

A Semantic Earth Observation Data Cube for Monitoring Environmental Changes during the Syrian Conflict

Hannah Augustin¹, Martin Sudmanns¹, Dirk Tiede¹ and Andrea Baraldi^{1,2}

¹University of Salzburg, Austria

²Italian Space Agency (ASI), Rome, Italy

Abstract

Freely and openly available, remotely-sensed Earth observation (EO) data are rapidly developing in terms of increased geo-spatial resolution and temporal frequency. This type of data requires automated workflows for handling, processing and analysis, including methods to convert data into valid information. This study presents a proof-of-concept implementation of a generic, semantic EO data cube with automated daily integration and semantic enrichment of Sentinel-2 data. The paper focuses on the technical implementation of an automated dataflow that enables semantic queries in an EO data cube. It proposes a transferable analytical environment with analysis-ready data to facilitate research on replicable extraction of EO-based indicators. Application-independent semantic data cubes can facilitate monitoring of land cover changes and the development of transferable, generic EO-based indicators to support decision-makers in the humanitarian sector, and international initiatives, such as the United Nations' Sustainable Development Goals. Syria was chosen as the study area because of the ongoing conflict and humanitarian demand, paired with the relatively cloud-free climate's suitability for optical EO time-series analysis.

Keywords:

remote sensing, Big Earth Data, data cube, semantic enrichment, reproducible

1 Introduction

Possibilities for remotely-sensed Earth observation (EO) data analysis were expanded considerably with the launch of the Landsat programme in 1972, and again by offering free and open public access to the archive in 2008 (Wulder et al., 2012). The launching of the Sentinel fleet by the Copernicus programme in 2014 has further increased the spatial resolution and temporal frequency of free and open EO data for analysis, with a promise of continued observation for many years to come (Drusch et al., 2012). We are currently witnessing a shift towards an alternative approach for storage and analysis using multi-dimensional data cubes for massive, gridded data (Baumann, 2017), bringing users to data,

rather than data to users. Google Earth Engine is another approach that is also bringing users to data. This enables access to massive amounts of geo-spatio-temporal data that are ready for analysis. Even in the isolated case of Sentinel-2, automated workflows are necessary to handle approximately 3.4TB of data captured every day (ESA, 2017), not to mention fusion with other similar sensors (e.g. Landsat) or integration with different datasets (e.g. radar, digital elevation models, socio-economic data).

Thanks to the consistent global coverage of free and open EO data, independent of political borders, EO data are ideal sources of evidence for generating useful information products to support decision-makers. This is especially the case when EO data sources are combined or integrated with additional sources, including non-EO ones. For example, night-time light EO data integrated with the Joint Research Centre's Global Human Settlement Layer (GHSL) and disaggregated population data were used to assess the humanitarian impact of the Syrian conflict (Corbane et al., 2016).

Indicator extraction from EO data is often necessary because the spatial resolution and acquisition frequency of EO data do not allow direct measurements of many objects or events on Earth (i.e. mixed pixels or relatively slow events). In the case of optical EO data, objects cannot be measured directly at all (e.g. pixels with similar reflectance values can represent different objects, surfaces, etc.). Non-physical entities (e.g. political boundaries) also cannot be directly measured. EO-based indicators can complement indicators or reports from other in-situ sources. Possibilities for the extraction of EO-based indicators are much more diverse if initial, generic semantic enrichment is applied, e.g. automatic spectral categorization (i.e. preliminary classification) into classes equal or inferior to land cover classes. Image understanding in the context of remote sensing is envisioned to include automated land cover classification, for which pre-classification would be a first step (Baraldi & Boschetti, 2012). Increased inclusion of such semantics transforms EO images into meaningful information in an automated way and allows queries through time, within a data cube model, for a plethora of target indicators.

Current setups of reproducible research for EO data cubes require significant time and financial investment and are limited to larger institutions, but this might change once appropriate technological development has been achieved. The importance of reproducible, transferable, interoperable, automated and repeatable workflows to process, handle and analyse massive EO data is becoming more apparent in a now data-rich world. With so much big data, it makes sense to avoid application-specific data (pre-)processing, which contradicts many big data principles. At the time of writing, three national-level Open Data Cube (ODC) implementations are operational, seven are in-development and twenty-nine are under review. The Committee on Earth Observation Satellites (CEOS) has set a goal of twenty operational national-scale data cubes by 2022 (ODC, 2017). One notable example is a framework for live monitoring of the Earth's surface (LiMES) proposed by Giuliani, Dao, et al. (2017), who are involved with the Swiss Data Cube (SDC) (Giuliani et al., 2017). The SDC is one of the operational national-level ODC implementations (Lewis et al., 2017). LiMES identified one of their main challenges in building a framework to be the turning of data into understandable information products.

