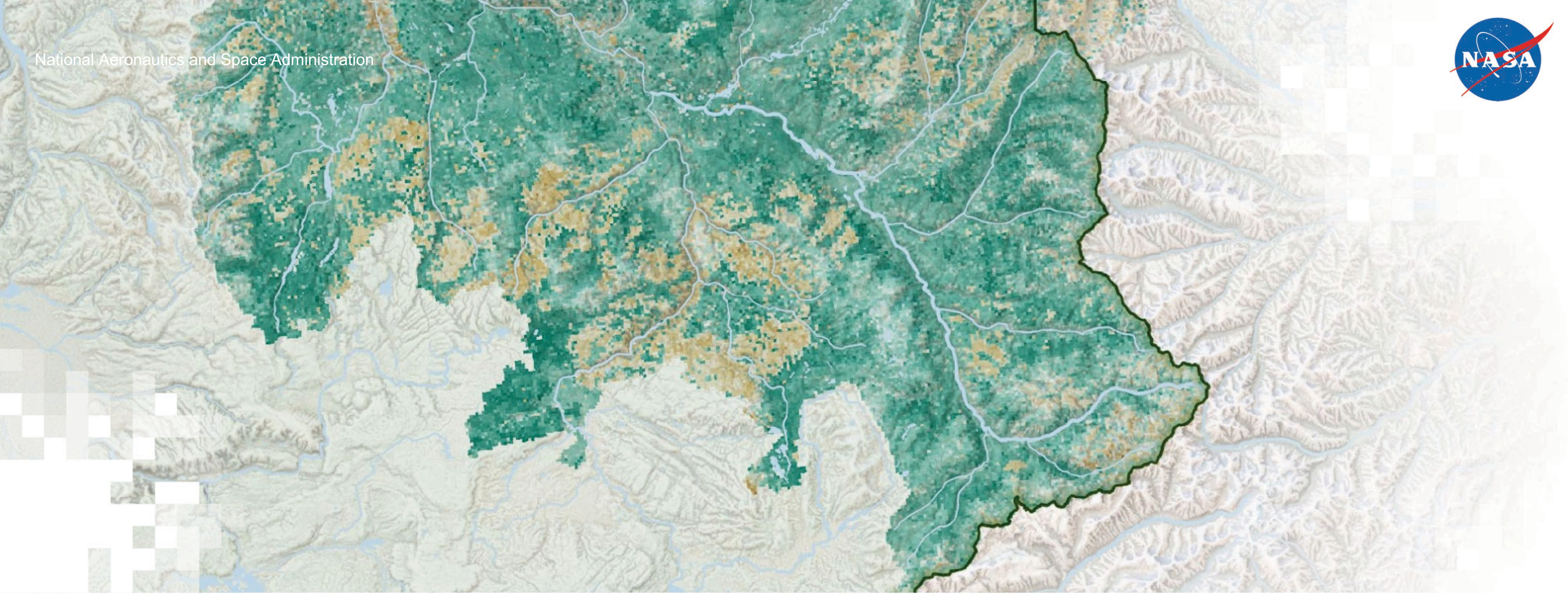
National Aeronautics and Space Administration



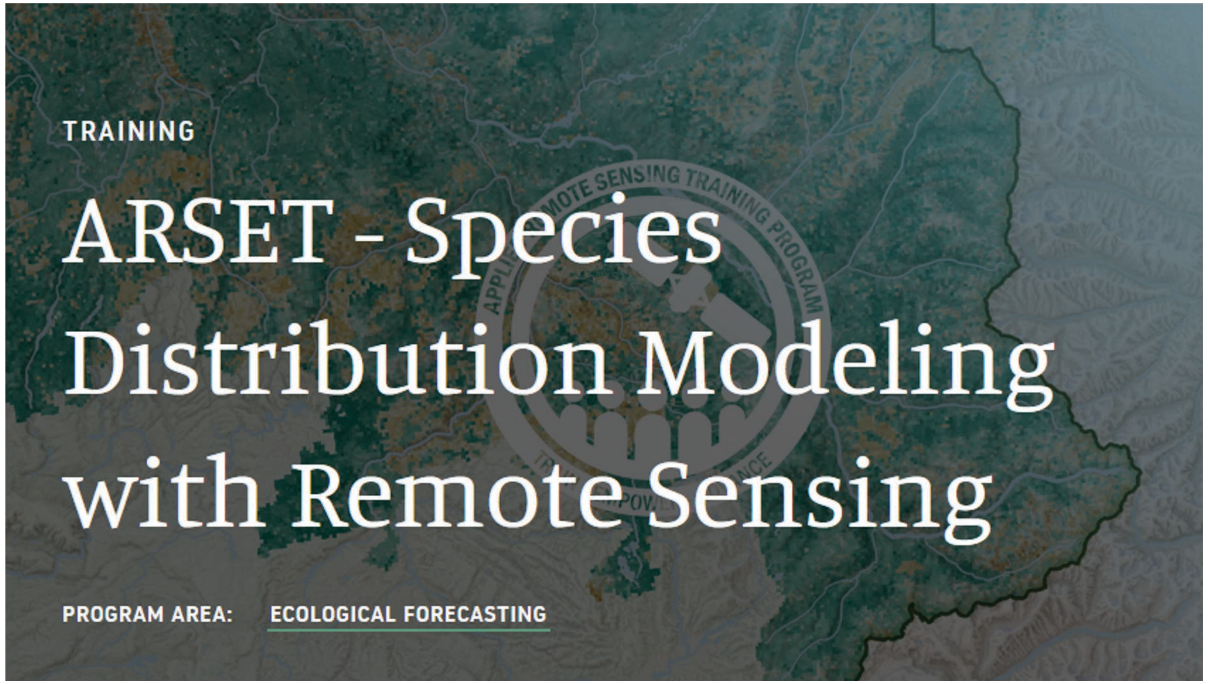
# Species Distribution Modeling with Remote Sensing

Amber McCullum, Juan Torres-Perez, Zach Bengtsson
Guest Speakers: Erica E. Johnson, Andrea Paz Velez, Mary E. Blair

Aug 17, 2021

# Course Structure and Materials

- Three 1.5-hour sessions on August 12, 17, & 19
- Sessions will be presented once in English 12:00-13:30 EDT
- Webinar recordings, PowerPoint presentations, and the homework assignment can be found after each session at:
  – https://appliedsciences.nasa.gov/join-mission/training/english/arset-species-distribution-modeling-remote-sensing
- Q&A following each lecture and/or by email at:
  – amberjean.mccullum@nasa.gov
  – juan.l.torresperez@nasa.gov
  – bengtsson@baeri.org




Amber McCullum

Juan Torres-Pérez

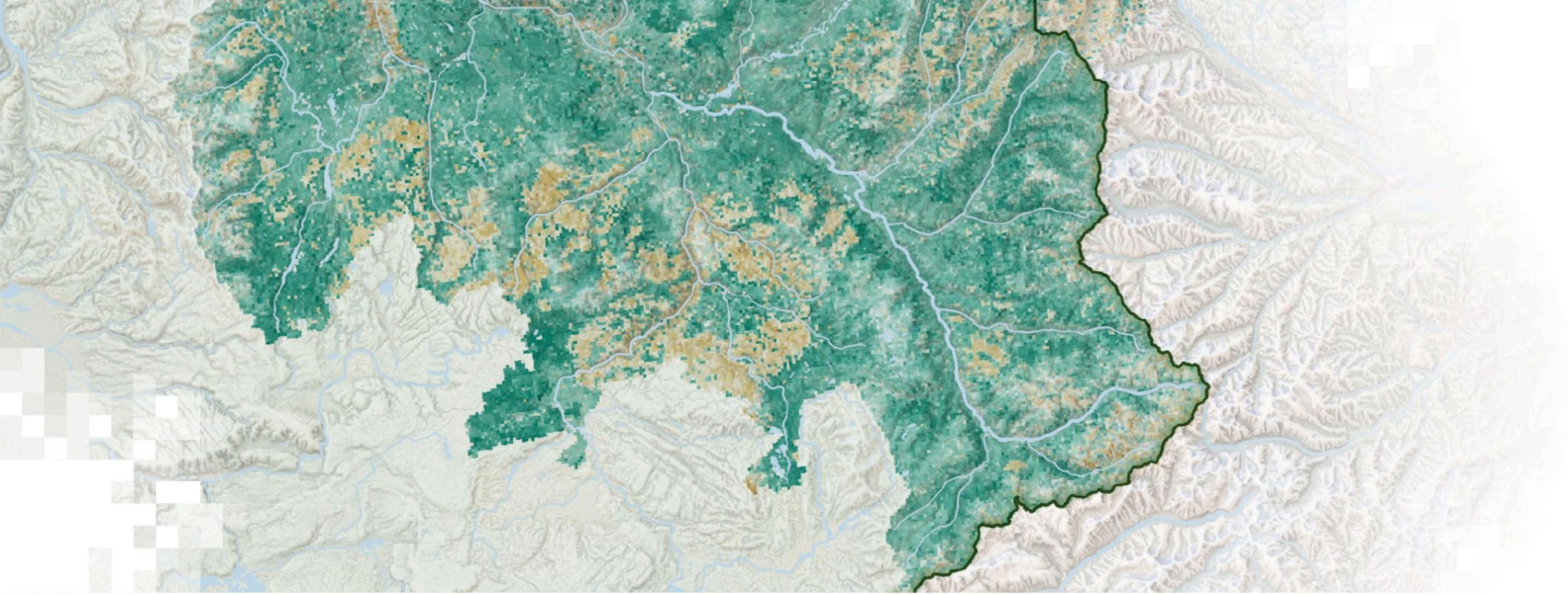
Zach Bengtsson

# Webinar Agenda

**Part 1: Overview of Species Distribution Models (SDMs)**

**Part 2: Using Wallace to Model Species Niches and Distributions**

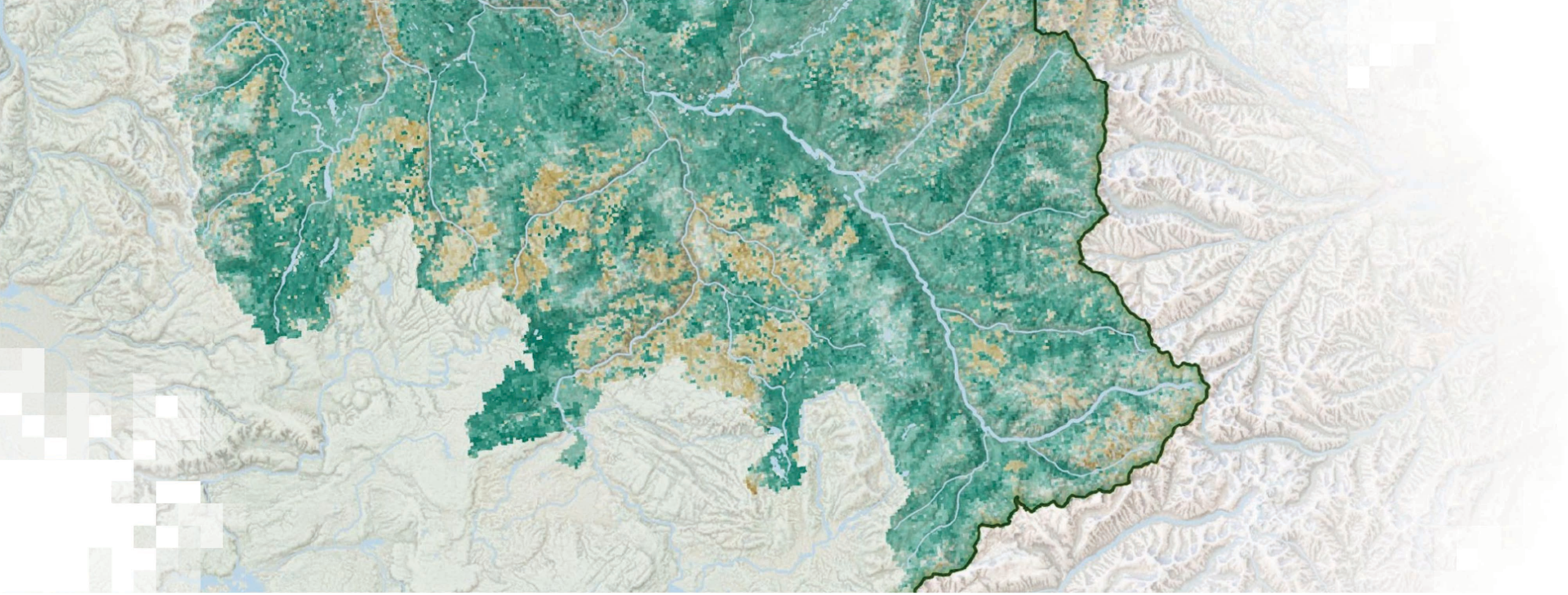
**Part 3: Additional SDM Tools and Techniques, ASP Projects, and Summary**



# Using `Wallace` to Model Species Niches & Distributions

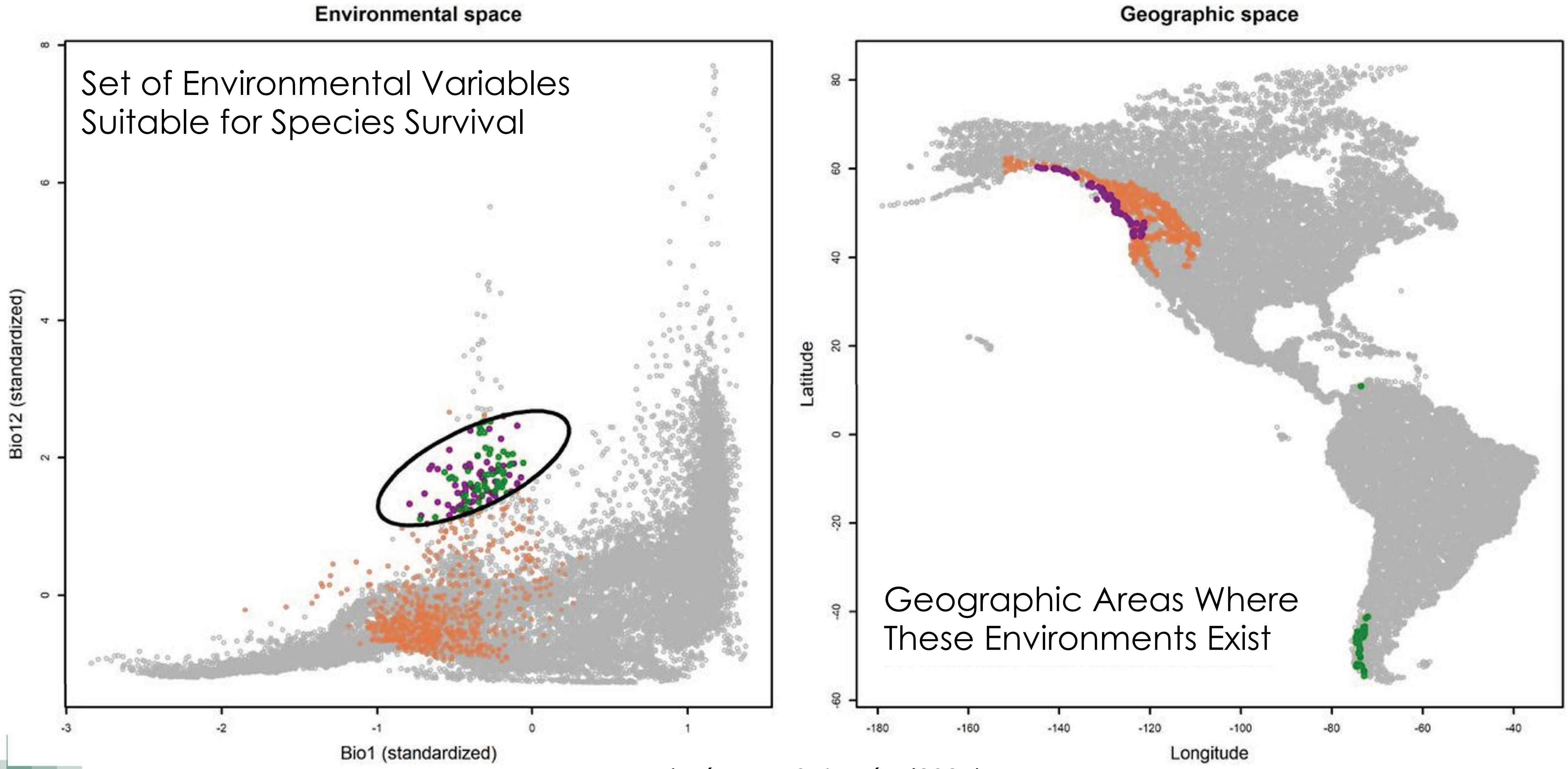
# Session 2 Overview

1. Species Distribution Models in **`Wallace`**
2. **`Wallace`** Walkthrough: Running a Full SDM Workflow
3. Redesign & Expansion: Making it Easier to Visualize Results and Add New Modules
4. Extensions: SDM Post-Processing for Conservation Decision-Making and Beyond



