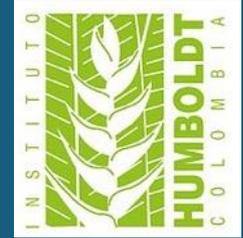




AMERICAN MUSEUM OF NATURAL HISTORY  
CENTER FOR BIODIVERSITY AND CONSERVATION



# Expanding *Wallace* species distribution modeling software to support national biodiversity change indicator calculations for GEO BON assessment and reporting

**Mary E. Blair, AMNH, [mblair1@amnh.org](mailto:mblair1@amnh.org)**

Robert P. Anderson, CCNY/CUNY;

Matthew Aiello-Lammens, Pace; Cory Merow, UCONN;

Peter Galante, Cecina Babich Morrow, Alex Moore,

Ned Horning, AMNH; Jamie Kass, Gonzalo Pinilla-

Buitrago, CCNY/CUNY; Beth Gerstner, Michigan State;

Jorge Velásquez, Elkin Noguera, Instituto Humboldt

The City College  
of New York



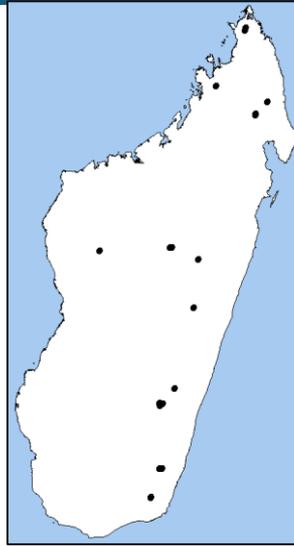
CU  
NY The City  
University  
of New York

PACE  
UNIVERSITY



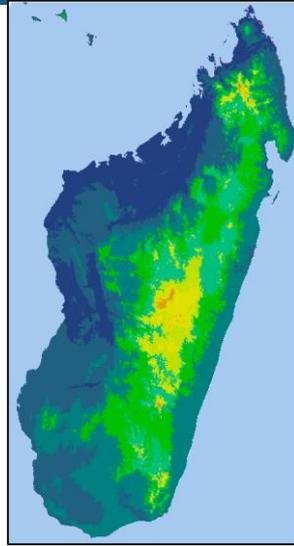
GOO GROUP ON  
EARTH OBSERVATIONS

# Species Distribution Models (SDMs)



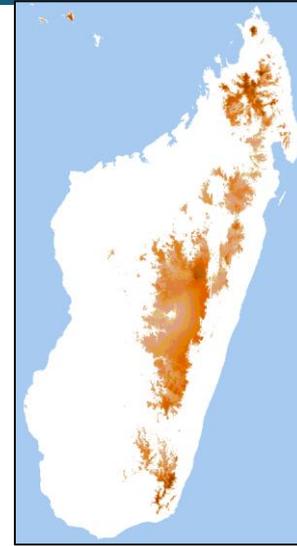
Occurrence  
localities

+



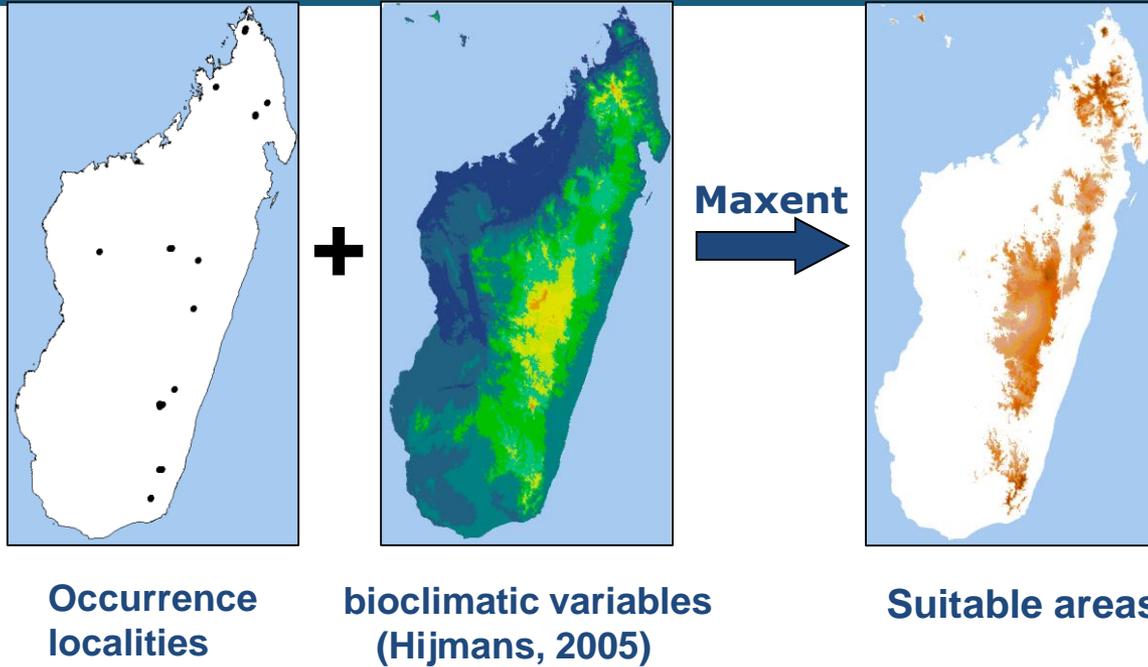
bioclimatic variables  
(Hijmans, 2005)

Maxent  
➔



Suitable areas

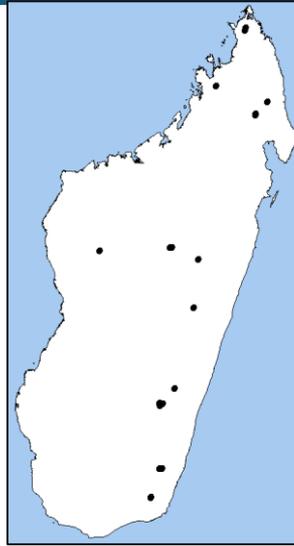
# Species Distribution Models (SDMs)



## SDMs for conservation decision-making:

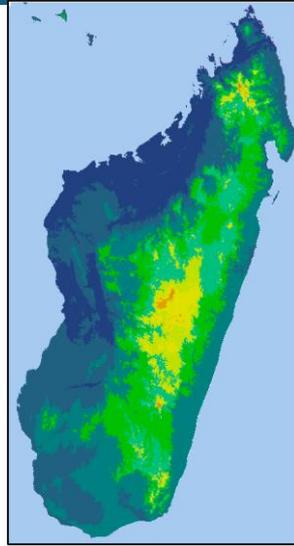
- Planning and prioritization
- Guiding surveys, monitoring and reintroductions
- Danger of misuse, false precision

# Species Distribution Models (SDMs)



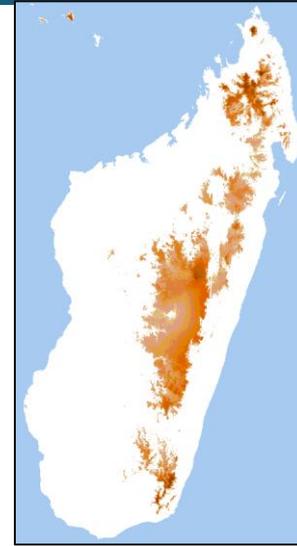
Occurrence localities

+



bioclimatic variables (Hijmans, 2005)