The semantic EO data cube presented in this study suggests a fully automated framework not only for storing data, but also for their preliminary classification as building blocks for semantic analyses allowing information production. This automated, reproducible framework is one example for handling and analysing massive EO data, and demonstrates the benefits of automated, knowledge-based semantic enrichment for environmental change detection and EO-based indicator extraction for various thematic domains.

2 Data, theory and methods

Sentinel-2

The European Earth observation programme, Copernicus, operates downstream services and owns the Sentinel satellite fleet, which is connected to other sensors called ‘contribution missions’. The Sentinel-2 satellites are equipped with a multi-spectral instrument (MSI) observing 13 spectral bands (443nm–2190nm). A swath width of approximately 290km is captured with spatial resolutions ranging from 10m to 60m: three visible bands and one near-infrared band (10m); six red-edge/shortwave-infrared bands (20m); and three atmospheric correction bands (60m).

Currently, two Sentinel-2 satellites continuously and systematically collect observations – Sentinel-2A and 2B, launched on 23 June 2015 and 7 March 2017, respectively. The nominal average revisit time at the equator is every 5 days, with more frequent data capture towards the poles. Data are processed and provided by the European Space Agency (ESA) as level-1C (L1C), which includes radiometric calibration to top-of-atmosphere (TOA) reflectance and geometric corrections (e.g. orthorectification and spatial registration). L1C scenes are available as granules (i.e. tiles), approximately 100km by 100km, containing around 500MB of data.

Area of interest

The study area is located in northwestern Syria, along the border with Turkey, and is covered by three adjacent Sentinel-2 granules (37SBA, 37SCA, 37SDA), over an area of more than 30,000km² (latitudes 36.01°–37.05°N; longitudes 35.67°–39.11°E), as depicted in Figure 1. These three granules are provided in the same projection (UTM zone 37N, EPSG: 32637). All Sentinel-2 L1C scenes available on the Copernicus Open Access Hub for these granules are included in the data cube, resulting in a dense time-series beginning 28 June 2015 until the most recent scenes (i.e. at the time of writing, 31 January 2018), a total of 479 Sentinel-2 scenes. These granules are captured by two Sentinel-2 relative orbits (78 and 121), resulting in temporally denser data where the orbits overlap (see Figure 1).

The study area’s data characteristics indicate suitability for optical time-series analyses. According to the Köppen-Geiger classification, the climate is mostly warm Mediterranean (Csa) in the western part of the study area, transitioning into warm and semi-arid (BSh) towards the east (Peel, Finlayson, & McMahon, 2007). The annual average cloud-cover percentage, extracted from ESA’s L1C metadata, also decreases from west to east. The

majority of scenes acquired from May to October have a cloud-cover percentage below 10%, while otherwise generally ranging between 20% and 40%.

