SEMANTIC DATA CUBES UTILISING FREE AND OPEN-ACCESS EO-DATA FOR GENERATING SPATIALLY-EXPLICIT EVIDENCE

APPLIED USE-CASE IN SYRIA BASED ON SENTINEL-2 DATA

Master Thesis

by
Hannah Augustin

supervised by
Dr. Dirk Tiede
Martin Sudmanns

Interfaculty Department of Geoinformatics

Salzburg, June 2018

Submitted to the Interfaculty Department of Geoinformatics in partial fulfilment of the requirements for the degree of Master of Science (MSc) in Applied Geoinformatics at the Paris-Lodron-University of Salzburg.

UNIVERSITÄT SALZBURG

Z_GIS

[May 2, 2018 at 21:38 – `classicthesis` version 1.0]

ABSTRACT

Areas exposed to events or processes such as armed conflict, natural disasters or even climate change generally lack up-to-date, accessible data due to a variety of barriers (e.g. security, politics, biased providers). Free and open-access Earth observation (EO) data are being increasingly generated with global coverage at higher spatial resolutions and temporal frequencies than ever before. High spatial resolution (5-30m) imagery enables monitoring of large-scale areas beneficial for monitoring such events or processes. This data requires automated workflows for handling, processing and analysis, including methods to convert data into valid information.

Indicator extraction is one way to translate this data into meaningful information.

because the spatial resolution does not allow direct measurements of most objects on Earth (i.e. mixed pixels).

Improved situation assessment of areas where barriers to in-field data collection exist may be achieved by developing suitable indicators. In the case of events or processes that affect human populations, indicators of livelihood stability are vital to decision-makers. Due to the global and consistent nature of EO-data, they are ideal candidates for use in crisis indicators, combined or integrated with additional non-EO-data sources. For example, night-time light EO-data integrated with the Joint Research Center's Global Human Settlement Layer (GHSL) was used to estimate the size and location of the affected population in Syria. Such crisis indicators are envisioned to provide evidence-based knowledge to support crisis monitoring and impact assessment, but very few have been developed, applied and validated to date.

This study reviews the state-of-the-art of existing and envisioned humanitarian crisis indicators utilizing EO-data with a focus on solutions concerned with livelihood security (e.g. changes in agricultural areas, droughts, floods, power shortages). Emphasis in this review is given to indicators that monitor larger-scale areas, apply semi-automated to fully-automated workflows and utilise Sentinel-1/2/3 data. The aim is to provide an overview of the current state of development of EO-based humanitarian crisis indicators, and to distill knowledge about the contributions high resolution multi-spectral images (e.g. Sentinel-2) may offer. Ideally, future developments in this field will offer more solutions that are integrated or combined with at least one non-EO data source and utilise increasingly automated workflows.

A proof-of-concept implementation of a generic, semantic EO data cube with automated daily integration and semantic enrichment of Sentinel-2 data is presented and applied to Syria, a country with low annual cloud cover percentages beneficial for surface analysis using multi-spectral data. Using this implementation, surface water dynamics can be queried and analysed, and changes or losses to irrigated agricultural land over time are detected as a suggested indicator for instability. Challenges for analysis are posed not only by the identification of significant crisis indicators, requiring a combination of deductive and inductive methods, but also by the development of large-scale, automated (repeatable and reliable) methods for extracting indicators from relatively unwieldy big EO data. Analysis is applied along the entire Western and Northern borders of Syria, utilizing Sentinel-2 data, with a focus on the automation of information extraction and integration of derived information with at least one additional data source in a semi-automatic workflow. The idea is that such application-independent, semantic data cubes can facilitate reproducible and repeatable monitoring of land cover changes and the development of transferable, generic EO-based indicators to support international initiatives, such as the United Nations' Sustainable Development Goals.

Keywords: remote sensing, big Earth data, data cube, semantic enrichment, reproducible research, crisis indicators, livelihood security

ZUSAMMENFASSUNG

Kurze Zusammenfassung des Inhaltes in deutscher Sprache...

Schlagwörter:

STATEMENTS

DECLARATION OF ACADEMIC INTEGRITY

I hereby confirm that this thesis is my own work and that if any text passages or diagrams from books, papers, the web or other sources have been copied or in any other way used, all references, including those found in electronic media, have been acknowledged and fully cited.

The thesis at hand has not yet been submitted as thesis in this or a similar form.

Salzburg, June 2018

Hannah Augustin

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

List financial support

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— **knuth:1974** [**knuth:1974**]

ACKNOWLEDGEMENTS

Professor Dr. Dirk Tiede for supportive and attentive supervision, many insightful discussions, constructive feedback and patience.

Martin Sudmanns, for his input and direction, both in shaping the topic of the work, but also for technical advice and resources.

Andrea Baraldi

Stefan Lang

Barbara Riedler

- funding support

Maren Mueller, simultaneous sidekick and hero, for putting up with rants that often made little sense to her, accepting my workaholic nature and her overall support, but especially for many delicious cooking adventures, getting me into the mountains for perspective and for consistently celebrating small victories together.

