Satellite image processing in Google Earth Engine cloud platform

Ulpu Leinonen
University of Turku, Department of Geography and Geology
ulpu.leinonen@utu.fi

CSC Geocomputing Seminar October 8th 2018
Topics of the presentation

• Introducing Google Earth Engine platform

• Case study of GEE / forest plantation mapping in Tanzania
With the masses of RS data available – how to manage, store, analyze?
Goodchild et al. (2012):

“The supply of geographic information from satellite-based and ground-based sensors has expanded rapidly, encouraging belief in a new, fourth, or “big data,” paradigm of science that emphasizes international collaboration, data-intensive analysis, huge computing resources, and high-end visualization.”
Google Earth Engine (GEE)

- A planetary-scale platform for Earth science that allows users to run geospatial analysis on data which is readily in Google’s infrastructure

https://earthengine.google.com/
What started with Google Earth was taken further..

“Often it turns out to be more efficient to move the questions than to move the data.”

-Jim Gray (1944-2007)
Google Earth Engine: petabyte-scale archive of satellite and geospatial data + analysis possibilities

Colocated Data + Computation + APIs

Credit of the slide: Nicholas Clinton / Google Earth Outreach
https://docs.google.com/presentation/d/1T9q6kWigm1MM3p7IEcvNQlpPykedW-IgCCrlqbNeis/edit?usp=sharing
The Earth Engine Public Data Catalog

https://developers.google.com/earth-engine/datasets/

> 200 public datasets
> 5 million images
> 4000 new images every day
> 5 petabytes of data

Credit of the slide: Nicholas Clinton / Google Earth Outreach
https://docs.google.com/presentation/d/1hT9q6kWigM1MM3p7lEcwNqIpPvedW-IgCCrlqbnis/edit?usp=sharing
Online IDE (JavaScript)  

https://code.earthengine.google.com/

```javascript
// Compute the trend of nighttime lights from DMSP.
// Add a band containing image date as years since 1.
function createTimeBand(img) {
  var year = ee.Date(img.get('system:time_start')).
  return ee.Image(year).byte().addBands(img);
}

// Fit a linear trend to the nighttime lights collection.
var collection = ee.ImageCollection('NOAA/DMSP-OLS/MMAP').
  .select('stable_lights').
  .map(createTimeBand);
var fit = collection.reduce(ee.Reducer.linearFit());

// Display a single image
```

Credit of the slide: Nicholas Clinton / Google Earth Outreach

https://docs.google.com/presentation/d/1hT9q6kWigM1M1Mn7IEcyNOIPlPvkedW-lgCCrlqbNels/edit?usp=sharing
Geospatial Datasets

Requests

Results

Algorithmic Primitives

Storage and parallel computing

add  mosaic  distance
filter  join  convolve
focal_min  reduce
After Earth Engine

Credit of the slide: Nicholas Clinton / Google Earth Outreach
https://docs.google.com/presentation/d/1hT9q6kWigM1MM3p7IEcvNQlpPvkedW-IgCCrlqbNeis/edit?usp=sharing
29 years of satellite data
2,068,467 landsat scenes analyzed
909 terabytes of data

More than 2M hours of computation over 66,000 computers

Elapsed time: ~1.5 days to build the mosaics
High-resolution mapping of global surface water and its long-term changes

Jean-François Pekel¹, Andrew Cottam¹, Noel Gorelick² & Alan S. Belward¹

doi:10.1038/nature20584

Image credit: New York Times

Credit of the slide: Nicholas Clinton / Google Earth Outreach

https://docs.google.com/presentation/d/1hT9q6kWigM1MM3p7IEcvNQlpvkedW-lgCCrlqNeis/edit?usp=sharing
Most important wood production area in Tanzania; extremely cloudy

Study area approx. 200,000km²

Existing forest plantation baseline has not been known outside the large industrial-scale plantations

• Potential of the smallholder owned plantations?

