

# Creando modelos de distribución de especies para usos en conservación con Wallace, la nueva aplicación Shiny basada en R

ERICA JOHNSON | [ejohnson2@gradcenter.cuny.edu](mailto:ejohnson2@gradcenter.cuny.edu)  
GONZALO PINILLA-BUITRAGO | [gpinillabuitrago@gradcenter.cuny.edu](mailto:gpinillabuitrago@gradcenter.cuny.edu)

City College of New York, CUNY

---

XII CVE, 14 DE FEBRERO 2020



# Contenido



1. Introducción a modelos de distribución de especies



2. Introducción a Wallace



3. Trabajando con Wallace: demostración interactiva de cómo construir modelos de distribución



4. Aplicaciones potenciales de modelos de distribución de especies en biología de la conservación

# Contenido



1. Introducción a modelos de distribución de especies

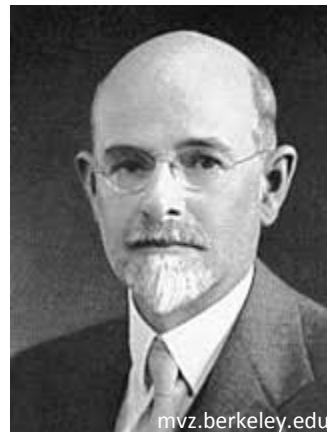
# Nichos ecológicos

**G.E. Hutchinson**



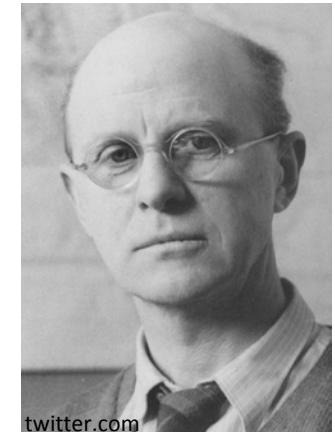
Hyper volumen de n-dimensiones conformado por variables ambientales dentro del cual una especie puede persistir de manera indefinida.

**J. Grinnell**



Conjunto de variables ambientales que restringe la distribución geográfica de una especie por sus efectos la fisiología de la misma.

**C. Elton**



El rol ecológico de una especie, incluyendo sus interacciones con otras especies

# Nichos ecológicos

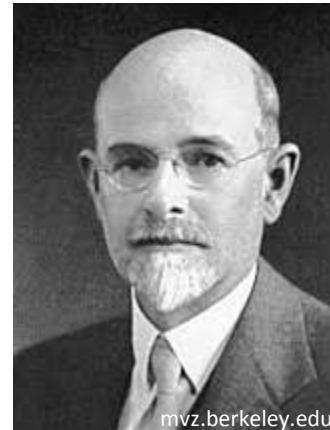
**G.E. Hutchinson**



[news.yale.edu](http://news.yale.edu)

Hyper volumen de n-dimensiones conformado por variables ambientales dentro del cual una especie puede persistir de manera indefinida.

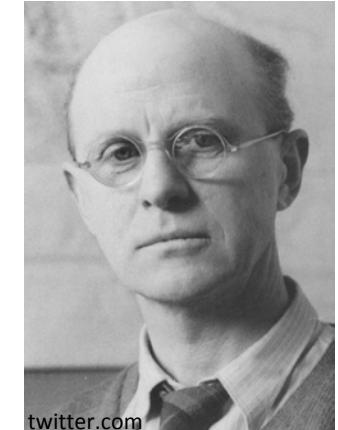
**J. Grinnell**



[mvz.berkeley.edu](http://mvz.berkeley.edu)

Conjunto de variables ambientales que restringe la distribución geográfica de una especie por sus efectos la fisiología de la misma.

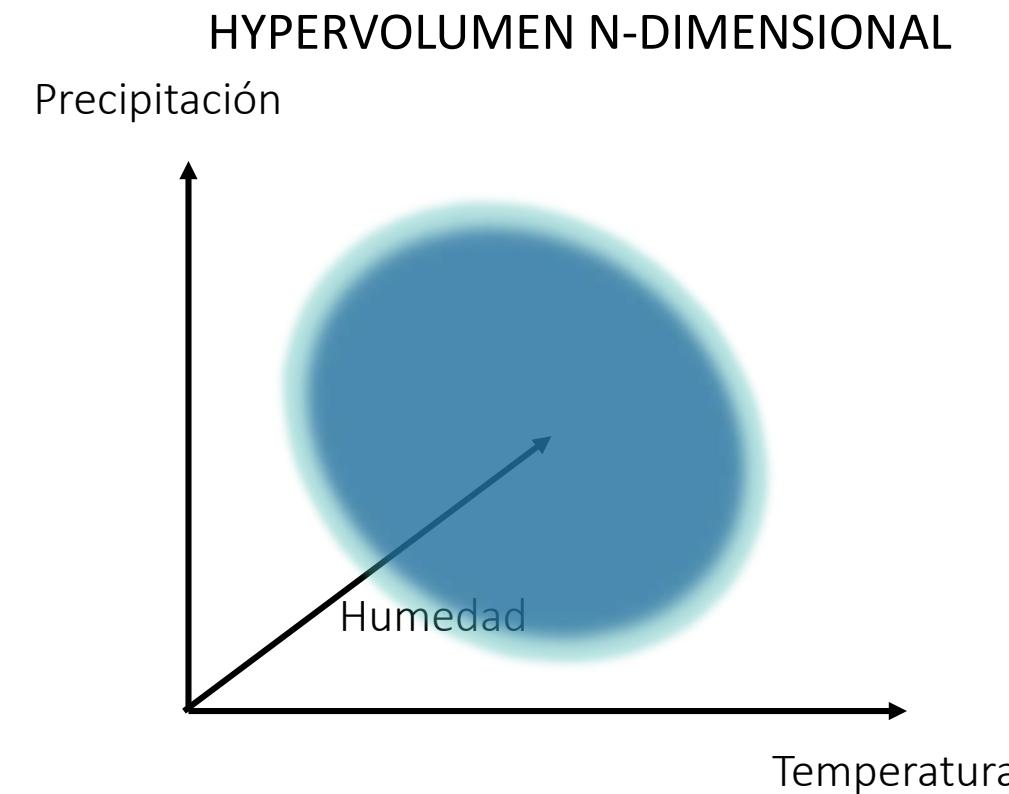
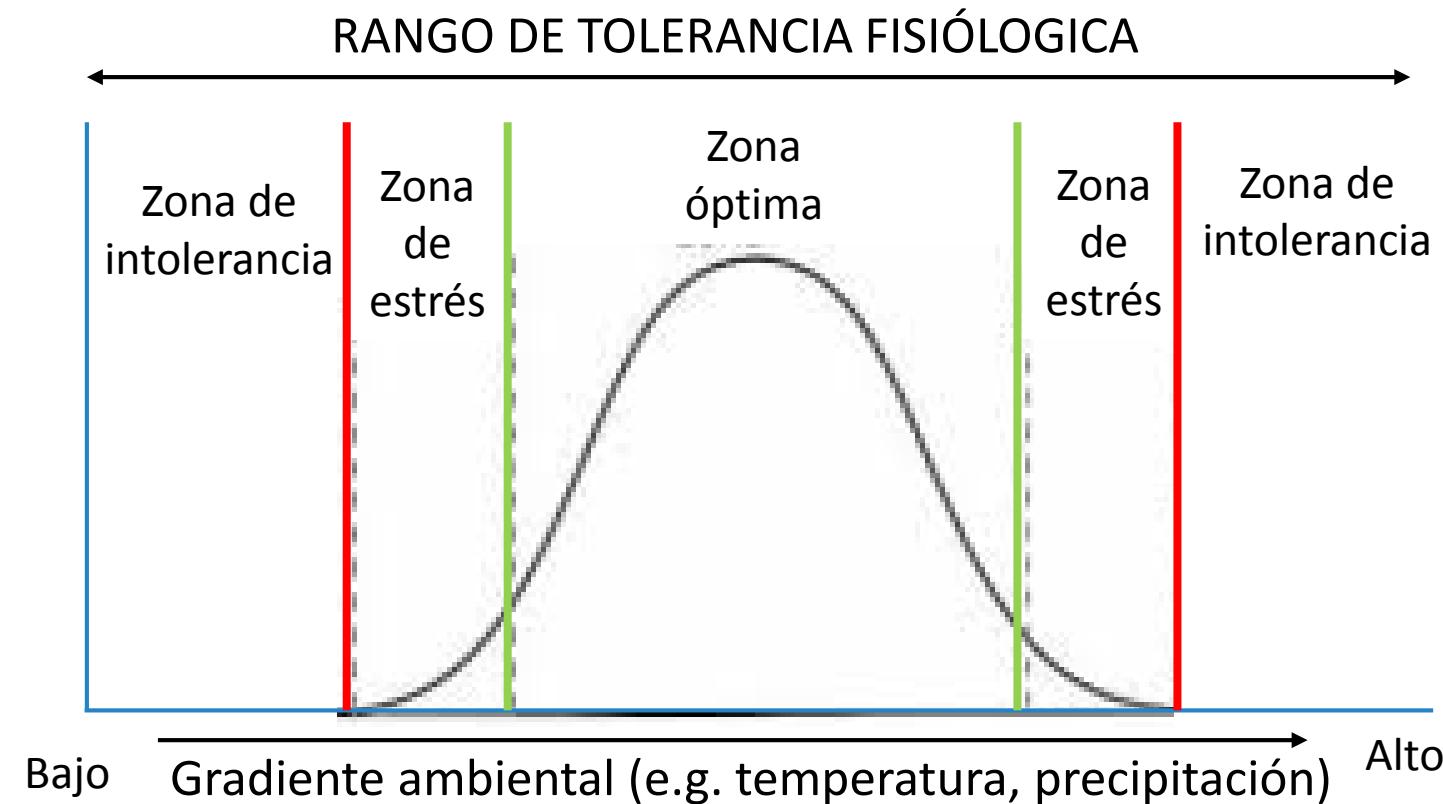
**C. Elton**



