Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

The Big Data Revolution for Sustainable Development

XI MEETING OF THE LATIN AMERICA AND THE CARIBBEAN MONITORING AND EVALUATION NETWORK

Santiago de Chile, June 28-30, 2016

Anupam Anand

Evaluation Officer

WHAT WE WILL TALK ABOUT

- What is big data?
- Why do we want big data for sustainable development?
- What questions can we answer with big data?
- Challenges, limitations and lessons from using big data

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

What is BIG DATA?

- No fixed definition
- 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

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day

6 BILLION PEOPLE
have cell phones

WORLD POPULATION: 7 BILLION

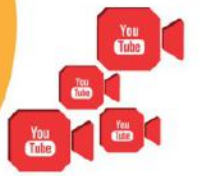
Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

Variety
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month



30 BILLION PIECES OF CONTENT
are shared on Facebook every month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Velocity
ANALYSIS OF STREAMING DATA

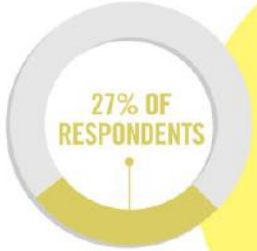
The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



in one survey were unsure of how much of their data was inaccurate

Poor data quality costs the US economy around
$3.1 TRILLION A YEAR

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

IBM

What can we use BIG DATA for?

- **Foster Decision Making and Accountability**
- Where are the funds going?
- Is funding going to the right places?
- **Monitoring & Evaluation**
- What changes occurred over time?
- Did the intervention cause the change?
- What other factors might have led to the outcome?

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

Q1: How many SDG Goals, Targets and indicators are there ?

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

A: SDGs- 17 goals, 169 targets and 230 indicators

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

Why use BIG DATA for SDG?

- Scarcer financial resources
  - Need to target interventions where most needed
- Greater demand for transparency and country ownership
- Monitoring of the progress
- Need objective evidence base for decision-making

IEO Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

SDGs and Earth Observation

European Space Agency

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

IEO Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

The GEF and the SDGs

- GEF support closely aligns with the SDGs on climate, oceans and marine resources, terrestrial ecosystems, forests, biodiversity and land degradation.
- The creation of more than **3,300 protected areas** covering **860 million hectares.**
- Conservation-friendly management of more than **352 million hectares** of productive landscapes and seascapes
- **790 climate change mitigation projects** contributing to **2.7 billion tonnes** of GHG emission reductions
- Sustainable management of **34 transboundary river basins** in **73 countries**

IEO Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

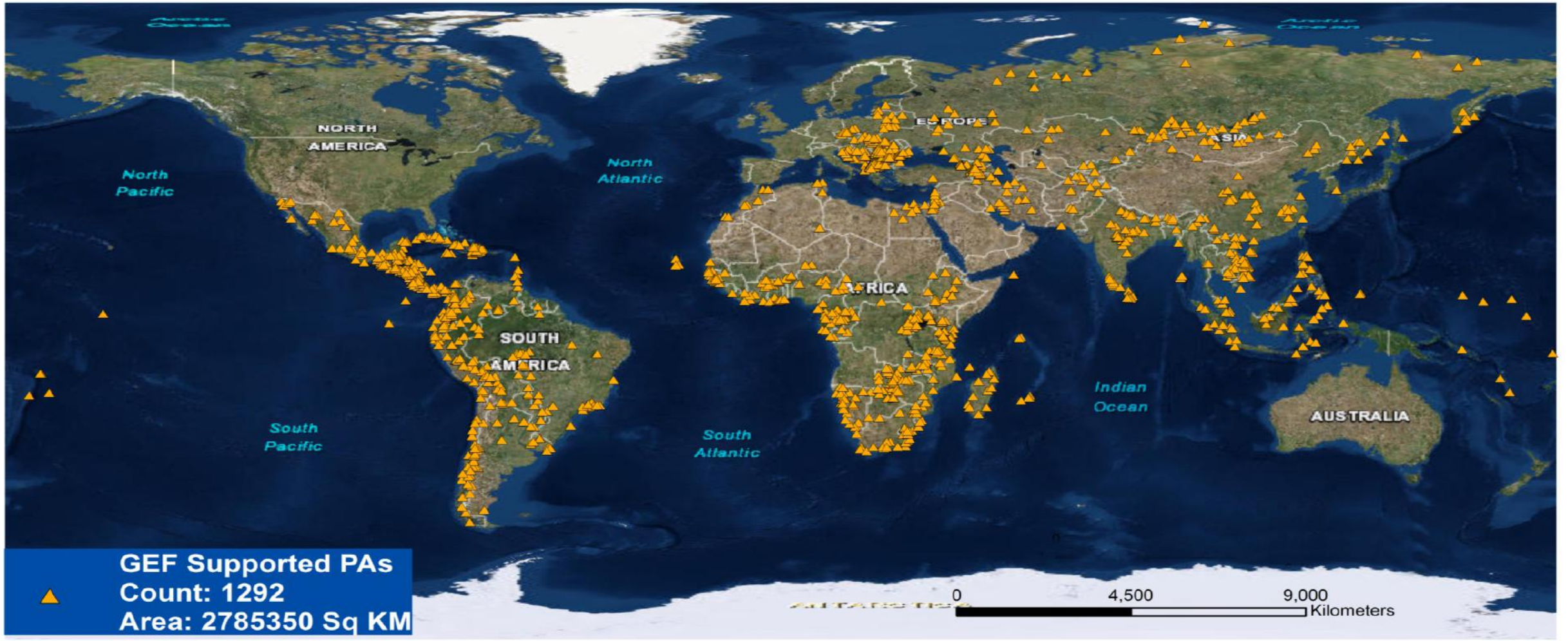
How are we leveraging Big Data at GEF-IEO

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

Big data for Biodiversity

- Goal 15: **Sustainably manage forests,** combat desertification, halt and reverse land degradation, **halt biodiversity loss**
- Indicators
  - Annual change in forest area and land under cultivation* – Geospatial data
  - Area of forest under sustainable forest management as a percent of forest area – Geospatial data/Administrative data
  - Red List Index - Telemetry Tracking Data/International monitoring
  - Protected areas overlay with key biodiversity areas(KBAs)

