



World Bank

19th Meeting of the DAC Network on
Development Evaluation

26-27 April 2016

Paris, France

**FOCUS SESSION ON USE OF NEW
TECHNOLOGIES IN M&E AND IMPLICATIONS
FOR EVALUATION**

Use of Big Data in Environmental Evaluation

Juha Uitto

Director, GEF IEO

WHAT WE WILL TALK ABOUT

- ❖ What is big data?
- ❖ Why do we want big data in evaluation?
- ❖ What questions can we answer with big data?
- ❖ Challenges, limitations and lessons from using big data

What is BIG DATA?

- ❖ data sets that are so large or complex that **traditional data processing** applications are **inadequate**
- ❖ Characterized by
 - **Volume** from various sources needing large storage
 - **Velocity** at which they are generated
 - **Variety** of unstructured formats needing additional processing
 - **Value** or meaning not immediately apparent

Adapted from Laney 2001, www.oracle.com and www.sas.com

Why use BIG DATA in evaluation?

- ❖ Scarcer financial resources
 - Need to target interventions where most needed
- ❖ Greater demand for transparency and country ownership
 - Need objective evidence base for decision-making
 - SDGs: 17 goals, 169 targets and 304 proposed indicators



SDGs and Earth Observation

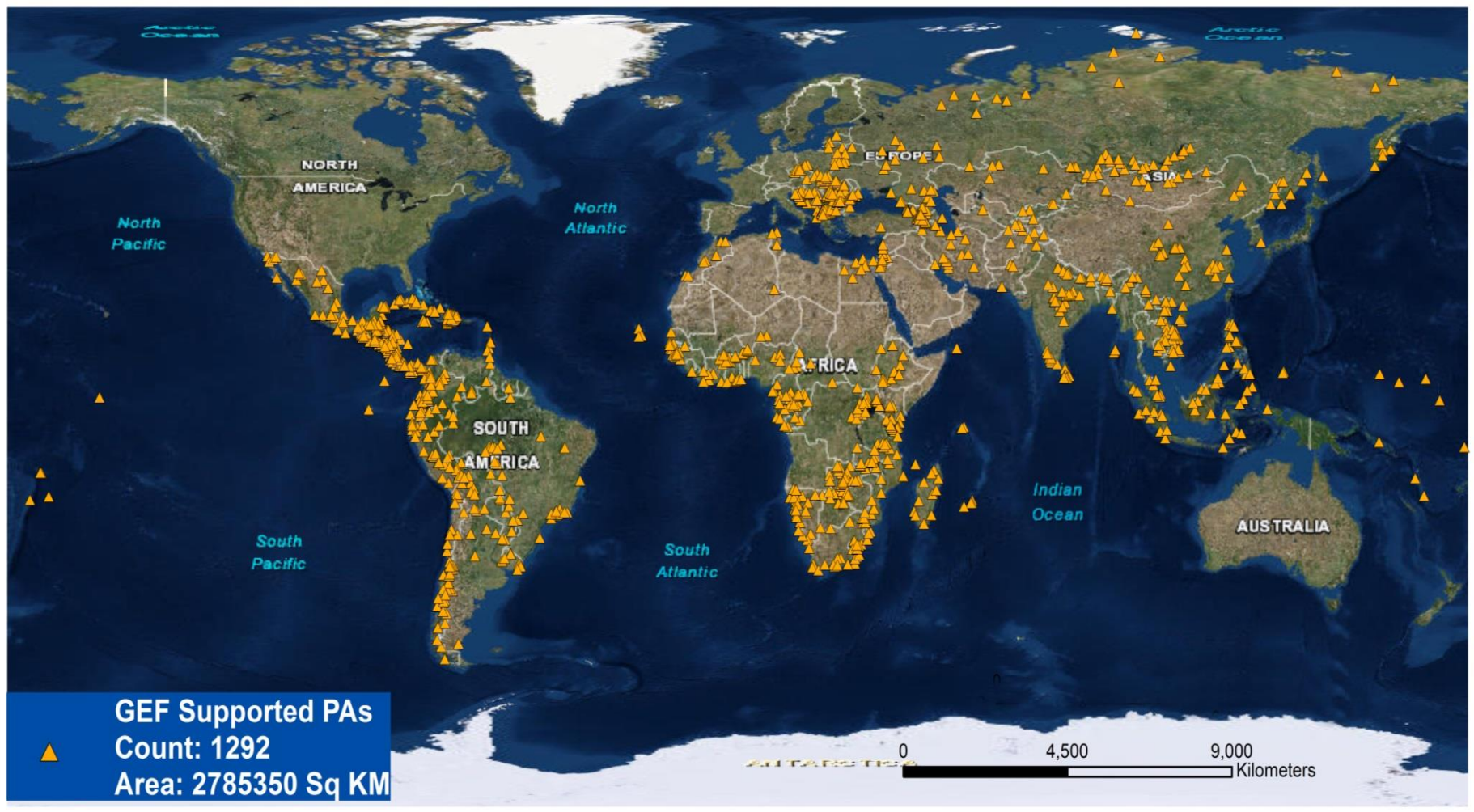


European Space Agency

Big data such as from satellite imagery and sensor networks make environment and development indicators increasingly measurable

What can BIG DATA tell us?

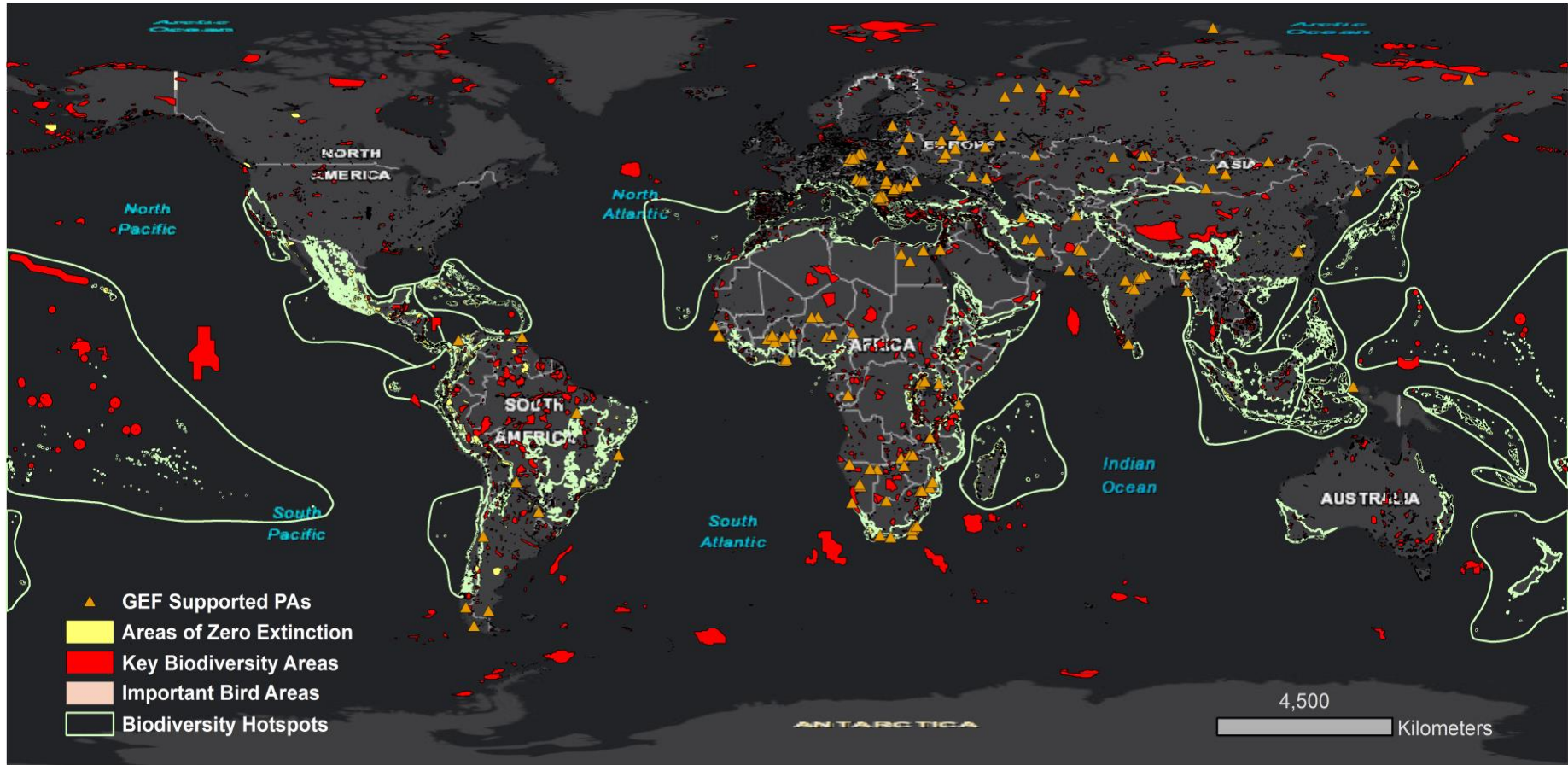
- ❖ Where are the funds going?
- ❖ Is funding going to the right places?
- ❖ What change occurred over time?
- ❖ Did the intervention cause the change?
- ❖ What other factors might have led to the outcome?



