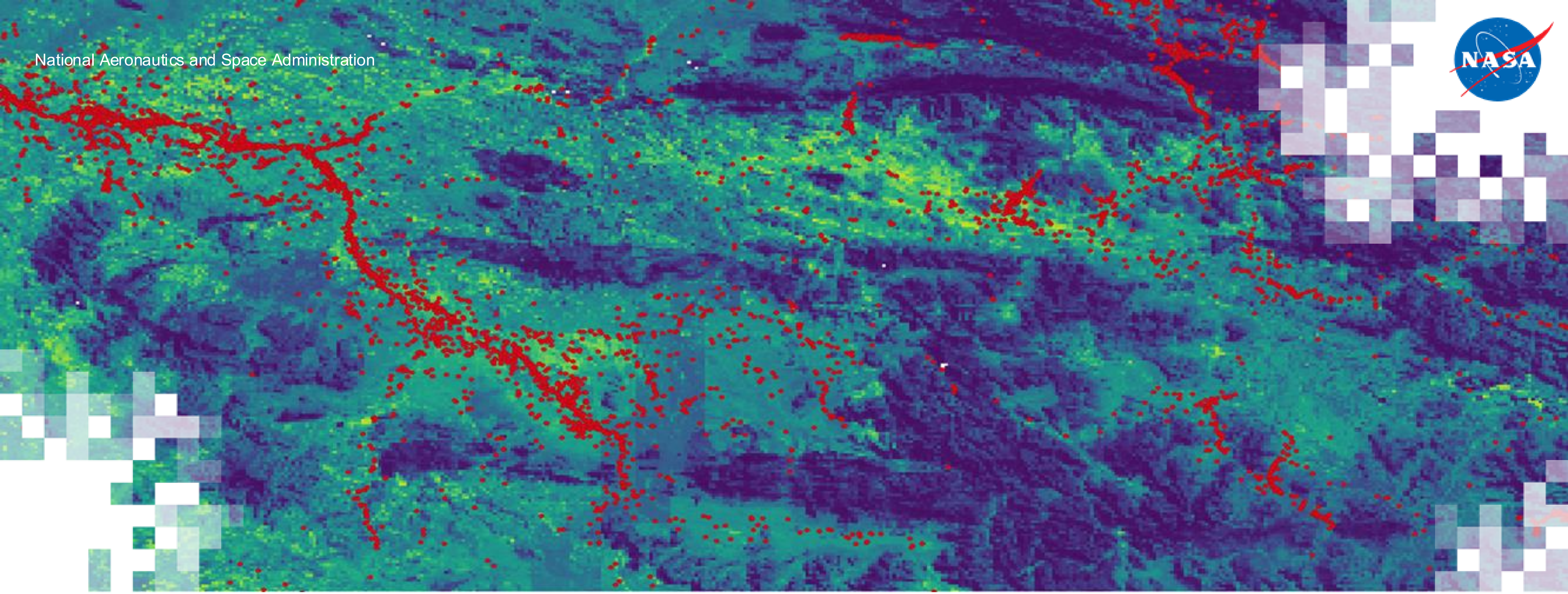
National Aeronautics and Space Administration



Species Distribution Modeling with Google Earth Engine

Part 1: Foundations of Species Distribution Modeling with GEE

ARSET Hosts: Justin Fain (NASA Ames/BAERI) & Sativa Cruz (NASA Ames/BAERI)
Guests: Ramiro Crego (University College Cork), Grant Connette (Smithsonian), Jared Stabach (Smithsonian)

July 07, 2026

ARSET Eco Team

Justin Fain

Research Scientist
BAER/NASA Ames Research Center

Sativa Cruz

Applied Scientist
BAER/NASA Ames Research Center

Juan Torres-Perez

Research Scientist
NASA Ames Research Center

Savannah Cooley

Research Scientist
BAER/NASA Ames Research Center

Part 1 – Trainers

Ramiro D Crego

Lecturer in Animal Ecology

University College Cork

Grant Connette

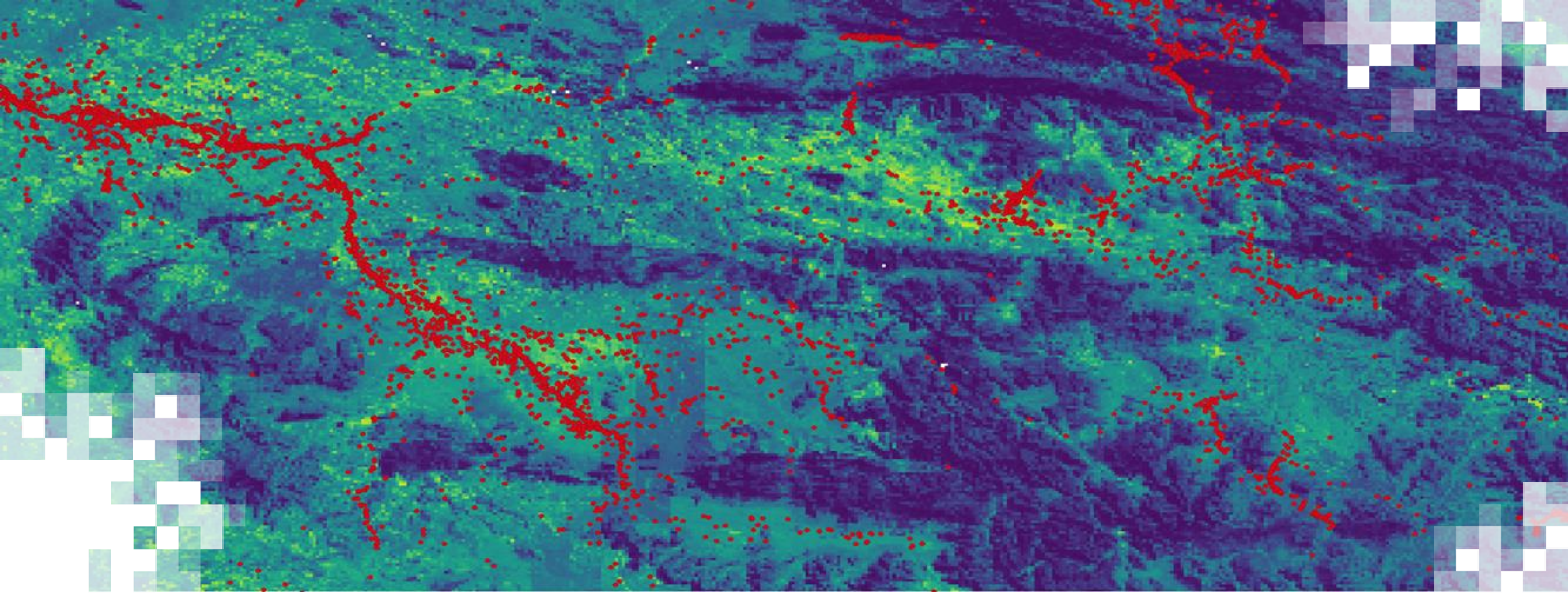
Research Ecologist

Smithsonian's National Zoo & Conservation Biology Institute

Jared Stabach

Research Ecologist

Smithsonian's National Zoo & Conservation Biology Institute



About ARSET

About ARSET

- **ARSET provides accessible, relevant, and cost-free training on remote sensing satellites, sensors, methods, and tools.**

- Trainings include a variety of applications of satellite data and are tailored to audiences with a variety of experience levels.

AGRICULTURE

DISASTERS

ECOLOGICAL CONSERVATION

HEALTH & AIR QUALITY

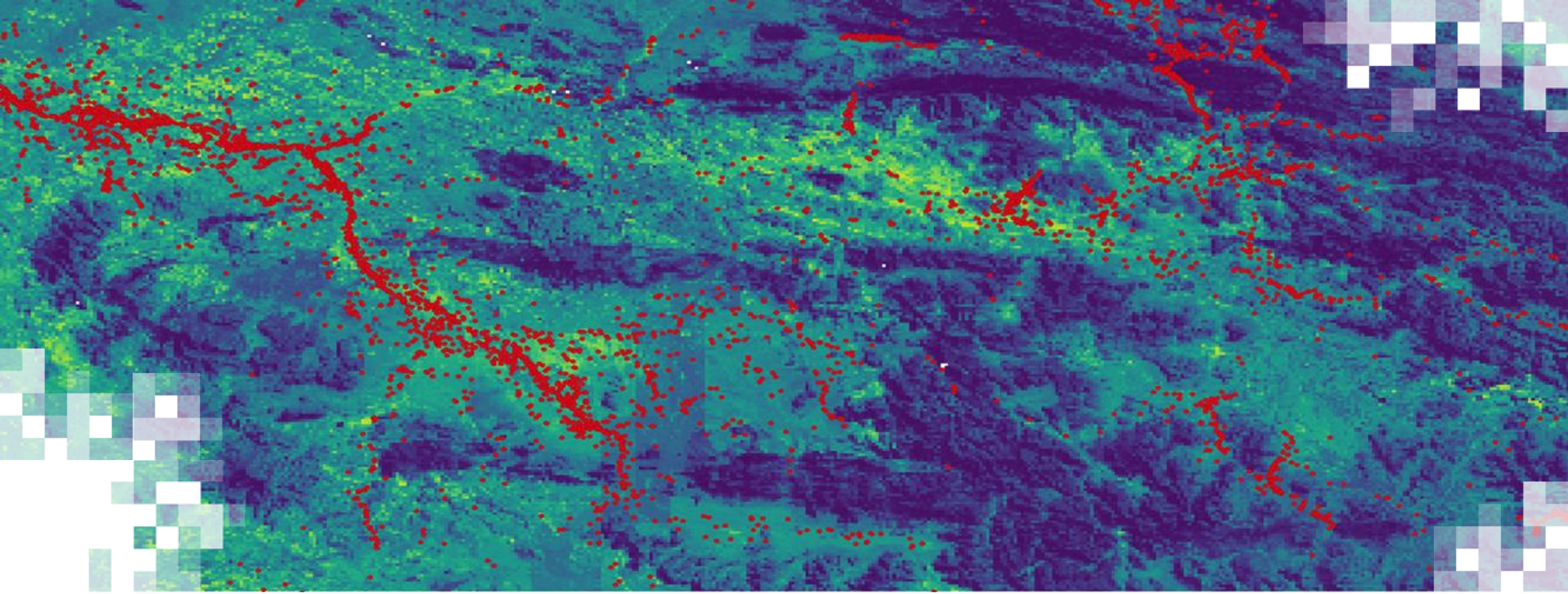
WATER RESOURCES

WILDLAND FIRES

About ARSET Trainings

- Online or in-person
- Live and instructor-led or asynchronous and self-paced
- Cost-free
- Bilingual and multilingual options
- Only use open-source software and data
- Accommodate differing levels of expertise

- Visit the ARSET website to learn more.



Species Distribution Modeling with Google Earth Engine Overview

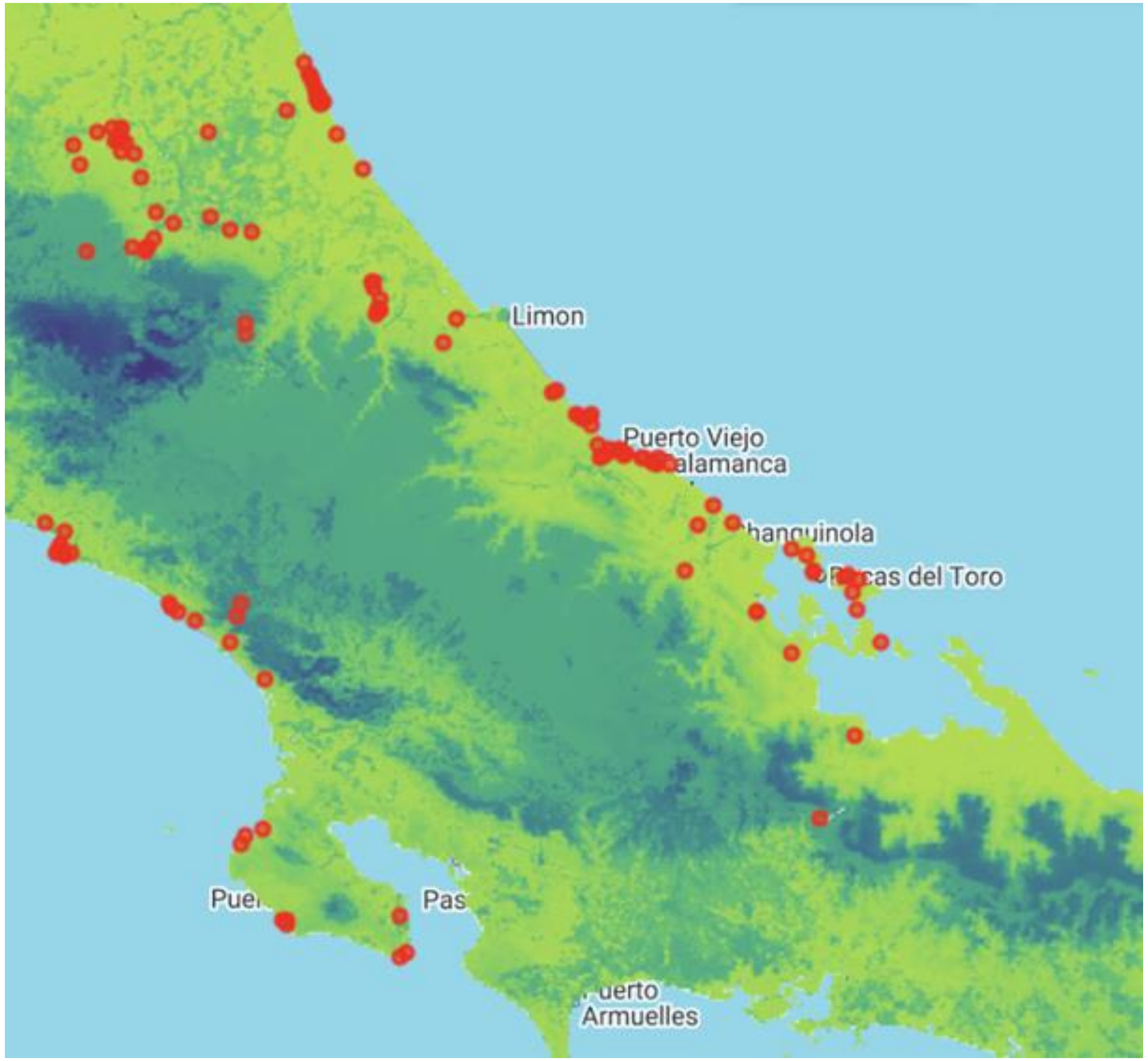
Training Learning Objectives

By the end of this training, participants will be able to:

- Conduct basic operations and functions in the Google Earth Engine (GEE) web interface using JavaScript code.
- Identify core concepts and applications of species distribution modeling (SDM).
- Assess key SDM workflow decisions to ensure occurrence point independence, to select background points, and to choose ecologically relevant predictors.
- Access and manage analysis-ready spatial data from the GEE catalog as predictor variables.
- Fit species distribution models using machine learning models (e.g., Random Forest) in GEE using a reproducible code-based workflow.
- Critically evaluate the appropriateness of the model design and interpret predicted distribution and habitat suitability results.

First Looks

- Predicted habitat suitability based on observed species presence



Prerequisites

- Part 2 of Visualizing Land Cover and Land Use Change with NASA Satellite Imagery
- General knowledge of GIS
- Some coding experience

Training Outline

Part 1

Foundations of Species Distribution Modeling with GEE

July 07, 2026

Session A: 9:00 a.m. to 10:30 a.m. EDT (UTC-4)

Session B: 4:00 p.m. to 5:30 p.m. EDT (UTC-4)

Part 2

Building, Evaluating, and Interpreting a SDM on GEE

July 14, 2026

Session A: 9:00 a.m. to 10:30 a.m. EDT (UTC-4)

Session B: 4:00 p.m. to 5:30 p.m. EDT (UTC-4)

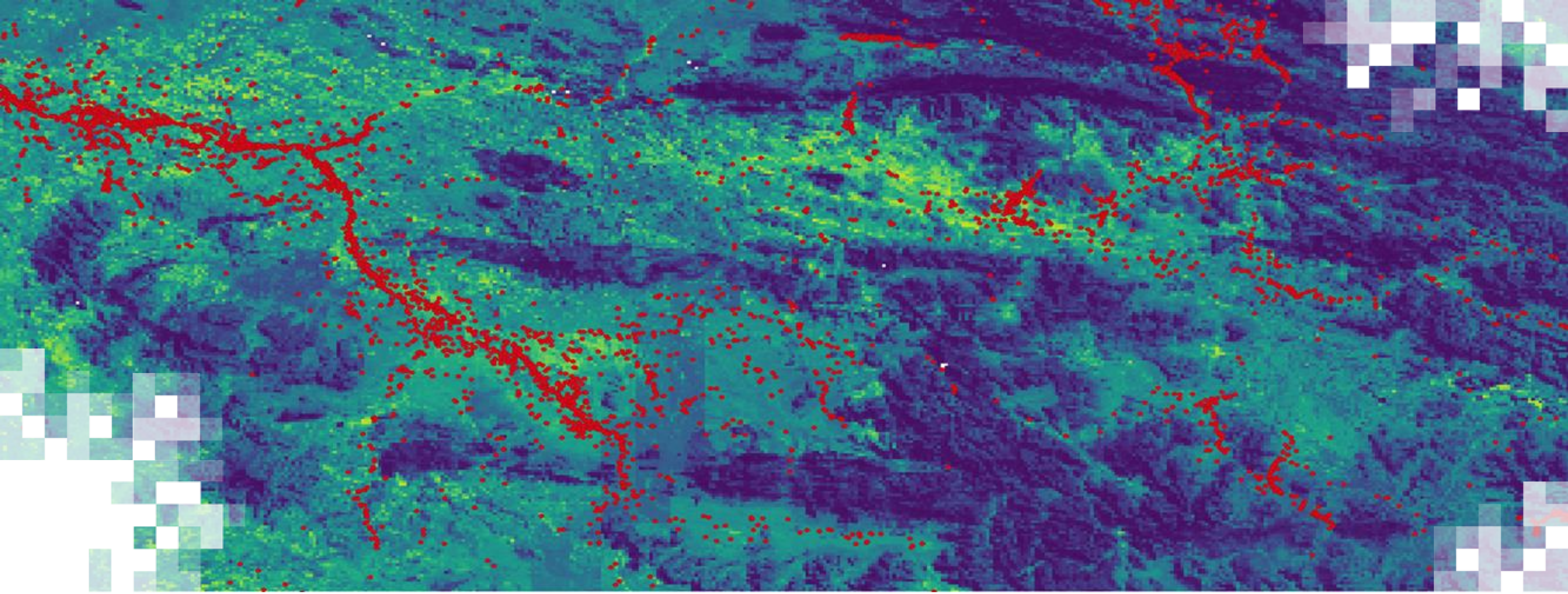
Homework

Opens July 14 – Due July 28 – Posted on Training Webpage

A certificate of completion will be awarded to those who attend all live sessions and complete the homework assignment(s) before the given due date.

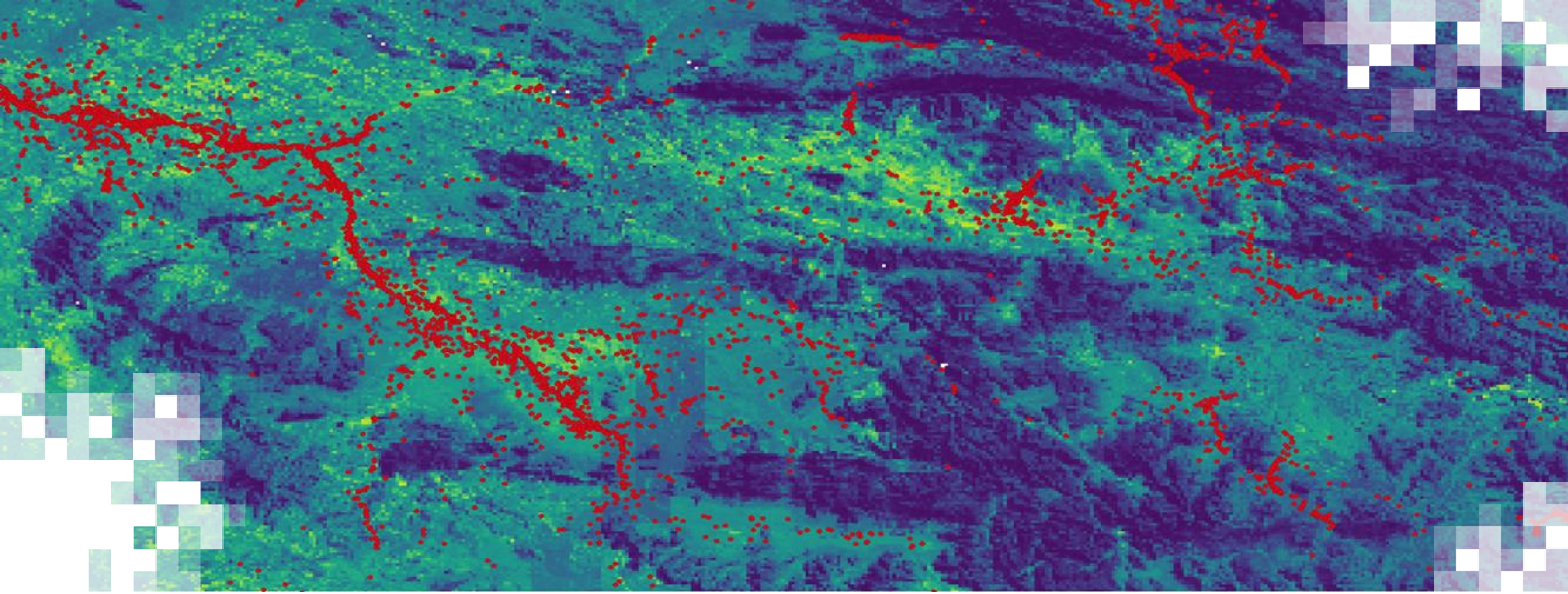
How to Ask Questions

- Please put your questions in the Questions box and we will address them at the end of the webinar.
- Feel free to enter your questions as we go. We will try to get to all of the questions during the Q&A session after the webinar.
- The remainder of the questions will be answered in the Q&A document, which will be posted to the training website about a week after the training.