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY



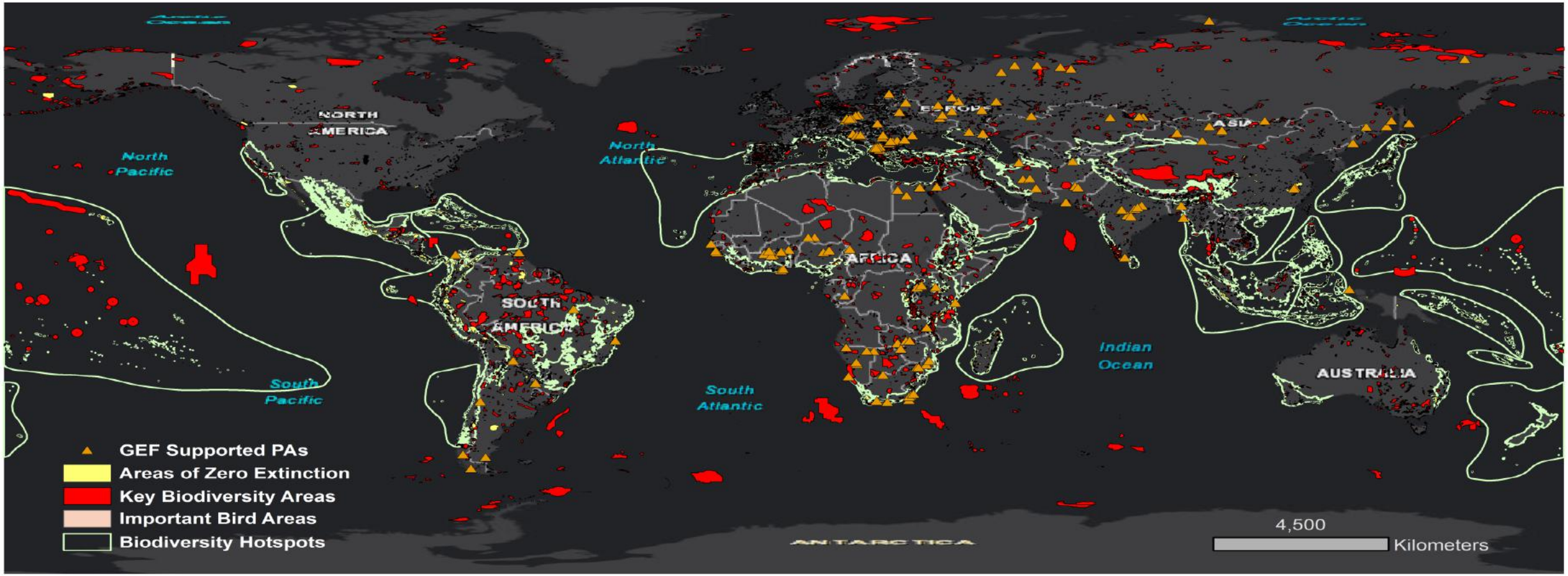
Where are the funds going?