Maxent



Suitable areas



```

library(RDlab)
library(sp)
library(spThin)
globalbio<-read(list.files(pattern = "\\.tif", path = "C:/Users/jgalante/Documents/Projects/vLayers/vc2-5", full.names = T))
e<-extent(c(91, 115, 5, 30))
vietnam<-crop(globalbio, e)
## LOAD THINMED OCCURRENCE RECORDS
llocs<- read.csv("C:/Users/jgalante/Documents/Projects/Vietnam/thinned_data_thin0.csv")
## Generate buffered MCP of llocs as background points
library(rgeos)
mcp <- function(xy) {
  xy <- as.data.frame(coordinates(xy))
  coords <- cbind(xy[, 1], xy[, 2])
  xy.bord <- xy[coords, ]
  xy.bord <- rbind(xy.bord[rown(xy.bord), ], xy.bord)
  return(SpatialPolygons(list(Polygons(list(Polygon(as.matrix(xy.bord))), 1)))
}
buff.dist <- 2
MCP.llocs <- mcp(llocs[,2:3])
MCP.buf <- gbuffer(MCP.llocs, width = buff.dist)
e1 <- mask(vietnam[[1]], MCP.buf)
background <- randomPoints(e1, 10000)
rm(globalbio);rm(e1); rm(MCP.buf); rm(MCP.llocs)
##### 1) CLIMATE ONLY
res<-RMValueLATE(lcc = llocs[,2:3], env = vietnam, NValues = seq(0.5,5,0.5), fc = c("L","Q","UQ"),
  method = "jackknife", bg.coords = background, parallel = T, numcores = 6, algorithm = "maxent")
with.mod <- maxent(
  xenv, # bio stack
  p=llocs[,2:3], # locality csv
  env = background, # background coords
  path = "", # path to save to
  args=c(
    'betamultiplier=1.5',
    'linear=true',
    'quadratic=true',
    'product=false',
  )
  )
  
```

- Easy use
- Inflexible/"Black box"
- Flexible
- Steep learning curve

Received: 9 October 2017 | Accepted: 16 November 2017

DOI: 10.1111/2041-210X.12945

**APPLICATION**

Methods in Ecology and Evolution  
BRITISH  
ECOLOGICAL  
SOCIETY

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

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

... and Mary E. Blair, Ned Horning, Jorge Velásquez-Tibatá, Gonzalo E. Pinilla-Buitrago, Beth E. Gerstner & Sarah I. Meenan



# Wallace

Obtain Occurrence Data **2**

Modules Available:

- Query Database
- User-specified Occurrences

---

Module: Query Database **2a**

via *spocc* package: Interface to Species Occurrence Data Sources

---

Choose Database: **2b**

- GBIF
- VertNet
- BISON

Enter scientific name (format: Genus species)

Tremarctos ornatus

Search Database

Maximum number of occurrences:

75 / 3,000

Download DB Occurrences

---

spocc references

Developers: Scott Chamberlain, Karthik Ram, Ted Hart

CRAN | documentation

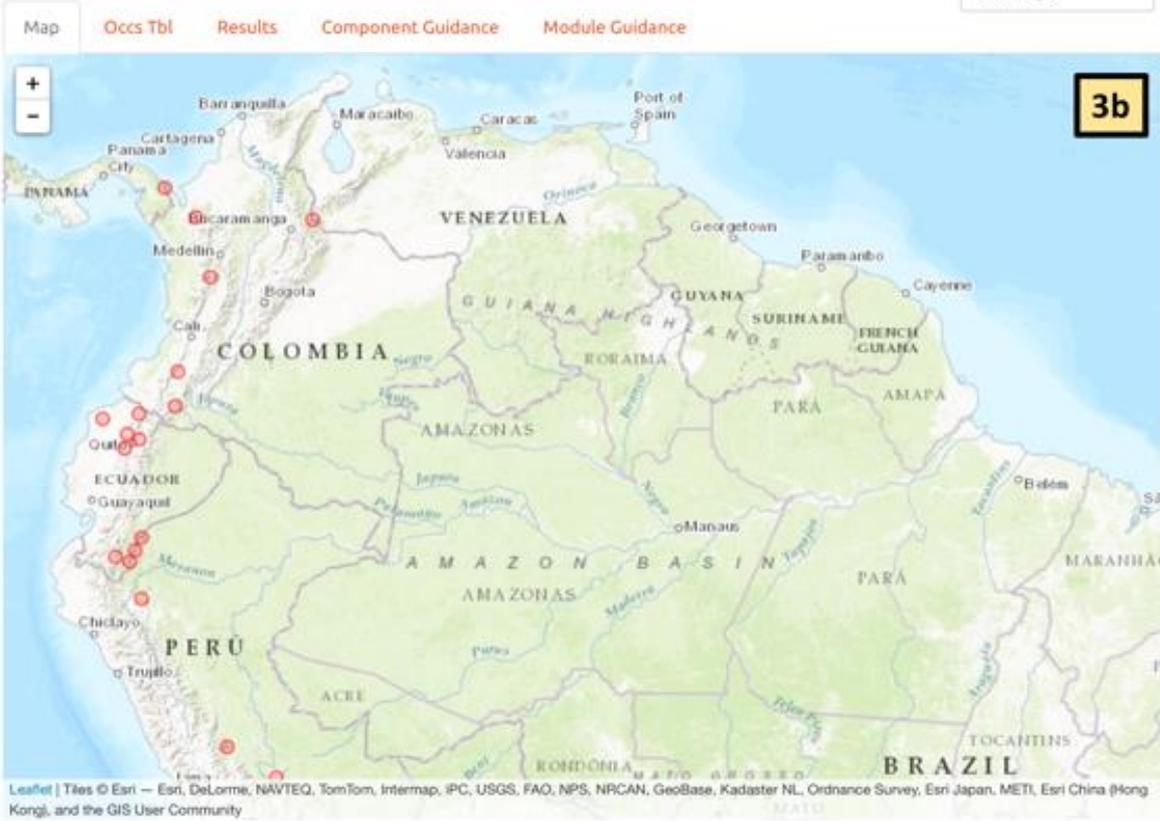
... Searching gbif ...

\* Total gbif records for Tremarctos ornatus returned [ 75 ] out of [ 209 ] total (limit 75 ). Records without coordinates removed [ 53 ]. Duplicated records removed [ 0 ]. Remaining records [ 22 ].