Species Distribution Modeling with Google Earth Engine

Part 1: Foundations of Species Distribution Modeling with GEE



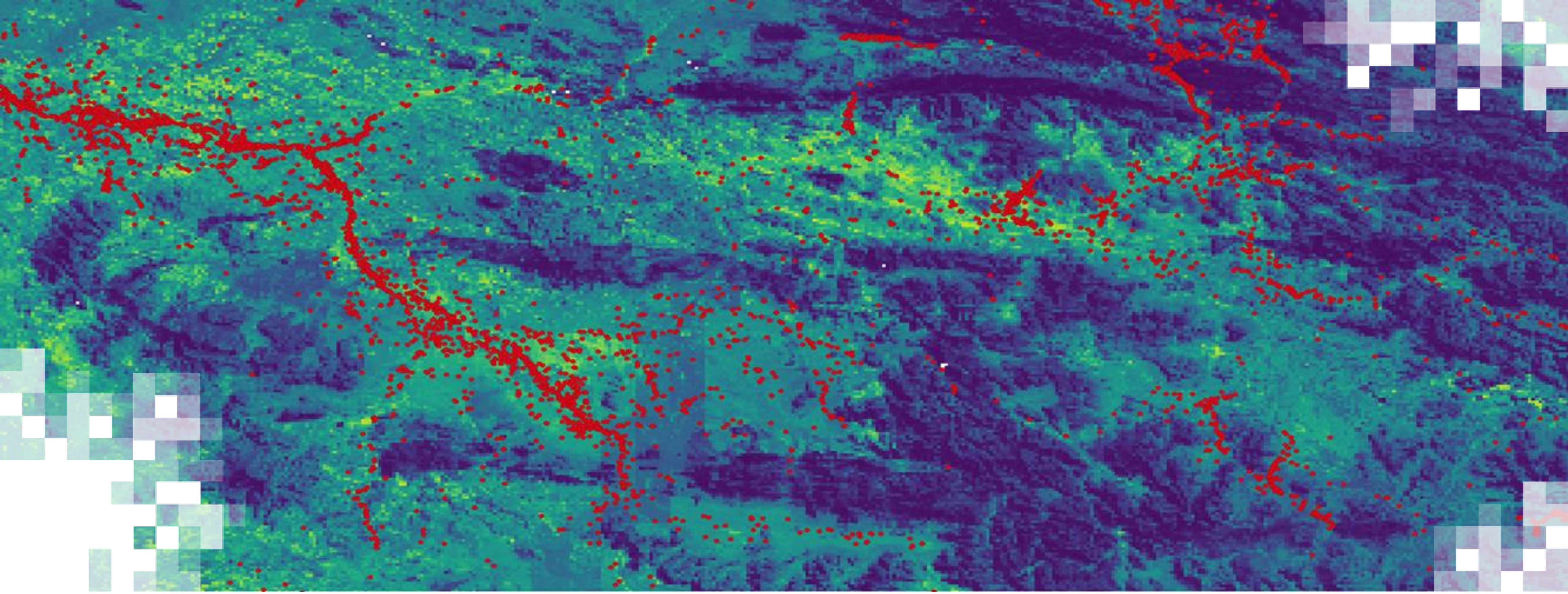
Species Distribution Modeling with Google Earth Engine

Ramiro Crego, Grant Connette, Jared Stabach

July 7, 2026

Learning Objectives

- Conduct basic operations and functions in Google Earth Engine (GEE) web interface using JavaScript code
- Identify core concepts and applications of Species Distribution Modeling (SDM)
- Assess key SDM workflow decisions to ensure occurrence point independence, select background points, and choose ecologically relevant predictors.
- Access and manage analysis-ready spatial data from the GEE catalog as predictor variables
- Fit species distribution models using machine learning models (e.g., Random Forest) in GEE using a reproducible code-based workflow
- Critically evaluate the appropriateness of the model design and interpret predicted distribution and habitat suitability results



Introduction to Species Distribution Models and Habitat Suitability


Species are adapted to particular environments

What Drives a Species Distribution?

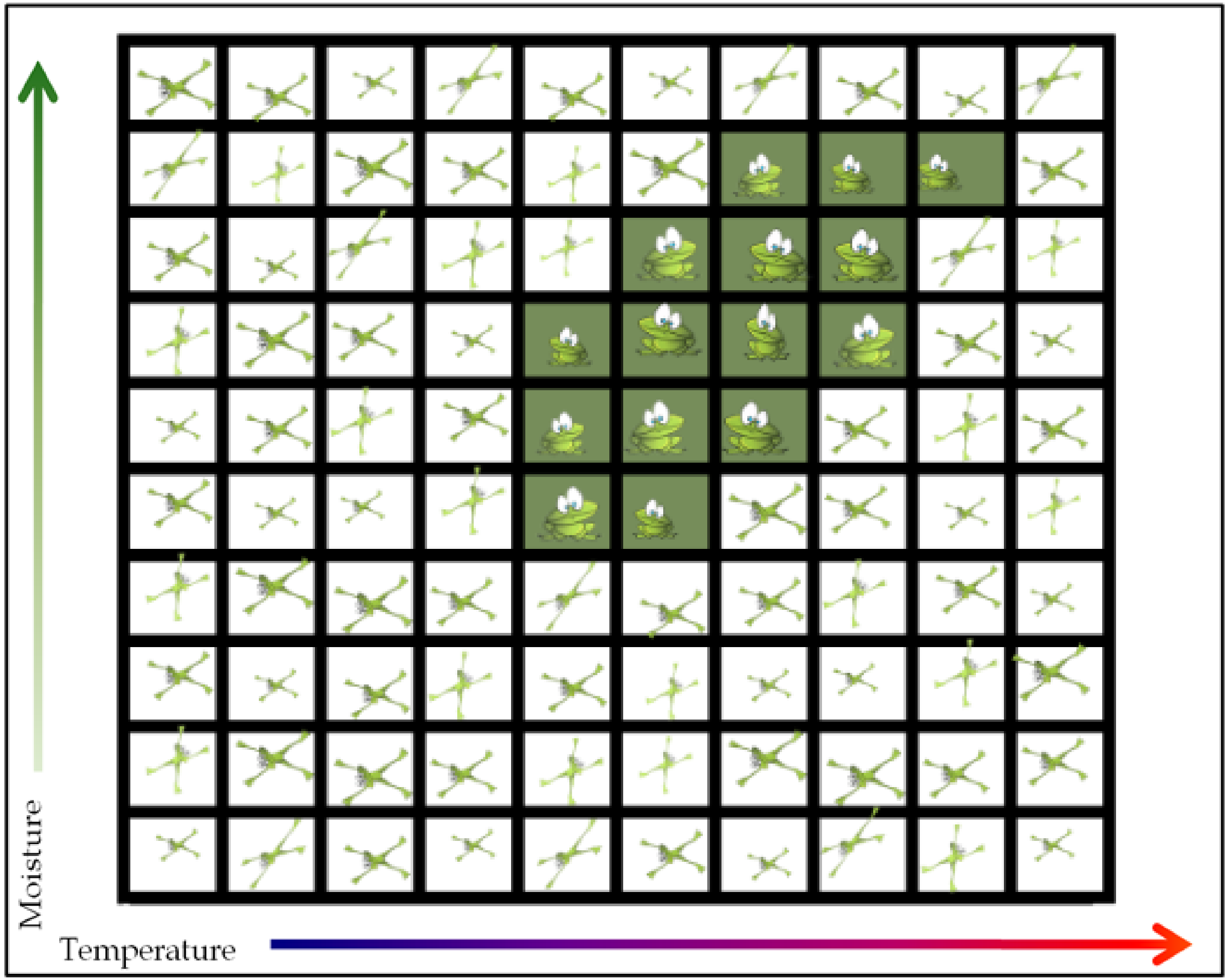
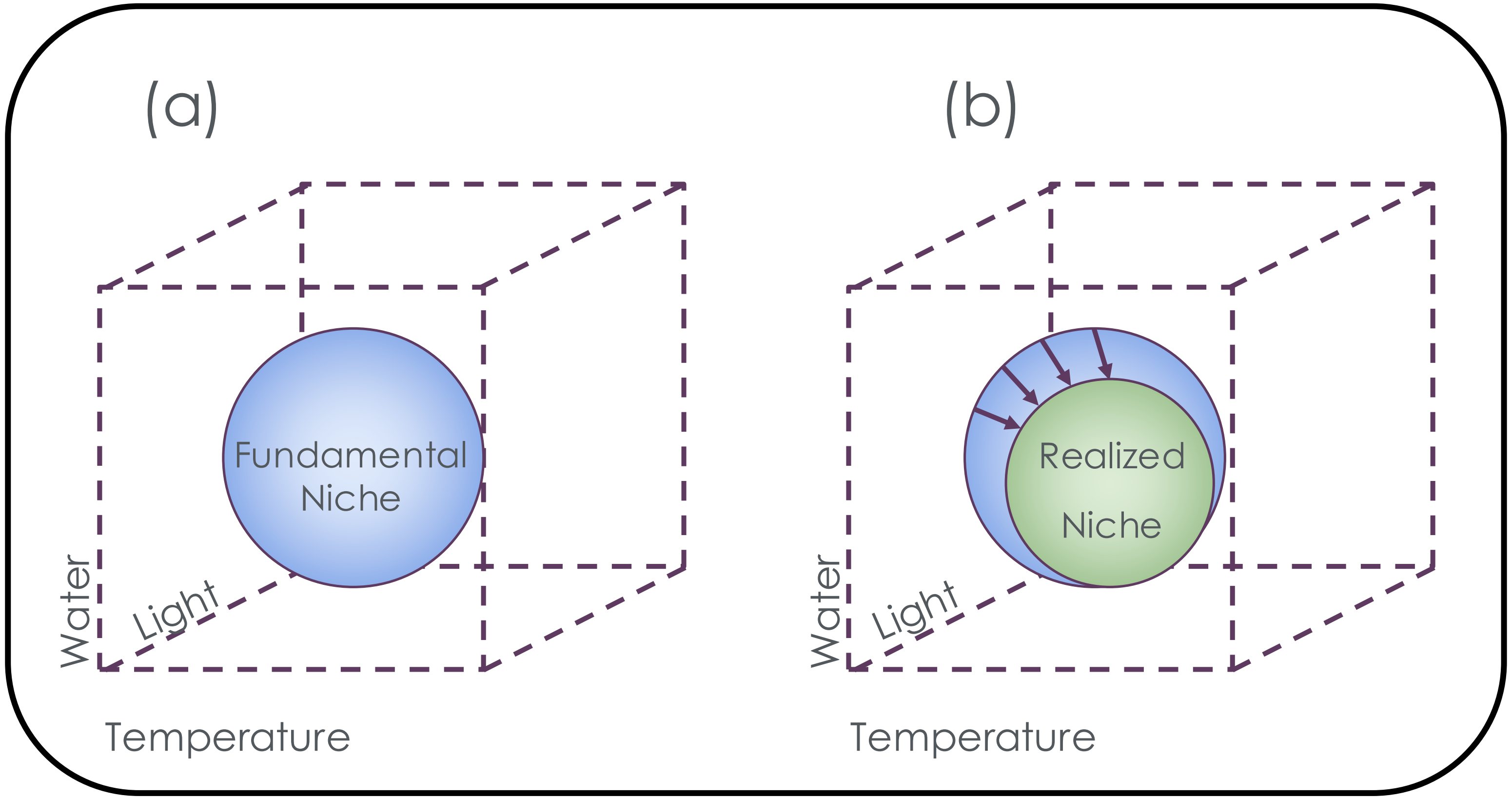


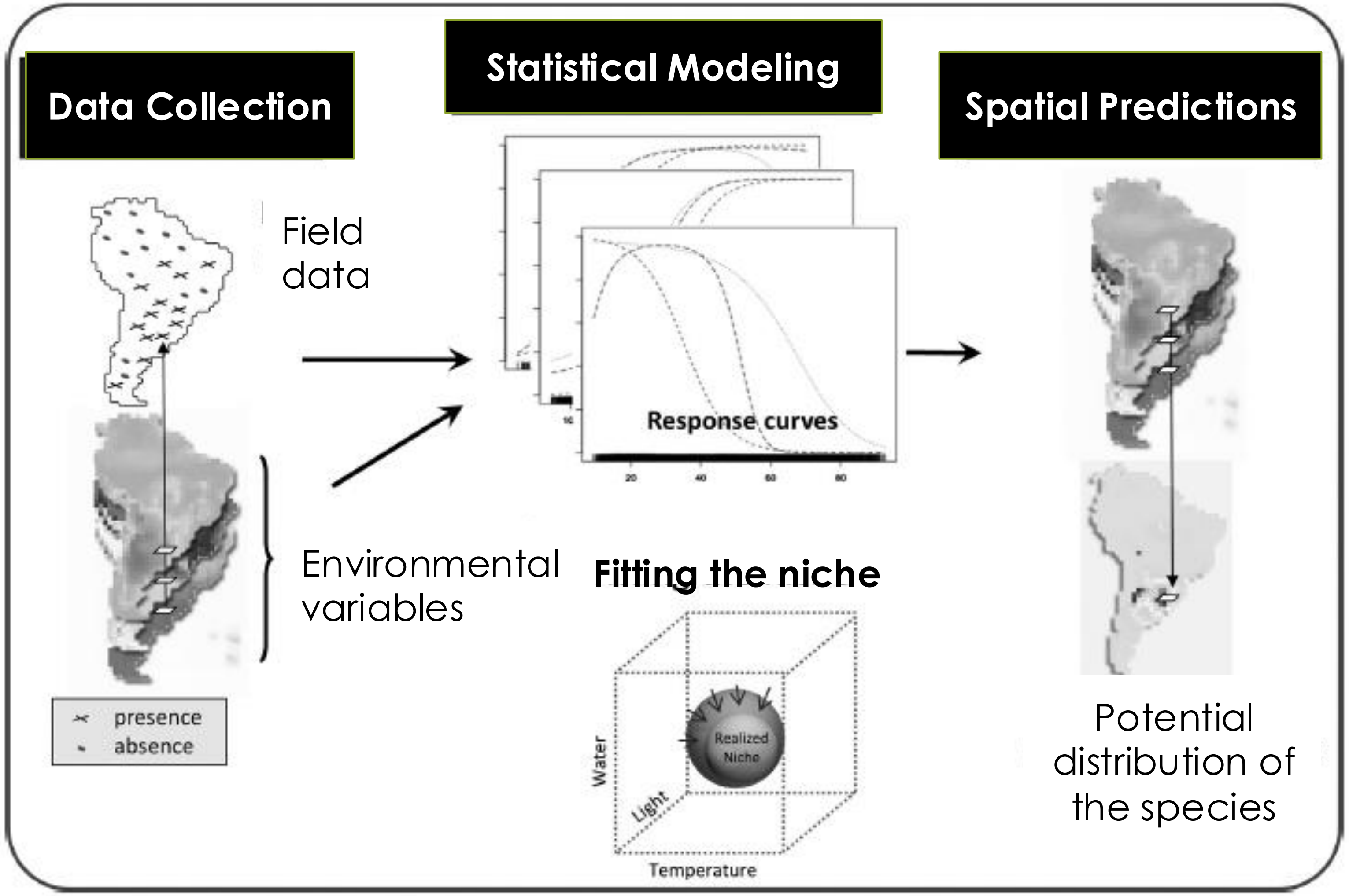
Image Source

Fundamental vs. Realized Niche



What is a Species Distribution Model?

- An empirical model that relates species occurrence with environmental predictor variables to assess habitat suitability
- Typically combine field-collected species occurrence data with raster layers (NDVI, Land Use, etc…) to map suitable habitat



Interpreting Species Distribution Models

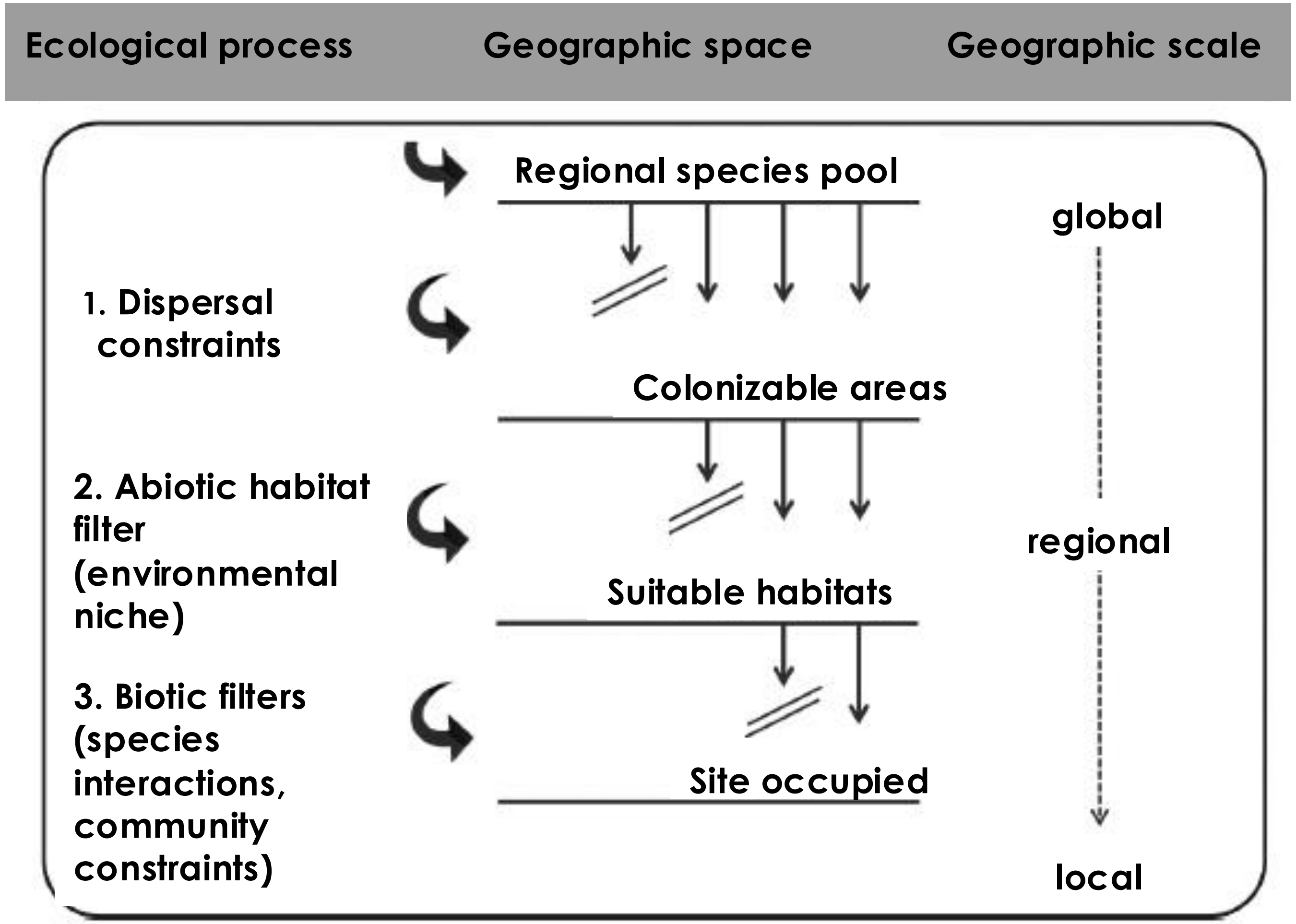
Identified Areas

The Reality:

- Based on environmental conditions of known occupied sites
- Not where a species is actually found

Why the Gap?

- Constrained by factors such as historical barriers, dispersal limitation, and biotic interactions
- Models without true absence data cannot tell us how habitat is actually occupied

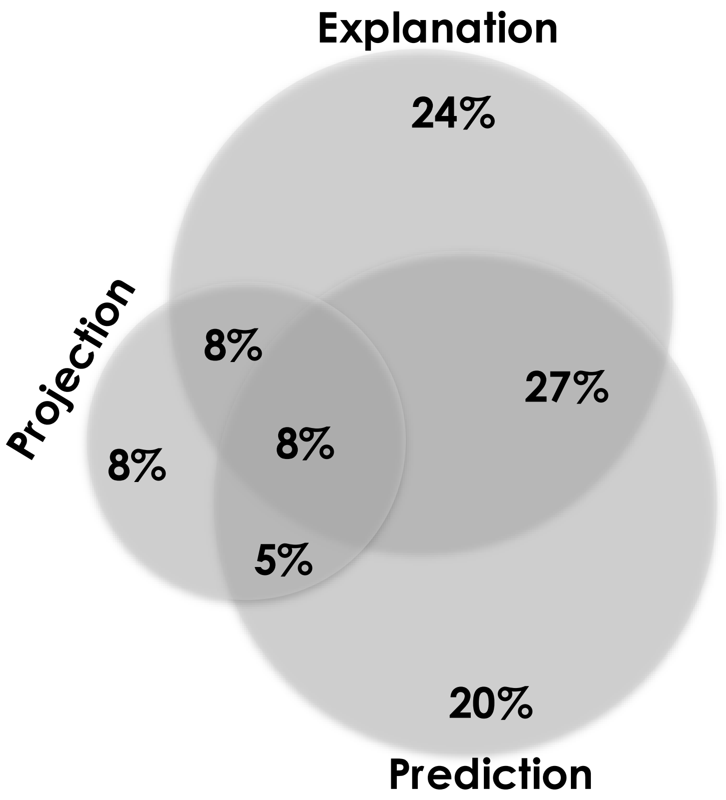


Why Map the Geographic Distribution of a Species?