Visualization of geographical context

1292 GEF-supported protected areas

~2.8 million km^2 in 137 countries

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

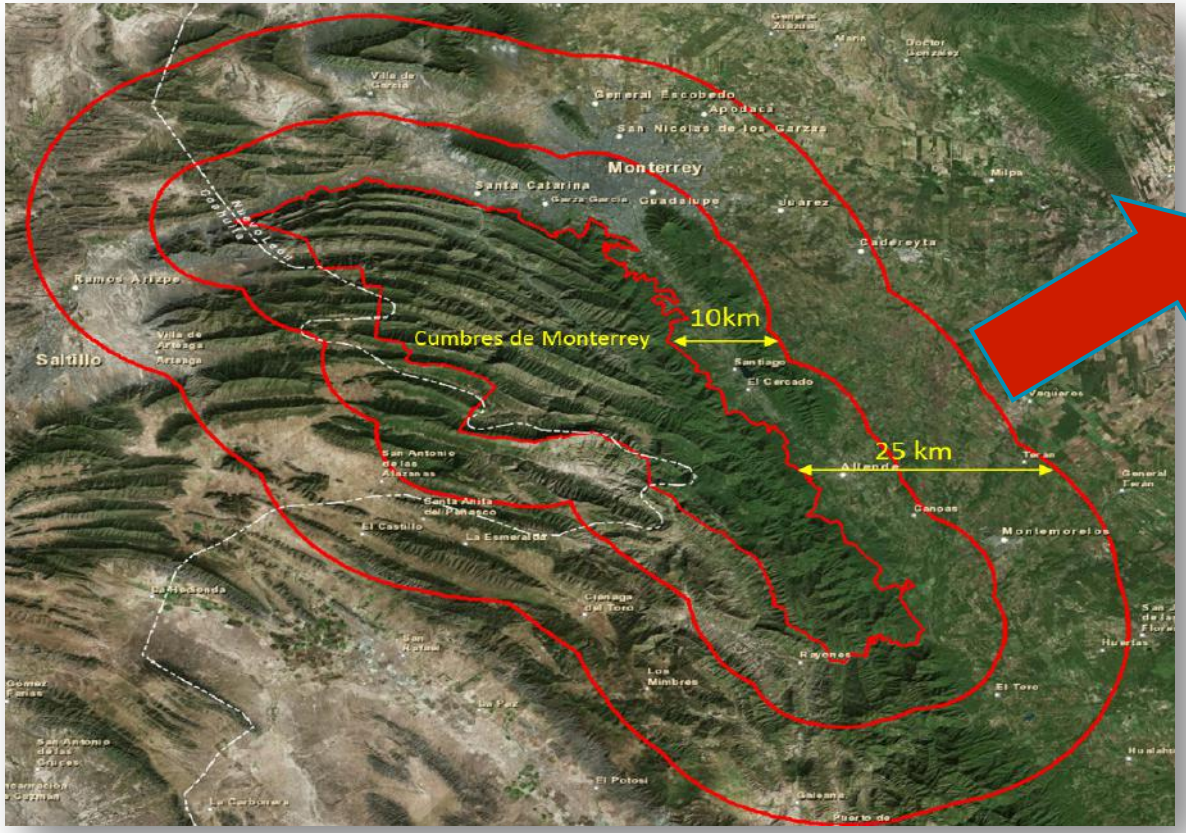


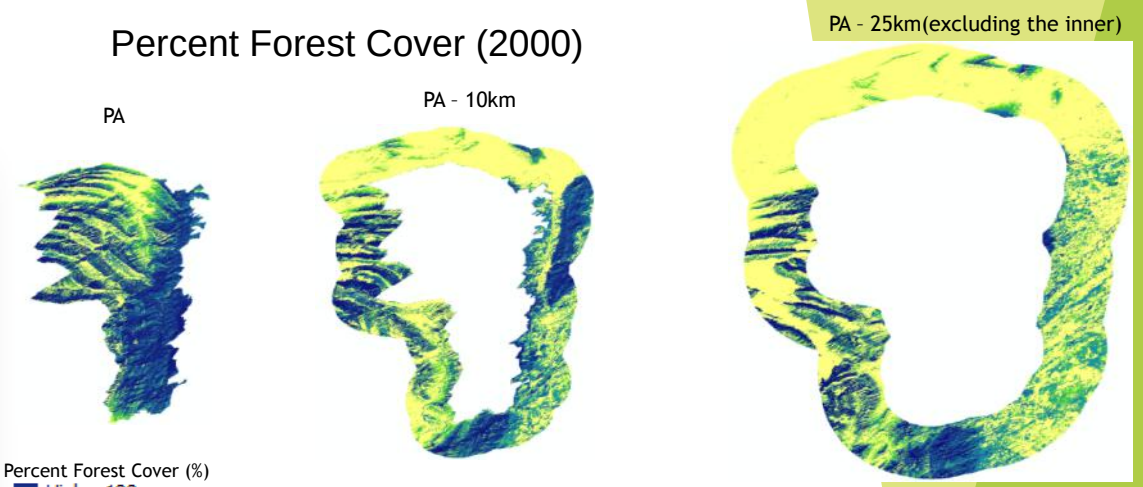
Is funding going to the right places?

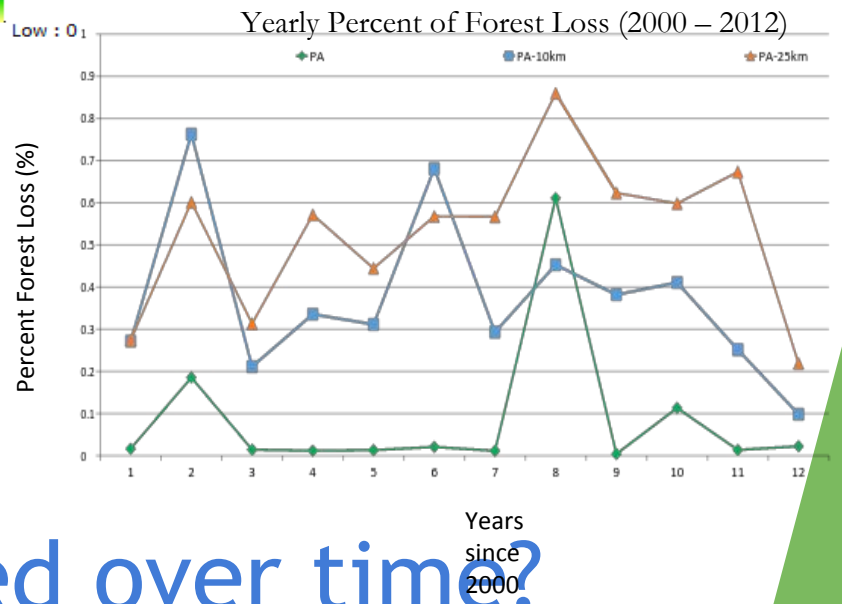
Overlay of project sites with scientific criteria

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

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY







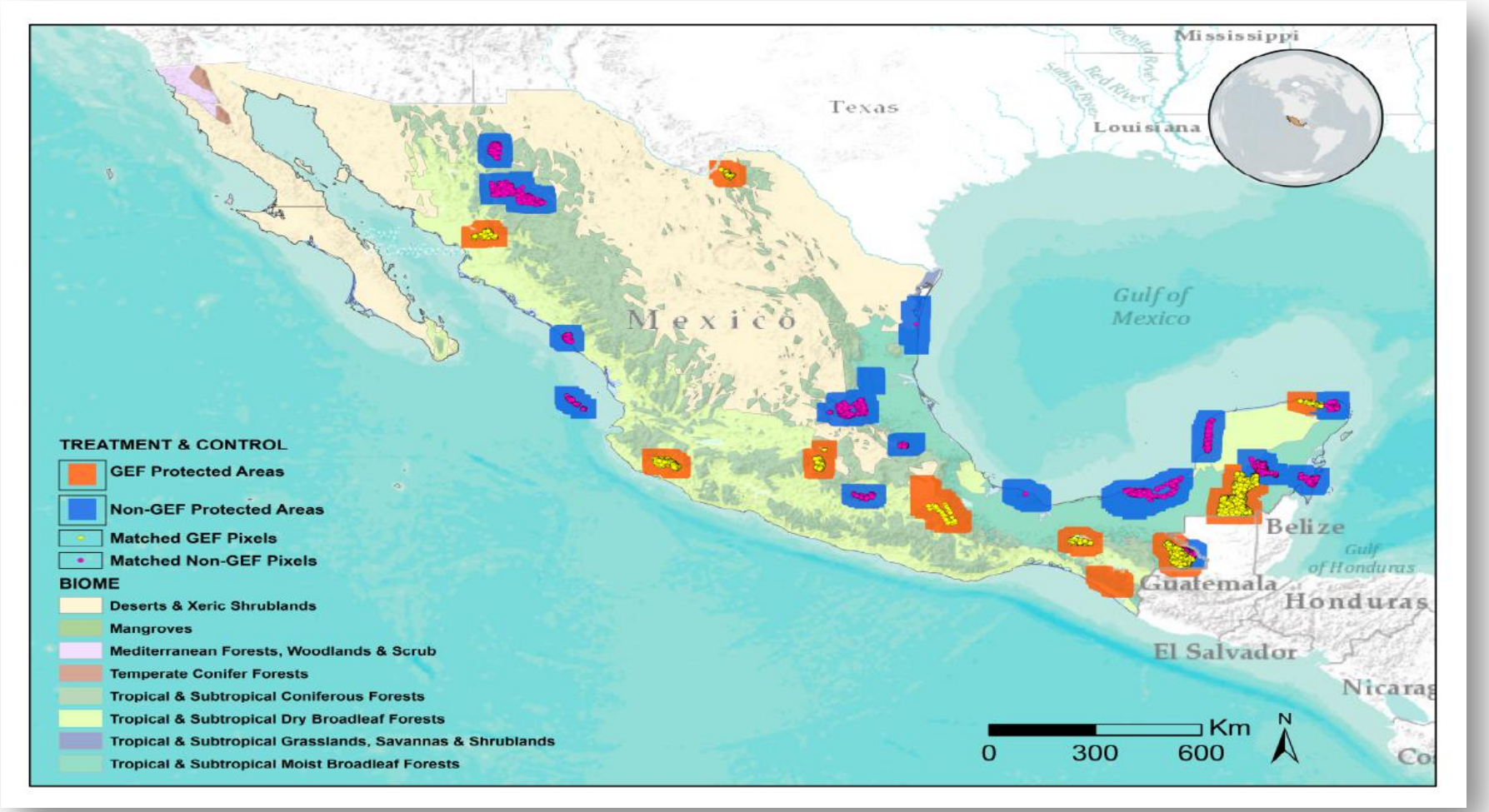
What changes occurred over time?

