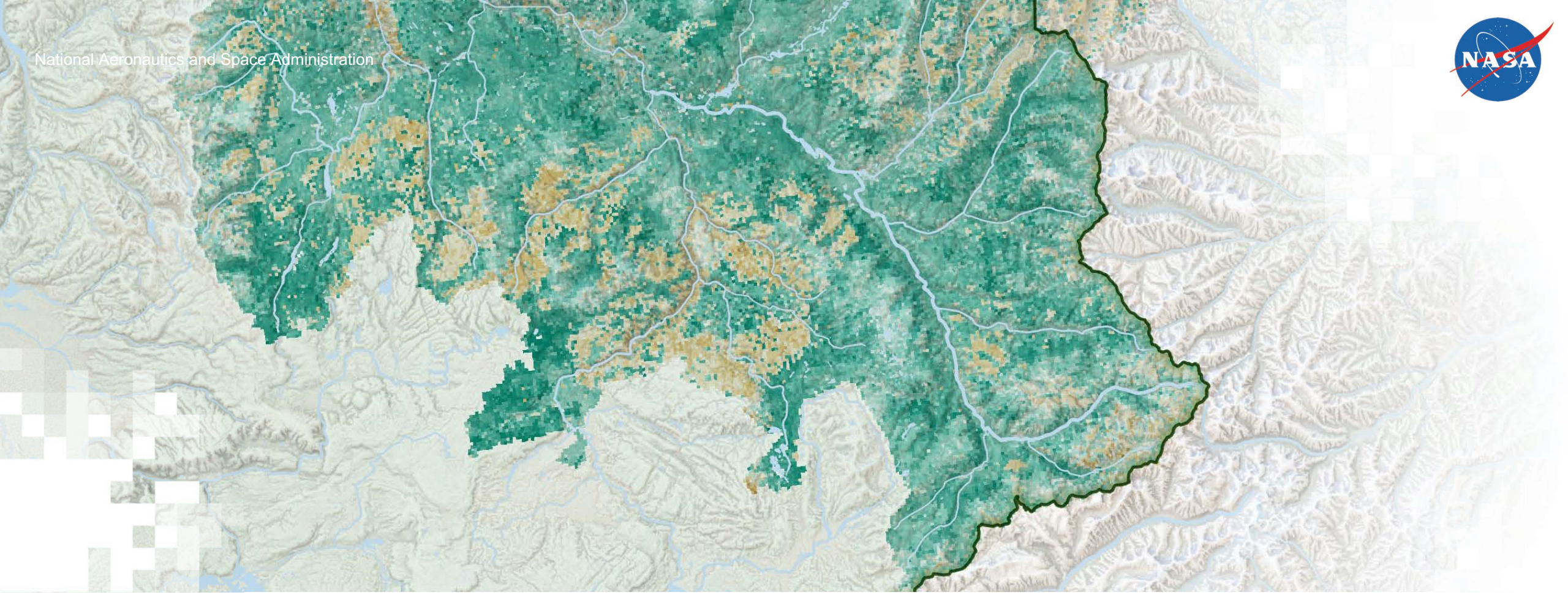


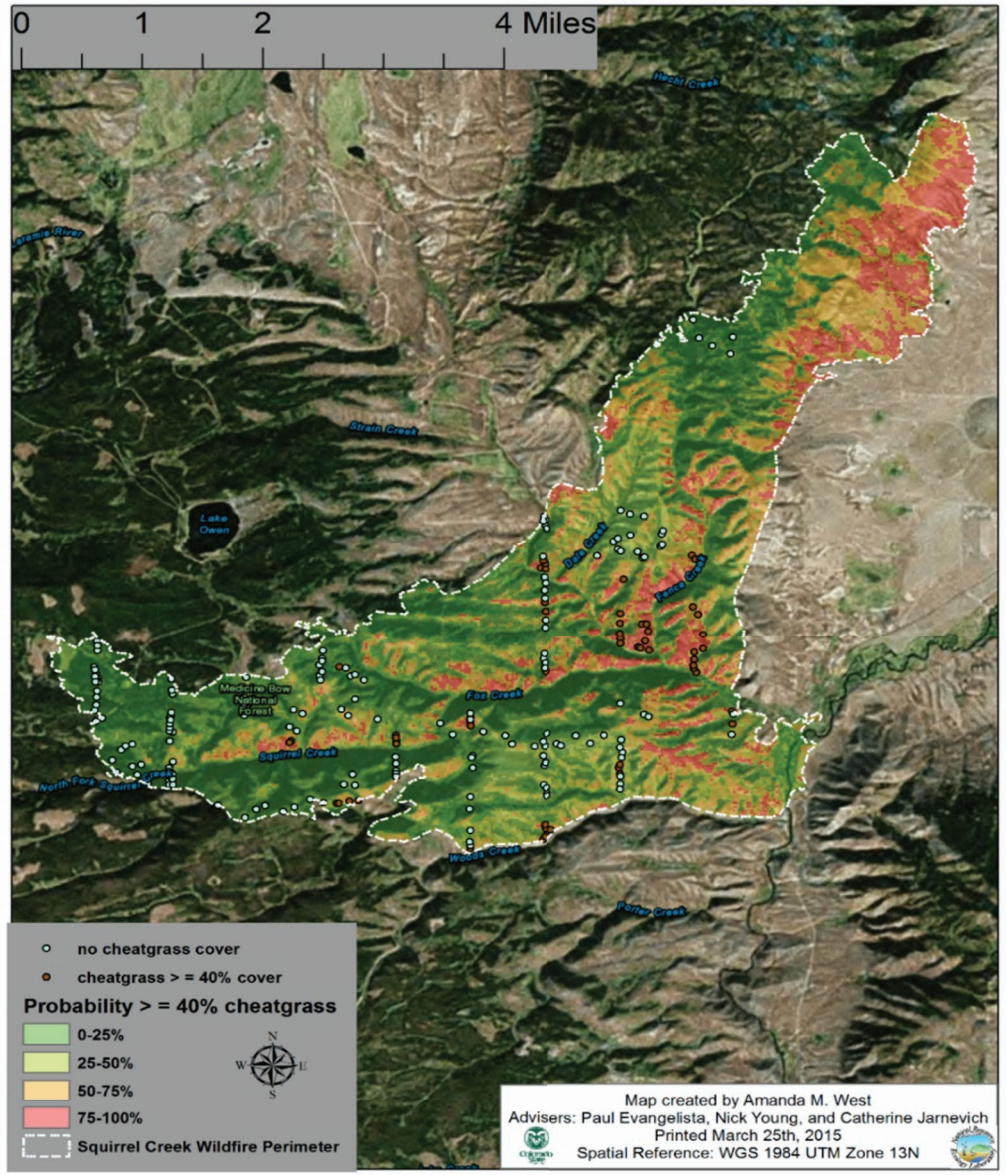
Species Distribution Modeling with Remote Sensing

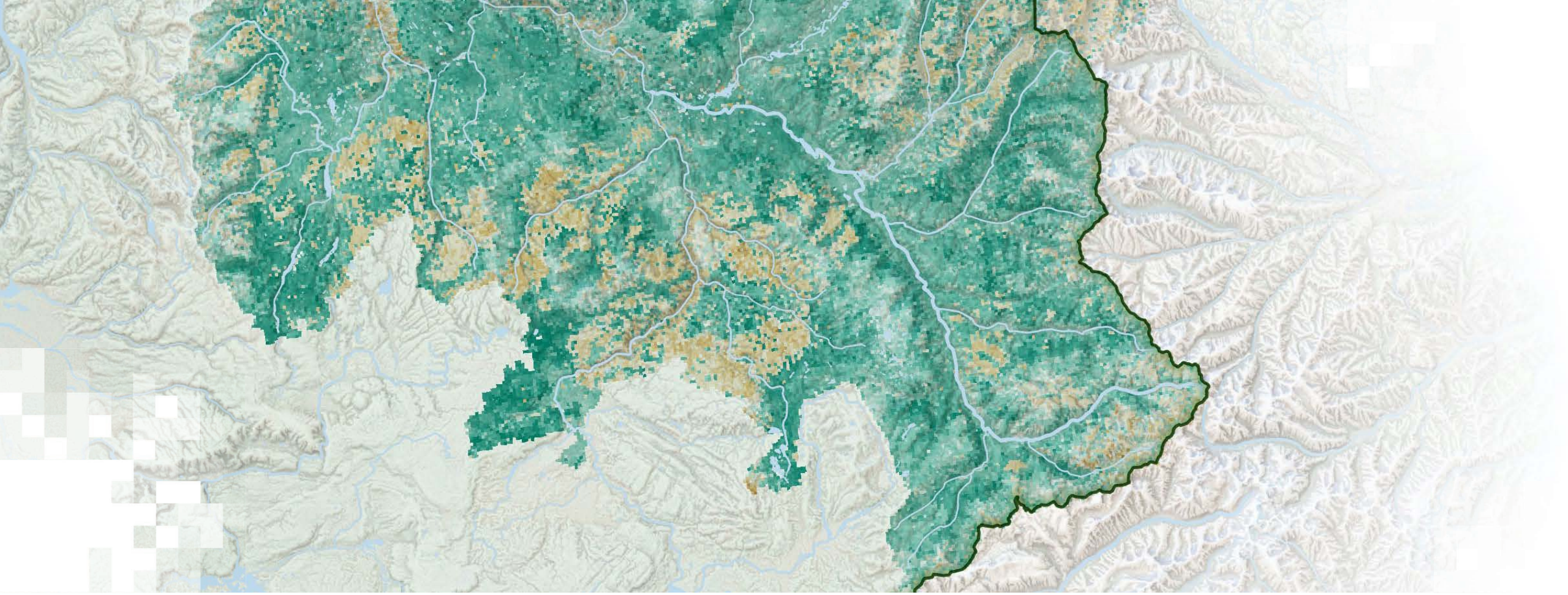
Amber McCullum, Juan Torres-Perez, Zach Bengtsson

Aug 12, 2021

Part 1 Overview

- Species Distribution Models (SDMs) Overview
- Environmental Variables
- Occurrence Data
- Methods and Models
- Case Study Examples
- Question and Answer Session

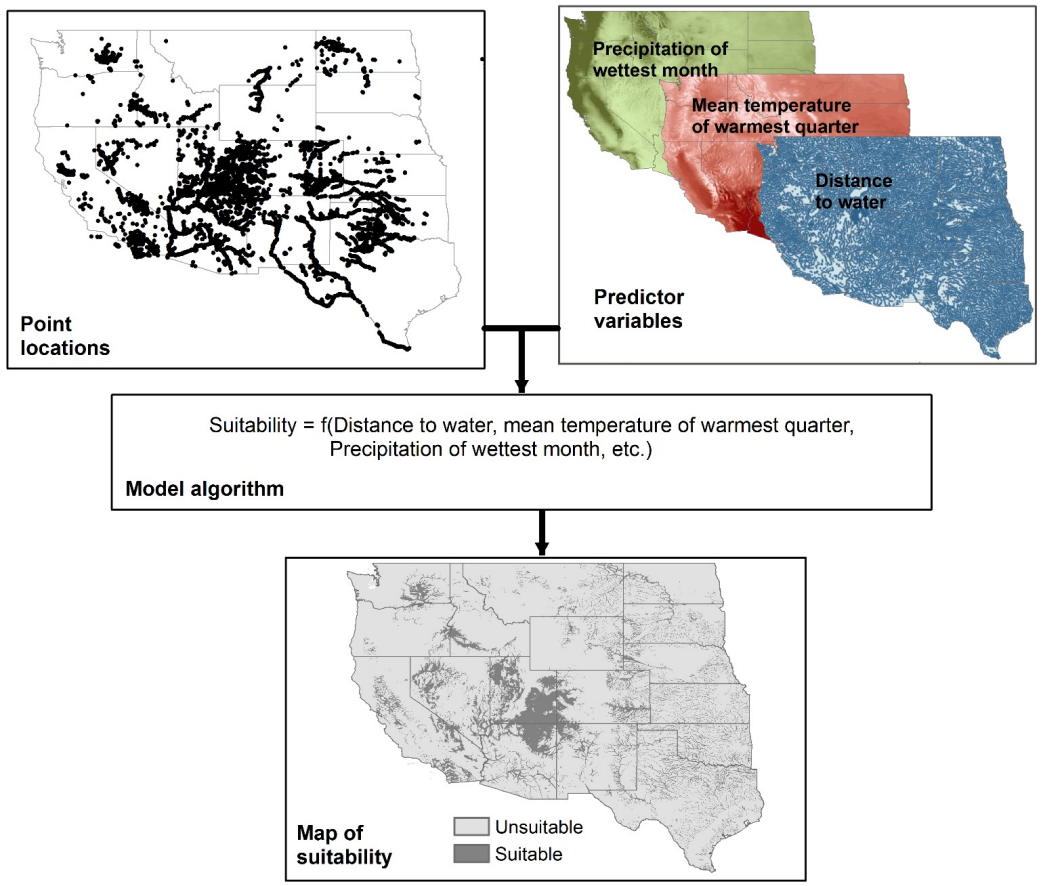




Species Distribution Models (SDMs) Overview

Species Distribution Models (SDMs)

- Species Distribution Models allow you to assess the suitability of a habitat for a species.
- The models use raster-based layers such as land use/land cover, elevation, and others as predictors of suitable habitats.
- The predictor data is combined with ground-collected presence-absence or abundance data in empirical statistical models.

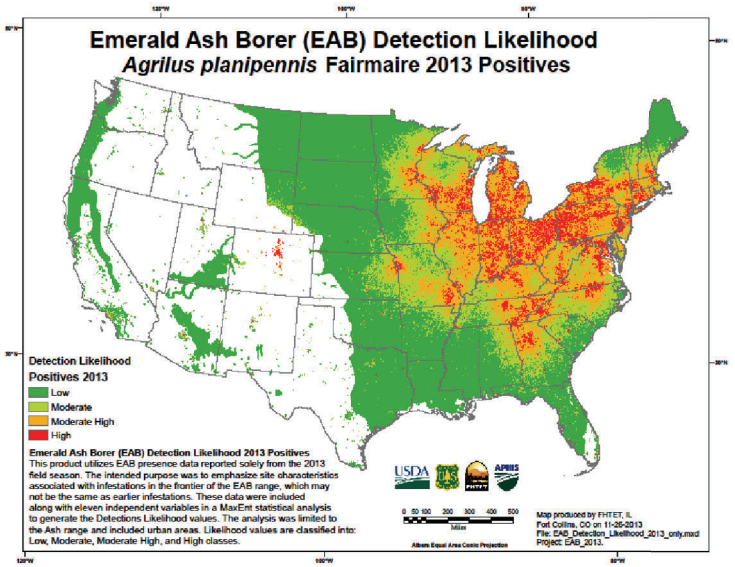


Jarnevich, C. S., T. J. Stohlgren, S. Kumar, J. T. Morrisette, and T. R. Holcombe, 2015, Caveats for Correlative Species Distribution Modeling: Ecological informatics, v. 29, p. 6-15.

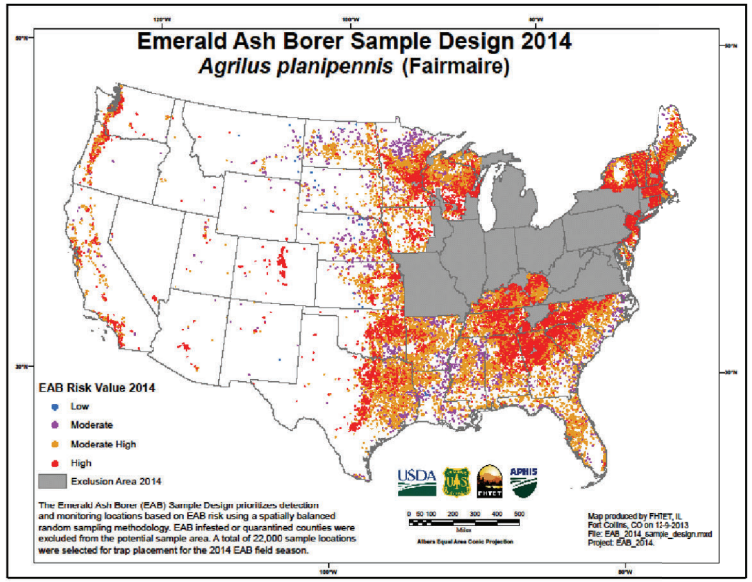
SDM Applications

- Invasive species control
- Risk assessment
- Conservation planning
- Monitoring strategies
- Scenario modeling under a changing climate

West et al. 2017, International Journal of Applied Earth Observation and Geoinformation



USDA – model for EAB



SDM Needs and Outputs

- **Environmental Variables**
  - Characteristics of suitable habitat for a particular species
- **Occurrence Data**
  - Point locations of species presence
- **Models/Algorithms**
  - Use of environmental and occurrence data to predict current and/or future species distributions
- **Output**
  - Habitat suitability (past, current, future)
    - Will need different types of data depending on the time period of interest

SDMs…

DO

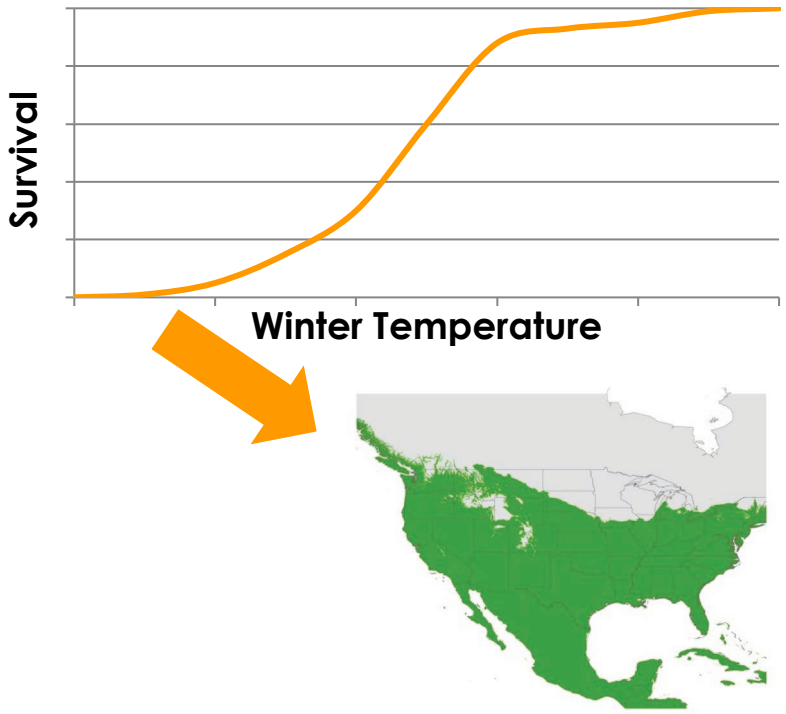
Identify areas with environmental conditions similar to where a species occurs

DO NOT

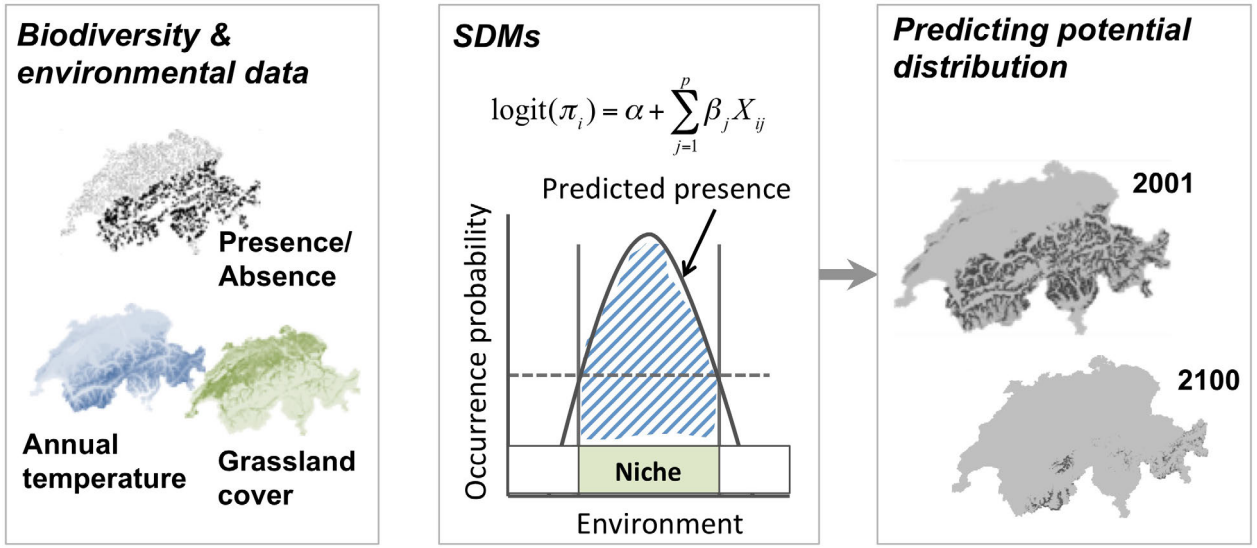
Identify where a species is actually found

SDM Types

- **Process-Based** – Mechanism; physiological constraints



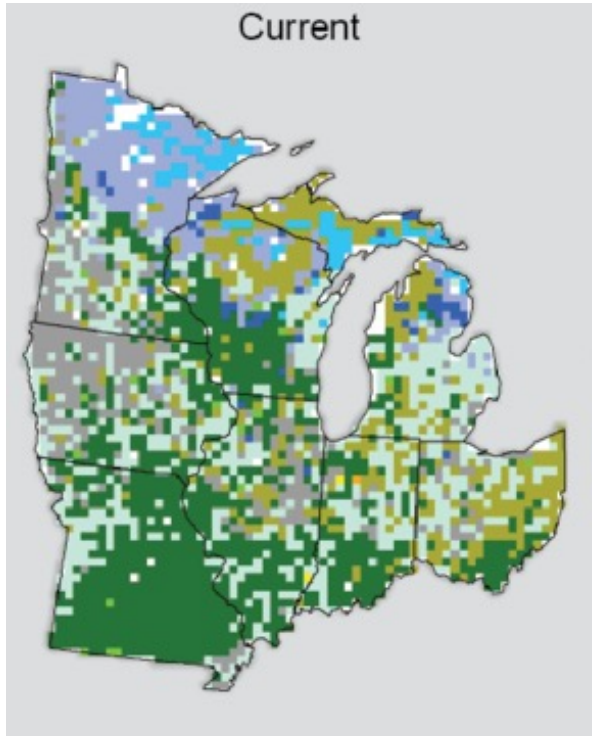
- **Correlation-Based** – Pattern; based on current locations

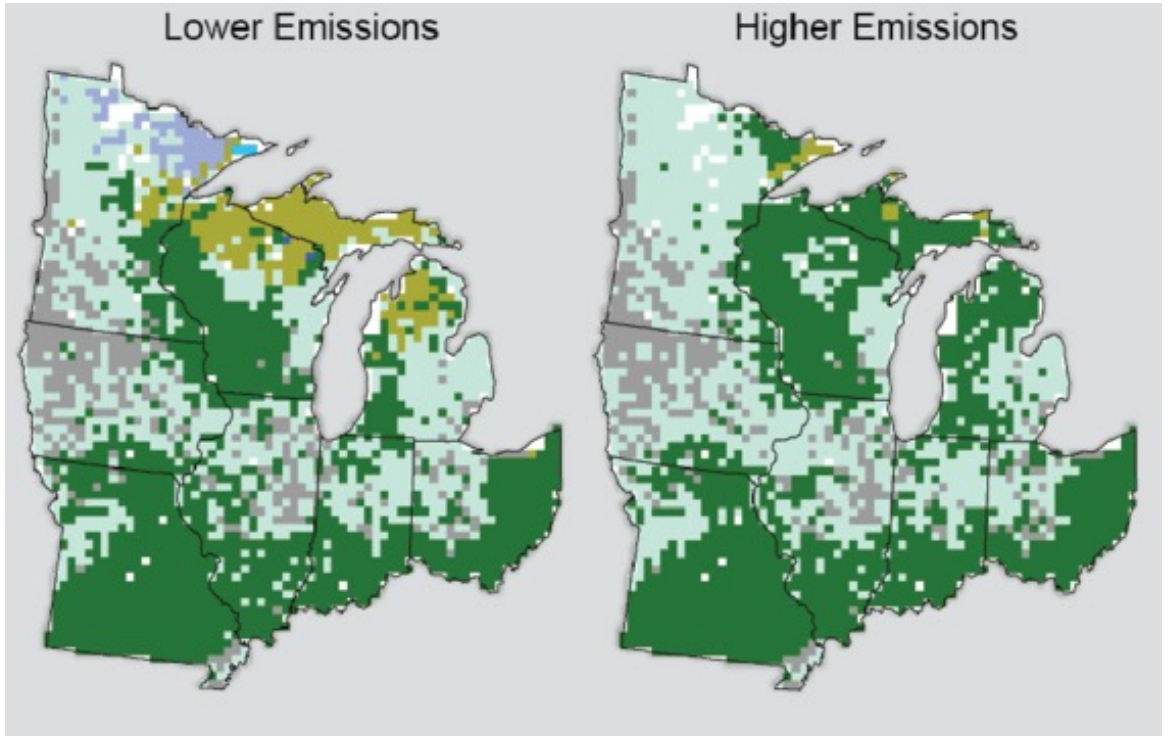


SDM in R: Zurell

SDM Types

- **Current** – Where is it now?
  - Mapping
- **Future** – Where might it be?
  - Potential