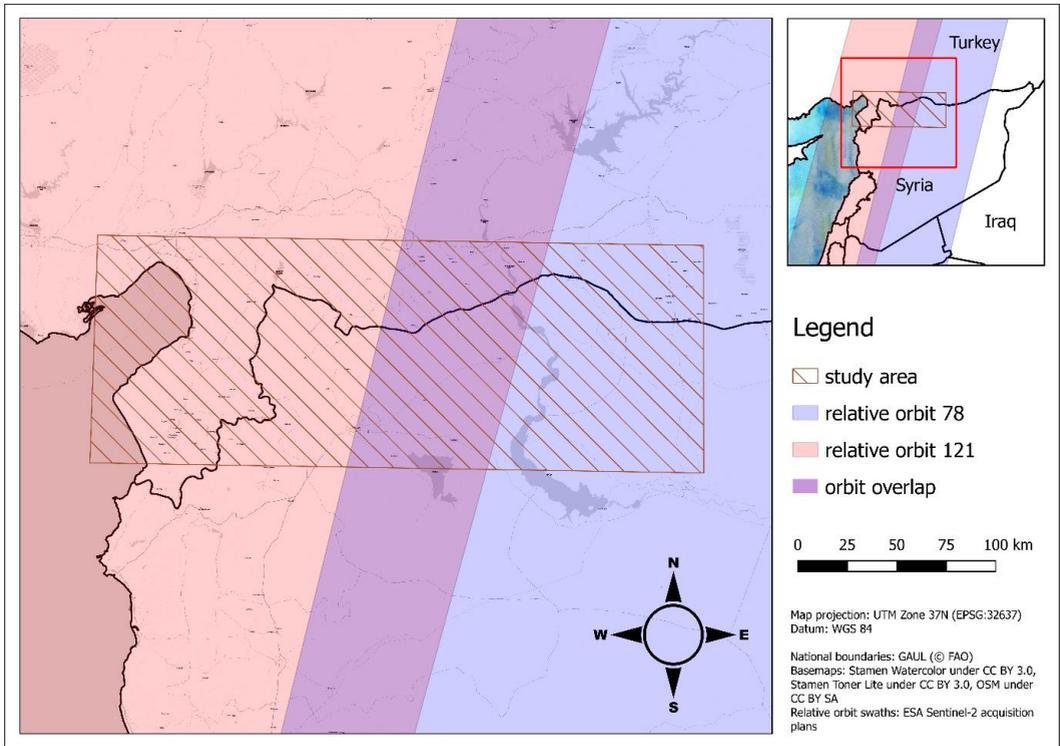


Figure 1: Overview of study area with Sentinel-2 relative orbits based on simplified acquisition swaths, showing an approximate orbit overlap in purple.

Preliminary classification

Image pre-classification is an initial classification of remotely-sensed images for use in image-understanding workflows. Spectral-based image pre-classification, as implemented by the Satellite Image Automatic Mapper™ (SIAM™, release 88v7), divides the feature space of a multi-spectral image into semantic semi-concepts using a knowledge-based approach, in contrast to data-driven approaches (e.g. supervised classification) (Baraldi et al., 2010a; Baraldi, 2011, 2018). A physical, model-based decision-tree using a priori knowledge of spectral profiles is applied to each pixel for the spectral categorization. Assuming images are calibrated to a minimum of TOA reflectance, these semi-concepts are comparable and therefore transferable between multiple images and optical sensors without any additional user-defined parametrization (i.e. the process is fully automatic).

According to Marr (1982), human vision begins with a pre-attentive first stage. The output is a symbolic primal sketch, including both a raw and a final version. The raw primal sketch is pure spectral differentiation of grey shades and colour tones, and the final primal sketch

groups similar shades and tones. Pre-classification is a primal sketch in Marr's sense, where semi-concepts are groups of spectrally similar pixels.

Semi-concepts are considered semi-symbolic in that they are an initial step in connecting sensory data (i.e. pixel values) to symbolic, semantic classes (Baraldi & Boschetti, 2012). They require further context, analysis or additional information to classify pixels into symbolic classes, such as land-cover classes. More general to more detailed semi-concepts may be considered a sort of multi-scale segmentation (Baraldi & Boschetti, 2012). Our implementation uses multiple semantic semi-concept granularities generated by SIAM™ to enable semantic queries (Baraldi, 2018).

Open Data Cube initiative

The Open Data Cube (ODC) initiative evolved from the Australian Geoscience Data Cube (AGDC) with the objective of providing a means to store, manage and analyse large volumes of EO data. The initiative has three goals: (i) increase the impact of EO; (ii) build a user community; and (iii) provide free and open software, including documentation (CEOS, 2017). Since CEOS is one of the founding members, the initiative seeks to align EO with overarching international agendas, such as the United Nations Sustainable Development Goals (SDGs). By fostering continental- and global-scale applications, the ODC initiative aims to improve decision-makers' use of EO-based information.

In following the data cube paradigm, ODC is conceptually comparable to array databases such as rasdaman (Baumann et al., 1998) and SciDB (Stonebraker et al., 2013), or database extensions such as SciQL (Kersten et al., 2011; Zhang et al., 2011). Indexing generates a metadata database including the location and properties of EO data, while ingestion generates its own data storage as NetCDF files. In both cases, the logical view offered to the user is a multi-dimensional data cube. Access is provided either by a Python application programming interface (API), e.g. usable in a Jupyter notebook (Kluyver et al., 2016), or by the ODC Web-based user interface. While a detailed review goes beyond the scope of this paper, the ODC has been selected because the data cube paradigm is best suited for the application, the technology has been proven to be robust and scalable, and it allows for transferable approaches through its open source licence.

3 Implementation

The automated workflow encompasses downloading Sentinel-2 data, re-formatting, preliminary classification with SIAM™ (i.e. information layer creation), indexing images, and ingesting information layers into an implementation of the ODC (Figure 2). This process runs automatically every day for each of the three study area granules. The result is daily incorporation of the most recently available data ready for analysis, including semantic queries. At the time of writing, 479 Sentinel-2 scenes from 28 June 2015 to 31 January 2018 have been processed.

Computing environment

The hardware used for this implementation is a Red Hat Enterprise Linux 7 virtual machine, with 16 virtual central processing units (CPUs) at 2.5 GHz clocking, 31 GB random-access memory (RAM) and 3 TB of generic, all-use storage.