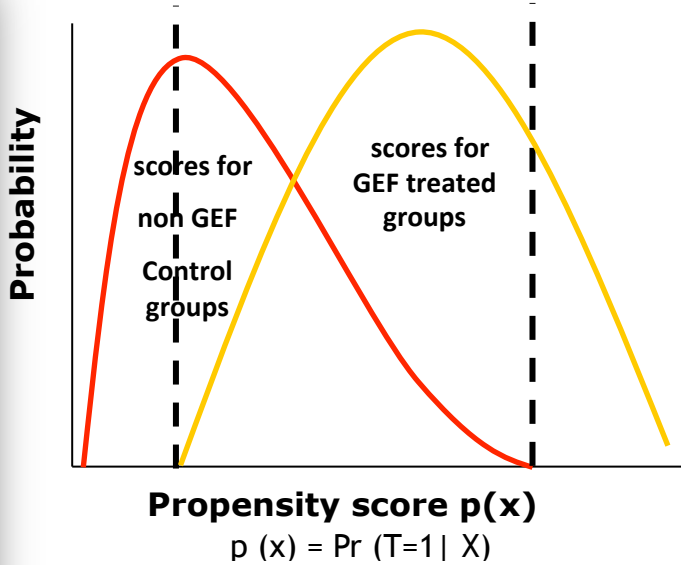
Analysis of forest cover change

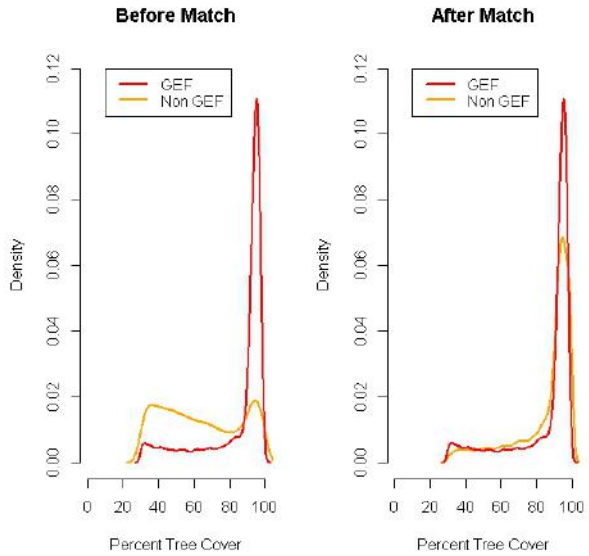
Extraction of satellite data for 30,000 GEF and non-GEF sites

30-m resolution (LANDSAT) for 12-year period

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY







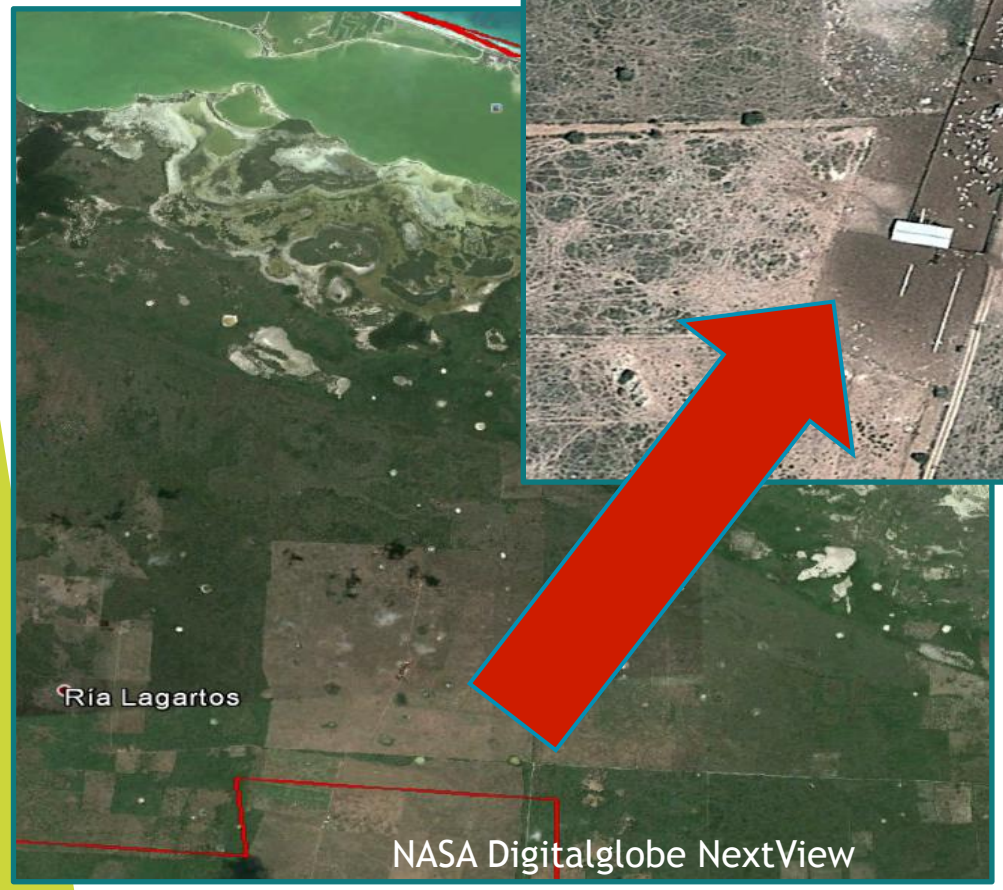
Did the intervention cause the change?