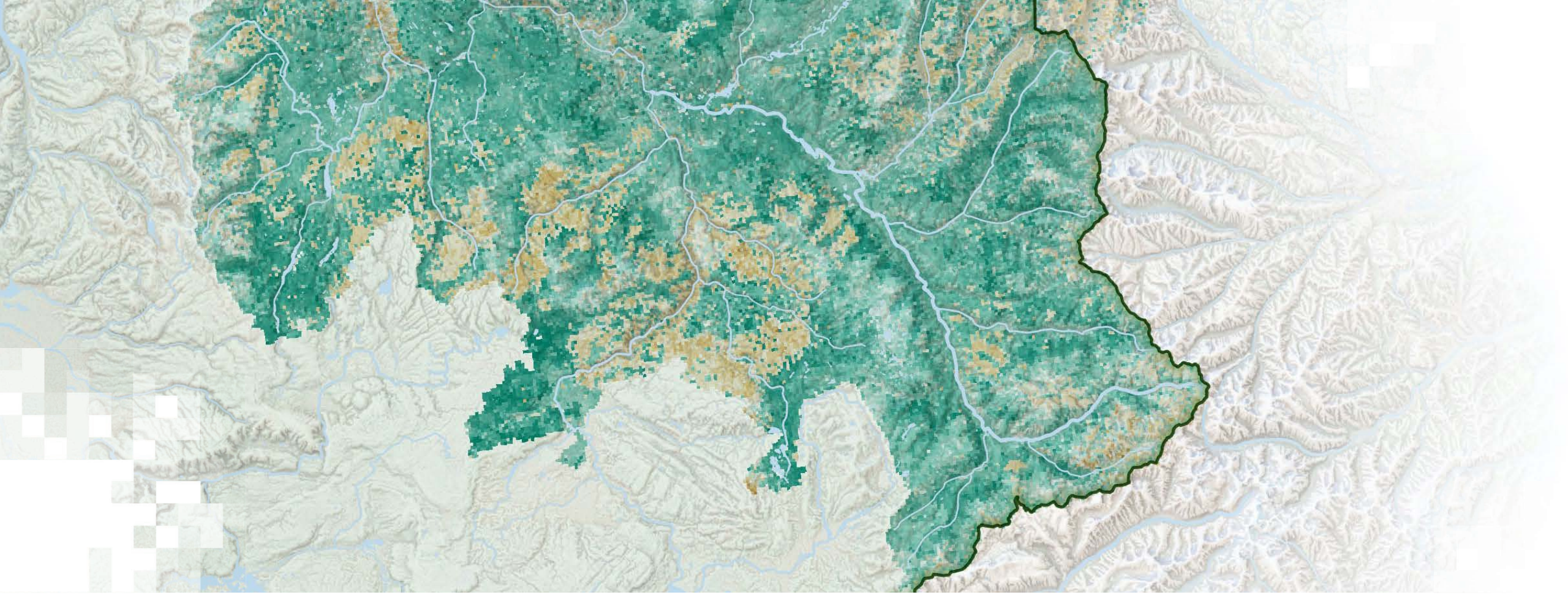




Forest Types

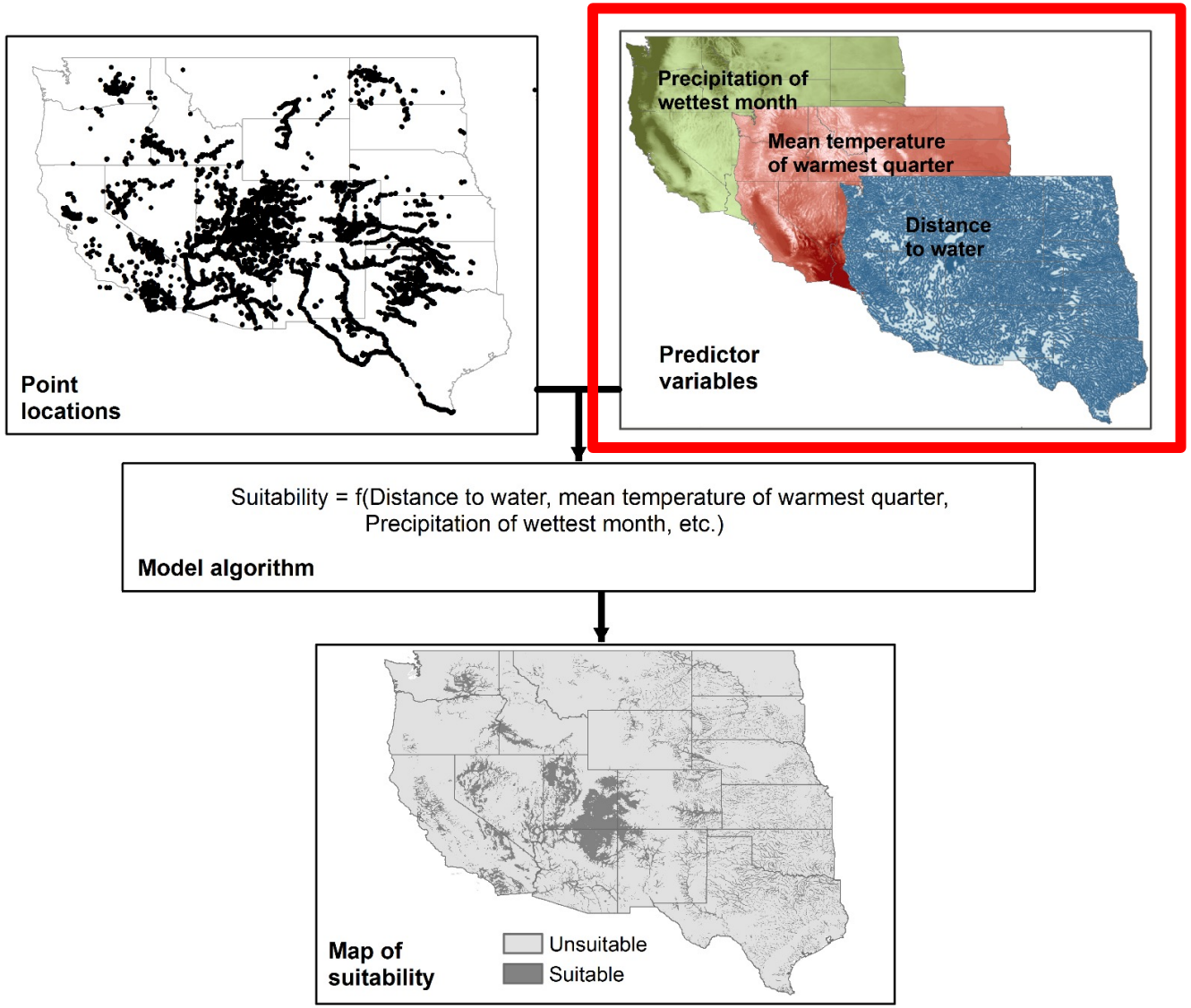
White/Red/Jack Pine
Loblolly/Shortleaf Pine
Oak/Hickory
Maple/Beech/Birch
Elm/Ash/Cottonwood
Spruce/Fir
Oak/Pine
Oak/Gum/Cypress
Aspen/Birch
No Data

Pryor, et al, 2014



Environmental Variables

Environmental Variables

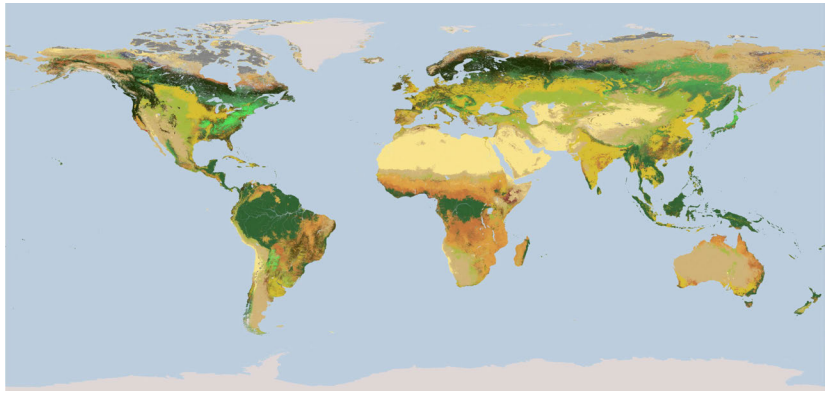


Jarnevich et al, 2015

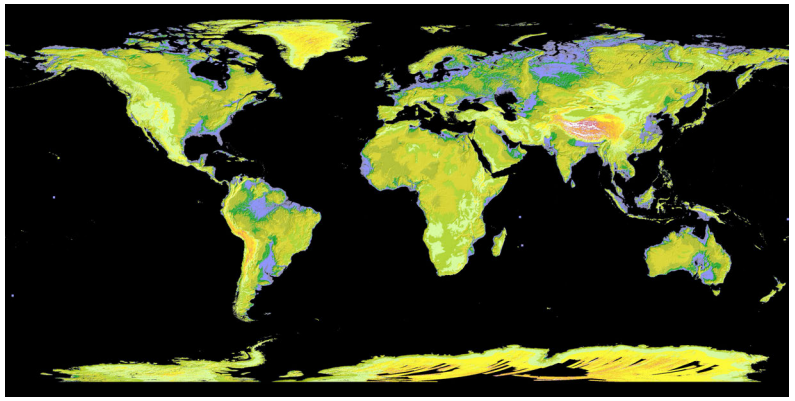
Environmental Variables

- Land Cover
- Phenology
- Vegetation Indices and Fractional Cover
- Tree Mortality (insect/disease)

- Topography: Elevation, Slope, Aspect
- Climatology: Temperature (min., max., mean, etc.), Precipitation (min., max., etc.)



MODIS Land Cover Product



ASTER Digital Elevation Model (DEM)

Land Cover and Species

- Land cover is one of the most important drivers of biodiversity.
- It affects patterns of species diversity, distributions, and ecological processes.
- Accurate land cover maps are essential for SDMs.
  - Land cover changes can dramatically affect species distributions.
- Many regional and global land cover maps are available.
- Creating your own land cover maps with ground-based data is ideal.

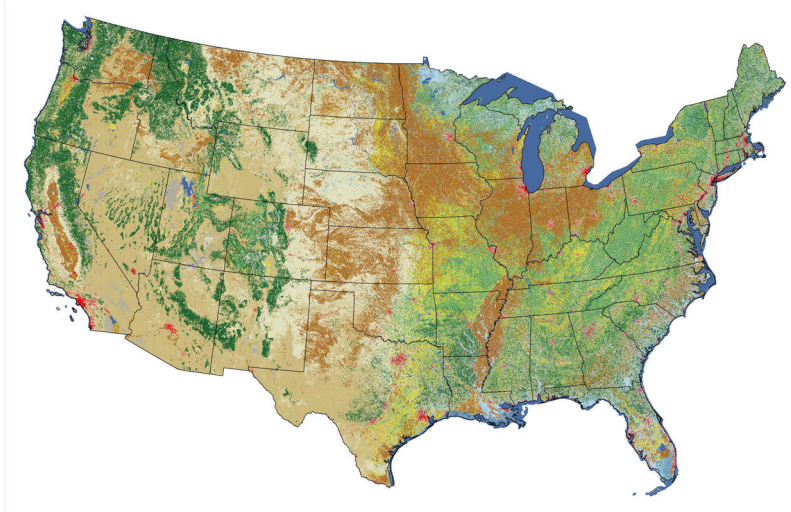
Environmental Variables: Land Cover Products

United States:

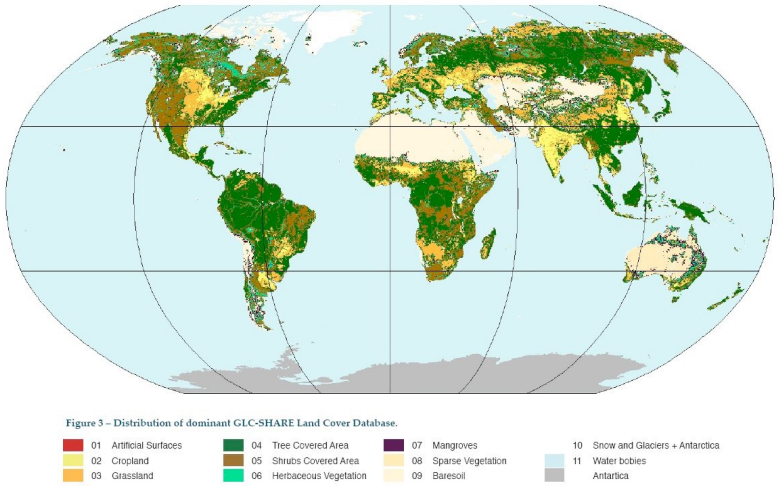
- National Land Cover Database (NLCD)
- GAP Analysis
- LANDFIRE

Global:

- MODIS Land cover product
- FAO Global Land Cover-SHARE
- ESA Climate Change Initiative Land Cover



National Land Cover Database 2011



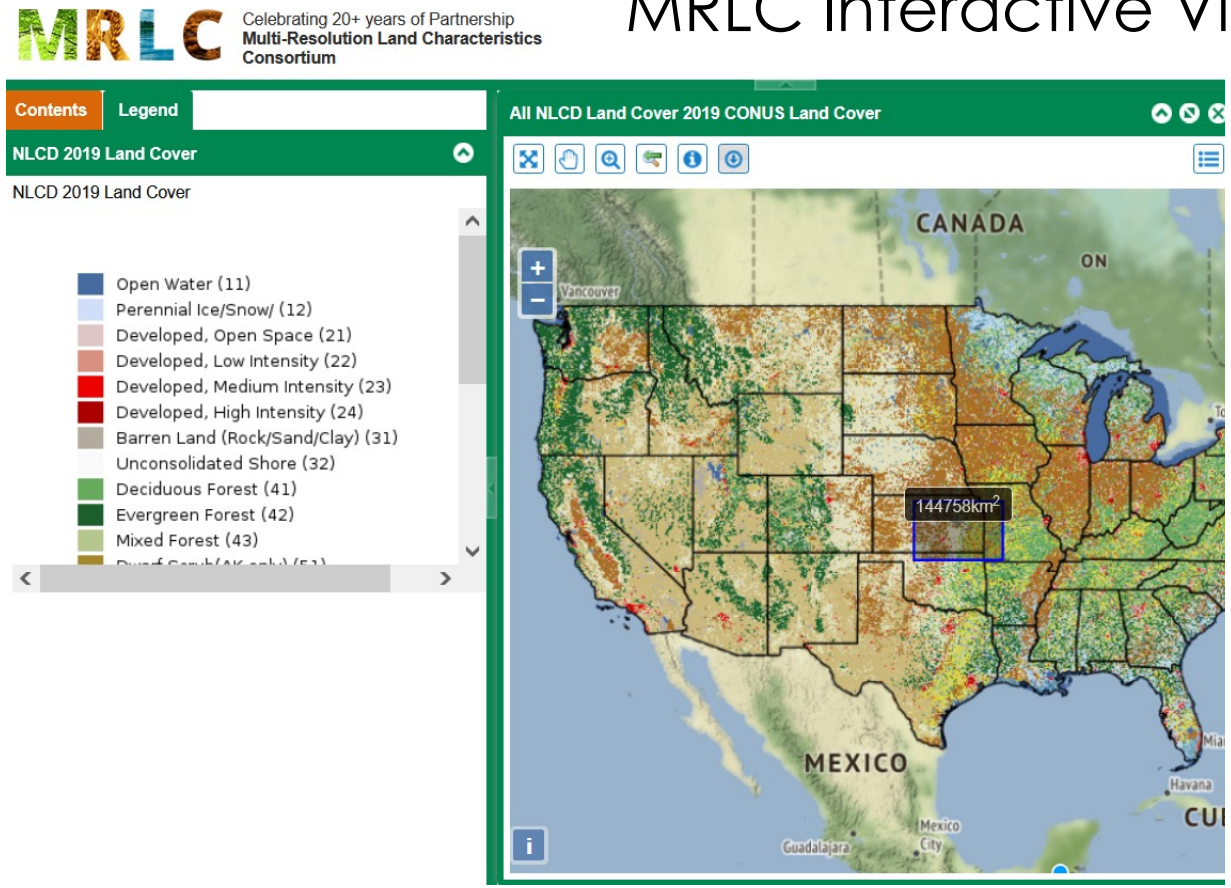
FAO Global Land Cover

Multi-Resolution Land Characteristics (MRLC) Consortium

https://www.mrlc.gov

- National Land Cover Database (NLCD)
- Landsat-based, 30 m resolution
- 16 land cover classes
- 2001, 2006, 2011, 2013, 2016, 2019
- Other products include:
  - Percent Tree Canopy (2011, 2016)
  - Impervious Descriptor (2001, 2004, 2006, 2008, 2011, 2013, 2016, 2019)
- Multiple interactive viewers for land cover, rangeland metrics, and advanced analysis

MRLC Interactive Viewer



Gap Analysis Project (GAP)

https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap

- United States Geological Survey (USGS)
- Data and analysis tools for species, land cover, and protected areas
  - Species range
  - Additional data (land cover, forest edge, human impact, slope, aspect, canopy cover, etc.)
- Land cover maps for U.S. that incorporate the Ecological System classification from NatureServe

Species Data

Land Cover Data

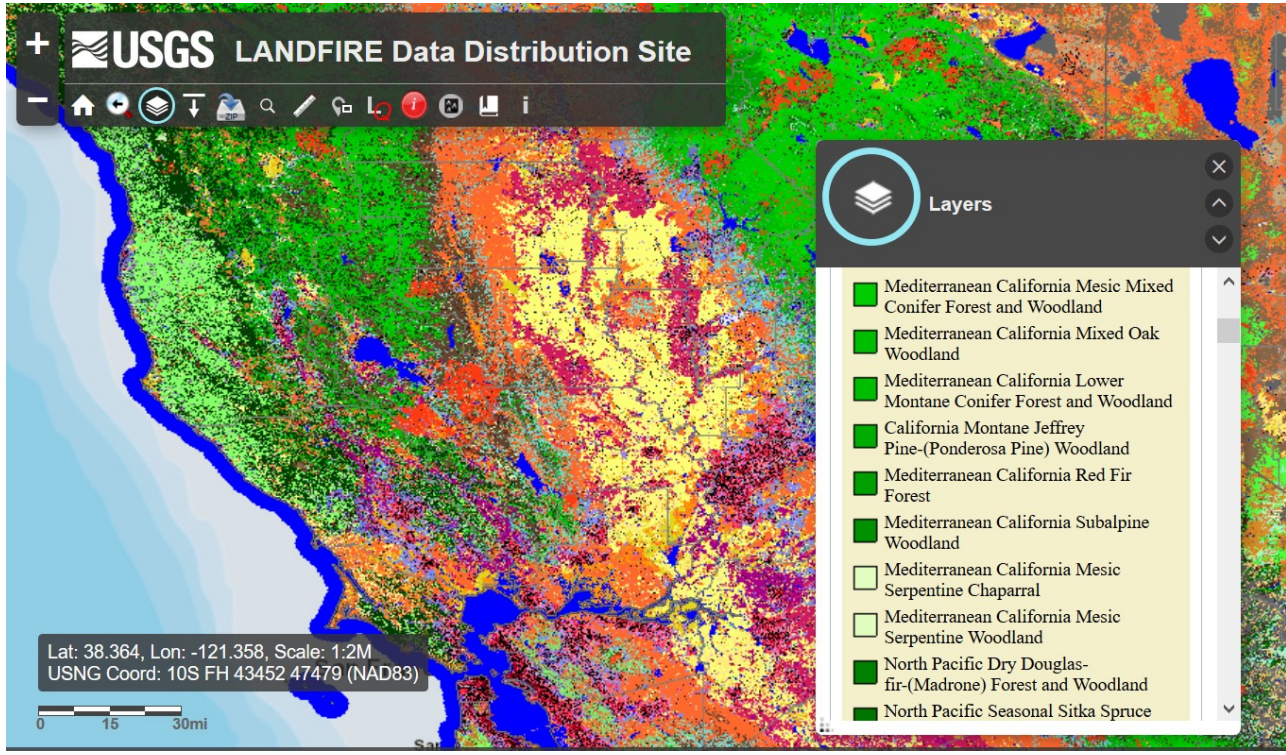
Protected Areas Data

Recent ARSET Fires training: https://appliedsciences.nasa.gov/join-mission/training/english/arset-satellite-observations-and-tools-fire-risk-detection-and

LANDFIRE

http://www.landfire.gov

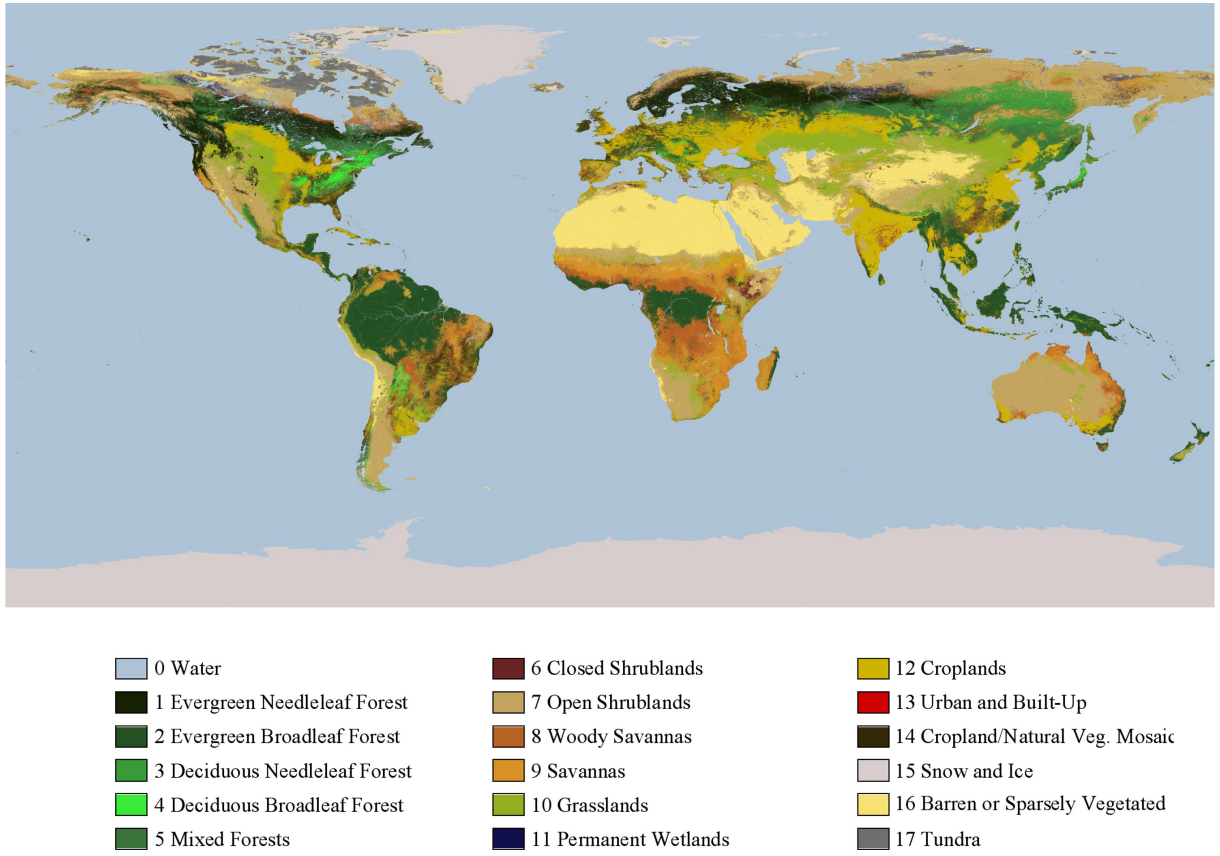
- Products:
  - Delivered at 30 m spatial resolution
  - Available from 1999-present
  - For the US only
- Vegetation Data Layers using Landsat Imagery:
  - Plant communities via NatureServe's terrestrial Ecological Systems Classification, through 2016
    - Mapped with models, field data, and Landsat
- Disturbance Data:
  - Fuel, vegetation, natural, and prescribed disturbance by type and year



Vegetation layers via LANDFIRE. Image Credit: https://www.landfire.gov/viewer

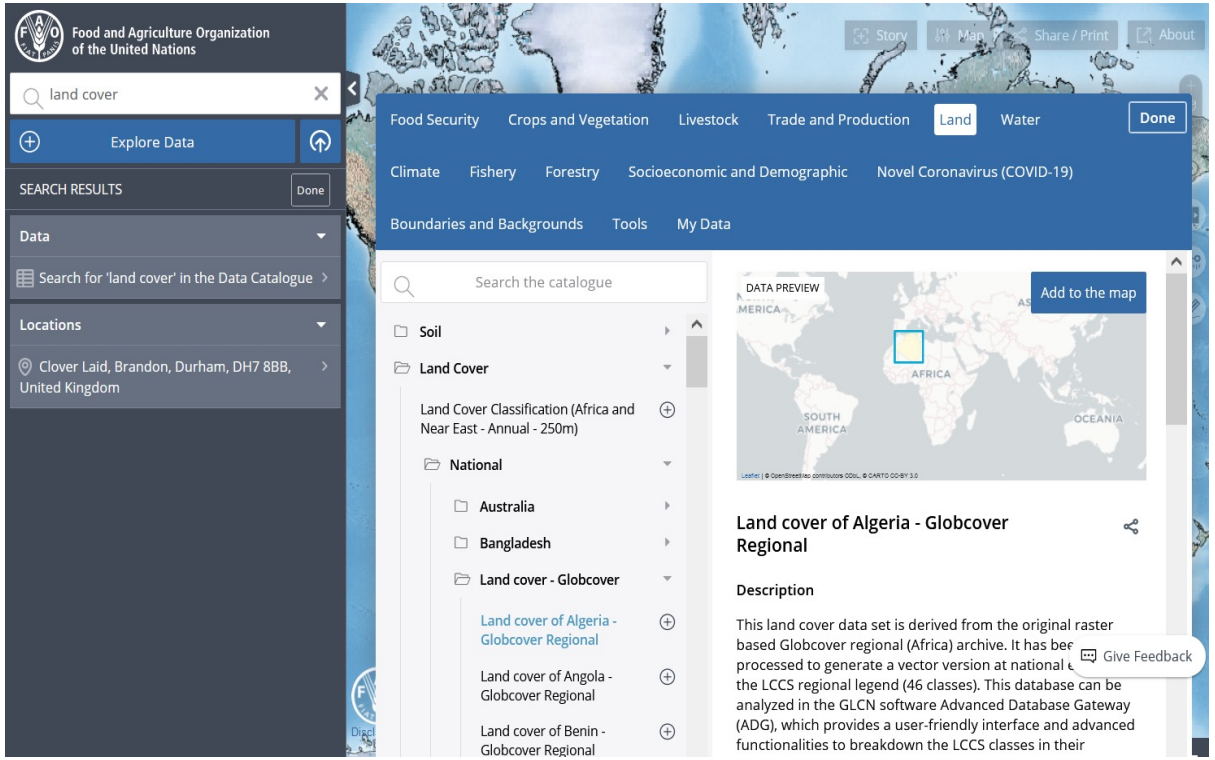
MODIS Land Cover

- Contains 5 classification schemes
  - Identifies 17 land cover classes identified by the International Geosphere Biosphere Programme, which includes 11 natural vegetation classes, 3 developed and mosaicked land classes, and 3 non-vegetated land classes
- Spatial Resolution: 500 m
- Temporal Coverage: 2001 – 2019 annually
- Download data from NASA's Earthdata Search: http://search.earthdata.nasa.gov