Two reproducible virtual conda environments for Python are used. Data download to processing with SIAM™ is automated using cron, Python scripts and a conda environment with Python 2.7.x, Scipy, Geospatial Data Abstraction Library (GDAL) and requests. Data cube indexing, ingestion, Python API access and resulting analysis are conducted with a conda environment recommended for ODC installations (Geoscience Australia, CSIRO, & NCI, 2017d). This Python 3.5.x environment includes the datacube, jupyter, matplotlib, Scipy, basemap and basemap-data-hires packages for working with existing Jupyter notebooks (CEOS-SEO, 2017) and ones created by the authors.

Image acquisition

A command line interface (CLI) was implemented in Python, similar to the work of Olivier Hagolle (2018), to query the Copernicus Open Access Hub's API. Temporal restraints are set based on start and end times for data acquisition or ingestion to the hub. Spatial restraints can be set based on a point, polygon or granule name, the last of which is based on an in-house API that returns the centre point of any existing Sentinel-2 granule by name. If this option is used, only the identified granule is downloaded from any matching results, including targeted extraction from older, multi-granule products in the archive, dating from before 6 December 2016. Results are automatically unzipped, and any products already located in the target directory are not downloaded again.

Data formatting for SIAM™

Following complete and successful download, the necessary bands from each newly acquired Sentinel-2 scene are automatically re-formatted. SIAM™ is sensor-independent, but input data format requirements are based on Landsat for high-resolution data. Six bands – blue, green, red, near infrared and two medium infrared bands – are used (i.e. 2, 3, 4, 8, 11, 12). Bands 11 and 12 are resampled from 20m pixels to 10m using SciPy. Sentinel-2's MSI does not collect a thermal band, so a constant is used to ignore thermal decision rules in SIAM™. Based on the assumption that pixels with a value of 0 in any of the input bands contain no data and not a measured value of 0 (see discussion), a no-data mask is generated. Finally, the six Sentinel-2 bands of TOA reflectance values are converted to an 8-bit range, stacked in ascending order and saved in the ENVI data format for SIAM™ using GDAL.

Information layer generation with SIAM™

Automated generation of four semi-concept granularities (i.e. 18, 33, 48 and 96 semi-concepts) and four additional information layers has been implemented. The additional layers are: (1) a binary vegetation mask based on vegetation-related semi-concepts; (2) a pentanary haze mask, a discretised continuous sub-symbolic variable; (3) a greenness index,

i.e. $(\text{NIR} / \text{R}) + (\text{NIR} / \text{MIR1}) - (\text{Vis} / \text{MIR1})$ (Baraldi et al., 2010b; Baraldi, 2018); and (4) a panchromatic brightness image, a linear combination of all multi-spectral input bands. Processing of each Sentinel-2 scene takes approximately 5–6 minutes.

Indexing images and information layers

A product description needs to be defined in the ODC implementation database in order to index data. Indexing links to externally stored data and is backed by PostgreSQL. Product descriptions identify metadata common to all datasets of that product (Geoscience Australia, CSIRO, & NCI, 2017b) and only need to be defined once.

Metadata necessary for indexing is automatically generated for each dataset. This has been implemented for Sentinel-2 and SIAM™-generated information layers by modifying existing Python scripts provided by ODC. This metadata includes spatio-temporal data extents, data format, projection, bands/layers, file paths relative to the metadata, and more. A copy of the source Sentinel-2 dataset's metadata is included in the information layer metadata to document provenance. Once metadata has been generated, indexing automatically follows using a Python script and the data cube's API.

Ingestion of information layers

Once data have been indexed, they can be ingested, resulting in automated tiling of an indexed product into NetCDF files for more efficient access, creating a gridded time-series data cube (Geoscience Australia, CSIRO, & NCI, 2017c). The data cube API automatically creates a new product description, re-projects the data if necessary, tiles them accordingly, creates the necessary metadata and indexes them, with automatic checks to avoid duplication.

In this implementation, automated ingestion of information layers in 100km² tiles (i.e. 10km by 10km by one time-step) occurs, keeping the original projection (i.e. UTM zone 37N, EPSG: 32637). At the time of writing, 58,477 tiles of ingested information layers have been created, a total of 144 GB.

Python API

Once data have been indexed and ingested, they can be accessed using a Python API (Geoscience Australia, CSIRO, & NCI, 2017a). This API accesses data from a given indexed or ingested product for a defined spatio-temporal extent and returns a Dataset object defined by the xarray Python package. This Python object is a multi-dimensional, in memory, array with dimension names, and is used for further analysis (e.g. in Jupyter notebooks).