Quasi-experimental analysis

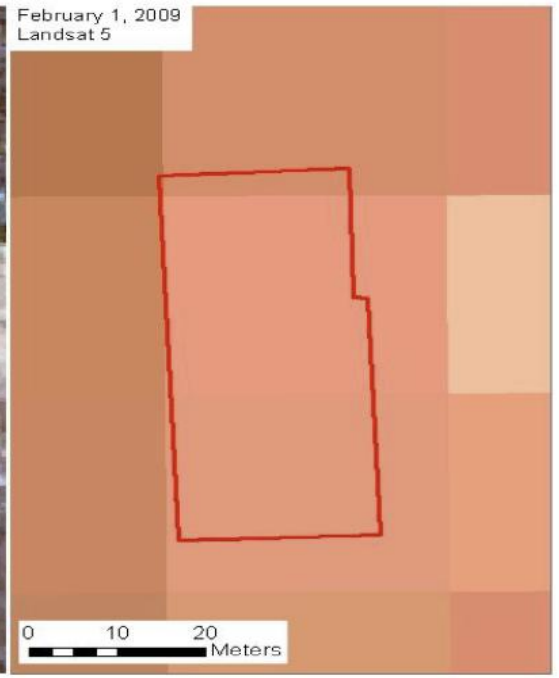
Propensity score matching found appropriate counterfactuals using 9 socioeconomic and biophysical variables

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

Identify the drivers



2.5 m



30 m zoomed in to 2.5 m

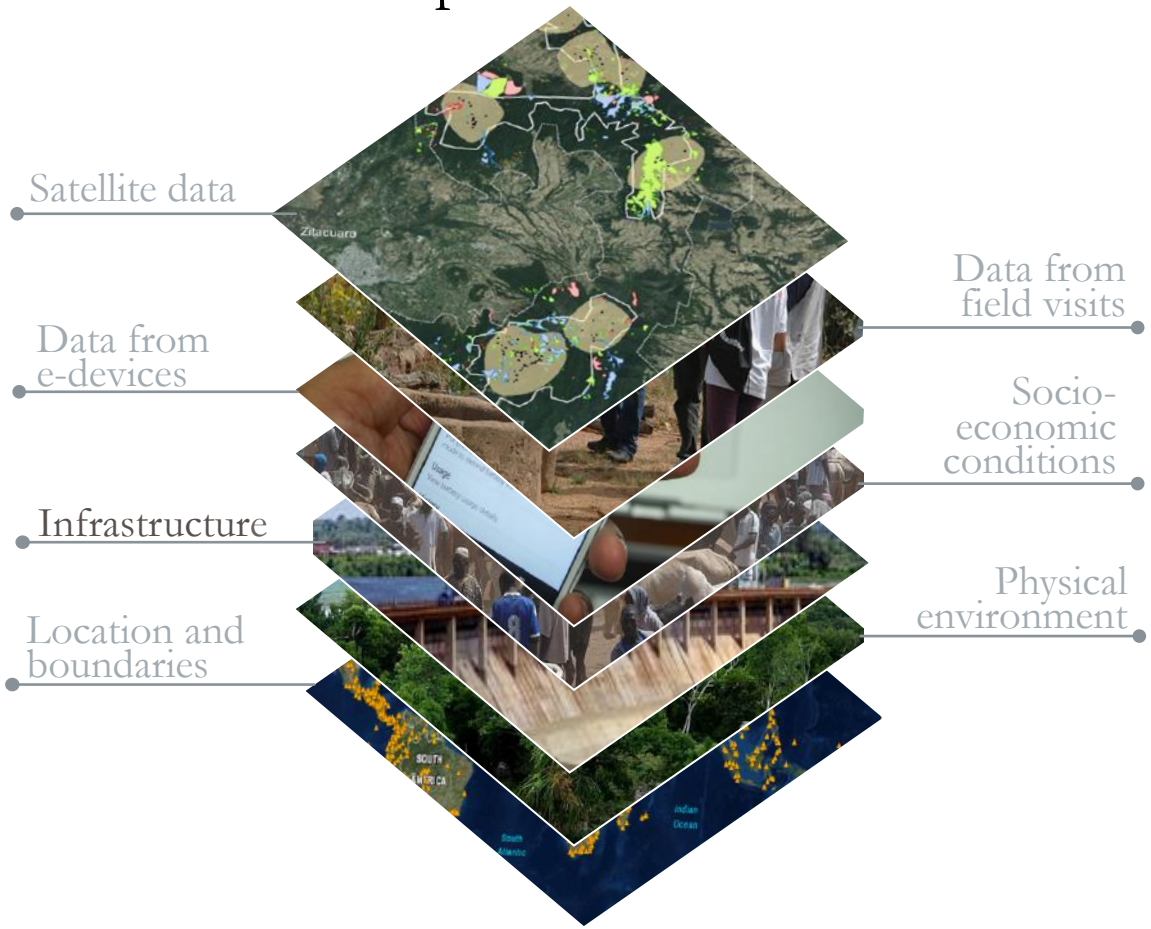
Images at 2.5 to 0.5 m resolution used to identify drivers of change that hinder success of GEF support

Real World

Problem-Driven

To assess
- Impacts
- Causes
- Trends

Spatial Model



What other factors might have led to the outcome?