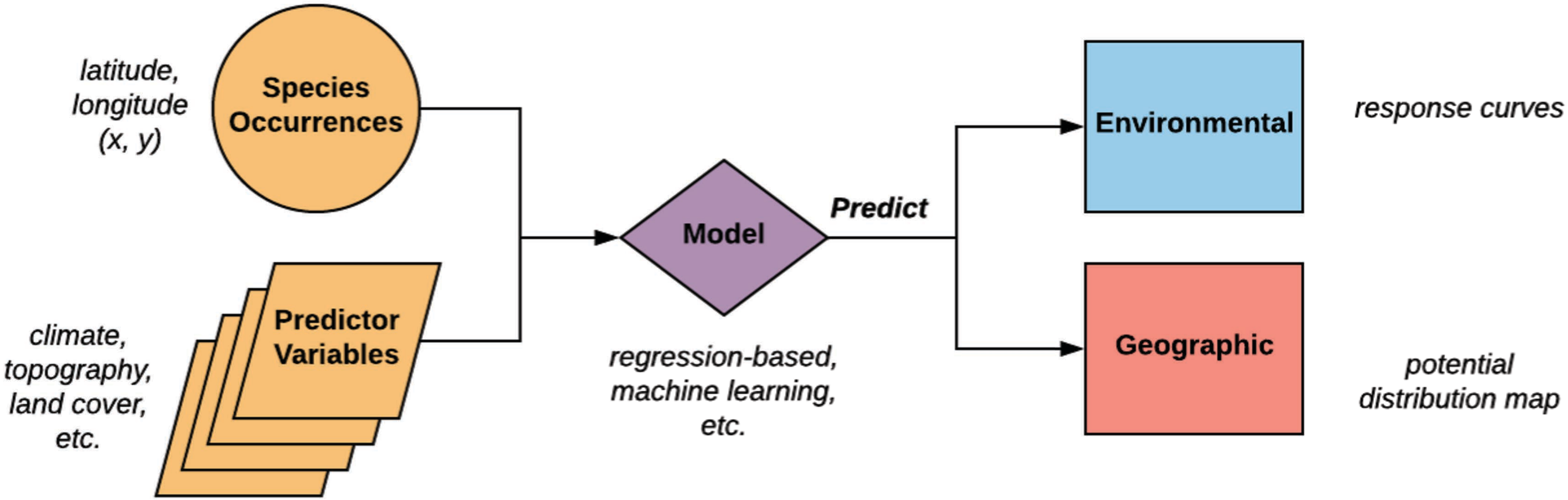
# Species Distribution Models in **Wallace**

# From Ecological Niches to Geographic Distributions




Jiménez & Soberón (2021)

# Species Distribution Modeling Overview




Elith & Graham 2009, Phillips et al. 2017; Guisan et al. 2017

# Needed in Species Distribution Modeling

Software that achieves a balance between automation and supervision by:

- **Automating** repetitive tasks
- **Forcing** the user to make critical biological and conceptual decisions
- **Being general** with respect to the algorithm(s) used

# Needed in Species Distribution Modeling

Software that achieves a balance between automation and supervision by:

- **Automating** repetitive tasks
- **Forcing** the user to make critical biological and conceptual decisions
- **Being general** with respect to the algorithm(s) used

Received: 9 October 2017 | Accepted: 16 November 2017

DOI: 10.1111/2041-210X.12945

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

APPLICATION

**WALLACE: A flexible platform for reproducible modeling of species niches and distributions built for community expansion**

Jamie M. Kass[1,2] | Bruno Vilela[3] | Matthew E. Aiello-Lammens[4] | Robert Muscarella[5] | Cory Merow[6] | Robert P. Anderson[1,2,7]

EBBE NIELSEN CHALLENGE

National Science Foundation

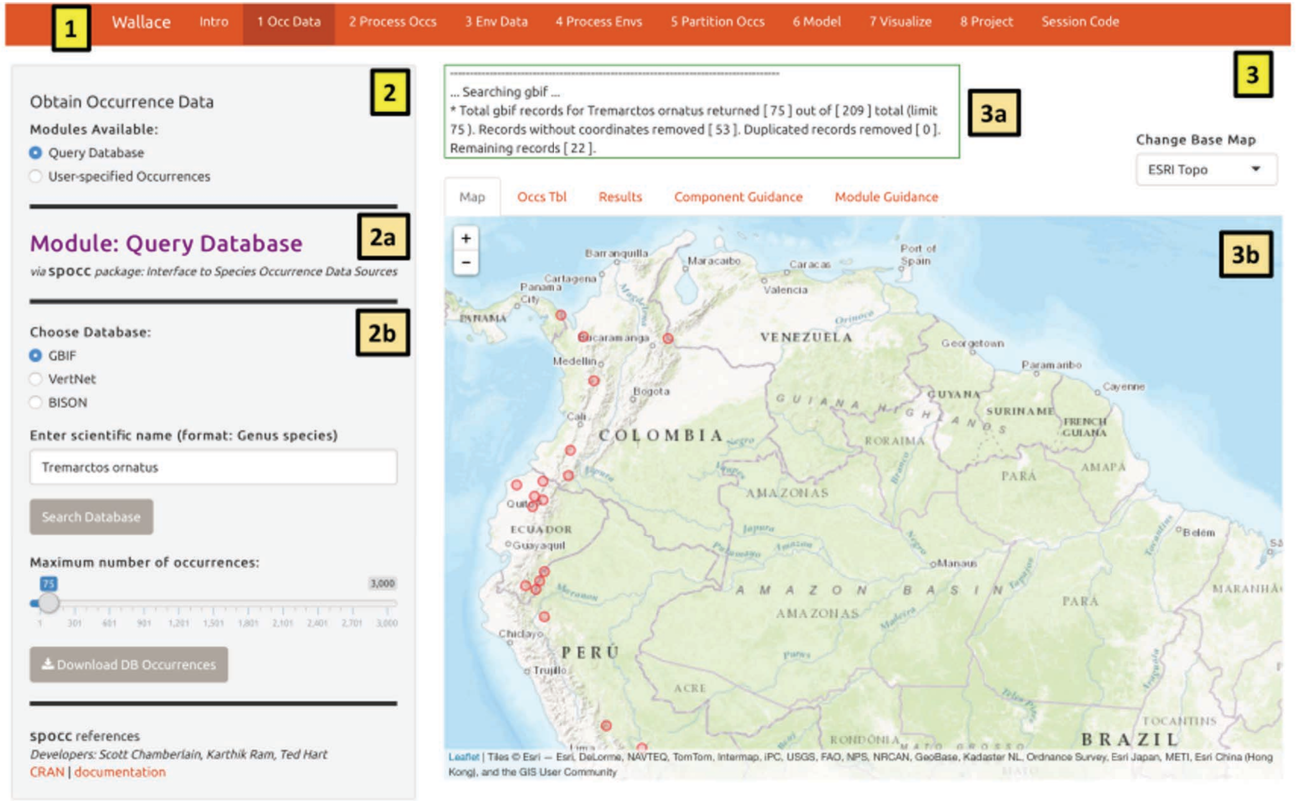
Anderson 2012; Kass et al. 2017

# Wallace: An Ecological Modeling Application

A **user-friendly** application for species distribution modeling that provides **guidance** towards following **best-practices** at each step.

**Wallace** is:

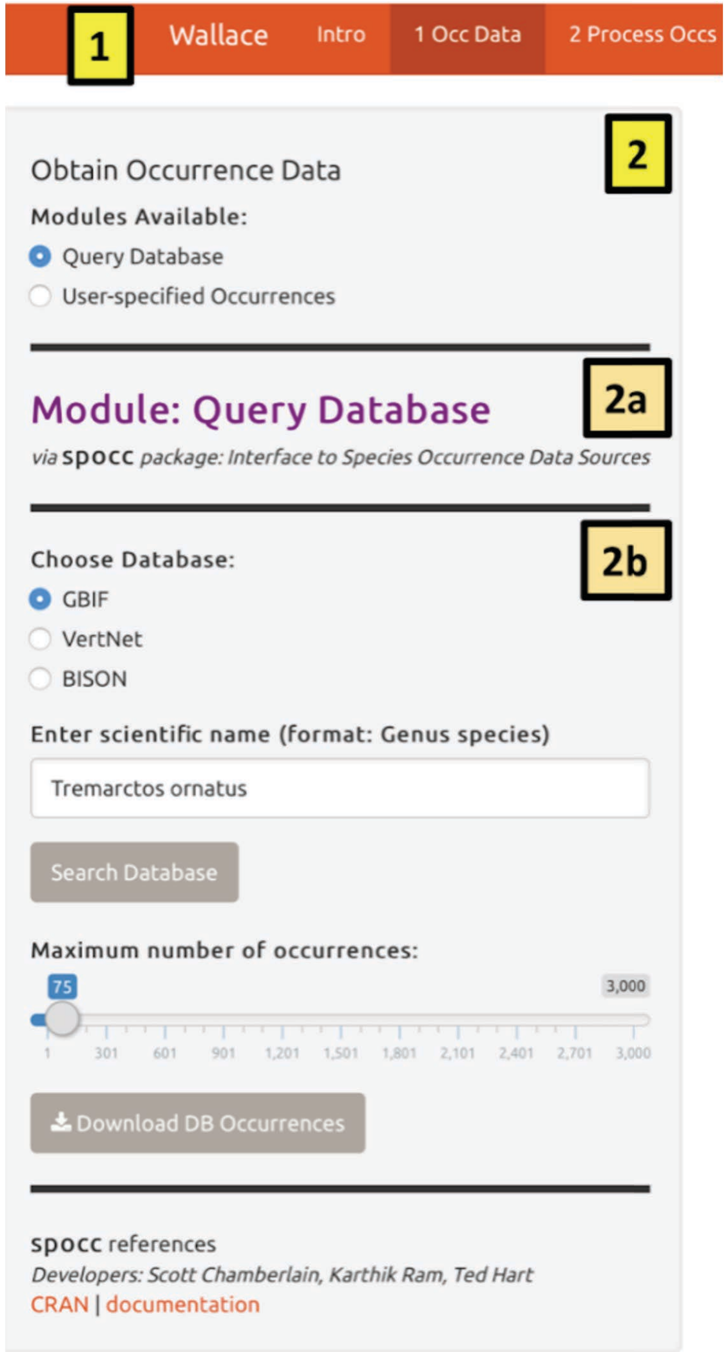
- **Accessible**
- **Instructive**
- **Flexible**
- **Interactive**
- **Reproducible**
- **Expandable**
- **Open**



# `Wallace` is:

## Accessible

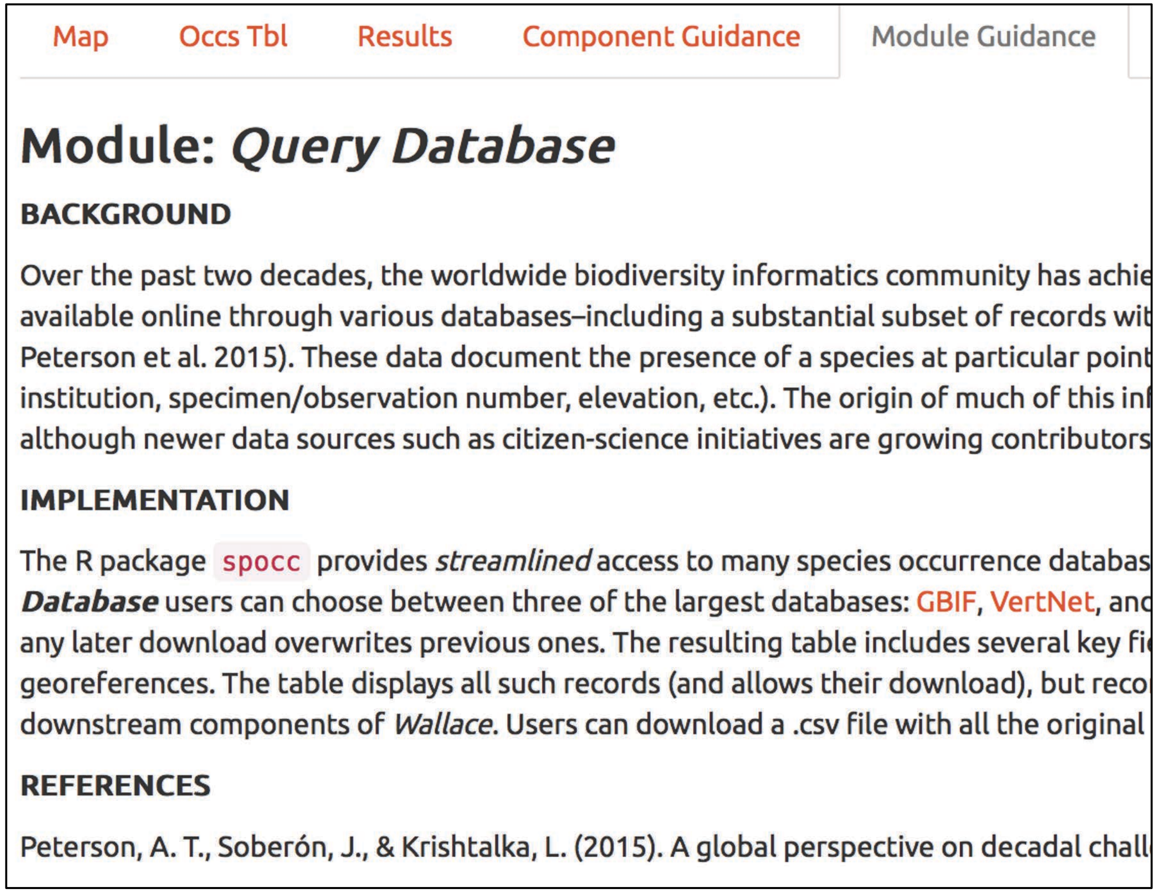
- Lowers barriers to implement cutting-edge SDM techniques
- Allows users to download occurrence & environmental data from diverse sources
- Users can find support from the community through our various networks
  - Google Groups
  - GitHub



# Wallace is:

## Instructive

- Provides educational & instructional resources addressing:
  - Concepts/Theories
  - Methods
- Guides users in following best-practices

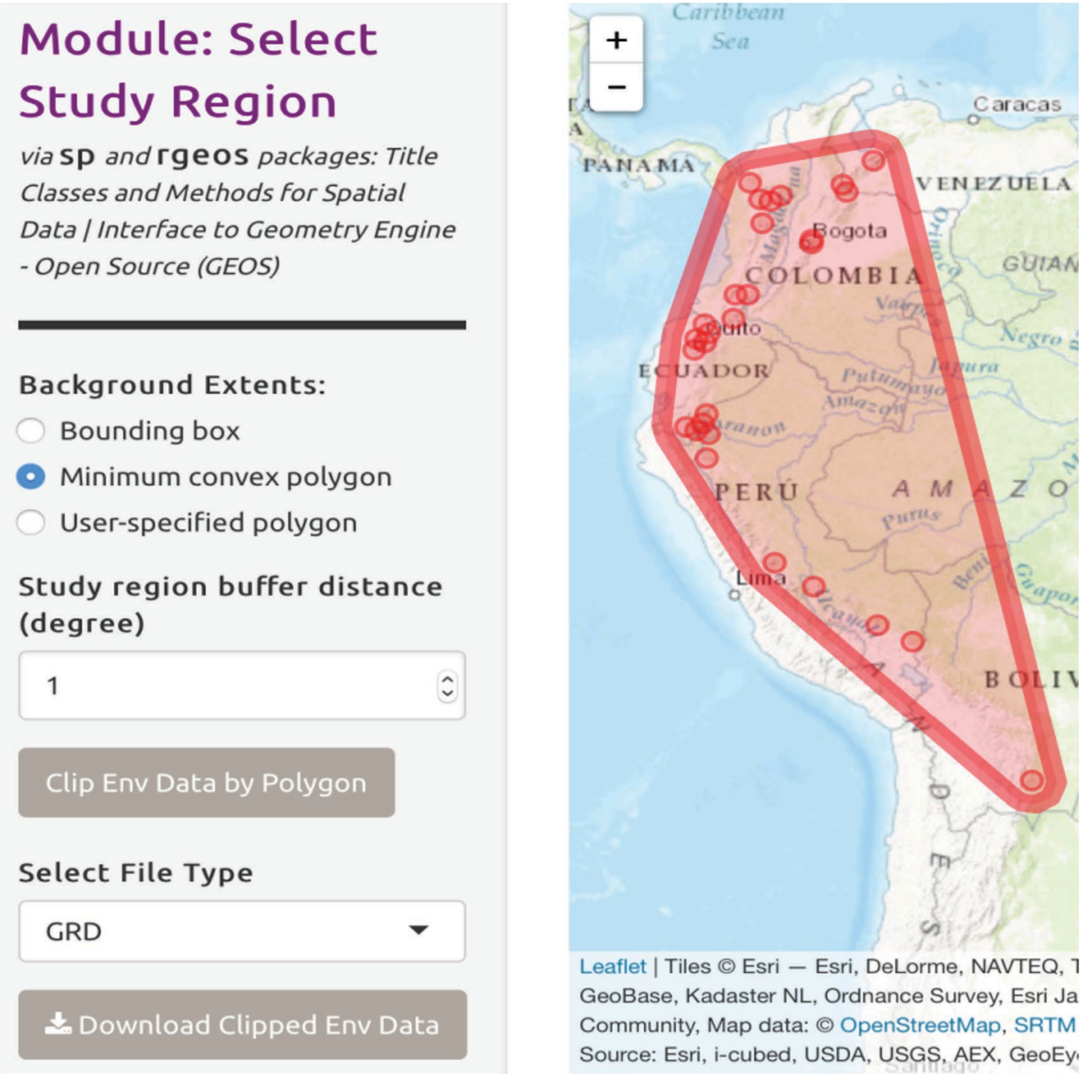


# `Wallace` is:

## Flexible

- Allows users to select from multiple:
  - Data Sources
  - Analytical Tools

- Designed to fit the user's needs

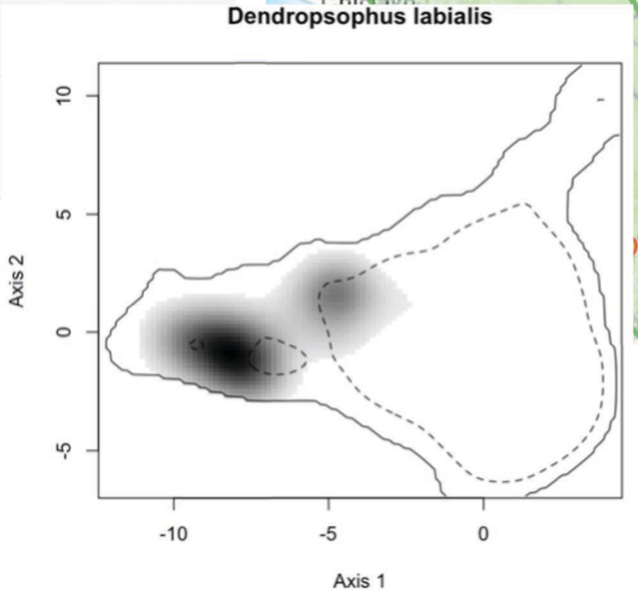


# **Wallace** is:

## Interactive

- Provides an assortment of dynamic visualizations to explore data and results
  - Maps
  - Tables
  - Graphs