GI Stammtisch crew in Salzburg, for celebrating all of the victories (even the not so victorious), lamenting all of the challenges and tolerating a bunch of awkward silence while lost in thought on a weekly basis

parents for supporting my decision not to return to the place I grew-up, instilling a sense of adventure, supporting my interest in learning German from a young age, even though they couldn't really understand it or even the language itself, and

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

Part I

CONTEXT

1 INTRODUCTION

1.1 MOTIVATION

- big data and big earth data
- free and open-access EO data
- trends towards utilisation to support international initiatives (SDGs)
- indicators of humanitarian crisis or generation of spatially-explicit evidence of humanitarian crisis

Earth Observation (EO) satellites capture data covering the entirety of Earth's surface using a multitude of sensors with differing spatial resolutions and re-visit frequencies. This data is often termed big Earth data. EO data offers a solution for large-scale, multi-temporal and persistent monitoring, especially of interest for areas that are otherwise difficult to reach (e.g. war zones). Various remote-sensing satellites with differing capabilities can offer data that can be turned into meaningful information for improving preparedness and response to migration. The sheer amount of data is rapidly increasing and demands a higher degree of automation for information extraction (e.g. automated-prior-knowledge based or machine learning classification procedures), especially when conducting analysis over large-scale areas or long, dense time-series. Freely available, open-access EO data include sources from radar, multi-spectral instruments and other instruments. Some options include the Sentinel satellite fleet from the European Space Agency (ESA), the Landsat program, co-managed by the United States Geological Survey (USGS) and National Aeronautics and Space Administration (NASA) and the Joint Polar Satellite System (JPSS) from NASA and the National Oceanic and Atmospheric Administration (NOAA). The JPSS includes the Visible Infrared Imaging Radiometer Suite (VIIRS), meant to replace the Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) sensors for tasks such as night-time light analysis, active fire detection and climate change monitoring. These data are accessible to the general public, often via online archives (e.g. Sentinels Scientific Data Hub, USGS Landsat archive).

High resolution optical imagery includes contributions from both the Sentinel and Landsat satellites. Sentinel's multi-spectral satellites, 2A and 2B, run as part of the Copernicus programme (formerly known

as GMES) led by the European Union (EU), together at full operational capacity have a revisit time of 5 days over equatorial areas and a relatively high spatial resolution (10-60m) with 13 spectral bands. Offering data already calibrated to top of atmosphere (TOA) reflectance, with a collective total of around one TB of data daily, the Sentinel-2 satellites are predominantly used to monitor water cover, vegetation, coastal areas, soils,natural disasters and other features of interest for land services. Landsat5/7/8 are other multi-spectral instruments with relatively lower spatial resolutions, less frequent re-visit times and without pre-processed TOA reflectance, that can be exploited in similar ways, especially where historical data are relevant for comparison.

Commercial satellite data has a clear focus on very high resolution (VHR) optical data, meaning generally a spatial resolution of less than 1m to about 4m.Examples include QuickBird from DigitalGlobe, IKONOS, WorldView, GeoEye, Kompsat, Formosat, Pleiades and SPOT.

Possibilities for remotely sensed Earth observation (EO) data analysis were drastically expanded with the launch of the Landsat programme in 1972, and again by offering free and open public access to the archive in 2008 (Wulder, Masek, Cohen, Loveland, & Woodcock, 2012). The launching of the Sentinel fleet by the Copernicus programme starting 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 geo-spatio-temporal data 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).

Due to the consistent global coverage of free and open EO data, independent of political borders, they are ideal sources of evidence for generating useful information products to support decision-makers. This is especially the case when combined or integrated with additional data sources, EO or otherwise. 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, Kemper, Pesaresi, Freire, & Louvrier, 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). 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 extraction of EO-based indicators are much more diverse with initial, generic semantic enrichment, e.g. automatic spectral categorisation (i.e. preliminary classification) into classes equal or inferior to land cover classes. Image understanding in the context of remote sensing is envisioned by the authors to include automated land cover classification, whereby 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 model queries through time for a plethora of target indicators within a data cube model.

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 status is reached. 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 et al., 2017b) who are involved with the Swiss Data Cube (SDC) (Giuliani et al., 2017a). 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 turning data into understandable information products.

The semantic EO data cube presented in this study suggests a fully automated framework towards storing not only data, but preliminary classification as building blocks for semantic analyses allowing information production. This contribution is an example of an automated, reproducible framework 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 differing thematic domains.

1.2 GENERAL OBJECTIVES

- knowledge of existing indicators that utilise EO-data
- develop and apply a highly automated, scalable workflow
- familiarity with data cube technologies
- semi-automated indicator extraction

1.3 USE-CASE SELECTION

- Northern Syria
- Justification for choice (Syrian conflict, weather/climate, Sentinel-2 data projection + overlap of swaths)
- After-the-fact Turkish attack on/invasion of Afrin, Syria
- Mention previous study with Landsat data

1.4 RESEARCH QUESTIONS

The following body of work aims towards answering the following:

What indicators exist in the realm of humanitarian-related monitoring or detection?

Are semi-concepts sufficient for semi-automated monitoring in a humanitarian domain, where time is of the essence and sample-based algorithms applicability to various climates/ geographic locations might be limited?

2

THEORY

2.1 DEFINITION OF TERMS

- big data
- big Earth data
- data cube
- Top of Atmosphere vs. BOA vs. SURF calibration
- ESA Level 1C - Level 2A
- semantic enrichment
- humanitarian “crisis”
- indicator
- livelihood

2.2 INDICATORS AND EVIDENCE

- Indicators for “humanitarian crisis”
- EO data and need for indicators
- Development of spatially-explicit indicators
- Indicators vs. evidence
- literature review of existing EO-based indicators or sources of evidence

Indicator development is imperative to leveraging the potential of EO data and transforming them into meaningful and actionable information, especially as big, open and free data sources, such as provided by the Sentinel-2 satellites, are collected over a longer timespan. Indicator extraction is necessary because the reflectance observed by a sensor is only a proxy for detecting, identifying and monitoring objects and processes, since pixels representing similar reflectance values can represent different objects, surfaces, etc. Optical EO data does not contain direct measurements of most objects or events on Earth (i.e. mixed pixels or relatively slow events). Non-physical entities (e.g. political boundaries) also cannot be directly measured. Replicable extraction of generic EO-based indicators can complement indicators or reports from other in-situ sources as evidence for consilience to support decision-makers. Since much EO-data is independent of political boundaries, if not global in coverage, indicators derived from them will especially be useful in supporting international initiatives in various thematic domains, such as the United Nation’s Sustainable Development goals.