[twitter.com](http://twitter.com)

El rol ecológico de una especie, incluyendo sus interacciones con otras especies

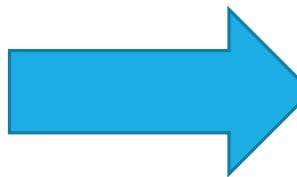
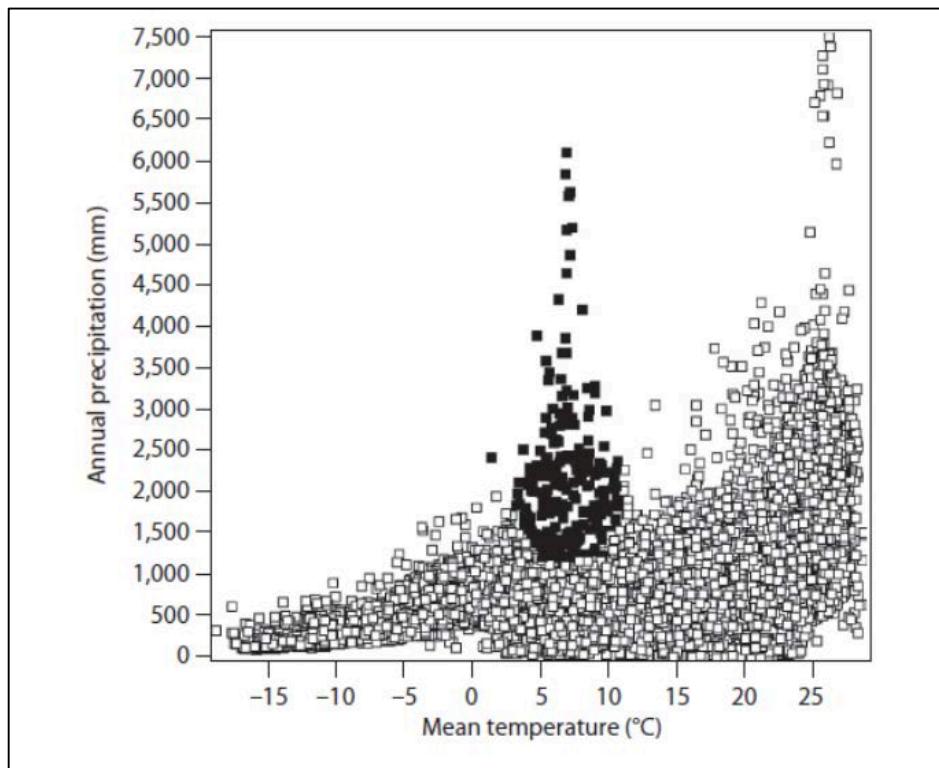
# Relación entre nichos ecológicos y distribución geográfica



