Detection and semantic description of changes in Earth Observation Time Series data

Daniela F. Milon-Flores,[0000−0002−3587−4838] Jérôme Gensel,[0000−0003−1398−7118], Gregory Giuliani,[0000−0002−1825−8865], and Camille Bernard[0000−0003−2246−6568]

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France
daniela.milon-flores@univ-grenoble-alpes.fr

Abstract. The overexploitation of natural resources and pollution are urgent concerns affecting the Earth's global system. Earth Observation (EO) data can be used to analyze the environmental impact of human activities. However, extracting meaningful insights from EO time series data requires domain expertise. In this position paper, we propose a methodology to improve the accessibility and understanding of environmental trends for a wide audience. Using Machine Learning (ML) technologies, we detect and describe in the Semantic Web (SW) changes in EO time series.

Keywords: Earth Observations · Semantic Web · Machine Learning.

1 Introduction

The overexploitation of natural resources such as forests and seas, as well as the pollution of air, soil, and water are urgent concerns affecting the Earth's global system and leading to climate change and loss of biodiversity. The Intergovernmental Panel on Climate Change (IPCC)[1] constantly reports the drastic consequences of inappropriate human behavior against the environment. To improve decision-making and implement effective environmental policies that counteract these negative trends, non-experts stakeholders, e.g., policy-makers and citizens, need access to Open Data that gives them insight into the environmental evolution of their municipality, here broadly referred to as Territorial Unit (TU).

Earth monitoring programs such as US Landsat[2] and European Copernicus[3] provide a free and open collection of satellite data depicting the Earth, also known as Earth Observation (EO) data. Due to the enormous amount of EO data, most state-of-the-art works [1, 8, 10] propose organizing EO images into Data Cubes. An Earth Observation Data Cube (EODC) is a massive multi-dimensional array organizing data to properly store, manage, and analyze the

1 www.ipcc.ch
2 landsat.gsfc.nasa.gov
3 www.copernicus.eu/en

EOs [3]. Although EODCs offer numerous benefits, it also presents certain challenges. For instance, managing EODCs to obtain information requires expertise. Specialists compute indices such as the Normalized Difference Vegetation Index (NDVI) to assess the environmental characteristics of specific areas. However, these indices are provided as raw time series, which requires metadata as well as processing and analysis to understand their meaning and evolution over time. Furthermore, EODCs are isolated from other data on the Web, difficulting their interoperability and reusability.

This position paper introduces our methodology focused on enhancing the accessibility and understanding of the environment's evolution over time (i.e., the environmental trajectory of Territory Units), by exploiting Machine Learning and Semantic Web technologies. In particular, this paper outlines three key areas of focus: (1) Structuring and semantizing EO data, (2) Automatically detecting significant changes in time series, and (3) Modeling environmental trajectories in the Semantic Web. This research is part of the TRACES project[4], an international collaborative research program between France and Switzerland, with a focus on building a Knowledge Graph (KG) that provides insights into the environmental evolution of municipalities for a wide audience.

2 Related work

Below, we present related work that is relevant to the problems we aim to address and that fits our proposed methodology:

1. **Structure and semantize EO data:** Semantic Sensor Network (SSN) [2] and RDF Data Cube Vocabulary (QB) [13] are standard ontologies that can be used to integrate EO data into the Semantic Web framework. SSN provides a means to describe sensors and their observations, encompassing satellite imagery as well. RDF Data Cube supports the publication of various multidimensional data, e.g., socioeconomic or environmental, and aligns with the OLAP cube concept used in Online Analytical Processing. Following, in projects like TELEIOS [9], novel methods for managing large EO data were devised. However, their focus is primarily on publishing image metadata on the Semantic Web using the non-standard stRDF ontology. The paper [5] introduced a method for publishing EO raster data at the pixel level using RDF Data Cube. In the context of our work, it is more appropriate to publish data at a local level such as municipalities, which is meaningful to the stakeholders. In the [15] study, the authors presented a modular ontology that contributes to the semantization of EO data. Their model reuses vocabularies such as SNN and the TSN ontology [4]. As a result, the integration allowed characterizing the Territory Units along with their land cover characteristics.
2. **Detect significant changes in time series:** Time series data often present change points such as trends and breaks, which indicate shifts in the behavior of the observations. Such changes, related to environmental indices, may