Basic

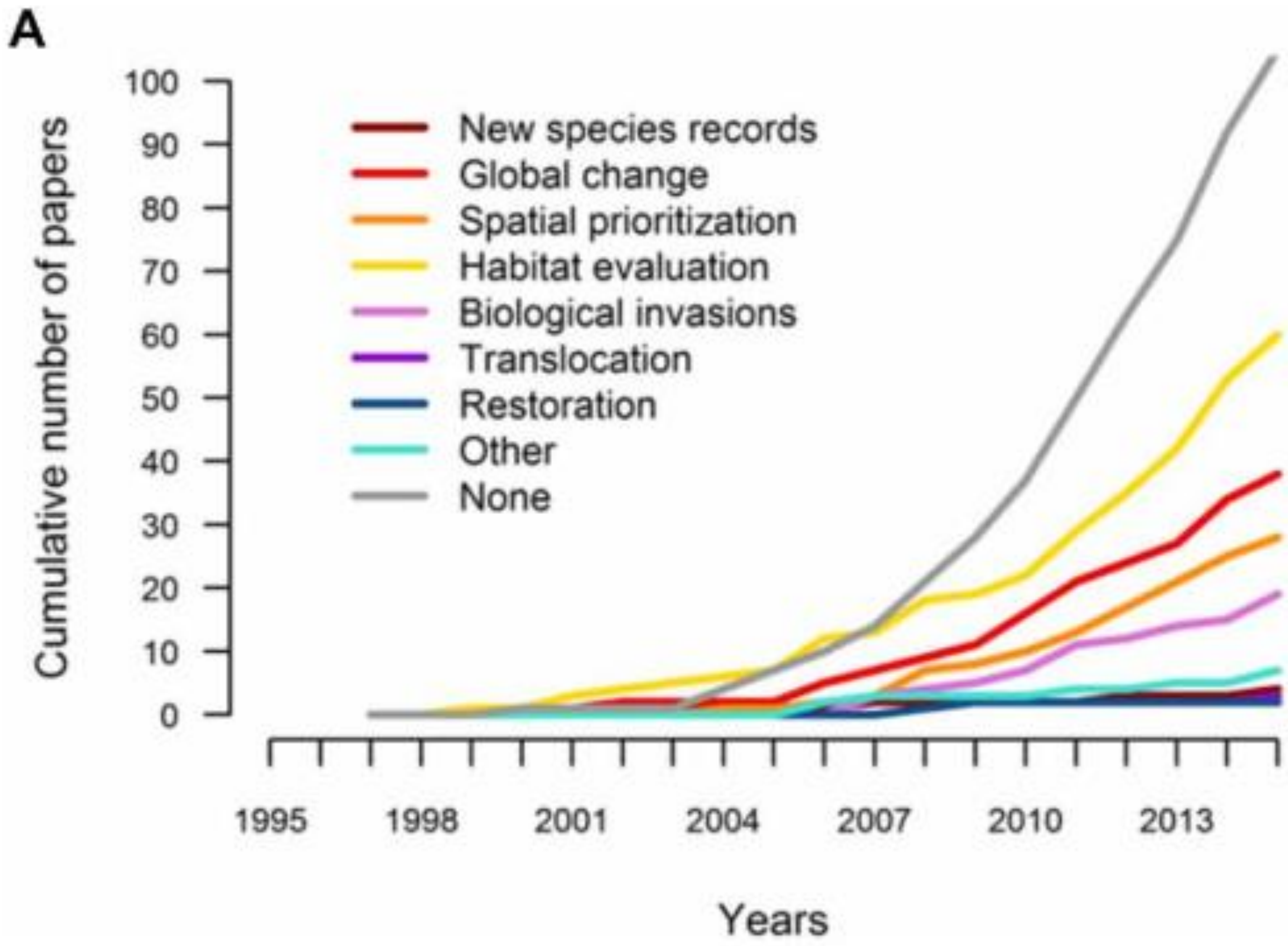
- Compare fundamental/realized niche (where a species **could** live vs. where it **actually** lives)
- Understand species ecology & evolution
- Factors/mechanisms controlling geographic distribution
- Find previously unknown populations
- Assess current threats to populations
- Model impacts of current and future threats
- Develop conservation/management plans

Applied

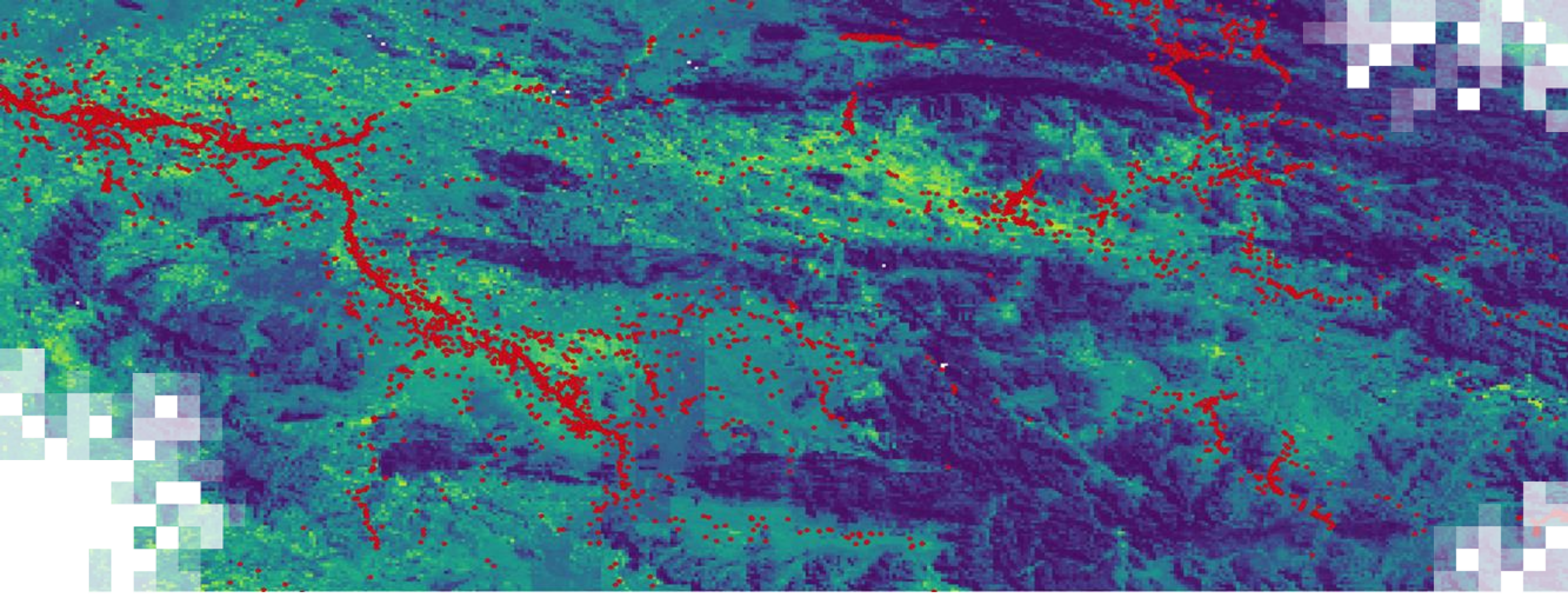


Proliferation of Species Distribution Modeling

Currently the most widely used modeling tool for analyzing occurrence data, and explaining, predicting and projecting habitat suitability.



Standards for distribution models in biodiversity assessments
Miguel B. Araújo *et al.*



Why SDMs in Google Earth Engine?

Why GEE? Technical Barriers

Too much data to download

Spatial data can consist of very large files, GB or TB.

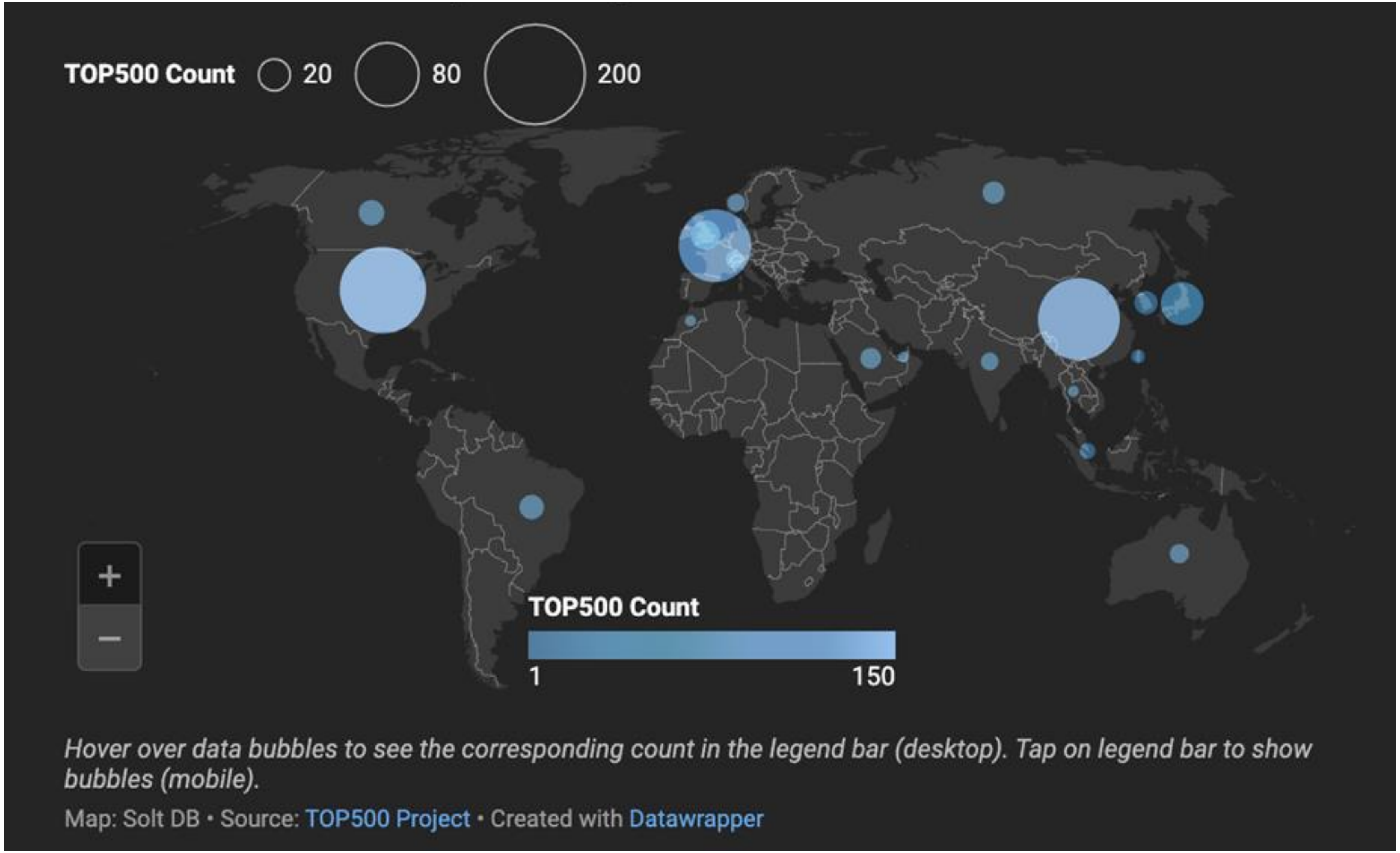
Too Slow

Often need lot of computing power and most researchers do not have access to it.

Coarse resolution

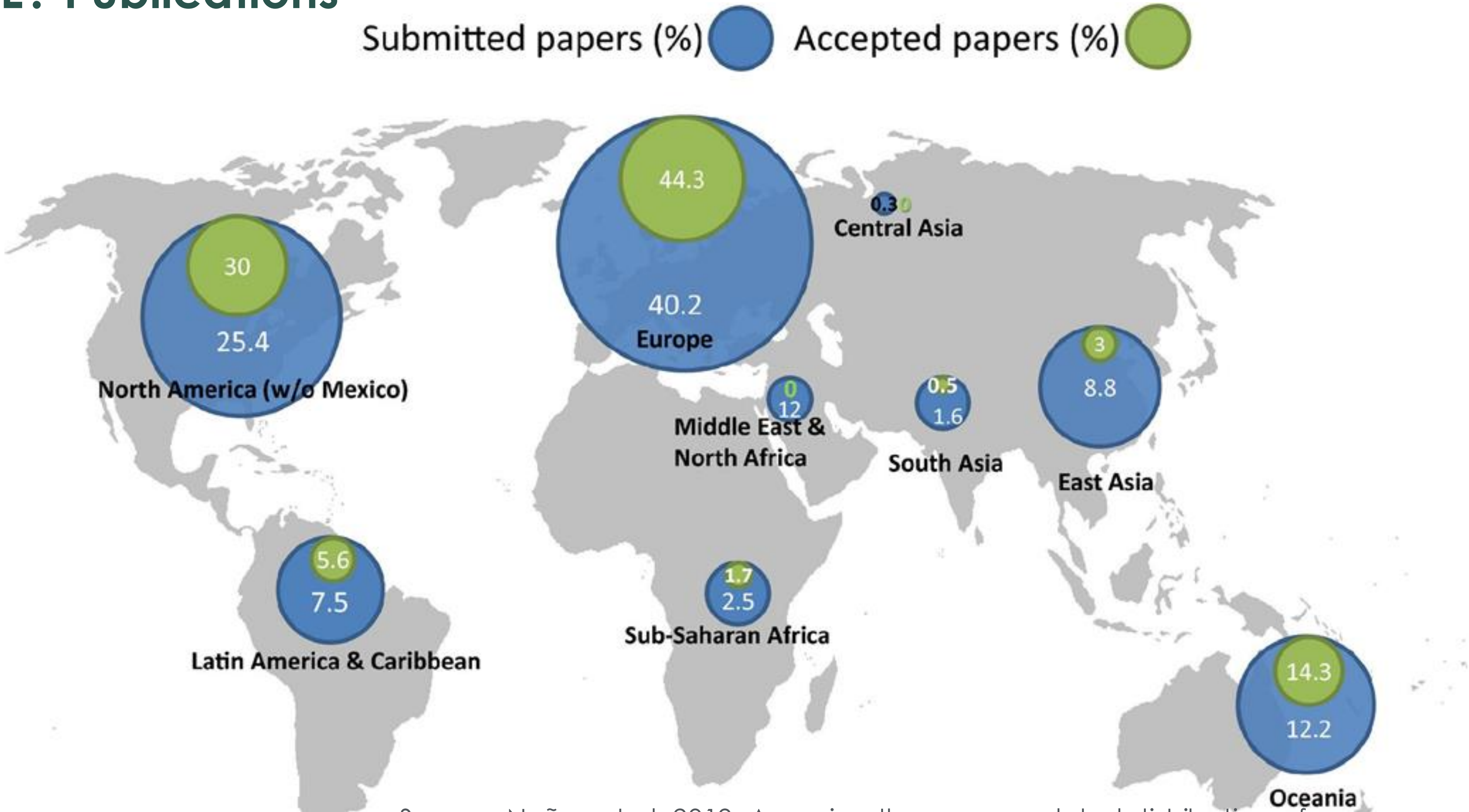
Lack of analytical capacity results in studies with low spatiotemporal resolution.

Why GEE? Global TOP500 Supercomputers Count



Source: [TOP500 Project](#) Created with [Datawrapper](#)

Why GEE? Publications



Source: Nuñez et al. 2019. Assessing the uneven global distribution of readership, submissions and publications in applied ecology: Obvious problems without obvious solutions

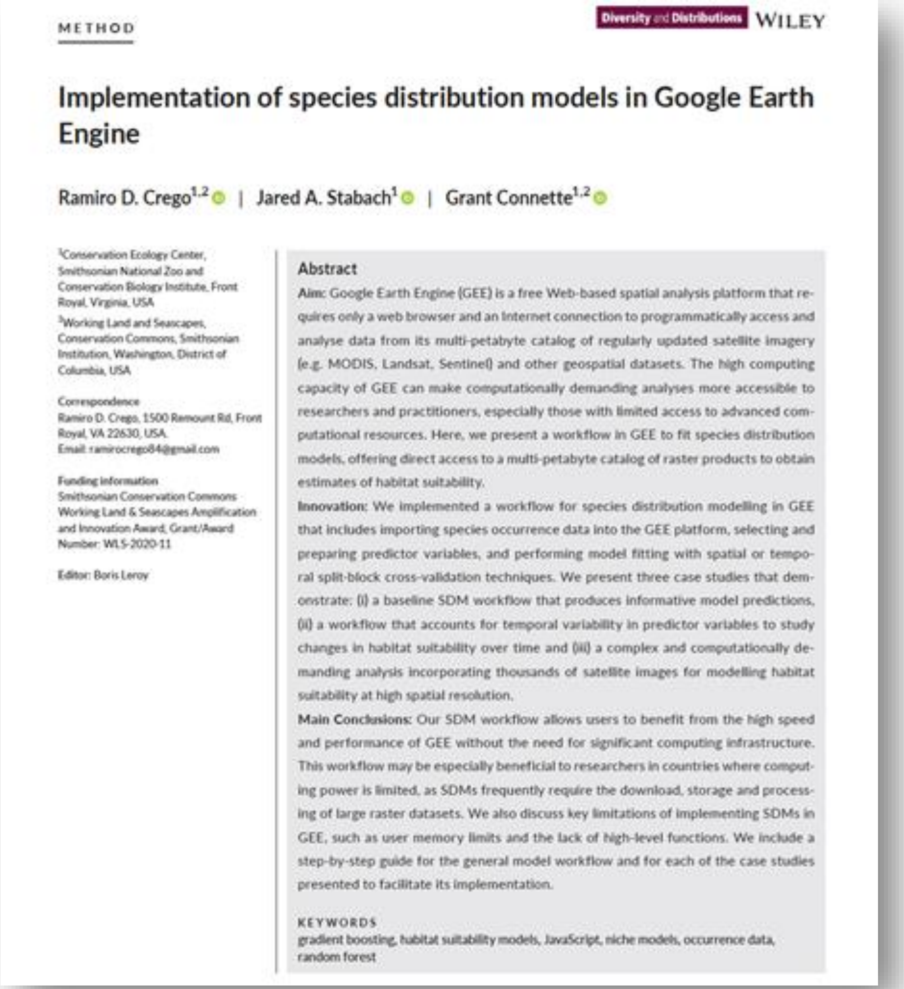
Why GEE? Google Earth Engine Benefits

Google Earth Engine

- It is a cloud-based platform hosted by Google that runs on Google's powerful computing network
- It was primarily designed to process satellite information for research, education and other nonprofit use
- **It is free for noncommercial use**

Why GEE? Analysis Process

1. Species data upload
2. Preprocessing
3. Predictor variables
4. Presence and absence
   a. Presence and absence
   b. Presence and pseudoabsence
   c. Presence and absence and pseudoabsence
5. Model fitting and model evaluation
6. Model predictions



METHOD

Diversity and Distributions WILEY

Implementation of species distribution models in Google Earth Engine

Ramiro D. Crego[1,2] | Jared A. Stabach[1] | Grant Connette[1,2]

[1]Conservation Ecology Center, Smithsonian National Zoo and Conservation Biology Institute, Front Royal, Virginia, USA
[2]Working Land and Seascapes, Conservation Commons, Smithsonian Institution, Washington, District of Columbia, USA

Correspondence
Ramiro D. Crego, 1500 Remount Rd, Front Royal, VA 22630, USA.
Email: ramirocrego84@gmail.com

Funding information
Smithsonian Conservation Commons Working Land & Seascapes Amplification and Innovation Award, Grant/Award Number: WLS-2020-11

Editor: Boris Leroy

Abstract

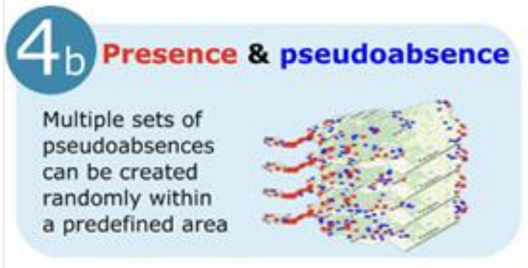
Aim: Google Earth Engine (GEE) is a free Web-based spatial analysis platform that requires only a web browser and an Internet connection to programmatically access and analyse data from its multi-petabyte catalog of regularly updated satellite imagery (e.g. MODIS, Landsat, Sentinel) and other geospatial datasets. The high computing capacity of GEE can make computationally demanding analyses more accessible to researchers and practitioners, especially those with limited access to advanced computational resources. Here, we present a workflow in GEE to fit species distribution models, offering direct access to a multi-petabyte catalog of raster products to obtain estimates of habitat suitability.

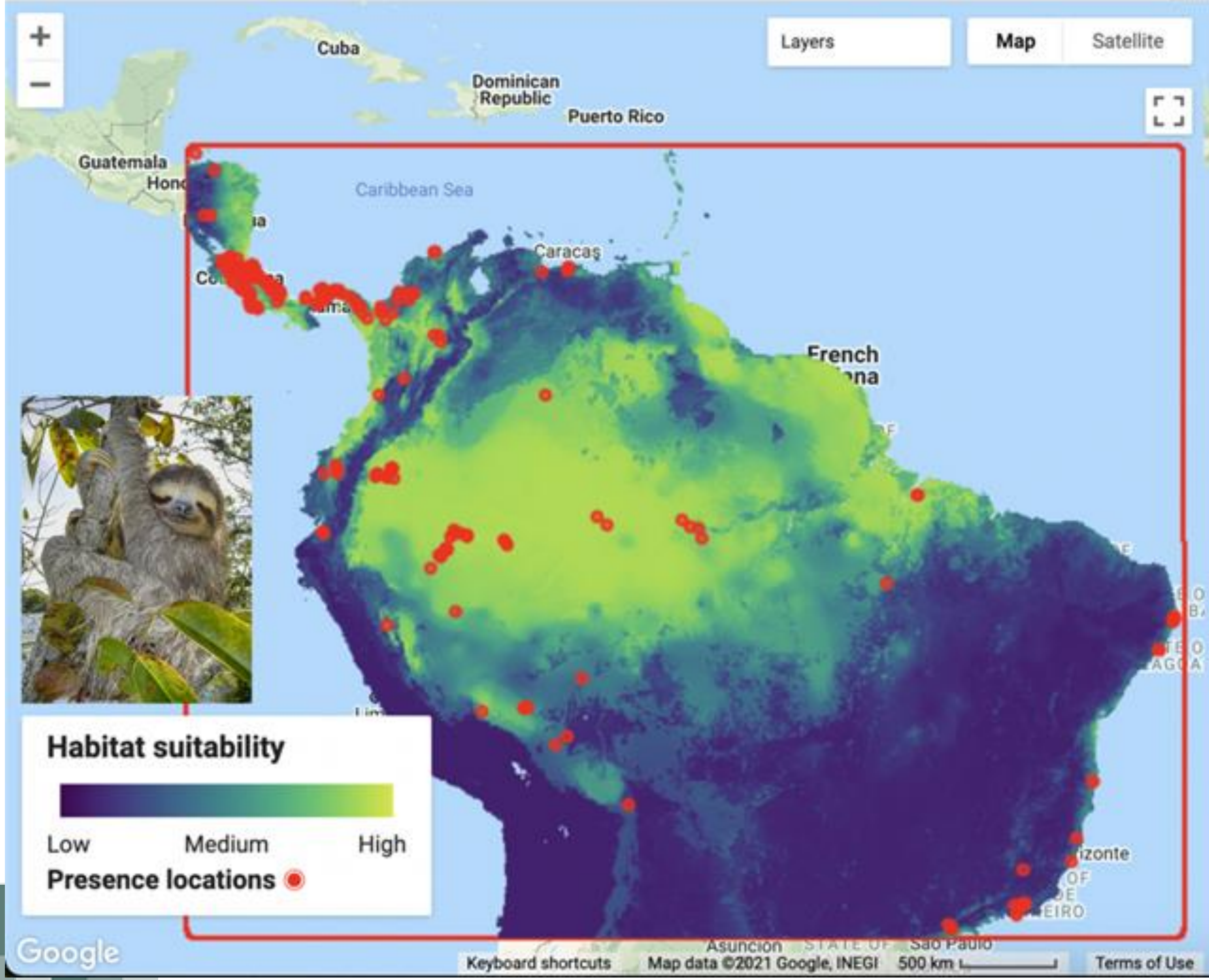
Innovation: We implemented a workflow for species distribution modelling in GEE that includes importing species occurrence data into the GEE platform, selecting and preparing predictor variables, and performing model fitting with spatial or temporal split-block cross-validation techniques. We present three case studies that demonstrate: (i) a baseline SDM workflow that produces informative model predictions, (ii) a workflow that accounts for temporal variability in predictor variables to study changes in habitat suitability over time and (iii) a complex and computationally demanding analysis incorporating thousands of satellite images for modelling habitat suitability at high spatial resolution.