# Relación entre nichos ecológicos y distribución geográfica

## E-espacio

Conjunto de variables ambientales tolerables para nuestra especie

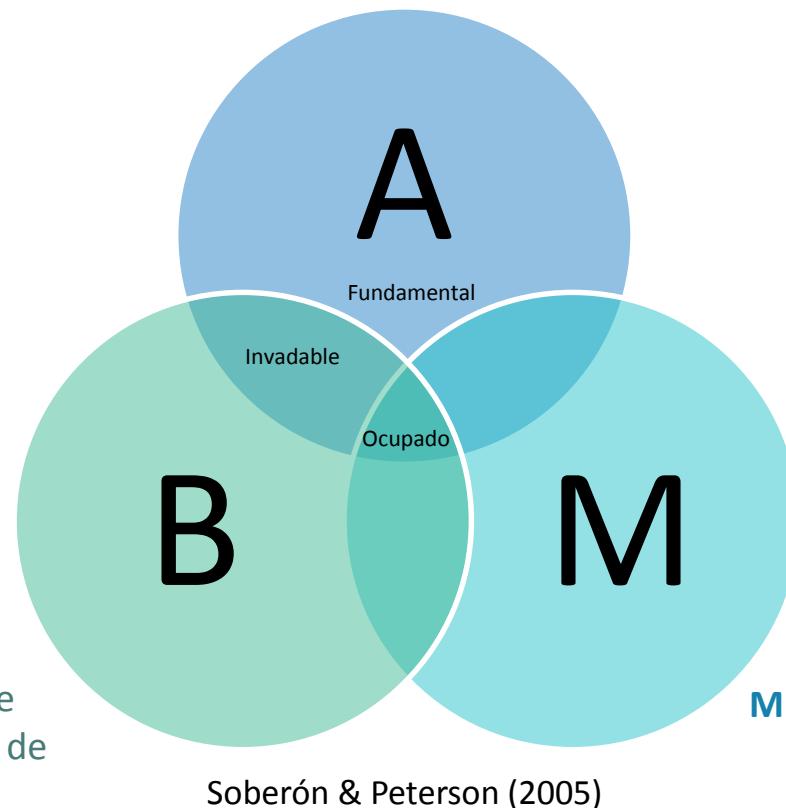


## G-espacio

Lugar geográfico donde ese conjunto de variables ambientales existe



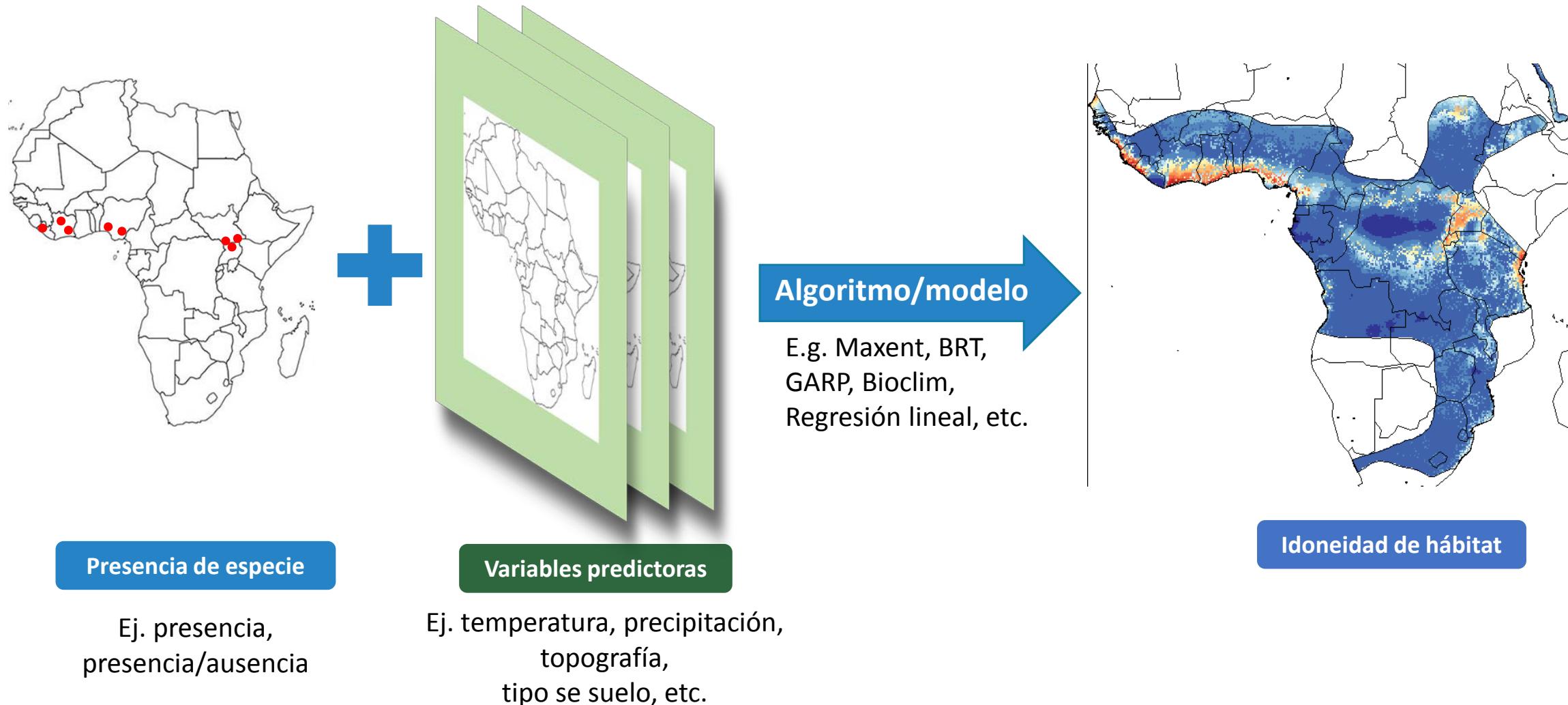
**A** = Rango de variables ambientales dentro del cual puede sobrevivir una especie



**B** = Interacciones bióticas que limitan/facilitan la persistencia de una especie

**M** = Espacio geográfico disponible para la dispersión de la especie

# Modelos de distribución de especies



# Aplicaciones

- Comprender las dimensiones del nicho ecológico de una especie.
- Anticipar la **distribución geográfica potencial** de las especies.
- Aproximar la **distribución geográfica real** de las especies.
- Evalúa la interacción potencial entre una especie y otra

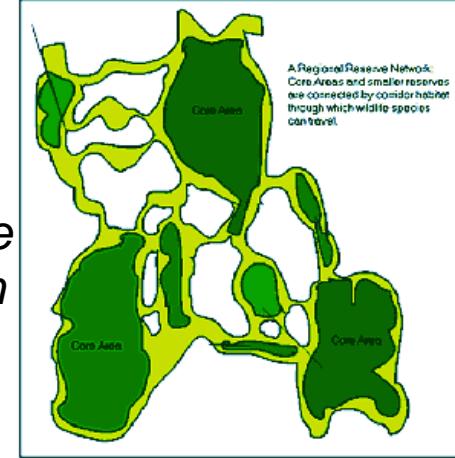


# Aplicaciones



*Especies invasoras*

*Planificacion para areas de conservacion*



*Enfermedades zoonoticas*



*Politicas de desarollo y uso de la tierra*

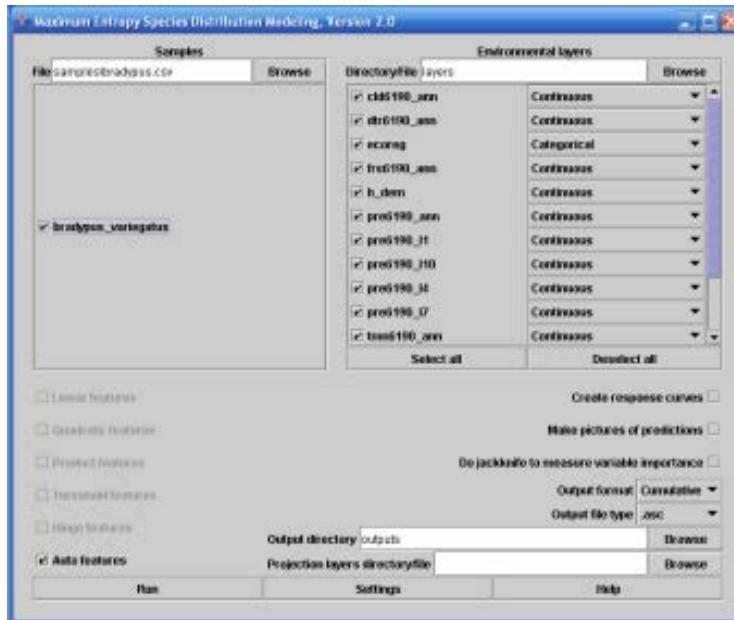
# Contenido



2. Introducción a Wallace

# Problemas analíticos comunes

## Interfaz gráfica (GUI)



- (**demasiado**) fácil → “caja negra”
- Configuración predeterminada
- Inflexibles
- Actualizados infrecuentemente
- Requiere múltiples aplicaciones/softwares