Use of contextual variables in different formats to assess correlations with changes

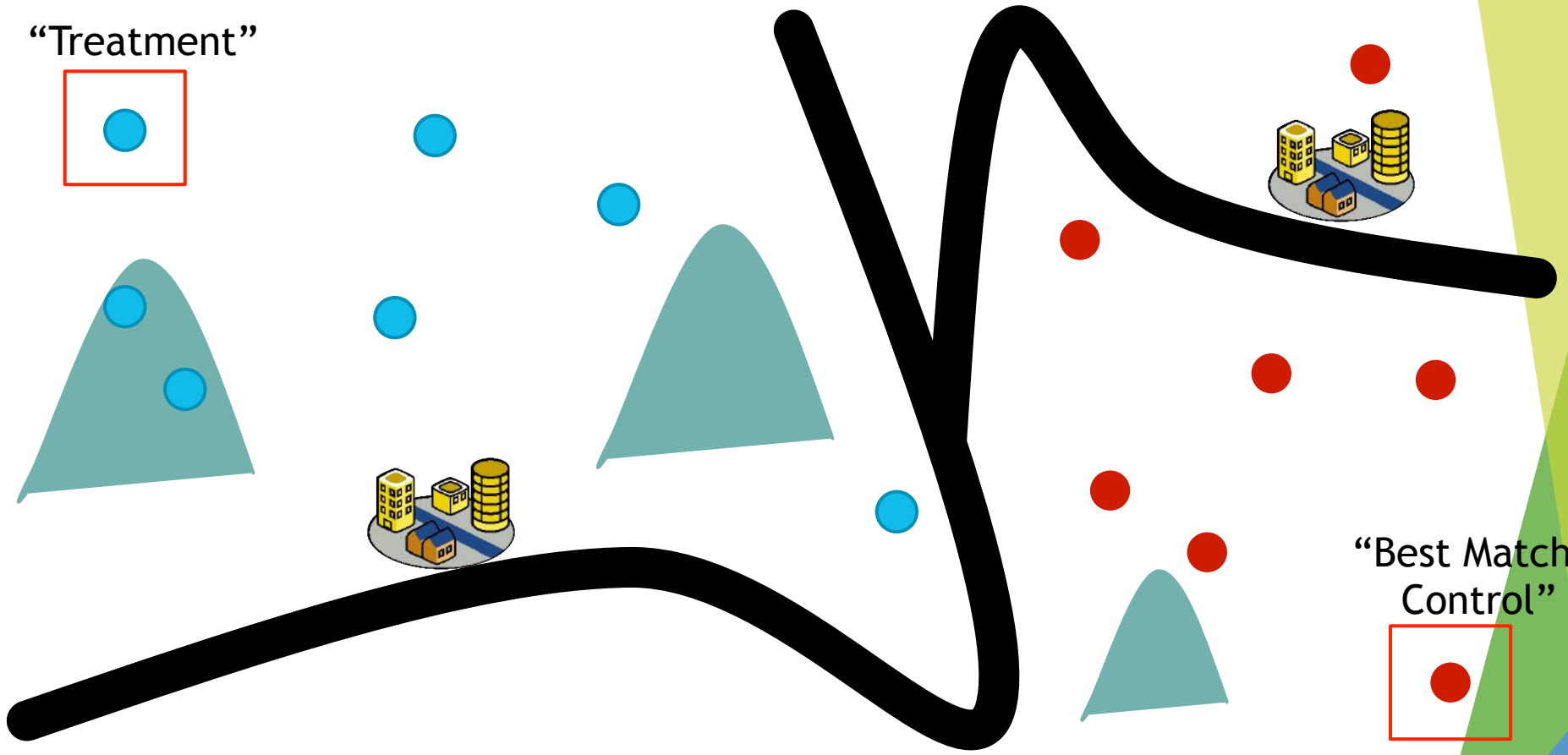
Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

Big data for Land degradation

- Goal 15: Sustainably manage forests, combat desertification, halt and reverse **land degradation,** halt biodiversity loss
  - Indicator for Goal 15
  - Annual change in degraded or desertified arable land (% or ha) - Remote sensing/ satellite and administrative data.
- UNCCD Indicators for Land Degradation Neutrality(LDN)
  - Vegetation productivity (NDVI)
  - Landuse and landcover change and
  - Carbon sequestration

