

# GEE TIMESERIES EXPLORER FOR QGIS – INSTANT ACCESS TO PETABYTES OF EARTH OBSERVATION DATA

Philippe Rufin<sup>1,2</sup>, Andreas Rabe<sup>1</sup>, Leon Nill<sup>1</sup>, Patrick Hostert<sup>1,2</sup>

<sup>1</sup> Geography Department, Humboldt-Universität zu Berlin, Germany - (philippe.rufin, andreas.rabe, leon.nill, patrick.hostert)@geo.hu-berlin.de

<sup>2</sup> Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Germany

**KEY WORDS:** Google Earth Engine, Remote Sensing, Validation, Python, Sampling, Education.

## ABSTRACT:

Earth observation analysis workflows commonly require mass processing of time series data, with data volumes easily exceeding terabyte magnitude, even for relatively small areas of interest. Cloud processing platforms such as Google Earth Engine (GEE) leverage accessibility to satellite image archives and thus facilitate time series analysis workflows. Instant visualization of time series data and integration with local data sources is, however, currently not implemented or requires coding customized scripts or applications. Here, we present the GEE Timeseries Explorer plugin which grants instant access to GEE from within QGIS. It seamlessly integrates the QGIS user interface with a compact widget for visualizing time series from any predefined or customized GEE image collection. Users can visualize time series profiles for a given coordinate as an interactive plot or visualize images with customized band rendering within the QGIS map canvas. The plugin is available through the QGIS plugin repository and detailed documentation is available online (<https://geetimeseriesexplorer.readthedocs.io/>).

## 1. INTRODUCTION

Recent advances in Earth observation were fuelled by increasing computational resources paired with open access to data archives provided by space agencies across the globe (Wulder et al., 2012; Zhu et al., 2019). Current applications for mapping land cover, land use, or biophysical quantities such as biomass, heavily rely on analyses of dense intra-annual or inter-annual time series (Wulder et al., 2018; Dong et al., 2019). Such workflows require mass processing of satellite data, with data volumes easily exceeding several terabytes, even for relatively small areas of interest.

While operational and openly available software solutions for mass processing of satellite data are available (Frantz, 2019), data volumes can define feasibility boundaries for users without access to appropriate computational resources. This may particularly affect early career researchers with limited access to - or funding for - computational infrastructure and thus generates barriers for scientific advances.

Cloud processing platforms such as Google Earth Engine (GEE) leverage accessibility to a variety of satellite image archives and thus facilitate time series analysis workflows for a broad user base (Gorelick et al., 2017). However, instant visualization of time series data is currently not implemented in GEE, and visually exploring time series commonly requires coding customized scripts or applications. Furthermore, the integration of locally stored datasets, such as reference datasets in vector format, can be challenging in cloud environments.

We here present the GEE Timeseries Explorer plugin which allows instant access to Earth Engine image collections within QGIS. The plugin seamlessly integrates the QGIS user interface with a compact widget for visualizing time series profiles from any GEE image collection as an interactive plot, or spatially as images with customized band rendering. These functionalities grant instant and flexible access to petabyte-scale data archives for a variety of users in research and higher education.

The remainder of this paper aims at presenting the key features of the GEE Timeseries Explorer, and to provide insights into exemplary applications from ongoing or recently published research.

## 2. GEE TIMESERIES EXPLORER

The GEE Timeseries Explorer builds on the existing Google Earth Engine QGIS plugin (<https://github.com/gee-community/qgis-earthengine-plugin>), therefore requiring a GEE account and QGIS version 3.16 or higher. It is currently available in the QGIS plugin repository, documentation and guidelines are available online (<https://geetimeseriesexplorer.readthedocs.io/>).

The following sections briefly describe the core features of the plugin, including 1) flexible access to GEE image collections, 2) dynamic time series visualization via PyQtGraph, 3) instant visualization of image data and time series aggregations with customized band rendering in the QGIS map canvas, as well as 4) download of sample-based time series data. We further performed benchmarking to indicate the performance of the GEE Timeseries Explorer.