Where are the funds going?

Visualization of geographical context

1292 GEF-supported protected areas
~2.8 million km² in 137 countries



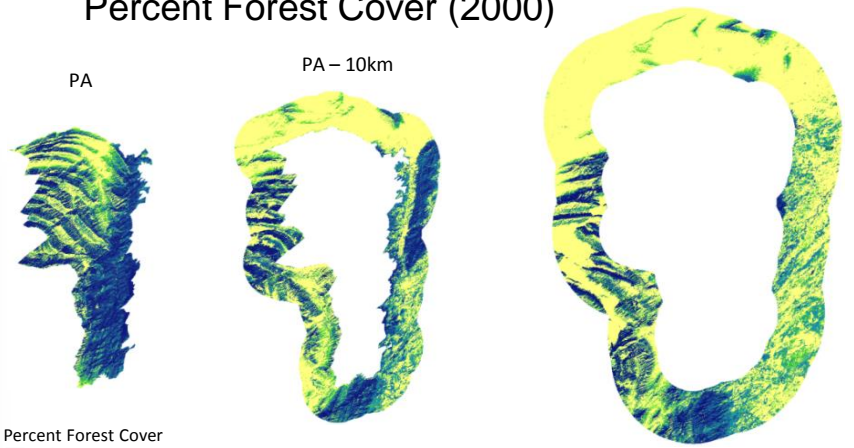
Is funding going to the right places?

Overlay of project sites with scientific criteria

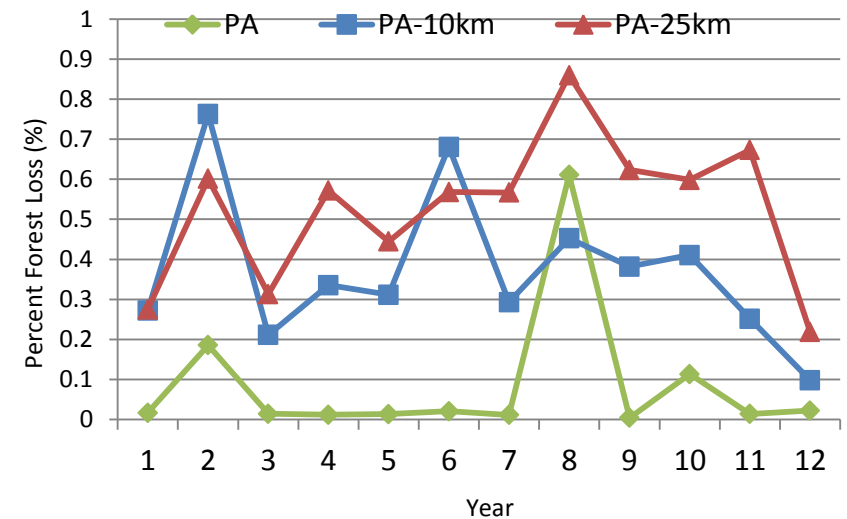
Use of global datasets + GIS analysis to determine overlaps of GEF support with critical sites

Percent Forest Cover (2000)

PA – 25km(excluding the inner)



Annual Percent Forest Loss (2000-2012)



What change occurred over time?