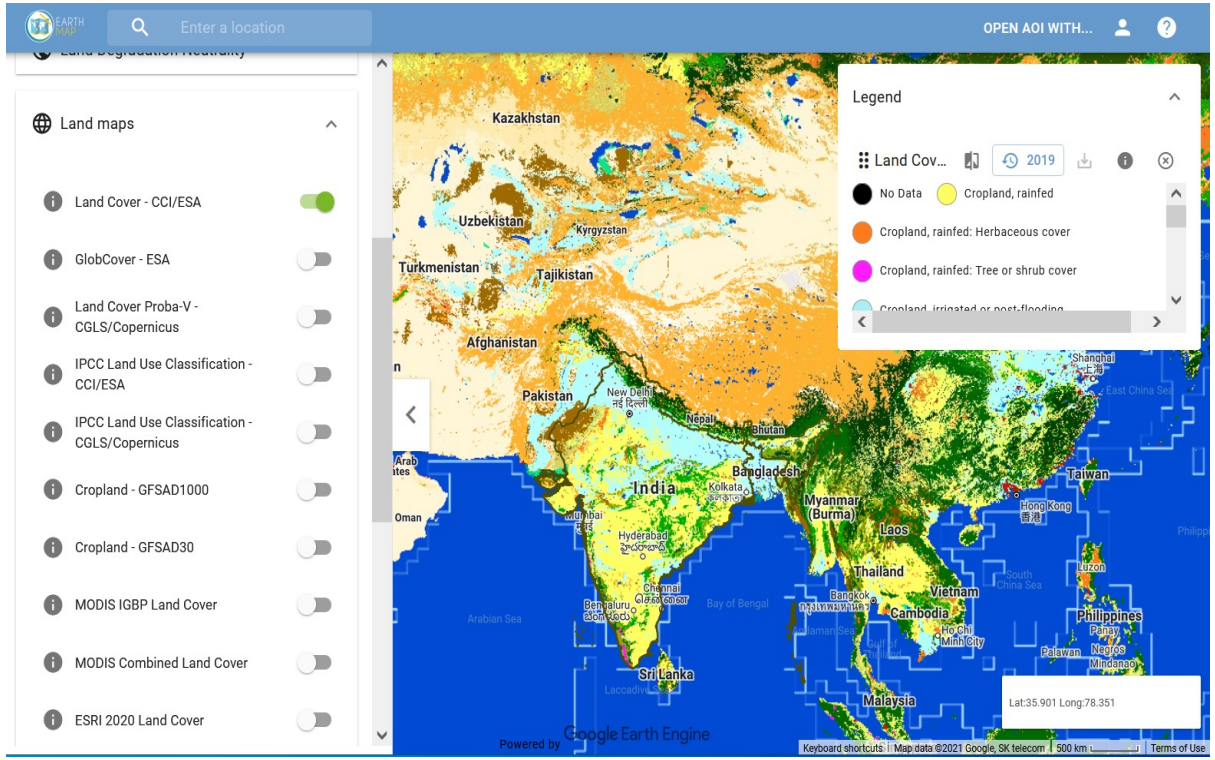
Food and Agriculture Organization

Hand-in-Hand Geospatial Platform



https://data.apps.fao.org

Earth Map

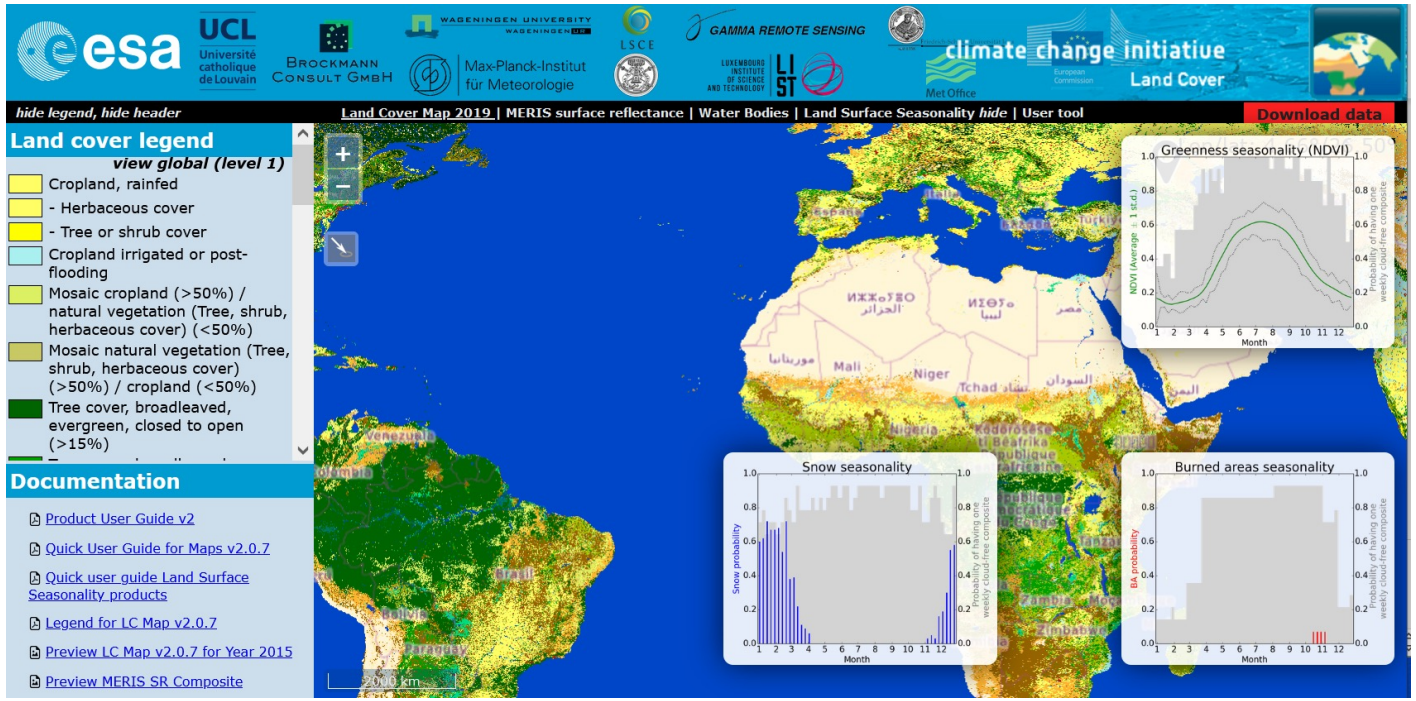


https://earthmap.org

ESA Climate Change Initiative Land Cover

http://www.esa-landcover-cci.org

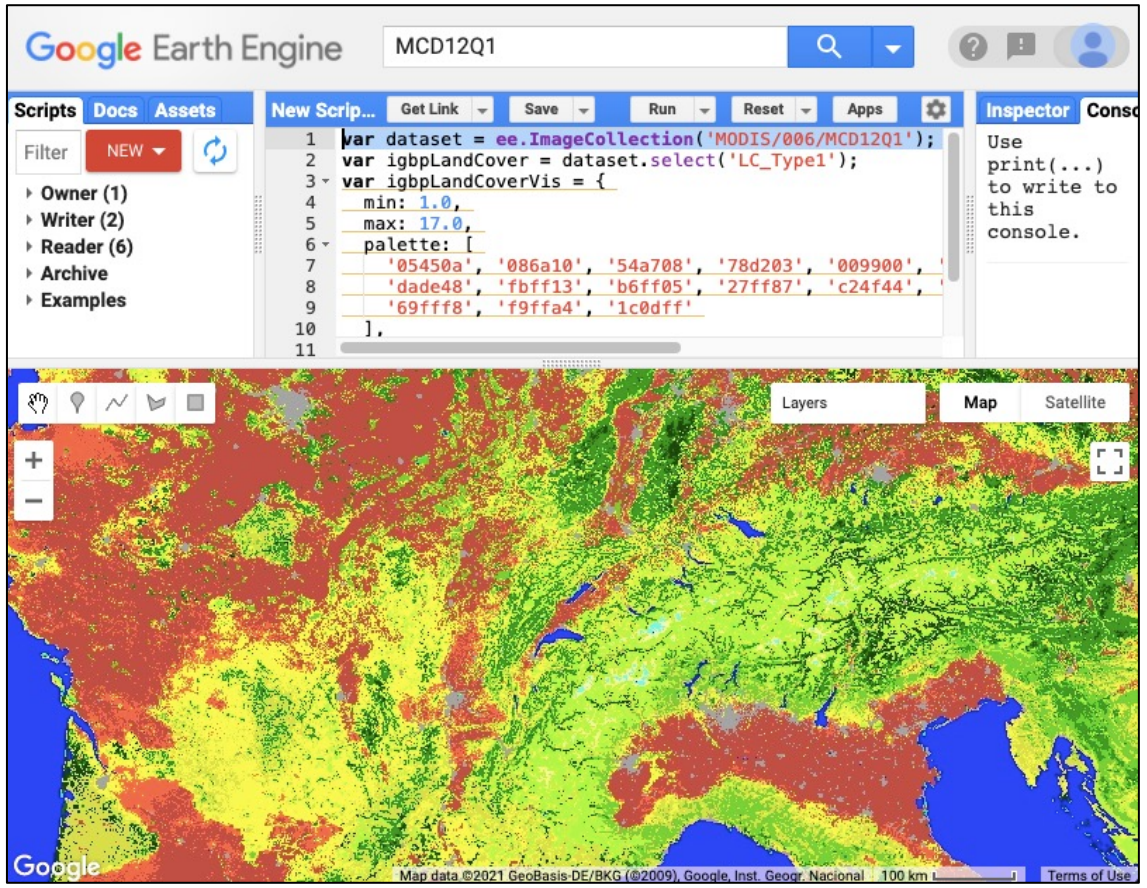
- Annual global land cover time series from 1992 – 2019
- Spatial Resolution: 300 m
- Remote Sensing Sources:
  - NOAA AVHRR
  - SPOT
  - ENVISAT
  - PROBA-V
- 22 land cover classes based on the UN Land Cover Classification System



Visualize and download with the CCI Land Cover viewer: http://maps.elie.ucl.ac.be/CCI/viewer/

Create Your Own Land Cover Map

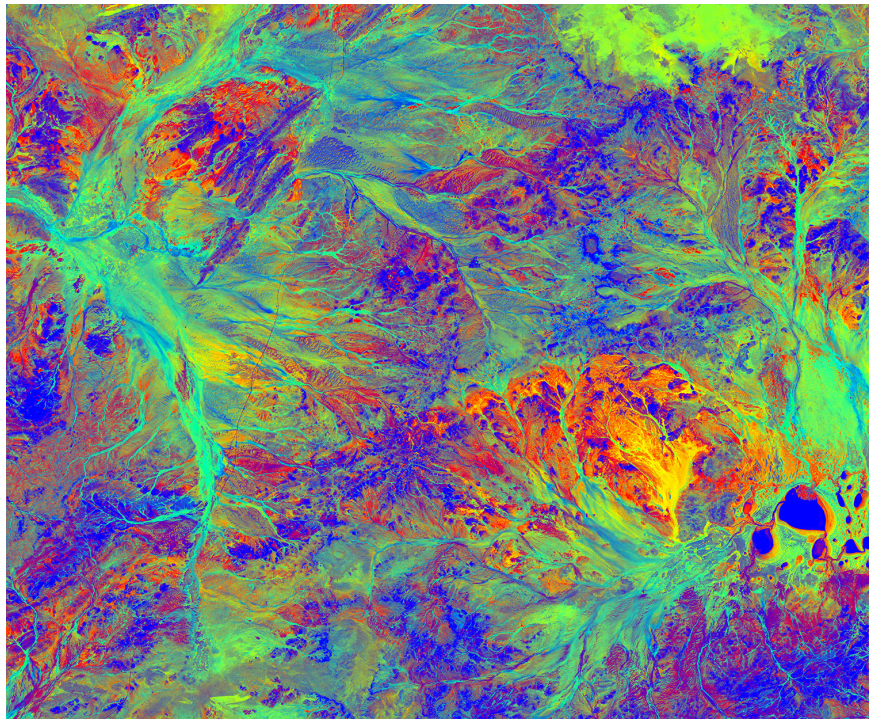
- Land cover classification is the process of grouping spectral classes and assigning them informational class names.
- Spectral Classes:
  - Groups of pixels that are uniform with respect to their pixel values in several spectral bands.
- Informational Classes:
  - Categories of interest to users of the data (like water, forest, urban, agriculture, etc.).
- Previous ARSET trainings:
  - Land Cover Classification with Satellite Imagery
  - Accuracy Assessment of a Land Cover Classification
  - Using Google Earth Engine for Land Monitoring Applications



Annual land cover. Credit: GEE Developers

Fractional Cover (FC)

- Estimation of the proportion of an area that is covered by each member of a pre-defined set of vegetation or land cover types
- Requires conducting a land cover classification
- Limitations for FC estimates in remote sensing:
  - Spatial resolution (one value per pixel or need for spectral unmixing techniques)
  - Errors in classification will propagate into FC estimates.
  - Requires hyperspectral data for vegetation species distinctions



Fractional Cover (FC) from Australia: Green (leaves, grass, and growing crops), brown (branches, dry grass or hay, and dead leaf litter), and bare ground (soil or rock). Image Credit: Digital Earth Australia

Land Surface Phenology (LSP)

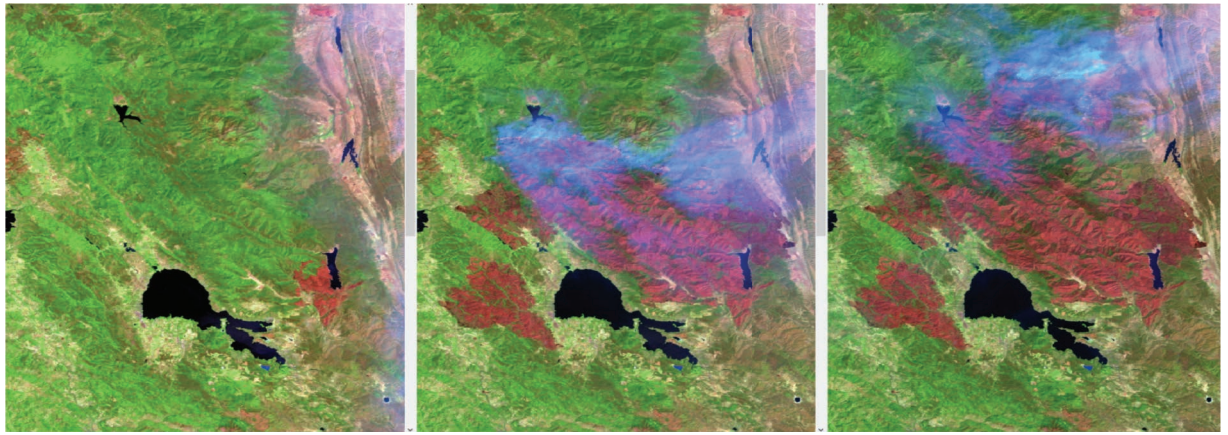
- Use of satellites and sensors to track seasonal patterns in vegetated land surfaces
  - Regular monitoring of the entire global land surface
  - Gather information on entire ecosystems: broad scale trends
  - Timing of seasonal patterns related to day length, temperature, and precipitation patterns
  - Impacts on species distributions
- Useful when linked to ground observation networks
- See previous ARSET training on Phenology: https://appliedsciences.nasa.gov/join-mission/training/english/arset-understanding-phenology-remote-sensing

Seasonal cycle of a tree, Image Credit: USGS/NPN)

Vegetation Indices

- Normalized Difference Vegetation Index (NDVI)
- Enhanced Vegetation Index (EVI)
- Normalized Difference Moisture Index (NDMI)
- Normalized Burn Ratio (NBR)
- And many more!

EVI in South America

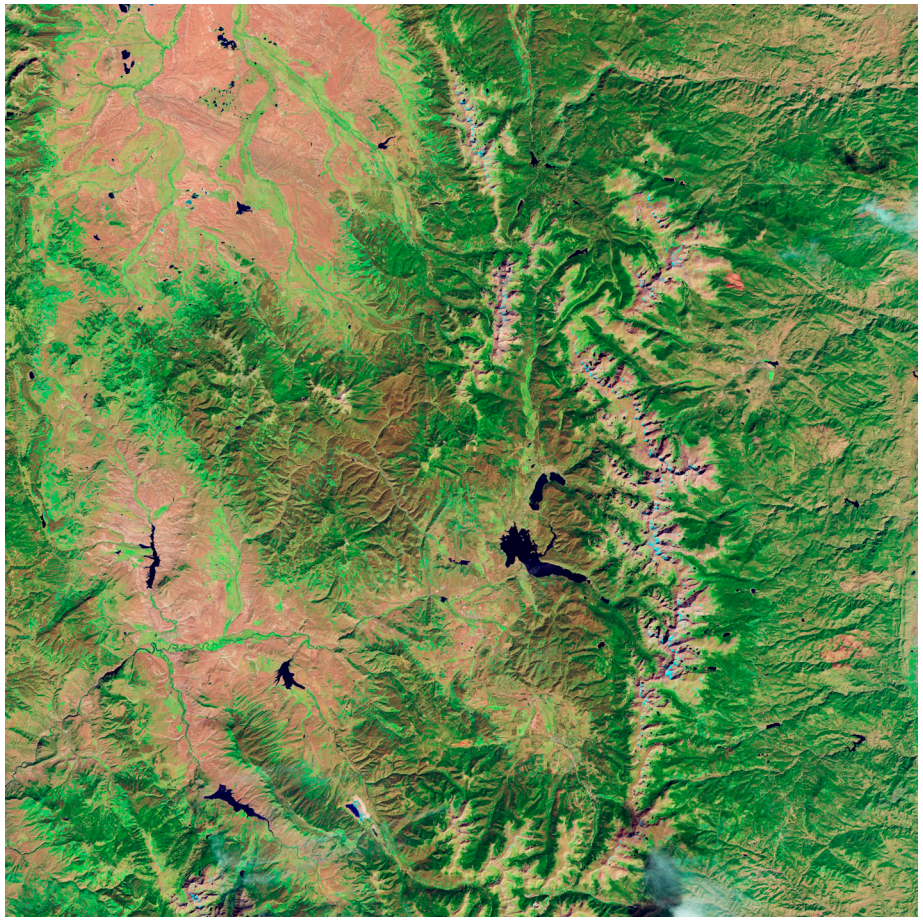


July 26 Aug 11 Aug 27

Mendocino Complex Fires, 2018

Tree Mortality/Vegetation Disturbance

- Change Detection and Time Series Analysis for:
  - Large-scale mortality or disturbance
  - Forest fragmentation and succession
  - Changes to habitat connectivity
- Tools:
  - Global Forest Watch
  - USDA Insect and Disease Survey
  - Landtrendr for Google Earth Engine
  - And many more!
- Previous ARSET trainings:
  - Google Earth Engine for Land Monitoring Applications
  - Investigating Time Series of Satellite Imagery
  - Change Detection for Land Cover Mapping

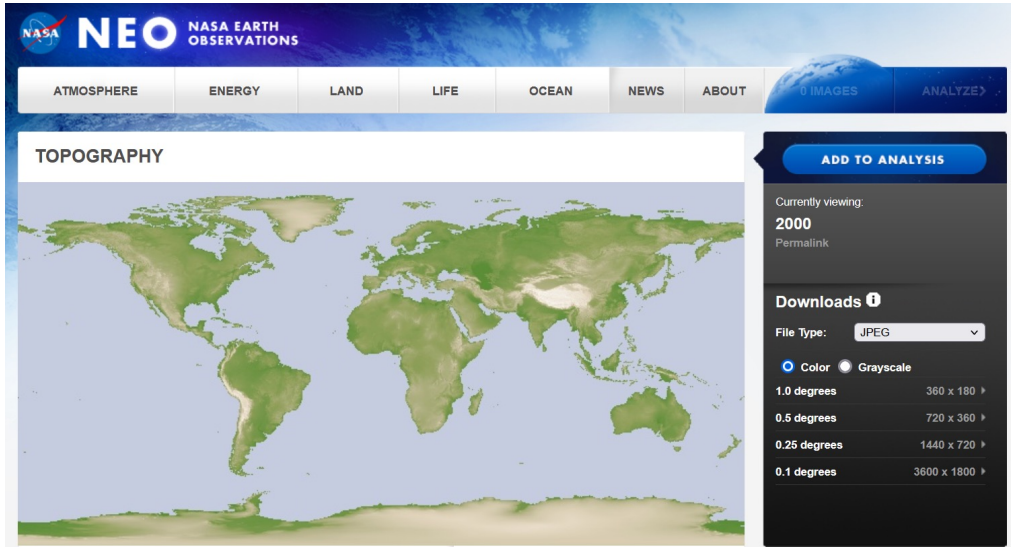


Landsat imagery of bark beetle epidemic in Lodgepole pine forests in Colorado. The top image was acquired in September 2005 and the bottom image in September 2011.

Topography

- **The Shuttle Radar Topography Mission (SRTM)**
  - Topographic (elevation) data of Earth's surface
  - SRTM used the technique of interferometry flown onboard the Space Shuttle Endeavour
  - Launched in February 2000
  - 30 m and 90 m spatial resolution
- Useful for ecosystem analysis and predictions of where specific plant and animal species exist
- Data Access: NASA Earth Observations (NEO): https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SRTM_RAMP2_TOPO

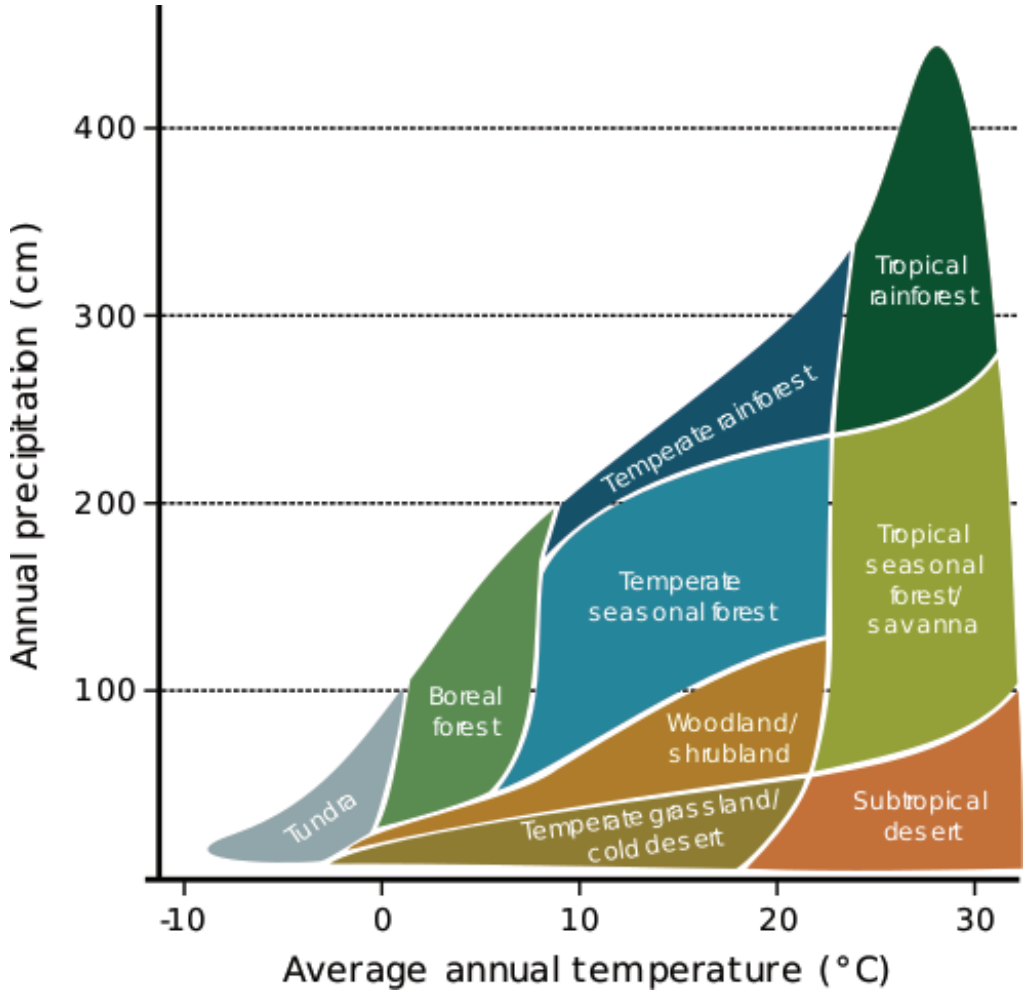
Plateau Pika, Image Credit: Smith et al, 2018



NASA NEO for topography data

Climate Variables

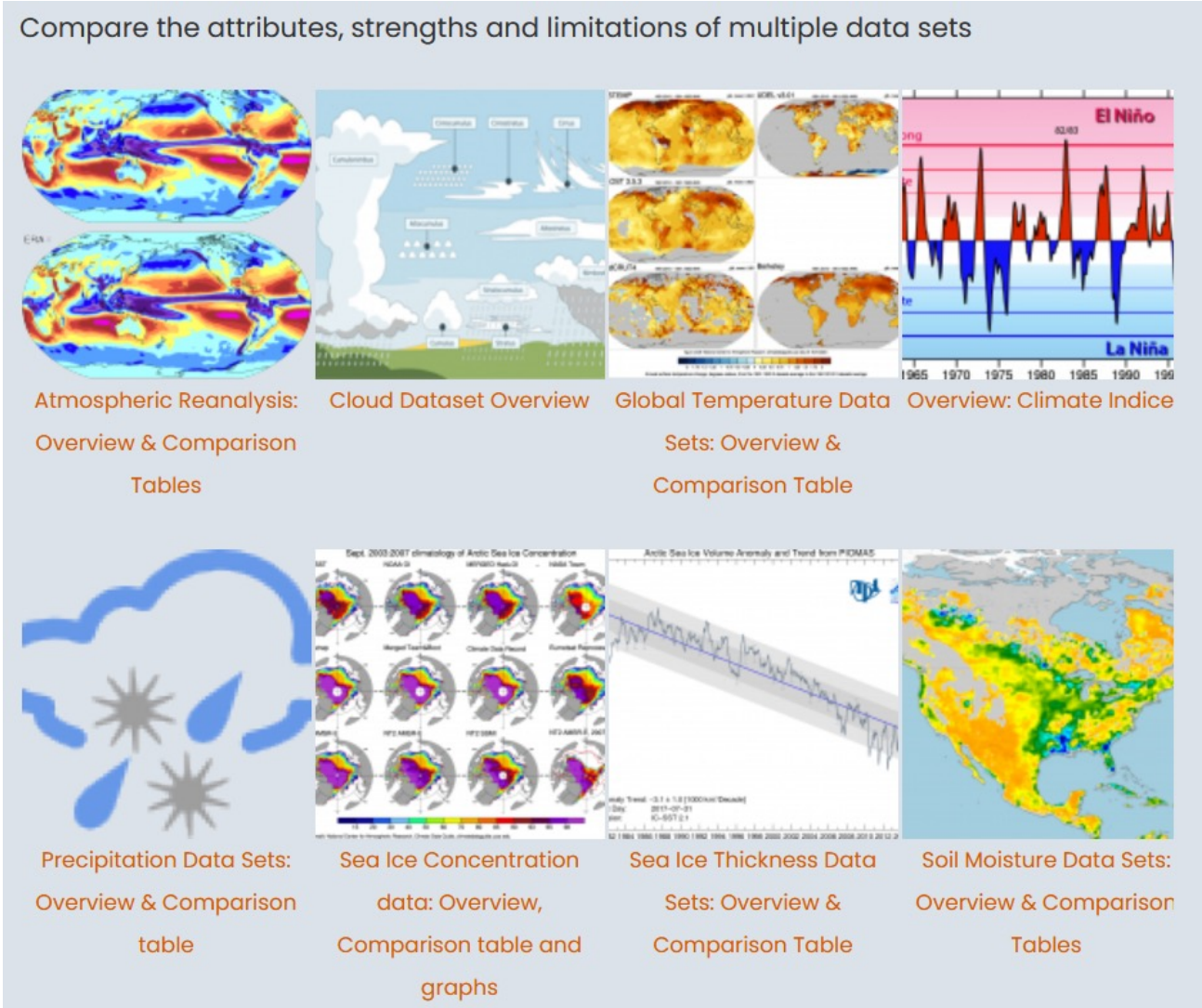
- Key drivers of ecological processes
  - Temperature
  - Precipitation
- Gridded Estimates: Derived from climate stations
  - Location matters
  - Increased uncertainty in data sparse regions
- Scenario Modeling: Use of climate change models for predicting future habitat



Whittaker's biome classification; image from Wikipedia, Garrabou et al. 2009 Global Change Biology