### 2.1 Accessing data archives

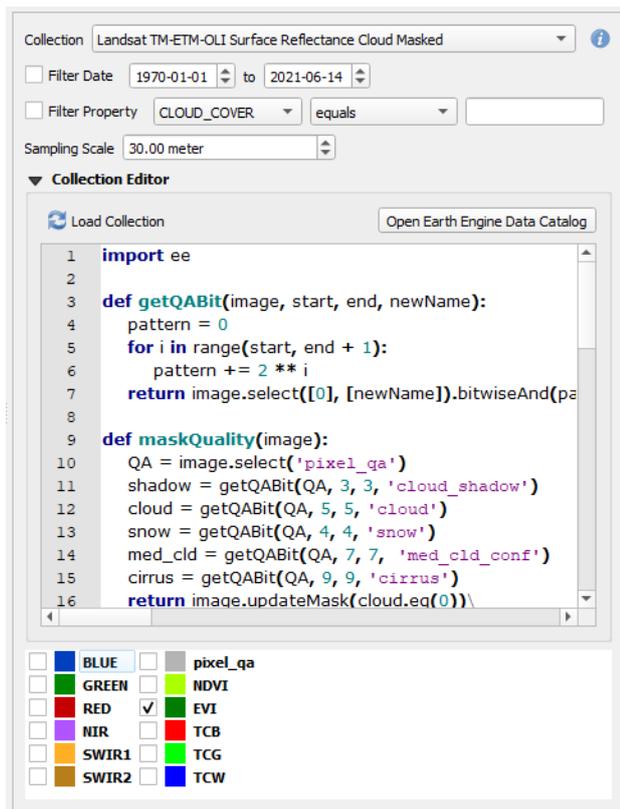
The GEE Timeseries Explorer offers flexible integration of any GEE image collection, such as the MODIS, Landsat, Sentinel-2, or Sentinel-1 product suites.

We added a set of pre-defined collections, including the quality-filtered MODIS vegetation indices product (MOD13Q1), integrated and cloud-masked Landsat TM, ETM+, OLI Level-2 surface reflectance, or cloud-masked Sentinel-2AB L2A surface reflectance products with enhanced cloud masking based on the Cloud Displacement Index (Frantz et al., 2018). Vegetation indices and Tasseled Cap transformation features were amended in the Landsat and Sentinel-2 collections for convenience. To ensure efficient access for users, a date range filter and a property filter were integrated into the user interface. Once the collection

is loaded, users can select one or multiple bands from the image collection for queries of time series data (Figure 1).

The built-in collection editor widget allows for straightforward integration of custom image collections, or modifications of existing collections, e.g., amending quality filtering, cloud- and cloud-shadow masking, or adding band indices or transformations. The collection editor is based on the Earth Engine Python API and thus uses its syntax (Figure 1).

Users are actively encouraged to contribute to the GEE Timeseries Explorer by sharing custom collections through the repository (<https://bitbucket.org/janzandr/geetimeseriesexplorer>) or the plugin homepage, which may be included in future versions.

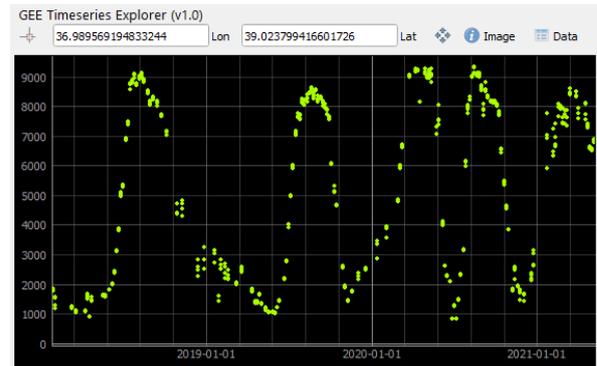


**Figure 1.** Collection settings and Python collection editor to define or modify image collections, here the predefined cloud-masked Landsat TM, ETM+, and OLI surface reflectance collection.

## 2.2 Time series visualization

The QGIS map canvas can be used to directly access pixel-level time series data by clicking on any map location. For efficient analysis of predefined samples, users can navigate through point locations stored in vector datasets, such as ground observations from fieldwork.

Retrieval of time series data occurs in raw text format and is instantly visualized via PyQTGraph (<http://www.pyqtgraph.org/>). The interactive nature of the plots allows adjusting point or line layout options, freely navigate and adjust axis scaling, which is particularly useful to identify outliers or inconsistencies in the time series or to investigate long intra-annual time series (Figure 2).