2.3 LIVELIHOOD-SPECIFIC EVIDENCE

- how can livelihood be addressed from an indicator perspective...
- some non-EO livelihood security indicators (existing or envisioned in literature)
- existing or envisioned EO-based livelihood indicators or sources of evidence

GDP growth has been estimated through measuring light emissions from satellite images.

2.4 STATE-OF-THE-ART

2.5 TAKEN FROM ELSEWHERE...

Big Earth data holds high potential for migration and emergency preparedness and response, especially due to its inherent independence from national or other human-imposed borders. Development of large-scale, automated (repeatable and reliable) methods for extracting information from huge amounts of data is the current trend in the field. This information can be used to improve situational awareness as well as regular, temporal monitoring and identification of changes. Data with lower spatial resolutions can be exploited by moving away from "direct" information extraction towards indicator-based approaches, whereas VHR data can be exploited using multi-scale approaches. Analysis of night-time lights data for various purposes related to conflicts, as well as using VHR data for visual monitoring of areas (e.g. IDP camps, borders) are application fields with considerable research that is also relevant for migration monitoring, but methods based on other data sources exist or are being actively explored (e.g.crisis indicator development). Particularly, methods to monitor phenomena or events that have historically resulted in migration would be a means to shift from measures reacting to migration to more prevention or preparedness measures.

Global night-time lights (NTL) data show the locations and brightness of light escaping into space. Most of these lights are electric and originate from human settlements,making NTL a useful data source for bridging social science and remote sensing. Since 2011, NTL data are being captured by the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB). NTL data has been used as an indicator for various socio-economic factors, but also for applications di-

rectly relevant to migration, such as estimating the number of affected or displaced people in the case of a crisis (Corbane et al. 2016) or early damaged area estimation (Kohiyama et al. 2004). Pre-processing requires removal of background noise and solar or lunar light contamination, cloud cover screening and exclusion of non-electric light sources (e.g. volcanoes, fires) (Elvidge et al. 2017). For 40 years, data was collected using the Defence Meteorological Satellite Program (DMSP) Operational Line Scan System (OLS). Methods exist for inter-calibrating DMSP with VIIRS data in order to gain longer time-series of images for detecting changes before VIIRS became operational in late 2011. Liet al. (2017) inter-calibrated DMSP/OLS and VIIRS night-time light images in order to retrospectively analyse changes that occurred to human settlement areas during the course of the Syrian civil war. Corbane et al. (2016) developed a methodology to estimate the number of people affected during a crisis utilising NTL data combined with the Joint Research Centre's (JRC)Global Human Settlement Layer (GHSL). Syria was the use-case and they demonstrated that a satellite-derived indicator from NTL data can potentially offer a relatively objective estimate of the number of people impacted by a humanitarian crisis in a timely manner. The establishment and growth of refugee or IDP camps may also be able to be detected based on NTL data. Most NTL studies up to now have been based on a few dates or annual image composites.Further research in this field would include focusing more on temporal dynamics in NTL, taking seasonal or hourly changes into consideration to better inform interpretations of results.

Conventional remotely sensed data (i.e. optical) are limited in the sense that they can only detect features that are visible (e.g. built structures, vegetation,agricultural fields, roads). These data sources can be utilised to monitor security of livelihood assets (e.g. food or water security), land conflicts, post-crisis structural damage assessment, climate change effects that could cause population pressures, and more, depending on their spatial resolution and temporal characteristics.

Free and open high resolution optical imagery (e.g. Sentinel-2, Landsat 5/7/8) lends itself well to information extraction for indicators due to the fact that pixels are mixed. One applied example comes from Tiede et al. (2014) in the scope of the EC-FP7 project G-SEXTANT (Geospatial services in support of EU external action). They demonstrated an automatic post-classification land cover change detection method based on Landsat imagery, focusing on changes in agricultural areas in at the Syrian-Turkish border as a potential indicator for livelihood security and ultimately regional stability in areas where the regional climate mandates irrigation to support crops. Further exploration into indicators for crisis, whether natural disasters or man-made conflicts, is an expanding field of research, including the development of automated methods for pro-longed monitoring of areas. Indicators based on high

resolution data have been envisioned for detecting or monitoring:burnt villages; informal urban growth; the development or growth of refugee or IDP camps (Wang et al. 2015) and their impact on the surrounding environment; changes in activity (e.g. new infrastructures such as roads or air fields);illicit crop establishment and growth (e.g. opium cultivation in Afghanistan); environment degradation; flood assessment or visible changes to water bodies; changes or loss of agricultural areas; deforestation or reforestation; and visible climate change or extreme weather event artefacts.

The ultimate aim of any data management system is to facilitate technical access and handling of data as rapidly as possible. Handling in this case refers also to typical and well-known database use-cases, which are including, but not limited to, projection (sub setting), selection (filtering) or joining (combining) of data. Typical applications of big EO images do not require processing all available images in an archive, but is usually selective regarding the area-of-interest, the time interval, quality levels (e.g. cloud cover) of the images but also their content itself (e.g. based on the legend of a scene classification map, i.e. water, vegetation, fire). The selection (filtering) might precede further operations, e.g. finding vegetation peaks over multiple years (a combination of projection and joining). Current EO image archives such as the ESA Scientific Data Hub or the USGS EarthExplorer do not provide these operators. Moreover, typical purely files-in-directories-based approaches are limited to fulfil the requirements for implementing all of these operators to reach the aforementioned aim. Typically, the files are referenced by their filenames only or by using a hierarchical folder-based system. These storage systems are reading the files sequentially and are therefore not suitable for managing, including processing, large amount of data.