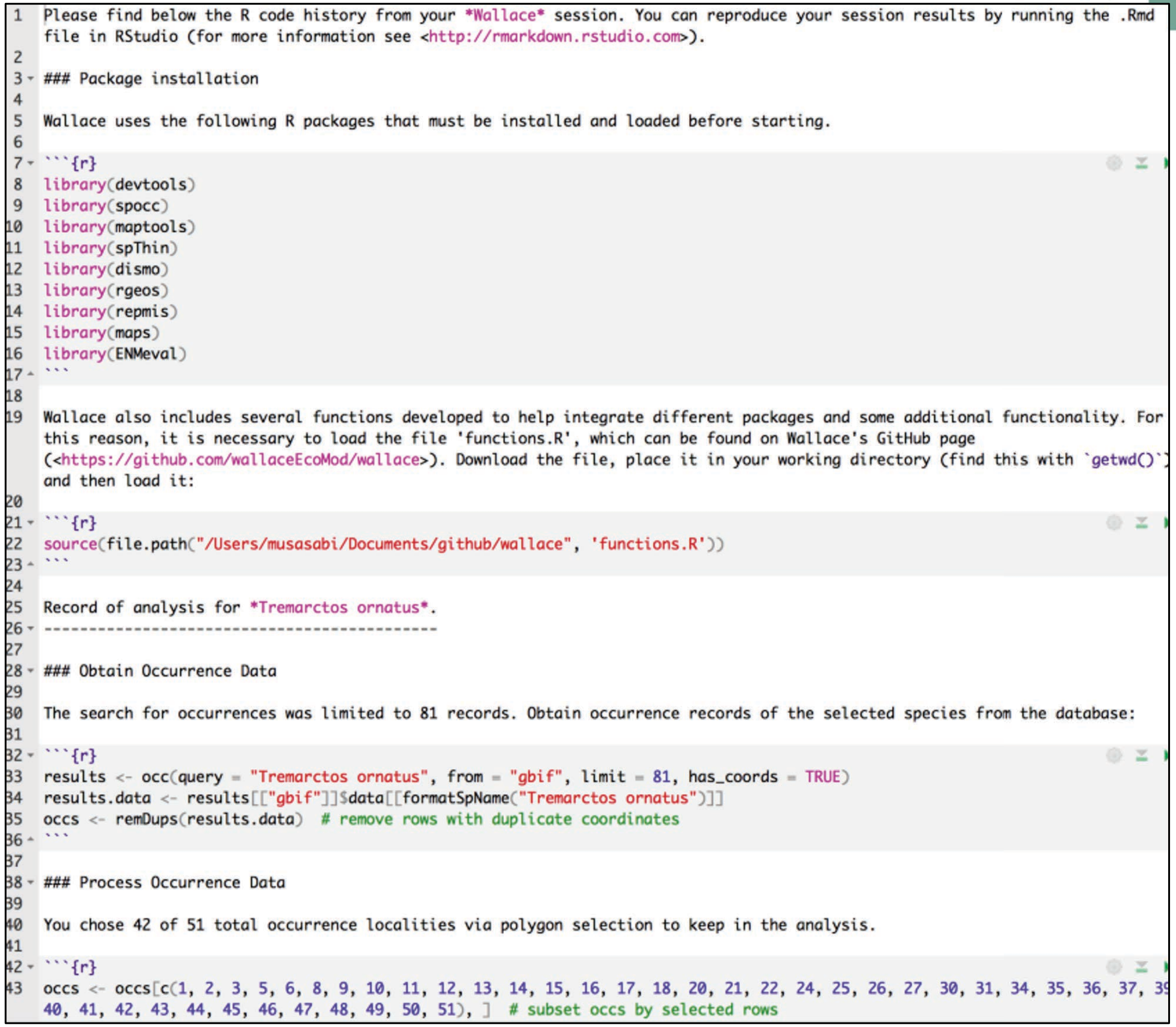

| occID | scientific_name | longitude | latitude | country |
|---|---|---|---|---|
| 1 | Dendropsophus labialis (Peters, 1863) | -73.87 | 4.77 | Colombia |
| 2 | Dendropsophus labialis (Peters, 1863) | -73.85 | | |
| 4 | Dendropsophus labialis (Peters, 1863) | -73.85 | | |
| 7 | Dendropsophus labialis (Peters, 1863) | -73.81 | | |



# Wallace is:

## Reproducible

- Allows users to:
  - Download executable code
  - Replicate analyses
  - Save work and load later (no need to redo steps)



```
Please find below the R code history from your *Wallace* session. You can reproduce your session results by running the .Rmd
file in RStudio (for more information see <http://rmarkdown.rstudio.com>).

### Package installation

Wallace uses the following R packages that must be installed and loaded before starting.

```{r}
library(devtools)
library(spocc)
library(maptools)
library(spThin)
library(dismo)
library(rgeos)
library(repmis)
library(maps)
library(ENMeval)
```

Wallace also includes several functions developed to help integrate different packages and some additional functionality. For
this reason, it is necessary to load the file 'functions.R', which can be found on Wallace's GitHub page
(<https://github.com/wallaceEcoMod/wallace>). Download the file, place it in your working directory (find this with `getwd()`)
and then load it:

```{r}
source(file.path("/Users/musasabi/Documents/github/wallace", 'functions.R'))
```

Record of analysis for *Tremarctos ornatus*.
--------------------------------------------

### Obtain Occurrence Data

The search for occurrences was limited to 81 records. Obtain occurrence records of the selected species from the database:

```{r}
results <- occ(query = "Tremarctos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
results.data <- results[["gbif"]]$data[[formatSpName("Tremarctos ornatus")]]
occs <- remDups(results.data)  # remove rows with duplicate coordinates
```

### Process Occurrence Data

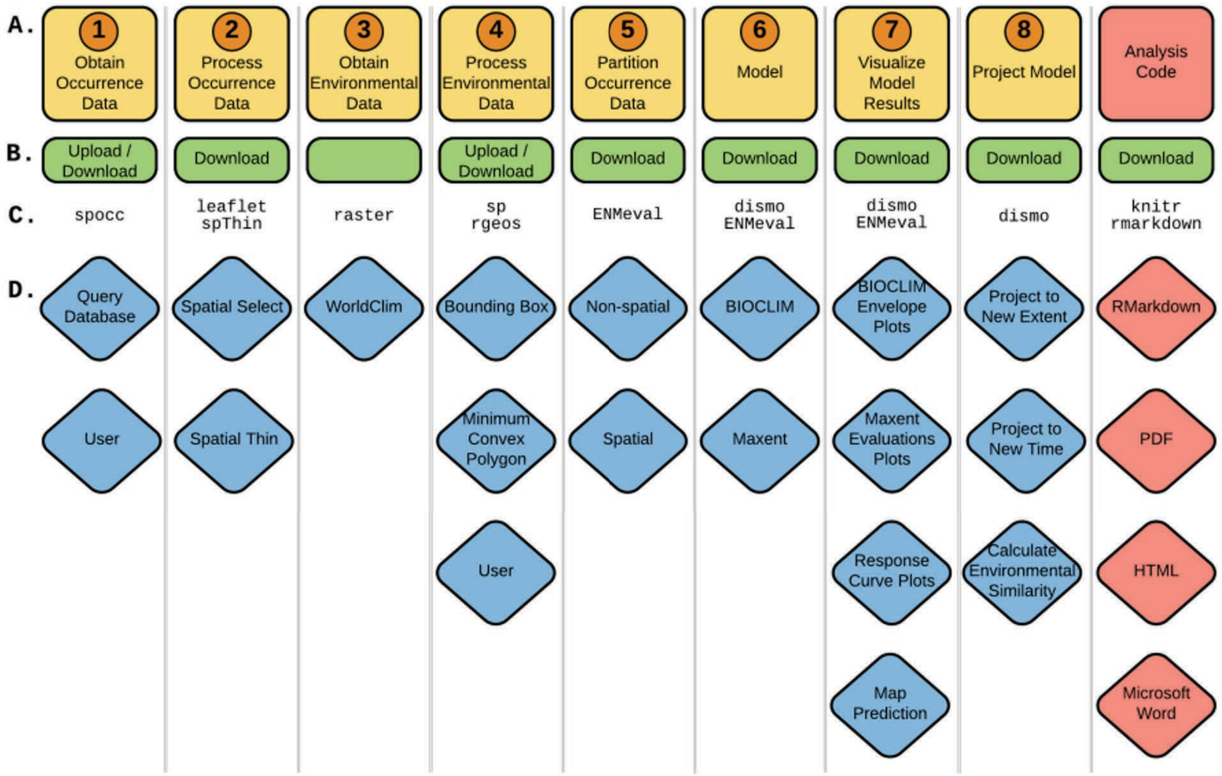
You chose 42 of 51 total occurrence localities via polygon selection to keep in the analysis.

```{r}
occs <- occs[c(1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, 26, 27, 30, 31, 34, 35, 36, 37, 3
40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51), ]  # subset occs by selected rows
```
```

# `Wallace` is:

## Expandable

- Users can contribute new functions & analytical tools to be integrated as:
  - Components (A)
  - Modules (D)

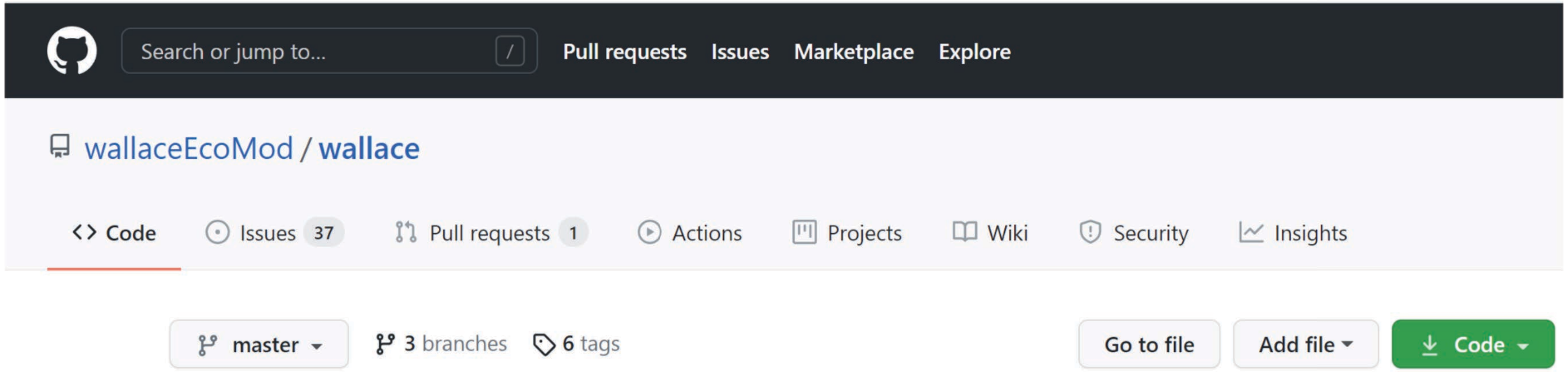


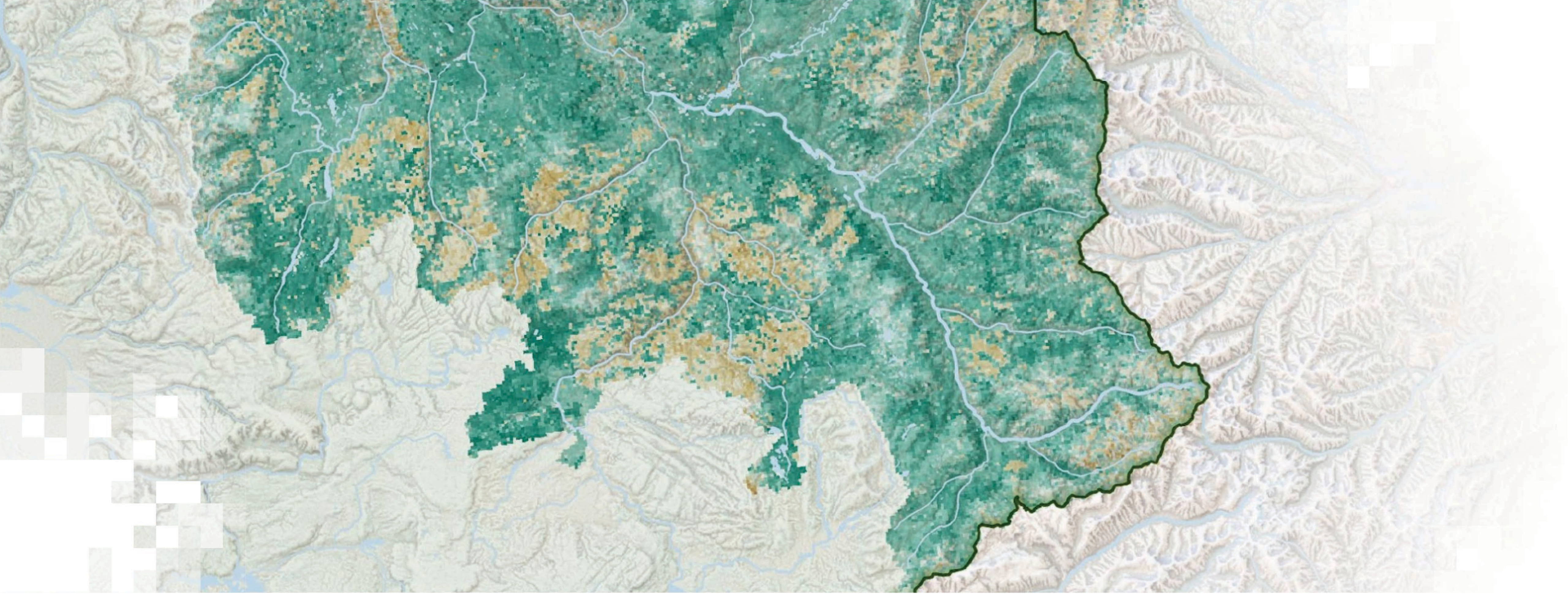
# Wallace is:

## Open

- Code is publicly available for users to:
  - Download (CRAN, GitHub)
  - Modify & suggest enhancements

- Enables users to become contributors



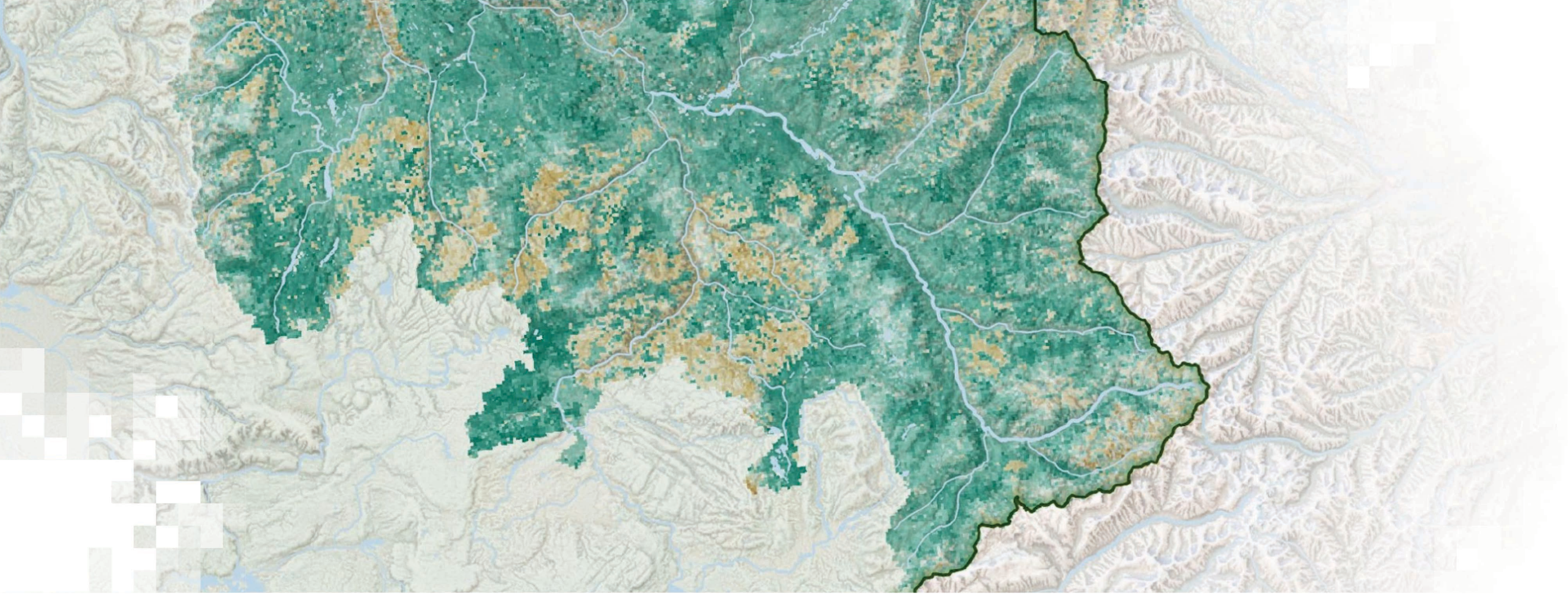


# Wallace Walkthrough: Running a Full SDM Workflow

# Install and Run `Wallace`

1. Install R (version ≥ v.3.5.0) and RStudio
2. Install the **`Wallace`** package from CRAN (v.1.1; stable)
3. Run **`Wallace`** in RStudio