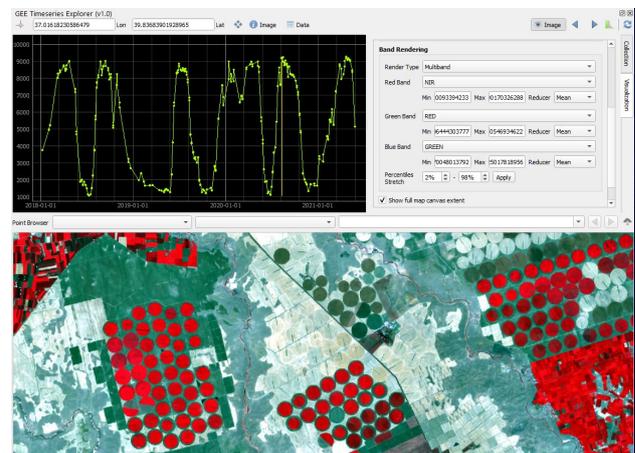


**Figure 2.** Exemplary NDVI time series (2018–2021) on intensively managed cropland derived from cloud-masked Sentinel-2AB surface reflectance product.

## 2.3 Image visualization

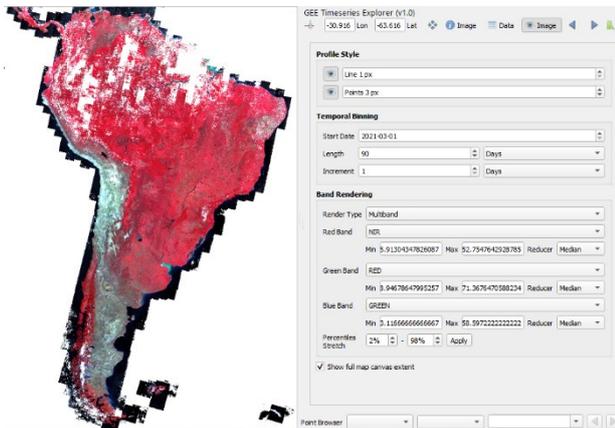
Users can visualize individual images of the time series within the QGIS map canvas by selecting the desired observation from the time series plot. Image data are retrieved as a WMS layer, which is added to the QGIS layer panel upon request.

Visualization options are based on default QGIS band rendering settings, allowing multi-band RGB and single-band palletted rendering. The image stretch can be freely adjusted using image percentiles or user-defined min/max values (Figure 3). Users can easily navigate through the time series and visualize observations subsequently using the built-in temporal navigation buttons, thus facilitating analyses of change.



**Figure 3.** Sentinel-2 NDVI time series and visualization settings (top), and image visualization (bottom, R/G/B: near infrared / red / green, acquisition date 10/08/2020).

Users can further define temporal bins and reducers to generate on-the-fly visualization of time series aggregations (often referred to as spectral-temporal metrics), such as mean, median, standard deviation, or selected percentiles of annual or seasonal surface reflectance across the entire map extent. Visualization of high-resolution images and time series derivatives across large areas is thus instantly feasible (Figure 4). Similar to the temporal navigation across original time series, users can define increments for navigating through time series aggregations, e.g., to revisit seasonal median reflectance across the Landsat data archive at annual intervals.



**Figure 4.** Landsat median reflectance across South America for the period 01/03/2021 – 31/05/2021 (R/G/B: near infrared / red / green).

## 2.4 Sample-based time series download

Time series data for selected locations can be downloaded directly from the user interface. For streamlining sample-based analyses, the GEE Timeseries Explorer features a parallel downloading functionality, which retrieves up to 50 time series in parallel. Data are stored as text files for each point feature, facilitating compatibility with all software packages and thus enabling straightforward integration of these time series data into existing workflows.

## 2.5 Benchmarking

The GEE Timeseries Explorer accesses the Earth Engine data catalog efficiently due to WMS technology for image visualization and parallel downloading capabilities. We ran multiple tests for common user operations and registered average response times for RGB visualization of single images, time series aggregations, and time series downloads.

RGB visualization of single cloud-masked Landsat images took on average 4.7 seconds, monthly mean reflectance across entire Germany was displayed on average in 16.6 seconds. Continental-scale Landsat-based visualizations resulted in longer waiting times, e.g., monthly mean reflectance across Europe was rendered in 46.7 seconds. We further assessed the data download performance by repeatedly downloading single-band Landsat time series for 1984–2020 at random locations with global coverage, resulting in an average download time of 4 minutes and 2 seconds per 1,000 points.

## 3. APPLICATIONS

In the following, we demonstrate the benefits of the GEE Timeseries Explorer using multiple application cases from research and higher education.