Current big EO image processing paradigms require systems "to bring the users to the data and not the data to the users" and allow "any query, any time". These paradigms are putting heavy requirements on software and hardware, especially in petabyte-scaled applications with data-intensive operations, which will be common in a few years. Therefore, data- and infrastructure providers are seeking the solution in cloud-based systems, where currently different approaches are existing side-by-side and are outperforming traditional methods. Arguably, the current prevalent big EO data handling approach is to use a map-reduce-approach (a prominent example is the Google Earth Engine, or Apache Hadoop as software package), or a database-approach, where native array databases are utilised (example technologies (in the sense of "tools") are Rasdaman and SciDB). Some approaches, such as the Australian Geoscience Data Cube, are combining both approaches, mainly by implementing and retrofitting database properties (user management, indexing, etc.).

Array-database-based approaches usually come with properties, which are well-known from relational databases and can be exploited for handling large amount of EO images as well. Core features of database systems are the centralised data management, improved data security, multi-user support, transparent query processing and the use of a declarative query language like SQL (Structured Query Language). Array databases have been applied for handling big Earth data in recent years [Planthaber et al. 2012].Examples are the EarthServer [Baumann et al. 2016], based on the Rasdaman database, and EarthDB which is based on the SciDB database.

Most of the currently available technology implement the so-called datacube. For example, an array database might instantiate OGC-compliant datacubes, where the semantics of the axes are defined using coordinate reference systems (CRS), e.g., spatial coordinate reference systems, known by the harmonisation efforts by the EPSG. For example,in a three-dimensional datacube, a one-dimensional time CRS overlays the two-dimensional CRS. In total six characteristics have been identified in the publicly available datacube manifesto (https://groups.google.com/forum/#!topic/rasdaman-users/Q3Zg7Tbc1_8).

Besides the technology-driven strategies for performing searching and processing on the database level, user-driven requirements are leading to on-demand web-based online processing of big EO data. In the last years, several technologies and standards, which can be used for online processing of EO data, have been developed and made available to the community [Petcu et al. 2010]. Two OGC standards are the Web Coverage Processing Service (WCPS) and the Web Processing Service (WPS), while technology implementations are Google Earth Engine [Google Earth Engine Team2015] or the Jupyter notebooks. Other examples of web-based platforms, which have been explicitly designed for processing and analysing EO data, include the Amazon Cloud AWS (Amazon Web Service) for processing of Landsat-8 data, with a free access to the API [Amazon 2016]. The Australian Geoscience Data Cube(AGDC) is using the National Computational Infrastructure (NCI) to provide Landsat images in the petabyte scale together with processing capabilities over the internet [Evans et al. 2015]. The Austrian Earth Observation Data Centre for Water Resources Monitoring (EODC), a collaboration between the technical university of Vienna, the Austrian Meteorological Service (ZAMG) and other companies, pursues a similar approach [Wagner et al. 2014].Big Earth Data is characterised by the (at least three) “V’s”: Volume, Velocity, Variety, where sometimes Veracity is added as fourth “V”. Taking into account these characteristics, compared to traditional EO image processing pipelines, Big Earth Data-“ready” systems have to consider some additional constraints, which are imposed by the “any query, any time” requirement. The exploitation of the value of Big Earth Data involves automation,

pre-processing, on-demand querying and compelling visualisation of the results. Massive processing power in the cloud and fast network connection is required, but not sufficient. Automation of intelligent workflows leading to pre-processing of data are important drivers for on-demand and ad-hoc querying to extract information in real time.Semantically enriched data allow also unexperienced users to formulate queries by means a high-level declarative language. Instead of having to translate an algorithm into software code manually, the query will be evaluated by the system and transformed into optimised physical access patterns. This approach can be realised by automatic (application independent) semantic enrichment of EO images in Big EO image databases, which are therefore "prepared" and "ready"for application specific queries in distributed array databases (with a declarative query language and a query optimiser). This approach avoids redundancy in data handling and repeated data (pre-) processing. The feasibility of this approach has been proven by Tiede et. al. [2016].

figure_IQ_18042016_300dpi

Figure6: EO image data are semantically enriched and stored as information layers in datacubes. In combination with declarative querying in array databases, ad hoc information extraction is possible by means of semantic querying

1.1.1 Global night-time lights monitoring

Global night-time lights (NTL) data show the locations and brightness of light escaping into space. Most of these lights are electric and originate from human settlements, making NTL a useful data source for bridging social science and remote sensing. Since 2011, NTL data are being captured by the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB). NTL data has been used as an indicator for various socio-economic factors, but also for applications directly relevant to migration, such as estimating the number of affected or displaced people in the case of a crisis (Corbane, Kemper, Freire, Louvrier, & Pesaresi, 2016) or early damaged area estimation (Kohiyama, et al.,2004). Pre-processing requires removal of background noise and solar or lunar light contamination, cloud cover screening and exclusion of non-electric light sources (e.g. volcanoes, fires) (Elvidge, Baugh,Zhizhin, Hsu, & Ghosh, 2017). For 40 years, data was collected using the Defence Meteorological Satellite Program (DMSP) Operational Line Scan System (OLS).Methods exist for inter-calibrating DMSP with VIIRS data in order to gain longer time-series of images for detecting changes before VIIRS became operational in late 2011. Li et al. (2017) intercalibrated DMSP/OLS and VIIRS night-time light images in order to retrospectively analyse changes that occurred to human settlement areas during the course of the Syrian civil war. (Corbane, Kemper, Freire, Louvrier, & Pesaresi, 2016)developed a methodology to estimate the

number of people affected during acrisis utilising NTL data combined with the Joint Research Centre's (JRC)Global Human Settlement Layer (GHSL). Syria was the use-case and they demonstrated that a satellite-derived indicator from NTL data can potentially offer a relatively objective estimate of the number of people impacted by a humanitarian crisis in a timely manner. The establishment and growth of refugee or IDP camps may also be able to be detected based on NTL data. Most NTL studies up to now have been based on a few dates or annual image composites. Further research in this field would include focusing more on temporal dynamics in NTL,taking seasonal or hourly changes into consideration to better inform interpretations of results.

