Democratising Earth Observation Big Data With *eo-learn*: Application to Water-Level Monitoring

Matej Aleksandrov
Matej Batic
Miha Kadunc
Grega Milcinski
Rok Mocnik
Devis Peressutti
Blaz Sovdat
Anze Zupanc
Sinergise d.o.o.
Slovenia

Klemen Kenda
Jozef Stefan Institute
Slovenia

ABSTRACT

This application showcases *eo-learn*, an open-source Python package that allows to seamlessly access and process Earth Observation data, such as satellite imagery available through the Copernicus programme (currently more than 5 PB of data). *eo-learn* allows end users to create value-added products from EO data for, for instance, land cover and land use classification, floods, droughts, and wildfires monitoring and management, and detection of other man-made or natural phenomena. *eo-learn* allows to streamline and parallelise complex workflows, facilitating the application of data mining techniques, and the development of machine learning models for forecasting and decision-making.

To showcase the potential of *eo-learn*, an open-source water-level monitoring system has been developed to automatically track the water-level of dams and reservoirs in South Africa for the past three years. The paper reports details on methods and results.

eo-learn is the result of a collaborative EU project, bringing together top-tier European experts in remote sensing, big data analytics and machine learning to improve accessibility and usability of EO big data.

KEYWORDS

Earth Observation, Big Data, Machine Learning, Open Source, Python

ACM Reference Format:
Matej Aleksandrov, Matej Batic, Miha Kadunc, Grega Milcinski, Rok Mocnik, Devis Peressutti, Blaz Sovdat, Anze Zupanc, and Klemen Kenda. 2018. Democratising Earth Observation Big Data With *eo-learn*: Application to Water-Level Monitoring. In *Proceedings of KDD 2018 conference (KDD 2018).* ACM, New York, NY, USA, 3 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

KDD 2018, August 2018, London, U.K.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Perceptive Sentinel (http://www.perceptivesentinel.eu) is a EU funded project bringing together partners from Slovenia (Sinergise, Jozef Stefan Institute, Agricultural Institute of Slovenia), Austria (GeoVille), France (Magellium) and Denmark (Landbrug & Fodevarer) with the aim of democratising Earth Observation (EO) big data to benefit small and medium enterprises and citizens. The consortium brings together top-tier European experts in remote sensing, big data analytics and machine learning to improve accessibility and usability of EO big data.

The main goal of the project is to create an open-source platform where EO big data, such as satellite imagery available through the Copernicus programme (currently more than 5 PB of data), can be easily and seamlessly accessed and processed. This would allow end users to create value-added applications from EO data for land cover and land use classification, floods, droughts, and wildfires monitoring and management, and detection of other man-made or natural phenomena, to name a few. Given the large amount of high spatial resolution data and high revisit frequency, techniques able to automatically extract complex patterns from spatio-temporal data are needed.

Towards achieving such goal, we have developed the *eo-learn* (https://github.com/sentinel-hub/eo-learn) Python package. *eo-learn*, in conjunction with other open-source Python packages, such as *sentinelhub-py*, *geopandas*, *numpy*, etc., allows to seamlessly build workflows to access, process and extract action-ready information from EO data. In detail, *eo-learn* is a collection of tasks that perform specific operations on EO data. Such data could be remotely sensed satellite imagery, or vectorised in-situ information about land parcels. Examples of tasks include downloading data from archives or accessing it from services, saving to disk, cloud detection in satellite images, spatio-temporal manipulation, and standard image processing operations. A complete workflow can be created by defining a relational acyclic graph of tasks, and *eo-learn* takes care of resolving the dependencies and execute the tasks in the correct manner. This allows to automate and parallelise analysis of large amounts of data. In short, *eo-learn* represents a tool to generate analysis ready data from EO spatio-temporal observations that could be directly used for statistical inference or modelling using,