**3a**

Change Base Map

ESRI Topo

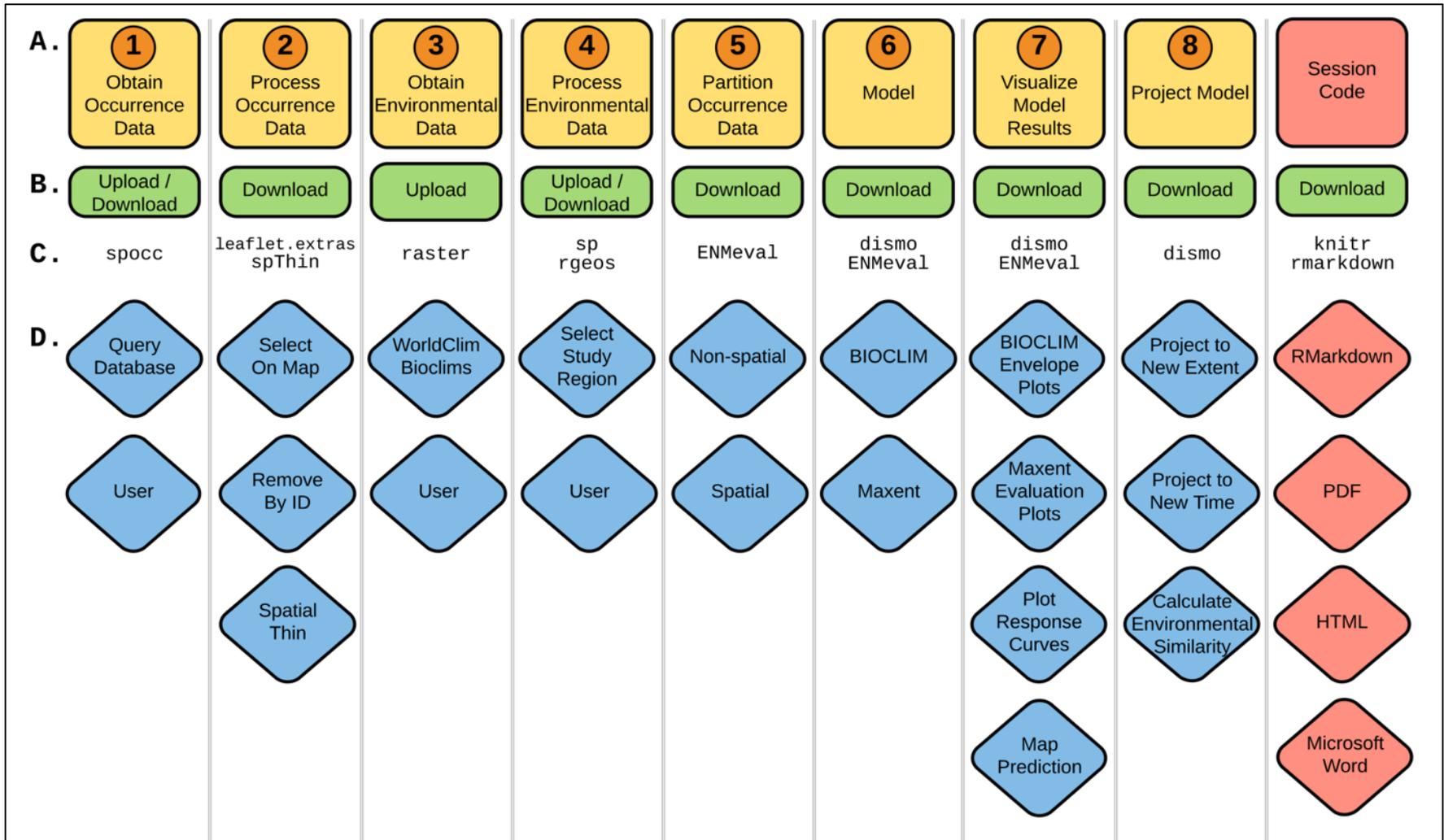


Open-source, flexible, reproducible

<https://wallaceecomod.github.io/>

# Wallace

## Components each with modules



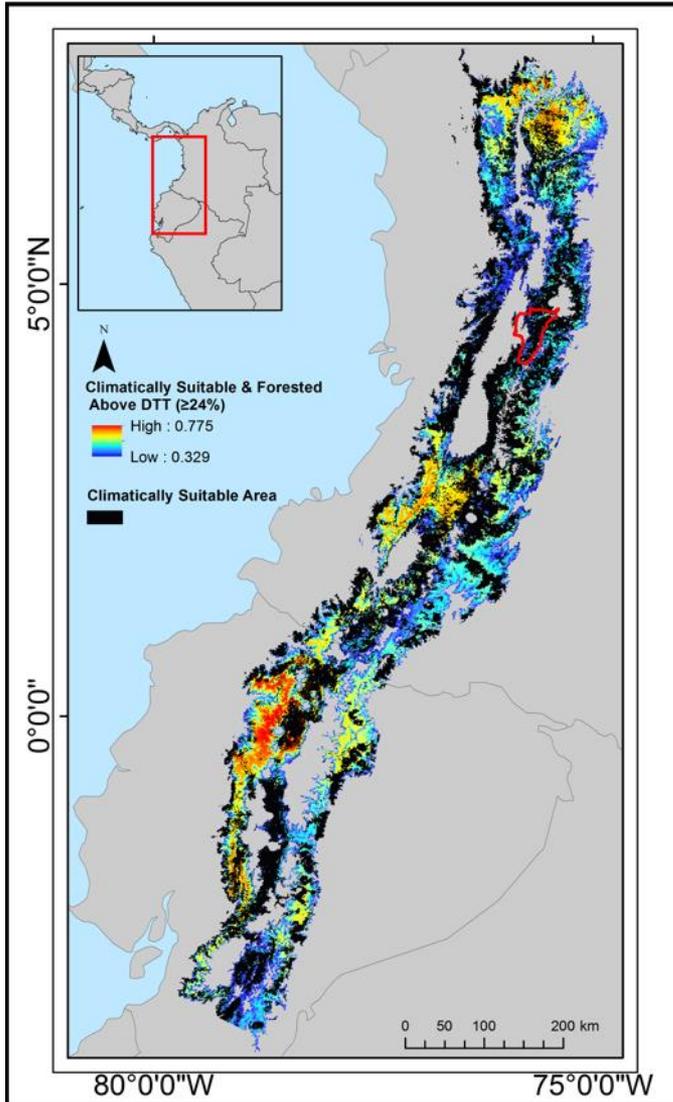
# We are expanding *Wallace* to:

1. Bridge the gap between SDM best practices and biodiversity conservation decision-making

# We are expanding *Wallace* to:

1. Bridge the gap between SDM best practices and biodiversity conservation decision-making
2. Facilitate responsible reporting on biodiversity change by national BONs, with the Colombia BON as a model.

# Develop new R packages to add to *Wallace*



1. `maskRangeR` – Estimates a species' current range
  - Post-processing of species distribution models with RS products



*"Thank you... it is not common for software developers to include such a broad range of end-users in their decisions."*



Wallace end-user consultation workshop in Bogotá, Colombia – April, 2018

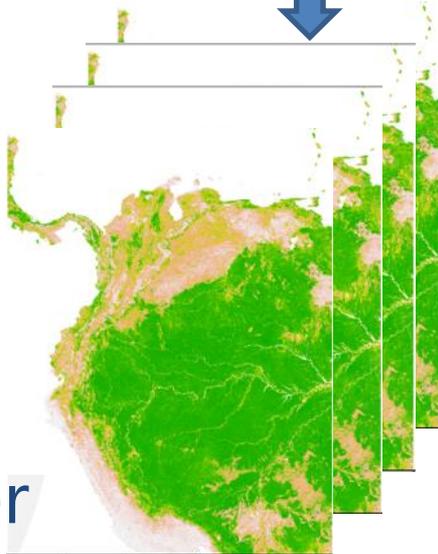
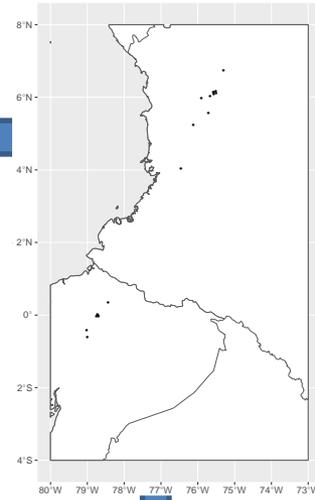
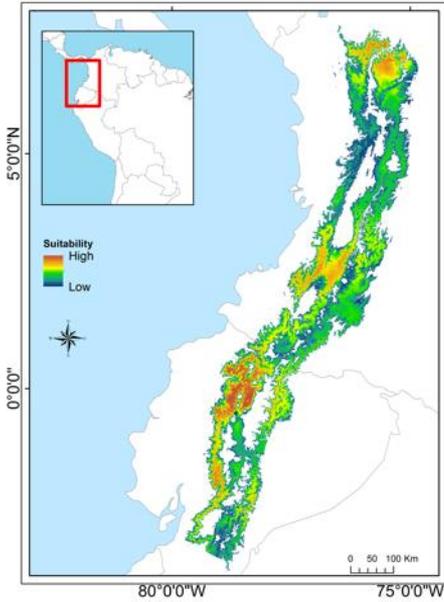
# *maskRangeR* example: Olinguito (*Bassaricyon neblina*)