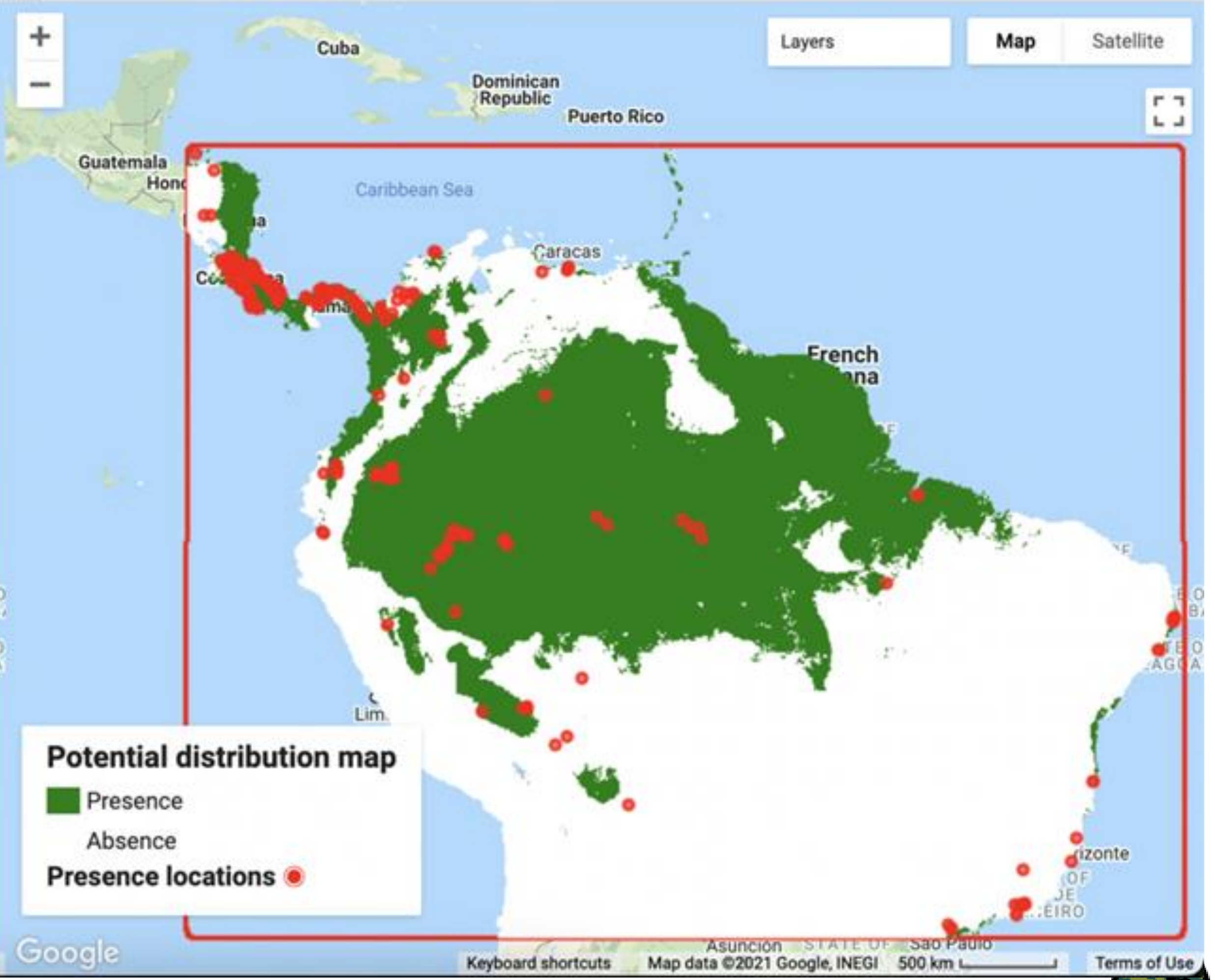
Main Conclusions: Our SDM workflow allows users to benefit from the high speed and performance of GEE without the need for significant computing infrastructure. This workflow may be especially beneficial to researchers in countries where computing power is limited, as SDMs frequently require the download, storage and processing of large raster datasets. We also discuss key limitations of implementing SDMs in GEE, such as user memory limits and the lack of high-level functions. We include a step-by-step guide for the general model workflow and for each of the case studies presented to facilitate its implementation.

KEYWORDS
gradient boosting, habitat suitability models, JavaScript, niche models, occurrence data, random forest

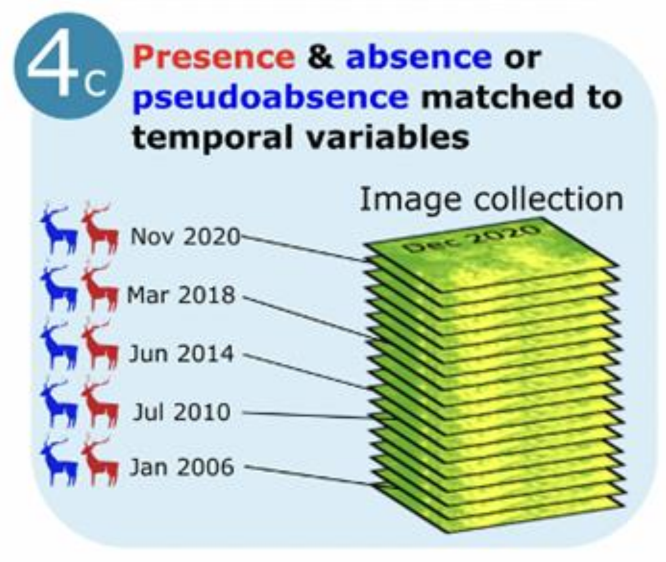
4b: Presence and Pseudo-absence SDMs



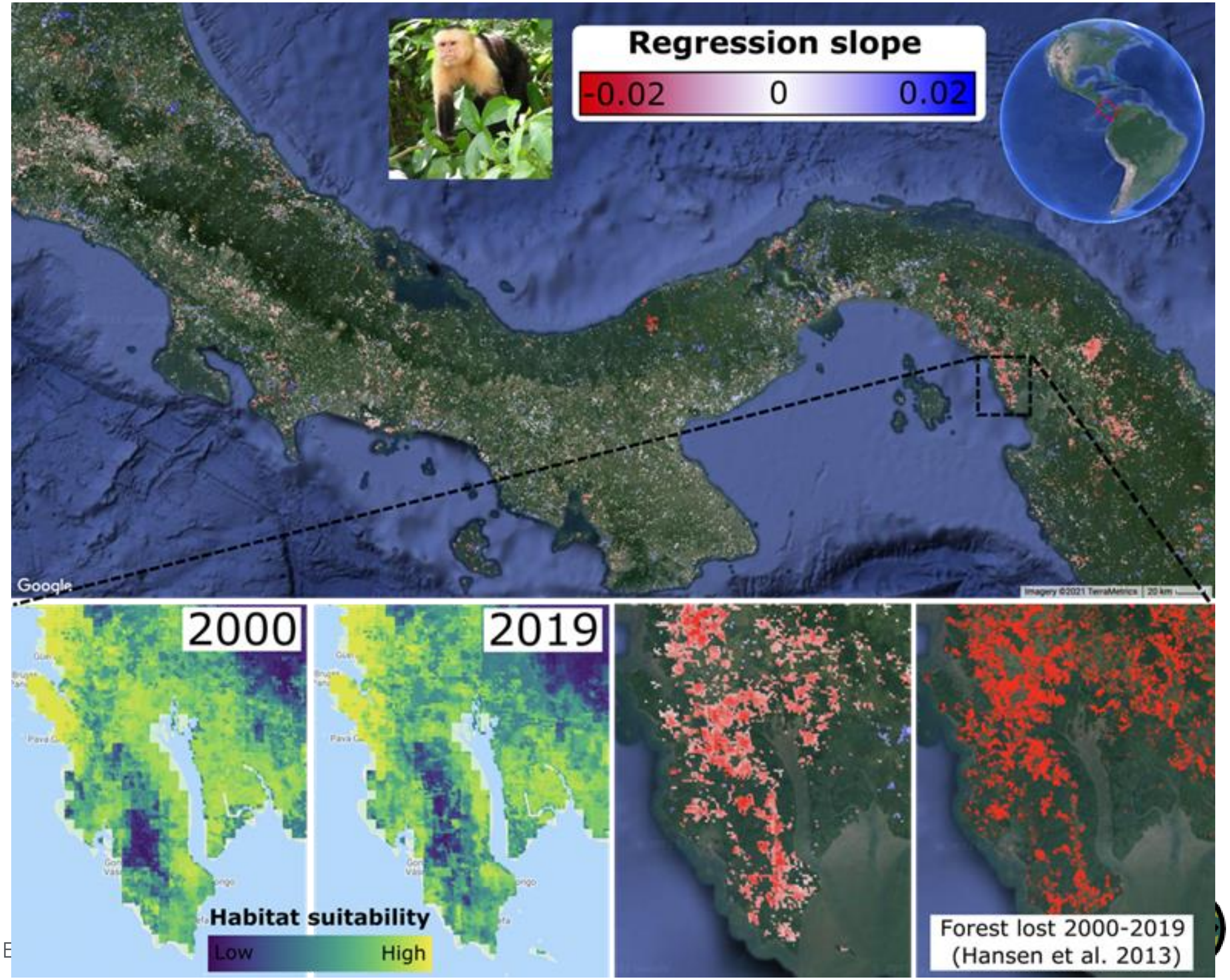




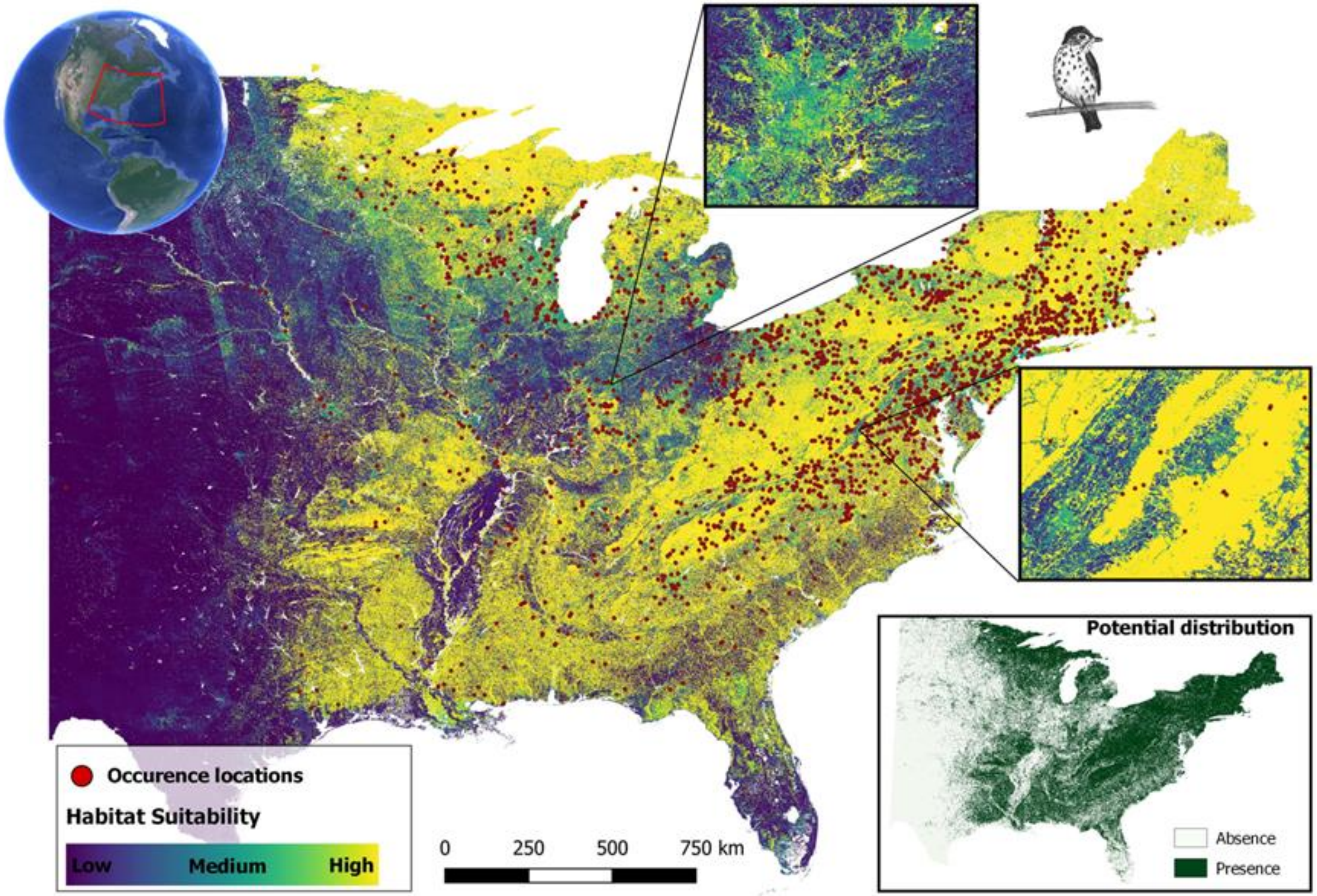
4c: Accounting for Temporal Resolution in SDMs