Gridded Climate Data Resources

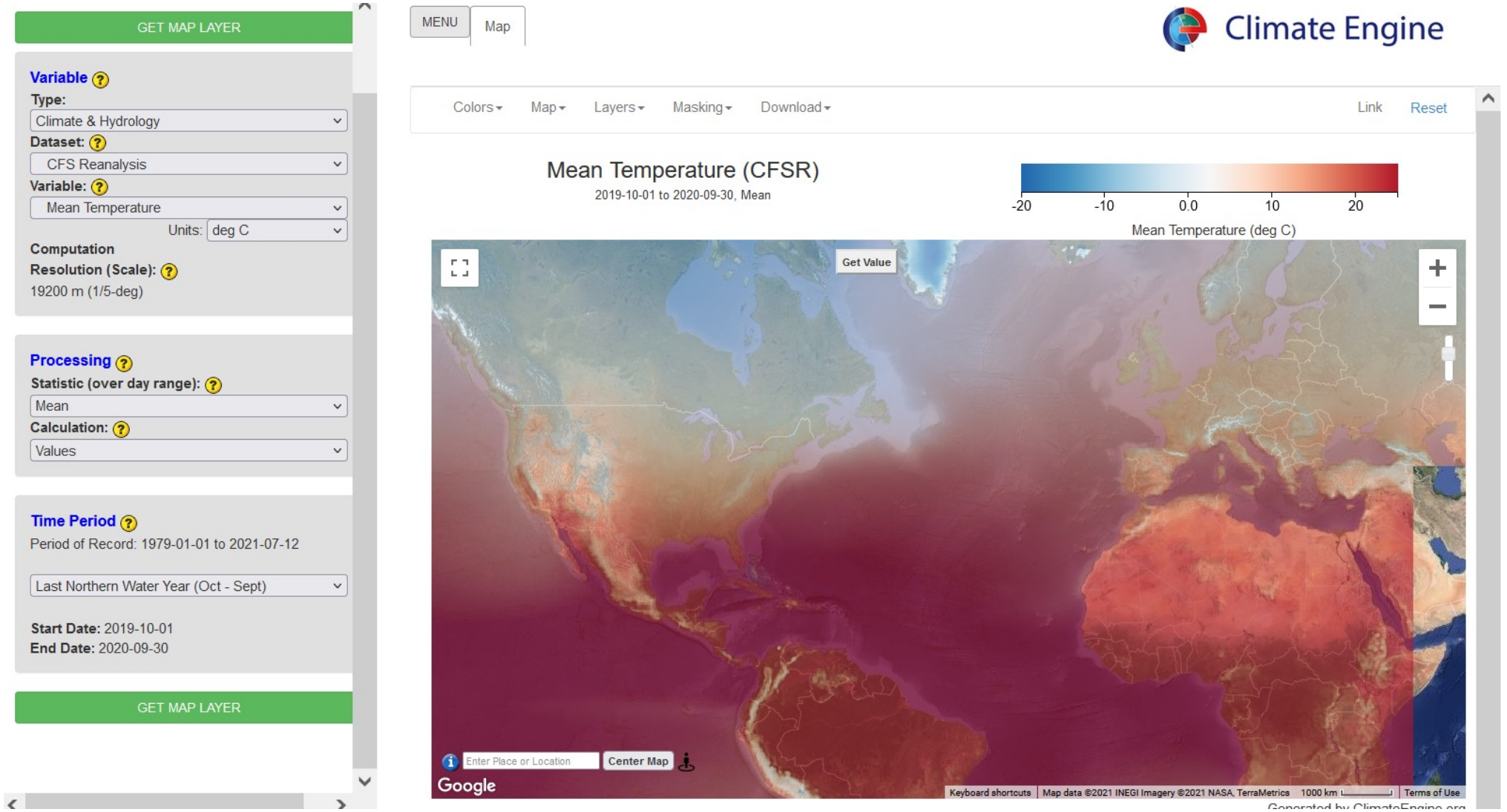
- NOAA Physical Sciences Laboratory
  - https://psl.noaa.gov/data/gridded/
- NCAR Climate Data Guide
  - https://climatedataguide.ucar.edu/
- PRISM Climate Data
  - https://prism.oregonstate.edu/
- gridMET
  - http://www.climatologylab.org/gridmet.html



NCAR Climate Data Guide

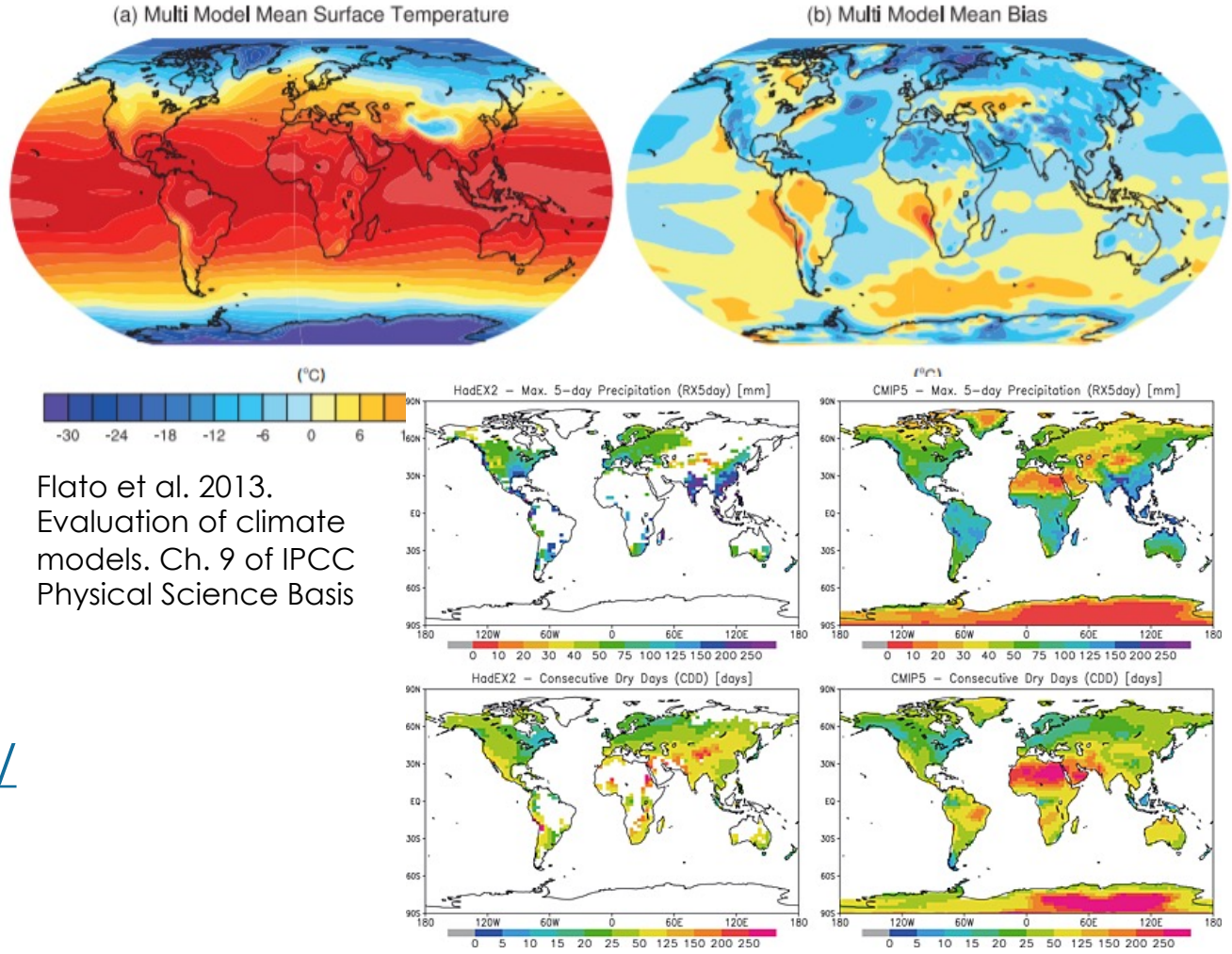
Gridded Climate Data with Climate Engine

http://climateengine.org/



Future Scenarios: Global Climate Models

- Modeling future species distributions can be challenging.
- The use of multiple climate models and downscaling techniques is often necessary.
- Uncertainties across models and variables can propagate.
- See previous ARSET training on Scenario Modeling (Part 2 in particular) for more information:
  - https://appliedsciences.nasa.gov/join-mission/training/english/arset-introduction-remote-sensing-scenario-based-ecoforecasting

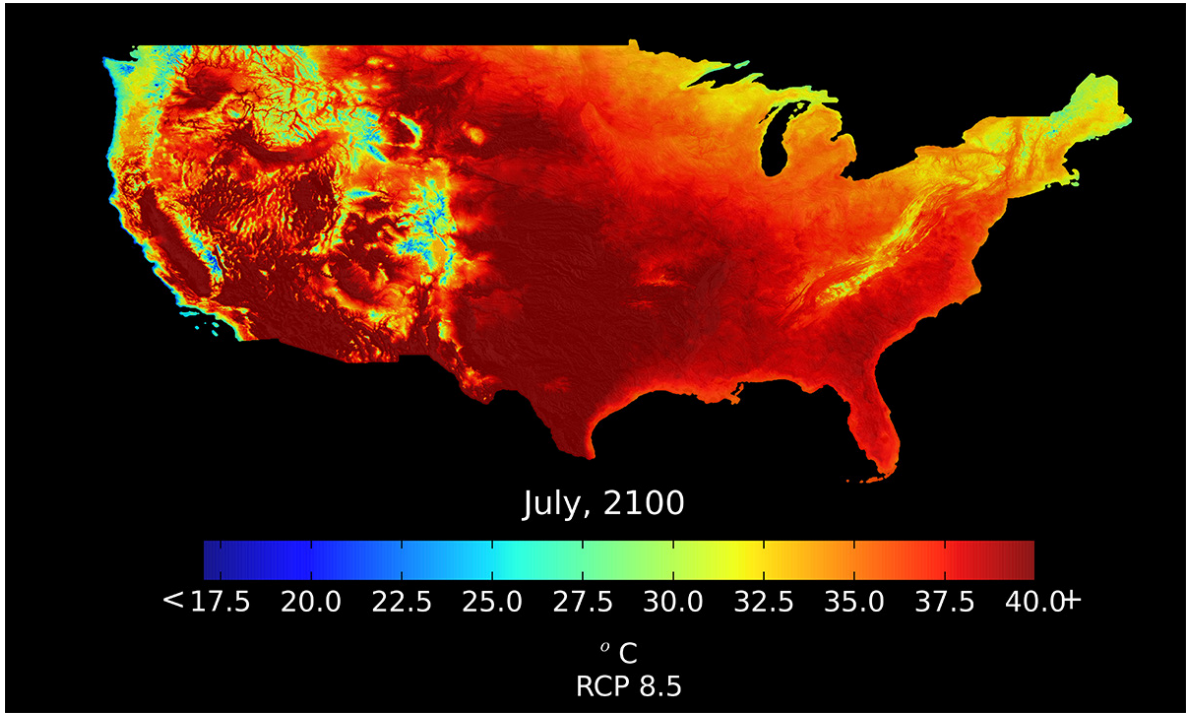


Flato et al. 2013. Evaluation of climate models. Ch. 9 of IPCC Physical Science Basis

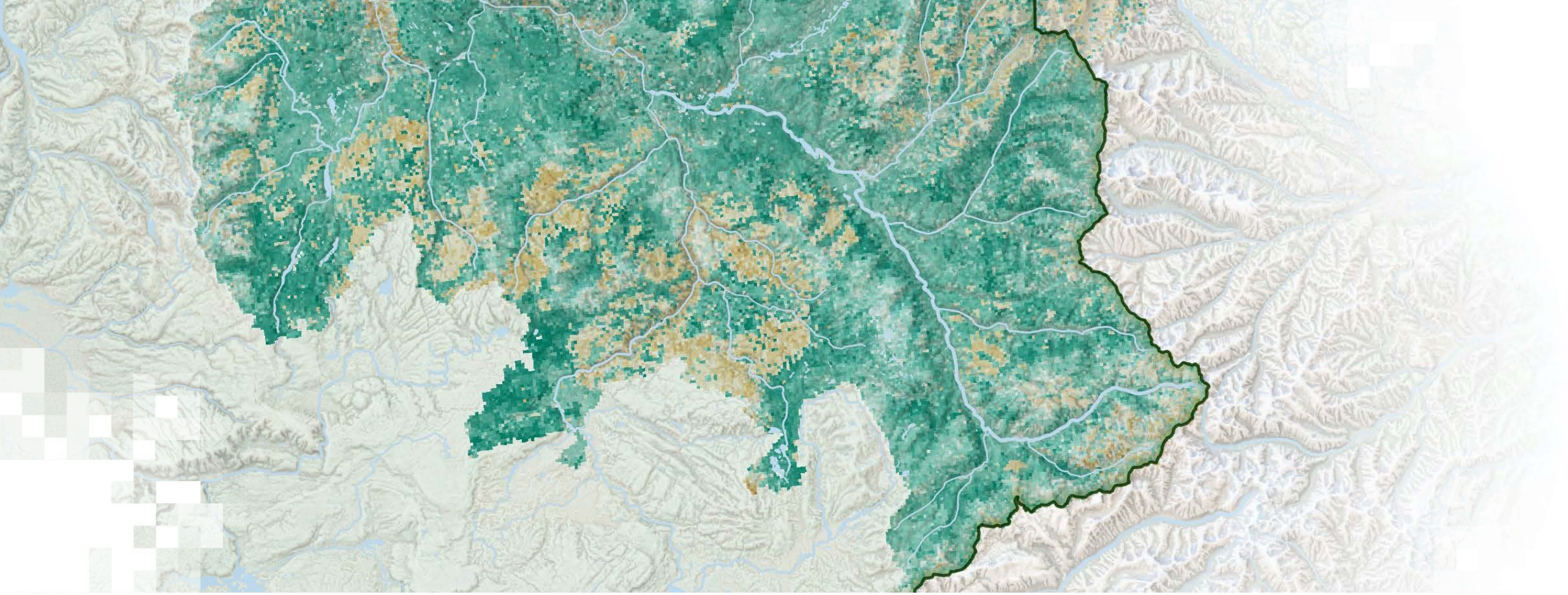
Sillmann et al. 2013 Journal of Geophysical Research: Atmospheres

Climate Models with the NASA Earth Exchange (NEX)

- Downscaled Climate Projections (NEX-DCP30)
  - Conterminous United States
  - Coupled Model Intercomparison Project Phase 5 (CMIP5)
  - Four greenhouse gas emissions scenarios known as Representative Concentration Pathways (RCPs)
  - Projections from 33 models and ensemble statistics
  - https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-dcp30

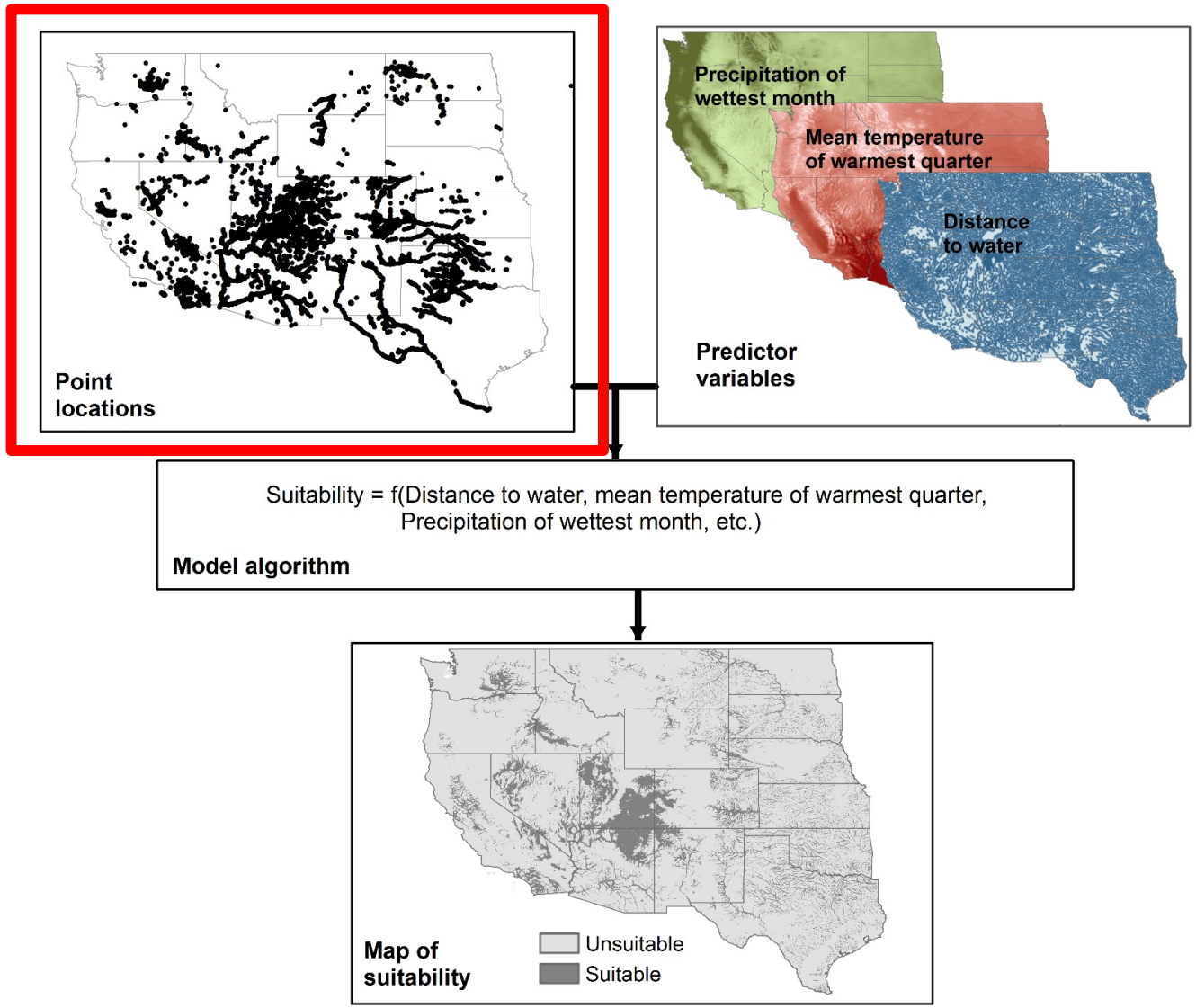


The NEX-DCP30 predicted monthly mean of the daily maximum near-surface air temperature for July 2100, downscaled to 800 meters. The image data is based on the Representative Concentration Pathway scenario 8.5 (RCP8.5). Image Credit: NEX



Occurrence Data

Occurrence Data

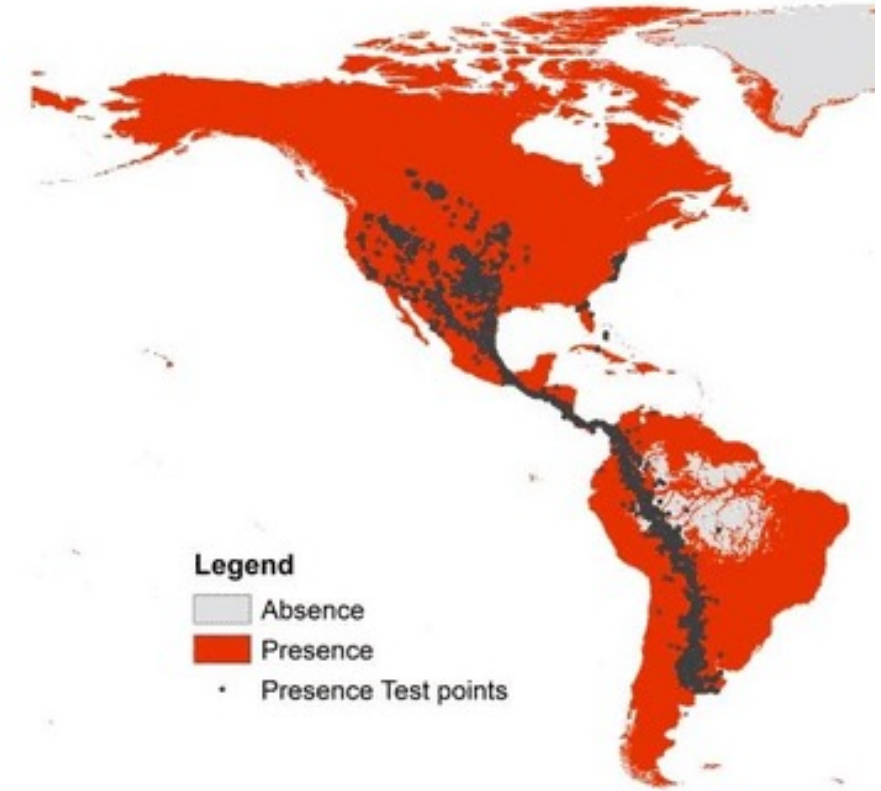


Jarnevich et al, 2015

Occurrence

- Ground-based observations
- Presence only OR presence and absence

Puma presence locations in the San Francisco Bay Area from GBIF



Presence/Absence map for Peregrine Falcon. Image Credit: Sriram and Huettmann, 2017

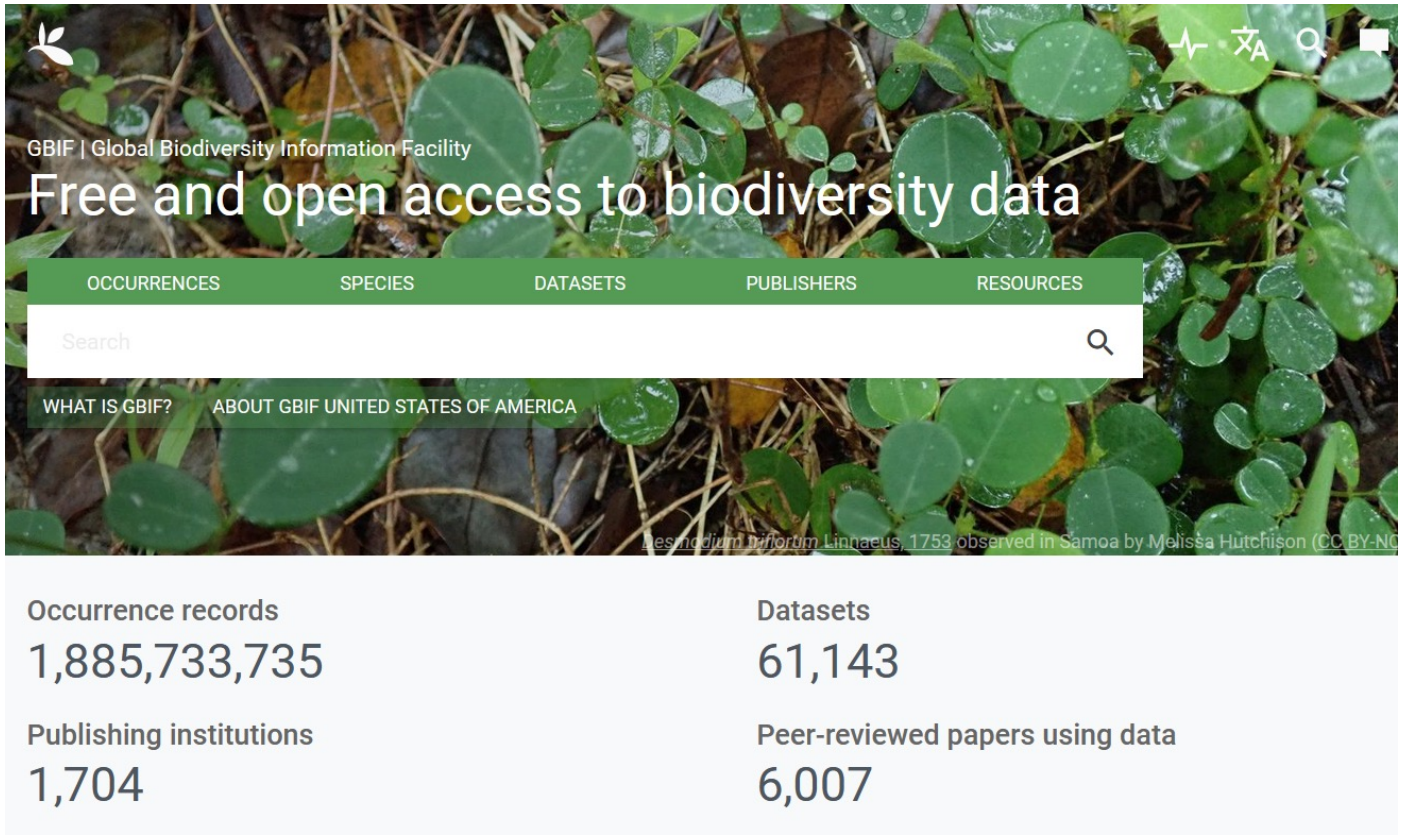
Absence Data

- A species may be classified as 'absent' for a number of reasons, but this does not necessarily denote the absence of suitable conditions.
  - The environment is truly unsuitable for the species. **True absence**
  - The species could not be detected, even though it was present. **False absence**
  - The species was absent, even though the environment is suitable (e.g., due to dispersal limitation or metapopulation dynamics). **False absence**
- Resource on pseudo-absence: Massin et al, 2012