1.1.2 HR/VHR EO sources

Conventional remotely sensed data (i.e. optical) are limited in the sense that they can only detect features that are visible (e.g. built structures, vegetation,agricultural fields, roads). These data sources can be utilised to monitor security of livelihood assets (e.g. food or water security), land conflicts, post-crisis structural damage assessment, climate change effects that could cause population pressures, and more, depending on their spatial resolution and temporal characteristics.

Free and open high resolution optical imagery (e.g. Sentinel-2, Landsat 5/7/8) lends itself well to information extraction for indicators due to the fact that pixels are mixed. One applied example comes from (Tiede, Luethje, and Baraldi, 2014) in the scope of the EC-FP7 project G-SEXTANT (Geospatial services in support of EU external action). They demonstrated an automatic post-classification land cover change detection method based on Landsat imagery, focusing on changes in agricultural areas in at the Syrian-Turkish border as a potential indicator for livelihood security and ultimately regional stability in areas where the regional climate mandates irrigation to support crops. Further exploration into indicators for crisis, whether natural disasters or man-made conflicts, is an expanding field of research, including the development of automated methods for pro-longed monitoring of areas.Indicators based on high resolution data have been envisioned for detecting or monitoring: burnt villages; informal urban growth; the development or growth of refugee or IDP camps (Wang, So, & Smith, 2015) and their impact on the surrounding environment; changes in activity (e.g. new infrastructures such as roads or airfields); illicit crop establishment and growth (e.g. opium cultivation in Afghanistan); environment degradation; flood assessment or visible changes to water bodies; changes or loss of agricultural areas; deforestation or reforestation; and visible climate change or extreme weather event artefacts.

VHR data can be used to monitor and map changes, including: IDP, temporary and refugee settlements (Laneve, Santilli, & Lingenfelder, 2006); damaged or burnt urban or village structures; characterizing

slums; border surveillance; and more detailed analysis of any land cover change. One critical issue in the optical data series processing is that preliminary cloud masking is required and also an accurate detection of haze conditions. In particular clouds make the optical data useless, while areas in the image affected by haze should be radiometrically corrected in order to avoid discarding of information. Change detection using optical data is usually a bigger challenge with respect to the SAR CD, depending on the information depth to be analysed.

The years have also seen the birth of constellation with tens of micro EO satellites able to capture images of the Earth at an unprecedented pace. One image per day and maybe more is no more a chimera like it was in the early 2000 years. Constellation like Planet and Terra Bella/SkyBox (now merged Planetto Acquire Terra Bella from Google, Sign Multi-Year Data Contract, https://www.planet.com/pulse/planet-to-acquire-terra-bella-from-google/)offers HR and VHR data commonly with a business model based on subscription which is exactly focused on monitoring purposes. As shown before in (Adam Van Etter, 2016), there are several efforts in order to exploit data coming from traditional VHR missions like DigitalGlobe ones and new space missions like Planet in order to extract automatically objects.

Part II

APPLIED USE-CASE

3

IMPLEMENTATION

3.1 BACKGROUND

- rationality
- why is this relevant for monitoring/detecting spatially-explicit evidence of humanitarian crisis?

3.2 STUDY AREA

The study area used in this proof-of-concept implementation is located in north-western Syria, along the border to Turkey. Three adjacent Sentinel-2 granules (37SBA, 37SCA, 37SDA) cover an area of more than 30,000km2 (latitudes 36.01°-37.05°N; longitudes 35.67°-39.11°E), as depicted in Figure 1, and ultimately define the exact extent of the study area. These three granules are provided in the same projection (UTM zone 37N, EPSG: 32637). All Sentinel-2 L1C data for these three granules that are available on the Copernicus Open Access Hub are continuously included in the data cube, resulting in a dense time-series beginning June 28, 2015 until the most recent image. At the time of writing, data is included up to ##### ###, 2018, which results in ### Sentinel-2 images. These granules are captured by two Sentinel-2 relative orbits (78 and 121), resulting in temporally denser data where the orbits overlap (see Figure 3.1).

Data characteristics of the study area 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 20-40% from October to May.

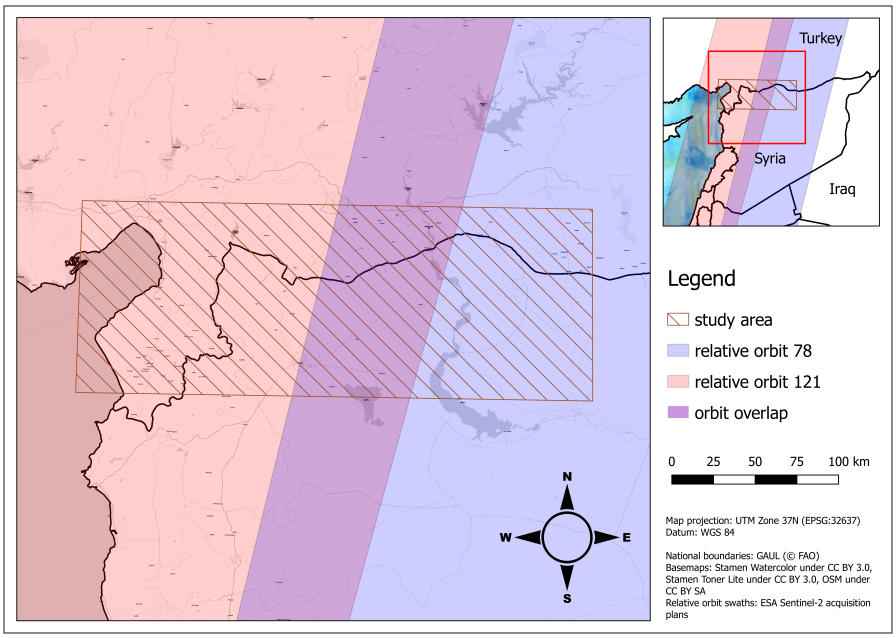

Figure 3.1: Overview of study area with Sentinel-2 relative orbits based on simplified acquisition swaths, showing an approximate orbit overlap in purple.