Analysis of forest cover change

Extraction of satellite data for 30,000 GEF and non-GEF sites
 30-m resolution (LANDSAT) for 12-year period

The screenshot displays the Google Earth Engine web interface. At the top, the 'Google Earth Engine' logo and a search bar are visible. The main interface is divided into several panels:

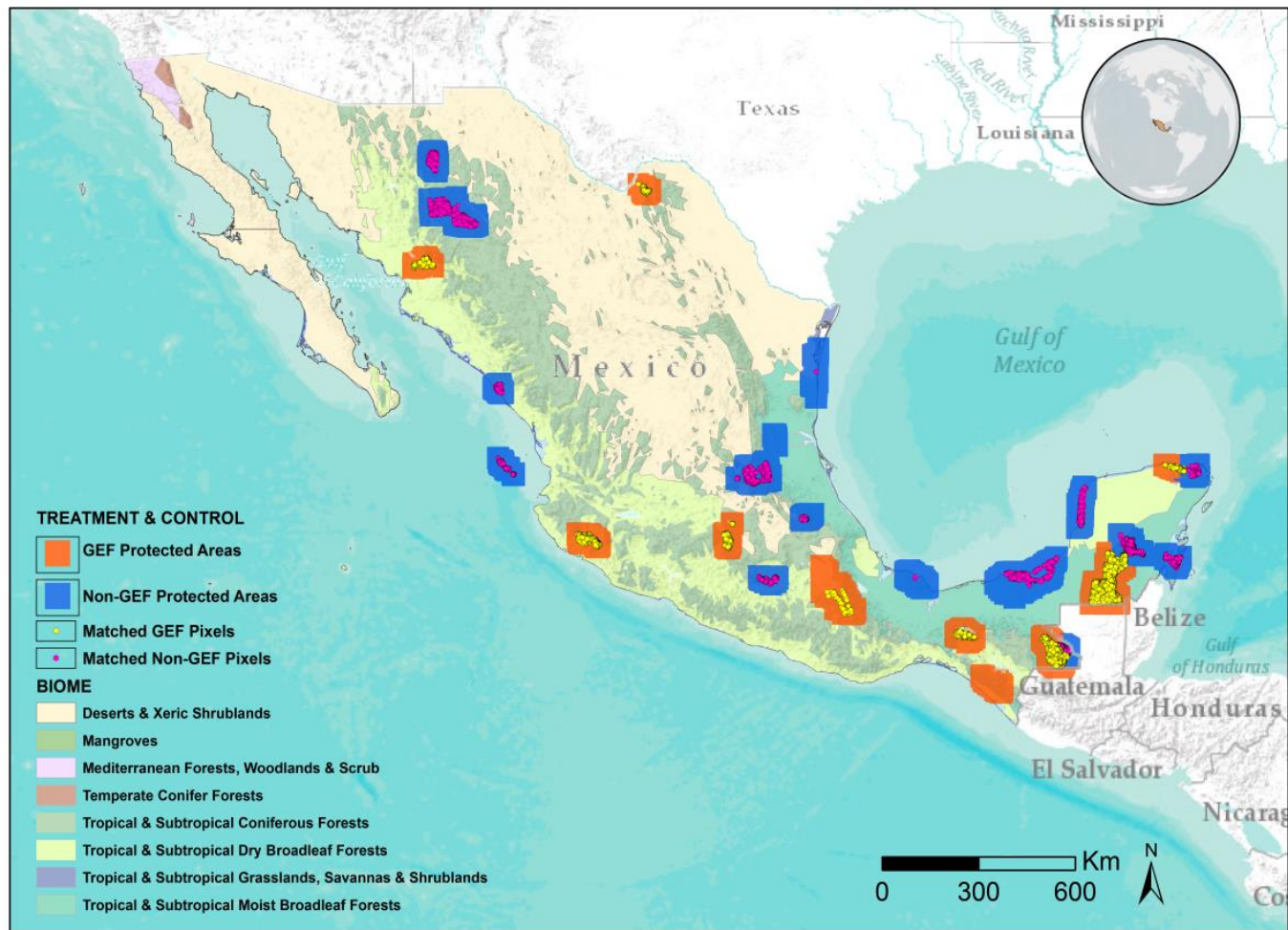
- Scripts Panel (Left):** A list of scripts under the 'Private' category, including 'Array-based spectral unmixing_DC', 'Biodiversity Hotspots from Conservation International', and 'Compute the trend of nighttime lights from DMSP'.
- Code Editor (Center):** A script titled 'Linear Fit' with the following code:

```
Imports (1 entry)
var geometry: GeometryCollection

1 // Compute the trend of nighttime lights from DMSP.
2
3 // Add a band containing image date as years since 1991.
4 function createTimeBand(img) {
5   var year = ee.Date(img.get('system:time_start')).get('year').subtract(1991);
6   return ee.Image(year).byte().addBands(img);
7 }
8
9 // Fit a linear trend to the nighttime lights collection.
10 var collection = ee.ImageCollection('NOAA/DMSP-OLS/NIGHTTIME_LIGHTS')
11   .select('stable_lights')
12   .map(createTimeBand);
13 var fit = collection.reduce(ee.Reducer.linearFit());
14
15 // Display a single image
16 Map.setCenter(30, 45, 4);
17 Map.addLayer(ee.Image(collection.select('stable_lights').first()),
18   {min: 0, max: 63,
19     'stable lights first asset'});
20
21 // Display trend in red/blue, brightness in green
```
- Inspector/Console (Right):** The 'Console' tab is active, showing the instruction: 'Use print(...) to write to this console.'
- Map Panel (Bottom):** A world map visualization showing the results of the script. The map is color-coded, with red and blue representing the linear fit trend and green representing the brightness. A 'Layers' panel on the right shows two checked layers: 'stable lights trend' and 'stable lights first asset'. The map also includes a scale bar, a 'Map/Satellite' toggle, and a 'Polygon drawing' tool.