% tree cover - Terra MODIS Vegetation Continuous Fields 250 m



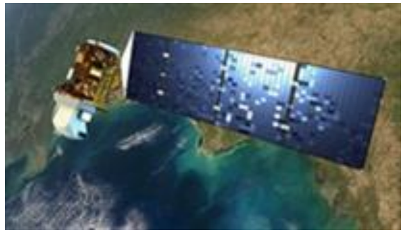
Using Raw Satellite Data in SDMs



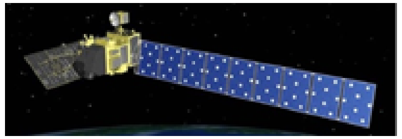
99,939 observations

GBIF

6,150 Landsat 8 images



3 Global radar datasets

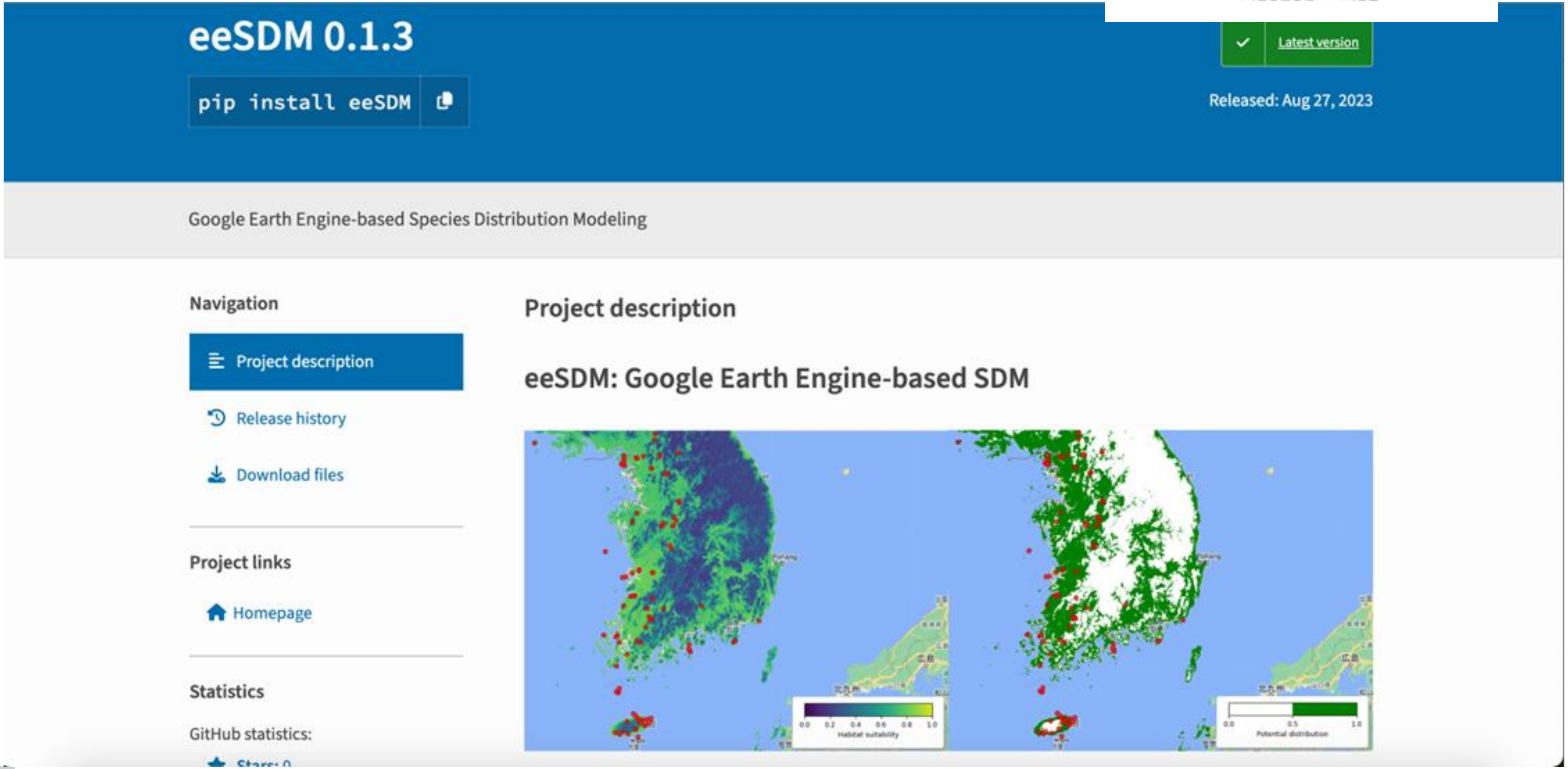


705,147,693 pixels of **90 m** over **4,606,284** km^2

22 hr to run

Things are Happening: eeSDM

Mr. Yu



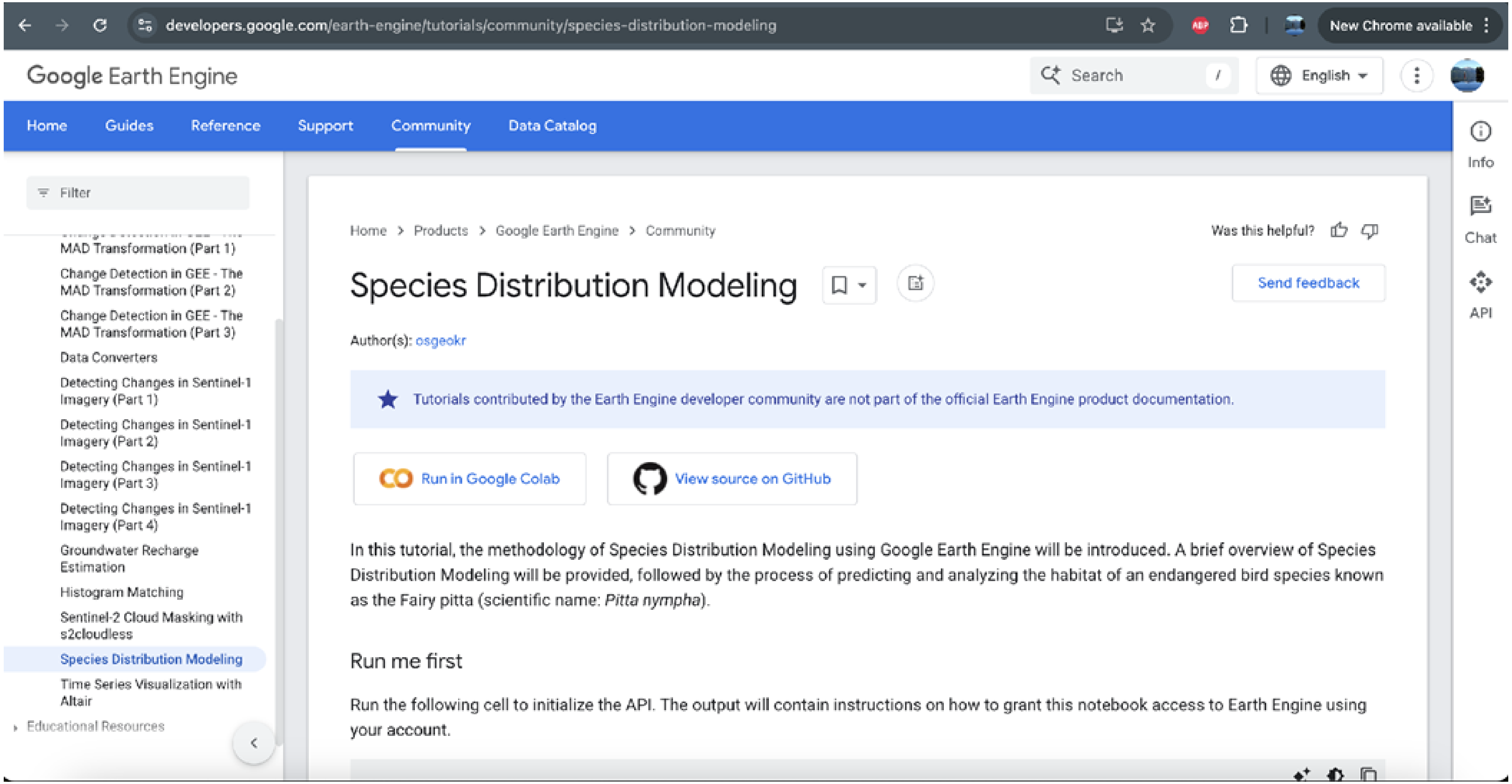
Things are Happening: GEE Tutorials

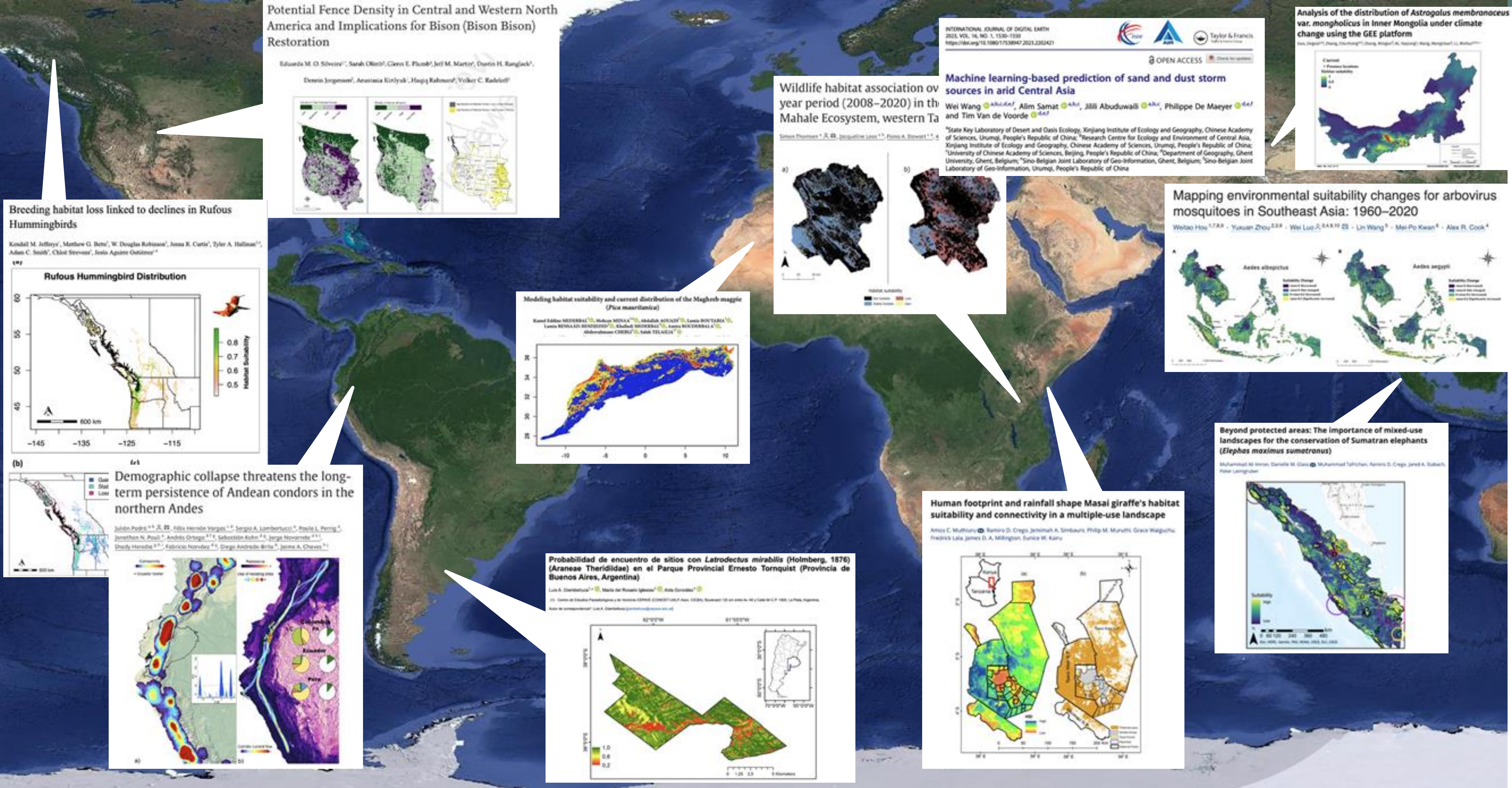
Mr. Yu

eeSDM: Google Earth Engine 기반 종 분포 모델링을 위한 Python 패키지

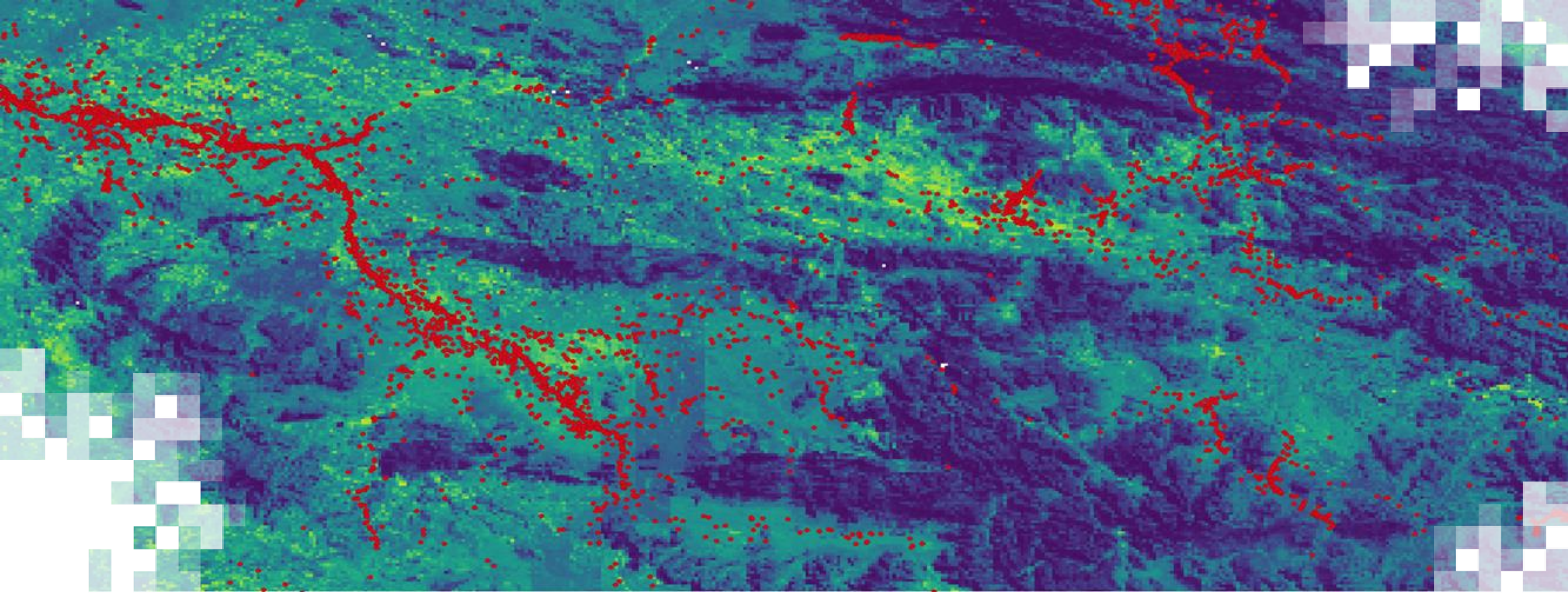
13:00~13:15(15분)

유병혁 | 공간데이터분석가
국립공원공단 ESG혁신실

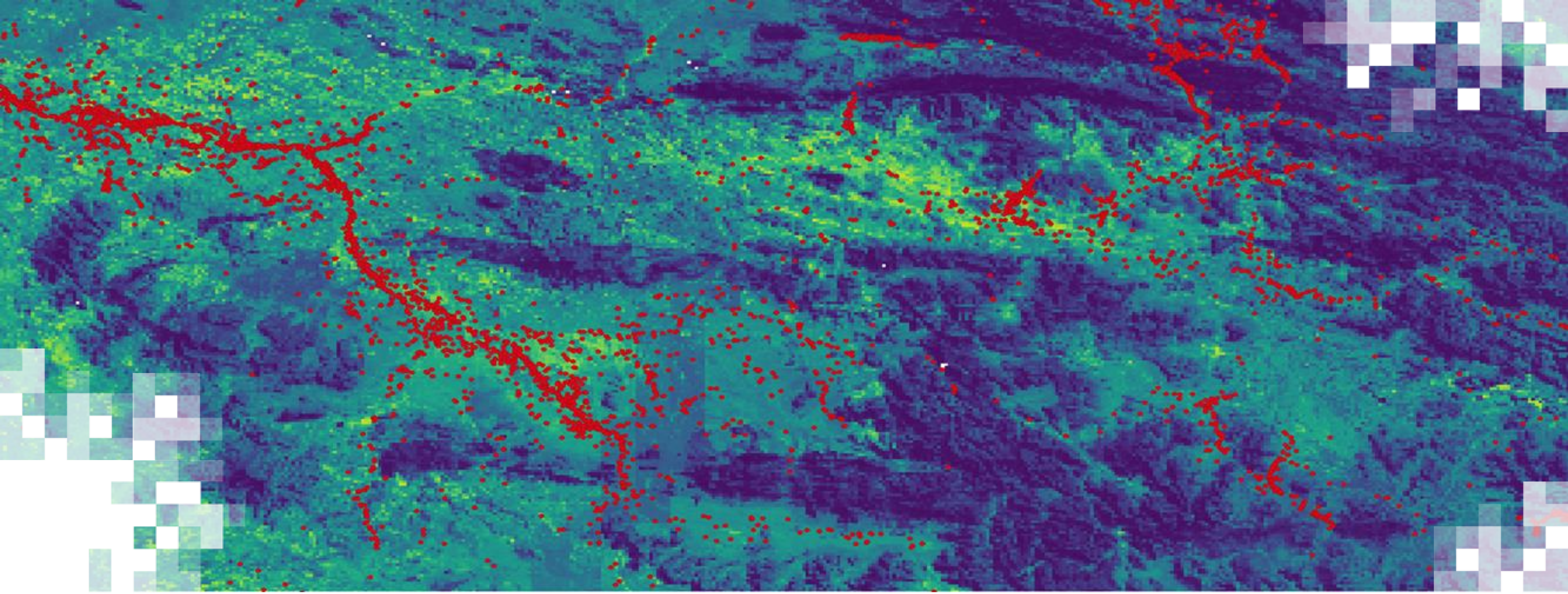




Potential Fence Density in Central and Western North America and Implications for Bison (Bison Bison) Restoration
Breeding habitat loss linked to declines in Rufous Hummingbirds
Rufous Hummingbird Distribution
Habitat Suitability
600 km
Demographic collapse threatens the long-term persistence of Andean condors in the northern Andes
Modeling habitat suitability and current distribution of the Maghreb magpie (Pica mauritanica)
Wildlife habitat association ov
year period (2008–2020) in th
Mahale Ecosystem, western Ta
INTERNATIONAL JOURNAL OF DIGITAL EARTH
Taylor & Francis
OPEN ACCESS
Machine learning-based prediction of sand and dust storm sources in arid Central Asia
Wei Wang, Alim Samat, Jilili Abuduwaili, Philippe De Maeyer and Tim Van de Voorde
Analysis of the distribution of Astragalus membranaceus var. mongholicus in Inner Mongolia under climate change using the GEE platform
Mapping environmental suitability changes for arbovirus mosquitoes in Southeast Asia: 1960–2020
Aedes albopictus
Aedes aegypti
Beyond protected areas: The importance of mixed-use landscapes for the conservation of Sumatran elephants (Elephas maximus sumatranus)
Human footprint and rainfall shape Masai giraffe's habitat suitability and connectivity in a multiple-use landscape
Probabilidad de encuentro de sitios con Latrodectus mirabilis (Holmberg, 1876) (Araneae Theridiidae) en el Parque Provincial Ernesto Tornquist (Provincia de Buenos Aires, Argentina)

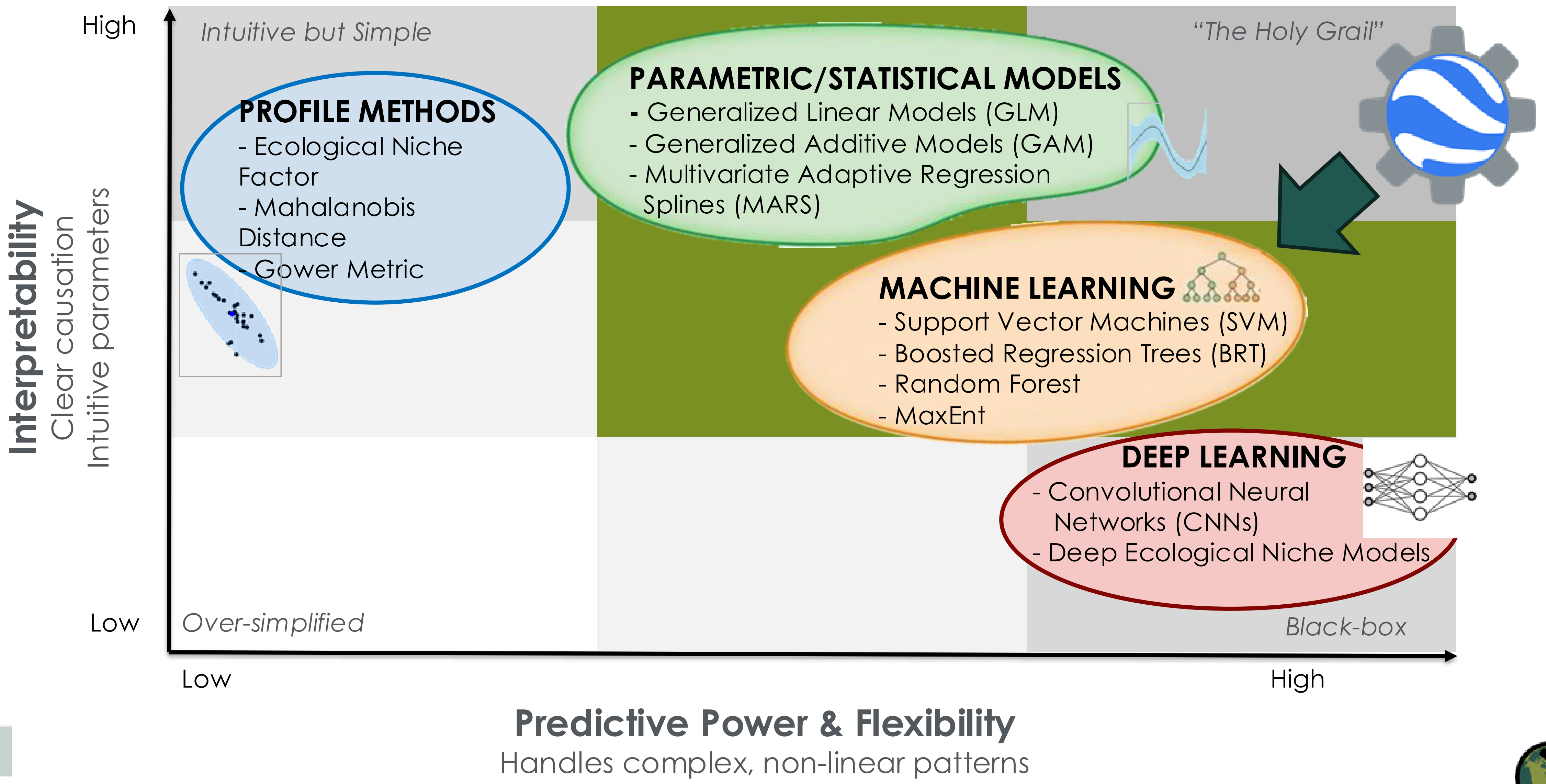


A Quick Demo



Understanding the Methods

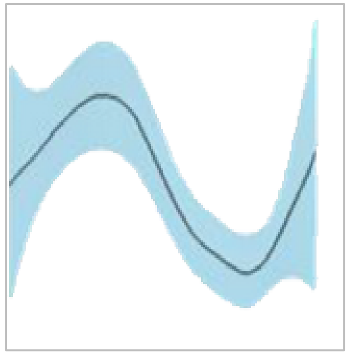
Overview of Some SDM Methods



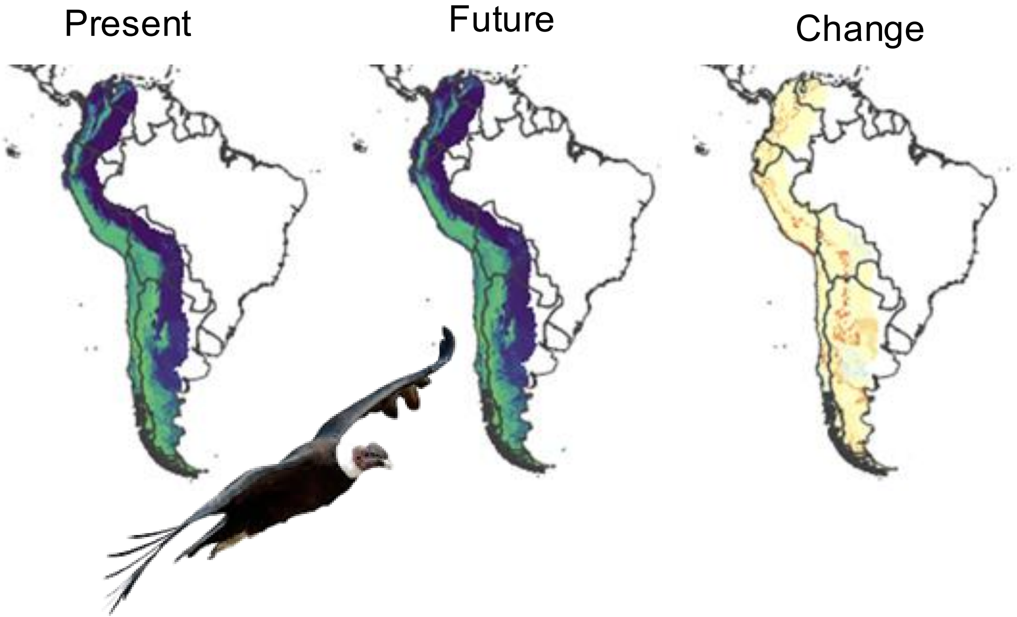
Inference vs. Prediction

SDMs have two prominent applications:

1. Understanding the drivers of species distributions



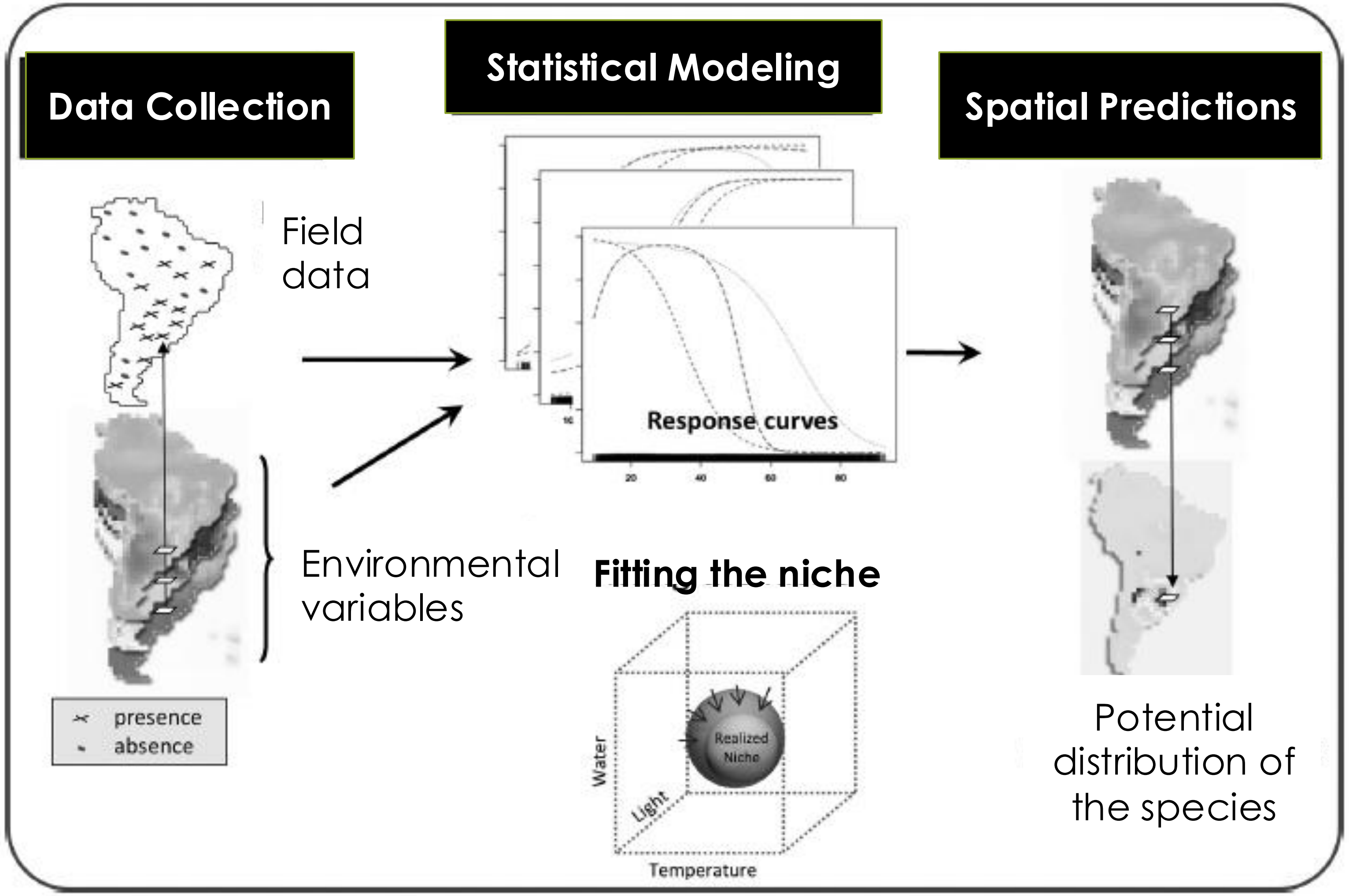
1. Predicting habitat suitability across time and space



The modelling workflow we present here employs non-parametric algorithms available in GEE for **predicting suitable habitats for single species.**

Model Inputs

- Occurrence data
- Absence data
- Environmental predictor variables



Absence Data

- Determining whether a species is absence from a site is not trivial
- An absence in the dataset could be do to:
  - The species is absent due to an environment that is unsuitable
  - The species is absent in suitable habitats due to other ecological factors
  - The species is present but was not detected (areas never sampled, observer error, or species not available to be detected during sampling)

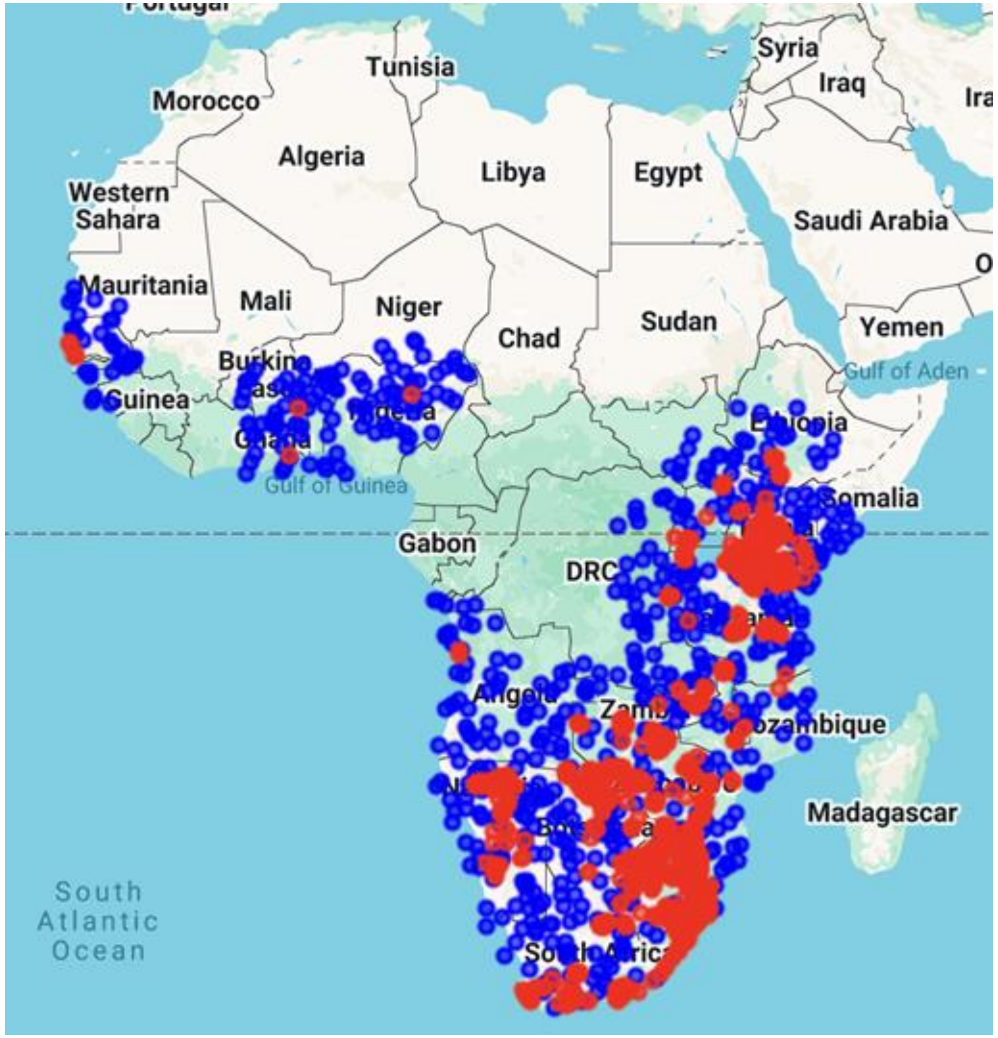
True Absences

- Definitive proof of species absence (with sufficient sampling effort)
  - Some models (e.g., occupancy models) can even account for false negatives
- The Gold Standard: **if you have it, use it!**
- Cost-intensive and difficult to acquire at large scales