FAO/University of Turku case study with GEE:

Mapping forest plantations in Southern Tanzania using Google Earth Engine

• Most important wood production area in Tanzania; extremely cloudy
• Study area approx. 200,000km²
• Existing forest plantation baseline has not been known outside the large industrial-scale plantations
Mapping methodology

- Reference data collection from high and medium resolution satellite imagery in Google Earth, Bing Maps and Earth Engine
  - Participation of 20 Tanzanian experts
Mapping methodology

- Reference data taken into GEE as a fusion table
- Possibility to upload shp added since
Mapping methodology

- Collected data used as training data in classification of satellite imagery (natural forest / planted forest / other land cover)
  - Different classifiers and input combinations tested with validation data
https://github.com/utu-tanzania/sh-plantations
Some results

- Combining optical and radar satellite data and Random Forest classifier provided the best result
  - Data sets used: Landsat-8, Sentinel-1, Sentinel-2 and SRTM elevation & slope
  - Overall accuracy 85±2%

<table>
<thead>
<tr>
<th></th>
<th>Forest plantation</th>
<th>Forest</th>
<th>Other</th>
<th>Total</th>
<th>Map area (ha)</th>
<th>Estimated area (ha)</th>
<th>User's accuracy</th>
<th>Producer's accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest plantation</td>
<td>0.0075</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0089</td>
<td>180011</td>
<td>239842 ± 87023</td>
<td>0.84 ± 0.07</td>
<td>0.96 ± 0.04</td>
</tr>
<tr>
<td>Forest</td>
<td>0.0044</td>
<td>0.3399</td>
<td>0.1100</td>
<td>0.4542</td>
<td>9200524</td>
<td>7132229 ± 425063</td>
<td>0.75 ± 0.04</td>
<td>0.95 ± 0.02</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0.0116</td>
<td>0.5252</td>
<td>0.5369</td>
<td>10874033</td>
<td>12882496 ± 420410</td>
<td>0.98 ± 0.01</td>
<td>0.76 ± 0.04</td>
</tr>
<tr>
<td>Total</td>
<td>0.0118</td>
<td>0.3521</td>
<td>0.6360</td>
<td>1</td>
<td>20254568</td>
<td></td>
<td>0.85 ± 0.02</td>
<td></td>
</tr>
</tbody>
</table>

Source: Koskinen et al., in review at the ISPRS Journal of Photogrammetry and Remote Sensing
On the experience of using GEE..

- Extremely powerful for large area satellite image processing
  - Access to a huge amount of data; currently one of the most significant free satellite image repositories in the world
- Computing in the cloud gives freedom
  - Creation of cloud-free composites for very cloudy regions, multi-temporal and multi-sensor analysis etc.
  - Repeatability and testing without limitations (almost!)
- Constantly updating (data and ready algorithms)
- JavaScript IDE
  - Control over the code; you know what happens behind the hood (ideally)
  - Not all algorithms/functionalities/processed data sets available; limitations for a beginner in coding
  - For example scrutinizing the results was difficult; leads to data transfer between different software
- Highly recommended especially if your study area is outside Finland
Google is responding to the geospatial big data paradigm

Google's mission:

"To organize the world's information and make it universally accessible and useful."

“Often it turns out to be more efficient to move the questions than to move the data.”

-Jim Gray (1944-2007)
To become a user

- [earthengine.google.com/signup](https://earthengine.google.com/signup)
  - The sign up means you request a **trusted tester access** to all the features of the API
  - Google account is needed, because then you can export data and outputs straight in to your Google Drive
- User guide and Help forum extremely useful, start with tutorials
- Basic GEE training at CSC later this year or next, stay tuned!
Contacts & links

• UTU Tanzania Team
  • Tanzania.utu.fi
  • Facebook: UTU Tanzania Team, @ututanzania
• Google Earth Engine
  • https://earthengine.google.com
  • https://code.earthengine.google.com/
• Forest plantation mapping results from the Southern Highlands, Tanzania
  • Participatory mapping of forest plantations with Open Foris and Google Earth Engine (Koskinen, J, Leinonen, U, Vollrath, A, Ortmann, A, Pekkarinen, A, & Käyhkö N, in review at the ISPRS Journal of Photogrammetry and Remote Sensing)
  • Data will become available at the time of publication at https://doi.pangaea.de/10.1594/PANGAEA.894892
Thanks for your attention!

ulpu.leinonen@utu.fi