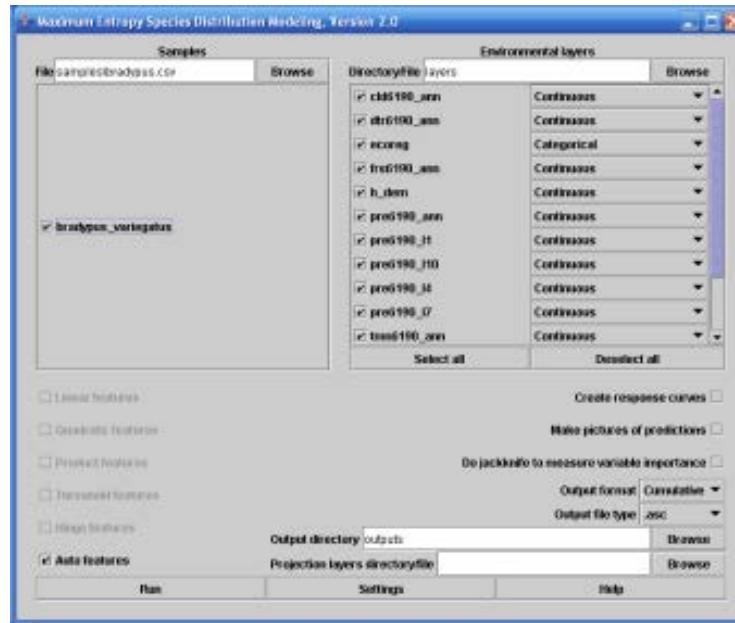
## Código computacional

```
32  #> 
33 results <- occ(query = "Tremarctos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
34 results.data <- results[["gbif"]]$data[[formatSpName("Tremarctos ornatus")]]
35 occs <- remDups(results.data) # remove rows with duplicate coordinates
36 ...
37
38 ## Process Occurrence Data
39
40 You chose 42 of 51 total occurrence localities via polygon selection to keep in the analysis.
41
42 ...
43 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, 2
44, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51), ] # subset occs by selected rows
```

- Propenso a errores
- Difícil de generalizar
- Curva de aprendizaje lenta

# Problemas analíticos comunes

## Interfaz gráfica (GUI)



## Código computacional

```
32  tr
33 results <- occ(query = "Tremarcitos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
34 results.data <- results[["gbif"]]$data[[formatSpName("Tremarcitos ornatus")]]
35 occs <- remDups(results.data) # remove rows with duplicate coordinates
36 ``
37
38 ## Process Occurrence Data
39
40 You chose 42 of 51 total occurrence localities via polygon selection to keep in the analysis.
41
42 ``{r}
43 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, 2
40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51), ] # subset occs by selected rows
```

Primer plano



Segundo plano



Un software que provee un equilibrio entre flexibilidad y facilidad de uso

# Wallace

## Methods in Ecology and Evolution



APPLICATION | Free Access

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

Jamie M. Kass , Bruno Vilela, Matthew E. Aiello-Lammens, Robert Muscarella, Cory Merow, Robert P. Anderson

First published: 07 December 2017 | <https://doi.org/10.1111/2041-210X.12945> | Cited by: 9



The City  
College  
of  
New York



University of  
Connecticut



AMERICAN  
MUSEUM  
OF  
NATURAL  
HISTORY



Yale  
AARHUS UNIVERSITY



UNIVERSIDADE  
FEDERAL DE GOIÁS

WALLACE | Intro | 1 Occ Data | 2 Process Occs | 3 Err Data | 4 Process Env | 5 Partition Occs | 6 Model | 7 Visualize | 8 Project | Session Code | Change Base Map | ESRI Topo

Obtain Occurrence Data

Modules Available:

- Query Database
- User-specified

Module: User-specified Occurrences

Upload Occurrence CSV

Browse... Tremarctos\_oncaus\_original\_gbif.csv Upload template

Load Occurrences

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

\*\*\*WELCOME TO WALLACE\*\*\*  
Please find messages for the user in this log window.  
User-specifiedOccurrenceCSVForTremarctos\_oncaus\_original\_gbif.csv

Map | Occs 1/1 | Results | Component Guidance | Module Guidance

Component: Obtain Occurrence Records

ORIENTATION

Niche distributional modeling analyses require georeferenced occurrence records for the species (e.g., with latitude/longitude). At present, Component Obtain Occurrence Records focuses on data documenting the presence of the species (i.e., not any information on its absence or non-detection; Franklin 2010 chap. 4; Peterson et al. 2011 chapter 5; Anderson 2012). Wallace currently allows users either to: 1) obtain occurrence records from selected online biodiversity databases and download the results (Module Query Database); or 2) upload their own dataset (Module User-specified Occurrences).

REFERENCES

Anderson, R. P. (2012). Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, 1260, 66–80.

Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Data for species distribution models: the biological data. In: *Mapping species distributions: spatial inference and prediction*. Cambridge: Cambridge University Press.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., & Araújo, M.B. (2011). Species Occurrence Data. In: *Ecological Niches and Geographic Distributions*. Princeton, New Jersey: Monographs in Population Biology, 49. Princeton University Press.

WALLACE | Intro | 1 Occ Data | 2 Process Occs | 3 Err Data | 4 Process Env | 5 Partition Occs | 6 Model | 7 Visualize | 8 Project | Session Code | Change Base Map | ESRI Topo

Obtain Occurrence Data

Module: User-specified

Module: User-specified Occurrences

Upload Occurrence CSV

Browse... Tremarctos\_oncaus\_original\_gbif.csv Upload template

Load Occurrences

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson

INSTITUTE VALIDATION

Please find messages for the user in this log window.

Component Guidance | Module Guidance

Map | Occs 1/1 | Results | Component Guidance | Module Guidance

A map showing a coastal area with several red dots indicating occurrence points. The map includes a legend and various geographical features like rivers and forests.