Global Biodiversity Information Facility (GBIF)

https://www.gbif.org

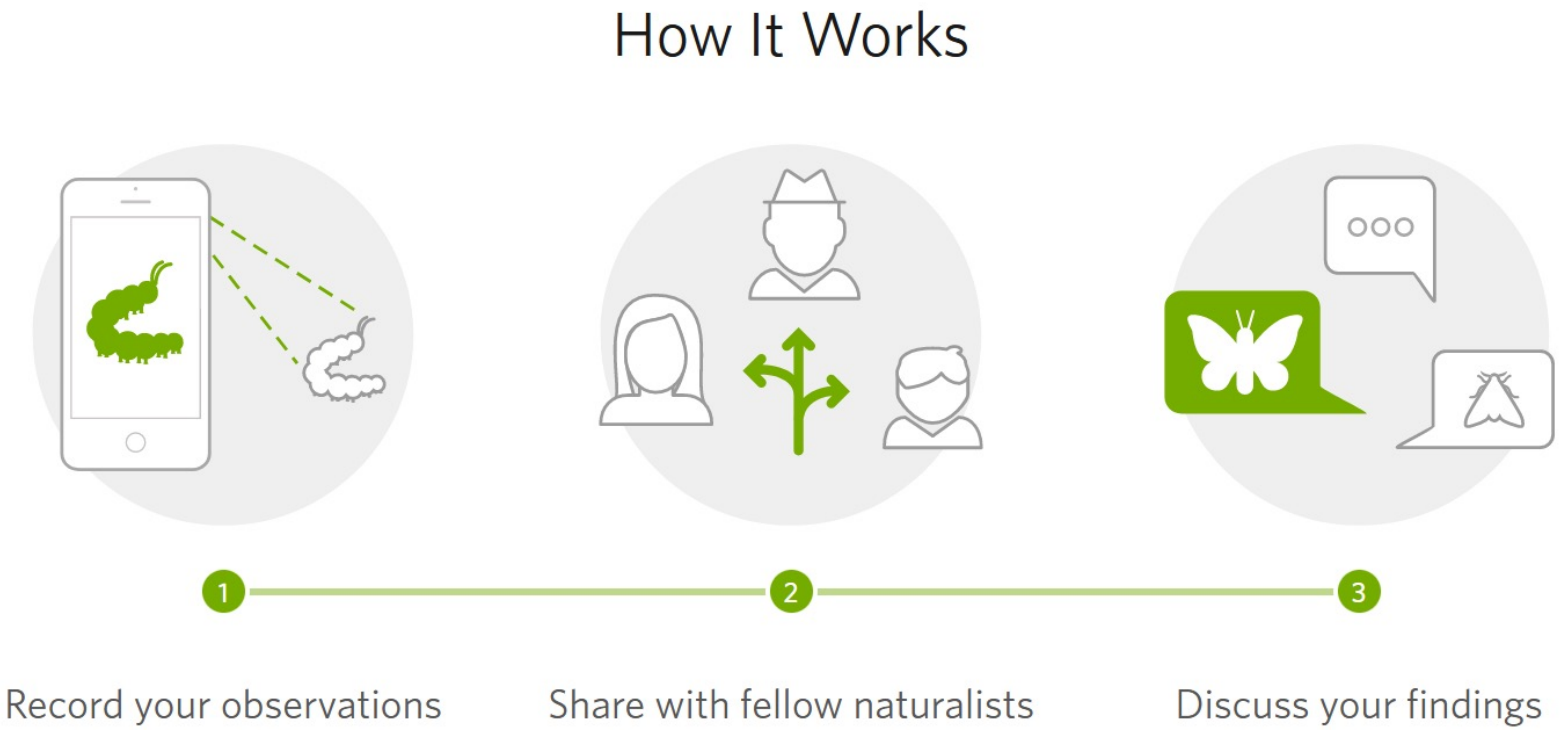
- International network and research infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth
- Nearly 1.5 billon occurrence records
- 60% of all named species
- Data standards and open access



iNaturalist

https://www.inaturalist.org

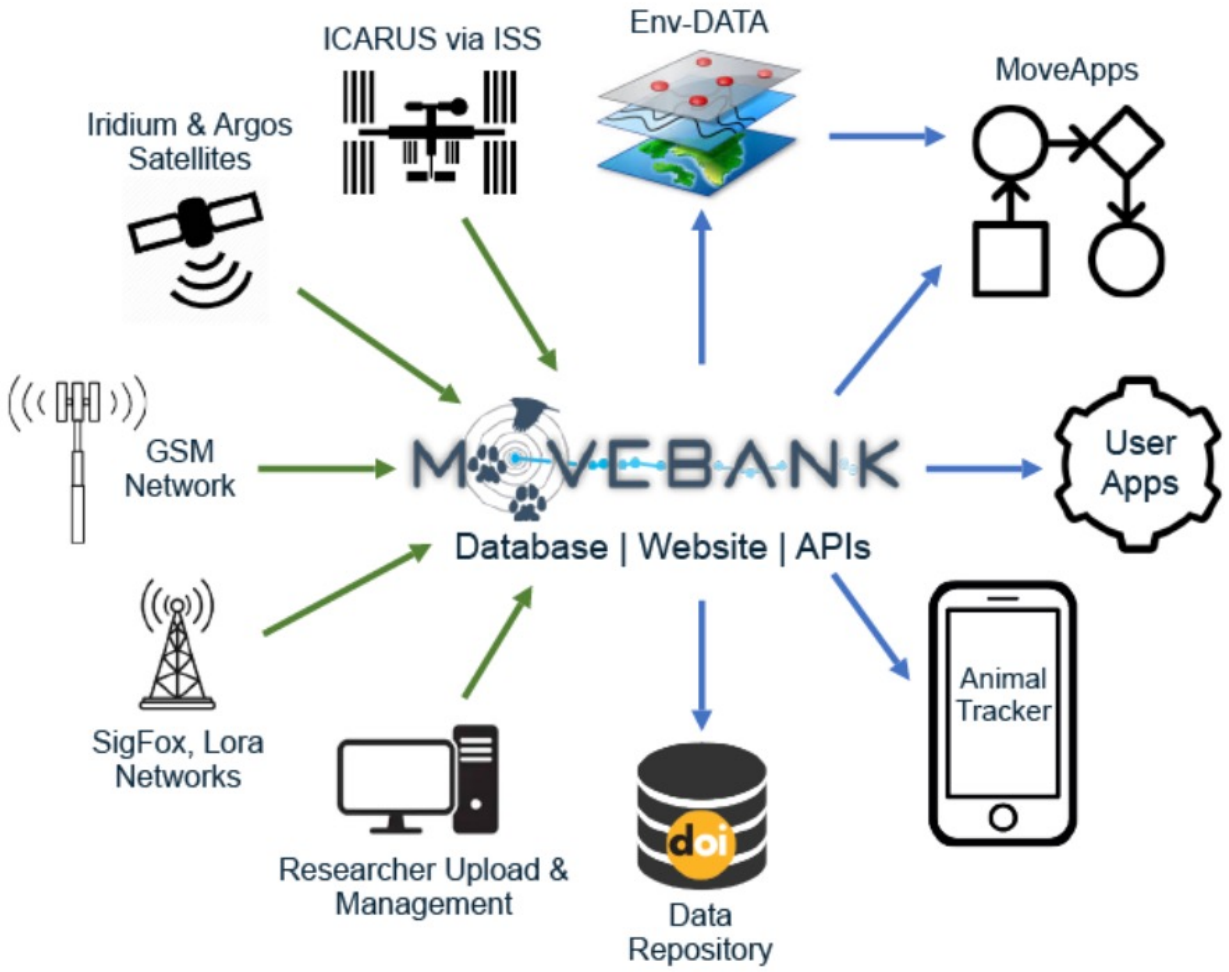
- Citizen science smartphone application for recording and sharing species information
- Connect with other observers
- Contribute to a specific project
- Hold events for field campaigns
- Share data with GBIF



Movebank

https://www.movebank.org/cms/movebank-main

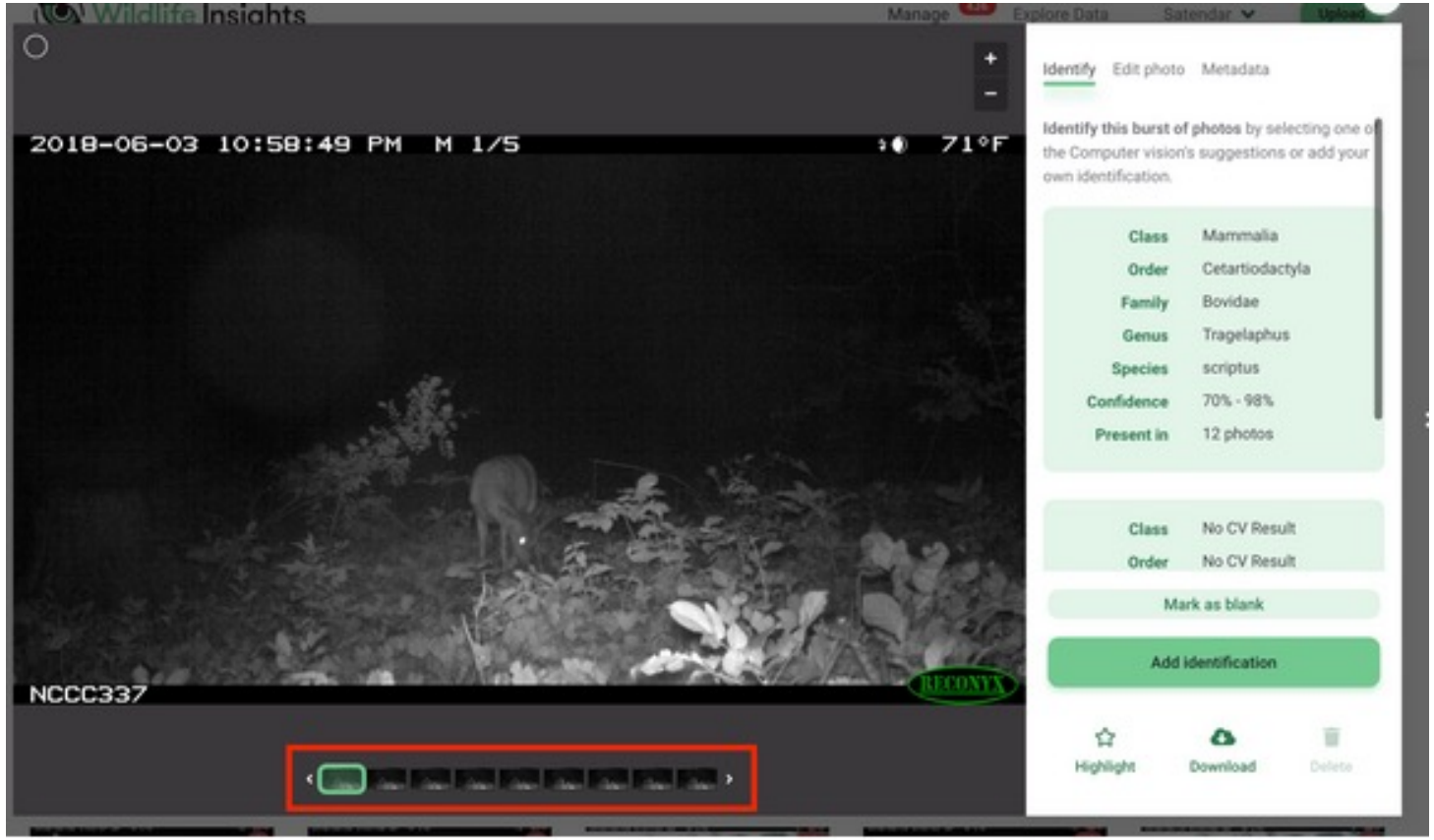
- Online platform that helps researchers and wildlife managers worldwide manage, share, analyze, and archive animal movement data
  - Archive animal movement
  - Enable collaborations
  - Help scientists address new questions
  - Promote open access



Wildlife Insights

https://www.wildlifeinsights.org

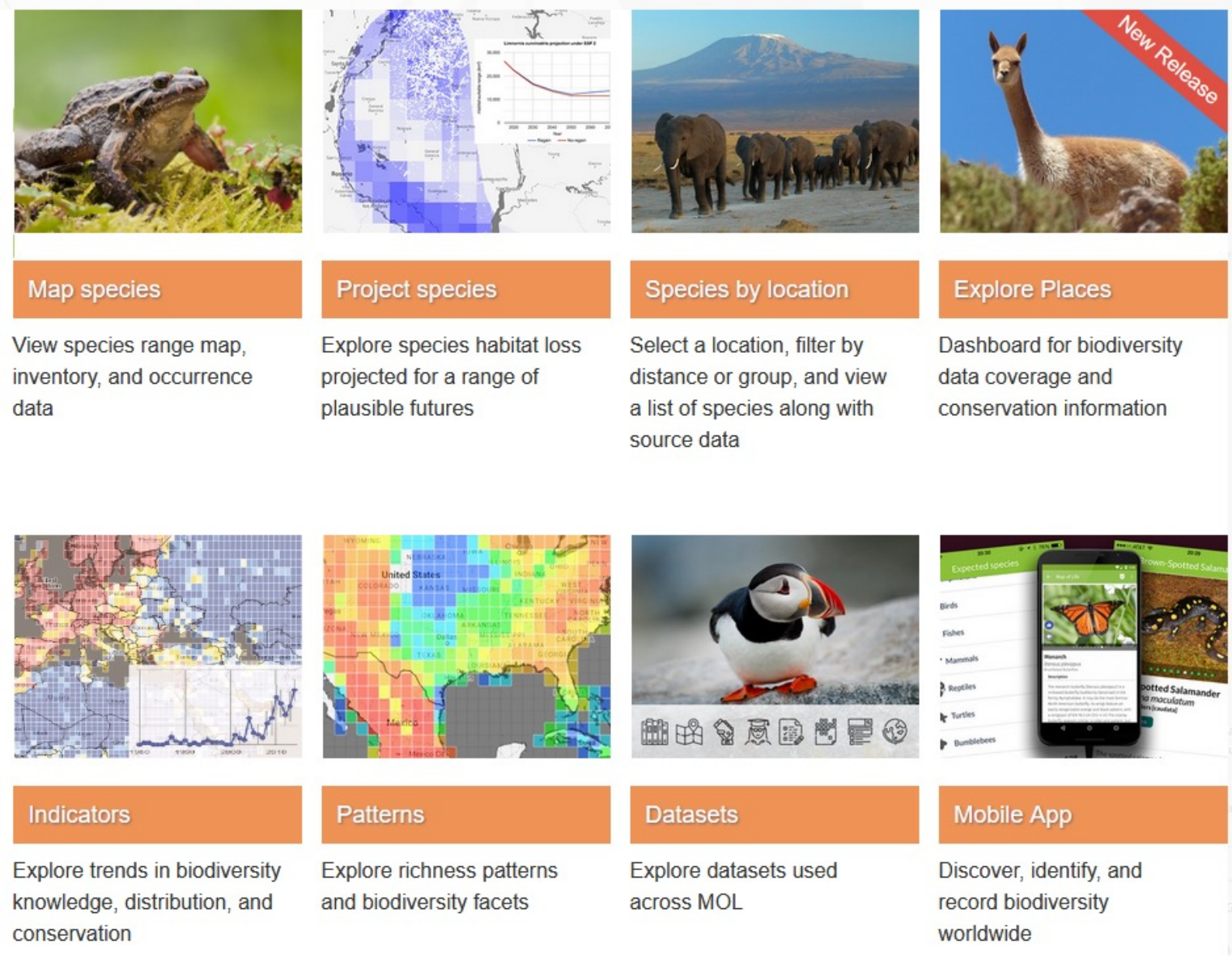
- Collection, dissemination, and analysis of camera trap data globally
- Combines field and sensor expertise, cutting edge technology and advanced analytics to enable people everywhere to share wildlife data and better manage wildlife populations
- Upload images to website for species identification with artificial intelligence



Map of Life (MOL)

https://mol.org

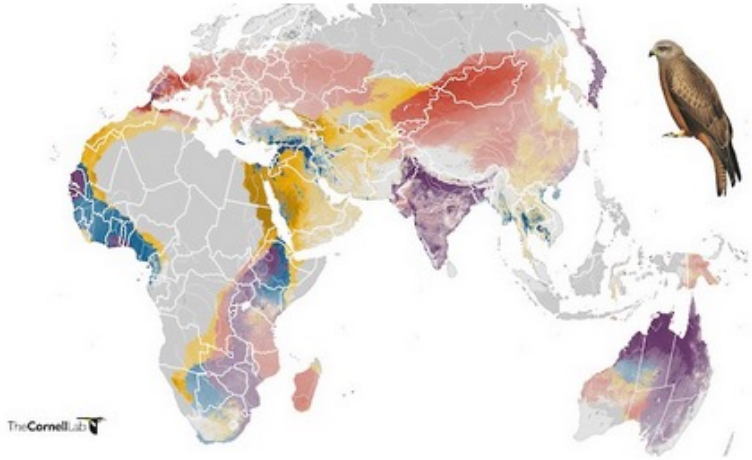
- Provides species range information and species lists for any geographic area
- Multiple tools for exploring species habitat and trends in biodiversity
- Mobile app for discovering, identifying, and recording biodiversity



eBird

https://ebird.org/home

- Gather and share bird information for science, conservation, and education
- Mange lists, photos, and recordings
- Real-time maps of species distributions
- Species alerts



eBird Status and Trends

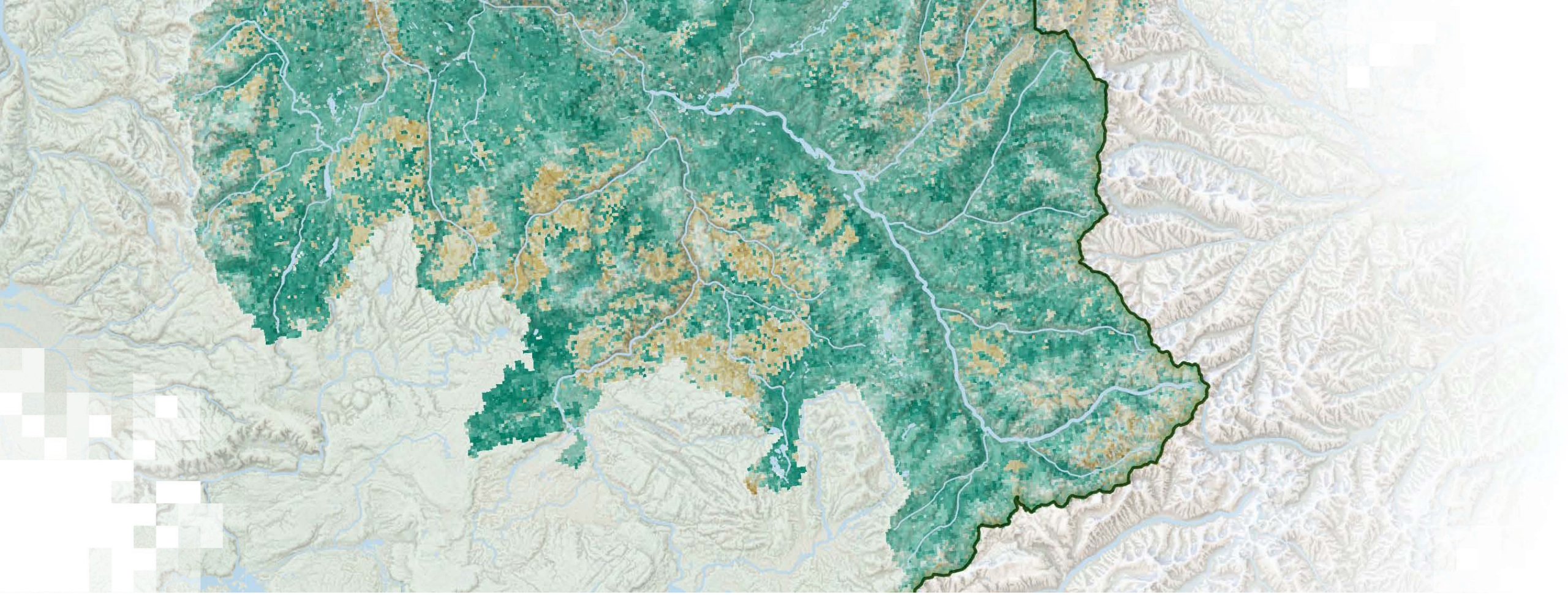
Use eBird data and tools

Research and conservation

Early Detection and Distribution Mapping System (EDDMaps)

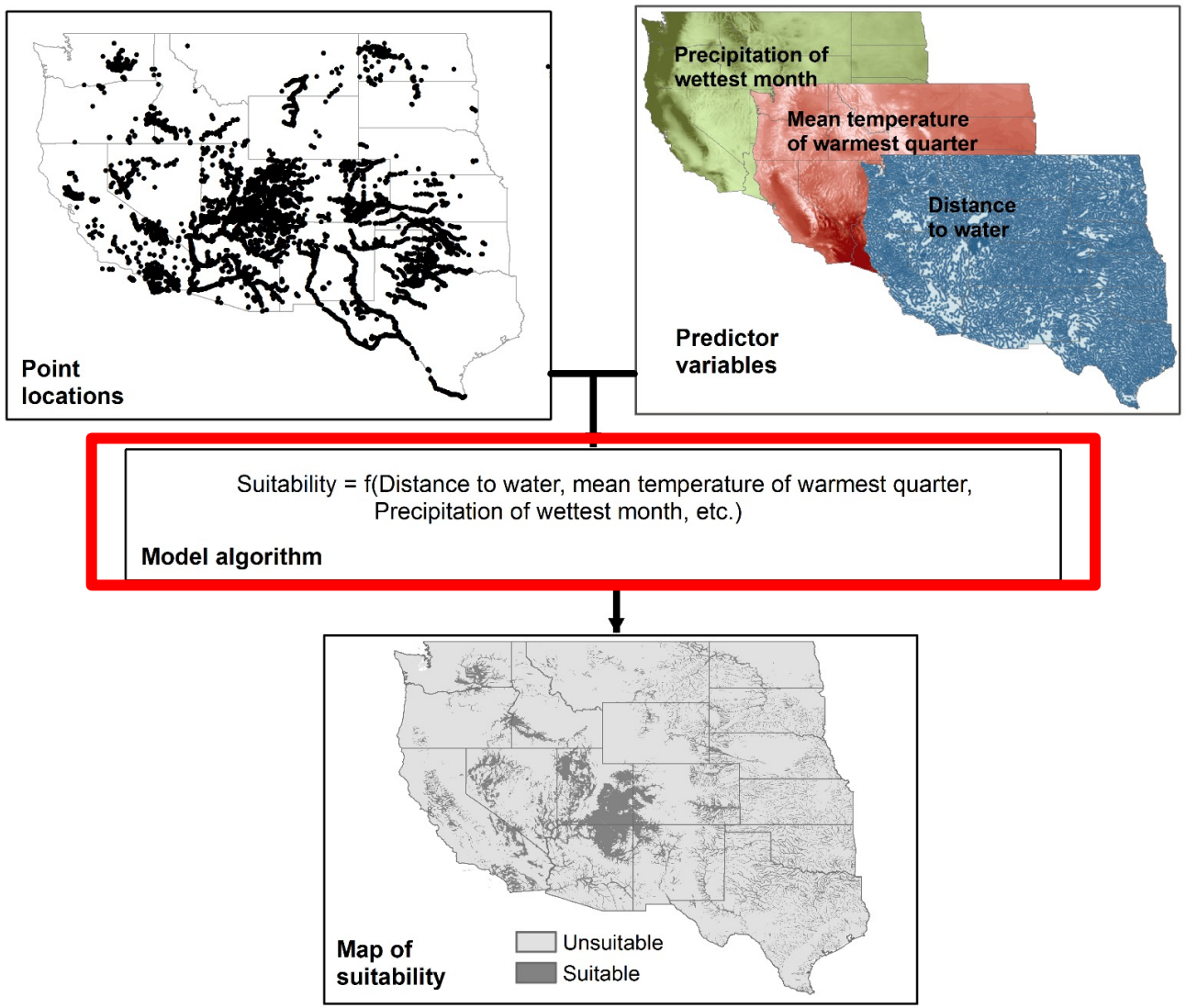
https://www.eddmaps.org

- Web-based mapping system for invasive species and pest distribution
- US and Canada
  - Interactive distribution maps
- Aggregates data from other databases and volunteer observations
- All data freely available



Methods and Models

Methods and Models



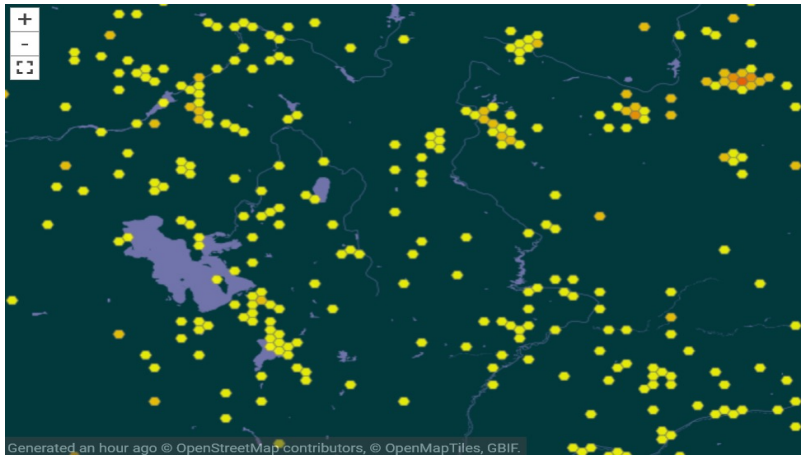
- Fit mathematical functions, describing species distributions in environmental space, for each of *n* environmental variables
- Then generate predictive maps of species distributions in geographic space using these functions

Jarnevich et al, 2015

Modeling Success: Key Factors

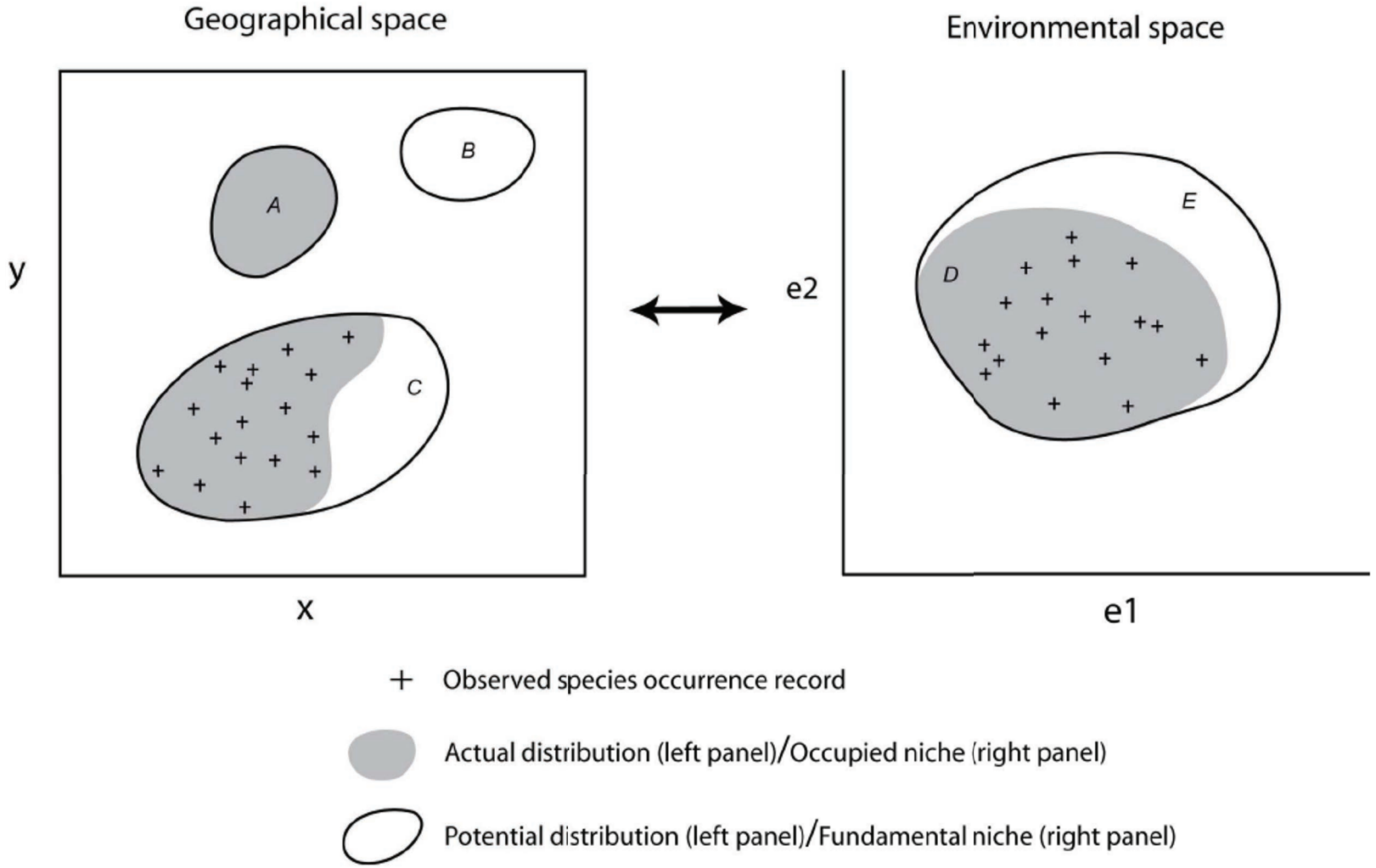
- **Input Data:** availability of sufficient and reliable species data
- **Equilibrium:** a species occurs in all suitable areas, while being absent from all unsuitable areas
  - Dependent on biotic interactions and dispersal ability
- **Sampling Adequacy:** The extent to which the observed occurrence records provide a sample of the environmental space
- **Model:** Type, complexity, and accuracy of the models used

Bromus Tectorum (cheatgrass) from iNaturalist



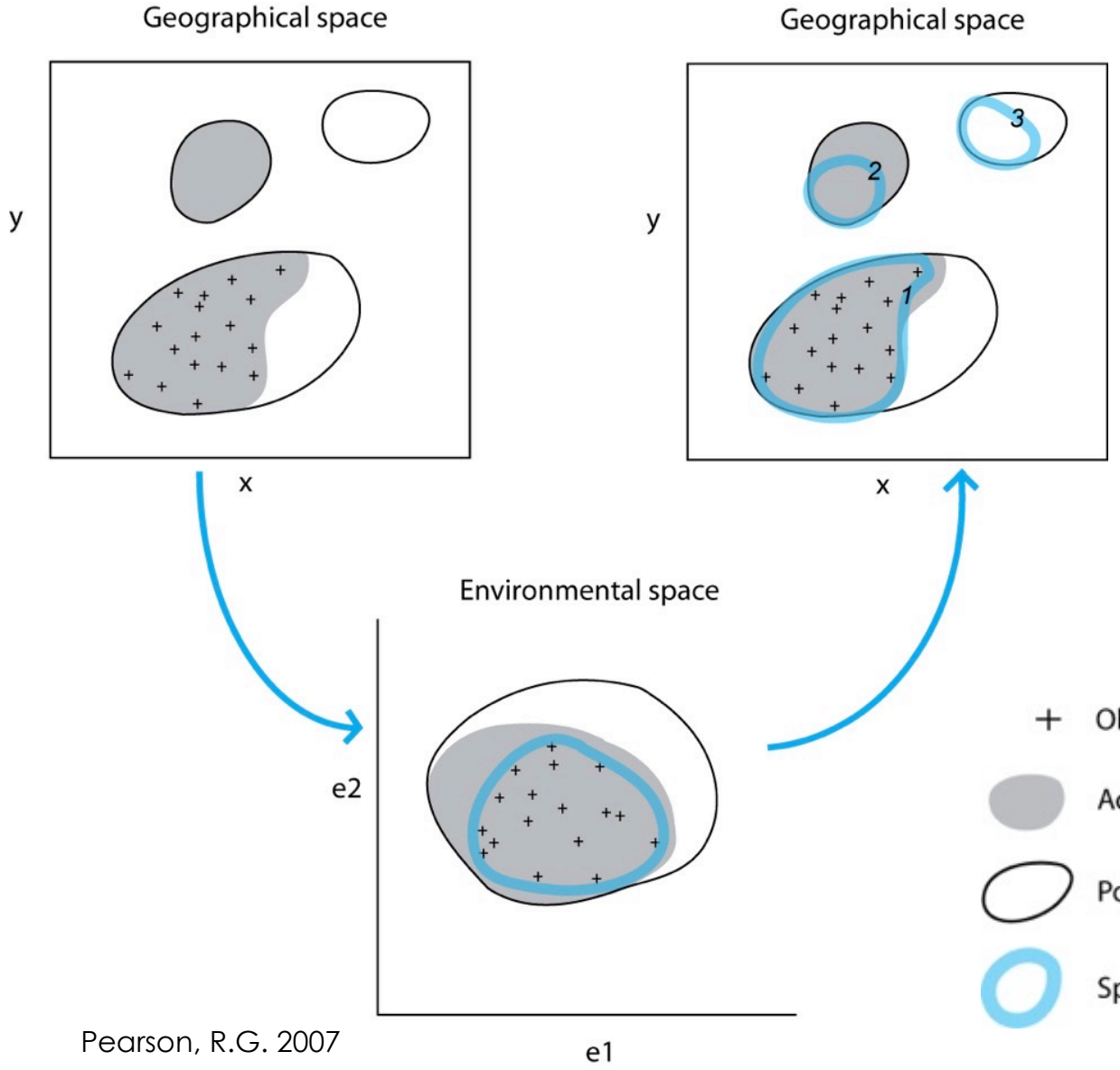
Bromus tectorum (cheatgrass) sampling from GBIF, near Salt Lake City Nevada.