### 3.1 Research proposals

Students and researchers can employ the plugin in the process of developing thesis or project proposals. The GEE Timeseries Explorer allows users to instantly assess data availability, data quality, and suitability in their areas of interest. Potential pitfalls in analysis workflows, such as data limitations during cloud-prone seasons, or during eras of altered sensor constellations can

thereby be identified and accounted for early on. Furthermore, users can investigate the spectral-temporal behaviour and separability of, e.g., specific land cover types. This allows for efficiently assessing the suitability of time series analysis strategies that rely, e.g., on minimum data densities or seasonal data distributions and can thus help to prevent resource investments into dead-end research projects.

The GEE Timeseries Explorer has, amongst others, been employed to design a study on mapping different crop types in smallholder agricultural systems of Nigeria (Ibrahim et al., in prep). Specifically, the plugin allowed for evaluating the suitability of Landsat and Sentinel-2 data in spatially fragmented systems and revealed the superiority of Sentinel-2 data for the purpose at hand. Furthermore, initial assessments of consistency and quality of data products revealed multitemporal geometric inconsistencies in the Sentinel-2 data archive, which could be identified early on and accounted for during data pre-processing (Rufin et al., 2021). The focus of the work was separating spectrally and phenologically similar crops, as well as mixtures thereof. Inspecting fieldwork locations with the GEE Timeseries Explorer allowed for comparing the phenological behaviour of the target crops and for designing analyses strategies that highlight differences in seasonality, ultimately enabling the production of detailed and accurate crop type maps in this complex smallholder system.

### 3.2 Sample-based analyses

The GEE Timeseries Explorer facilitates sample-based time series analyses without the need to download and process large volumes of image data. Complete time series can be efficiently downloaded for thousands of points stored in vector layers using efficient parallel download capabilities. Downloaded time series can then be integrated into existing workflows for further processing, classifying, and modelling. This is a particular asset in studies relying on rich ground truth datasets from field campaigns, but also standard classification or regression workflows can profit from locally stored point time series for reference data locations to run initial analyses and test model performance on a sample basis before even downloading image data or implementing workflows in a cloud environment.

Recently, the GEE Timeseries Explorer was used in the context of a sample-based mapping of land use in pivot irrigation plots across the Cerrado Biome in Brazil (ANA and INPE, 2021). In this study, time series data for 152,000 samples were downloaded and fed into a processing chain for deriving phenological information and ultimately classification of land use (Bendini et al., 2019). This workflow has been used for Brazil's irrigation atlas (ANA, 2021) and will contribute to an operational monitoring system to assess water consumption across Brazil.

### 3.3 Reference data collection

The GEE Timeseries Explorer allows for efficient reference data collection and labelling for applications requiring time series information for determining, e.g., land cover or land use. The plugin offers a ready-to-use interface for integrating a variety of data sources for reference data collection, while efficiently navigating through large validation samples, e.g., point vector files. Interactive plots allow for assessing detailed intra-annual time series over decadal time frames, such as the predefined cloud-masked Landsat TM, ETM+, and OLI surface reflectance collection covering all years since 1984. Visualizing single images or time series metrics can further inform reference data collection in periods lacking very high-resolution imagery.

In this context, the GEE Timeseries Explorer was used for labelling reference samples to validate maps on long-term agricultural land use around the Aral Sea in Central Asia (Müller et al., 2021). A set of 2,187 validation samples was labelled at annual intervals across the period 1987 through 2019 by eight trained interpreters. The resulting dataset allowed for state-of-the-art accuracy assessment (Olofsson et al., 2014) and deriving unbiased area estimates of irrigated cropland in the region.

### 3.4 Education

The GEE Timeseries Explorer is suitable for remote sensing education such as university seminars or capacity building workshops, where it can be an entry point to the GEE cloud computing platform for inexperienced students, while at the same time offering functionalities relevant to intermediate and experienced students across a diversity of higher education programs.

The plugin enables instant access to a wealth of datasets without the need for proficiency in any programming language. The Python collection editor functionality, however, allows maximum flexibility through the integration of additional collections, or modifications of predefined collections. Students may, e.g., add band indices to existing collections, or modify quality screening procedures. As such, GEE Timeseries Explorer provides an entry point to programming for students in STEM and non-STEM disciplines.

## 4. CONCLUSIONS

We here presented the GEE Timeseries Explorer, a QGIS plugin for seamless integration of Google Earth Engine and QGIS. We demonstrated key functionalities of the plugin and its benefits in the context of research and higher education. The plugin facilitates exploratory analyses of data availability and quality which are crucial for the design and implementation of remote sensing workflows. The GEE Timeseries Explorer enables users to conduct sample-based analyses across large areas without the need to download large volumes of image data and facilitates the generation of reference datasets which rely on time series information, such as land use or land cover change. Lastly, the GEE Timeseries Explorer carries potential for education, offering an entry point for remote sensing students across STEM and non-STEM disciplines, or allows intermediate and advanced students to enhance their remote sensing knowledge and skills.