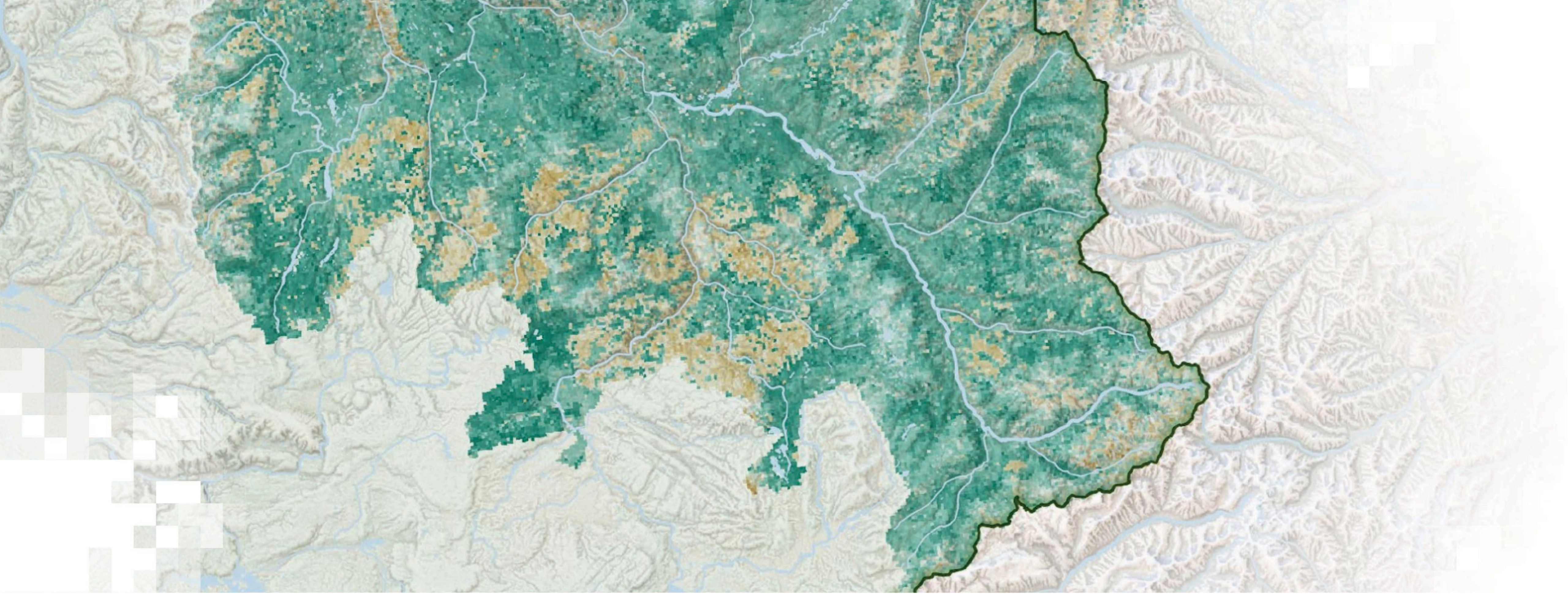
```r
# install the package
install.packages('wallace')
# load the package
library(wallace)
# run the app
run_wallace()
```



# Redesign & Expansion: Making it Easier to Visualize Results and Add New Modules

# Wallace V2 Coming Soon

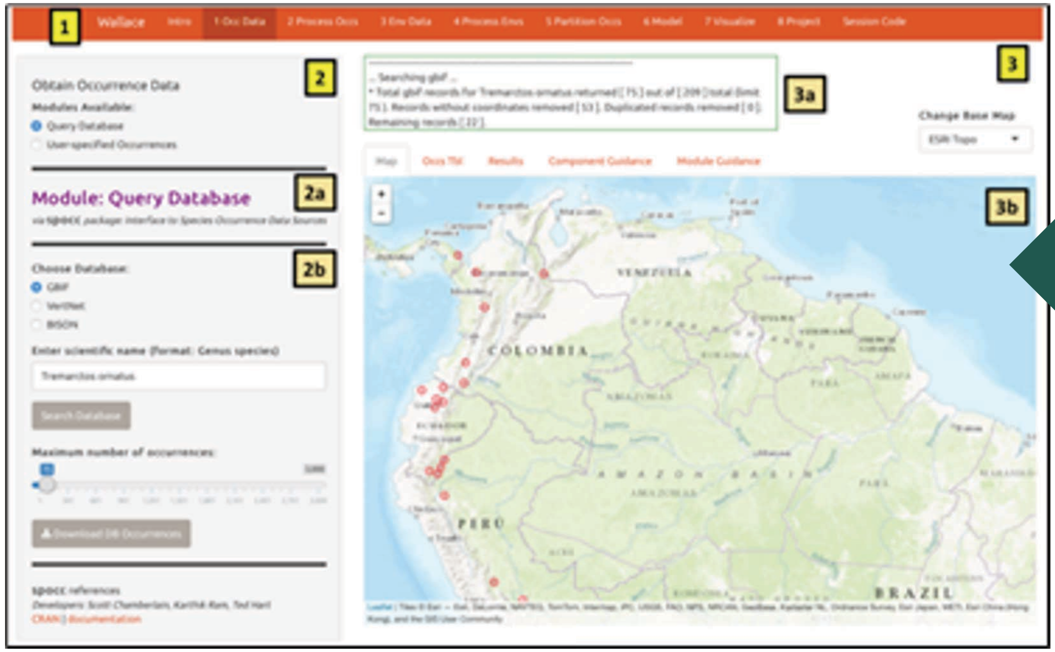
- **New modules** added with help from collaborators around the world
  - More data sources
  - More analytical features
- Re-designed to make **module additions easier**
- Now based on RMMS **metadata** (Merow et al. 2019)
- Development guided by **user feedback**:
  - Conference workshops, emails, Google group



# Extensions: SDM Post-Processing for Conservation Decision-Making and Beyond

# Wallace V3 Partnership for Conservation

## Wallace

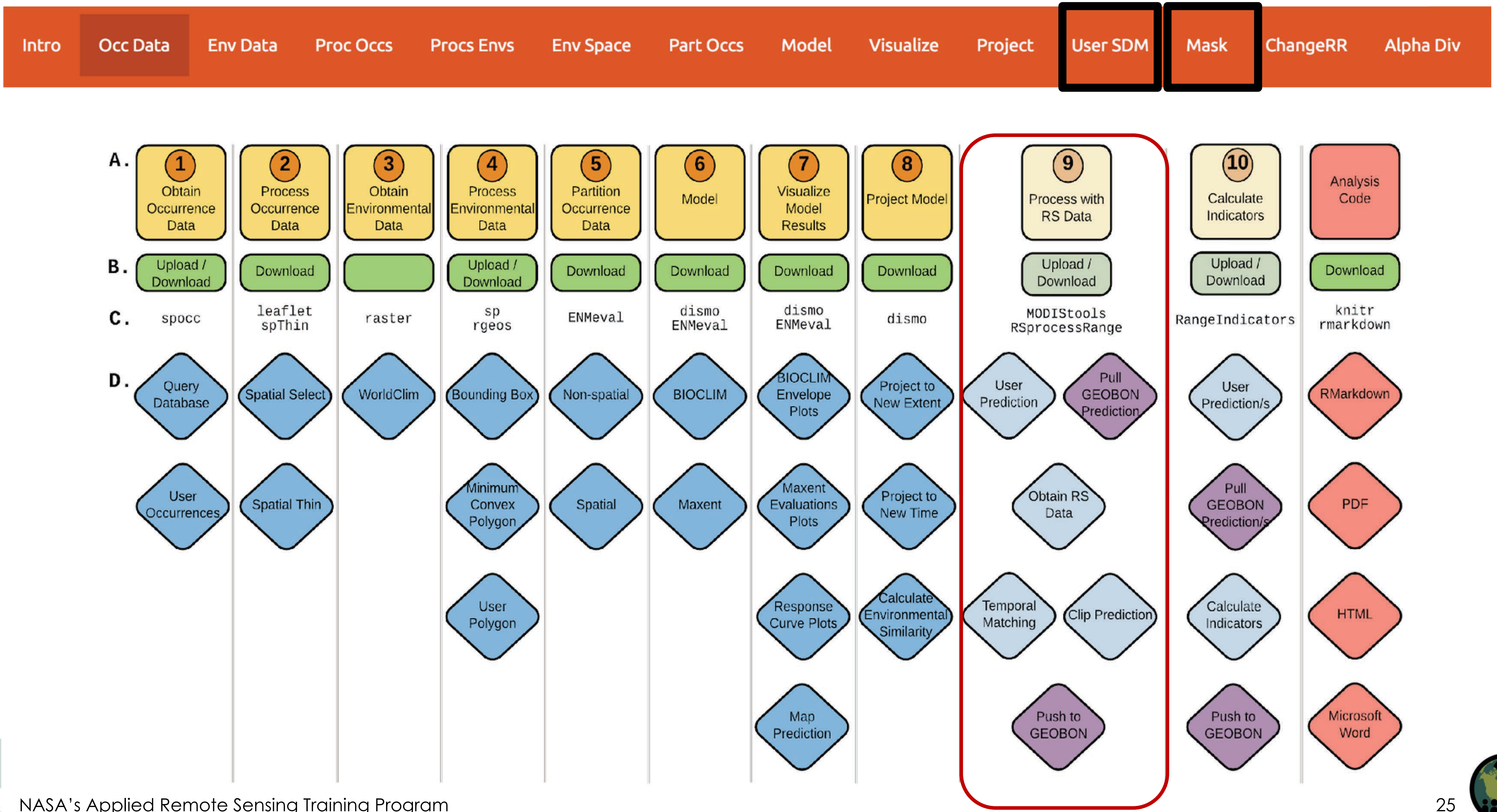




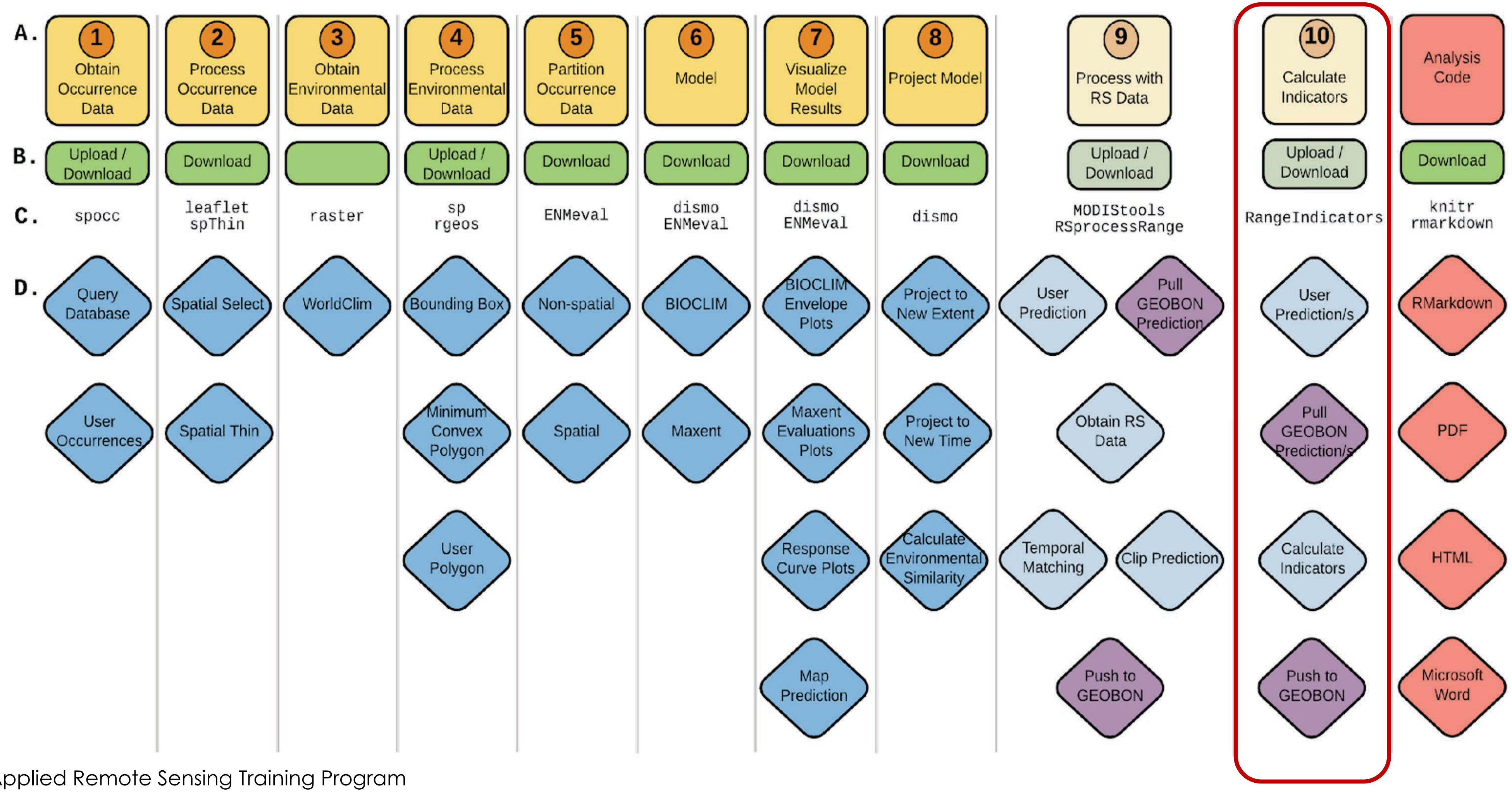
GEO BON BON in a Box

Galante et al. *in prep.*

# New Components Added to Wallace (V3)

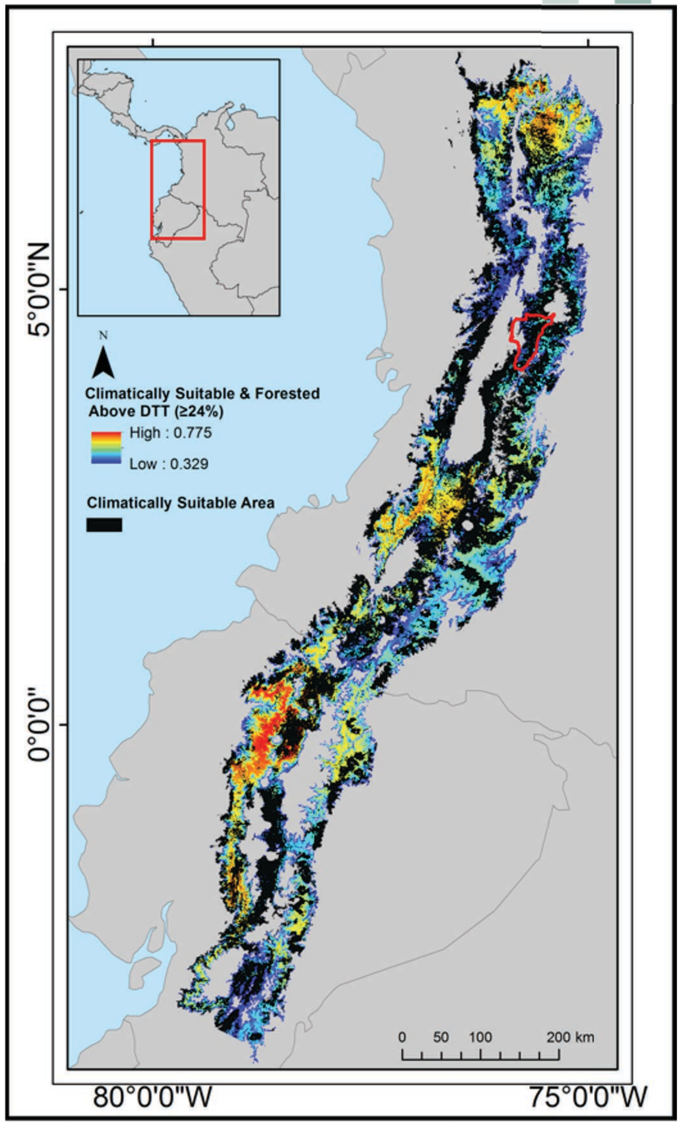


# New Components Added to Wallace (V3)

# Mask SDM (Based on maskRangeR)

- Post-processes SDMs to estimate current ranges:
  - RS Products
    - E.g., Forest Cover, Urbanization, Cultivars
  - User-Defined Polygons
    - E.g., Protected Areas, Land Cover
  - Expert-Defined Maps/Ranges




