Copernicus Master in Digital Earth

Student: Emanuel Goulart Farias

Student number: 12413874

Big Earth Data Concepts

Assignment 1: ARD and Level-2 data.

1. Analysis-ready-data

1.1. Why satellite data provider found it necessary to investigate how ARD can be created and delivered to users.

Analysis-ready data (ARD) has become a must-do approach in the field of Earth Observation (EO). With the advancements in big data and the exponential increase in data flow, it is inefficient to repeat unnecessary processing. For the continuous advancement of EO, users need to engage in data analysis rather than repeatedly pre-processing data. The barrier, where EO users typically invest a large proportion of their effort into data preparation, is replaced with the concept of ARD to ensure the success of space-based data and major global initiatives.

As we have observed, the focus on time-series data in EO field has advanced rapidly because a significant amount of information can be extracted from these analytics. However, different satellite providers have their own technical specifications for each sensor, with varying parameters. A common approach to make this data comparable between different image collections relies on the standardization of ARD. In other words, ARD standardizes and make it possible to compare data from different satellite providers to assure quality, correctness, and controllable sources of error. This ensures that when users compare time-series data, they are comparing the same requirements and observing changes at a pixel scale, such as ground and water, rather than changes due to different data processing methods like cloud cover interference.

1.2. Name and explain some common characteristics of the three different ARD definitions / approaches.

CEOS – Committee of Earth Observing Satellites

The Committee of Earth Observing Satellites (CEOS) is attempting to organize international government imagery providers under the product family frameworks of its CEOS Analysis Ready Data for Land initiative. CEOS ensures international coordination of civil space-based Earth observation programs and promotes the exchange of data to optimize societal benefits and inform decision-making for securing a prosperous and sustainable future for humankind.

The core elements under the family frameworks are:

- General Metadata
- Per-pixel Metadata
- Radiometric Corrections and Accuracy
- Geometric Corrections and Accuracy

General Metadata allows users to assess the overall suitability of datasets with a human-readable filename structure, while per-pixel metadata contains pixel information, allowing users to filter unsuitable observations. Radiometric corrections specify the approach to correct top-of-atmosphere satellite data to ground level to meet measurement and accuracy requirements, providing consistent data that reflect surface changes and not changes due to sensor variations. Geometric corrections deal with geometric accuracy and corrections using digital terrain elevations.

CEOS has two levels of requirements for ARD: the 'Threshold' and 'Target'. The Threshold level is the minimum requirement, whereas the Target level represents the leading edge and could be interpreted as aspirational requirements. The pre-processing steps include geolocation and spatial alignment in the Albers Equal Area Conic map projection referenced to the **WGS84 datum**, radiometric calibration to provide consistent data reflecting surface changes rather than sensor changes, and atmospheric correction to reduce the variable radiometric influence of the atmosphere and per-pixel quality flags. Another level that assures quality and accuracy is the peer review by CEOS experts, which is reviewed and updated periodically to ensure they evolve in line with technological trends and user needs.

Planet

Planet, as a data provider, has its own pre-processing definitions. Besides sharing, the core pre-processing that is sourced by CEOS, such as Radiometric Correction, Geometric Correction, and Metadata, Planet also has a broader scope when it comes to pre-processing. The following pre-processing approaches are:

- Radiometry – Calibration and Atmospheric Correction: Keep radiometry consistent by cross-calibrating Doves, RapidEye, and SkySats using the same ground calibration targets, and having a consistent process for converting data to surface reflectance.
- Geometry – Orthorectification: Correcting image geometry, such as the removal of distortion caused by view angle and terrain, so that the coordinates of each pixel are projected to an accurate location on Earth. Planet uses a highly accurate, high-resolution ground control point (GCP) dataset to achieve positional accuracy of less than 10 meters RMSE globally.
- Metadata – Usable Data Masks (UDM): The UDM asset includes information on whether a pixel in an image contains usable imagery, cloud, no-data, or “anomalous” pixels.
- Interoperability – Sensor-to-Sensor Normalization: The equivalent of spectral band adjustment factors for all the electro-optical data sets available on the platform. If you are trying to compare results of an NDVI when using different data sets, it is helpful to normalize these differences to avoid added uncertainty in the comparison.
- Framing – Image Mosaicking and Clipping: For bulk orders of imagery, ensuring the user is only downloading the pixels they need with a single API call already clipped over the area of interest.

USGS

The United States Geological Service (USGS) has reprocessed the entire global Landsat archive, and its acquisitions sensed by Landsat 1 through Landsat 8 following the approach given by CEOS. The ARD are generated using the Collection 1 Tier data processing and reprojected into the Albers projection. They have also applied per-pixel quality assessment information, geographic correction, radiometric correction, ARD filename convention format, and metadata documentation.

2. Summarize in a short paragraph the main elements of the ESA Sentinel-2 Level-2A specifications and how they are processed.