## ACKNOWLEDGEMENTS

This research contributes to the Landsat Science Team 2018-2023 (<http://www.usgs.gov/land-resources/nli/landsat/landsat-science-teams>) and the Global Land Programme (<https://glp.earth/>). We gratefully acknowledge the open cloud processing platform provided by Google.

## REFERENCES

ANA, 2021. Atlas Irrigação 2021: Uso da Água na Agricultura Irrigada (2ª edição). Edited by Agência Nacional de Águas (ANA). Agência Nacional de Águas (ANA). Brasília.

ANA; INPE, 2021. Colaboração ANA-INPE para o Atlas Irrigação 2021: Monitoramento de pivôs centrais nos polos de agricultura irrigada do Cerrado. Edited by Agência Nacional de

Águas (ANA). Agência Nacional de Águas (ANA). Brasília (NOTA TÉCNICA CONJUNTA, N° 3/2021/SPR/INPE).

Bendini, Hugo N.; Garcia Fonseca, Leila Maria; Schwieder, Marcel; Sehn Körting, Thales; Rufin, Philippe; Del Arco Sanches, Ieda et al., 2019. Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. *International Journal of Applied Earth Observation and Geoinformation* 82, 101872. doi.org/10.1016/j.jag.2019.05.005.

Dong, Jinwei; Metternicht, Graciela; Hostert, Patrick; Fensholt, Rasmus; Chowdhury, Rinku Roy, 2019. Remote sensing and geospatial technologies in support of a normative land system science: status and prospects. *Current Opinion in Environmental Sustainability* 38, 44–52. doi.org/10.1016/j.cosust.2019.05.003.

Frantz, David, 2019. FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing* 11 (9), 1124. doi.org/10.3390/rs11091124.

Frantz, David; Haß, Erik; Uhl, Andreas; Stoffels, Johannes; Hill, Joachim, 2018. Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sensing of Environment* (215), 471–481. doi.org/10.1016/j.rse.2018.04.046.

Gorelick, Noel; Hancher, Matt; Dixon, Mike; Ilyushchenko, Simon; Thau, David; Moore, Rebecca, 2017. Google Earth Engine. Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202, 18–27. doi.org/10.1016/j.rse.2017.06.031.

Ibrahim, E. S.; Nill, L.; Kamali, B.; Rufin, P.; Nendel, C.; Hostert, P., in prep. Mapping crop types and cropping systems in Nigeria with Sentinel-2.

Müller, Daniel; Dara, Andrey; Krause, Christopher; Peña-Guerrero, Mayra Daniela; Schmitz, Tillman; Umirbekov, Atabek et al., 2021. Post-Soviet changes in irrigated crop production in the Amu Darya Basin. *EGU General Assembly Conference Abstracts*. doi.org/10.5194/egusphere-egu21-16266.

Olofsson, Pontus; Foody, Giles M.; Herold, Martin; Stehman, Stephen V.; Woodcock, Curtis E.; Wulder, Michael A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42–57. doi.org/10.1016/j.rse.2014.02.015.

Rufin, Philippe; Frantz, David; Yan, Lin; Hostert, Patrick, 2021. Operational Coregistration of the Sentinel-2A/B Image Archive Using Multitemporal Landsat Spectral Averages. *IEEE Geosci. Remote Sensing Lett.* 18 (4), 712–716. doi.org/10.1109/LGRS.2020.2982245.

Wulder, Michael A.; Coops, Nicholas C.; Roy, David P.; White, Joanne C.; Hermosilla, Txomin, 2018. Land cover 2.0. *International Journal of Remote Sensing* 39 (12), 4254–4284. doi.org/10.1080/01431161.2018.1452075.

Wulder, Michael A.; Masek, Jeffrey G.; Cohen, Warren B.; Loveland, Thomas R.; Woodcock, Curtis E., 2012. Opening the archive. How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* 122, 2–10. doi.org/10.1016/j.rse.2012.01.010.

Zhu, Zhe; Wulder, Michael A.; Roy, David P.; Woodcock, Curtis E.; Hansen, Matthew C.; Radeloff, Volker C. et al., 2019. Benefits of the free and open Landsat data policy. *Remote Sensing of Environment* 224, 382–385. doi.org/10.1016/j.rse.2019.02.016.