Merow et al. *in rev.*

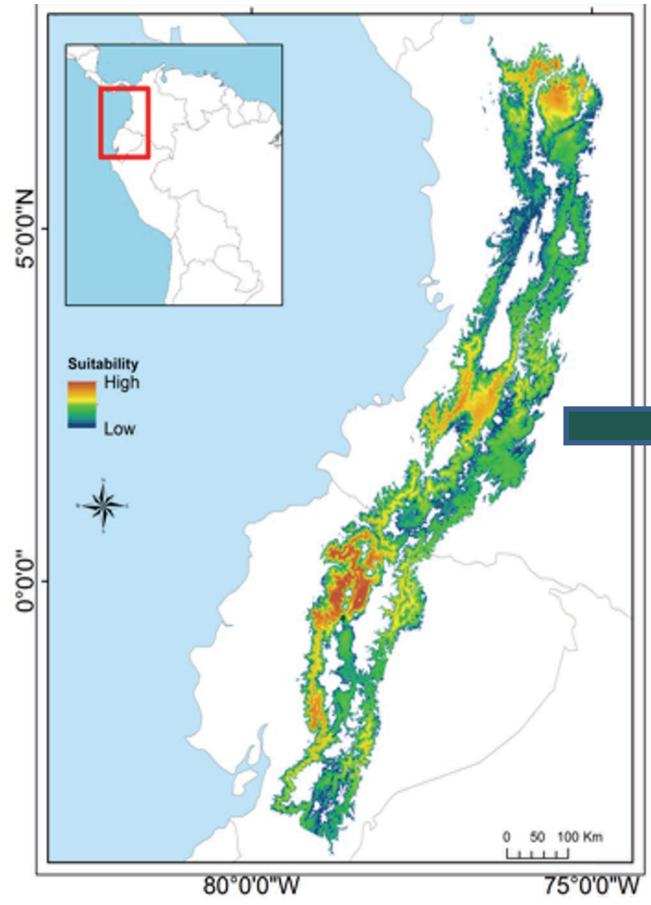
# Mask Example: Olinguito (*Bassaricyon neblina*)

- Recently described small carnivore
- Limited to high-altitude cloud forests
- Data-poor
- Needs IUCN status update given recent deforestation


Merow et al. *In rev.*

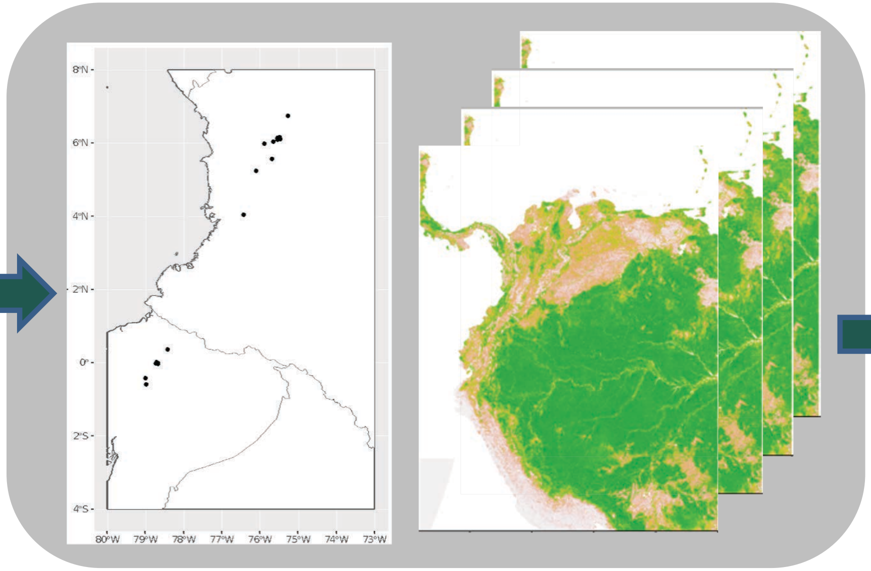
# Mask Example: Olinguito (*Bassaricyon neblina*)

## SDM

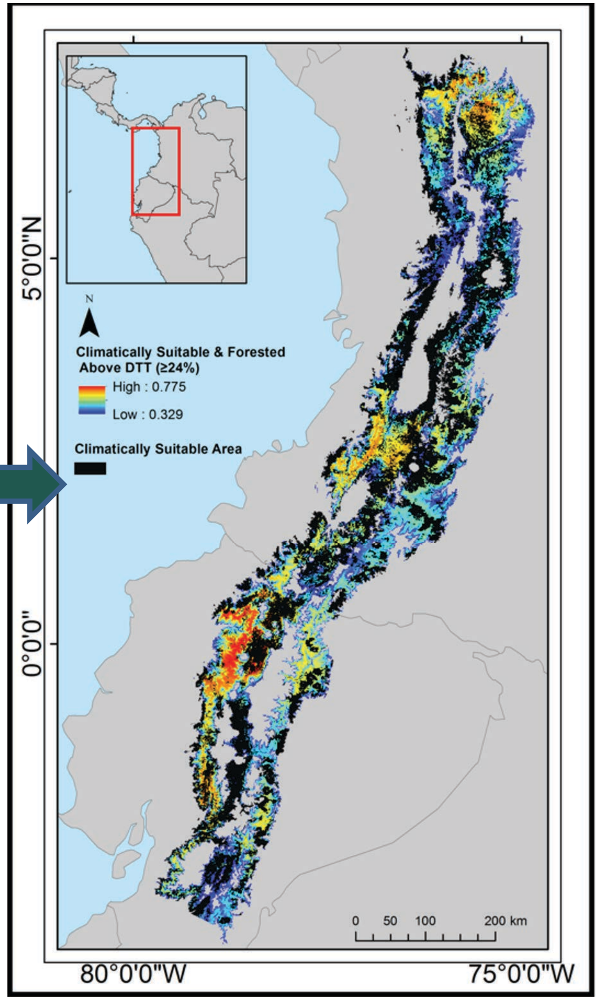



Recent Occurrences

MODIS Yearly Forest Cover



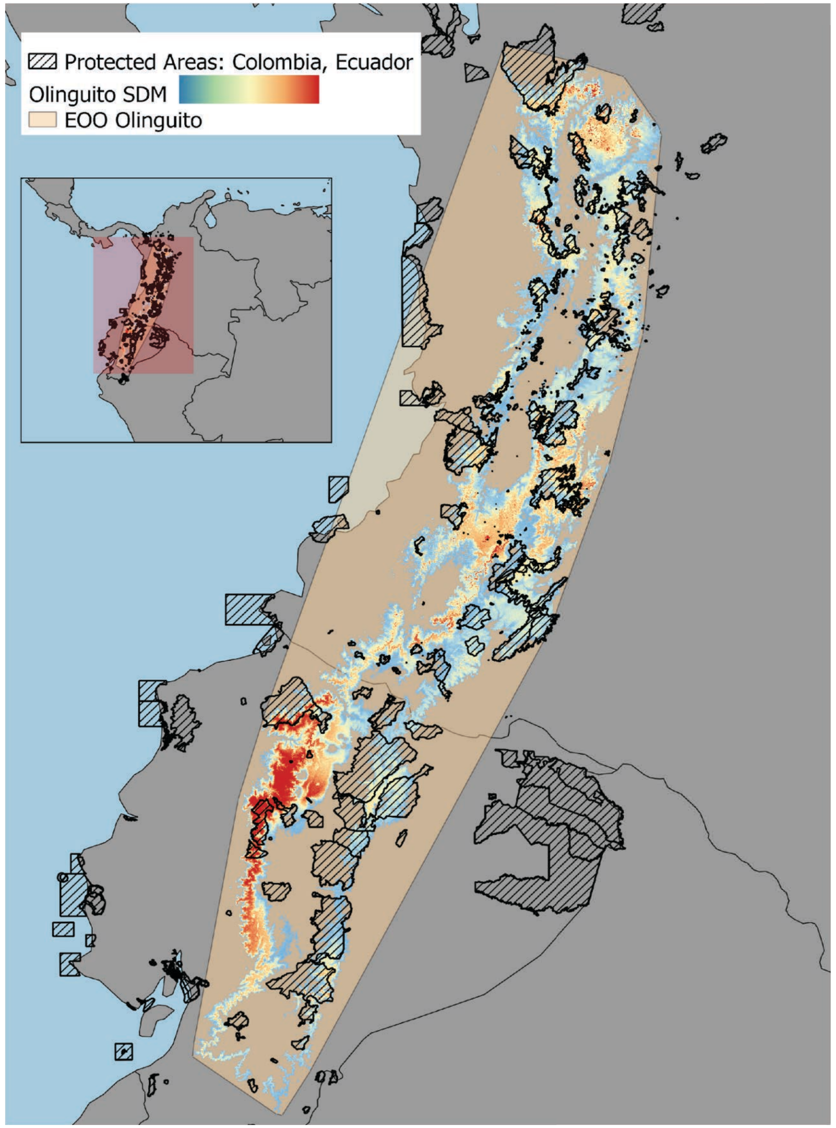

## Masked Distribution




Gavrutenko et al. 2021, Merow et al. *in rev.*

# Calculate Indicators (Based on changeRangeR)

- Calculate key biodiversity change indicators
  - IUCN AOO & EOO
  - Percentage of Suitable Land Cover
  - Protected Area Representativeness




Galante et al. *in prep.*

# Calculate Indicators Example: Olinguito (*Bassaricyon neblina*)

- Changes in Range, AOO, EOO given GIS data:
  - Changes over time/space
  - E.g., deforestation over time

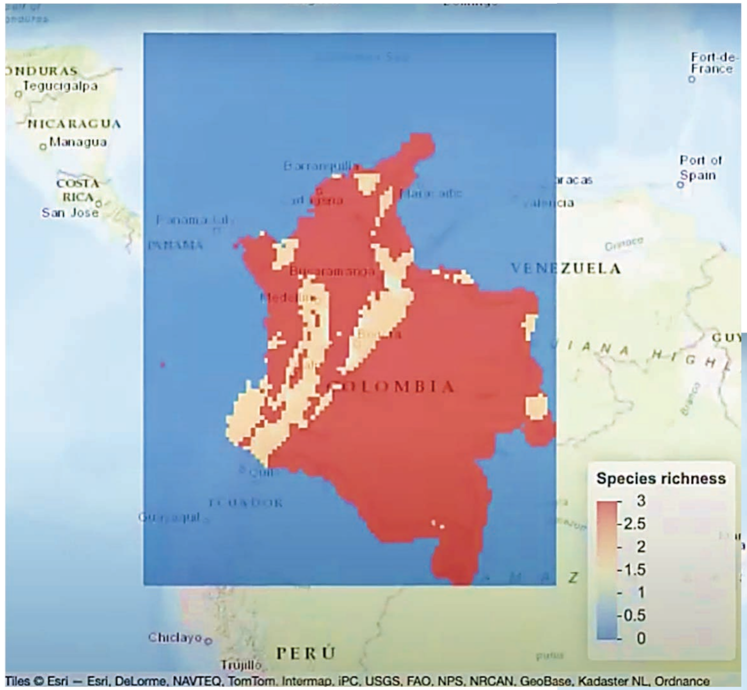

Galante et al. *in prep.*

# Calculate Indicators Example: Colombian Primates

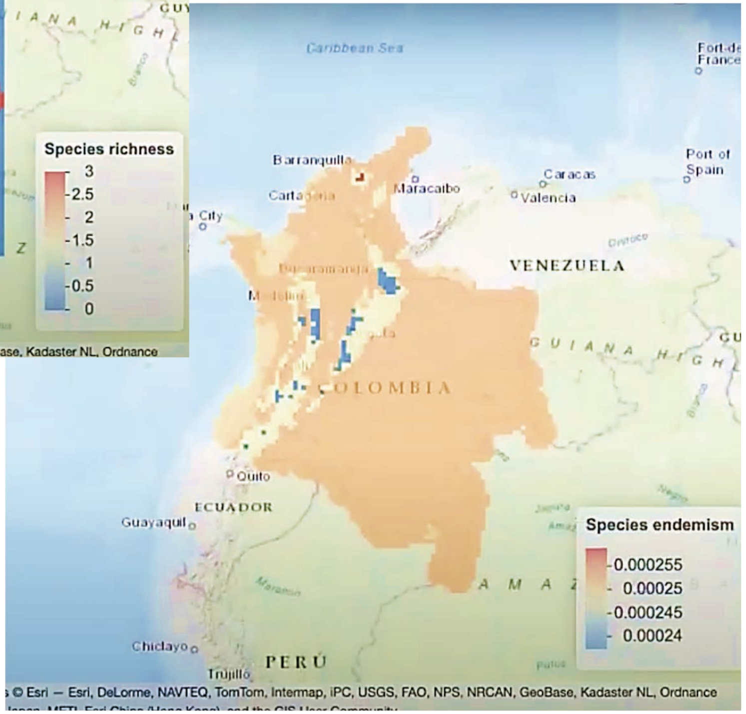
- Multispecies Diversity Metrics:
  - Species Richness
  - Species Endemism

- Soon in **`Wallace`**:
  - Phylogenetic Diversity
  - Phylogenetic Endemism
  - Complementarity

Richness




Endemism




Galante et al. *in prep.*

# Summary

- **`Wallace`** is a **user-friendly** application for species distribution modeling that provides **guidance** towards following **best-practices** at each step.
  - Accessible, Instructive, Flexible, Interactive, Reproducible, Expandable, Open
- **`Wallace V2`** (coming soon)
  - Provides additional SDM data sources & analytical tools.
  - Facilitates module contributions (from user community)
- **`Wallace V3`** (in development)
  - Added tools for conservation applications
    - Estimate species ranges
    - Calculate biodiversity indicators.
- Next Session: Additional SDM Tools and Techniques