- Recently described small carnivore
- Limited to high altitude, cloud forests
- Data-poor, in need of status update given recent deforestation

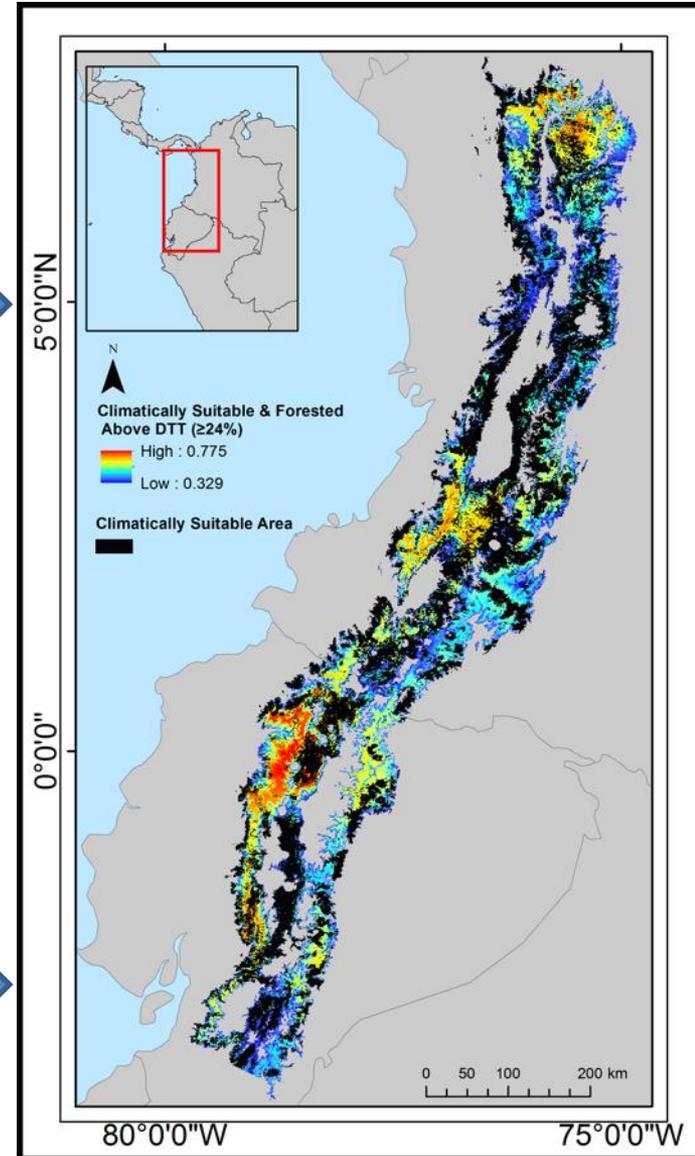


# Olinguito SDM

# Recent occurrences



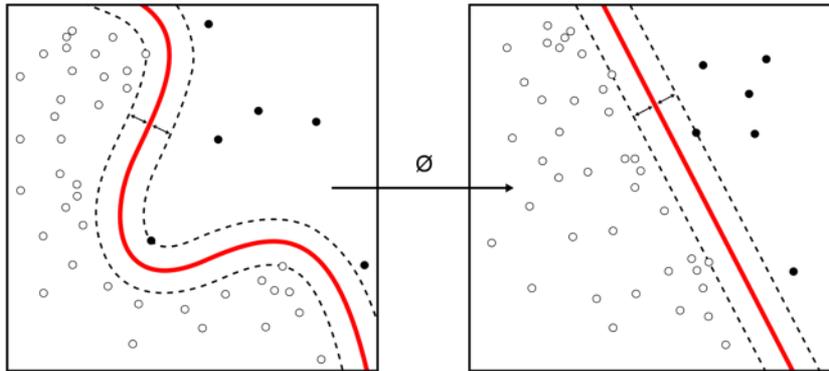
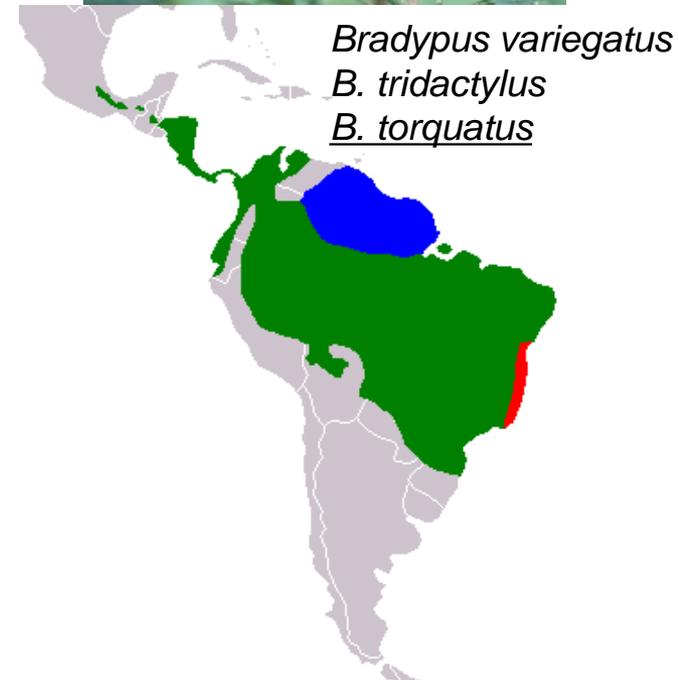
MODIS  
yearly  
forest cover



Masked Distribution

# *maskRangeR* example: parapatric species

- Use biotic information to improve modeling outputs
  - 3 species of three-toed sloths as example



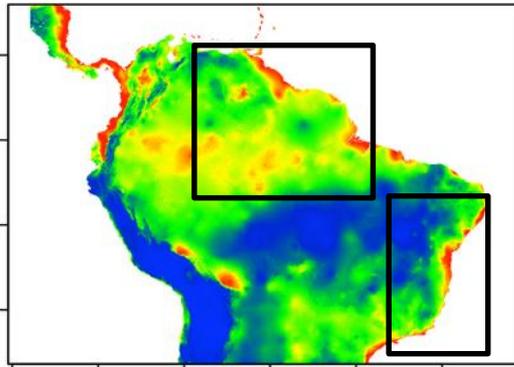
Support vector machine  
(SVM) [https://en.wikipedia.org/wiki/Support-vector\\_machine#/media/File:Kernel\\_Machine.svg](https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:Kernel_Machine.svg)

# *maskRangeR* example:

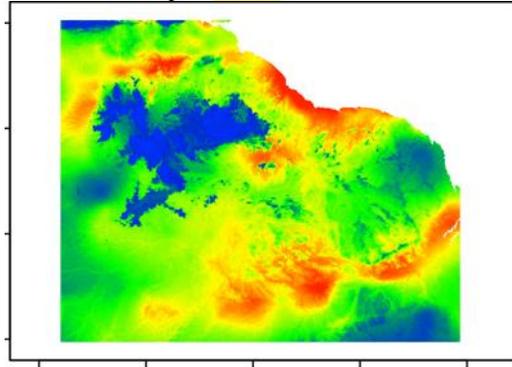
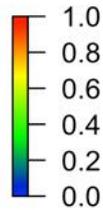
## parapatric species