Planetary level cloud computing with Google Earth Engine

10 years desktop computing = 7 days cloud computing



Did the intervention cause the change?

Quasi-experimental analysis

Propensity score matching found appropriate counterfactuals among 35,351 pixels using 9 socioeconomic and biophysical variables

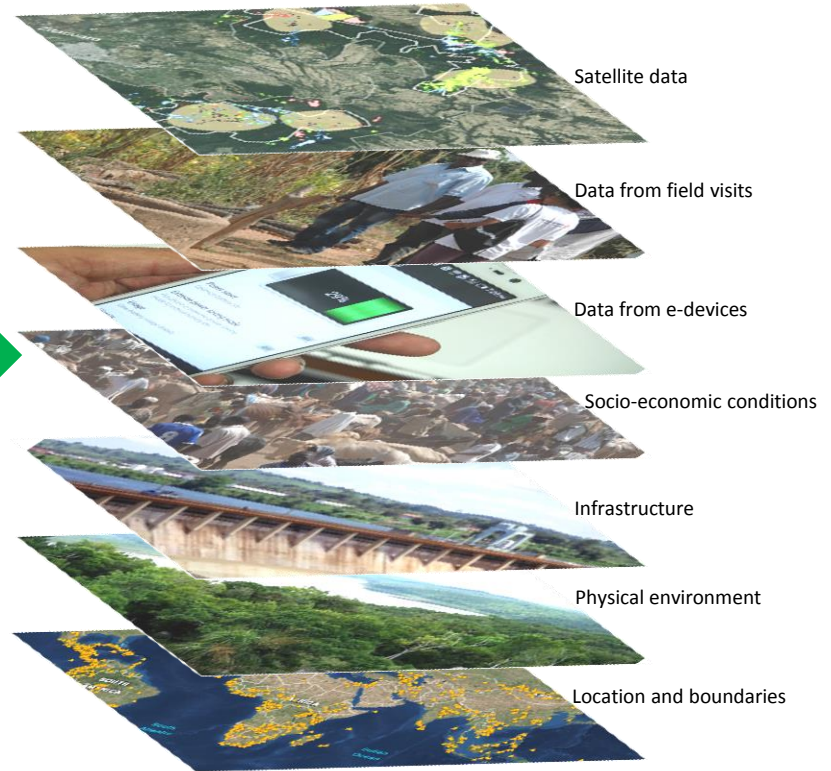
GIS Model

Real World



Problem-Driven

- *Impacts*
- *Causes*
- *Trends*

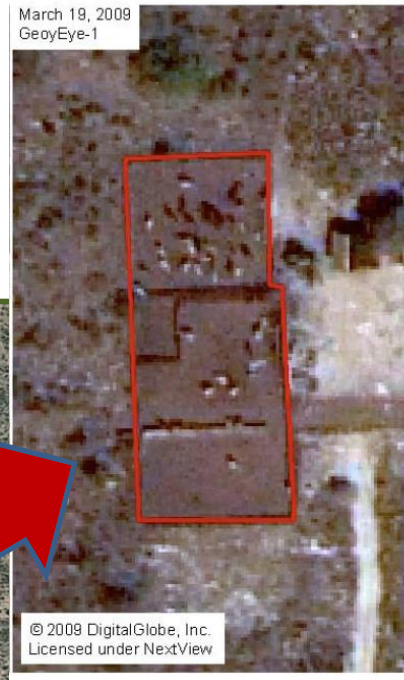


What other factors might have led to the outcome?

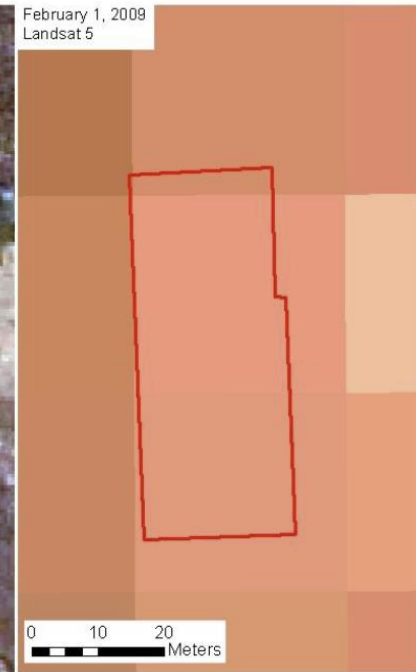
Machine learning and modelling

Data-hungry algorithms required multiple global datasets of contextual variables in different formats to assess correlations with changes

Analysis of high-resolution images



2.5 m



30 m zoomed in
to 2.5 m

Images at 2.5 to 0.5 m resolution used to identify drivers of change that hinder success of GEF support

Challenges and Limitations

- ❖ High computing power and technical skills needed
- ❖ Uneven availability and accuracy of contextual variables
 - often vary widely across countries and sites
- ❖ Cannot answer “how” and “why” questions
- ❖ Data only as good as available resolution
 - still need to do field verification/ groundtruthing
- ❖ Still need to account for possible biases in data collection methods

Solutions and Lessons

- ❖ Partner with global institutions with access to and infrastructure for using big data
- ❖ Used mixed approaches and methods
 - complemented global analyses with case study and portfolio analyses to triangulate findings
- ❖ Continue exploring use of new technology
 - drones, deep learning, internet of things, social media analysis, etc.
- ❖ Approach evaluation as a dynamic learning process
 - new data sets, approaches, issues will always emerge!

GEF Independent Evaluation Office with partners

- Global Land Cover Facility, University of Maryland
- WCPA-SSC Joint Task Force on Biodiversity and Protected Areas at IUCN
- National Aeronautics and Space Administration (NASA)
- AidData