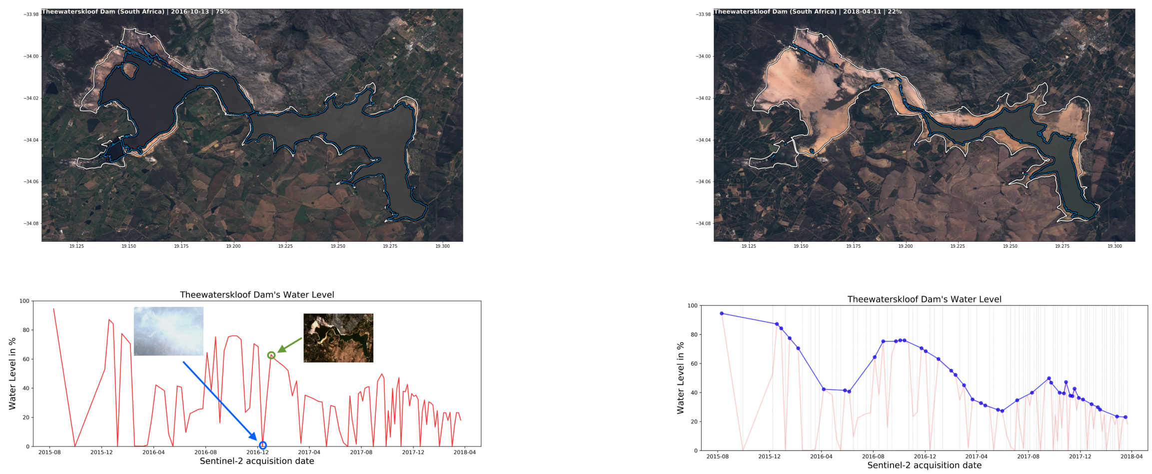

Figure 1: (Top). Sentinel-2 true color images of the Theewaterskloof dam acquired in 2016 (left) and 2018 (right). The white contour shows the nominal surface area. (Bottom) Surface area quantification for each Sentinel-2 image acquisition from 2015 to 2018 (left). By removing observations with clouds or artefacts from the analysis, trends in water-level can be observed and used for action-making (right).

for instance, machine learning algorithms. Furthermore, being an open-source project, users can contribute by creating new tasks and workflows.

In summary, the newly developed *eo-learn* library acts as a bridge between EO and remote sensing domain and Python ecosystem for data science and machine learning. Its aim is to lower the entry barrier to the field of remote sensing for non-experts on one hand, and bring state-of-the-art tools for computer vision, machine learning, and deep learning already existing in Python ecosystem to remote sensing experts on the other.

2 USE CASE: WATER-LEVEL MONITORING

In this proposal, we showcase the use of *eo-learn* to set up an open-source water-level monitoring service. The issue of water availability is now relevant more than ever, in particular in recent years with increasing changes in climate across the globe, and an increasing water usage by the population. In the past months for instance, South Africa faced a Day Zero threat, meaning that water in dams and reservoirs was running at a record low, and water supply to the population would be cut out. A service for monitoring water level in dams and reservoirs would help forecast and manage water availability.

To this end, *eo-learn* was used to build a water-level monitoring service for dams and reservoirs in South Africa. The service is set up as follows:

(1) retrieve the location and nominal water-levels for the dams and water reservoirs in South Africa;
(2) for each water body, use satellite imagery to compute water surface area for each available acquisition;
(3) display the surface area information for each water body in a way that can be directly used for analysis and action-making.

An *eo-learn* workflow automatically implements part (2) of the service, and consists of:

- downloading the satellite images for a region of interest comprising the water body for each available acquisition;
- filter out unusable acquisitions containing, for instance, clouds or imaging artefacts;
- segment the water body from the images;
- compute water surface area and compute ratio with respect to the nominal value.

The workflow is run automatically and can be parallelised across the water bodies and acquisition dates for efficient execution. Part (1) of the service is implemented using the Open Street Map data. Part (3) of the service is currently implemented using a JupyterLab notebook, but the service will be soon moved to a public server and the information made available through interactive maps and dashboards, which will be showcased at the KDD conference. The JupyterLab notebook can be found here (https://github.com/sentinel-hub/eo-learn/tree/master/examples). Figure 1 shows the water-level for the Theewaterskloof dam for the past three years. More information can also be found in this blog post.

Current work involves extending this service to a global scale, by running the workflow on a global collection of water bodies. The main challenge in scaling the service to a global scale is to develop a pipeline robust to inaccuracies and variance of the data, both the reference nominal surface level and the satellite imagery. Efficient management and visualisation of the analysis-ready results are also important requirements for a global monitorign system that need to be properly addressed.

Future work on this monitoring service will involve user engagement, where citizens could, for instance, provide feedback and report inaccuracies in the displayed information, or add new water bodies. The service could also be used by non-governative organisations for a more efficient planning of humanitarian interventions.

3 CONCLUSIONS

In conclusion, *eo-learn* represents a first step in allowing free and easy access and manipulation of EO big data. This opens up new possibilities to create EO applications with minimal costs, therefore facilitating uptake of EO big data and creation of new value-added products to market, and benefiting end users with transparent and accessible applications. Consortium partners are committed to develop *eo-learn* further, by developing and implementing processing tasks, and by creating graphical user interfaces to simplify the access and use of the platform for users not familiar with application development.