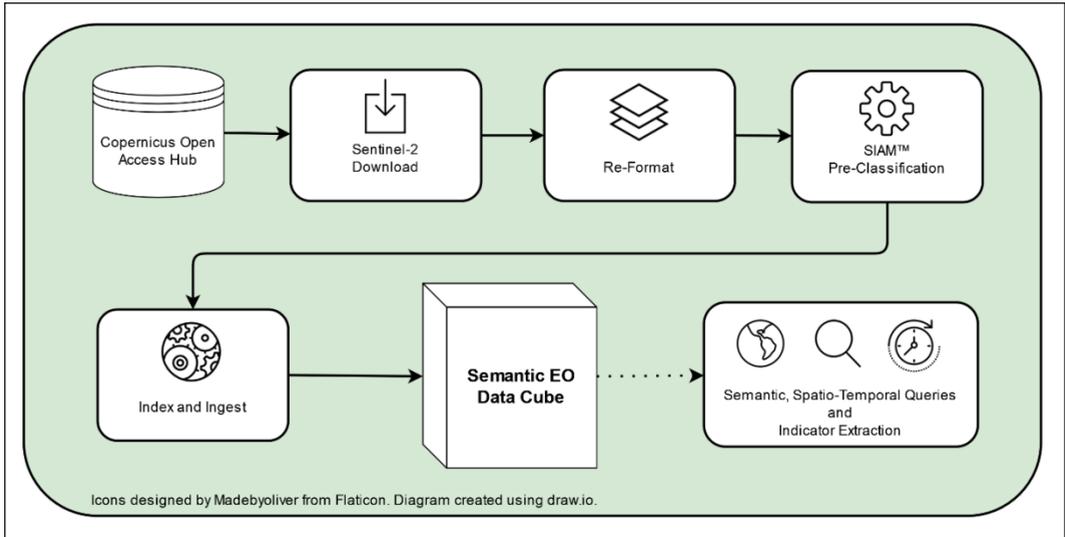


Figure 2: Automated workflow overview from download to queries and indicator extraction, the latter using the Python API.

4 Result and discussion

Semantic data cube

The largest benefit of the semantic data cube implemented here is that it fully automates data acquisition, semantic enrichment and access to data ready for analysis. Generic, application-independent semantic enrichment allows queries and EO-based indicator extraction for a variety of thematic tasks, and ensures reproducible results and repeatable analysis. An additional benefit of using SIAM™ is that it can be applied to data from multiple optical sensors, as long as they have been calibrated to TOA reflectance. Future incorporation of additional sensors would, in particular, expand the temporal extent of possible queries and analysis. Other EO data can also be incorporated (e.g. digital elevation model (DEM), gridded precipitation data) in order to further analysis possibilities.

One challenge is that processing using the Python API occurs predominantly in memory. This complicates implementation on the current hardware as it requires loading the complete dataset prior to analysis, which is a hardware limitation for smaller institutions. Even if many processes can be chunked, not all processes lend themselves to being divided, or they may produce similar but differing results (e.g. in the case of data-dependency in image-wide analyses).

A default exclusion of pixels with a value of 0 in any of the six Sentinel-2 bands used as input for SIAM™ (i.e. no-data masking) has been implemented because information about pixels not containing data for each band is not yet supplied with Sentinel-2 products. Even if the acquisition footprint is supplied in the metadata, each band's measurements at the edge of a swath are most often not perfectly aligned. Pixels with a measured value of 0 in any of the six

bands are excluded from semantic enrichment to reduce faulty semi-concept assignment, for example at the edge of a swath. Nevertheless, the authors are aware that this assumption may occasionally exclude meaningful information (i.e. when a valid measurement has a value of 0). Querying to test this assumption can, however, be done within the existing implementation, since the original Sentinel-2 bands are also indexed in the data cube. This will be conducted in the future to better assess the ramifications of this assumption.

Applications

Many new applications exist or are being envisioned for EO data cubes. These applications range from creating custom mosaics or composites (i.e. most recent cloud-free over a user-defined time span, seasonal composites), to various time-series analyses. Much research has been invested in looking at the dynamics of water. Surface water is a feature that can be relatively well differentiated from other types of land cover, using radar or optical data sources. Mueller et al. (2016) analysed 25 years of Landsat data using an implementation of the ODC, calculating a pixel-based normalized percentage of water detection, excluding no-data and clouds. Figure 3 demonstrates a similar method applied to the semantic data cube implemented here, but using water-like semi-concepts from SIAM™ instead of the Australian water-detection algorithm (i.e. Water Observation from Space). Here, pixels masked as no-data as well as cloud-like and unclassified semi-concepts have been excluded from the analysis, such that only pixels deemed to be clear observations are included and considered valid. It shows the normalized percentage of water semi-concept pixels related to other valid semi-concepts (e.g. vegetation-like) from 31 January 2016 to 31 January 2018. The percentage was obtained by analysing a total of 103 observations along the dimension of time.