SVM for range boundary estimation

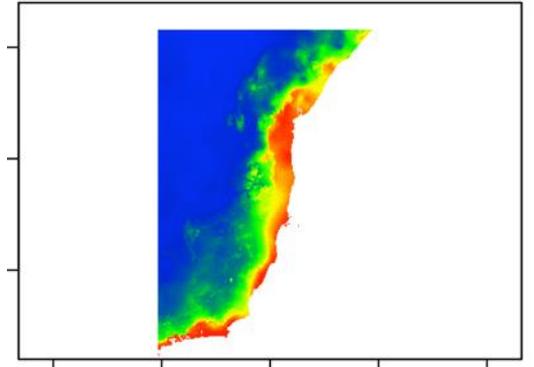
SDMs



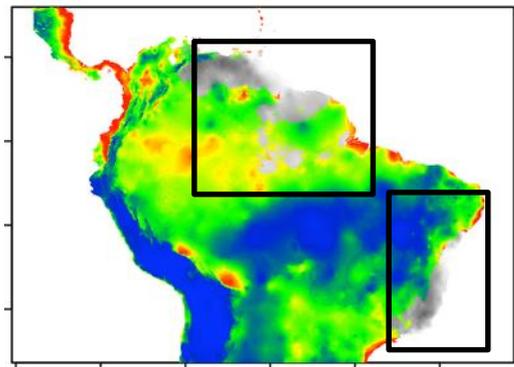
*B. variegatus*



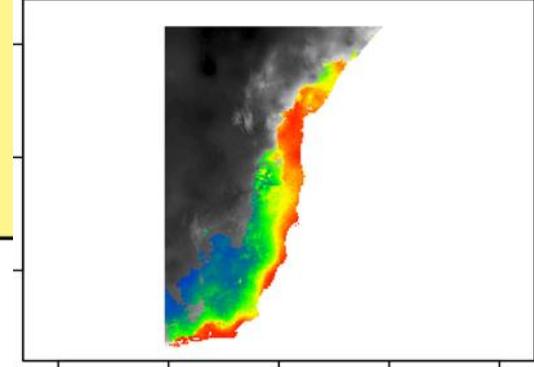
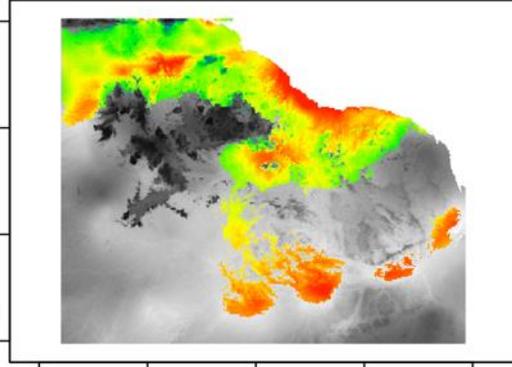
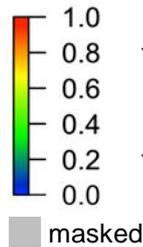
*B. tridactylus*



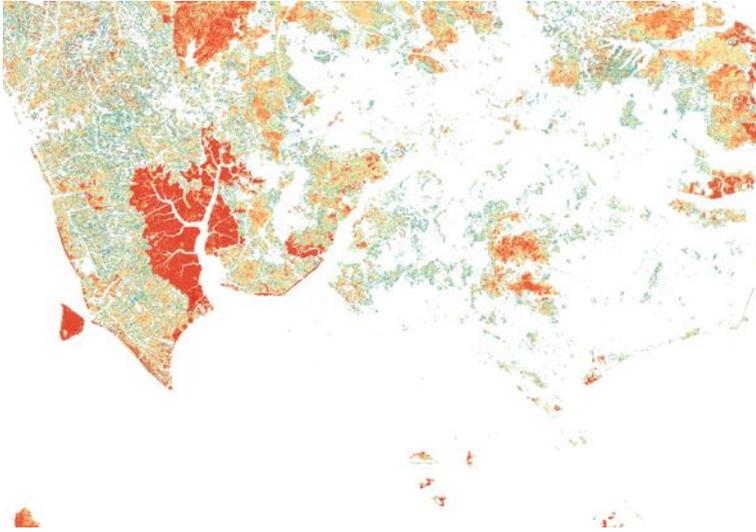
*B. torquatus*



Masked Distributions



# Expert-driven use cases



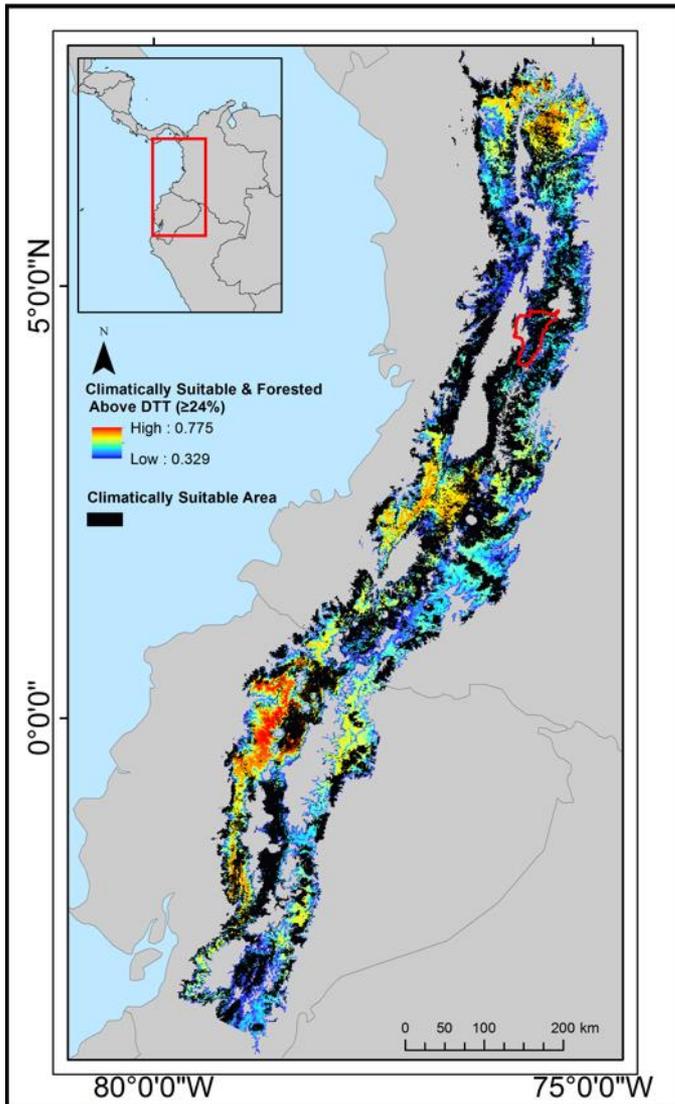
# Next, add to *Wallace* as new modules



# Second end-user consultation workshop June 11-12, 2019 in Bogotá



# Develop new R packages



1. maskRangeR – Estimate a species' current range
  - Post-processing of species distribution models with RS products
2. Package B – Calculate key biodiversity change indicators
  - IUCN AOO and EOO upper bounds
  - % suitable land cover
  - Protected Area representativeness, taxonomic diversity...