3.3 DATA

3.3.1 *Sentinel-2*

Specifications

Copernicus is a European Earth observation program, previously known as Global Monitoring for Environment and Security (GMES). It owns the fleet of Sentinel satellites, currently which is connected to other sensors called "contribution missions" and operates downstream services. The Sentinel-2 satellites are equipped with a multi-spectral imager (MSI) observing 13 spectral bands (443 nm-2190 nm). Data is captured with a swath width (i.e. field of view) of approximately 290km and spatial resolution ranging from 10-60m: three visible bands and one near-infrared band at 10m; six red-edge/shortwave infrared bands at 20m; and three atmospheric correction bands at 60m (see Figure 3.2).

Currently two Sentinel-2 satellites, known as Sentinel-2A and -2B, are continuously and systematically collecting observations. They were launched on June 23, 2015 and March 7, 2017, respectively, with Sentinel-2C and -2D already planned in the future to extend the longevity of Sentinel-2 observations. Looking at observations from both satellites together, the nominal average revisit time at the

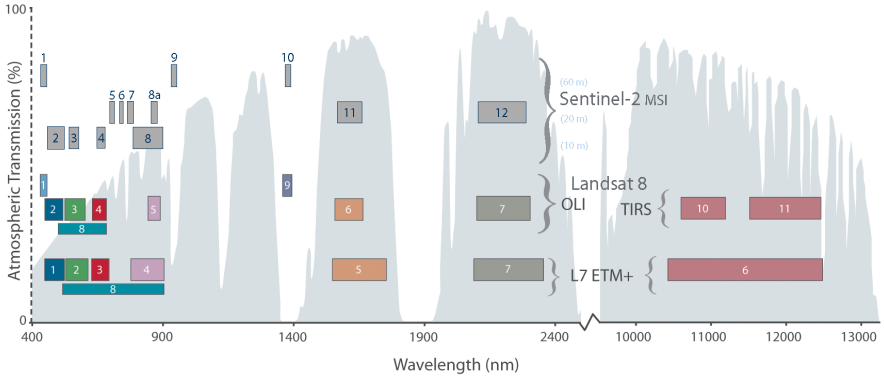

Figure 3.2: Spectral comparison of Landsat 7 and 8 bands with Sentinel-2 (retrieved on 25 April 2018 from https://landsat.gsfc.nasa.gov/sentinel-2a-launches-our-compliments-our-complements/)

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, spatial registration). L1C scenes are available as granules (i.e. tiles). These granules each cover approximately 100km by 100km and contain around 600MB of data, including all spectral bands, metadata and some quality indicators generated by ESA.

Data Used

- selected metadata in annex
- mention automated workflow described below

3.3.2 *Auxiliary data*

- Irrigated Areas (GMIA or Irrmap)
- Syria Refugee Sites (https://data.humdata.org/dataset/syria-refugee-sites)
- Precipitation or drought data
- mention acquisition/ingestion, etc. or reference below

3.4 METHODS

- overview of workflow, covering what was programmed by author (focus on automation and big data)

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 images from June 28, 2015 until January 31, 2018 have been processed.

3.4.1 *Indicator development*

- reference water in Australia and apply/situate in framework referenced in section 2
- describe two indicators (water, vegetation ...)

3.4.2 *Automatic knowledge-based spectral categorisation*

Theory

Image pre-classification is an initial classification of remotely sensed images for use in image understanding workflows. 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 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 the Marr 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).

SIAM(TM)

This implementation uses multiple semantic semi-concept granularities generated by the Satellite Image Automatic Mapper™ (SIAM™, release 88v7) to enable semantic queries (Baraldi, 2018). Spectral-based image pre-classification, as implemented by SIAM™, 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 the foundation applied to each pixel for the spectral categorisation. 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 parametrisation (i.e. fully automatic).

3.4.3 *Hardware*

- describe server

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 3TB of generic, all-use storage.

3.4.4 *Software*

Open-Source

- Python, virtual environments, etc.
- ODC
- Jupyter notebooks

LINUX

PYTHON

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, 2016/2017) and ones created by the authors.

GIT

OPEN DATA CUBE (ODC)

The Open Data Cube (ODC) initiative evolved from the Australian Geoscience Data Cube (AGDC) with the objective to provide 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 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 the use of EO-based information by decision-makers.

By 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 with 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 using 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 allows for transferable approaches through its open source license.

JUPYTER NOTEBOOKS

OPEN SCIENCE FRAMEWORK

Other

SIAM™

3.4.5 *Automated Workflow: Sentinel-2 Data Acquisition to ODC Ingestion*

- (scripts in annex)
- big EO data
- scalable – perhaps add some sort of benchmark

Accessing and Acquiring Sentinel-2 data

A command line interface (CLI) was implemented in Python, similar to the work of Olivier Hagolle (2015/2018), 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 latter of which is based on an in-house API that returns 171 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, multiple granule products in the archive from before December 6, 2016. Results are automatically unzipped, and any products already located in the target directory are not downloaded again.