[4] http://traces-anr-fns.imag.fr

describe an important event or phenomenon. Various studies have proposed methods to detect these changes in time series. One widely used algorithm is the *Breaks For Additive Seasonal and Trend* (BFAST) [16] algorithm, which decomposes time series into trend, seasonal, and remainder components while identifying changes within the data. BFAST is versatile and applicable to different types of time series data. BFAST lite is an alternative version that improves speed and flexibility by utilizing a multivariate piecewise linear regression approach and handling missing data without interpolation [11]. Continuous Change Detection and Classification (CCDC) is commonly used for near-real-time change detection [19]. BFAST is the most popular technique for detecting trends and breaks in environmental indices like EO land cover. In the work conducted by [7], BFAST was employed to detect forest clear-cuts and burnt areas in a specific region of central Portugal. In the research paper [18], the authors monitored methane emissions from wetlands in China between 2002 and 2018 to observe the impact of climate change. Additionally, in [17], BFAST was applied to detect changes in 16-day NDVI images taken in a forested study area in southeastern Australia.

3. **Modeling environmental trajectories in the Semantic Web:** Few works focus on describing environmental trajectories in the Semantic Web. In the study of [14], a modular ontology was proposed to monitor land cover changes over time. In the paper of [12] is introduced an ontological design for modeling "trajectories" and their explanatory factors. Although the study focused on life trajectory data, it introduced vocabulary terms applicable to environmental evolution, e.g., "episode", "event", and "trajectory". In the work of [6], the author proposed the use of BFAST to detect events in big EO data. Then, he presents a hierarchy of terms such as "trajectory", "pattern", and "event" in order to describe land cover changes over time. This work is relevant to our proposal but has not been developed so far.

3 Methodology

Based on the RW (section 2), we present our methodology that consists in first, preparing the input data by aggregating the EOs by municipalities to be as closer as possible to the stakeholders; second, opening and structuring the aggregated data, in time and space, by exploiting Semantic Web and Machine Learning technologies such as the RDF Data Cube vocabulary and the BFAST algorithm respectively; and third, building a Knowledge Graph describing the environmental trajectory of municipalities based on our produced RDF data cubes and the detected breaks and trends. Additionally, Figure 1 (a) illustrates the different phases involved in our approach.

1. Input data: The TRACES project partners are the creators of the Swiss Data Cube (SDC)[5]. The SDC data cover all of Switzerland and part of France. In this research, three significant case studies were selected from the SDC

[5] `https://www.swissdatacube.org/`

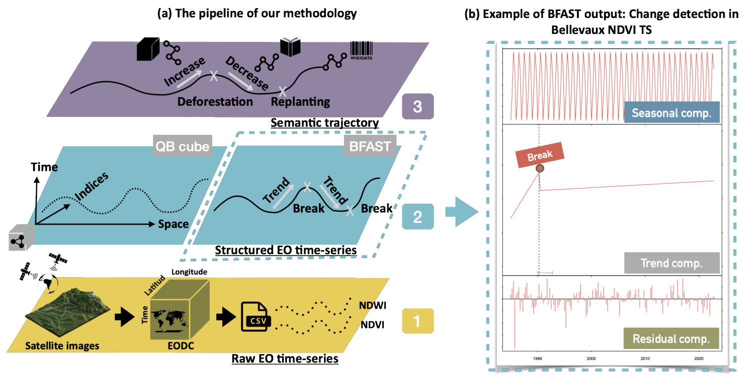

Fig. 1. Overview of our contributions and preliminary results.

as our study area, i.e., Fribourg, Evian, and Grand Geneve. Subsequently, Land Cover indices were calculated for each study area and delivered as raw time series. More specifically, three families of indices compose our case studies. Landsat Indices (LIS), Landsat Surface Temperature (LST), and Corine Land Cover (CLC). All the indices were calculated using the SDC. Furthermore, LIS and LST are available at seasonal, monthly, and daily levels covering 38 years, from 1985 to 2022, while CLC has five versions (1990, 2000, 2006, 2012, and 2018). Refer to Table 1 for more information.