# Wallace: Características

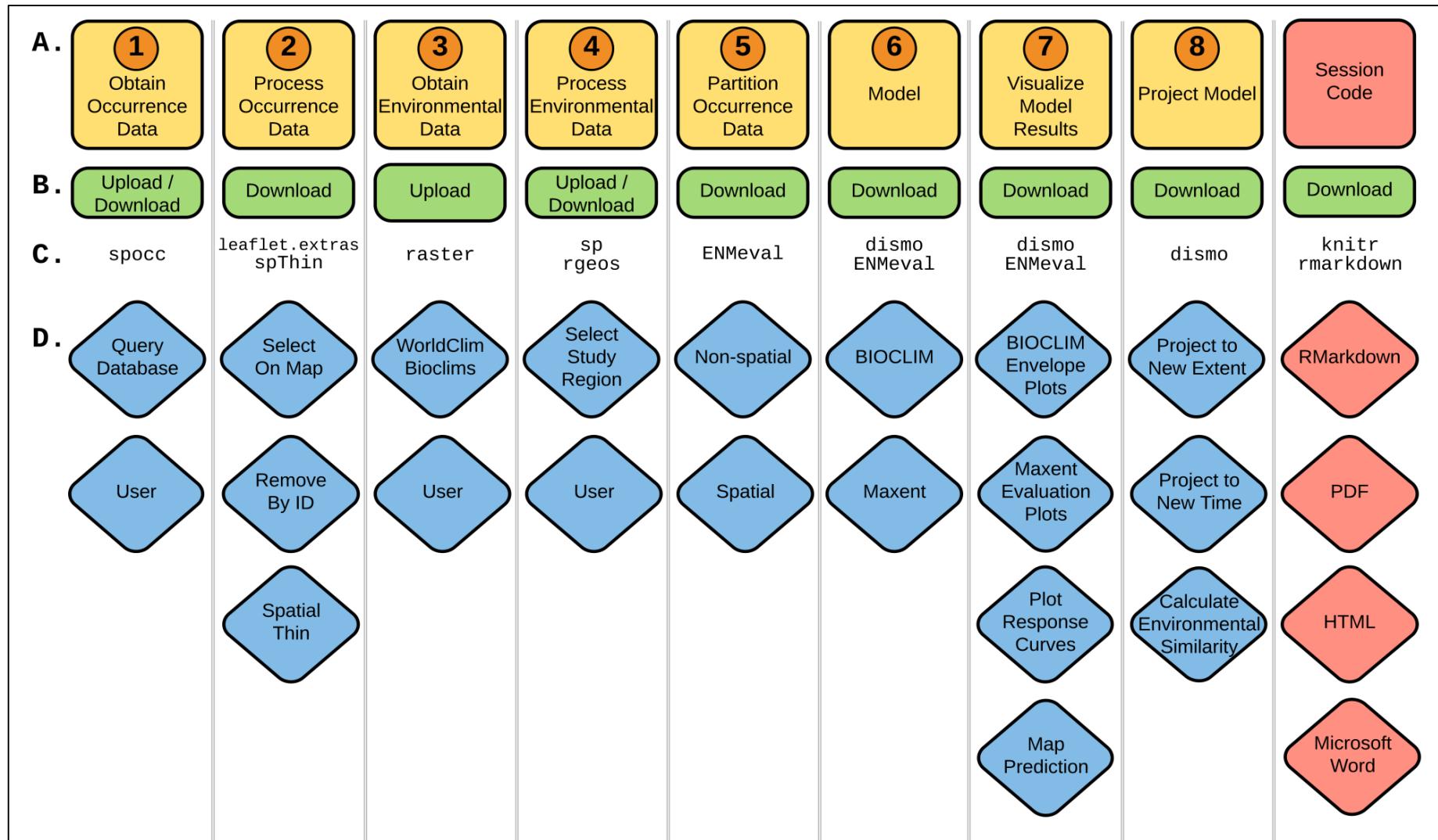
- **Gratis y de código abierto**
- **Instructivo:** Proporciona información guía sobre temas teóricos y metodológicos
- **Flexible:** Provee al usuario múltiples opciones analíticas y permite la descarga de datos
- **Interactivo:** Visualizaciones de mapas y gráficos interactivos para explorar modelos y sus predicciones
- **Reproducible:** Permite la descarga del código R ejecutable para su sesión
- **Expansible:** Módulos que pueden ser contribuidos por la comunidad de usuarios

The image shows the Wallace software interface with two main panels.

**Obtain Occurrence Data Panel:** This panel is titled "Module: Query Database (Present)". It includes dropdown menus for "Choose Database" (set to "Zamia\_amazonum") and "Version" (set to "0.5"). A search bar for "Enter species scientific name" contains "format: Genus species". A "Set maximum number of occurrences" input field is empty. A "Query Database" button is present. Below the form, there is a section for "Module Developers" and "spocct references".

**Process Environmental Data Panel:** This panel is titled "Module: Select Study Region". It includes dropdown menus for "Choose Background Extent" (set to "bounding box") and "Study region buffer distance (degree)" (set to "2"). A "Batch" checkbox is checked. A "Sample" button is present. Below the form, there is a section for "Step 2: Sample Background Points". A map of Colombia is displayed with a large circular buffer around the city of Bogotá, labeled "Tremarctos ornatus (FC-Covier, 1825) study extent: bounding box, buffered by 2 degrees". The map also shows state/province boundaries and major cities.

# Wallace: Estructura modular



# Contenido



3. Trabajando con Wallace: taller interactivo para construir modelos de distribución

# Instalar y correr Wallace

1. Instalar R (versión  $\geq$  v.3.5.0) y Rstudio
2. Instalar el paquete ‘wallace’ desde CRAN (v.1.0.6; estable)

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

OPCIONAL: Si desea utilizar ‘Maxent’ del paquete ‘dismo’\*\*

1. Instalar Java
2. Instalar [maxent.jar](#) y meterlo en la carpeta de dismo
3. ‘Troubleshooting’ en la pagina de [Github](#)

# Wallace: Interfaz gráfica

Wallace v1.9.9.0    Intro    Occ Data    Env Data    Process Occs    Process Envs    Partition Occs    Model    Visualize    Project    Data for Post-Processing    Post-processing

Use Wallace to build niche model/distribution model?  
TRUE

**WORKFLOW**  
Wallace (v1.9.9\*) currently includes the following components and modules:

- 1: Obtain Occurrence Data
  - Query Database
  - User-specified Occurrences
- 2: Process Occurrence Data
  - Select Occurrences on Map
  - Remove Occurrences by ID
  - Spatial Thin
- 3: Obtain Environmental Data
  - WorldClim
  - EcoClimate
  - User-specified Environmental Data
- 4: Process Environmental Data
  - Select Study Region
  - User-specified Study Region
- 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
  - Calculate Environmental Similarity
- 9: Post-Processing
  - Expert Driven: Editing using expert maps
  - Data Driven: Masking by land cover

Download Session Code

Intro    About

## What is Wallace?

Welcome to Wallace, a flexible application for reproducible ecological modeling, built for community expansion. The current version of Wallace (v1.9.9\*) 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 J. M., Vilela B., Aiello-Lammens M. E., Muscarella R., Merow C., Anderson R. P. (2018). Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods Ecol Evol.* 2018, 9: 1151-1156. <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
- **expandable:** users can author and contribute modules that enable new methodological options
- **flexible:** options for user uploads and downloads of results
- **interactive:** includes an embedded zoomable leaflet map, sortable R data tables, and visualizations of results
- **instructive:** features guidance text that educates users about theoretical and analytical aspects of each step in the workflow
- **reproducible:** users can download an markdown .Rmd file that when run reproduces the analysis

## Wallace website

For more information and relevant links see our website.

## Watch webinars about Wallace

The following webinar was the "37th Global Online Biodiversity Informatics Seminar" in the Biodiversity Informatics Training Curriculum organized by Town Peterson. Kass, J. M. 9 May 2018. "WALLACE: A flexible platform for reproducible modeling of species niches and distributions built for community expansion." Broadcast from the City College of New York, City University of New York. Watch on YouTube.

The following webinar was part of the "Modelado de Distributions Potenciales" series, organized by Angela Cuervo. Anderson, R. P. 21 May 2018. "El software Wallace para modelar nichos y distribuciones: Un coche con motor R, volante de ratón y cerebro de humano." Broadcast from the City College of New York, City University of New York. Watch on YouTube.

## Contribute to Wallace

Contributors should submit pull requests to the Wallace GitHub account for module authorship or significant code contributions to either the UI or server files. Also, please connect on Github to post code-related issues and the Google Group for methodological and other broader-scope questions, thoughts, or suggestions for improvement.

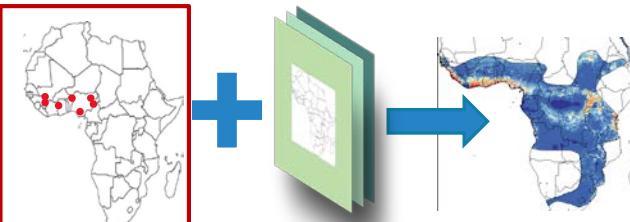
## Contact us

Please email us with any other questions.