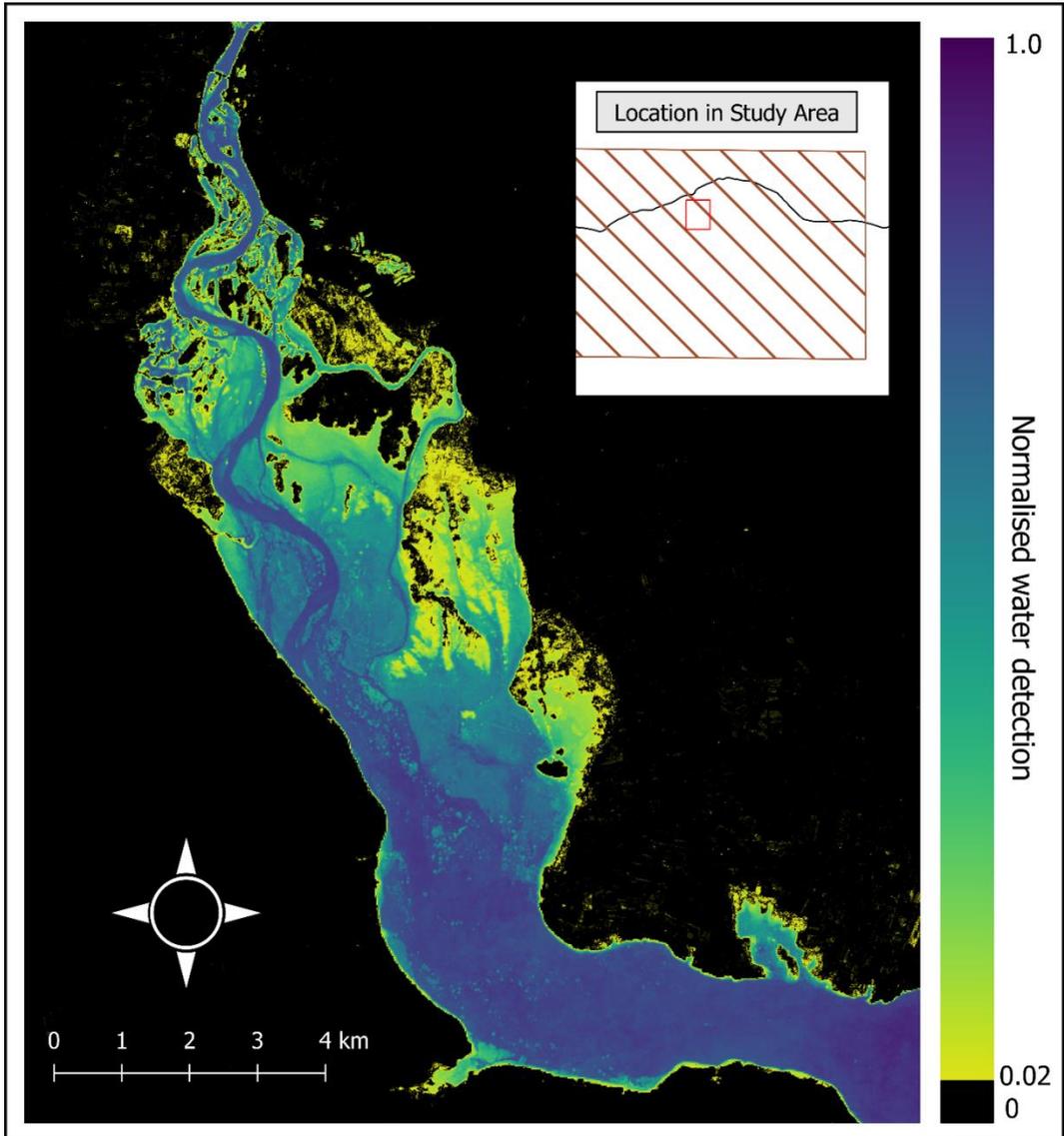


Figure 3: Normalized water detection based on water semi-concepts in Syria near the border with Turkey, excluding no-data, cloud-like and unclassified pixels, from 31 January 2016 to 31 January 2018 (103 time observations). Method similar to Mueller et al. (2016).

Pekel et al. (2016), in a similar application, carried out a global JRC study of water, but using Google Earth Engine. The most recent results available are from 2015, since results cannot be calculated in an automated way, unlike analysis using the Australian ODC and the ODC implementation featured in this paper. Pekel et al.'s results are, however, global.

In the implementation presented here, all of the application areas mentioned in the course of this paper and many more can be covered based on user-generated queries without requiring

the re-processing of the original data. The generic initial semantic enrichment in conjunction with flexible queries through time allows inferring new information layers or higher semantic levels (see Tiede et al., 2017).