Sentinel Level-2A provides atmospherically corrected Surface Reflectance (SR) products derived from Level-1C products. Sentinel Level-1C contains top-of-atmosphere data, which, once processed through atmospheric correction, becomes the data provided in the Sentinel-2A collection. The tiles in Sentinel Level-2A and Level-1C are 110x110 km² ortho-images in UTM/WGS84 projection. Earth is subdivided into a predefined set of tiles, defined in UTM/WGS84 projection, using a 100 km step. However, each tile has a surface of 110x110 km² to provide a large overlap with the neighbouring tiles, ensuring at least a 10 km overlap.

Regarding the analysis level of the images for Sentinel, Level-0 means telemetry analysis, preliminary quick-look, and cloud mask generation, compression, and SWIR pixels. Level-1A applies radiometric corrections such as dark signal correction and blind pixels removal, decompression, and compression to JPEG2000. Level-1B involves resampling and conversion to reflectance, aiming to generate TOA calibrated radiance from input Level-1A TOA radiance data, including radiometric corrections such as blind pixels removal, SWIR rearrangement, and geometric refinement.

The first level released to users is Level 1C. Level-2A includes the atmospheric correction applied to TOA Level-1C orthoimage products. Additionally, the level is processed by algorithms that handle scene classification and allow the detection of clouds, snow, and cloud shadows, generating a classification map with three different classes for clouds and six classes for shadows and vegetation. The classification algorithm uses threshold-based indexes. There are limitations with the classification algorithm, such as over-detection of clouds over bright targets, under-detection of semi-transparent clouds, and misclassification of snow and clouds.

2.1. Are Level-2A data available for all areas at the moment?

From the mid-March 2018, the Level-2A became an operational product, beginning with coverage of the Euro-Mediterranean region and since 2018 the global coverage has started for land areas and most of the islands except for high latitudes spots. However, the level-2A does not cover ocean areas.

2.2. What is “Sen2Cor”

The Sentinel-2 Level 2A processor, called Sen2Cor, is used for product generation and formatting. It performs atmospheric, terrain, and cirrus correction of Top-of-Atmosphere (TOA) Level 1C input data to create Bottom-of-Atmosphere or Surface Reflectance (SR) data. Optionally, the processor can generate terrain and cirrus corrected reflectance images, as well as aerosol optical thickness, water vapor, and scene classification maps. Its output product format is equivalent to the Level 1C User Product, providing JPEG 2000 images with three different resolutions: 60, 20, and 10 meters.

3. Describe how FORCE produces Level-2A data

FORCE stands for Framework for Operational Radiometric Correction for Environmental monitoring, providing an all-in-one solution for the mass processing and analysis of Landsat and Sentinel-2 image archives. FORCE processes Sentinel-2 Level-1 products to generate Level-2 to Level-4 products.

- Level-1: In the FORCE framework, this level involves radiometrically calibrated and georectified data.
- Level-2: Includes atmospheric correction.
- Level-3: Involves temporal aggregates of Level-2 data provided in different spatial references. It is the first level mapped on a regular grid, whereas lower-level products are in georectified swath geometry.
- Level 4: provides time series preparation and analysis functionality

The software components follow main steps where Level-1 acquires images from space agencies, using different routines for Landsat and ESA. Level-2 generates harmonized, standardized, and radiometrically consistent products with per-pixel information. The FORCE framework goes beyond what is standardized by CEOS by providing Level-3 and Level-4 products. FORCE aims to condense multi-temporal observations into a more manageable amount of spatially complete data, referred to as highly analysis-ready data (hARD).

3.1. Describe the main differences in the production chains compared to the ESA Level-2A data

The main difference between the FORCE framework and ESA production is with respect to the gridding aspect. Sentinel data already includes gridding at Level-1, which means that products are mapped on a regular grid and not only georectified swath geometry. This ESA distribution strategy contributes to effective gridding from the first level, efficiently saving time and effort in the subsequent processing levels and product generation. FORCE adapts gridding at Level-2, reprojecting all generated products into one coordinate system.

FORCE uses methodology from Sentinel-2 to process data. However, for Sentinel-2 data, the cloud displacement index was developed to compensate for missing thermal information using the parallax effect, an improvement on the Fmask algorithm that separates clouds from bright surfaces based on parallax effects. The Level-2 processing system applies Top-of-Atmosphere transformation to Surface Reflectance (SR), cloud detection, radiometric correction, and quality assurance information. The data are reprojected to a custom projection and then split into image chips using a custom grid with rectangular tiles, representing data cubes. Redundancy is prevented by aggregating same-day/same-sensor data on output.

The FORCE framework also provides high analysis-ready data (hARD) which differs from the Sentinel-2 diffusion strategy, meaning that Level-3 is capable of producing best-available pixel compositions or spectral temporal metrics. Best-available-pixel composites are produced by selecting the optimal observation according to defined criteria. Spectral temporal metrics are produced band-wise and are ideal predictors for machine-learning techniques that require independent features. FORCE also implements time-series analysis at Level-4, providing time-series preparation and analysis functionality, offering products such as mean, min, and max aggregation on a weekly, monthly, and annual basis, along with trend parameters.