The structure of products and metadata has seen a few modifications since the first images offered to the public in 2015. Most notably, Sentinel-2 products were served in packages of multiple granules prior to December 6, 2016, with sizes sometimes exceeding 6 GB.

Formatting data for SIAM™

Following complete and successful download, the necessary bands from each newly acquired Sentinel-2 image 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.

Generating information layers 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) binary vegetation mask based on vegetation-related semi-concepts; (2) pentanary haze mask, a discretised continuous sub192 symbolic variable; (3) ratio greenness index, i.e. (NIR / R) + (NIR / MIR1) – (Vis / MIR1) (Baraldi et al., 2010b; Baraldi, 2018); and (4) panchromatic brightness image, a

linear combination of all multi-spectral input bands. Processing of each Sentinel-2 image takes approximately 5-6 minutes.

Indexing images and information layers

A product description needs to be defined in the ODC implementation database 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 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.

Ingesting information layers

Once data has been indexed, it can be ingested, meaning 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 (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.

3.4.6 *ODC Python API*

Once data has been indexed and ingested, they can be accessed using a Python API (Geoscience Australia, CSIRO, & NCI, 2017a). This API retrieves data from a given indexed or ingested product for a defined spatio-temporal extent as a Dataset object from the xarray Python package. This Python object is a multi-dimensional, in memory, array

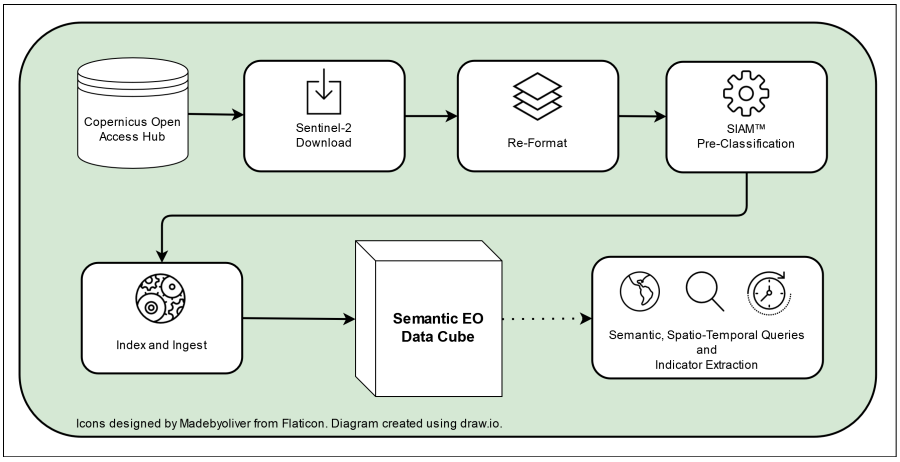

Figure 3.3: Automated workflow overview from download to queries and indicator extraction, which utilises the Python API.

with dimension names and is used for further analysis (e.g. in Jupyter notebooks).

3.4.7 *Jupyter Notebook for Indicator Calculation*

- Include screenshots

3.4.8 *Method of validation/agreement*

- random samples and an external dataset

4

RESULTS

4.1 MAPS AND CHARTS

- display and describe

4.2 VALIDATION RESULTS USING EO/NON-EO SOURCES

- (FAO food security services, statistics from Syria. . .)

5 DISCUSSION

5.1 INTERPRETATION OF RESULTS

- Discuss relevance of results

5.2 DISCUSSION OF METHODS

- (incl. challenges, successes, reconsiderations and shortcomings)
- repeatability
- reproducibility
- transferability

5.2.1 *GI_Forum Paper extract*

4.1 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 to 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 particularly 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) to further analysis possibilities.

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

The assumption that pixels with a value of 0 in any of the six Sentinel-2 bands used as input for SIAM™ be excluded (i.e. no-data masking) may be faulty more often than assumed, but information on pixels not containing data for each band in an image is not yet supplied with Sentinel-2 products. Even if the image footprint is supplied in the metadata, each band's measurements at the edge of an orbit swath are most often not identical. Pixels with a measured value of 0 in any of the six bands are thus excluded from semantic enrichment. The authors have found this assumption to prove useful in reducing faulty semi-concept assignment to pixels lacking valid data in any of the six bands within an image at a given time, for example, at the edge of an orbit swath. The authors are aware that the 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.

4.2 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 discerned from other types of land cover, whether 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 normalised percentage of water detection, excluding no-data and clouds. Figure ## 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 analysis such that only pixels deemed clear observations are included and considered valid. It shows the normalised percentage of water semi-concept pixels related to other valid semi-concepts (e.g. vegetation-like) from January 31, 2016 to January 31, 2018, which is from a total of 103 observations along the dimension of time.

Another similar application was completed by a global JRC study of water, but using Google Earth Engine (Pekel et al., 2016). The results calculated in this product are not produced in an automated way, unlike analysis using the Australian ODC and ODC implementation featured in this paper, so the most recent results available are from 2015. The results are, however, global.

In the presented implementation, all of these application areas and many more can be covered based on user-generated queries without

requiring re-processing 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 SIAM™ semi-concepts by Tiede, Lüthje, & Baraldi (2014), which introduced an automated post-classification change detection related to vegetation. More specifically, it focused on irrigated agriculture in Syria following the beginning of the still 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.