# 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

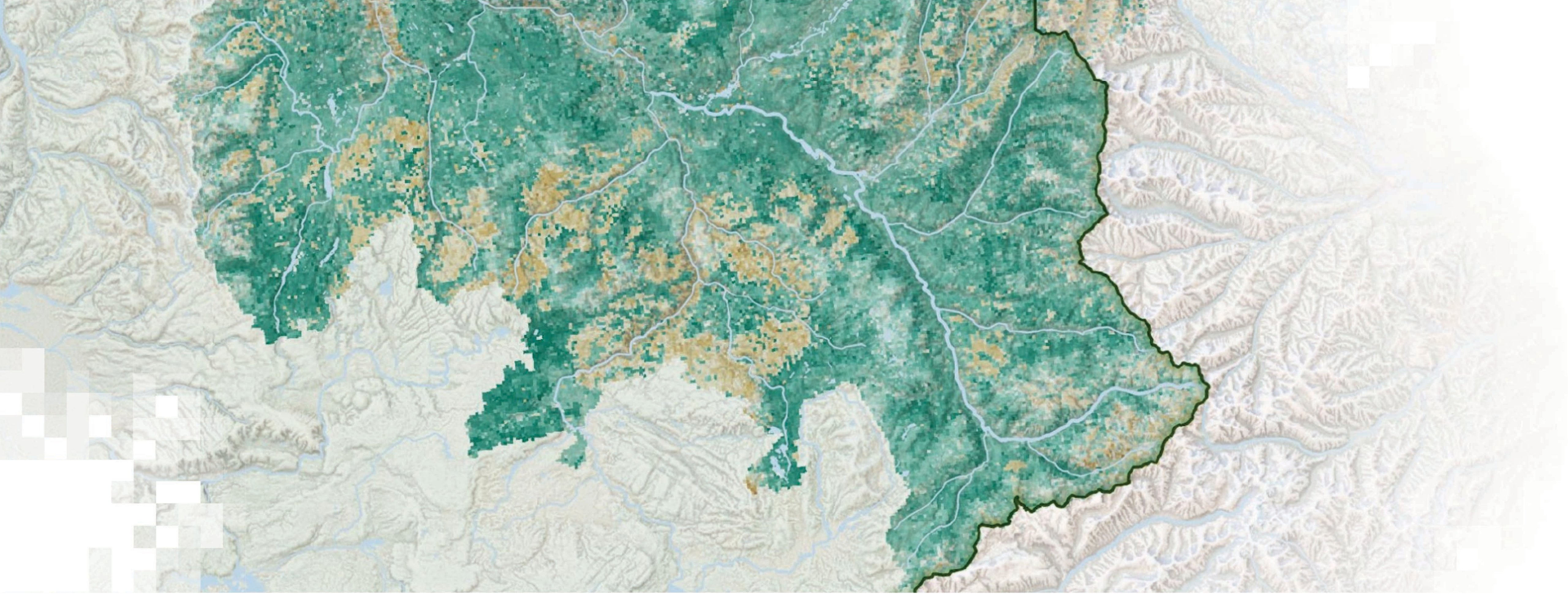
18-GEOBON-0020

DBI 1650241
DBI 1661510

# Thank You!

wallaceecomod.github.io

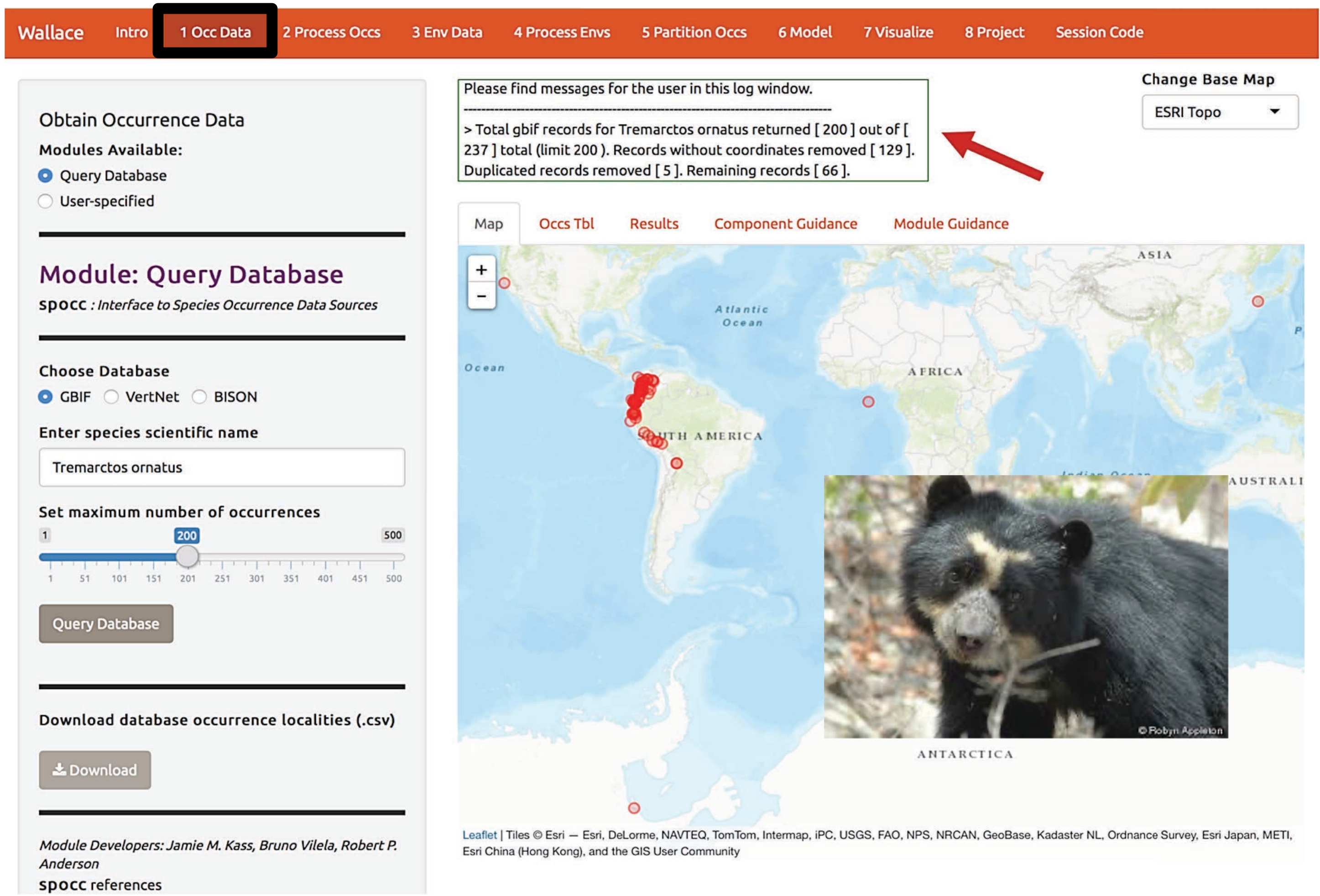
groups.google.com/g/wallaceecomod



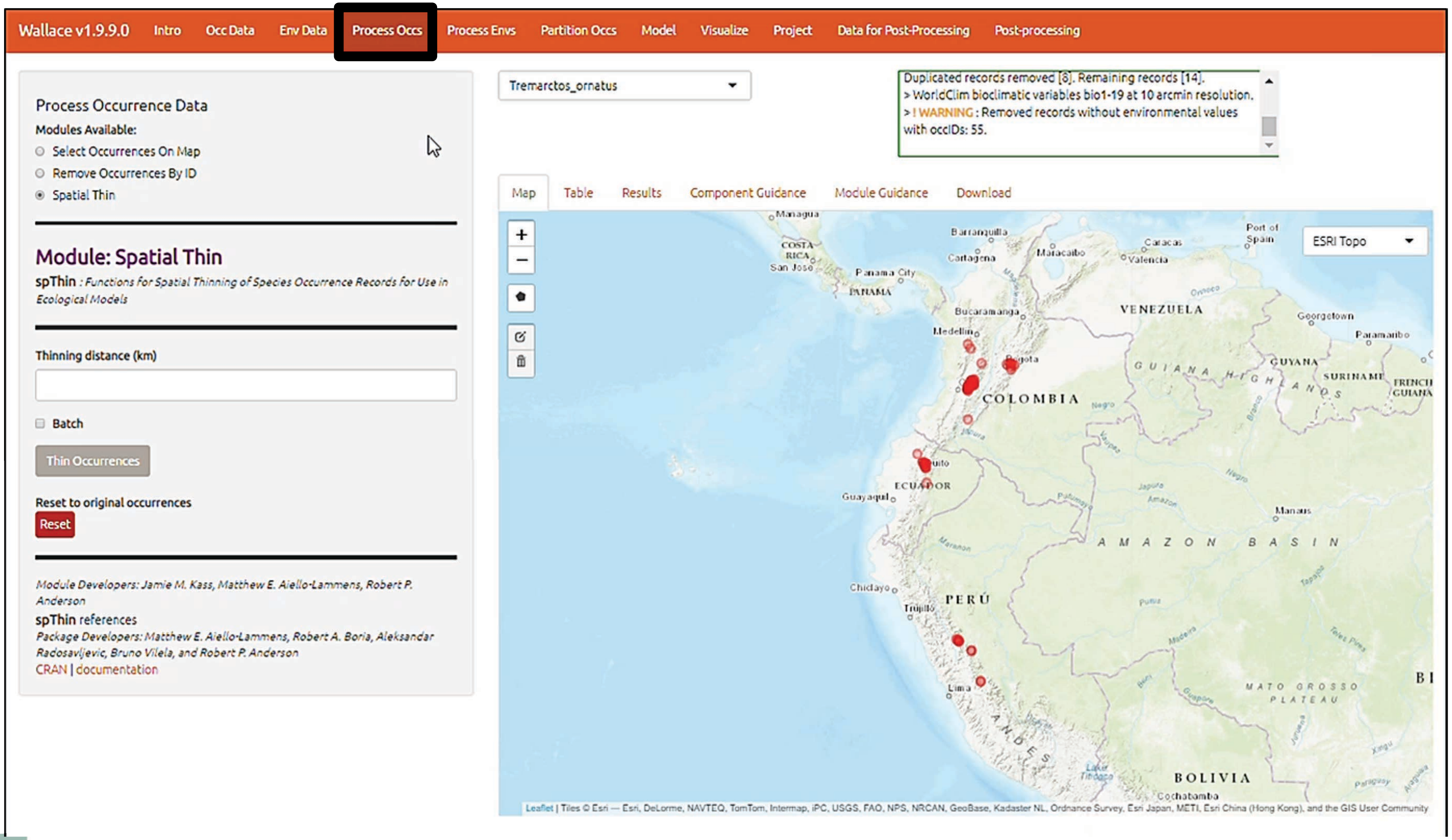
# Supplemental: `Wallace` Walkthrough Slides

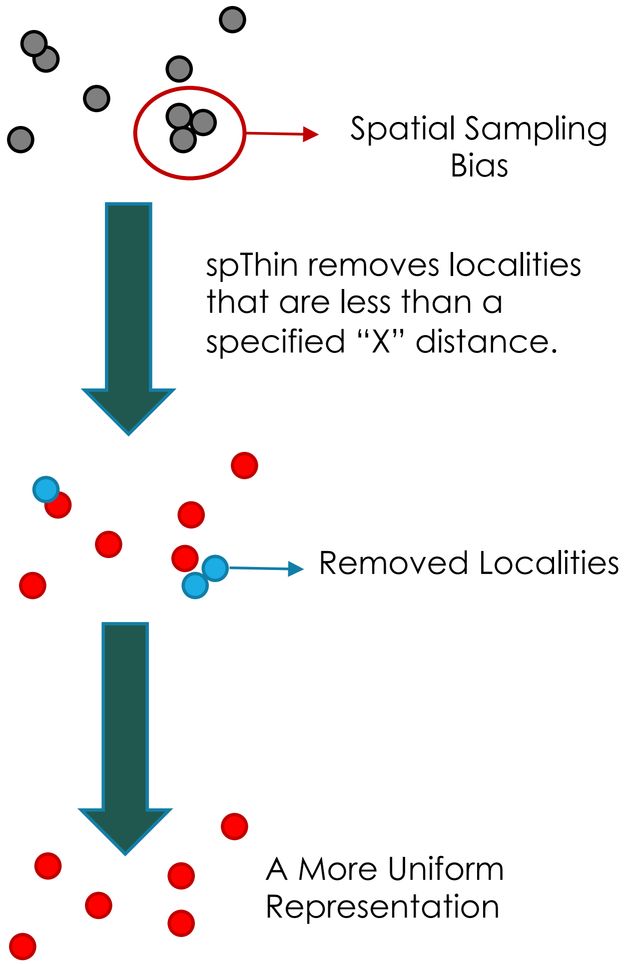
# Component 1: Obtain Presence Data (Occ Data)

- User-specified
- Download from online databases
  - GIBF
  - VertNet
  - USGS BISON



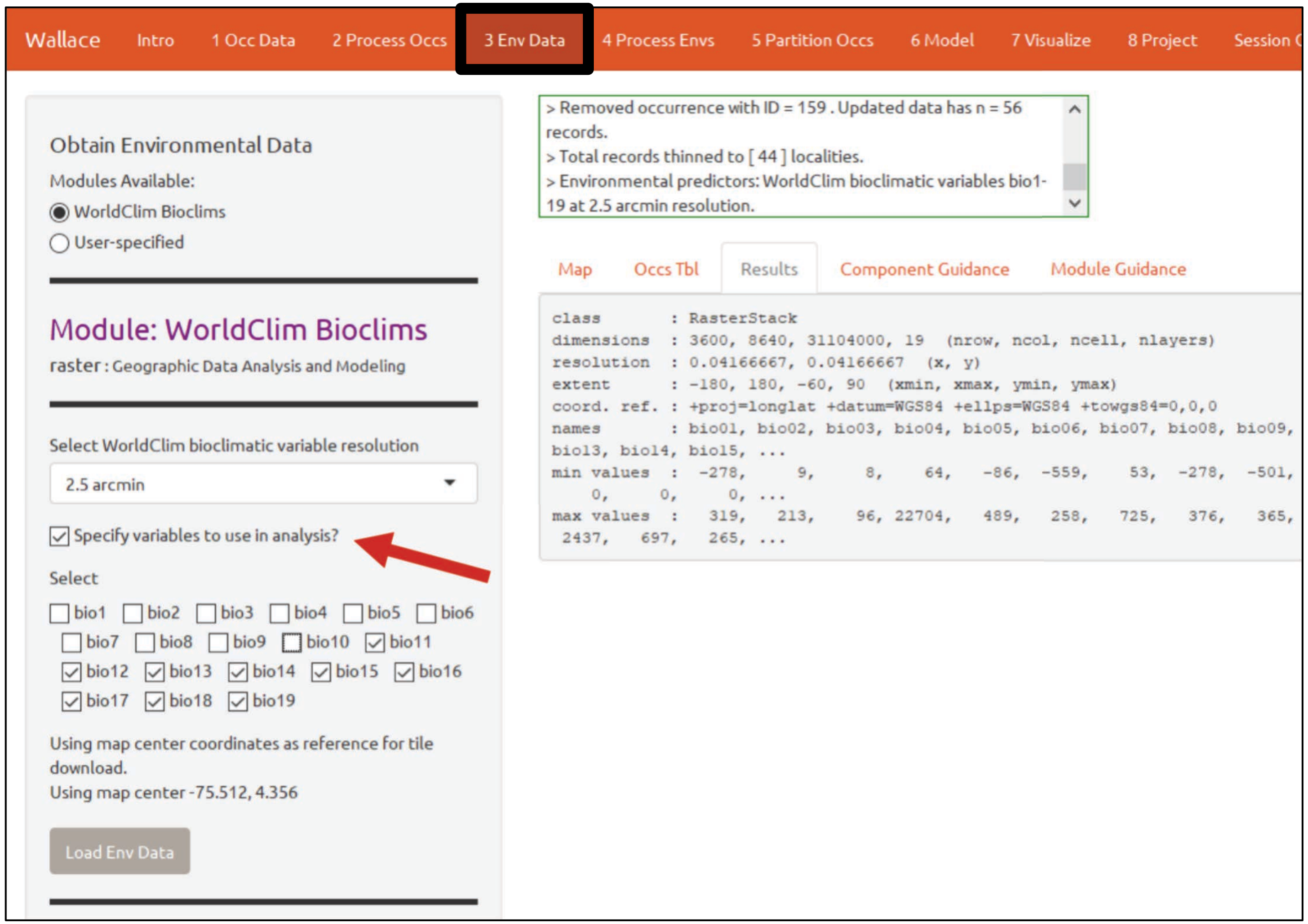
# Component 2: Process Presence Data (Process Occ)



# Component 3: Obtain Environmental Data (Env Data)

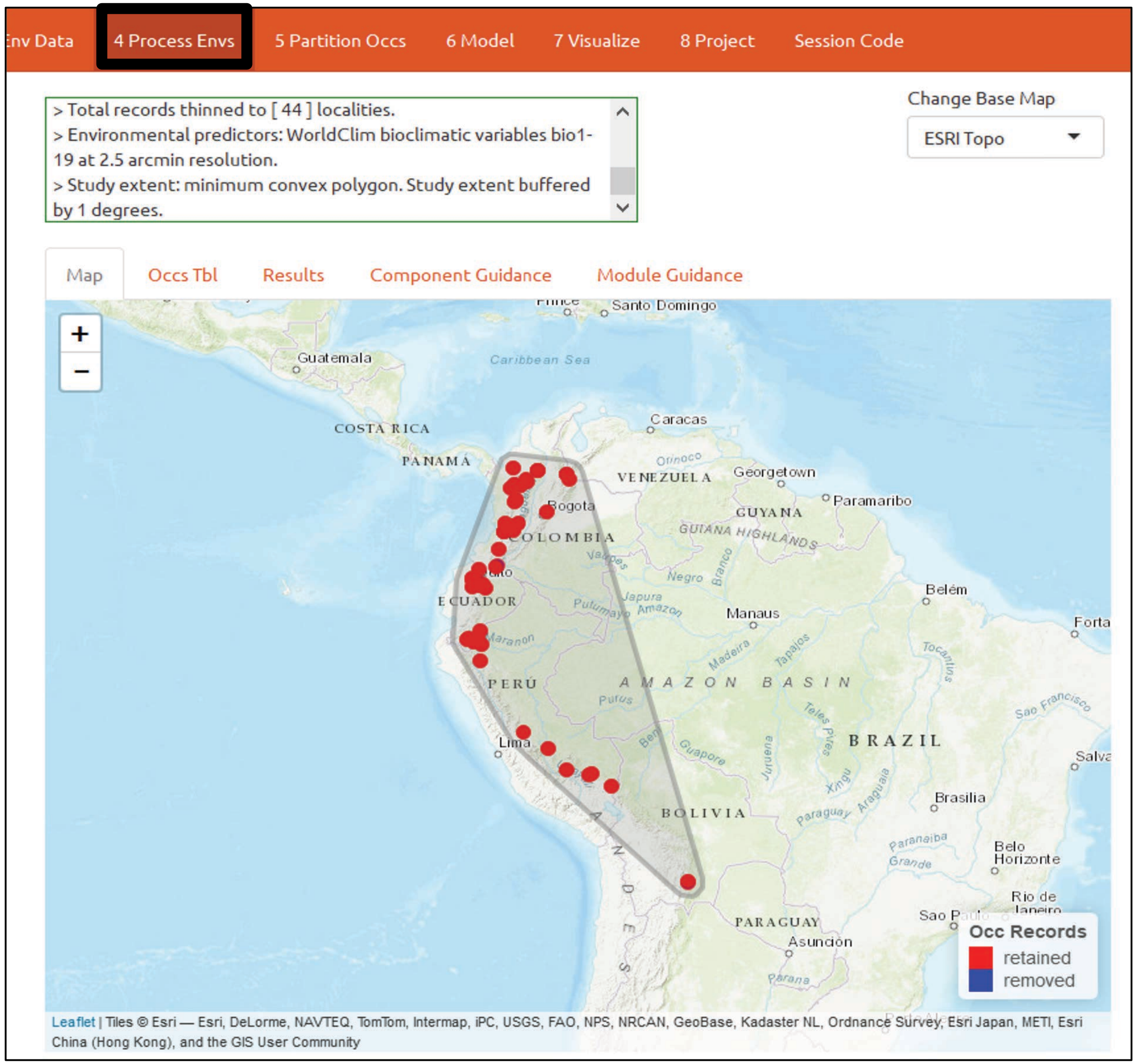
- User-specified
- Download from online databases
  - WorldClim



# Component 4: Process Environmental Data (Process Envs)

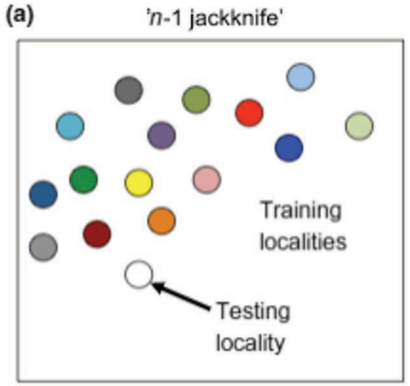
- Select Study Region
  - User-Specified
  - Bounding Box
  - Minimum Convex Polygon
  - Point-Buffers

- Sample Background Points



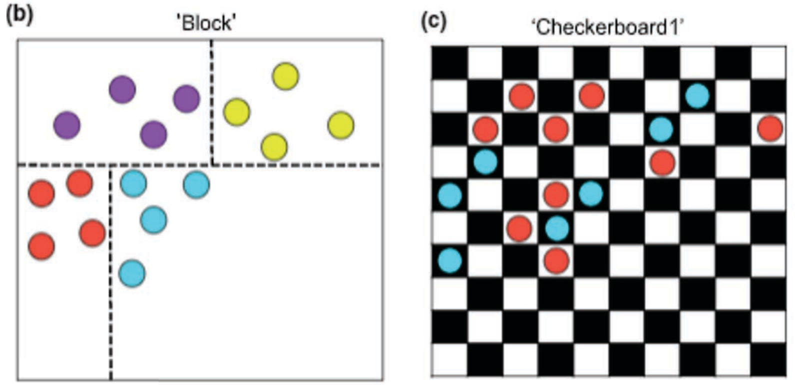
# Component 5: Partition Occurrence Data (Partition Occs)

- Non-Spatial (Random)