# Supporting the Colombia BON



BON in a Box Tools include software, protocols, and other tools that facilitate effective biodiversity change reporting

[/boninabox.geobon.org/](http://boninabox.geobon.org/)



# Integrate *Wallace* & *BioModelos*

The screenshot shows the Wallace web application interface. The top navigation bar includes 'Wallace' and several numbered steps: 1 Occ Data, 2 Previous Occs, 3 Env Data, 4 Previous Env, 5 Pathways Occs, 6 Models, 7 Visualizer, 8 Project, and 9 Session Code. The main content area is divided into a left sidebar and a right main panel. The sidebar contains sections for 'Obtain Occurrence Data' (with a yellow box '2'), 'Module: Query Database' (with a yellow box '2a'), 'Choose Database:' (with a yellow box '2b'), and 'Maximum number of occurrences:' (set to 3000). The main panel shows a search for 'Tremarctos ornatus' (with a yellow box '3a') and a map of South America (with a yellow box '3b'). The word 'Wallace' is overlaid in large black text on the map. A 'Download DB Occurrences' button is visible at the bottom left of the sidebar.



The BioModelos logo and banner. The logo consists of the text 'BioModelos' in a bold, sans-serif font, enclosed in a thin red rectangular border. Below the logo, the text 'MEJORES MODELOS CON EL APOYO DE EXPERTOS' is displayed in a smaller, black, sans-serif font. At the bottom of the banner, the website URL 'biomodelos.humboldt.org.co' is shown in a red, sans-serif font. To the right of the text is a colorful, stylized illustration of a bird, possibly a toucan, with vibrant red, green, and blue feathers.

# Create interactive web-based training materials

1 Wallace Intro 1 Occ Data 2 Process Occs 3 Env Data 4 Process Envs 5 Partition Occs 6 Model 7 Visualize 8 Project Session Code

2 Obtain Occurrence Data  
Modules Available:  
 Query Database  
 User-specified Occurrences

2a Module: Query Database  
via **spocc** package: *Interface to Species Occurrence Data Sources*

2b Choose Database:  
 GBIF  
 VertNet  
 BISON

Enter scientific name (format: Genus species)  
Tremarctos ornatus  
Search Database

Maximum number of occurrences:  
3,000  
Download DB Occurrences

3a ... Searching gbif ...  
\* Total gbif records for Tremarctos ornatus returned [ 75 ] out of [ 209 ] total (limit 75 ). Records without coordinates removed [ 53 ]. Duplicated records removed [ 0 ]. Remaining records [ 22 ].

3 Change Base Map  
ESRI Topo

3b Map Occs Tbl Results Component Guidance Module Guidance

spocc references  
Developers: Scott Chamberlain, Karthik Ram, Ted Hart  
CRAN | documentation

Network of  
Conservation  
Educators  
and  
Practitioners



[ncep.amnh.org](http://ncep.amnh.org)



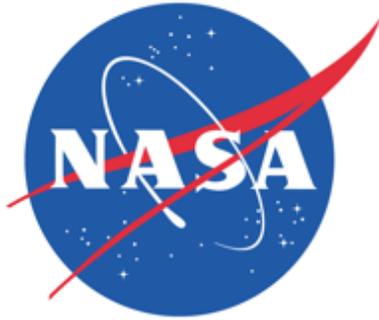
# Bridging the gap to decision-making with co-creation



*Wallace end-user consultation workshop in Bogotá,  
Colombia – April, 2018*

<https://wallaceecomod.github.io/>

Thanks to:



GROUP ON  
EARTH OBSERVATIONS 16-GEO16-0027



DBI 1650241  
DBI 1661510



GBIF



GC Digital Initiatives

# Wallace software

The screenshot displays the Wallace software interface. At the top, a navigation bar includes tabs for Wallace, Intro, 1 Occ Data, 2 Process Occs, 3 Env Data, 4 Process Envs, 5 Partition Occs, 6 Model, 7 Visualize, 8 Project, and Session Code. The main interface is divided into several sections:

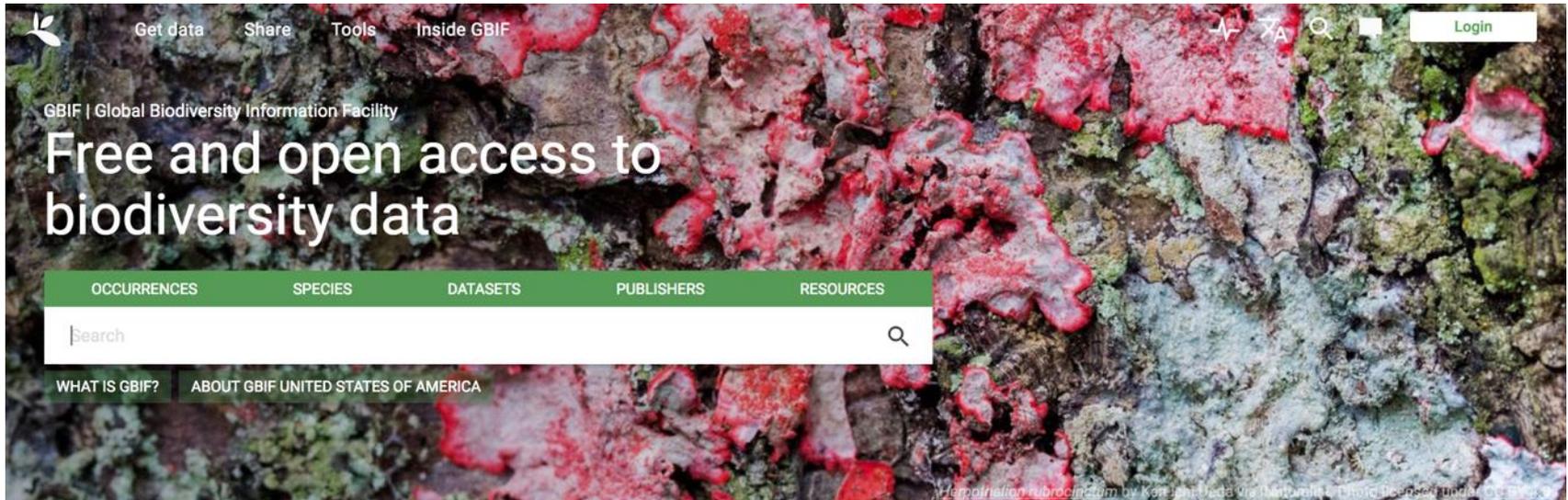
- Obtain Occurrence Data (2):** This section contains options for 'Query Database' (selected) and 'User-specified Occurrences'.
- Module: Query Database (2a):** Below this, it states 'via spocc package: Interface to Species Occurrence Data Sources'.
- Choose Database: (2b):** This section offers three database options: GBIF (selected), VertNet, and BISON.
- Enter scientific name (format: Genus species):** A text input field contains 'Tremarctos ornatus', with a 'Search Database' button below it.
- Maximum number of occurrences:** A slider control is set to 75, with a maximum value of 3,000.
- Download DB Occurrences:** A button to download the search results.
- spocc references:** A section listing developers (Scott Chamberlain, Karthik Ram, Ted Hart) and providing links to CRAN and documentation.

On the right side, a search status box (3a) indicates: '... Searching gbif ... \* Total gbif records for Tremarctos ornatus returned [ 75 ] out of [ 209 ] total (limit 75). Records without coordinates removed [ 53 ]. Duplicated records removed [ 0 ]. Remaining records [ 22 ].' Below this is a 'Change Base Map' dropdown menu set to 'ESRI Topo'. A map (3b) shows the Amazon basin region, with red dots representing occurrence records for Tremarctos ornatus. The map includes labels for countries like Venezuela, Colombia, Ecuador, Peru, and Brazil, as well as major cities and geographical features.