Image Source

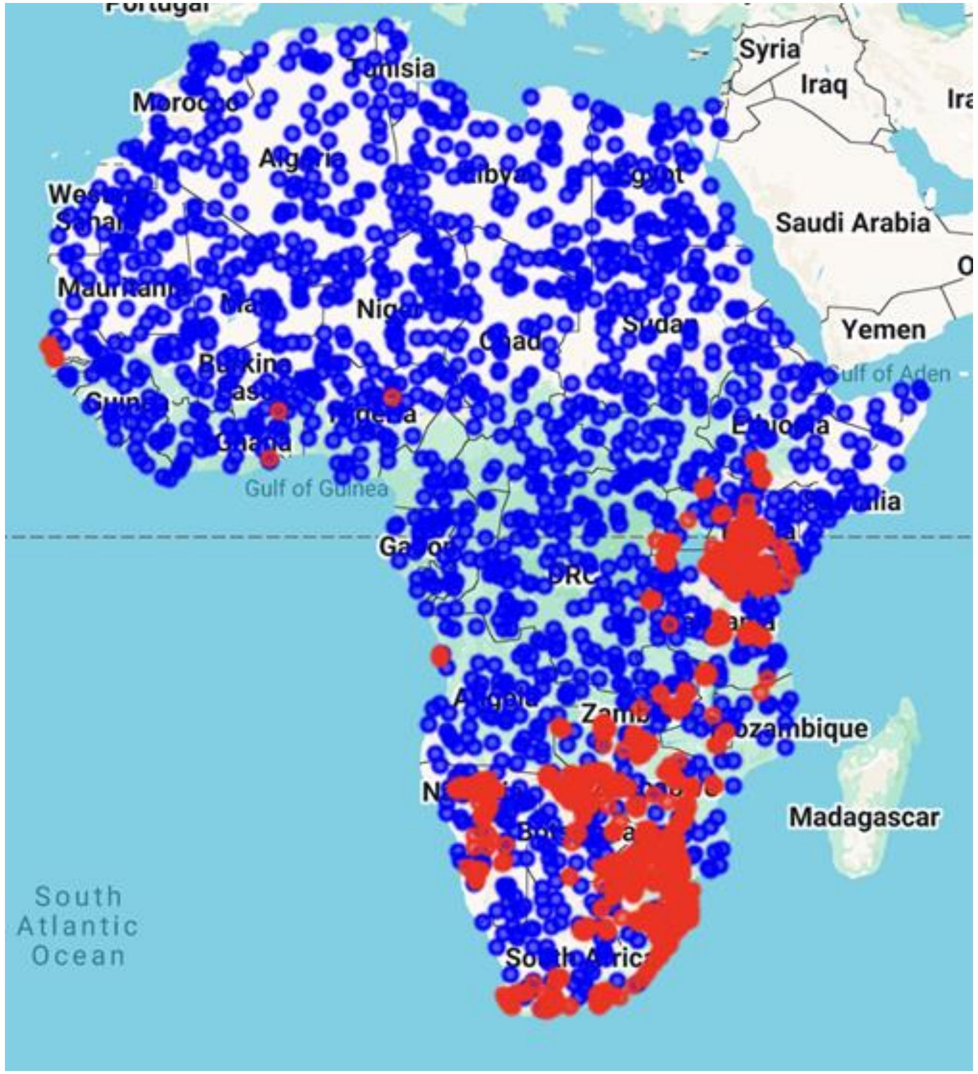
Pseudo-absence

- Synthetically generated points treated as “0s” in the model when true absence data are unavailable
- Essential for presence-absence algorithms like GLMs or Random Forests when species data is presence-only
- Selection strategy: Points are often picked far from presence locations or in environmentally unsuitable areas
  - Minimally, these should not fall on the same raster cells as presence points
  - Common to require a minimum distance from a presence point (e.g., > 10km)
  - Can be constrained to maximum distances or sites with strong environmental dissimilarity



Background Points

- A random sample of the study region's available environment
- Unlike pseudo-absences, they are not assumed to be absences. They provide the "average" conditions to contrast against the "used" presence locations
- Used in Maxent



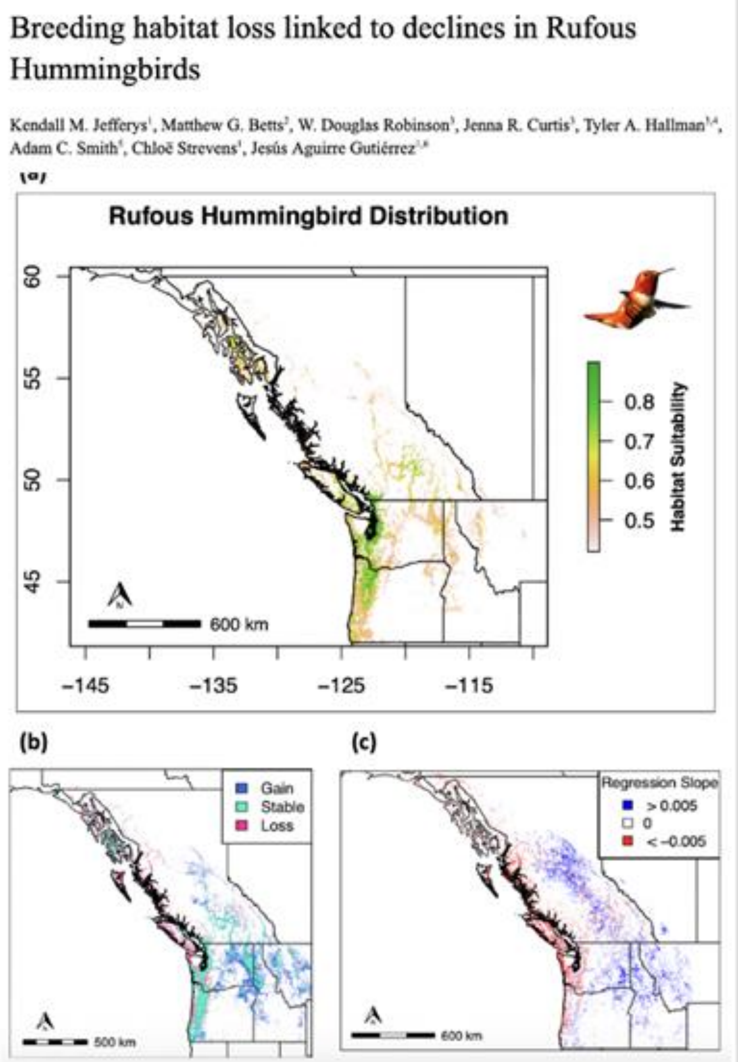
Best Practices

Spatial Matching

Sample pseudo-absence points from areas where related species were surveyed to account for sampling bias. This isolates drivers of underlying distribution from drivers of sampling intensity

Possible to do with eBird data, for example

eBird

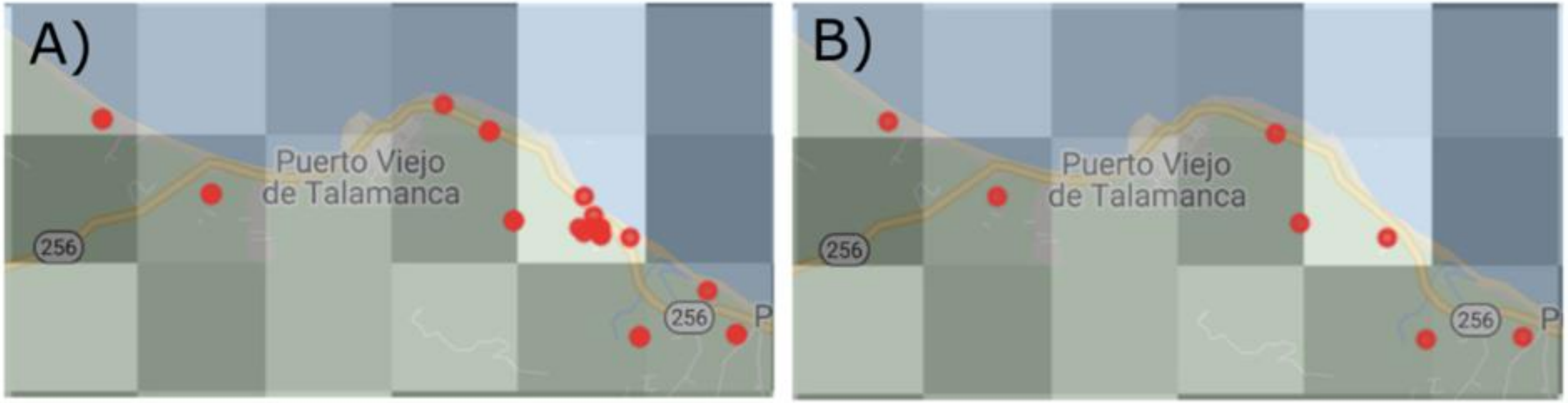

Breeding habitat loss linked to declines in Rufous Hummingbirds

Kendall M. Jefferys[1], Matthew G. Betts[2], W. Douglas Robinson[3], Jenna R. Curtis[3], Tyler A. Hallman[3,4], Adam C. Smith[5], Chloë Strevens[1], Jesús Aguirre Gutiérrez[1,6]

Best Practices

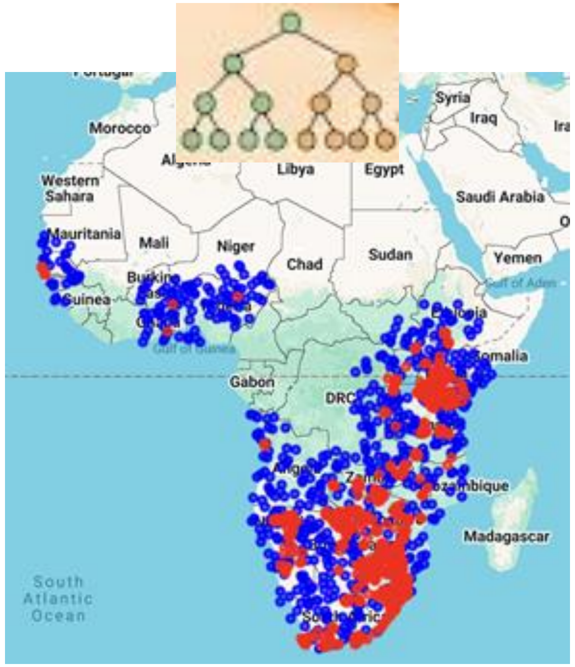
Strategic Spatial Thinning

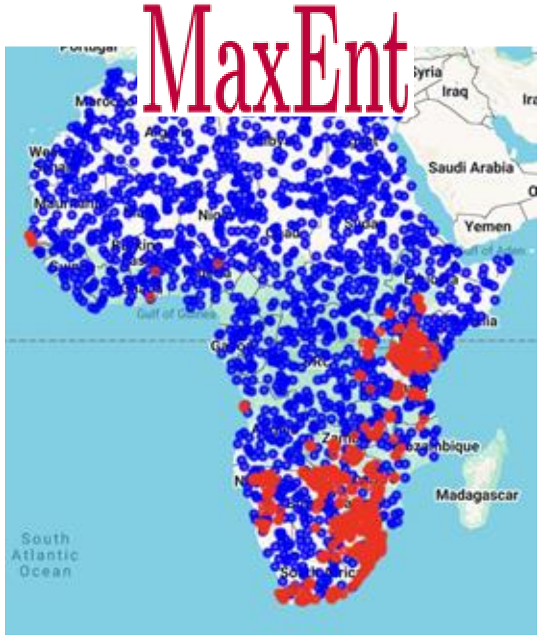
Thin presence points to eliminate spatial clustering due to sampling bias in opportunistic data (e.g., GBIF) *only* when you are not using spatial matching of pseudo-absence points. Be aware that thinning can reduce sampling bias but results in information loss by removing valid environmental data.



Best Practices

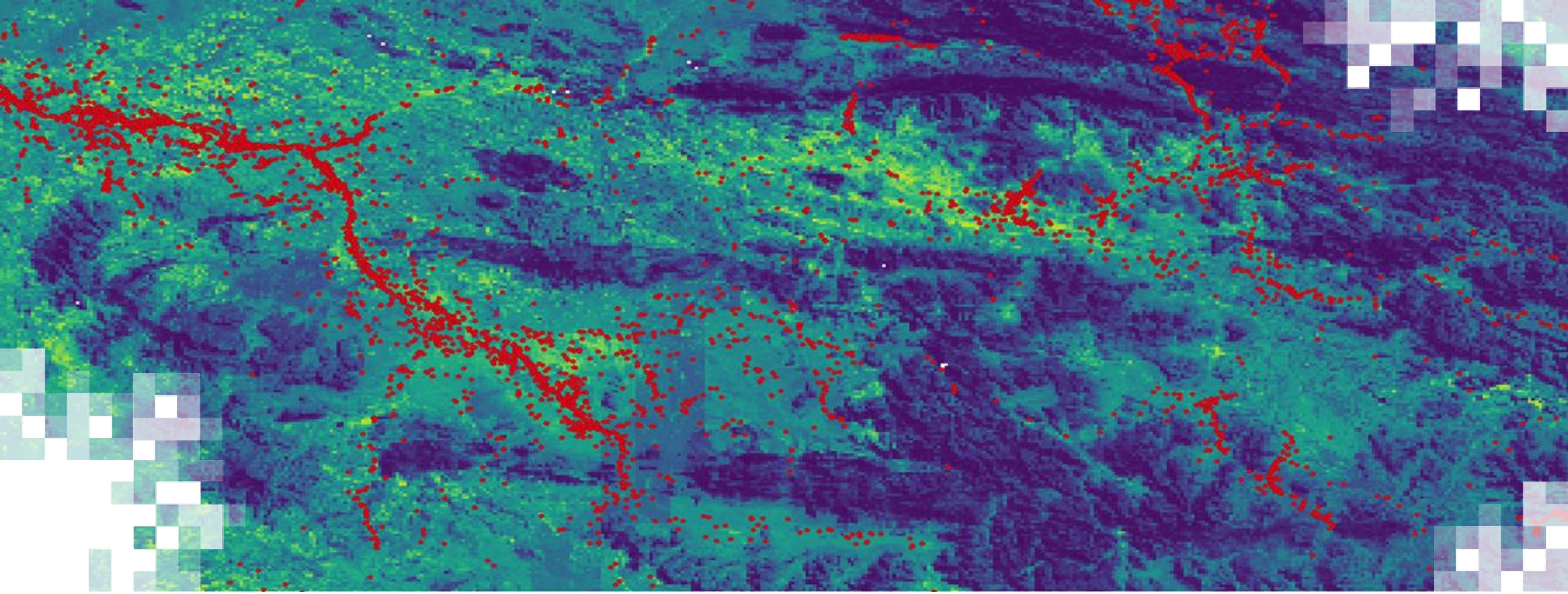
Model-Data Alignment: Match your absence data type to your algorithm and the ecology of your species of interest. Use background points for profile models or MaxEnt and true absences or pseudo-absences for binary classifiers (e.g., GLMs, Random Forests).





Best Practices

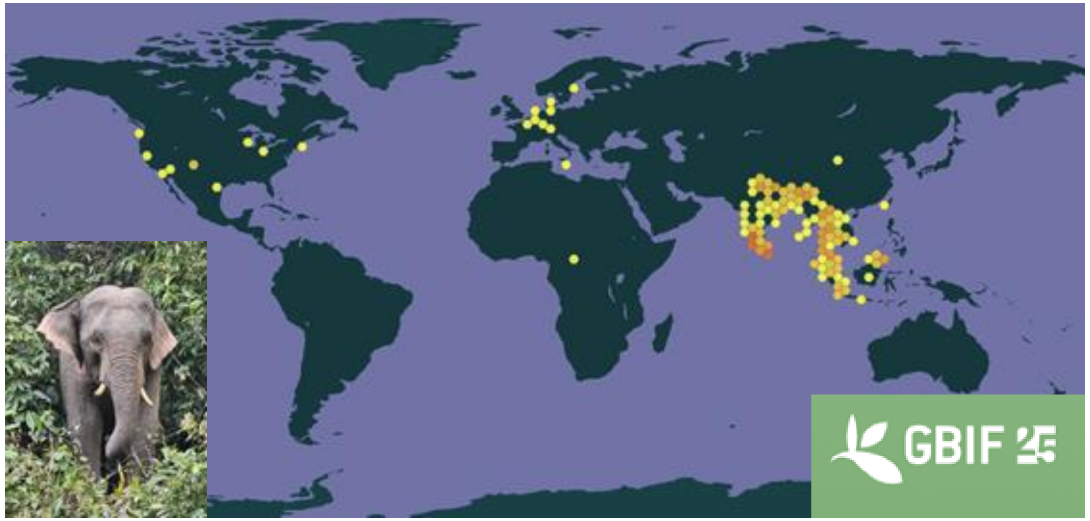
Know the ecology of the species of interest: This should be obvious, but it is good to remember that models need to make biological sense.



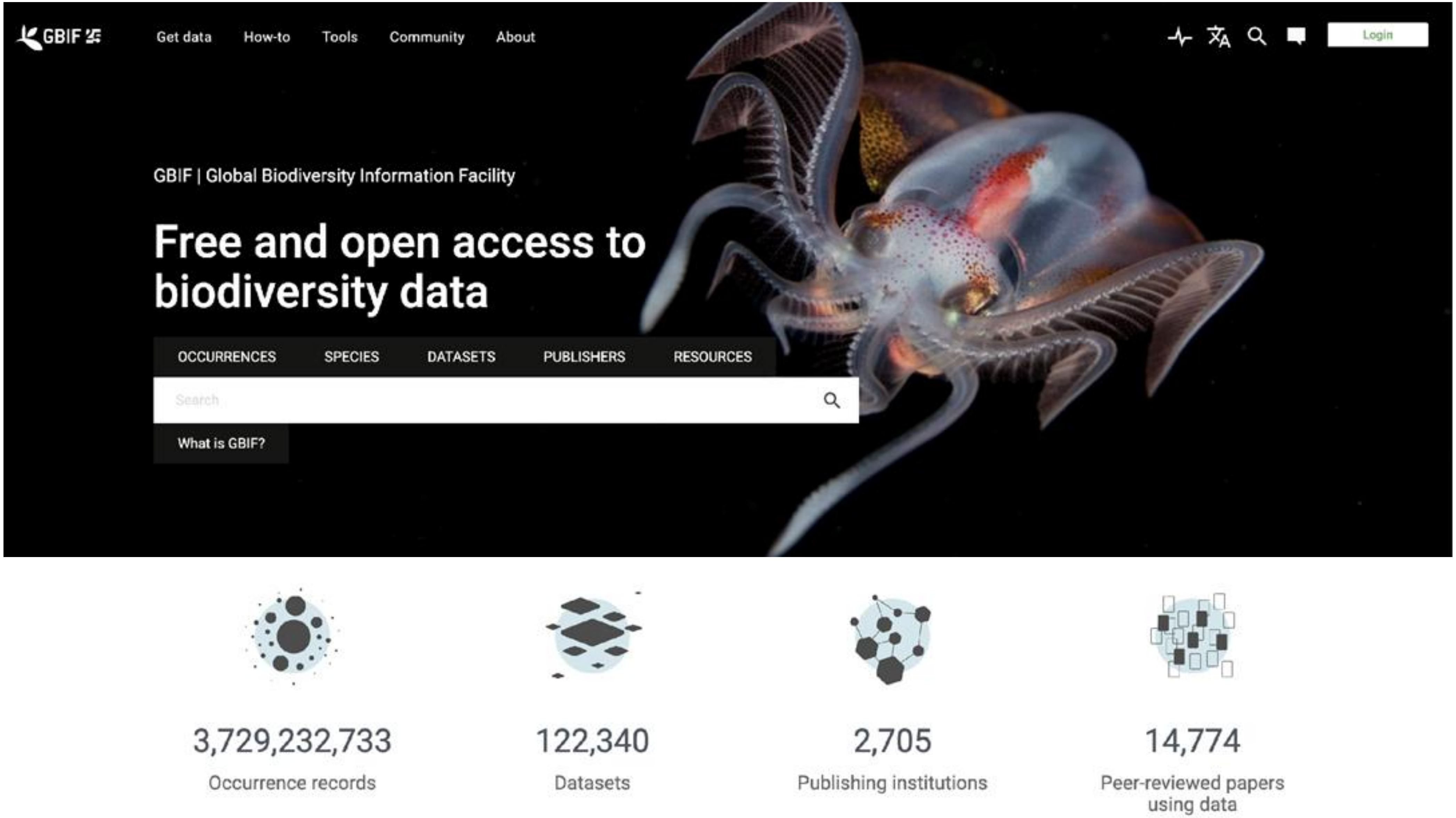
Model Inputs - Occurrence Data

Where to Access Occurrence Data (GBIF or Otherwise)

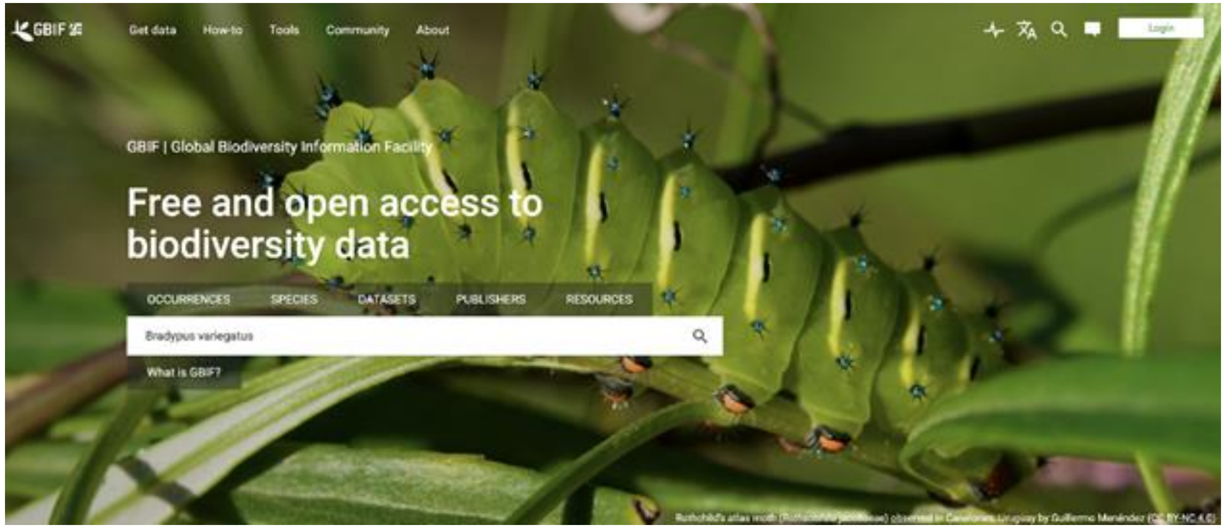
- Get species data(Global Biodiversity Information Facility, eBird, iNaturalist, movebank, your own field data)
  - Occurrence data
- Plot and inspect
- Clean
  - Remove duplicates
  - Correct outlier GPS coordinates
  - Georeferencing

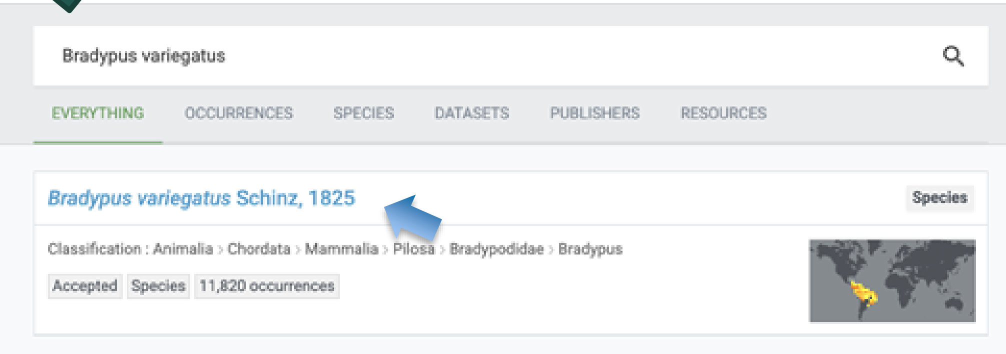


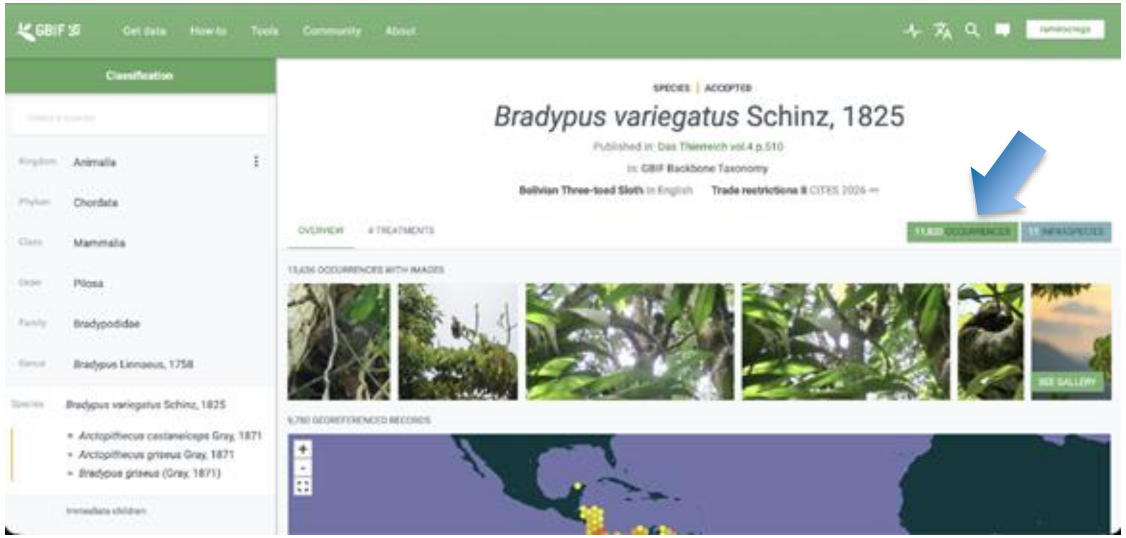
GBIF - Global Biodiversity Information Facility



GBIF – How to Obtain Data?