Modeling Approach



Pearson, R.G. 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

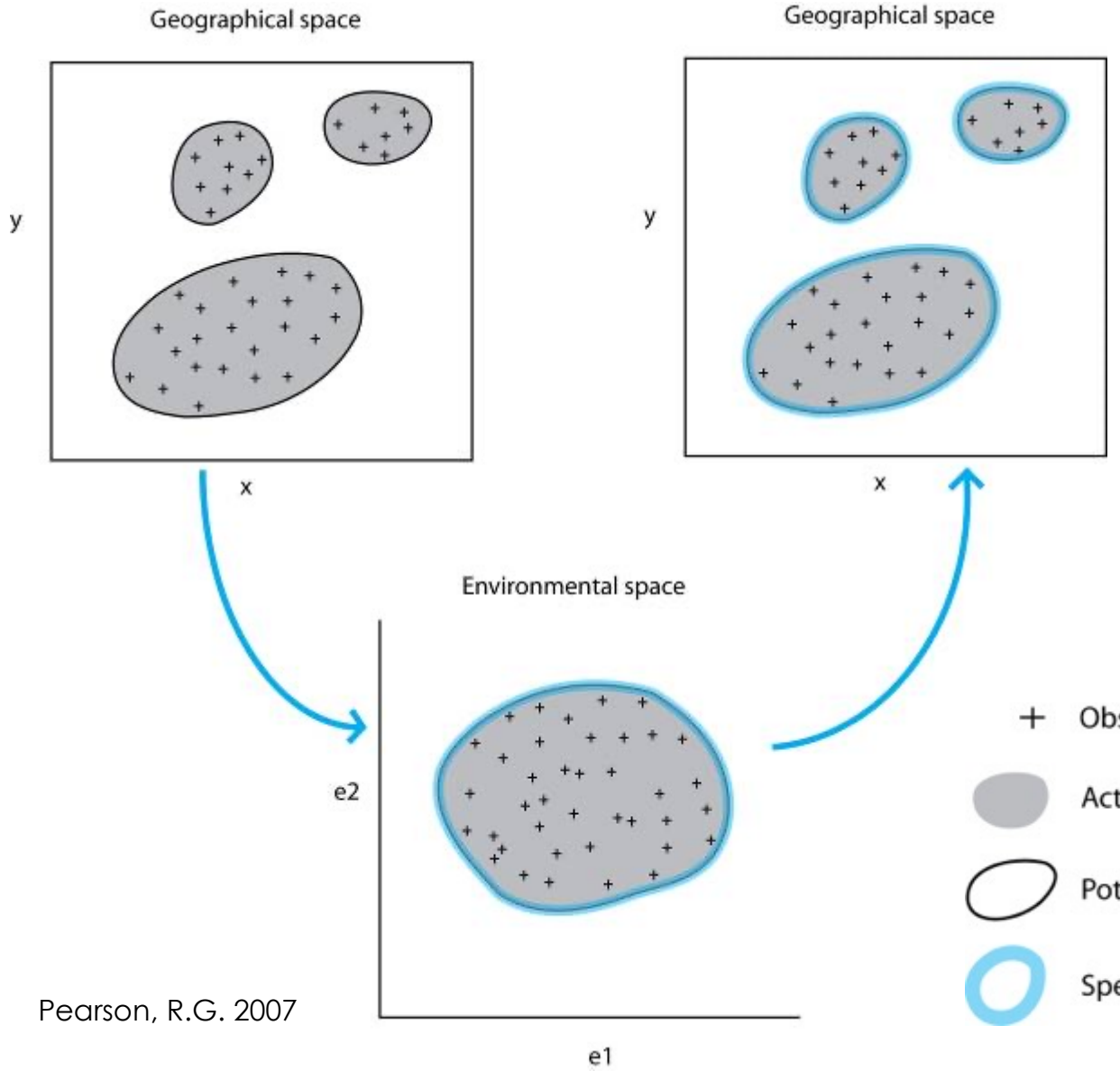
Modeling Approach



Pearson, R.G. 2007

General approach

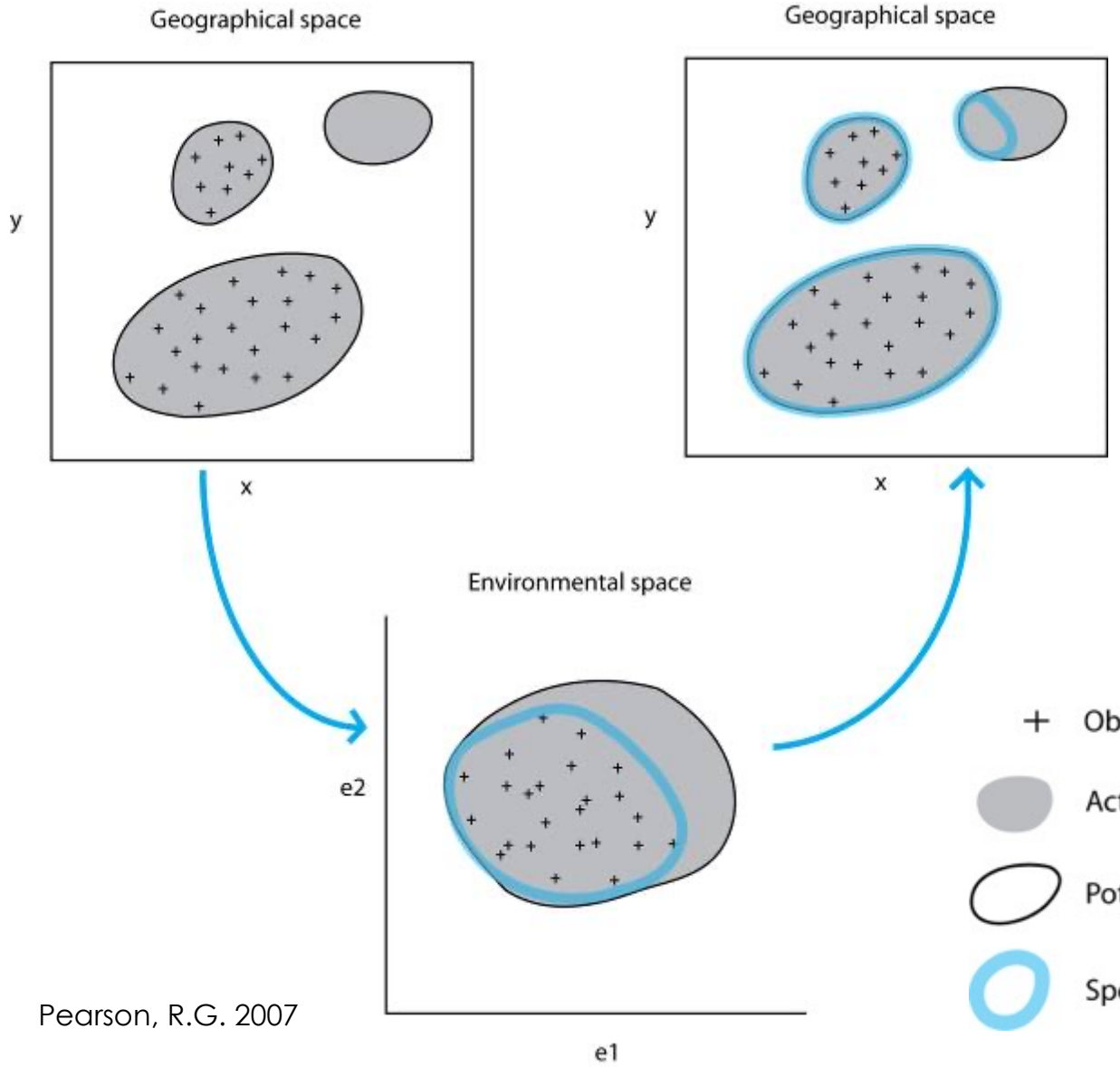
Modeling Approach



Pearson, R.G. 2007

Equilibrium and good sampling

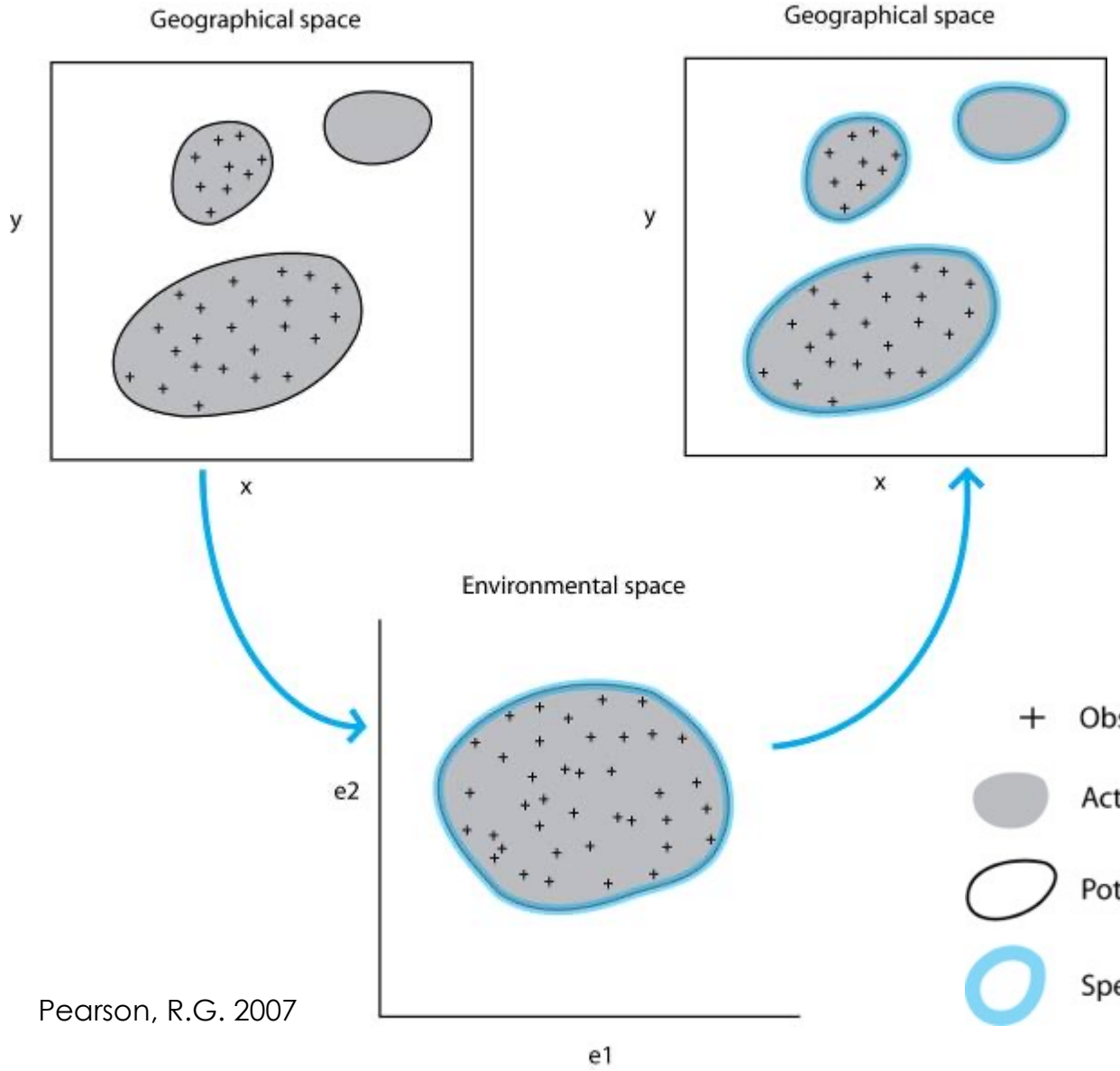
Modeling Approach



Pearson, R.G. 2007

Equilibrium and poor sampling
(in both geographic and environmental space)

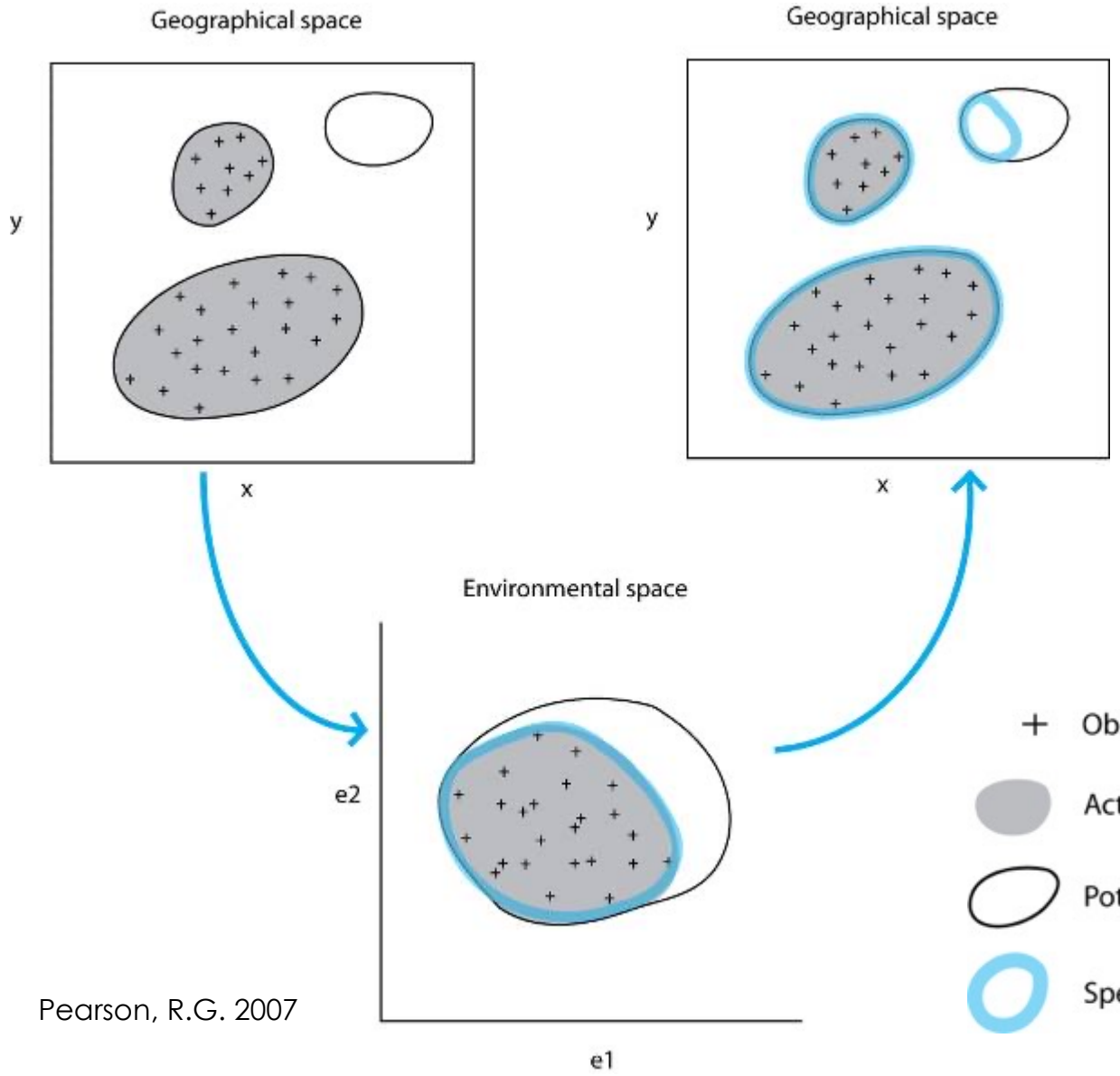
Modeling Approach



Pearson, R.G. 2007

Equilibrium and poor sampling in geographic space and good sampling environmental space

Modeling Approach



Pearson, R.G. 2007

Low Equilibrium and good sampling
(in both geographic and environmental space)

Methods and Models

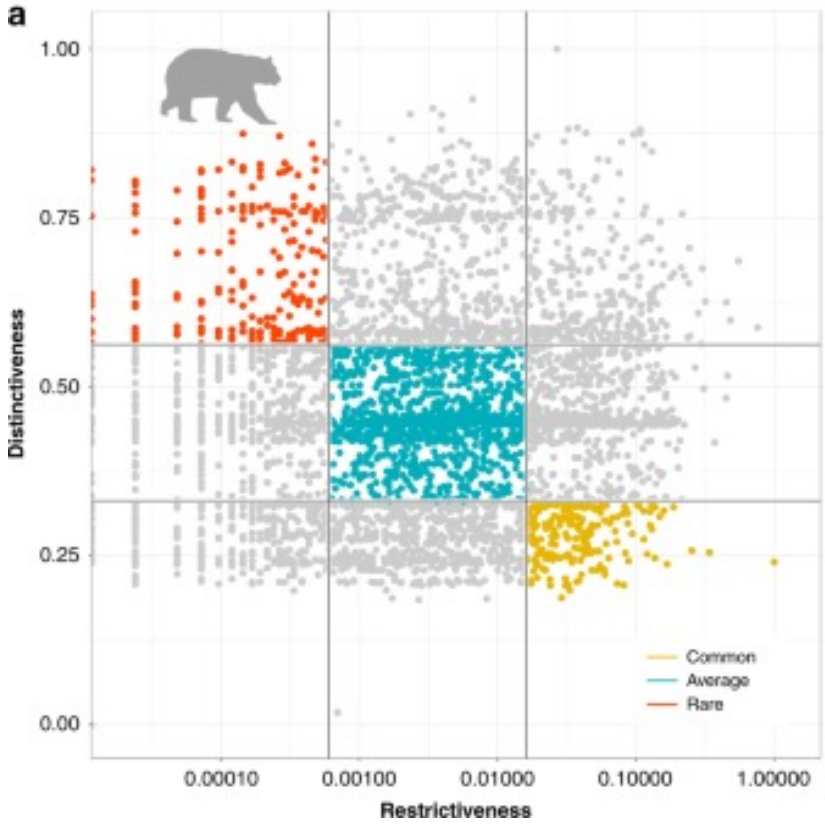
- **Profile Techniques:** Use environmental distance to know sites of occurrence
  - Gower Metric (DOMAIN)
  - Ecological Niche Factor Analysis (BIOMAPPER)

- **Regression-Based Techniques:** Linear Models
  - Generalized Linear Models (GLMs)
  - Generalized Additive Models (GAMs)

- **Machine-Learning Techniques:** Iterative "learning" algorithms, inspired by the workings of the brain
  - Genetic Algorithm for Rule-set Production (GARP)
  - MAXENT

Gower Metric

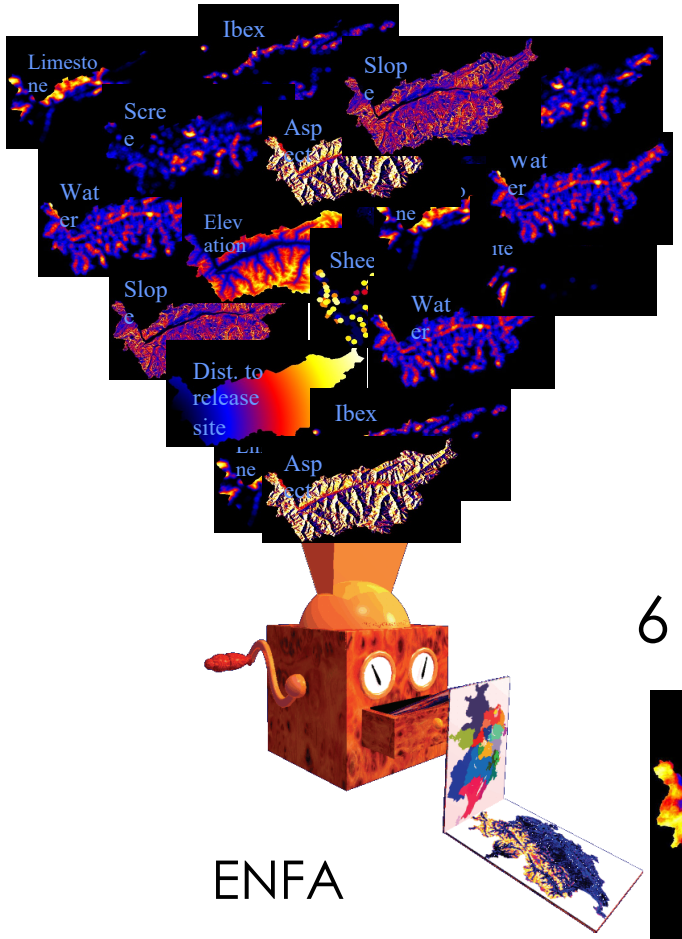
- Correlation between response and predictor variables
  - Clusters
- Assigns each cell in the output layer an average multivariate distance (Gower Metric) between that cell and the closest presence cell in the training set
- Uses only presence points
- Distance rescaling used to compare with probability-based techniques
- Example: DOMAIN



Geographical Restrictiveness: Each species is represented by a dot, ecologically rare species are in red, average in blue, and common in orange. Image Credit: Konowalik and Nosol, 2021

Ecological Niche Factor Analysis (ENFA)

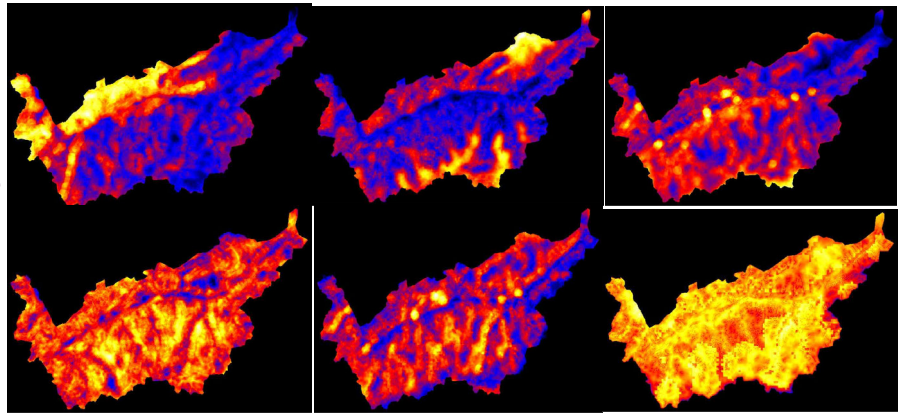
- Computation of the factors explaining the major part of a species environmental distribution
- Presence-only data
- Habitat Suitability Index (HSI)
  - Value inversely proportional to weighted mean distance of the cell to the median of each ENFA factor
- Example: BIOMAPPER https://www2.unil.ch/biomapper/