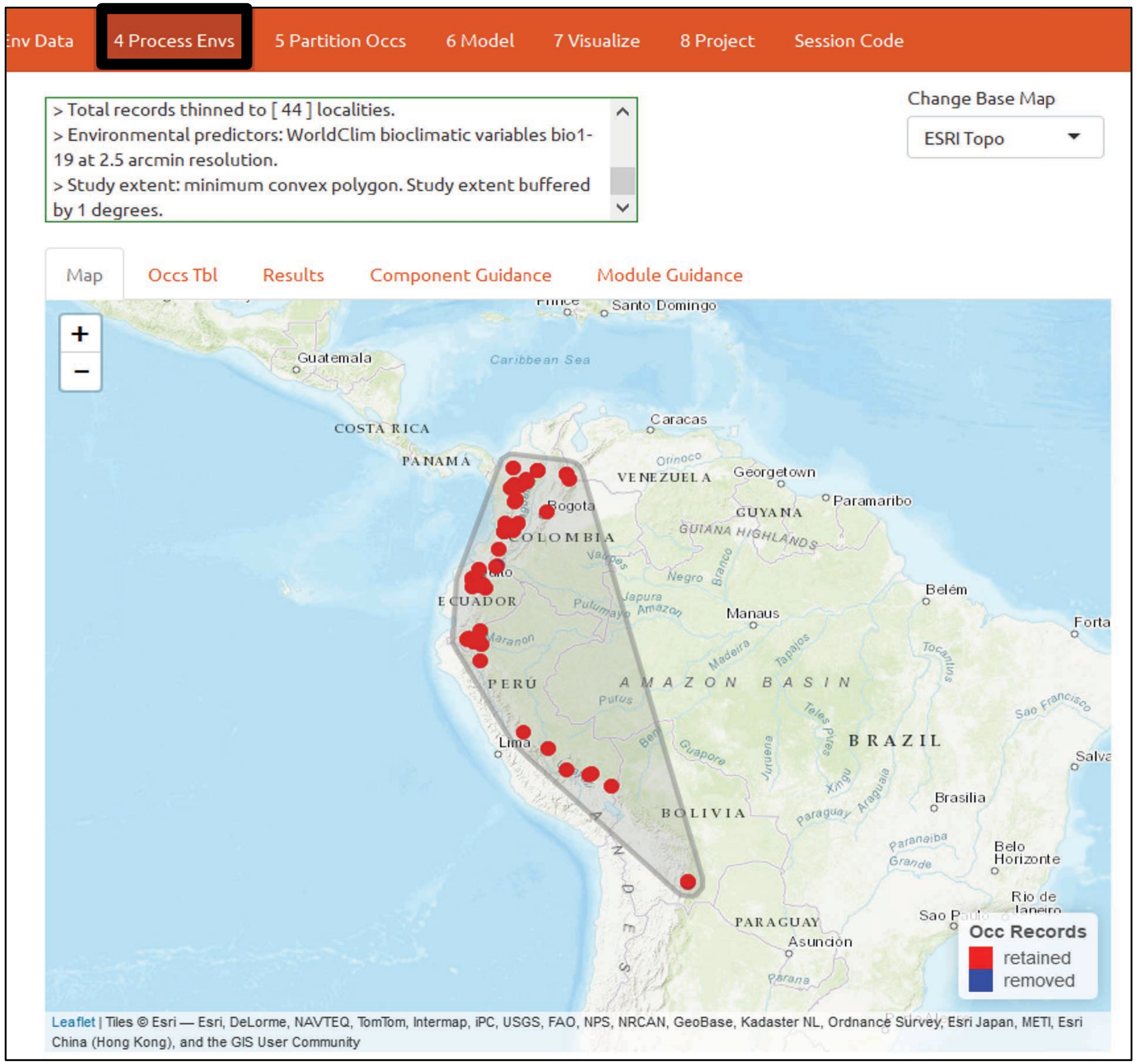



Muscarella et al. 2014

- Spatial



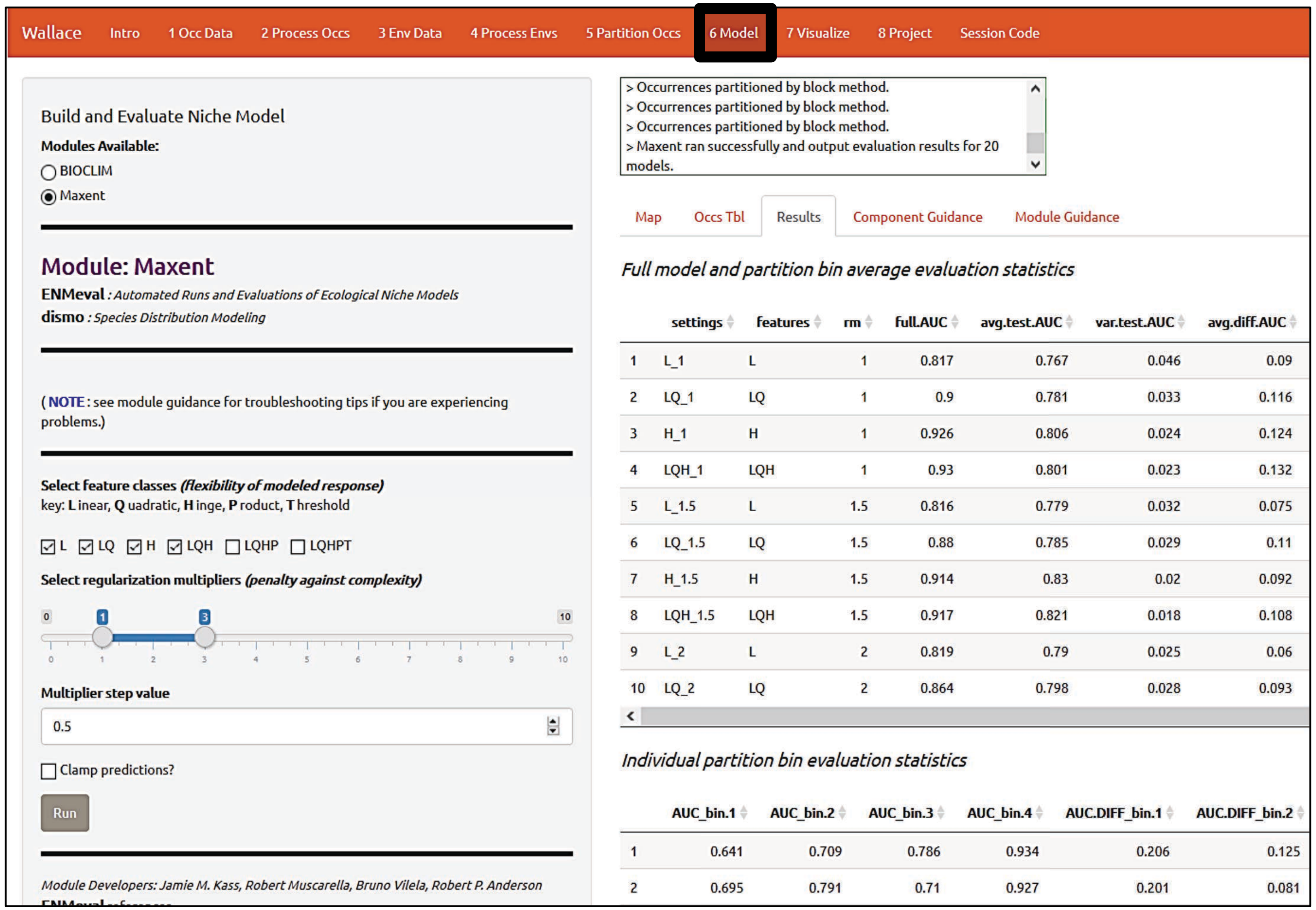

Muscarella et al. 2014




Muscarella et al. 2014

# Component 6: Model Building & Evaluation (Model)

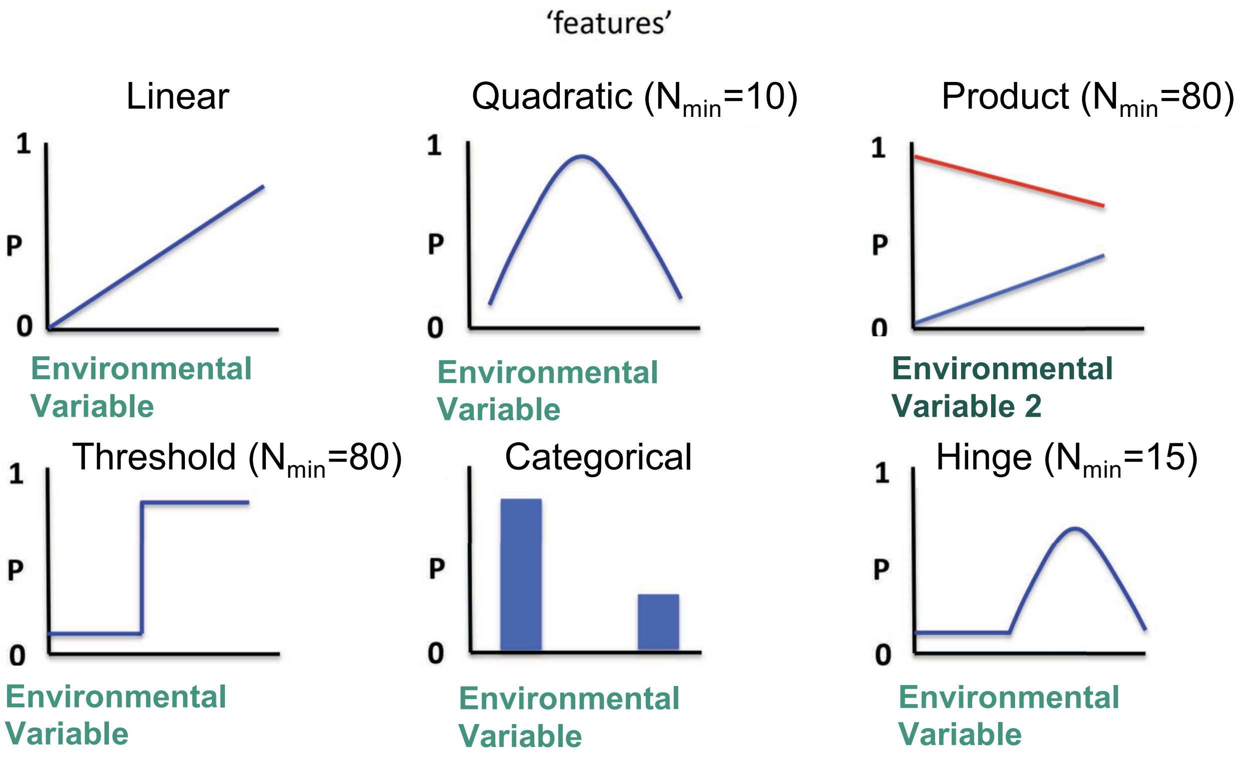
- Modeling Algorithms:
  - Bioclim
  - Maxent
    - maxent.jar
    - Maxnet
- Model Tuning:
  - Test parameter combinations
  - Find best-performing model
  - Reduce complexity



# Component 6: Model Building Background

Feature Classes

- L
- Q
- H
- P
- T
- :



# Component 6: Model Building Background

Regularization Multiplier

0.5

1

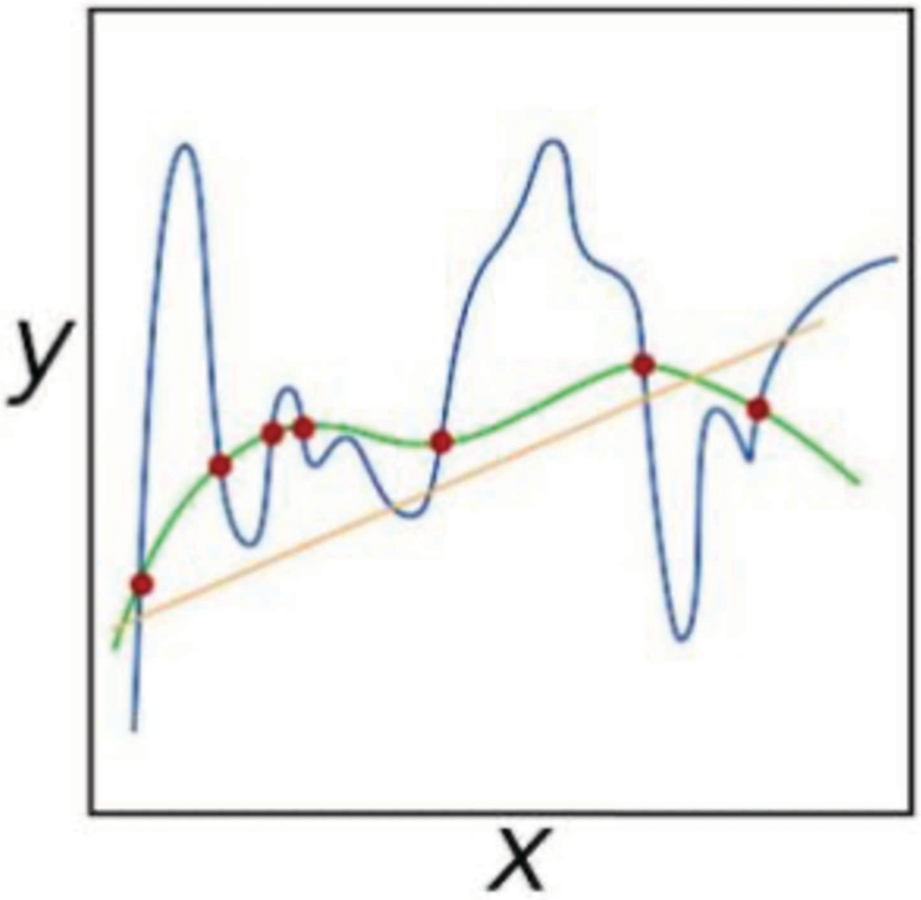
1.5

2

2.5

:

RMs penalize higher complexity.




Overfit Model (blue)

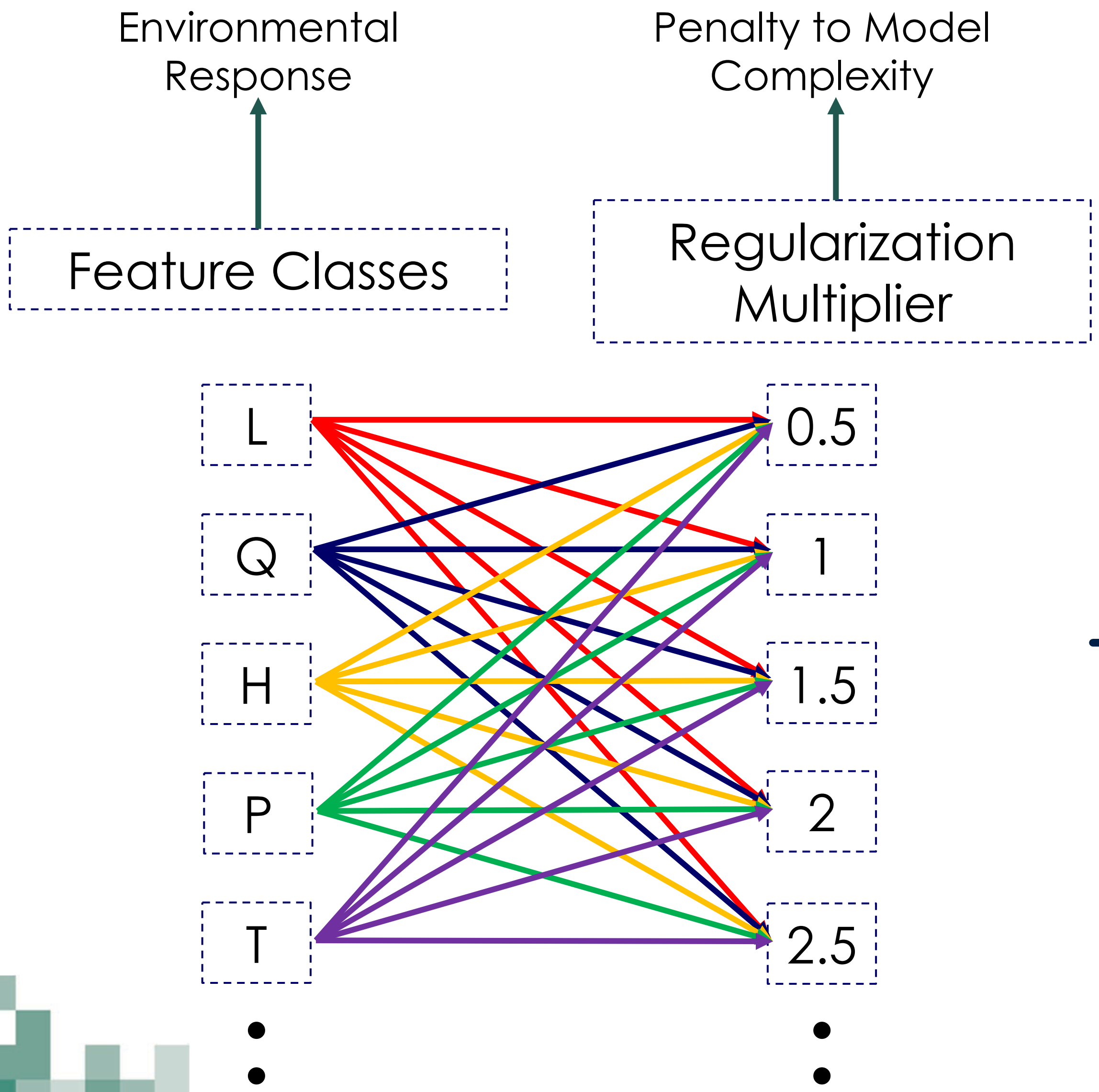
Underfit Model (yellow)