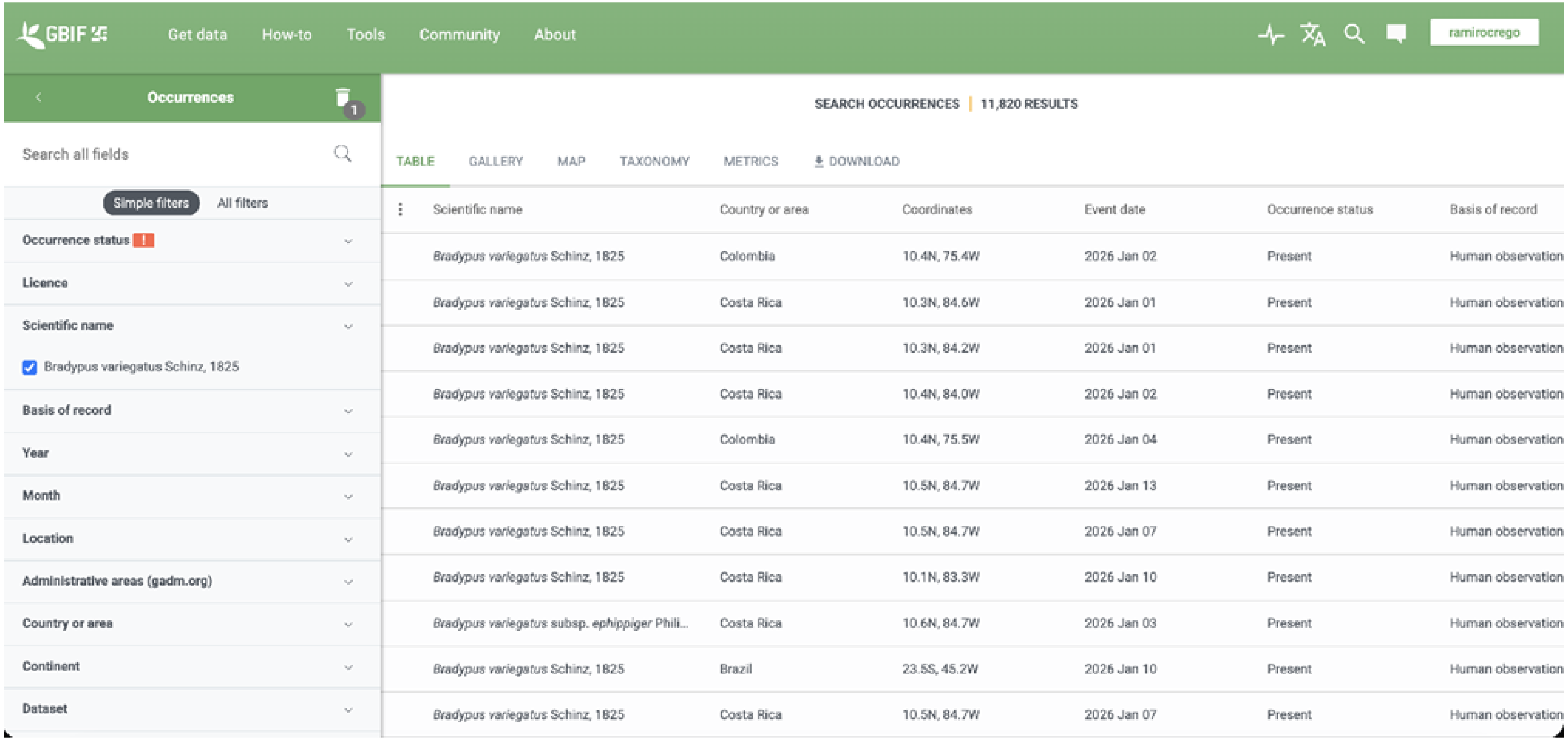






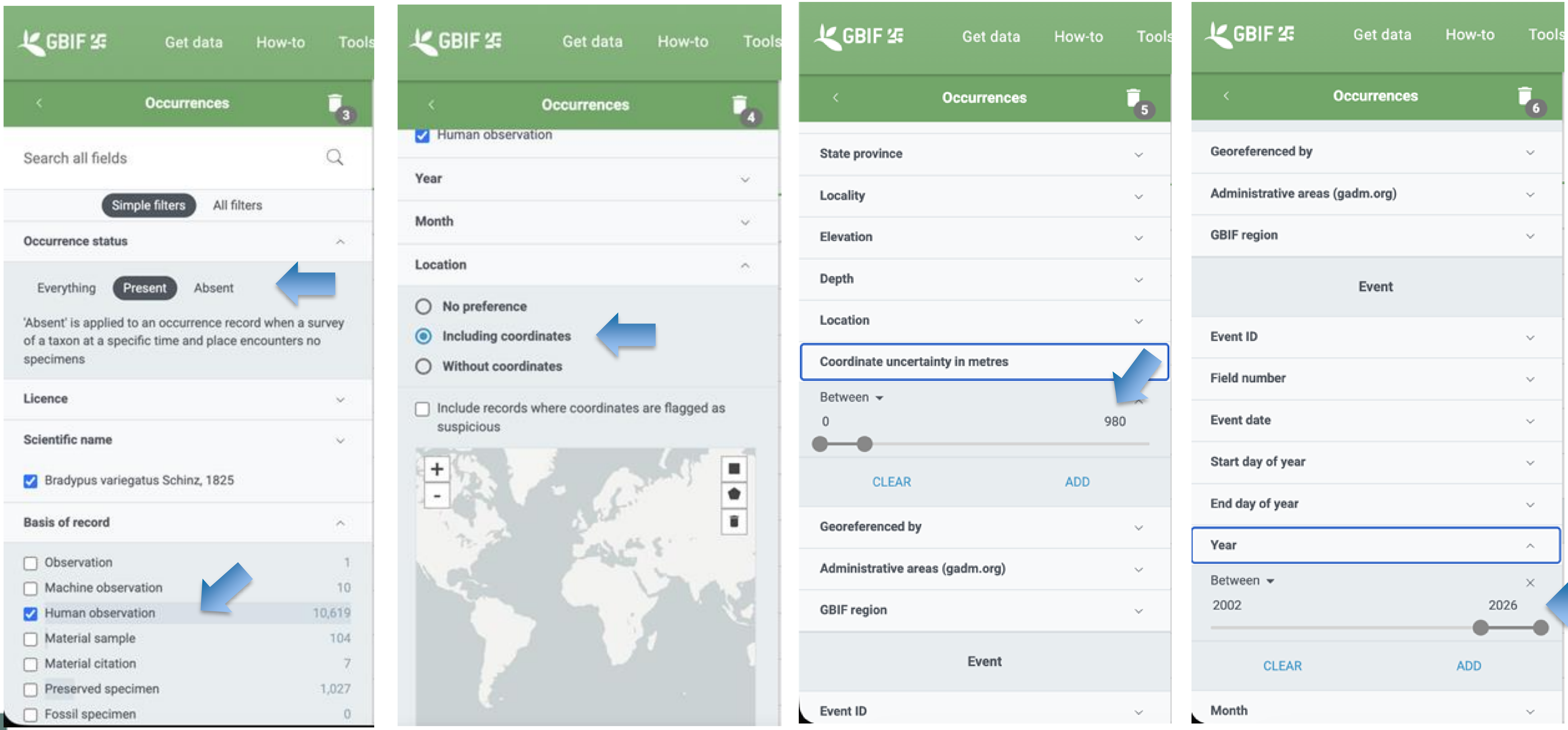
GBIF – How to Obtain Data?

Filter dataset

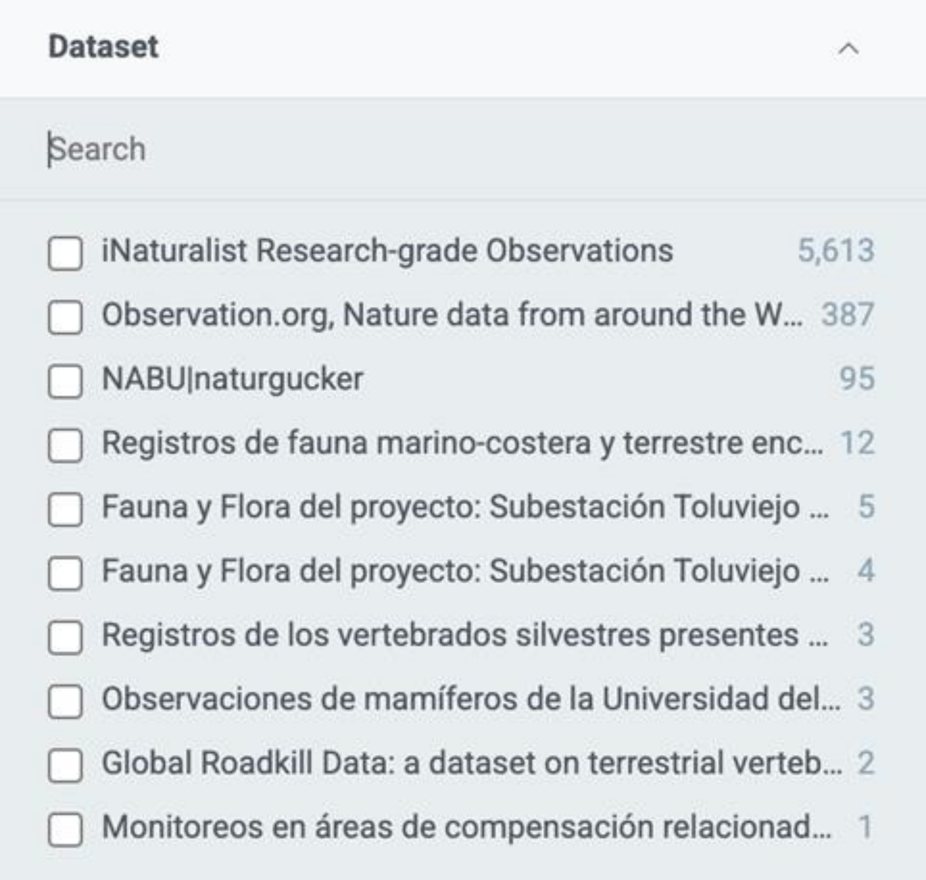


Scientific name	Country or area	Coordinates	Event date	Occurrence status	Basis of record
Bradypus variegatus Schinz, 1825	Colombia	10.4N, 75.4W	2026 Jan 02	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.3N, 84.6W	2026 Jan 01	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.3N, 84.2W	2026 Jan 01	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.4N, 84.0W	2026 Jan 02	Present	Human observation
Bradypus variegatus Schinz, 1825	Colombia	10.4N, 75.5W	2026 Jan 04	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.5N, 84.7W	2026 Jan 13	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.5N, 84.7W	2026 Jan 07	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.1N, 83.3W	2026 Jan 10	Present	Human observation
Bradypus variegatus subsp. *ephippiger* Phili...	Costa Rica	10.6N, 84.7W	2026 Jan 03	Present	Human observation
Bradypus variegatus Schinz, 1825	Brazil	23.5S, 45.2W	2026 Jan 10	Present	Human observation
Bradypus variegatus Schinz, 1825	Costa Rica	10.5N, 84.7W	2026 Jan 07	Present	Human observation

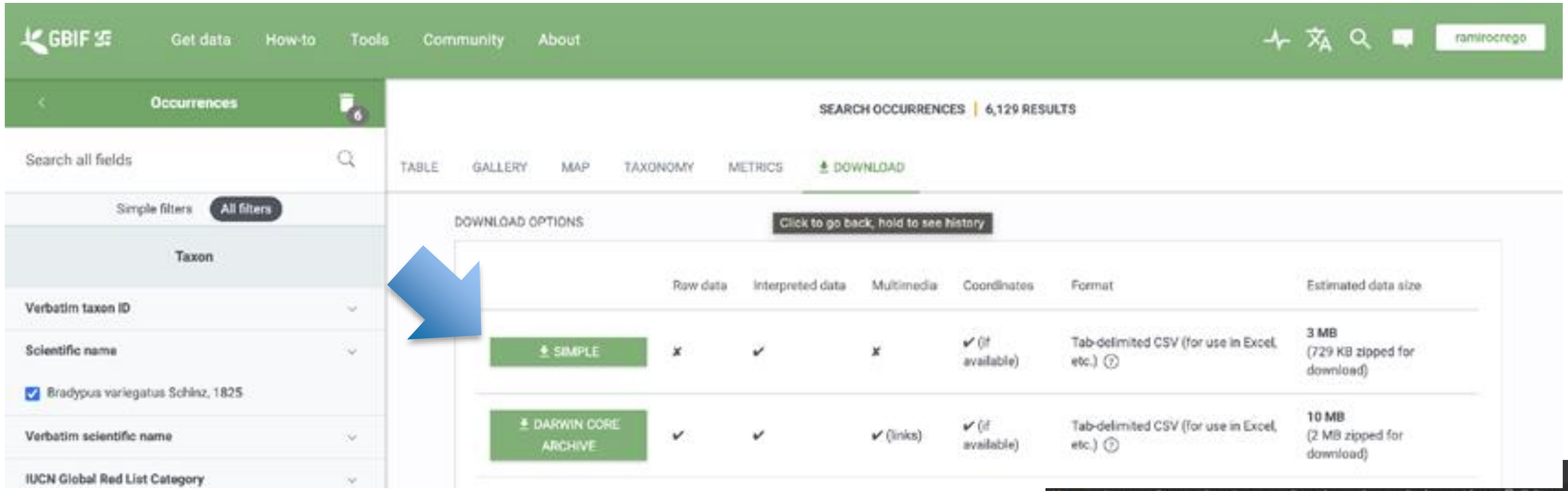
GBIF – How to Obtain Data? Filter Datasets



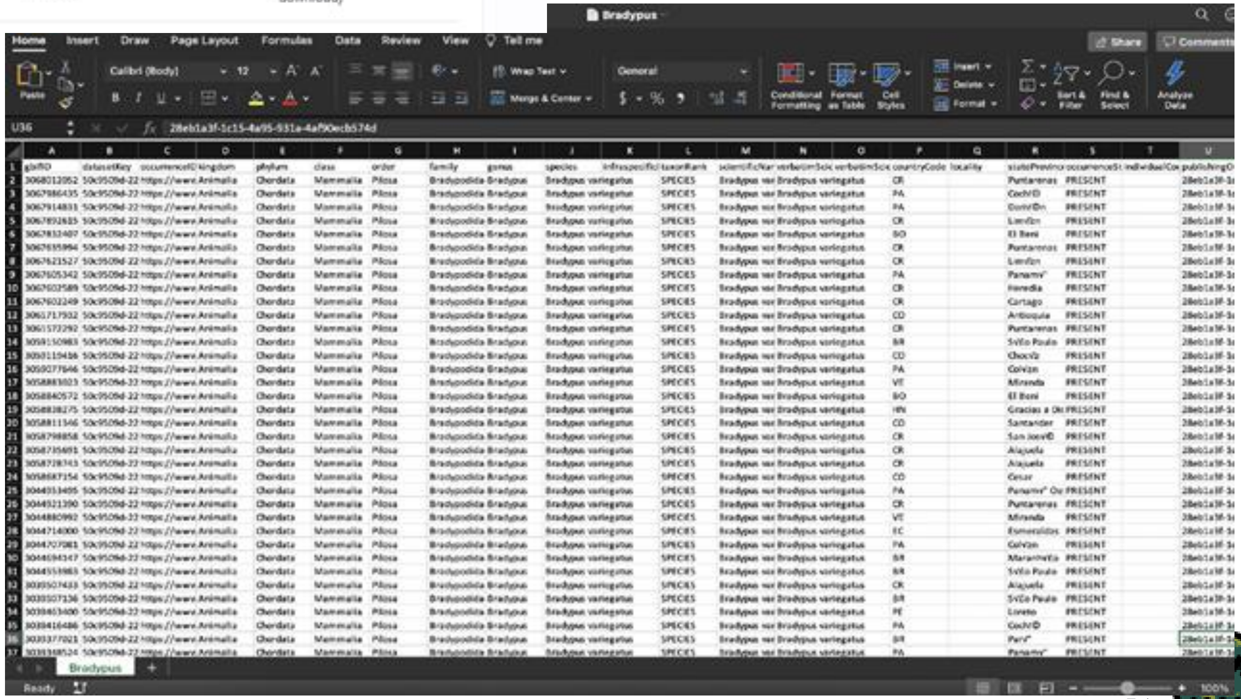
GBIF – How to Obtain Data? Search Occurrences

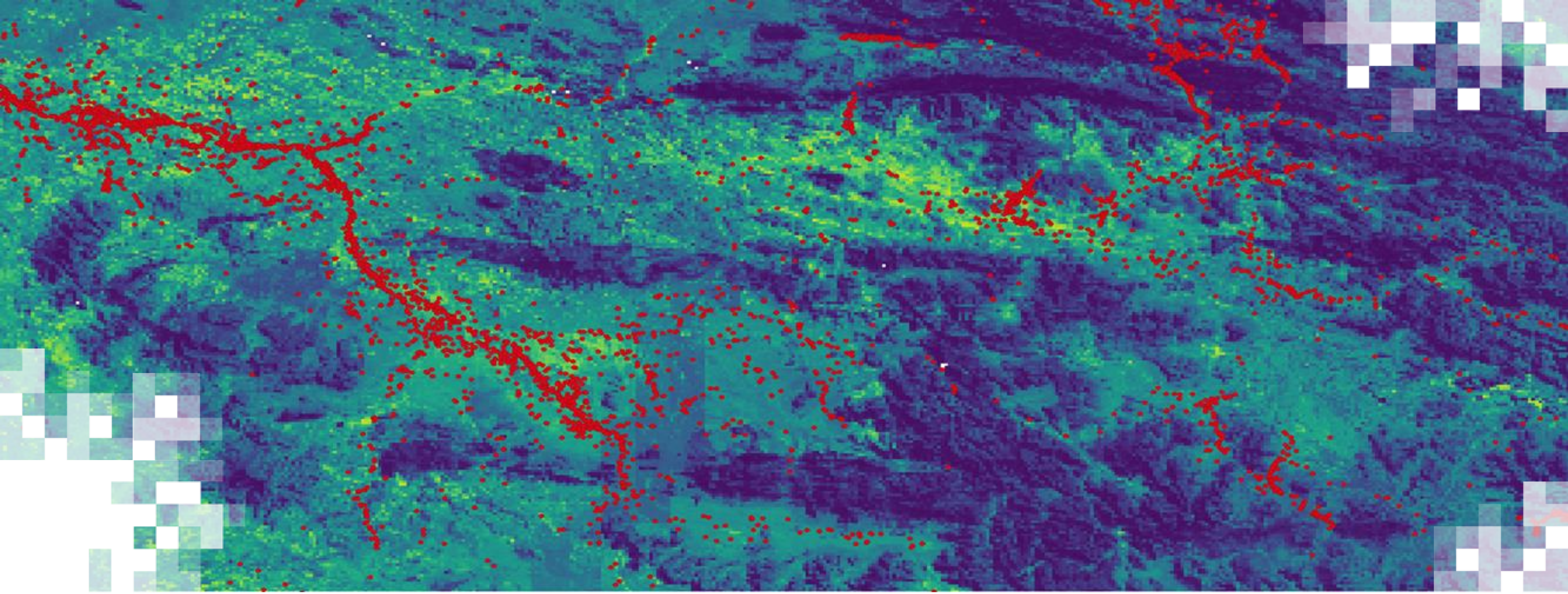


GBIF – How to Obtain Data? Postprocess



Postprocessing is often required to manually remove erroneous locations such as animals in zoos or records on top of cities

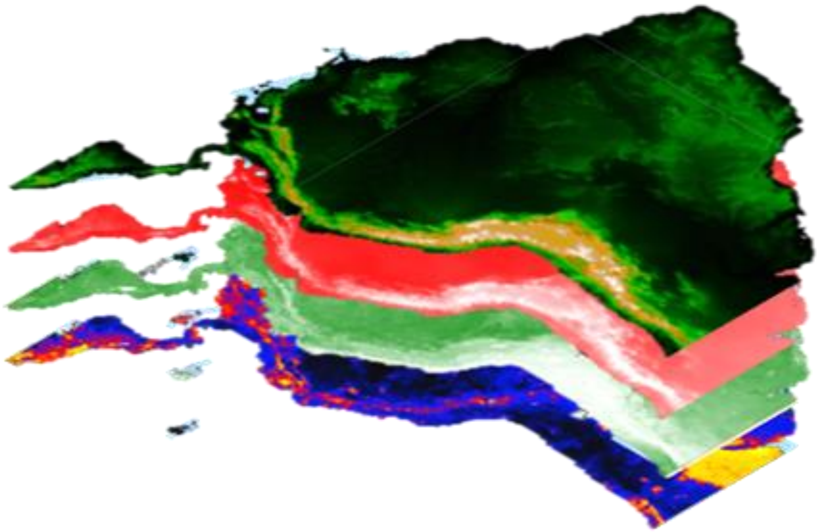




Model Inputs – Environmental Predictors

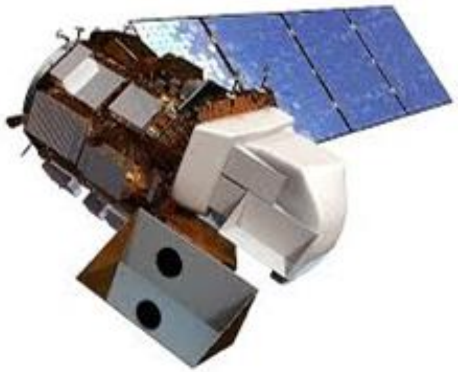
Predictor Variables

- Usually organized as raster data
- Each raster layer representing one predictor variable
- In GEE, raster data is known as image or image collections. We will learn how to use them shortly
- Images will serve as predictor variables for the models
- Extracting environmental data at species and background locations