ENFA

24 Predictors

6 Factors = 80% of Information

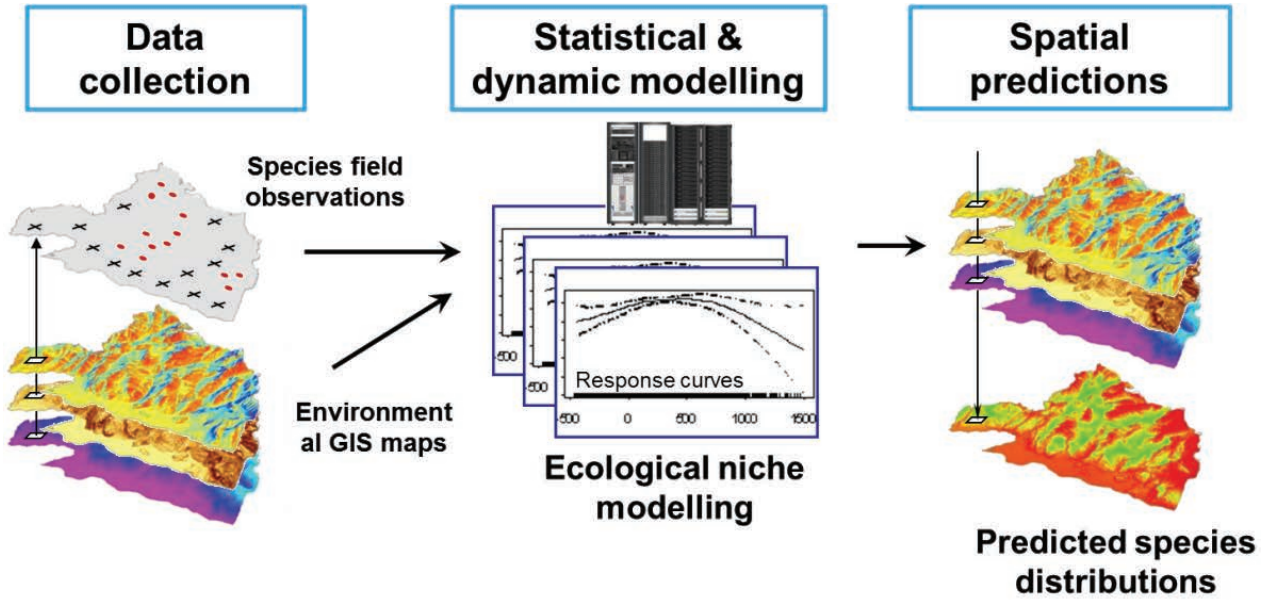


Regression Analysis

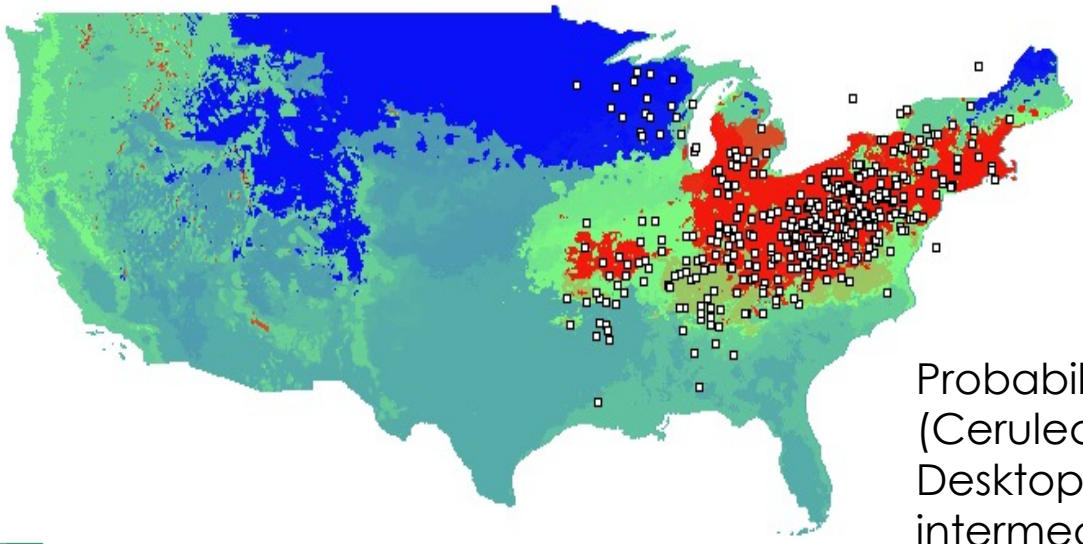
- **Generalized Linear Models (GLMs)**
  - Based on relationship between mean of the response variable and the linear combination of explanatory variables
  - Multiple types of distributions can be used
  - Flexible
- **Generalized Additive Models (GAMs)**
  - Similar for GLMs but assumes functions are additive and components are smooth
  - Data-driven
  - Can handle non-linear relationships
- Review paper on regression analysis for SDMs (Guisan et al, 2002)
  - https://www.sciencedirect.com/science/article/pii/S0304380002002041

Genetic Algorithm for Rule-set Production (GARP)

- Ecological niche model for species
  - Similar to the first SDM: BIOCLIM
- Inputs: Presence only point locations and geographic layers of environmental parameters
- Output: Predicted species distribution



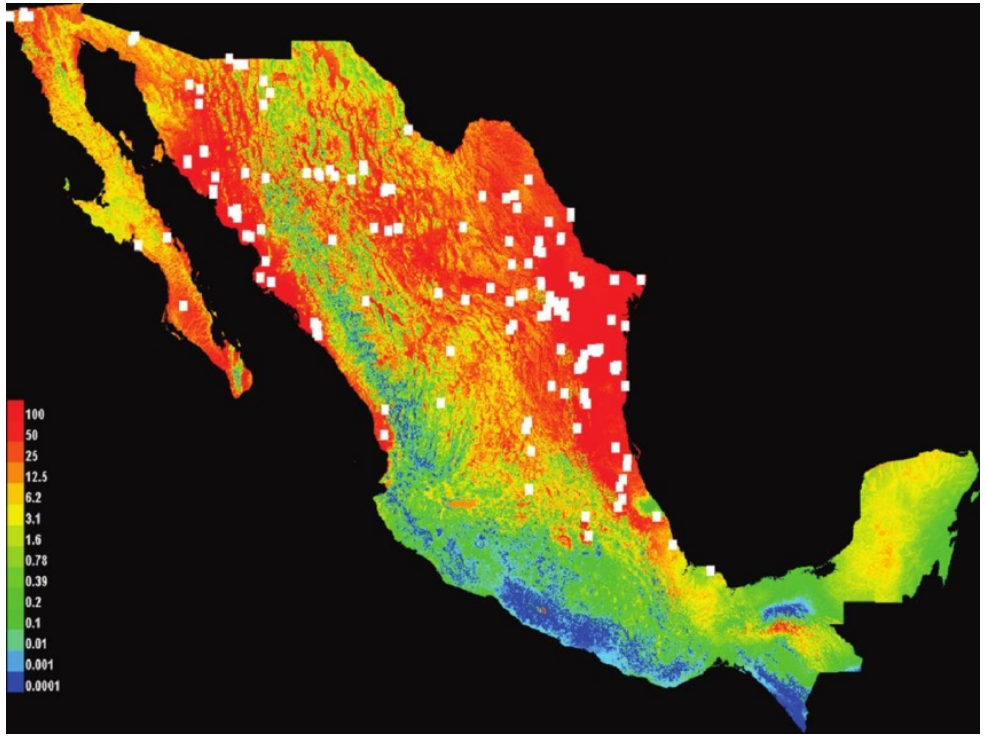
Peterson, A. T. (2003)



Probability distribution map of a bird species (Cerulean warbler, *Dendroica cerulea*) created by DesktopGarp. Red is high probability, green intermediate, and blue is low.

MAXENT

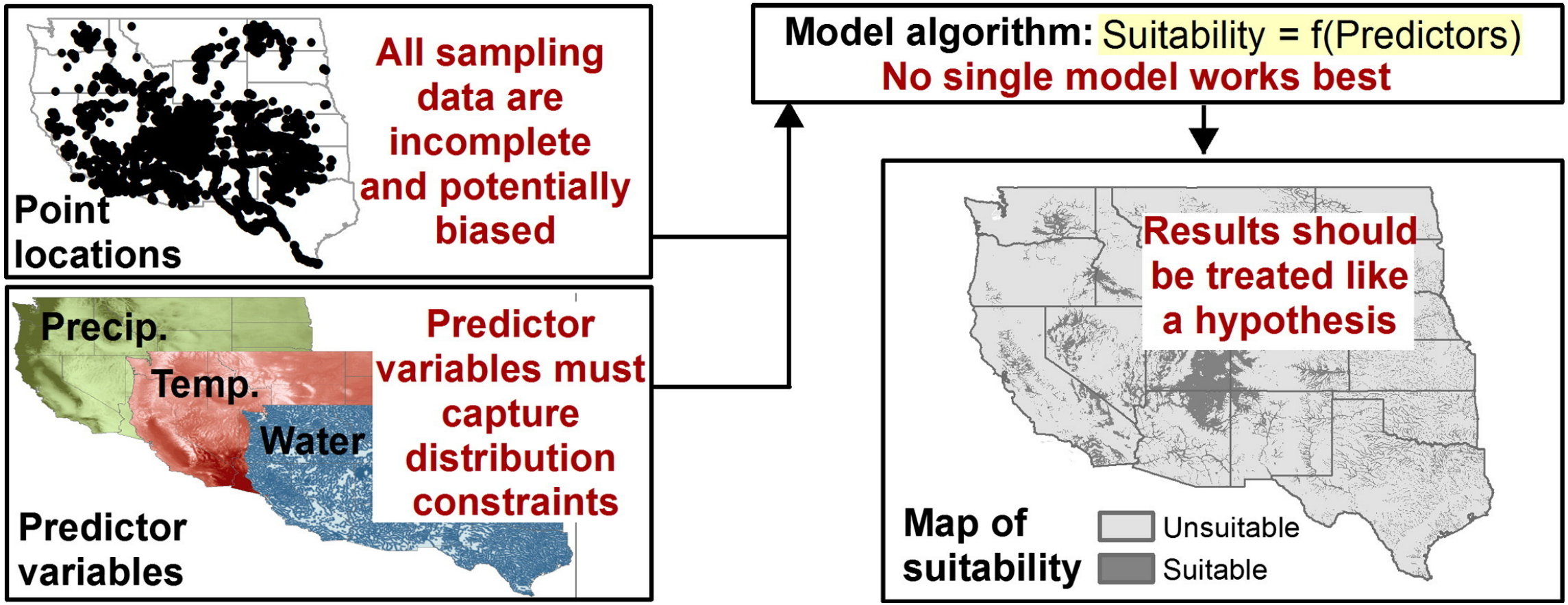
- Predicts the potential distribution of a species
  - Explores the relative suitability of one place over another using the maximum entropy principle
  - Entropy = Randomness
  - Maximizes randomness by removing patterns
- Uses machine Learning
- Uses only presence points



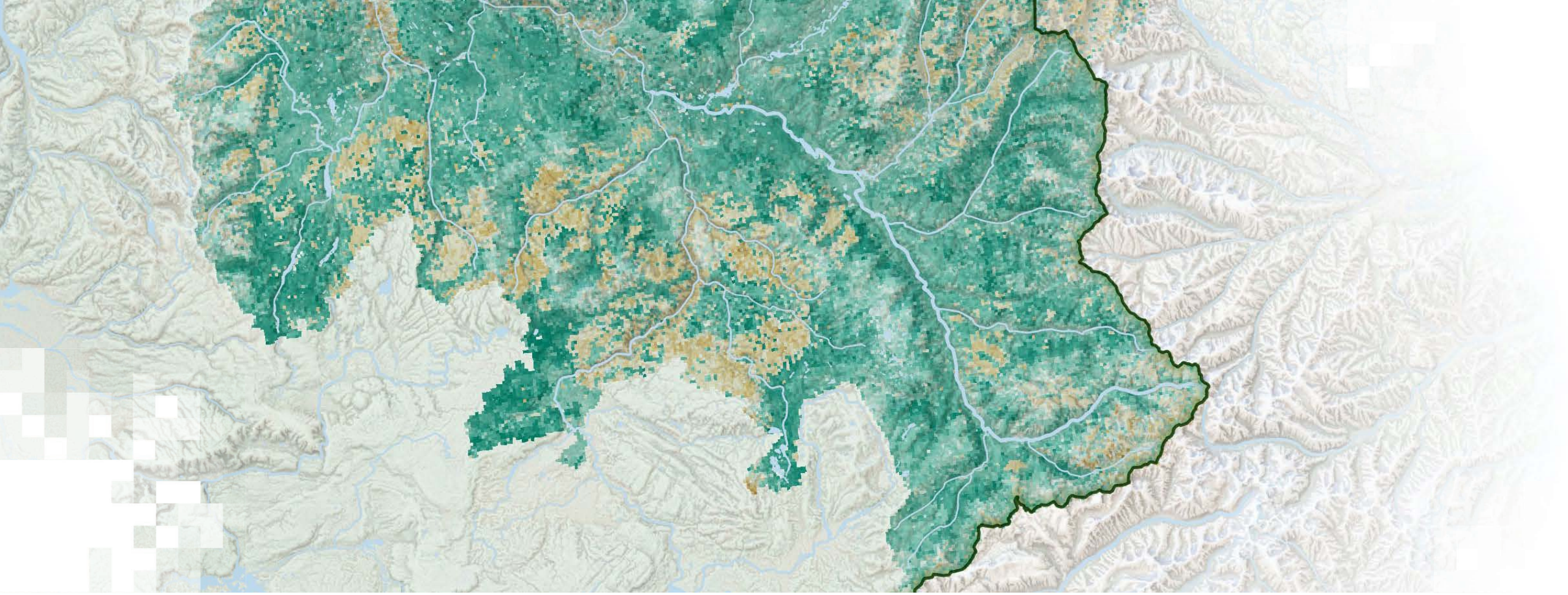
Areas in red are the most likely to have suitable habitat for sunflowers. The white squares represent modern wild sunflower collections. Image Credit: Lentz et al, 2008

https://biodiversityinformatics.amnh.org/open_source/maxent/

Limitations of SDMs



Jarnevich et al, 2015



Case Study Examples

SDM Case Study Example: Mosquitos in Europe

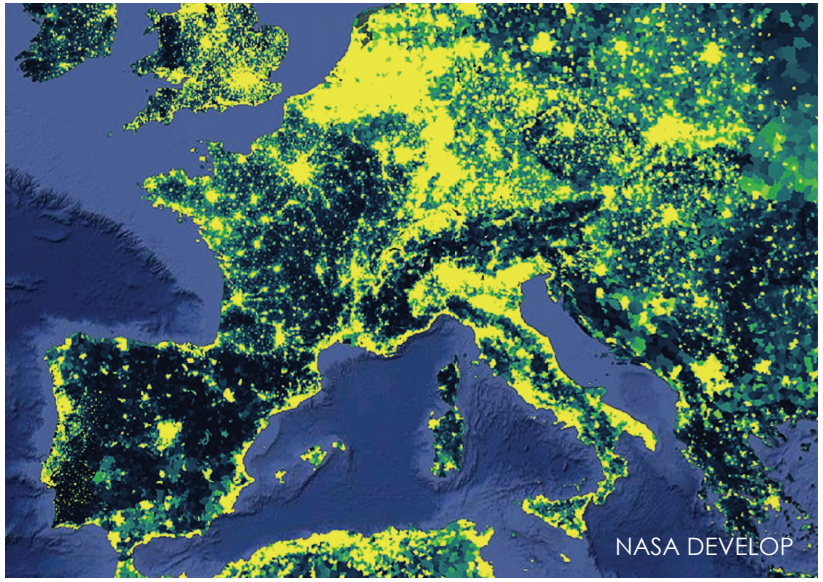
- **NASA DEVELOP Project**
  - An Interactive Model of Mosquito Presence and Distribution to Assist Vector-Borne Disease Management in Western Europe
  - https://develop.larc.nasa.gov/2018/spring/WesternEuropeHealthAQII.html

- **Community Concerns:**
  - Mosquito-borne disease kills over one million people a year
  - Zika, Dengue, Chikungunya, Yellow Fever, Malaria
  - Mosquito habitat range is spreading

Mosquito Alert Spain

Mosquitos in Europe: Project Overview

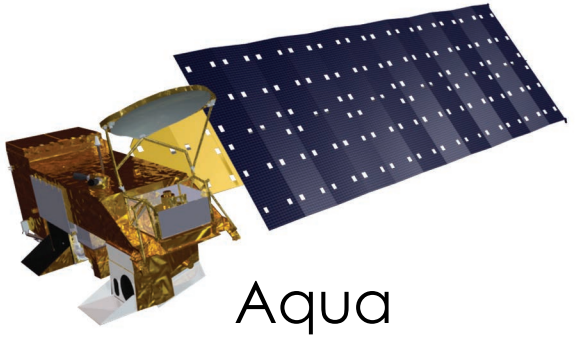
- **Project Goals:**
  - Integrate citizen science data and NASA Earth observations to make the information publicly accessible
  - Create an interactive map showing mosquito habitat suitability
  - Overlay results with transportation, population, and public health data
- **Study Area:** Western Europe
- **Study Period:** June 2016 – September 2017

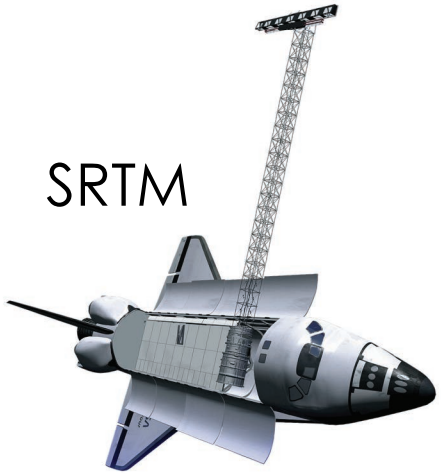


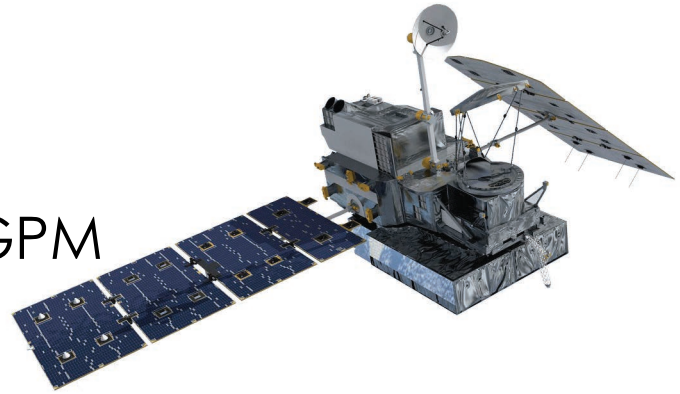
Mosquitos in Europe: Environmental Variables

- Elevation
- Humidity
- Land Surface Temperature
- Normalized Difference Vegetation Index (NDVI)
- Precipitation
- Soil Moisture
- Land cover


Terra

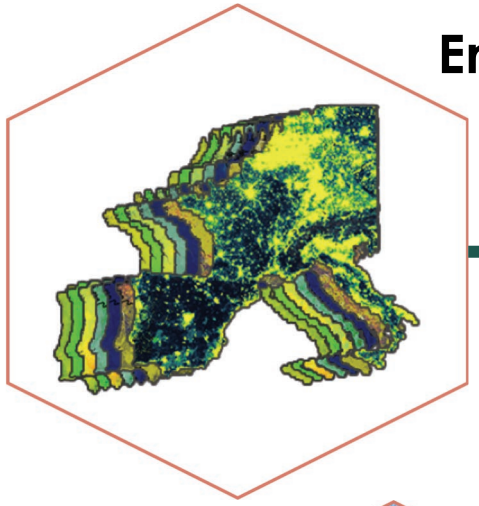

Aqua

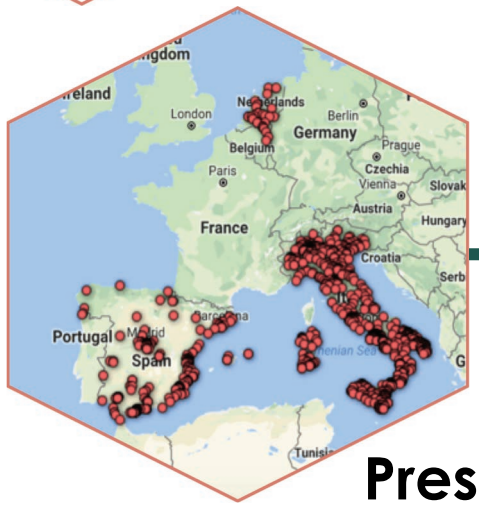

SRTM


GPM

Mosquitos in Europe: Methodology

Environmental Variables

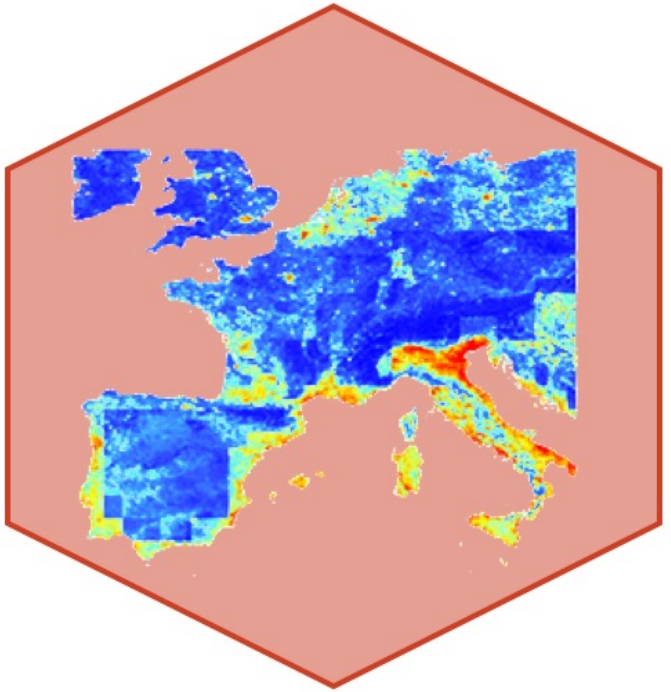




Presence Locations via Citizen Science

Model Algorithm (MaxEnt)

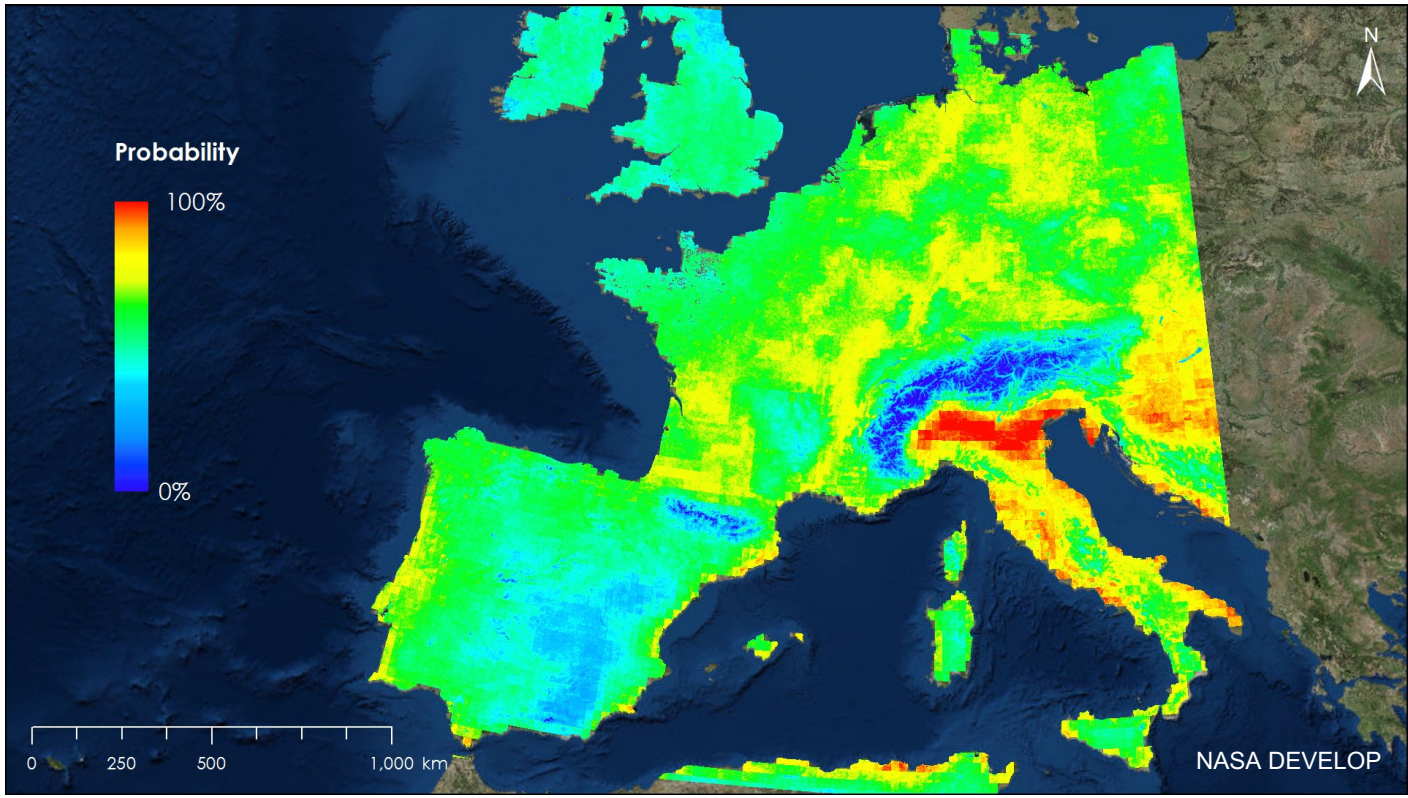
Habitat Suitability

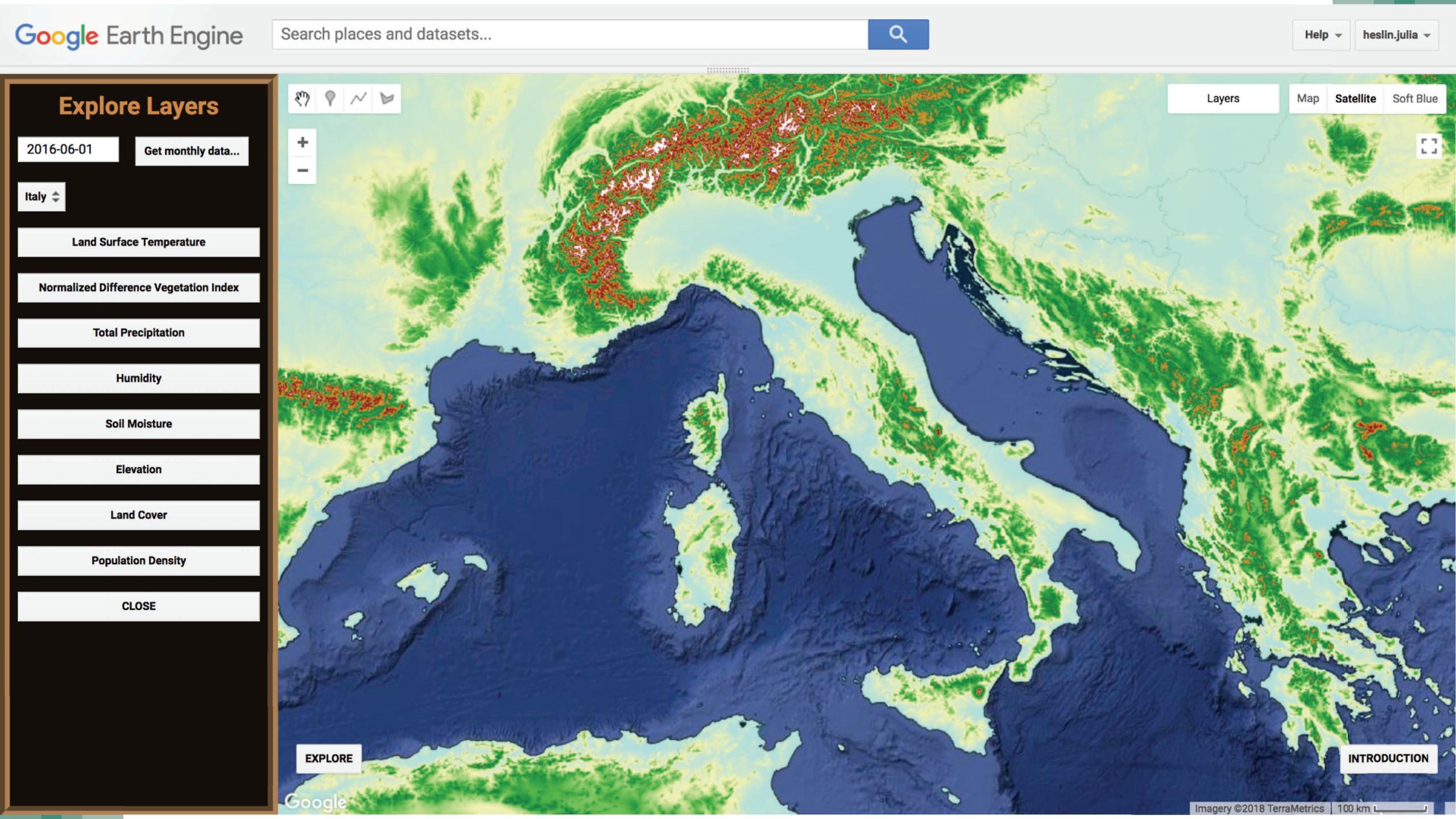


Mosquitos in Europe: Habitat Suitability

- The MaxEnt model revealed relationships between mosquito presence and the environmental variables.