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

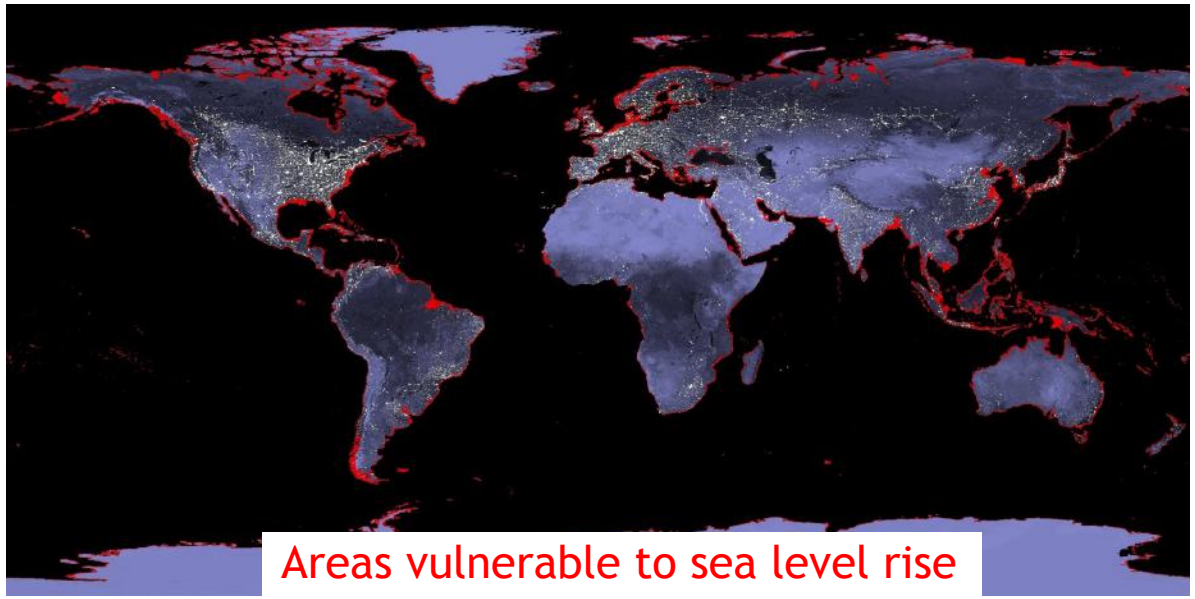
Geospatial Impact Evaluation



- GEF Project Locations (i.e., area under restortion project)
- Candidate Control Locations (can search across any set of relevant geographies)

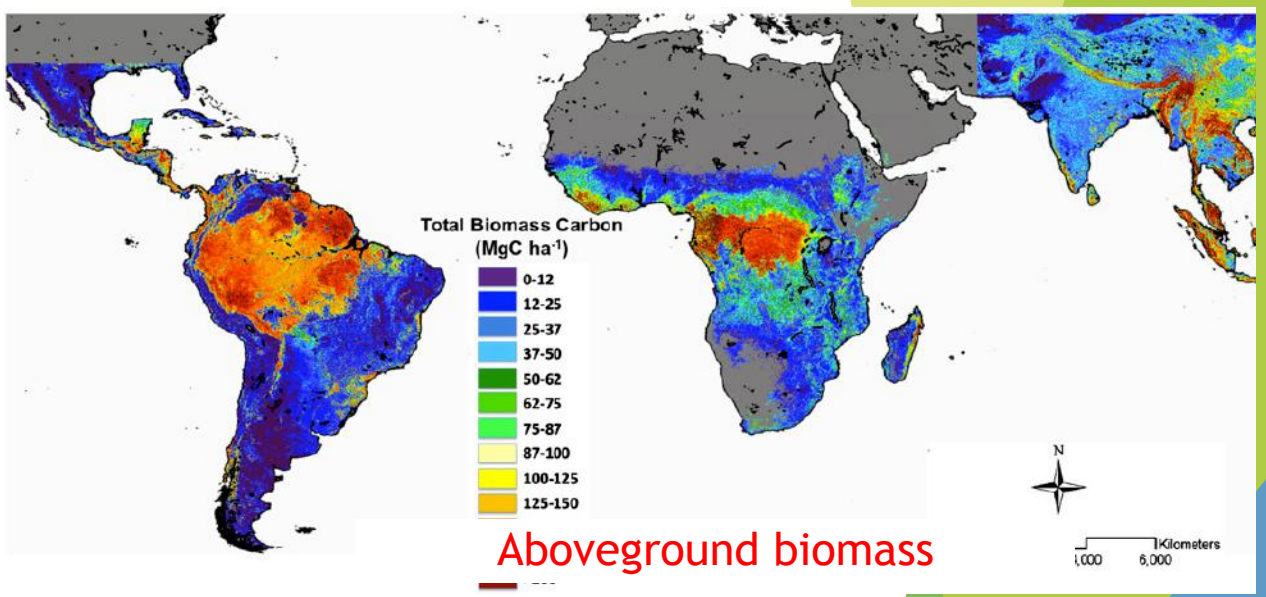
Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