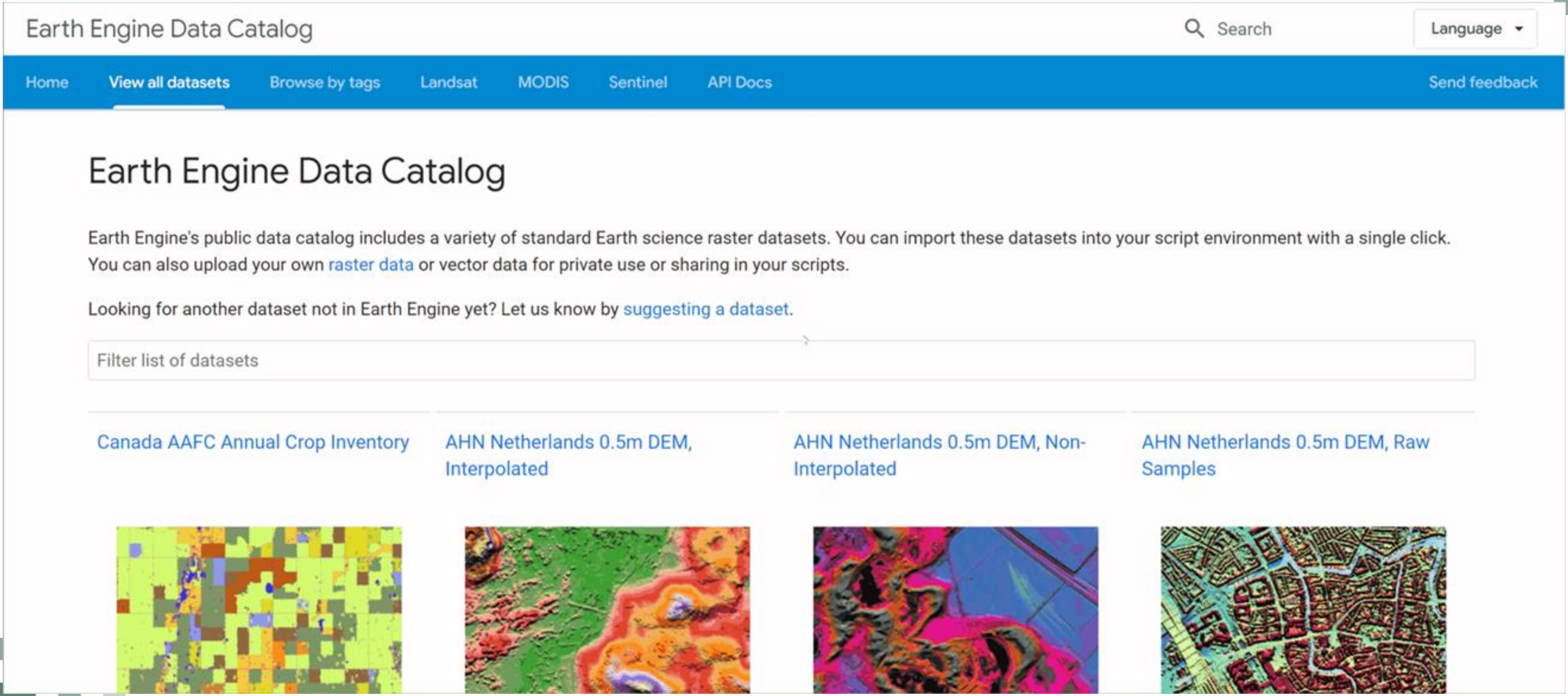
- E.g.:
  - A digital elevation model
  - Average annual temperature
  - A vegetation index
  - Or maybe the bands of a Landsat image

SRTM



Landsat-8

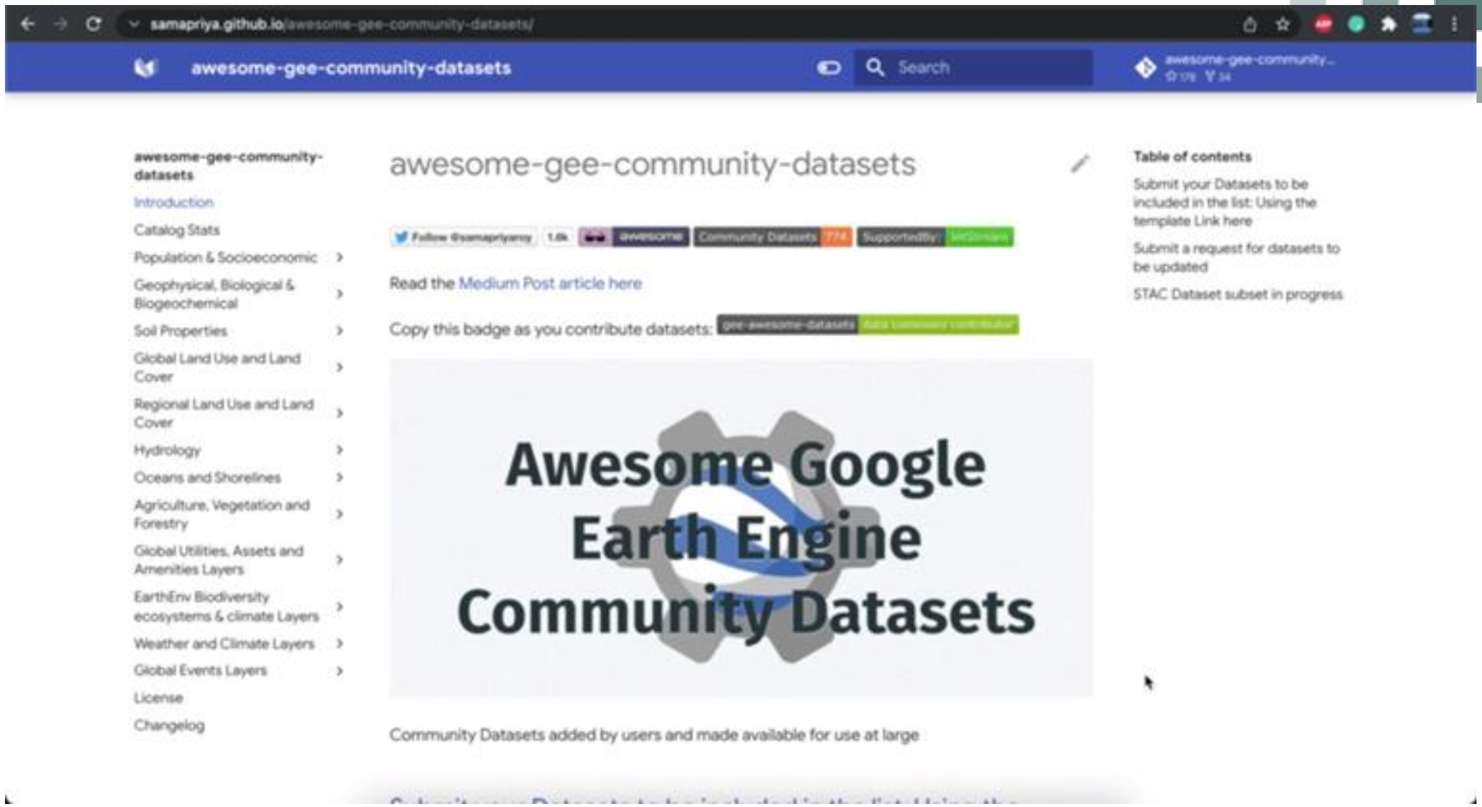
The GEE Catalog



There is More

Awesome GEE Community Catalog

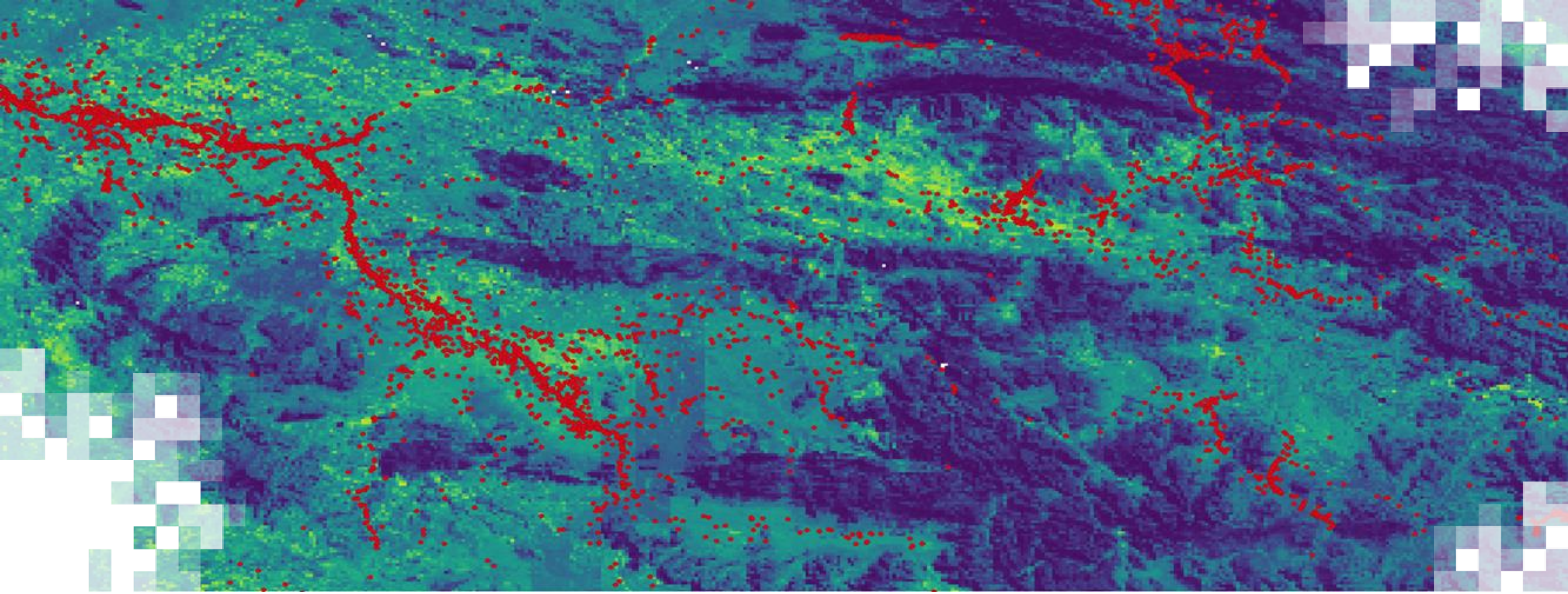
Samapriya Roy and Erin Dawn Trochim



The awesome-gee-community-catalog is an **unfunded open-source grassroots project** with a mission to help collect **community sourced** and **community generated** geospatial datasets.

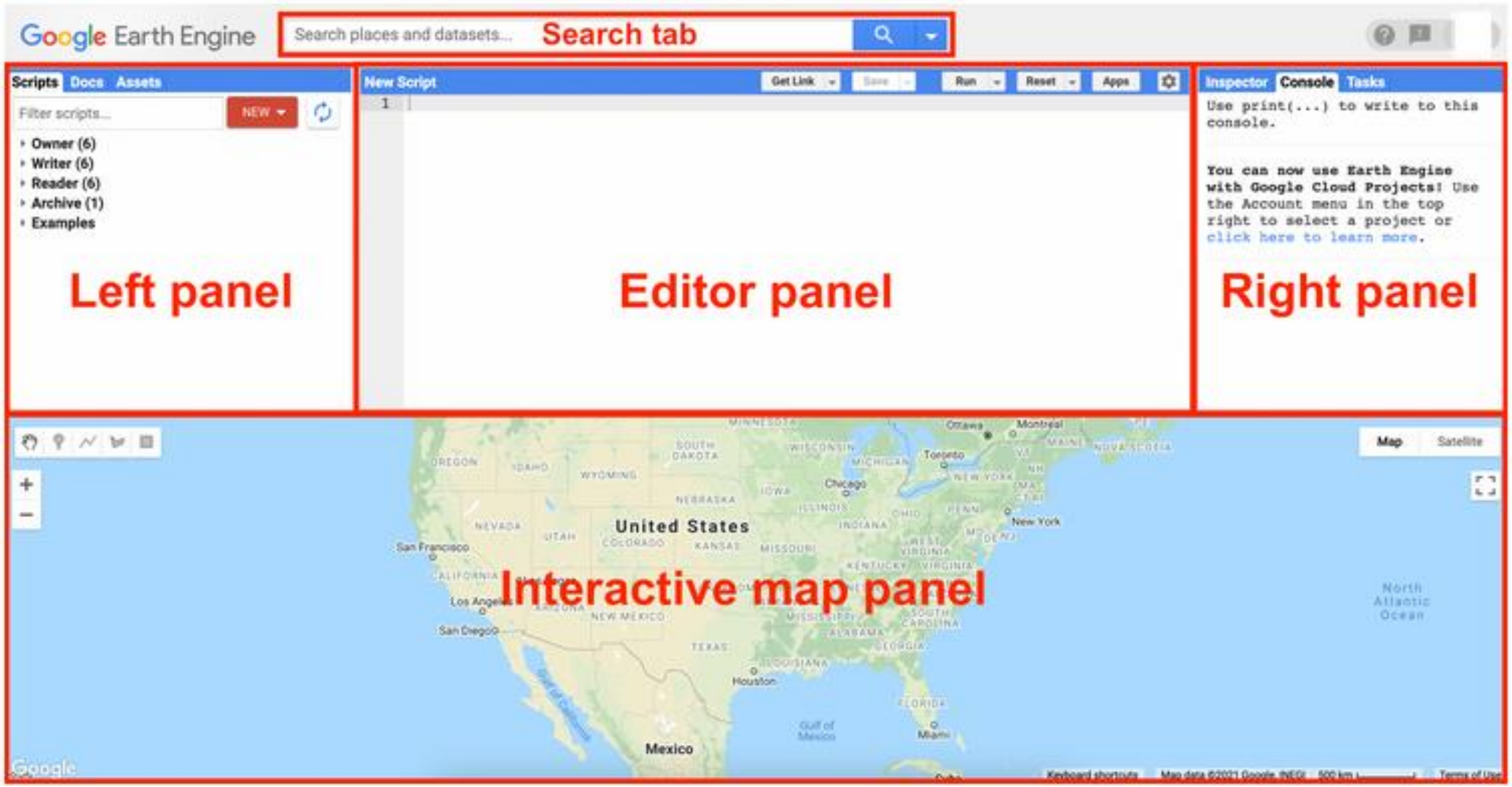
Our goal is to make data **accessible** and tie it to an analysis platform **fostering accessibility** and **reducing digital divide**.

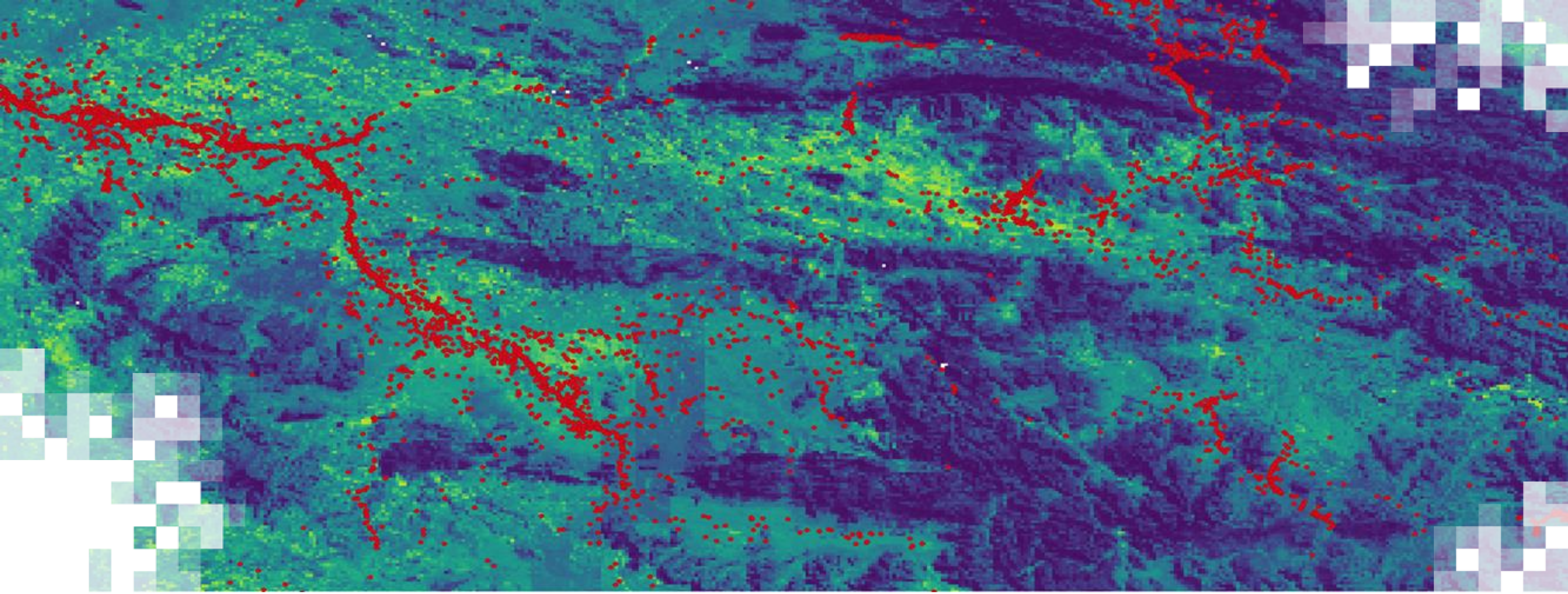
This catalog lives and serves alongside the [Google Earth Engine data catalog].



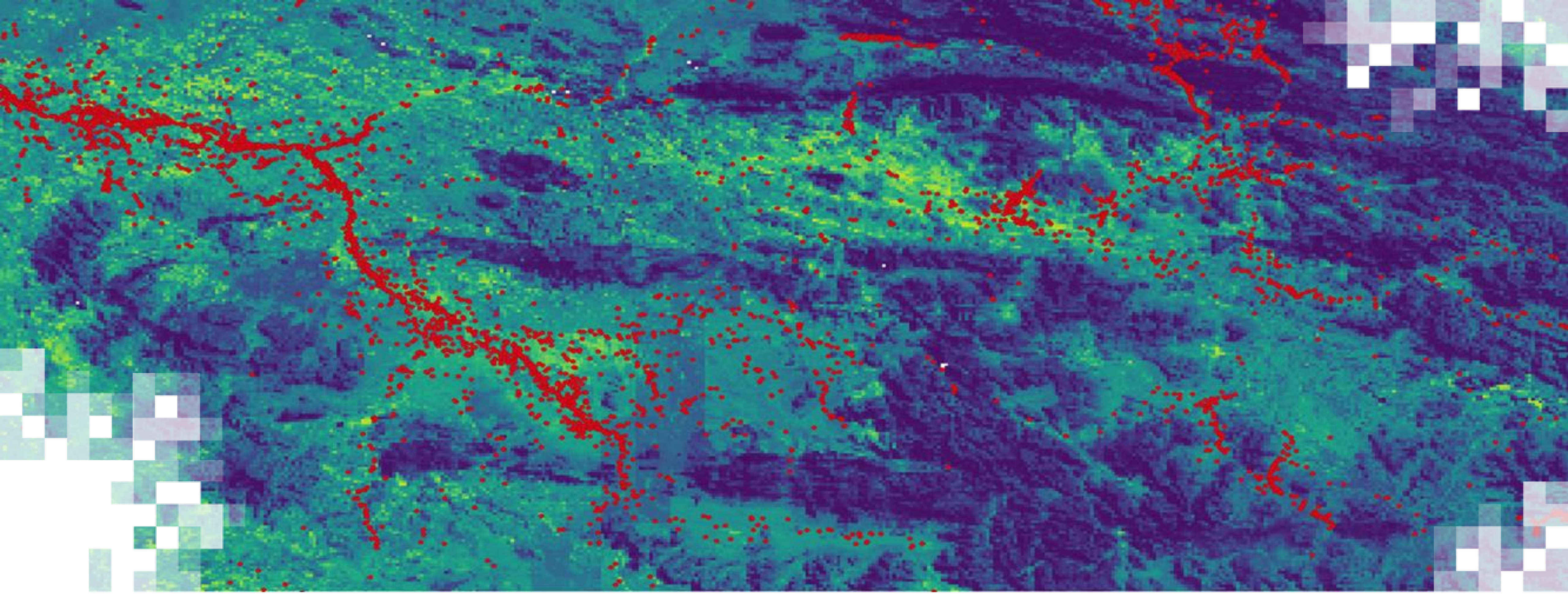
Brief Overview of GEE

The Code Editor





Live Demo



Species Distribution Modeling with Google Earth Engine

Part 1: Introduction to Species Distribution Models and Habitat Suitability

Summary

Part 1 Summary

- Provided an introduction to Species Distributions Models (SDM) including overview of core concepts, applications, methods, model inputs and interpreting SDM.
- Discussed benefits and limitations of Google Earth Engine for applied ecology.
- Explored model inputs, such as occurrence data in GBIF - Global Biodiversity Information Facility and environmental predictors in the GEE Catalog.
- Outlined best practices for producing habitat suitability maps.
- Explored the Google Earth Engine (GEE) web interface.

Looking Ahead to Part 2

- Building, Evaluating, and Interpreting a SDM on GEE

- July 14, 2026
  Session A: 9:00 a.m. to 10:30 a.m. EDT (UTC-4)
  Session B: 4:00 p.m. to 5:30 p.m. EDT (UTC-4)

Resources

- Original manuscript: Implementation of species distribution models in Google Earth Engine
- The Google Earth Engine code and data used in the manuscripts are freely available at the following Google Earth Engine repository.
- All the Google Earth Engine code used in this tutorial can be accessed in the following links: Intro to GEE and Data management in GEE.

Homework and Certificates

- **Homework:**
  - One homework assignment
  - Opens on 07/14/2026
  - Access from the training webpage
  - Answers must be submitted via Google Forms
  - **Due by 07/28/2026**

- **Certificate of Completion:**
  - Attend two live webinars (attendance is recorded automatically)
  - Complete the homework assignment by the deadline
  - You will receive a certificate via email approximately two months after completion of the course.

Contact Information

Hosts:

- Justin Fain
  - justin.j.fain@nasa.gov
- Sativa Cruz
  - sativa.cruz@nasa.gov

- ARSET Website
- ARSET YouTube

For questions, comments, or to share how you have applied our trainings to your work or studies, email nasa.arset@gmail.com.

Join our mailing list to stay up-to-date on our latest trainings. Visit our Contact page to subscribe.

APPLIED REMOTE SENSING TRAINING PROGRAM

TRAIN - EMPOWER - ADVANCE

Thank You!