- Mosquito presence positively correlated with:
  - NDVI
  - Homogenous land cover and greenness
  - Temperature




Google Earth Engine
Search places and datasets...
Help
heslin.julia
Explore Layers
2016-06-01
Get monthly data...
Italy
Land Surface Temperature
Normalized Difference Vegetation Index
Total Precipitation
Humidity
Soil Moisture
Elevation
Land Cover
Population Density
CLOSE
Layers
Map
Satellite
Soft Blue
EXPLORE
INTRODUCTION
Google
Imagery ©2018 TerraMetrics
100 km

SDM Case Study Example: Red Spruce in West Virginia

- **NASA DEVELOP Project:** Forecasting Red Spruce Restoration Using NASA Earth Observations to Support the USFS Monongahela National Forest
  - https://develop.larc.nasa.gov/2019/summer/MonongahelaNationalForestEco.html

- **Community Concerns:**
  - Red Spruce declines due to coal mining, logging, and wildfires
  - Red Spruce restoration is critical for:
    - Biodiversity
    - Water Quality
    - Flood Mgmt.
    - Carbon Stocks

- **Project Goals:**
  - **Identify** historical extent of red spruce from 1989 to 2018
  - **Create** maps of red spruce locations and suitable locations in Monongahela Forest, West Virginia
  - **Forecast** site suitability and forest cover to 2040

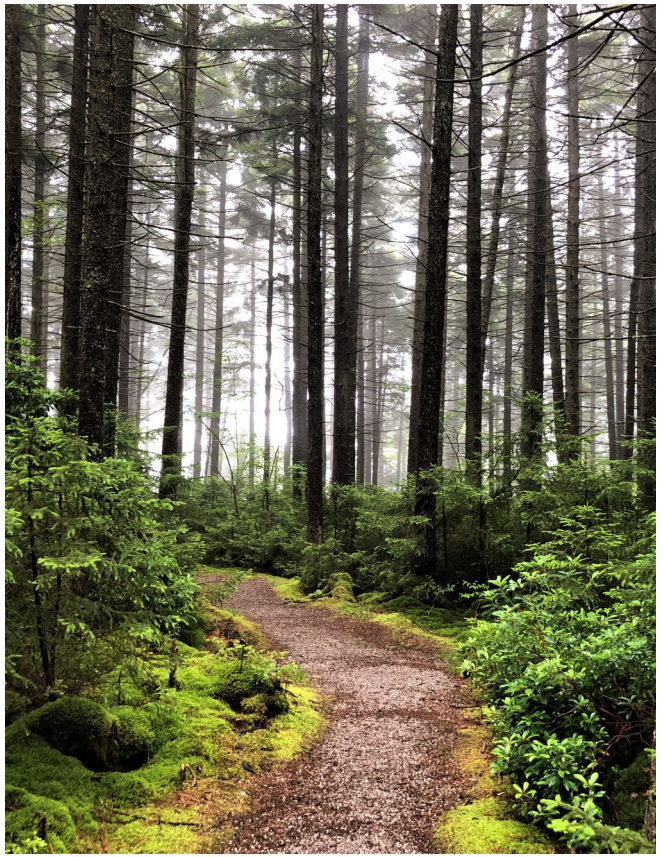
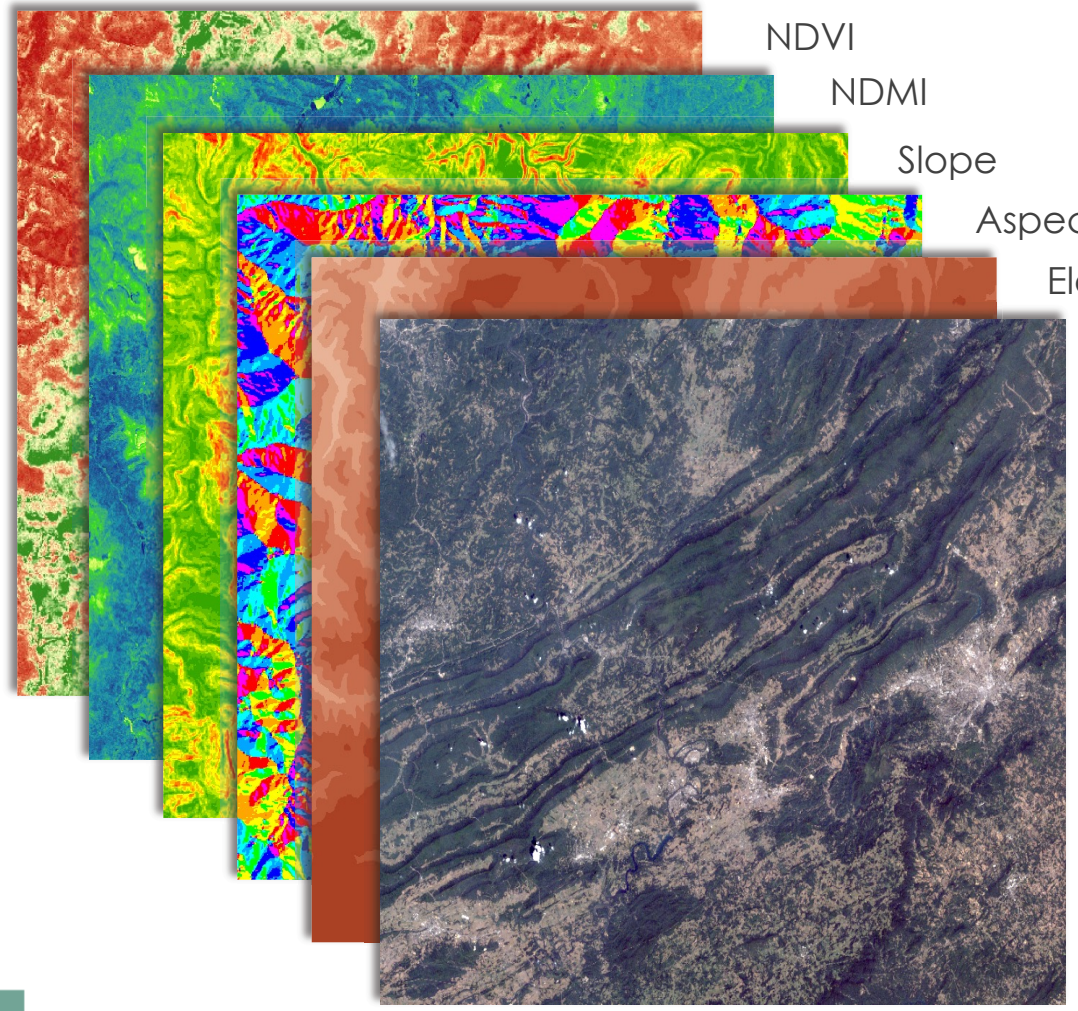
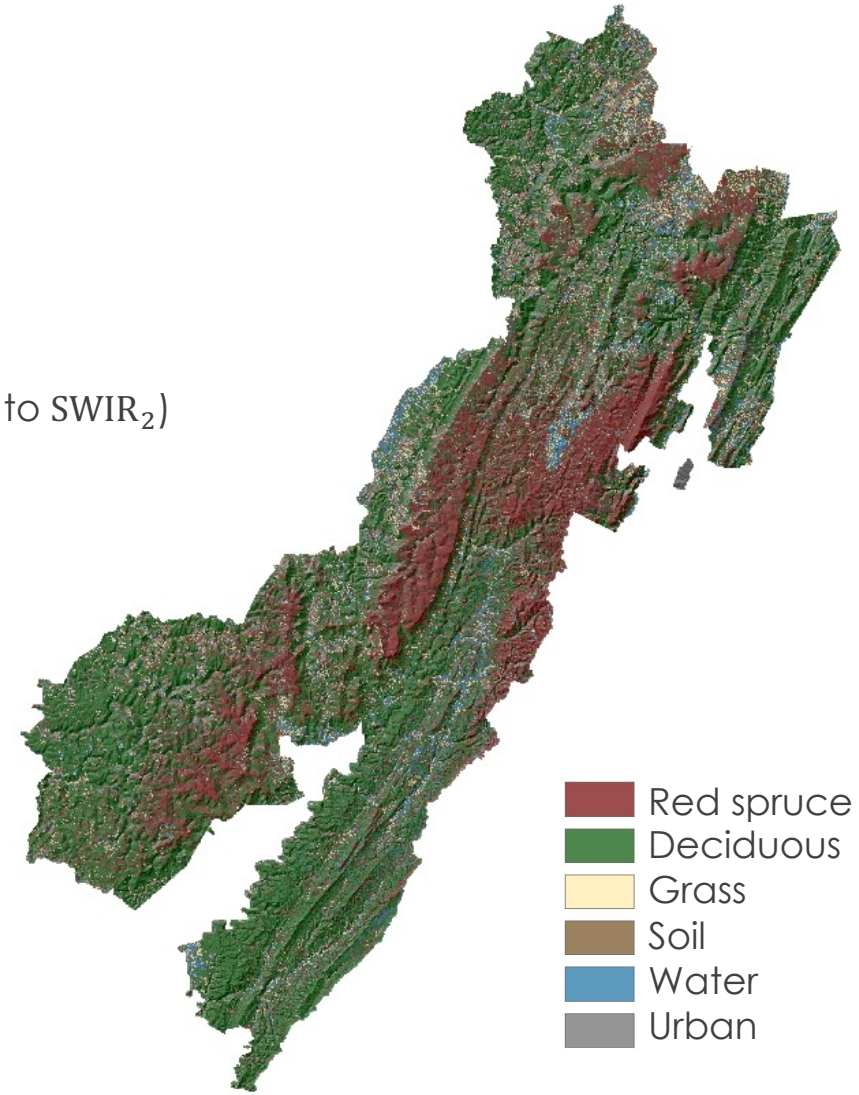


Image Credit: USFS

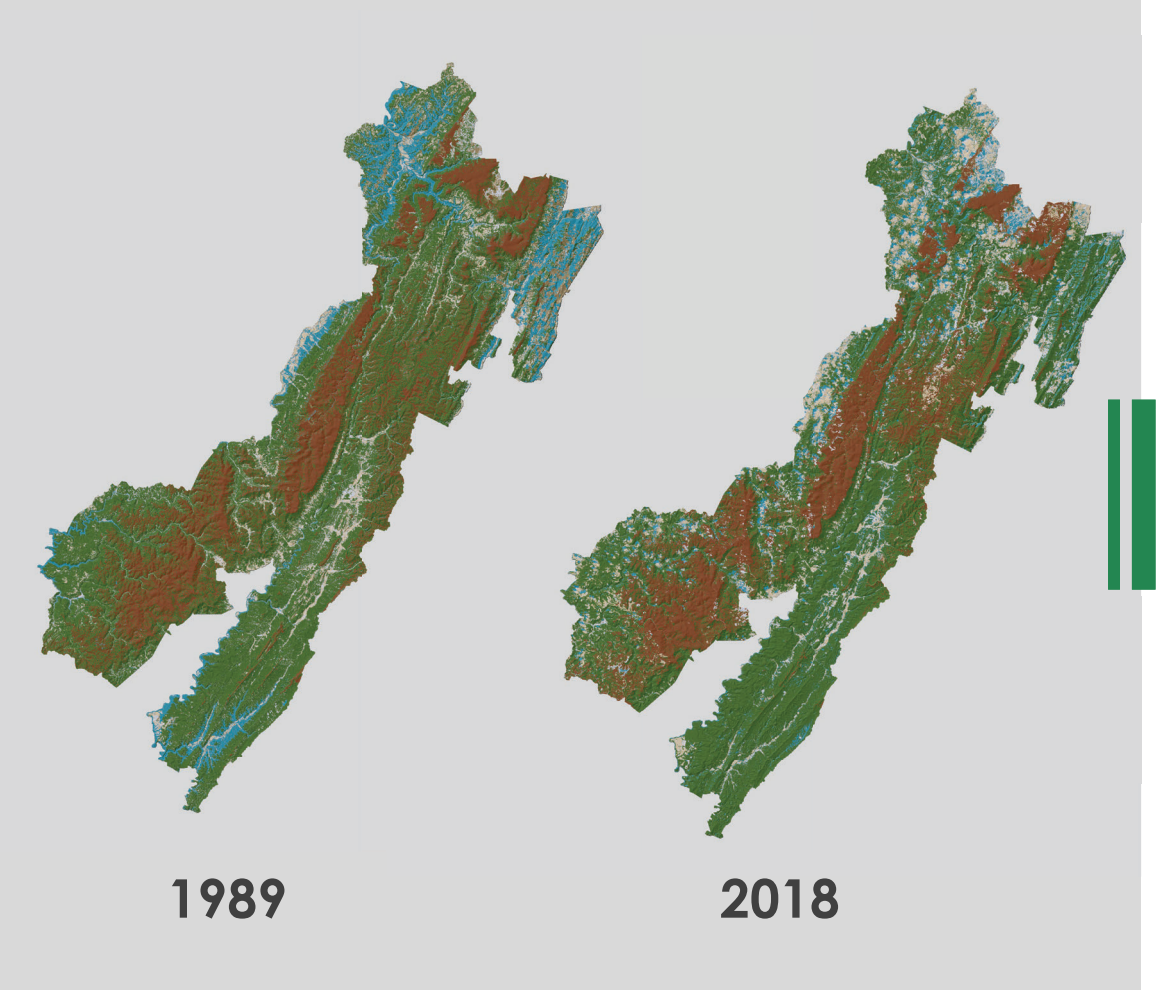
Red Spruce in West Virginia: Land Cover Mapping



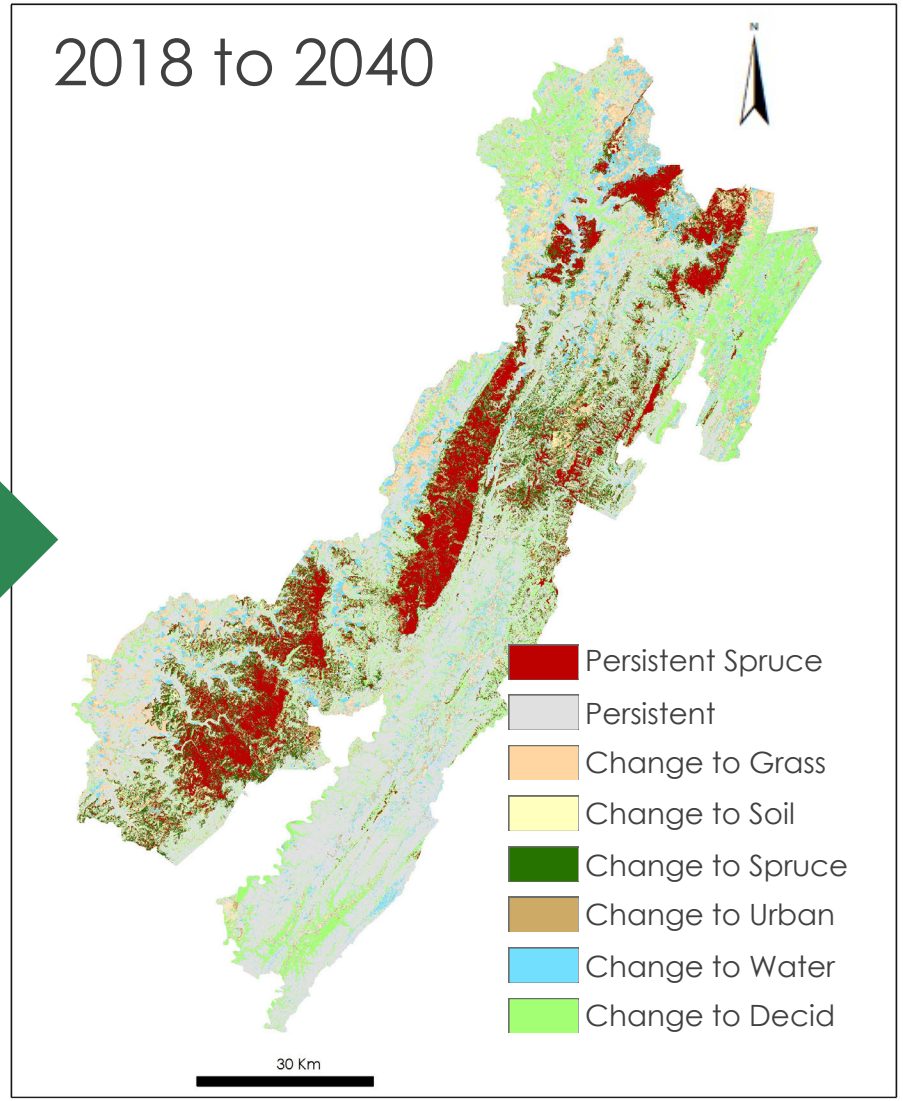
CTA



Red Spruce in West Virginia: Time Series Analysis

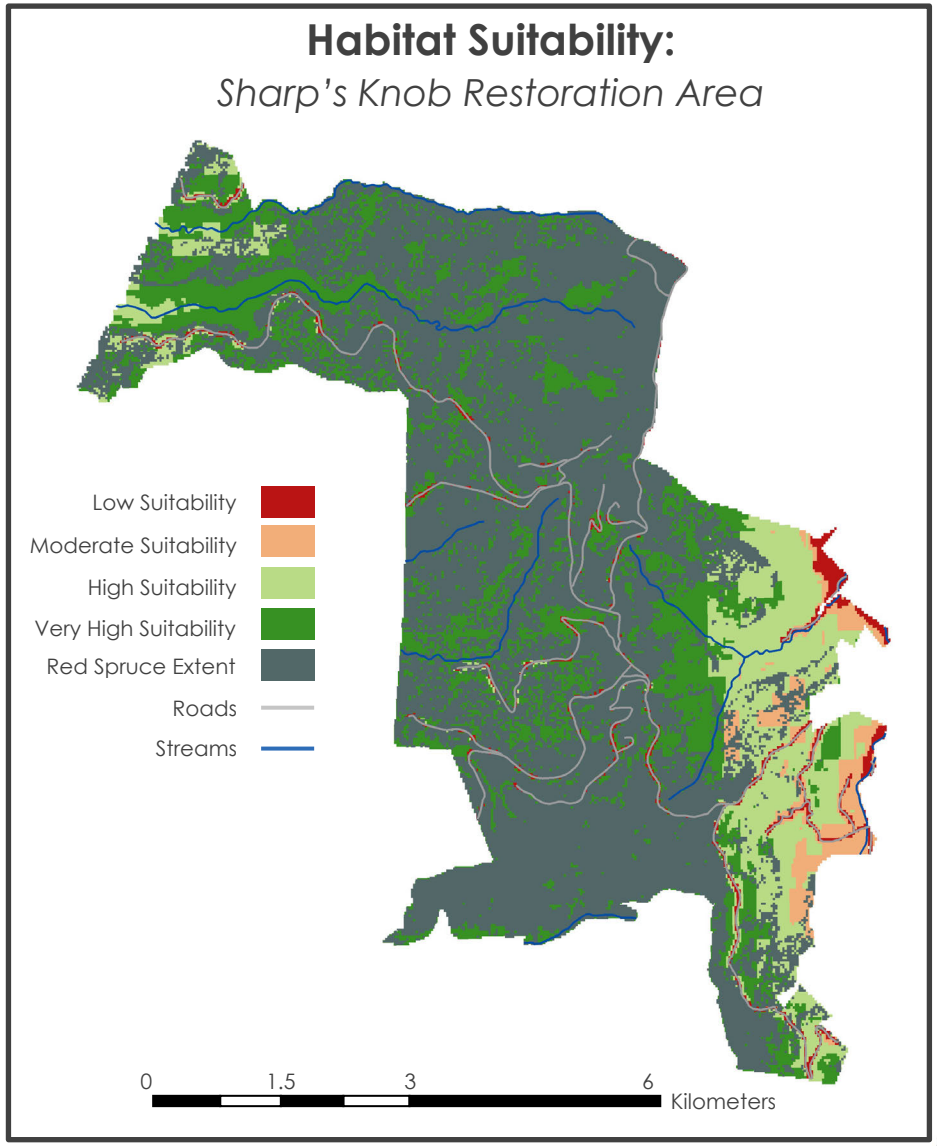


LAND CHANGE MODELER

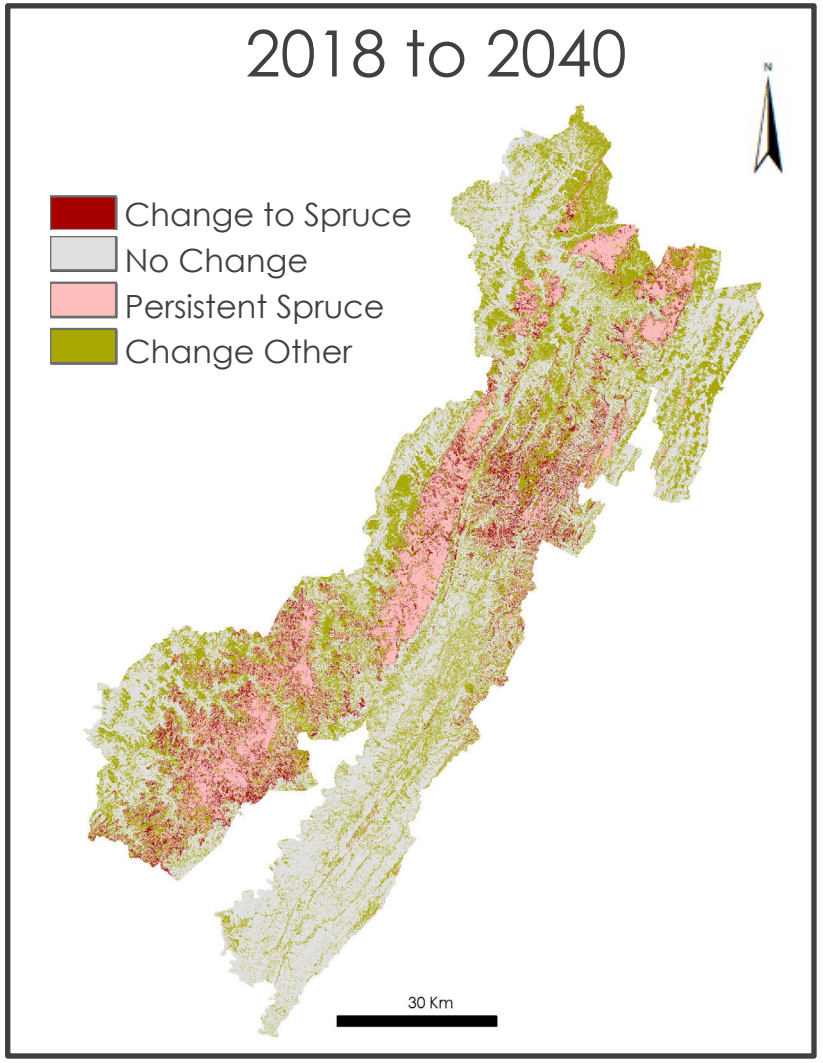


Red Spruce in West Virginia: Suitability Analysis

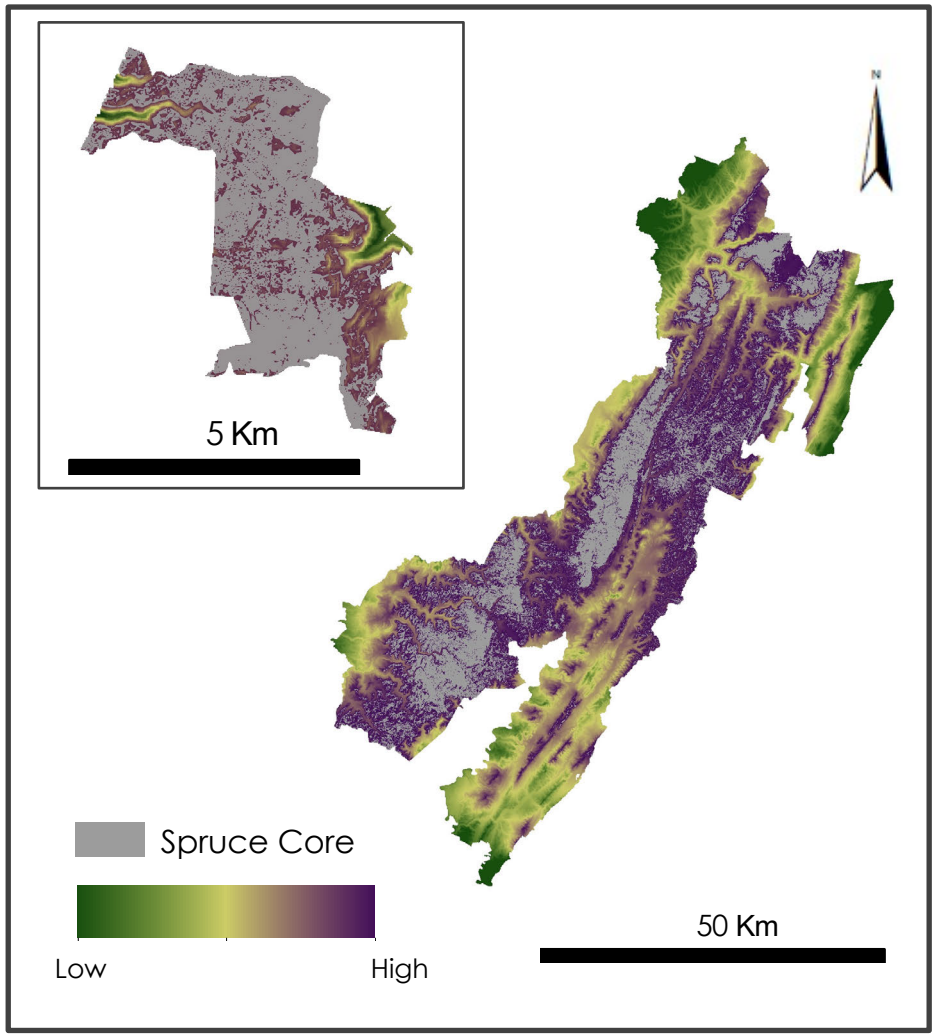
Inputs	Fuzzy Memberships
Elevation	Linear
Slope	Small
Frost Days	Large
Growing Degree Days	Small
Average Annual Temperature	Small
Average Annual Precipitation	Large
Distance to Road	Linear
Soil pH	Small



Red Spruce in West Virginia: Forecasting and Connectivity



Forecast Modeling



Connectivity Potential

Summary

- Species Distribution Models (SDMs) allow us to assess the suitability of a habitat for a species.
- SDMs primarily use environmental data and occurrence data to build a model for predictions of habitat suitability.
  - Remotely sensed data can be used for a variety of environmental data inputs.
- There is no universally correct or universally applicable method.
- Methodologies must be adapted to:
  - Ecological and biogeographical situation
  - Meet the study goals
  - Use available data
- Next Session: **Using Wallace to Model Species Niches and Distributions**

Contacts

- Trainers:
  - Juan Torres-Pérez: juan.l.torresperez@nasa.gov
  - Amber McCullum: amberjean.mccullum@nasa.gov
  - Zach Bengtsson: bengtsson@baeri.org

Follow us on Twitter
@NASAARSET

- Training Webpage:
  - https://appliedsciences.nasa.gov/join-mission/training/english/arset-species-distribution-modeling-remote-sensing
- ARSET Website:
  - https://appliedsciences.nasa.gov/what-we-do/capacity-building/arset


APPLIED REMOTE SENSING TRAINING PROGRAM
TRAIN - EMPOWER - ADVANCE

Thank You!