Open-source, flexible, reproducible

<https://wallacecomod.github.io/>





Get data Share Tools Inside GBIF [Login](#)

GBIF | Global Biodiversity Information Facility

# Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search

[WHAT IS GBIF?](#) [ABOUT GBIF UNITED STATES OF AMERICA](#)

*Temprallia rubrocinctum* by Ken Spivey via Wikimedia Commons. Photo: iStockphoto.com

Occurrence records  
1,017,750,309

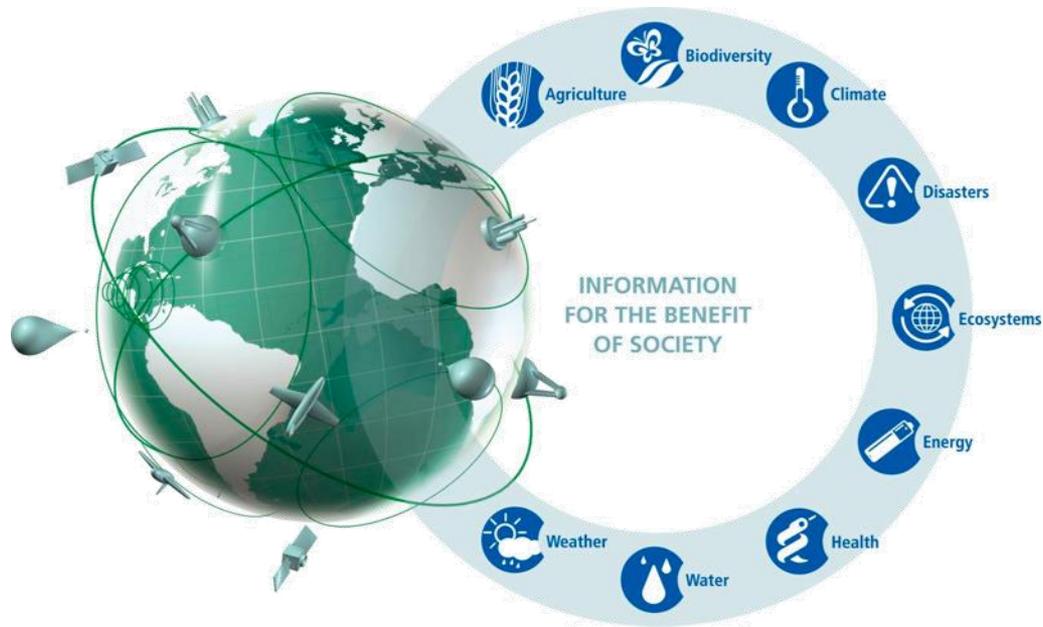
Datasets  
40,645

Publishing institutions  
1,265

Species  
Learn more about the number of species covered by data in GBIF.org.



Group on Earth Observations Biodiversity  
Observation Network



# Overview and guidance

Wallace Intro 1 Occ Data 2 Process Occs 3 Env Data 4 Process Envs 5 Partition Occs 6 Model 7 Visualize 8 Project Session Code

## WORKFLOW

Wallace (v1.0.4) currently includes the following components and modules:

### 1: Obtain Occurrence Data

- Query Database
- User-specified

### 2: Process Occurrence Data

- Select Occurrences on Map
- Remove Occurrences by ID
- Spatial Thin

### 3: Obtain Environmental Data

- WorldClim
- User-specified

### 4: Process Environmental Data

- Select Study Region
- User-specified

### 5: Partition Occurrence Data

- Non-spatial Partition
- Spatial Partition

### 6: Build and Evaluate Niche Model

- BIOCLIM
- Maxent

### 7: Visualize Model Results

- BIOCLIM Envelope Plot
- Maxent Evaluation Plots
- Plot Response Curves
- Map Prediction

### 8: Project Model

- Project to New Area
- Project to New Time

Intro About

## What is Wallace?

Welcome to Wallace, a flexible application for reproducible ecological modeling, built for community expansion. The current version of Wallace (v1.0.4) steps the user through a full niche/distribution modeling analysis, from data acquisition to visualizing results.

The application is written in [R](#) with the web app development package [shiny](#). Please find the stable version of Wallace on [CRAN](#), and the development version on [Github](#). We also maintain a [Wallace website](#) that has some basic info, links, and will be updated with tutorial materials in the near future.

Wallace is designed to facilitate spatial biodiversity research, and currently concentrates on modeling species niches and distributions using occurrence datasets and environmental predictor variables. These models provide an estimate of the species' response to environmental conditions, and can be used to generate maps that indicate suitable areas for the species (i.e. its potential geographic distribution; Guisan & Thuiller 2005; Elith & Leathwick 2009; Franklin 2010a; Peterson et al. 2011). This research area has grown tremendously over the past two decades, with applications to pressing environmental issues such as conservation biology (Franklin 2010b), invasive species (Ficetola et al. 2007), zoonotic diseases (González et al. 2010), and climate-change impacts (Kearney et al. 2010).

Also, for more detail, please see our paper in *Methods in Ecology and Evolution*.

Kass JM, Vilela B, Aiello-Lammens ME, Muscarella R, Merow C, Anderson RP. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods Ecol Evol.* 2018;00:1–6.

<https://doi.org/10.1111/2041-210X.12945>

## Who is Wallace for?

We engineered Wallace to be used by a broad audience that includes graduate students, ecologists, conservation practitioners, natural resource managers, educators, and programmers. Anyone, regardless of programming ability, can use Wallace to perform an analysis, learn about the methods, and share the results. Additionally, those who want to disseminate a technique can author a module for Wallace.

## Attributes of Wallace

- **open**: the code is free to use and modify (GPL 3.0), and it gives users access to some of the largest public online biodiversity databases
- **expandible**: users can author and contribute modules that enable new methodological options
- **flexible**: options for user uploads and downloads of results