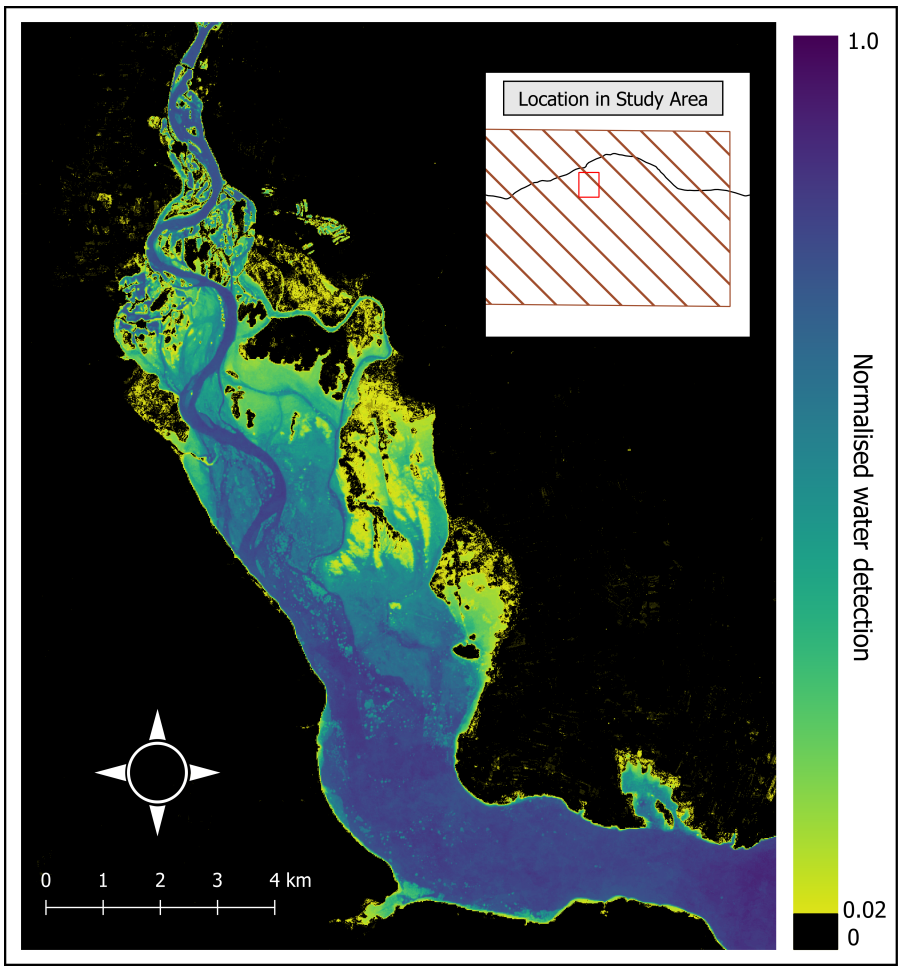

Figure 5.1: Normalised water detection based on water semi-concepts in Syria near the border of Turkey, excluding no-data, cloud like and unclassified pixels from January 31, 2016 until January 31, 2018 (103 time observations). Method similar to Mueller et al. (2016).

Part III

LOOKING FORWARD

6 OUTLOOK

- Outlook/further considerations/prospects – other indicators

6.1 DATA

- trend of free and open-access data
- set-backs due to proprietary algorithms in "open" data
- ARD

6.2 REPRODUCIBLE EO-ANALYSIS

transferability

repeatability

6.3 DATA CUBES

6.4 SEMANTICS

6.5 INDICATORS

- big Earth data and SDGs

6.6 PRIVACY AND MONITORING

- at least something about spatial resolution, temporal frequency and public vs. private sector
- what is privacy in terms of regular global EO data collection?
- maybe reference other kinds of big spatial data (e.g. Strava)

7

CONCLUSION

- absolute distillation of achievements in max. 2 pages

7.0.1 *GI_Forum extract*

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 utilisation of the semantic data cube for a surface water dynamics extraction in a use case located in Syria. Due to the generic approach, the same pre-processed data can be queried without changes for other EO-based indicator extraction in a wide variety of thematic domains. This avoids application and data specific classification algorithms as commonly proposed in recent open data cube literature. With the availability of free and open spatially and temporally high resolution data, we are expecting a general movement away from bi-temporal change analysis to change through or utilising 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 more than just a few years' time. 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, impermanent lakes, irrigated agriculture patterns, uncharacteristic vegetation removal), and offer evidence for supporting the requirements, actions and goals of multiple existing global initiatives.

8

REFERENCES

Baraldi, A., & Boschetti, L. (2012). Operational Automatic Remote Sensing Image Understanding Systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA). Part 2: Novel system Architecture, Information/Knowledge Representation, Algorithm Design and Implementation. *Remote Sensing*, *4*(9), 2768–2817. https://doi.org/10.3390/rs4092768

Baumann, P. (2017). The Datacube Manifesto. Retrieved January 30, 2018, from http://www.earthserver.eu/tech/datacube-manifesto

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

Giuliani, G., Chatenoux, B., Bono, A. D., Rodila, D., Richard, J.-P., Allenbach, K., ... Peduzzi, P. (2017a). 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. (2017b). 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

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

(ODC), O. D. C. (2017). Opendatacube | CEOS. Retrieved January 30, 2018, from https://www.opendatacube.org/ceos

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

Part IV

APPENDIX

A

APPENDIX TEST

A.1 DATA

A.1.1 *Data availability statement*

Equidem detraxit cu nam, vix eu delenit periculis. Eos ut vero constituto, no vidit propriae complectitur sea. Diceret nonummy in has, no qui eligendi recteque consetetur. Mel eu dictas suscipiantur, et sed placerat oporteat. At ipsum electram mei, ad aeque atomorum mea. There is also a useless Pascal listing below: Listing A.1.

More dummy text.
More dummy text.
More dummy text.
More dummy text.

A.2 CODE

A.2.1 *License Information for Code*

(open source/creative commons)

Python note!

Listing A.1: A floating example (`listings` manual)

```
for i in range(x, y):
    print 3
```