Land cover indices				
Index Family	**Index var**	**Name**	**Total number of indices**	**Obs. per index**
LIS	NDSI	Normalized Difference Snow Index	20	Seasonal: 149 aprox. Monthly: 380 aprox. Daily: 897 aprox.
	NDVI	Normalized Difference Vegetation Index		
	WRI	Water Ratio Index		
	NDBI	Normalized Difference Built-Up Index		
	...			
LST	st	Surface Temperature	1	
CLC	clc-11	Urban fabric	44	5-time observations.
	clc-22	Permanent crops		
	clc-31	Forests		
	...			

Table 1. Three families integrate the selected Land Cover indices for our case studies: LIS, LST and CLC.

2. Open and structure EO time series: To overcome the challenges of isolation, interoperability, and reusability typically found on EODCs, we utilize the W3C standard RDF Data Cube vocabulary (QB). QB enables the organization and publication of EO time series in a multidimensional structure on the Semantic Web. QB also allows the linking and sharing of cube dimensions with Linked Data resources. Simultaneously, we use Machine Learning techniques, such as trends and breaks detection algorithms, to structure the

successive data measures over time into segments that characterize the evolution of a given Territory Unit. These consecutive segments are referred to as Environmental Trajectories of Territories (ETT).
3. Semantic trajectory: After identifying the segments that best represent the environmental evolution of a given Territory Unit (i.e., ETT), our goal is to define, with the support of experts, a vocabulary that describes these segments on the Semantic Web. We refer to this vocabulary as the *Semantic Environmental Trajectories of Territorial units* (SETT) Ontology. SETT may include terms such as "increase", "decrease", "deforestation", "replanting", "forest", "trajectory" and "trend" in order to describe the segments in a precise but easy-to-understand format. Later, by populating the SETT ontology with data from our three case studies and linking it to other Linked Data resources, including our produced RDF data cubes, we obtained the SETT Knowledge Graph. Our main objective with SETT is to enable the creation of a semantic trajectory that provides a comprehensive understanding of the environmental changes over time expressed as an RDF Graph for wide audiences, such as citizens, associations, and policymakers.

4 Modeling SETT using trends and breaks detection

To understand the environmental trajectory of a Territory Unit, it is essential to analyze the changes in the EO time series. Environmental changes can be linked to important events. For example, a decline in the NDVI index might suggest a reduction in vegetation cover due to deforestation policies. To detect significant changes, the Machine Learning algorithm BFAST is used. It separates the time series data into three components: trend, seasonal, and remainder. The trend component reflects the long-term changing pattern, while the seasonal component captures recurring patterns like seasonal variations. The remainder component represents the unexplained parts of the data. By analyzing the trend and seasonal components, BFAST identifies breakpoints, which indicate shifts or changes in the behavior of the time series. However, it is important to note that BFAST has limitations. It does not work when the time series contains unknown measurements. In such cases, an interpolation process must be applied to complete the missing values. In addition, the selection of parameters, such as the "break threshold", influences the results of the analysis. This is why we want to test other change detection algorithms, such as BFAST lite or the CCDC algorithm, in the future.

As shown in Figure 1 (b), we applied BFAST to the NDVI index of Bellevaux, a French municipality. As a result, the time series was decomposed into seasonal, trend, and residual components. Moreover, only one significant breakpoint was detected in the trend component between the years 1990 and 1992, representing a decrease in vegetation cover in Bellevaux. The reader must note that this experiment was performed for one index at one municipality, while in this work, we have to deal with large data, i.e., 373 municipalities and 65 indexes with data covering 38 years.

