MSI ESL team. *Sentinel-2 Annual Performance Report – Year 2022*. European Space Agency, 2023. URL: sentinels.copernicus.eu/documents/247904/4893455/OMPC.CS.APR.001+-+i1r0+-+S2+MSI+Annual+Performance+Report+2022.pdf.