Image: Aiddata



Areas vulnerable to sea level rise

Image: NASA



Aboveground biomass

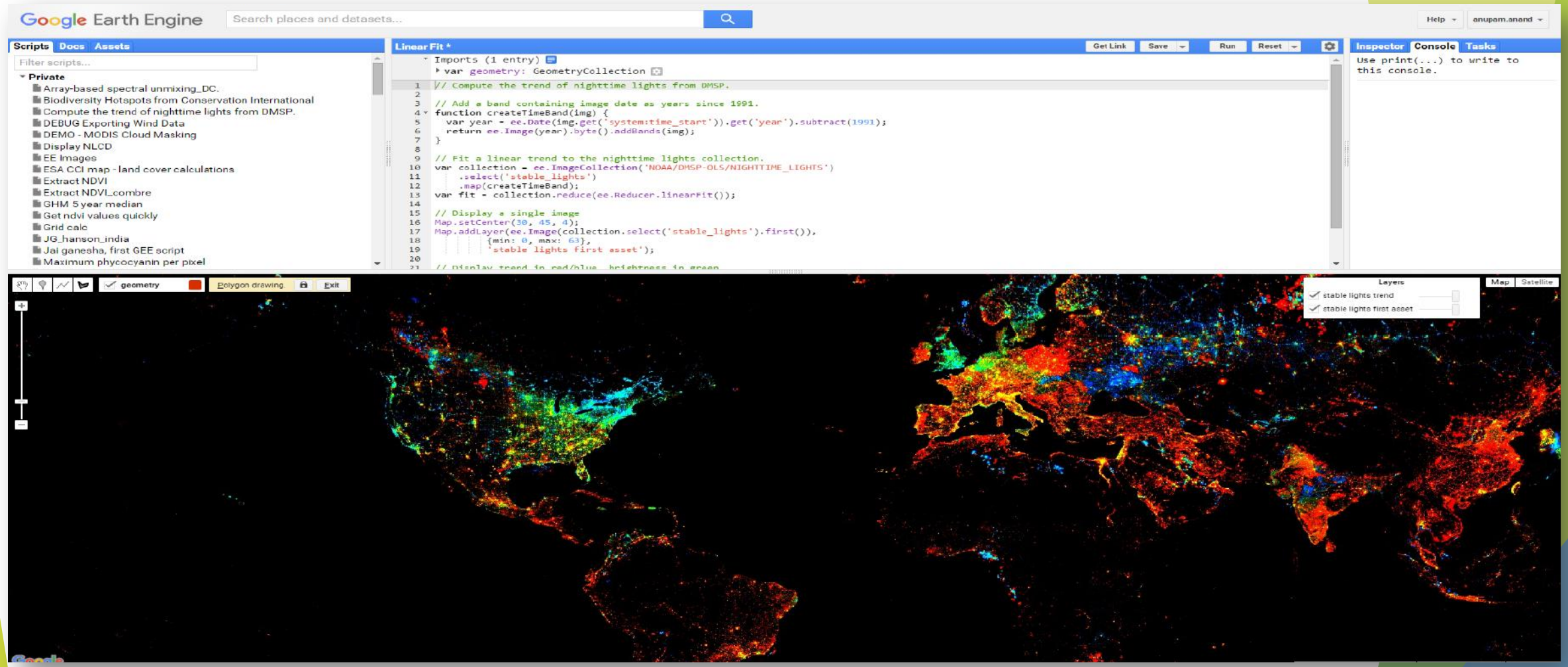
Saatchi et al, PNAS, 2011

Big data for Climate Action

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

13 CLIMATE ACTION

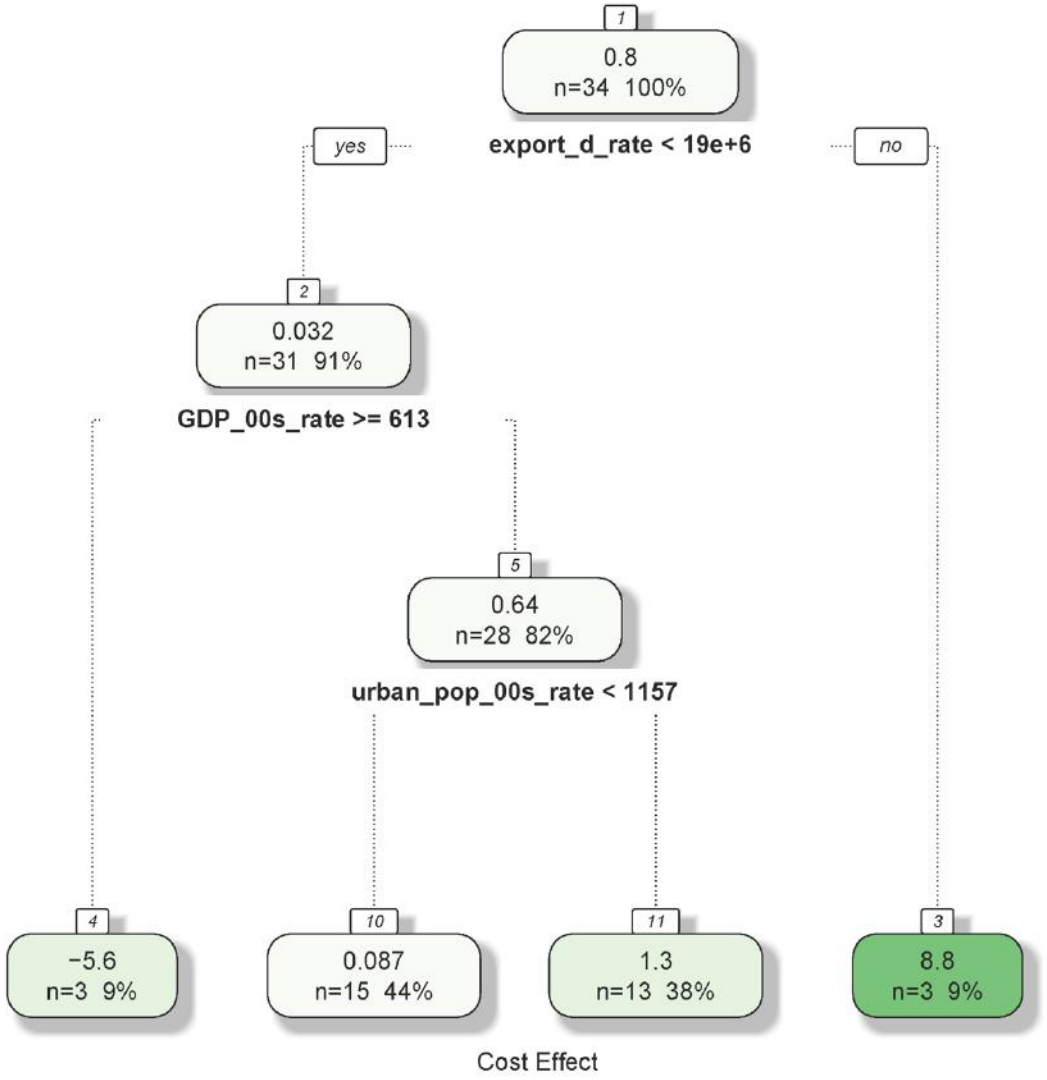
Big data needs big tools



Planetary level cloud computing with Google Earth Engine

10 years desktop computing = 7 days cloud computing

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY



Data	Sources
agricultural production	FAO, 2012
export of agricultural product	FAO, 2012
trade of agricultural product	FAO, 2012
urban population	FAO, 2012
rural population	FAO, 2012
Gross domestic product	world bank, 2015
rule of law	world bank, 2013
control of corruption	world bank, 2013
monitoring capacity	Romijin et al (2012)
International aid	Aid data (2010)

Variables and its sources used in the regression and decision tree analysis

Machine learning and modelling

Data-hungry algorithms required multiple global datasets of

Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

24

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/ ground truthing
- Still need to account for possible biases in data collection methods
- Legal issue

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, sentiment analysis, social media analysis, etc.
- Approach evaluation as a dynamic learning process
  - new data sets, approaches, issues will always emerge!

PARTNERS

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

IUCN

IEO Independent Evaluation Office
GLOBAL ENVIRONMENT FACILITY

GLCF landcover.org

UNIVERSITY OF MARYLAND 18 56

NASA

For more information, visit **www.gefieo.org**