# Componente 1: Obtener datos de presencia (Occ Data)

1. Especificado por el usuario

2. Descarga desde base de datos disponibles



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

Change Base Map  
ESRI Topo ▾

Obtain Occurrence Data

**Modules Available:**

Query Database  
 User-specified

---

**Module: Query Database**  
spocc : Interface to Species Occurrence Data Sources

---

**Choose Database**

GBIF    VertNet    BISON

**Enter species scientific name**

Format: Genus species

**Set maximum number of occurrences**

1      500

1   51   101   151   201   251   301   351   401   451   500

**Query Database**

---

**Download database occurrence localities (.csv)**

Download

---

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson  
spocc references

\*\*\*WELCOME TO WALLACE\*\*\*  
Please find messages for the user in this log window.

Map   Occs Tbl   Results   Component Guidance   Module Guidance

## Component: Obtain Occurrence Records

### ORIENTATION

Niche/distributional modeling analyses require georeferenced occurrence records for the species (e.g., with latitude/longitude). At present, Component **Obtain Occurrence Records** focuses on data documenting the *presence* of the species (i.e., not any information on its absence or non-detection; Franklin 2010 chap. 4; Peterson et al. 2011 chapter 5; Anderson 2012). Wallace currently allows users either to: 1) obtain occurrence records from selected online biodiversity databases and download the results (Module **Query Database**); or 2) upload their own dataset (Module **User-specified Occurrences**).

### REFERENCES

Anderson, R. P. (2012). Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*. 1260: 66-80.

Franklin J. (2010). Mapping Species Distributions: Spatial Inference and Prediction. Data for species distribution models: the biological data. In: Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press.

Peterson A.T., Soberón J., Pearson R.G., Anderson R.P., Martinez-Meyer E., Nakamura M., & Araújo M.B. (2011). Species Occurrence Data. In: *Ecological Niches and Geographic Distributions*. Princeton, New Jersey: Monographs in Population Biology, 49. Princeton University Press.

# Componente 1: Obtener datos de presencia (Occ Data)

## MAPA INTERACTIVO

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

Obtain Occurrence Data  
Modules Available:  
 Query Database  
 User-specified

**Module: Query Database**  
spocc : Interface to Species Occurrence Data Sources

Choose Database  
 GBIF  VertNet  BISON

Enter species scientific name  
Tremarcos ornatus

Set maximum number of occurrences  
1 200 500  
1 51 101 151 201 251 301 351 401 451 500

Query Database

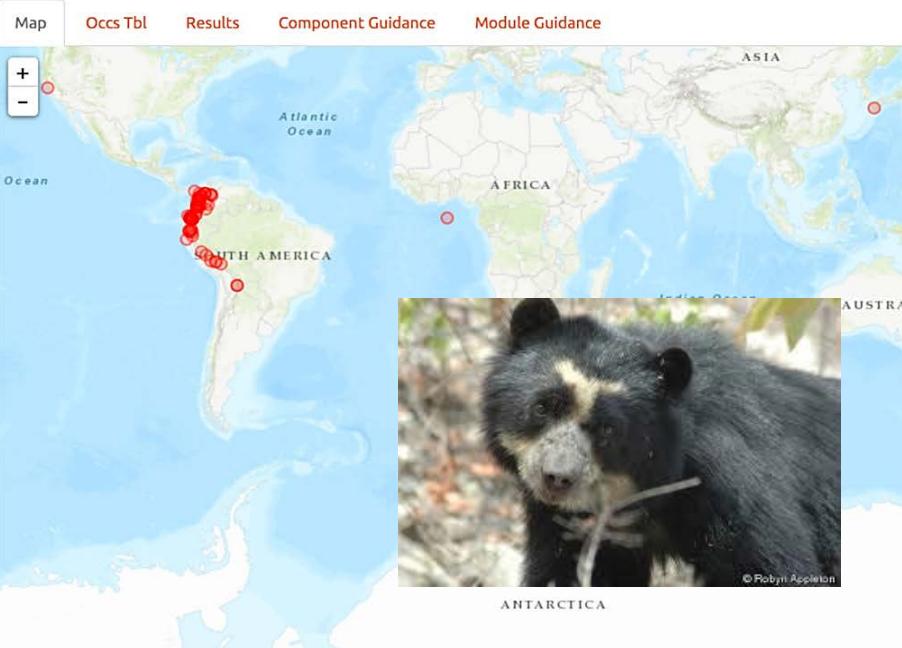
Download database occurrence localities (.csv)  
[Download](#)

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson  
spocc references

Please find messages for the user in this log window.  
> Total gbif records for Tremarcos ornatus returned [ 200 ] out of [ 237 ] total (limit 200 ). Records without coordinates removed [ 129 ]. Duplicated records removed [ 5 ]. Remaining records [ 66 ].

Change Base Map  
ESRI Topo

Map Occs Tbl Results Component Guidance Module Guidance



## TABLA CON DATOS DE PRESENCIA

Data 4 Process Envs 5 Partition Occs 6 Model 7 Visualize 8 Project Session Code

Please find messages for the user in this log window.  
> Total gbif records for Tremarcos ornatus returned [ 200 ] out of [ 237 ] total (limit 200 ). Records without coordinates removed [ 129 ]. Duplicated records removed [ 5 ]. Remaining records [ 66 ].

Change Base Map  
ESRI Topo

Map Occs Tbl Results Component Guidance Module Guidance

Show 10 entries Search:

name	occID	longitude	latitude	year	institutionCode	country	stateProvince
Tremarcos ornatus	1	-75.89	5.54	2018	iNaturalist	Colombia	
Tremarcos ornatus	2	-76.56	3.94	2017	CPR	Colombia	Valle del Cauca
Tremarcos ornatus	3	-75.63	4.07	2017	CPR	Colombia	Tolima
Tremarcos ornatus	4	-77.89	-0.51	2017	iNaturalist	Ecuador	
Tremarcos ornatus	5	-72.44	-13.15	2017	iNaturalist	Peru	
Tremarcos ornatus	6	-76.17	6.49	2017	iNaturalist	Colombia	
Tremarcos ornatus	7	-78.79	0.05	2017	iNaturalist	Ecuador	
Tremarcos ornatus	8	-78.73	0.19	2017	iNaturalist	Ecuador	
Tremarcos ornatus	9	-78.2	-0.37	2017	iNaturalist	Ecuador	
Tremarcos ornatus	10	-78.02	-0.3	2016	iNaturalist	Ecuador	

# Componente 2: Procesar datos de presencia (Process Occs)

Wallace v1.9.9.0   Intro   Occ Data   Env Data   **Process Occs**   Process Envs   Partition Occs   Model   Visualize   Project   Data for Post-Processing   Post-processing

Process Occurrence Data  
Modules Available:  
 Select Occurrences On Map  
 Remove Occurrences By ID  
 Spatial Thin

**Module: Spatial Thin**  
spThin : Functions for Spatial Thinning of Species Occurrence Records for Use in Ecological Models

Thinning distance (km)  
  
 Batch

Module Developers: Jamie M. Kass, Matthew E. Aiello-Lammens, Robert P. Anderson  
spThin references  
Package Developers: Matthew E. Aiello-Lammens, Robert A. Boria, Aleksandar Radosavljevic, Bruno Vilela, and Robert P. Anderson  
[CRAN documentation](#)

Duplicated records removed [8]. Remaining records [14].  
> WorldClim bioclimatic variables bio1-19 at 10 arcmin resolution.  
> **WARNING :** Removed records without environmental values with occIDs: 55.