Generalized Model (green) **Desired**



https://mylearningsinaiml.wordpress.com/concepts/overfitting-underfitting/

# Component 6: Model Tuning




| | rm | fc |
|---|---|---|
| 1 | 1 | L |
| 2 | 1 | LQ |
| 3 | 1.5 | L |
| 4 | 1.5 | LQ |
| 5 | 2 | L |
| 6 | 2 | LQ |
| 7 | 2.5 | L |
| 8 | 2.5 | LQ |
| 9 | 3 | L |
| 10 | 3 | LQ |

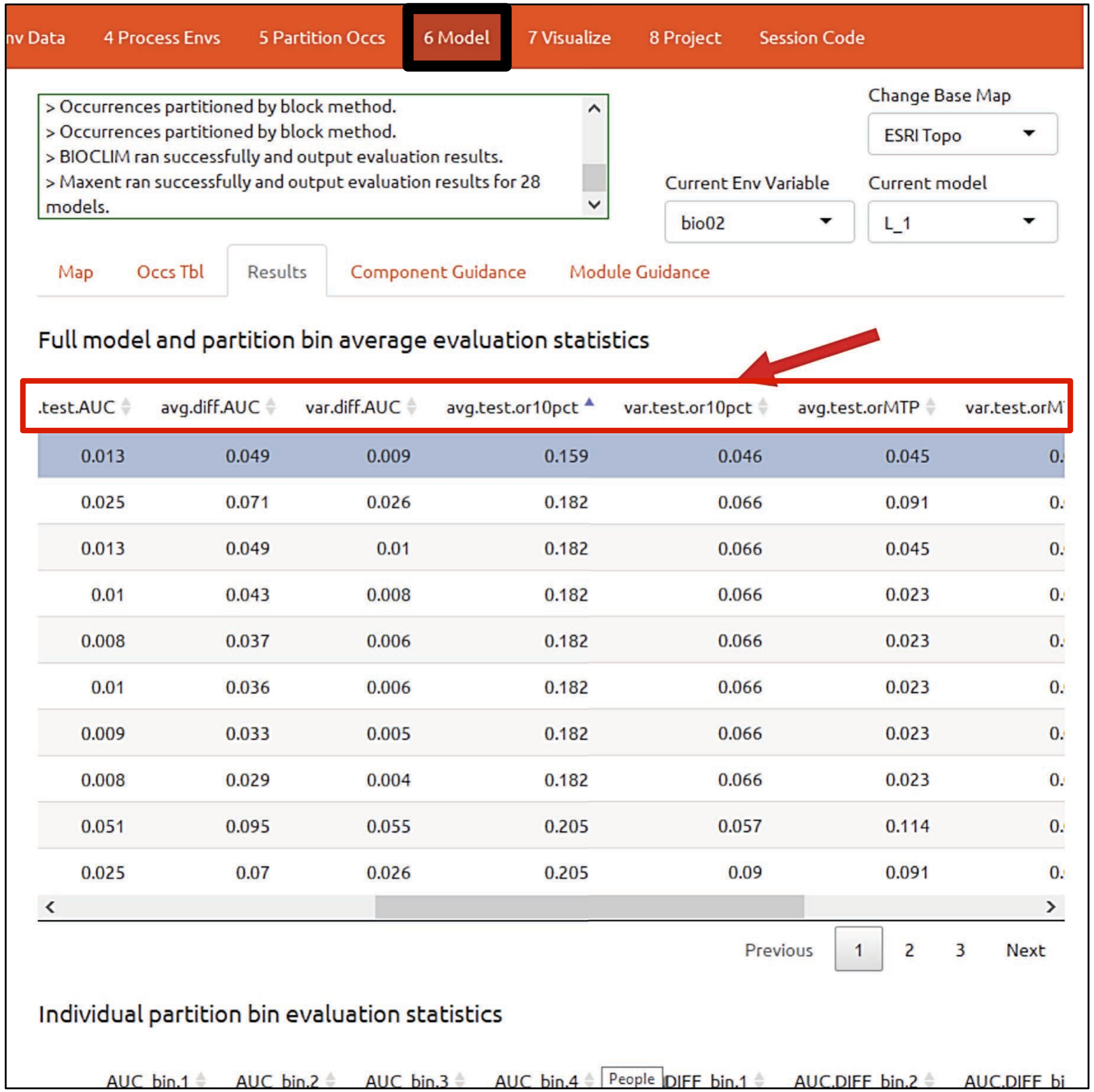
## Recommendations:

- >80 occurrence records:
  All Feature Classes
- 15 – 79 occurrence records: L, Q, H
- 10 – 14 occurrence records: L and Q
- < 10 occurrence records: L only

# Component 6: Model Evaluation

- Modeling Evaluation Metrics
  - AUC
  - Omission Rate
  - AICc
  - N Parameters

- Applicable To:
  - Entire Model
  - Each Partition




| .test.AUC | avg.diff.AUC | var.diff.AUC | avg.test.or10pct | var.test.or10pct | avg.test.orMTP | var.test.orM |
|---|---|---|---|---|---|---|
| 0.013 | 0.049 | 0.009 | 0.159 | 0.046 | 0.045 | 0. |
| 0.025 | 0.071 | 0.026 | 0.182 | 0.066 | 0.091 | 0. |
| 0.013 | 0.049 | 0.01 | 0.182 | 0.066 | 0.045 | 0. |
| 0.01 | 0.043 | 0.008 | 0.182 | 0.066 | 0.023 | 0. |
| 0.008 | 0.037 | 0.006 | 0.182 | 0.066 | 0.023 | 0. |
| 0.01 | 0.036 | 0.006 | 0.182 | 0.066 | 0.023 | 0. |
| 0.009 | 0.033 | 0.005 | 0.182 | 0.066 | 0.023 | 0. |
| 0.008 | 0.029 | 0.004 | 0.182 | 0.066 | 0.023 | 0. |
| 0.051 | 0.095 | 0.055 | 0.205 | 0.057 | 0.114 | 0. |
| 0.025 | 0.07 | 0.026 | 0.205 | 0.09 | 0.091 | 0. |

# Component 6: Model Evaluation

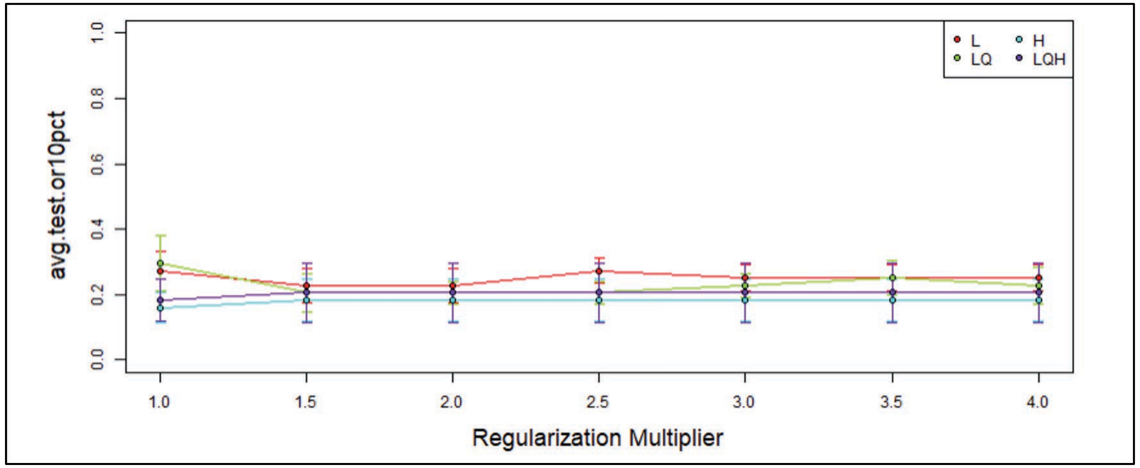
| AUC | orMTP/10pct | AICc | Nparam |
|---|---|---|---|
| Measures a model's ability to distinguish true & false presences | Measures the % of true presence points predicted as absences | Measures the relative quality of a model by balancing fit & complexity | More parameters increase model complexity |

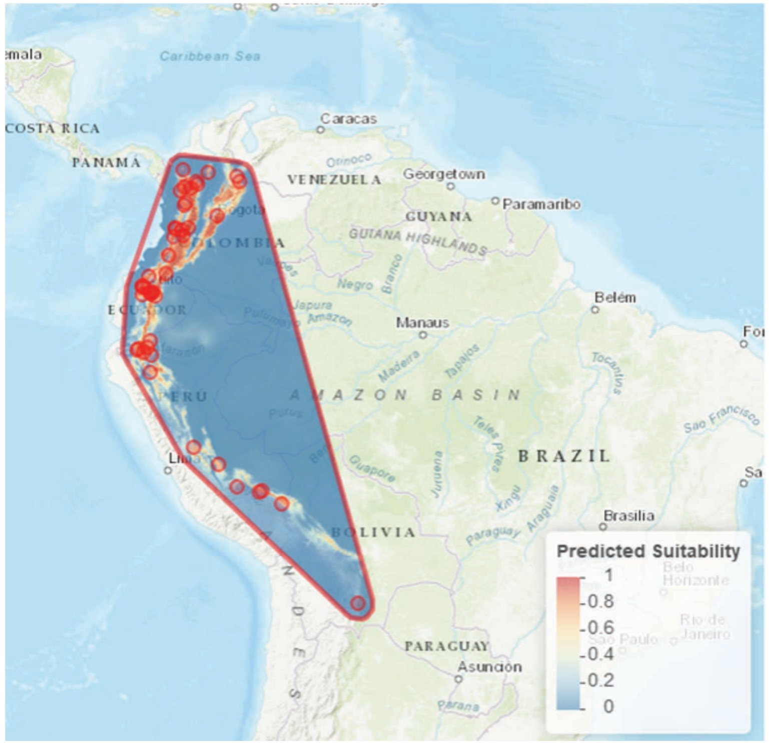
Generally, we want models with:

- High AUC
- Low Omission Rate
- Low AICc
- Fewer Parameters

# Component 7: Visualizing Results (Visualize)

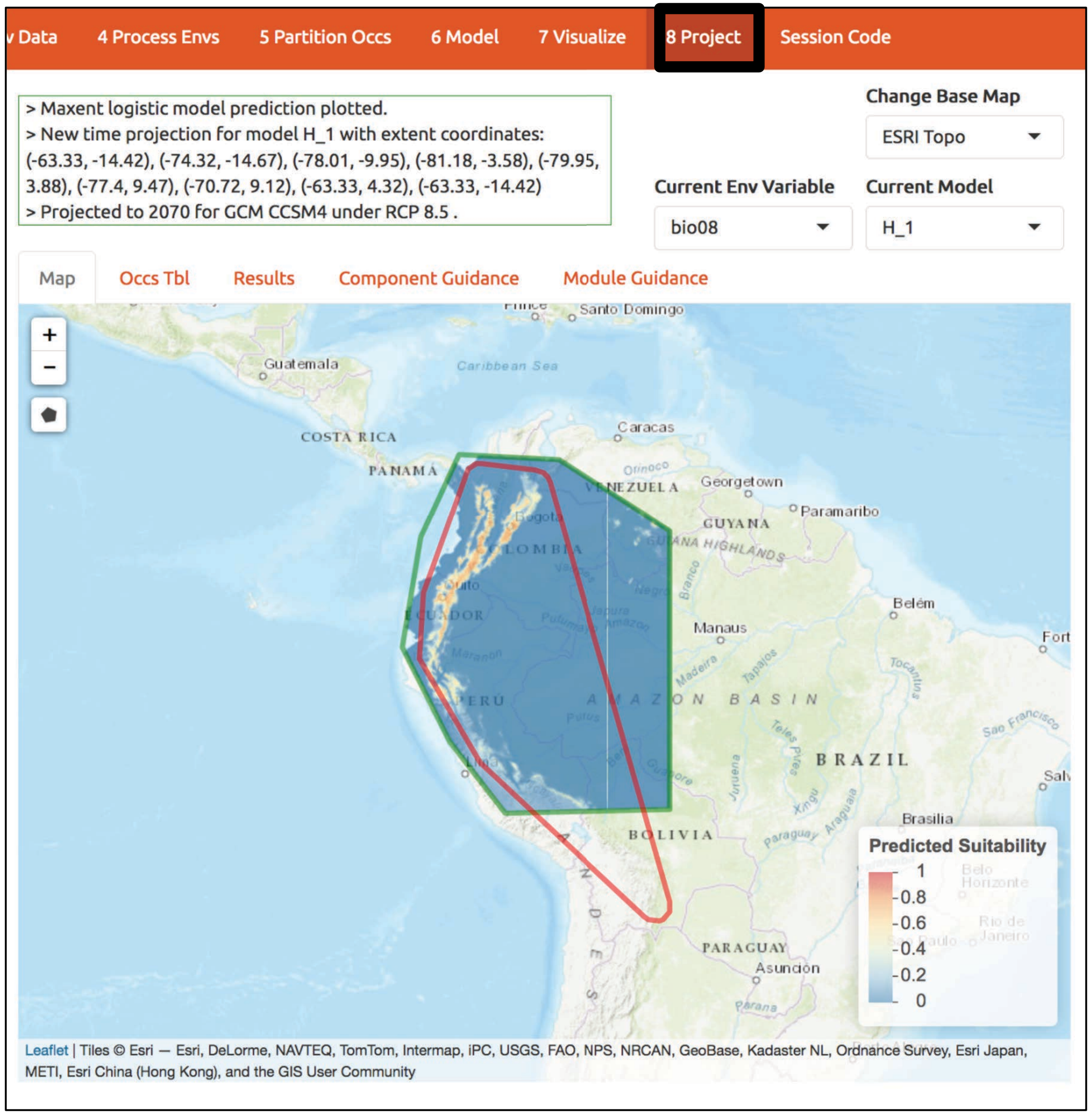
- Visualize results as:
  - Model Evaluation Plots
  - Response Curves
  - Prediction Maps
    - Continuous
    - Binary



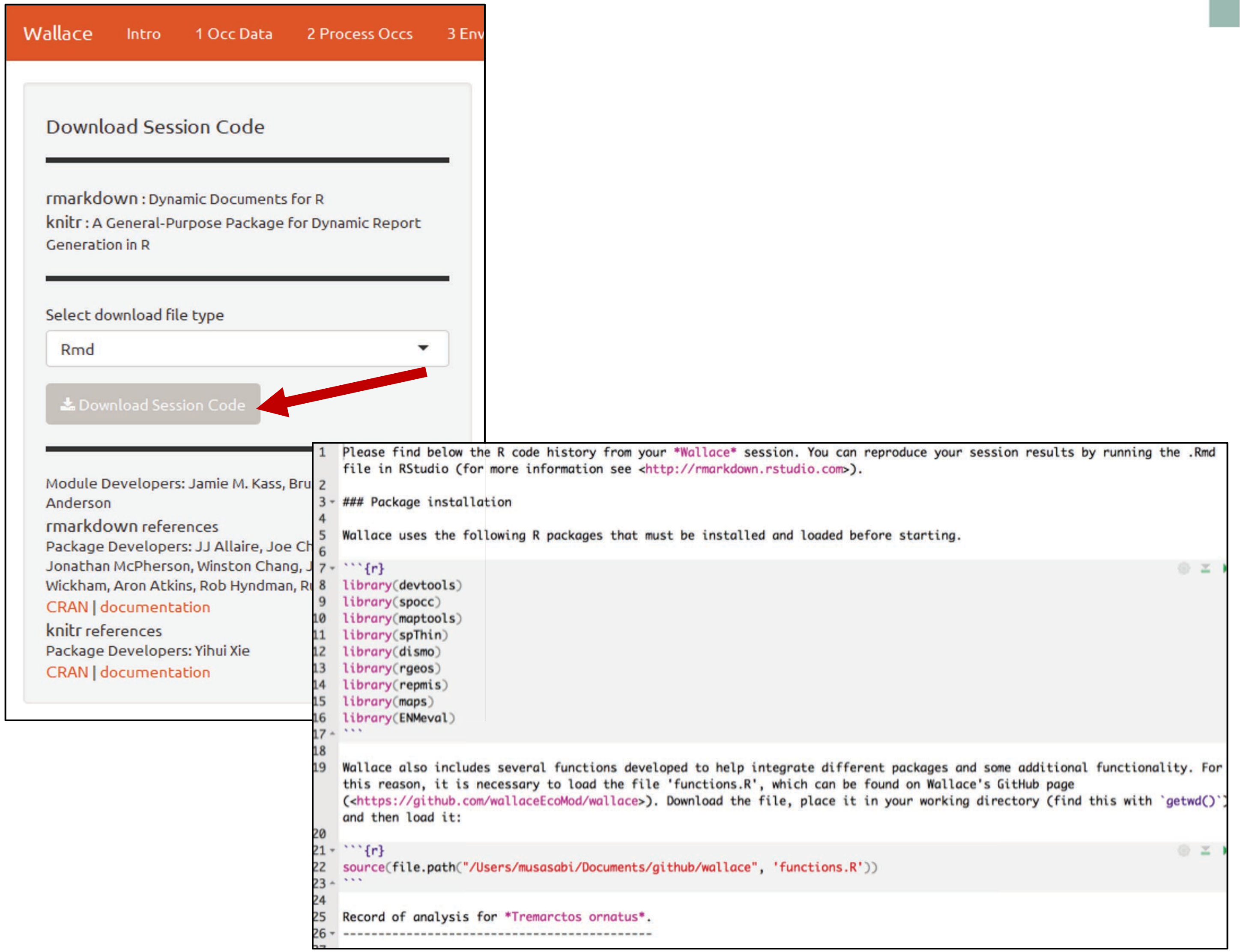
# Component 8: Spatial & Temporal Projections (Project)

- Allows us to project models to:
  - New Geographic Extents
  - Different Time Periods
    - Future
    - Past



# Component 9: Download the R Code (Session Code)

- Download the R code.
- Replicate analyses in R.



**Thank You!**