One planned application of the approach presented here will complement and greatly extend an initial example based on Landsat data and SIAMTM semi-concepts by Tiede, Lüthje, & Baraldi (2014), which introduced automated post-classification change detection related to vegetation. More specifically, it focused on irrigated agriculture in Syria after the start of the on-going conflict as an indicator of conflict-related changes. This particular example could greatly benefit from access to an automated, reproducible data cube infrastructure, moving away from bi-temporal change towards incorporating data in a temporally dense way over the period of interest, either within one year, or between multiple years. Such a shift inherently moves in the direction of developing indicators based on various sources of evidence to support decision-making.

5 Conclusions

Data cubes are on the rise as a viable solution for remotely-sensed big EO data storage and analysis. They allow users to access the same pre-processed data, supporting reproducible analysis, and facilitate analysis using dimensions beyond the spatial (e.g. time) as additional axes in the data cube.

The innovation presented here is the set-up of a semantic data cube, which, in contrast to existing data cubes, stores information together with the data, thus allowing ad-hoc semantic queries. This is made possible via a fully automated workflow, including generic semantic enrichment of data to information layers, which adds this functionality to the ODC. Exemplarily shown is the utilization of the semantic data cube for extracting surface water dynamics in a use case located in Syria. Due to the generic approach, the same pre-processed data can be queried for other EO-based indicator extraction in a wide variety of thematic domains. This avoids application- and data-specific classification algorithms commonly proposed in recent open data cube literature. With the availability of free and open data having high spatial resolution and temporal frequency, we expect a general movement away from bi-temporal change analysis to change through or utilizing time, rather than controlling for it, as has been seen in EO-data analysis up to now. This implementation enables multi-temporal queries and analysis.

Indicator development based on dense EO time-series, or seasonal slices, is the next step to leveraging the potential of EO data, especially as data sources, such as Sentinel-2, are collected over periods of more than just a few years. Indicator extraction is necessary because the reflectance is only a proxy for detecting and identifying objects. This implementation can assist in detecting, monitoring, quantifying and even discovering new visible land-cover dynamics and processes (e.g. meandering rivers, ephemeral lakes, irrigated agriculture patterns or uncharacteristic vegetation removal), and offer evidence for supporting the requirements, actions and goals of multiple existing global initiatives.

Acknowledgements

This research was supported by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) under the programme ‘ICT of the Future’ within the project SemEO (contract no: 855467).

References

- Baraldi, A. (2011). Satellite Image Automatic Mapper™ (SIAMTM) - A Turnkey Software Executable for Automatic Near Real-Time Multi-Sensor Multi-Resolution Spectral Rule-Based Preliminary Classification of Spaceborne Multi-Spectral Images. *Recent Patents on Space Technology*, 1(2), 81–106.
- Baraldi, A. (2018, January 8). SIAM_Report_BACRES_v1.18.pdf. Retrieved 29 January 2018, from http://siam.andreabaraldi.com/content/Documentation/SIAM_Report_BACRES_v1.18.pdf
- Baraldi, A., & Boschetti, L. (2012). Operational Automatic Remote Sensing Image Understanding Systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA). Part 1: Introduction. *Remote Sensing*, 4(9), 2694–2735. <https://doi.org/10.3390/rs4092694>
- Baraldi, A., Durieux, L., Simonetti, D., Conchedda, G., Holecz, F., & Blonda, P. (2010a). Automatic Spectral Rule-Based Preliminary Classification of Radiometrically Calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery -- Part II: Classification Accuracy Assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1326–1354. <https://doi.org/10.1109/TGRS.2009.2032064>
- Baraldi, A., Durieux, L., Simonetti, D., Conchedda, G., Holecz, F., & Blonda, P. (2010b). Automatic Spectral-Rule-Based Preliminary Classification of Radiometrically Calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery – Part I: System Design and Implementation. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1299–1325. <https://doi.org/10.1109/TGRS.2009.2032457>
- Baumann, P. (2017). The Datacube Manifesto. Retrieved 30 January 2018, from <http://www.earthserver.eu/tech/datacube-manifesto>
- Baumann, P., Dehmel, A., Furtado, P., Ritsch, R., & Widmann, N. (1998). The Multidimensional Database System RasDaMan. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data* (pp. 575–577). New York, NY, USA: ACM. <https://doi.org/10.1145/276304.276386>
- CEOS. (2017). The CEOS open data cube initiative. Retrieved from https://docs.wixstatic.com/ugd/f9d4ea_1aea90c5bb7149c8a730890c0f791496.pdf
- CEOS-SEO. (2017). data_cube_notebooks: Jupyter Notebook examples for our Data Cube capable algorithms and functions. Jupyter Notebook. Retrieved from https://github.com/ceos-seo/data_cube_notebooks (Original work published 2016)
- Corbane, C., Kemper, T., Pesaresi, M., Freire, S., & Louvrier, C. (2016). Monitoring the Syrian Humanitarian Crisis with the JRC’s Global Human Settlement Layer and Night-Time Satellite Data. <https://doi.org/10.2788/297909>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... Bargellini, P. (2012). Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- ESA. (2017, January 27). Sentinel High Level Operations Plan (HLOP): COPE-S1OP-EOPG-PL-15-0020. Retrieved from https://earth.esa.int/documents/247904/685154/Sentinel_High_Level_Operations_Plan