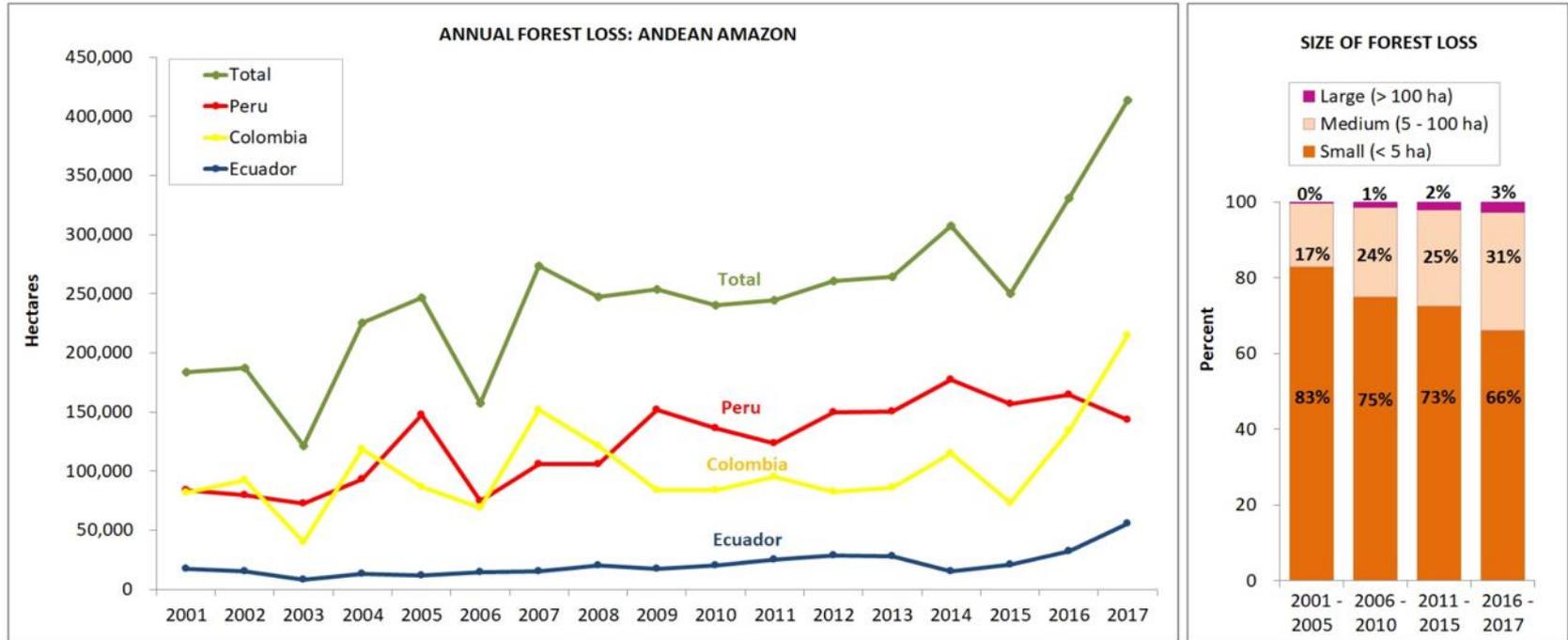
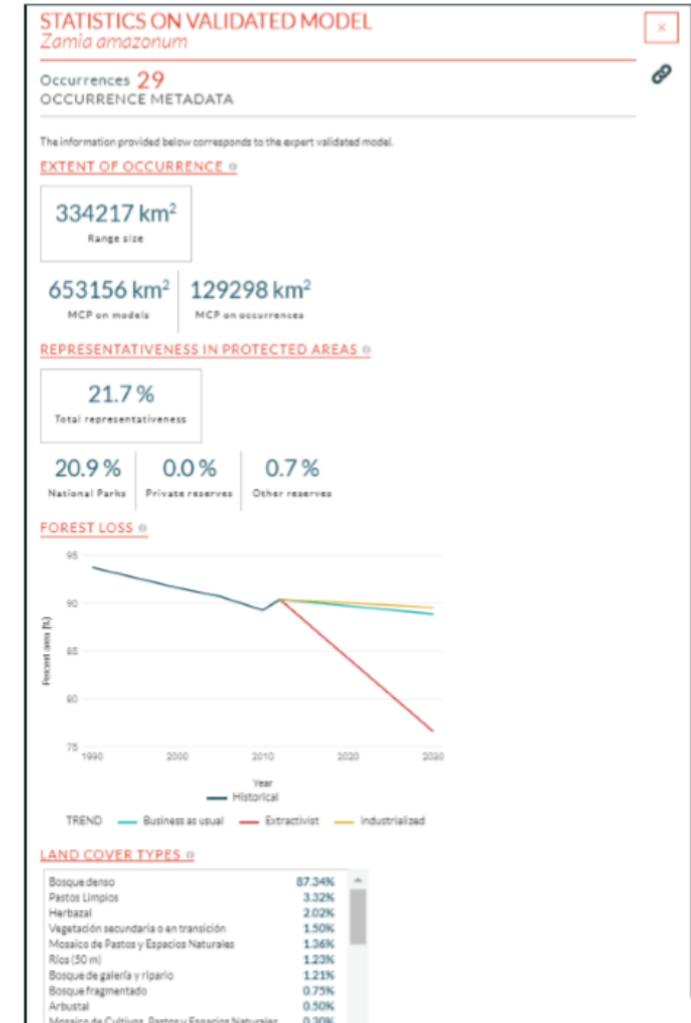


Image 1. Annual forest loss by country and size. Data: Hansen/UMD/Google/USGS/NASA, UMD/GLAD, Global Forest Watch, MINAM/PNCB, RAISG.

Source: MAAPProject.org

# BioModelos

- Connects biodiversity experts to species distribution models (SDMs)
- Facilitates:
  1. Curating records
  2. Editing models
  3. Suitable ecological variables
  4. Model approval
- Model statistics
  - Extent of occurrence
  - Percent in Protected Areas
  - Change under projected forest loss



# maskRangeR vignette example: Swamp Forest Crab



IUCN defined range



Elevation



Temperature



MODIS derived  
tree cover



# maskRangeR vignette example: Swamp Forest Crab



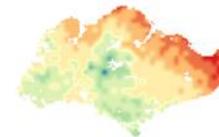
IUCN defined range



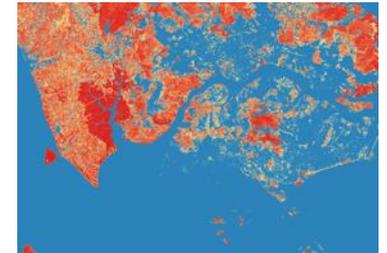
Elevation



Temperature



MODIS derived  
tree cover



<u>Layer</u>	<u>Minimum Value</u>	<u>Maximum Value</u>
Elevation	16	26
Temperature	26	34
Hansen tree cover	75	100

# maskRangeR vignette example: Swamp Forest Crab

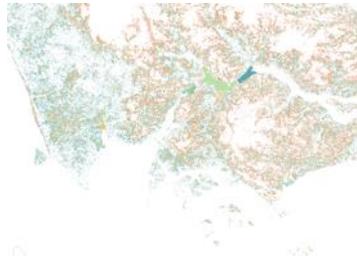
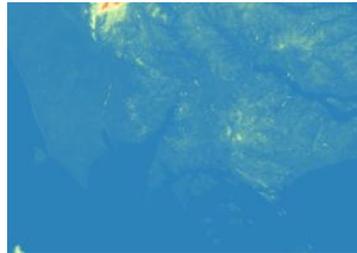


IUCN defined range



Limited to expert  
Indicated values:

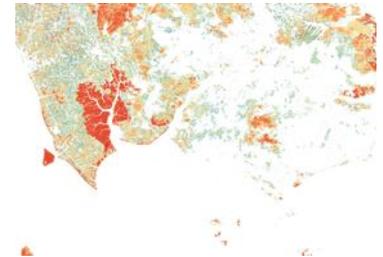
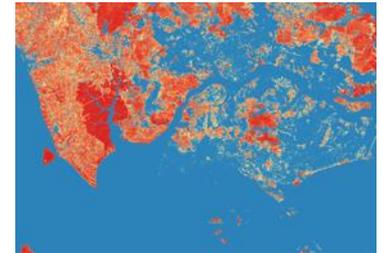
Elevation



Temperature



MODIS derived  
tree cover



# maskRangeR vignette example: Swamp Forest Crab



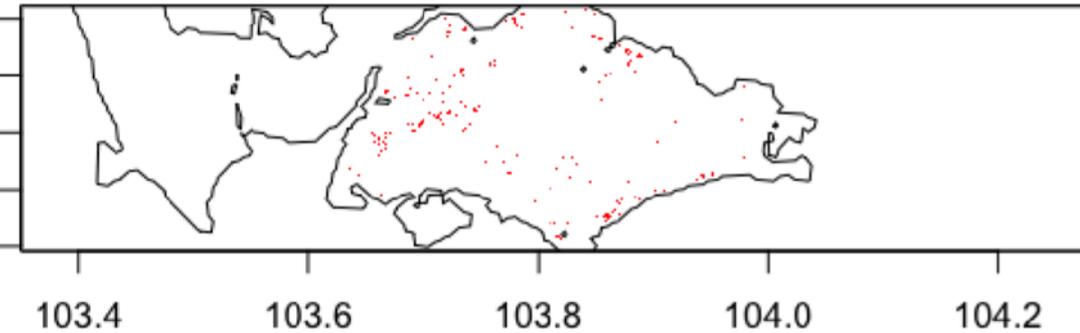
IUCN defined range



Elevation



Temperature



MODIS derived  
tree cover