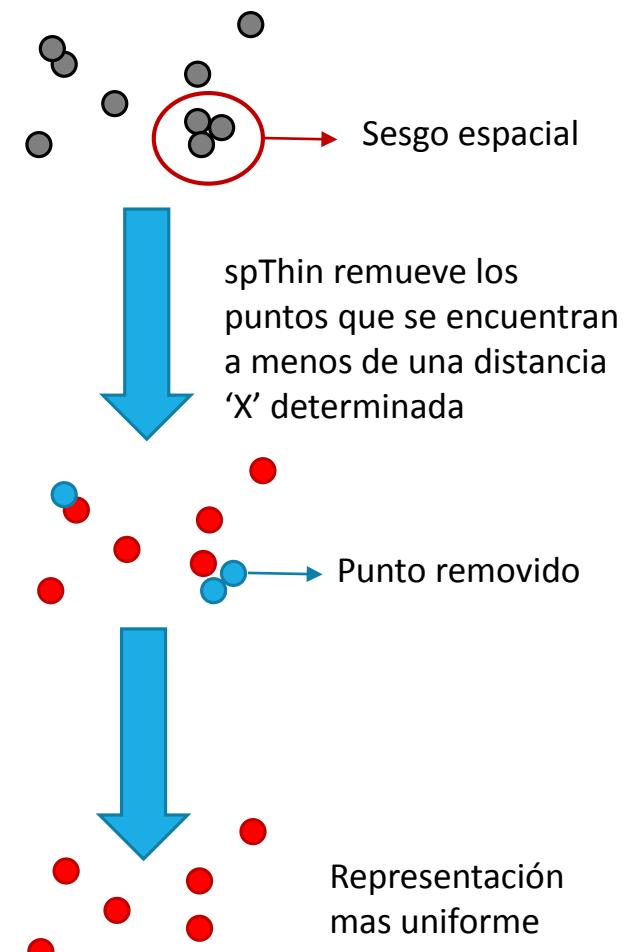
Map   Table   Results   Component Guidance   Module Guidance   Download

Managua   Barranquilla   Caracas   Port of Spain  
San Jose   Panama City   Cartagena   Valencia  
Bucaramanga   Georgetown   Paramaribo  
Medellin   Bogota   French Guiana  
Quito   Guayaquil   Manaus  
Lima   Chidado   Lake Titicaca

ESRI Topo

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

## Reducir sesgos espaciales



# Componente 2: Procesar datos de presencia (Process Occs)

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

Process Occurrence Data  
Modules Available:  
 Select Occurrences On Map  
 Remove Occurrences By ID  
 Spatial Thin

**Module: Spatial Thin**  
spThin : Spatial Thinning of Species Occurrence Records

Thinning distance (km)  
10

Thin Occurrences

Reset to original occurrences

Reset

Download processed occurrence localities (.csv)

Download

Module Developers: Jamie M. Kass, Matthew E. Aiello-Lammens, Robert P. Anderson

spThin references

Package Developers: Matthew E. Aiello-Lammens, Rob A. Boria, Alex Radosavljevic, Bruno Vilela, Robert P. Anderson

> Removed occurrence with ID = 147 . Updated data has n = 57 records.  
> Removed occurrence with ID = 159 . Updated data has n = 56 records.  
> Total records thinned to [ 44 ] localities.

Change Base Map  
ESRI Topo

Map   Occs Tbl   Results   Component Guidance   Module Guidance

Punto removido

Punto retenido

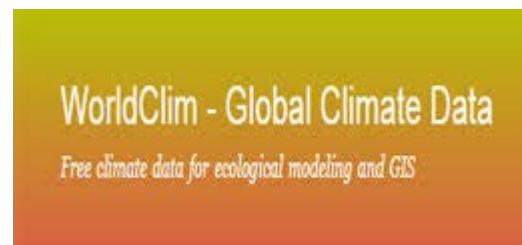
Occ Records  
red retained  
blue removed

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, IPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

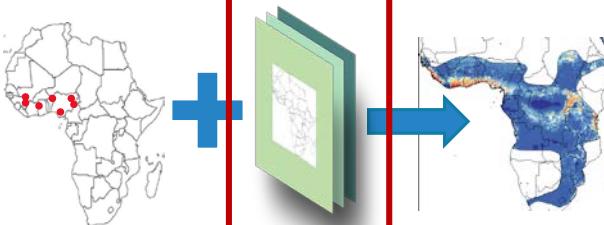
# Componente 3: Obtener datos ambientales (Env Data)

## 1. Especificado por el usuario

## 2. Descarga desde base de datos disponibles



<https://www.worldclim.org/>



Screenshot of the R package Wallace interface, specifically the "3 Env Data" tab. The interface includes a sidebar for "Obtain Environmental Data" with "WorldClim Bioclims" selected. The main area shows the "Module: WorldClim Bioclims" and its raster reference. A dropdown menu for resolution is set to "2.5 arcmin". A checkbox labeled "Specify variables to use in analysis?" is checked, with a red arrow pointing to it. Below the checkbox, a list of bioclimatic variables (bio1 through bio19) has several checked boxes. At the bottom, there is a "Load Env Data" button and information about module and package developers.

Obtain Environmental Data

Modules Available:

WorldClim Bioclims

User-specified

Module: WorldClim Bioclims

raster: Geographic Data Analysis and Modeling

Select WorldClim bioclimatic variable resolution

2.5 arcmin

Specify variables to use in analysis? →

Select

bio1  bio2  bio3  bio4  bio5  bio6  
 bio7  bio8  bio9  bio10  bio11  
 bio12  bio13  bio14  bio15  bio16  
 bio17  bio18  bio19

Using map center coordinates as reference for tile download.  
Using map center -75.512, 4.356

Load Env Data

Module Developers: Jamie M. Kass, Robert P. Anderson  
raster references  
Package Developers: Robert J. Hijmans, Jacob van Etten, Joe Cheng, Matteo Mattiuzzo, Michael Sumner, Jonathan A. Greenberg, Oscar Peripan Lamiguero, Andrew Bevan, Etienne B. Racine, Ashton Shortridge

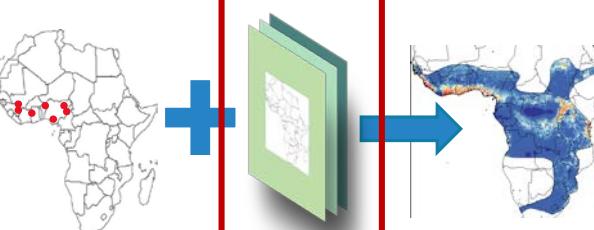
> Removed occurrence with ID = 159 . Updated data has n = 56 records.  
> Total records thinned to [ 44 ] localities.  
> Environmental predictors: WorldClim bioclimatic variables bio1-19 at 2.5 arcmin resolution.

Change Base Map  
ESRI Topo

# Componente 4: Procesar variables ambientales (Process Envs)

## 1. Escoger la region de estudio

## 2. Muestrar puntos de 'trasfondo' (background)



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

Change Base Map  
ESRI Topo

Process Environmental Data  
Modules Available:  
 Select Study Region  
 User-specified

**Module: Select Study Region**  
sp : Title Classes and Methods for Spatial Data  
rgeos : Interface to Geometry Engine - Open Source (GEOS)

**Step 1: Choose Background Extent**  
Background Extents:  
 Bounding box  
 Minimum convex polygon  
 Point buffers  
Study region buffer distance (degree)  
1  
Select

**Step 2: Sample Background Points**  
Mask predictor rasters by background extent and sample background points  
No. of background points  
10000

Total records thinned to [ 44 ] localities.  
Environmental predictors: WorldClim bioclimatic variables bio1-19 at 2.5 arcmin resolution.  
Study extent: minimum convex polygon. Study extent buffered by 1 degrees.