- Geoscience Australia, CSIRO, & NCI. (2017a). Data Access API — Open Data Cube 1.4.0+0.g7682003 documentation. Retrieved 30 January 2018, from <https://datacube-core.readthedocs.io/en/stable/dev/api.html>
- Geoscience Australia, CSIRO, & NCI. (2017b). Indexing Data — Open Data Cube 1.4.0+0.g7682003 documentation. Retrieved 29 January 2018, from <https://datacube-core.readthedocs.io/en/stable/ops/indexing.html>
- Geoscience Australia, CSIRO, & NCI. (2017c). Ingesting Data — Open Data Cube 1.4.0+0.g7682003 documentation. Retrieved 29 January 2018, from <http://datacube-core.readthedocs.io/en/stable/ops/ingest.html>
- Geoscience Australia, CSIRO, & NCI. (2017d). Miniconda (recommended) — Open Data Cube 1.4.0+0.g7682003 documentation. Retrieved 29 January 2018, from <http://datacube-core.readthedocs.io/en/stable/ops/conda.html>
- Giuliani, G., Chatenoux, B., Bono, A. D., Rodila, D., Richard, J.-P., Allenbach, K., ... Peduzzi, P. (2017). Building an Earth Observations Data Cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data*, 0(0), 1–18. <https://doi.org/10.1080/20964471.2017.1398903>
- Giuliani, G., Dao, H., De Bono, A., Chatenoux, B., Allenbach, K., De Laborie, P., ... Peduzzi, P. (2017). Live Monitoring of Earth Surface (LiMES): A framework for monitoring environmental changes from Earth Observations. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.05.040>
- Hagolle, O. (2018). Automated download of Sentinel-2 L1C data from ESA (through wget). Python. Retrieved from <https://github.com/olivierhagolle/Sentinel-download> (Original work published 2015)
- Kersten, M., Zhang, Y., Ivanova, M., & Nes, N. (2011). SciQL, a Query Language for Science Applications. In *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases* (pp. 1–12). New York, NY, USA: ACM. <https://doi.org/10.1145/1966895.1966896>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... others. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing* (pp. 87–90). IOS Press.
- Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., ... Wang, L.-W. (2017). The Australian Geoscience Data Cube — Foundations and lessons learned. *Remote Sensing of Environment*, 202(Supplement C), 276–292. <https://doi.org/10.1016/j.rse.2017.03.015>
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., ... Ip, A. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment, Complete*(174), 341–352. <https://doi.org/10.1016/j.rse.2015.11.003>
- Open Data Cube (ODC). (2017). *opendatacube | CEOS*. Retrieved 30 January 2018, from <https://www.opendatacube.org/ceos>
- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, 11(5), 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418. <https://doi.org/10.1038/nature20584>
- Stonebraker, M., Brown, P., Zhang, D., & Becla, J. (2013). SciDB: A Database Management System for Applications with Complex Analytics. *Computing in Science Engineering*, 15(3), 54–62. <https://doi.org/10.1109/MCSE.2013.19>

- Tiede, D., Baraldi, A., Sudmanns, M., Belgiu, M., & Lang, S. (2017). Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases. *European Journal of Remote Sensing*, 50(1), 452–463. <https://doi.org/10.1080/22797254.2017.1357432>
- Tiede, D., Lüthje, F., & Baraldi, A. (2014). Automatic post-classification land cover change detection in Landsat images: Analysis of changes in agricultural areas during the Syrian crisis. In Seyfert, E., Gülch, E., Heipke, C., Schiewe, J., Sester, M. (Eds.), *Band 23: Geoinformationen Öffnen Das Tor Zur Welt*, 34. Jahrestagung in Hamburg 2014. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF) e.V., Potsdam.
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122, 2–10. <https://doi.org/10.1016/j.rse.2012.01.010>
- Zhang, Y., Kersten, M., Ivanova, M., & Nes, N. (2011). SciQL: Bridging the Gap Between Science and Relational DBMS. In *Proceedings of the 15th Symposium on International Database Engineering & Applications* (pp. 124–133). New York, NY, USA: ACM. <https://doi.org/10.1145/2076623.2076639>