In this research, we will work with TRACES project members, specialists in Land Cover and Land Change Science, to address challenges such as large EO data management, setting suitable parameter values in change detection algorithms, and defining a specific vocabulary for SETT oriented to broad audiences. Regarding the latter, terms such as "increase", "decrease", "deforestation", "replanting", "forest", etc, can be used to describe the environmental trajectories and avoid complex terms such as "breakpoints" and "normalized difference vegetation index". Subsequently, the SETT ontology will be linked to other Linked Data resources and metadata. This whole integration enables the creation of a semantic trajectory that provides a comprehensive understanding of the environmental changes over time for municipalities.

5 Conclusions

This position paper aims to enhance the accessibility and comprehension of EO data, at the municipality level, for a broad audience. To face this issue, we proposed a methodology that constructs complex artifacts called SETT which require the exploitation of Semantic Web and Machine Learning technologies.

References

[1] Marius Appel and Edzer Pebesma. "On-demand processing of data cubes from satellite image collections with the gdalcubes library". In: *Data* 4.3 (2019).

[2] Krzysztof Janowicz Armin Haller and Simon Cox. *Semantic Sensor Network Ontology*. W3C Recommendation. https://www.w3.org/TR/vocab-ssn/. W3C, 2017.

[3] Peter Baumann. *The Datacube Manifesto*. 2017.

[4] Camille Bernard et al. "An ontology-based algorithm for managing the evolution of multi-level territorial partitions". In: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2018.

[5] Dmitry Brizhinev et al. "Publishing and using Earth observation data with the RDF data cube and the discrete global grid system". In: *W3C Working Group Note and OGC Discussion Paper W3C* 20170928 (2017).

[6] Gilberto Camara. "On the semantics of big Earth observation data for land classification". In: *arXiv preprint arXiv:2204.11082* (2022).

[7] Hugo Costa, Anny Giraldo, and Mário Caetano. "Exploring BFAST to detect forest changes in Portugal". In: *Image and Signal Processing for Remote Sensing XXVI*. Vol. 11533. SPIE. 2020.

[8] Gregory Giuliani et al. "Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard)". In: *Big Earth Data* 1.1-2 (2017).

[9] Manolis Koubarakis et al. “Linked Earth Observation Data: The Projects TELEIOS and LEO”. In: *Proceedings of the Linking Geospatial Data Conference*. 2014.

[10] Adam Lewis et al. “The Australian geoscience data cube—foundations and lessons learned”. In: *Remote Sensing of Environment* 202 (2017).

[11] Dainius Masiliūnas et al. “BFAST Lite: A Lightweight Break Detection Method for Time Series Analysis”. In: *Remote Sensing* 13.16 (2021).

[12] David Noel et al. “Design patterns for modelling life trajectories in the semantic web”. In: 2017.

[13] Dave Reynolds Richard Cyganiak and Jeni Tennison. *The RDF Data Cube Vocabulary*. W3C Recommendation. https://www.w3.org/TR/vocab-data-cube/. W3C, 2014.

[14] Ba-Huy Tran et al. “An Approach for Integrating Earth Observation, Change Detection and Contextual Data for Semantic Search”. In: *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2020.

[15] Ba-Huy Tran et al. “Semantic integration of raster data for earth observation: An RDF dataset of territorial unit versions with their land cover”. In: *ISPRS International Journal of Geo-Information* (2020).

[16] Jan Verbesselt et al. “Detecting trend and seasonal changes in satellite image time series”. In: *Remote Sensing of Environment* 114 (2010).

[17] Jan Verbesselt et al. “Phenological change detection while accounting for abrupt and gradual trends in satellite image time series”. In: *Remote Sensing of Environment* 114.12 (2010).

[18] Yuanyuan Yang and Yong Wang. “Using the BFAST Algorithm and Multitemporal AIRS Data to Investigate Variation of Atmospheric Methane Concentration over Zoige Wetland of China”. In: *Remote Sensing* (2020).

[19] Zhe Zhu and Curtis E. Woodcock. “Continuous change detection and classification of land cover using all available Landsat data”. In: *Remote Sensing of Environment* 144 (2014), pp. 152–171. ISSN: 0034-4257.