Map   Occs Tbl   Results   Component Guidance   Module Guidance

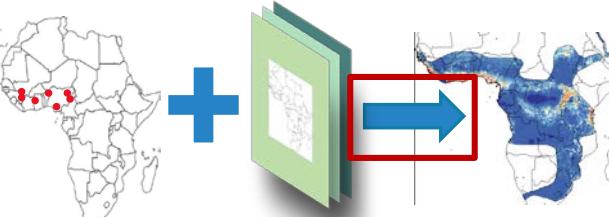
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCan, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

Occ Records  
retained removed

# Componente 5: Partición de datos (Partition Occs)

Los datos de presencia se pueden particionar de la siguiente manera para calibrar y evaluar el modelo:

1. Aleatoria (no espacial)
2. Espacial



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

Change Base Map  
ESRI Topo

Partition Occurrence Data  
Modules Available:  
 Non-spatial Partition  
 Spatial Partition

Module: Non-spatial Partition  
ENMeval: Automated Runs and Evaluations of Ecological Niche Models

Options Available:  
Random k-fold

Number of Folds  
4

Partition

Download occurrence and background localities with partition values (.csv)  
Download

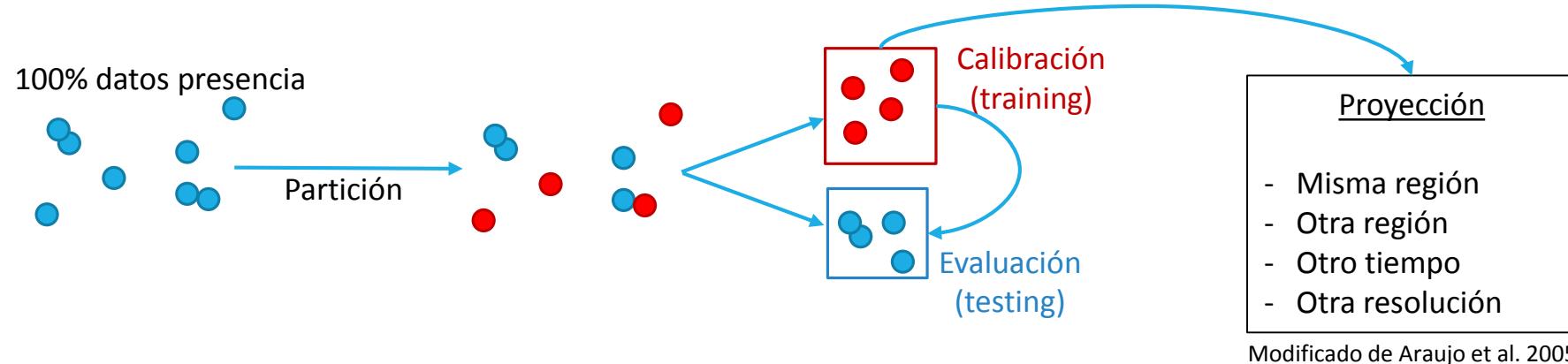
Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson  
ENMeval references  
Package Developers: Robert Muscarella, Peter J. Galante, Mariano Soley-Guardia, Robert A. Boria, Jamie M. Kass, Maria Uriarte, Robert P. Anderson  
[CRAN documentation](#) | [Software note](#)

> Environmental data masked.  
> Random background points sampled ( $n = 10000 : 0.36\%$  of cells with values).  
> Occurrences partitioned by jackknife method.  
> Occurrences partitioned by random k-fold ( $k = 4$ ).

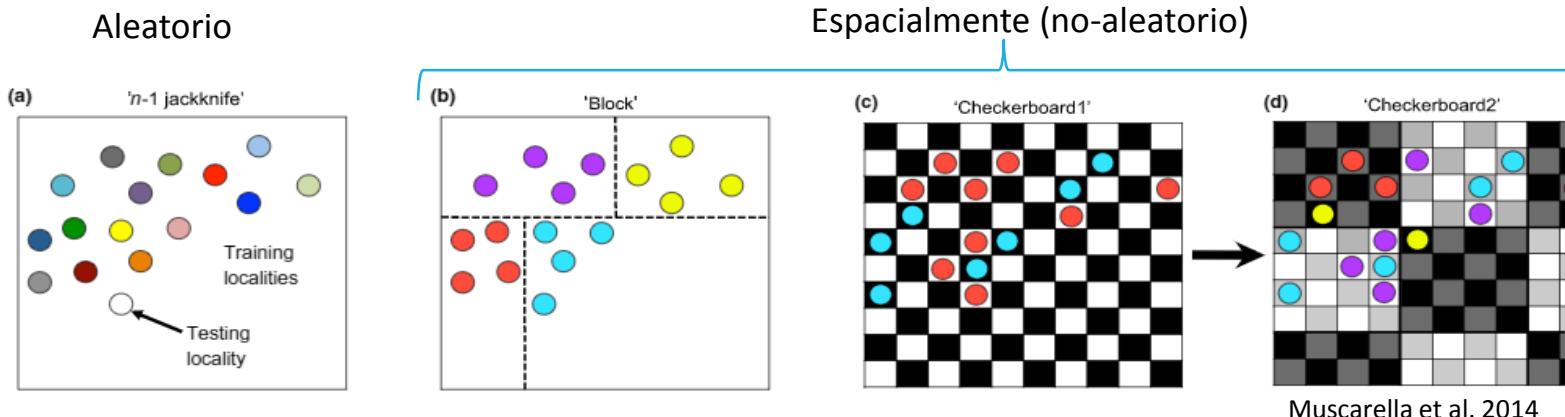
Map   Occs Tbl   Results   Component Guidance   Module Guidance

# Componente 5: Partición de datos (Partition Occs)

Un buen modelo debe tener capacidad predictiva

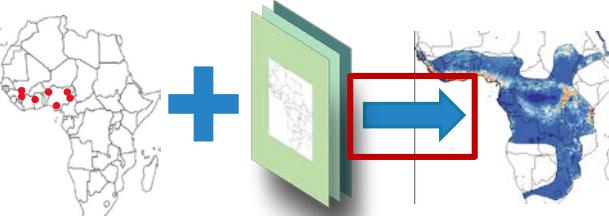


Muchas maneras de hacer esta partición



# Componente 6: Construcción y evaluación de modelos (Model)

Permite ajustar complejidad del modelo y sintonizar los parámetros (“model tuning”)



Screenshot of the ENMeval software interface showing the Maxent module for building and evaluating niche models.

The interface includes:

- Header menu: 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.
- Module Available: BIOCLIM (radio button), Maxent (radio button, selected).
- Module: Maxent (ENMeval: Automated Runs and Evaluations of Ecological Niche Models, dismo: Species Distribution Modeling).
- Note: (NOTE: see module guidance for troubleshooting tips if you are experiencing problems.)
- Select feature classes (flexibility of modeled response): L, LQ, H, LQH, LQHP, LQHPT.
- Select regularization multipliers (penalty against complexity): A slider from 0 to 10 with values 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. The value 1 is highlighted.
- Multiplier step value: Input field set to 0.5.
- Clamp predictions? checkbox (unchecked).
- Run button.
- Output sections:
  - > Occurrences partitioned by block method.  
> Occurrences partitioned by block method.  
> Occurrences partitioned by block method.  
> Maxent ran successfully and output evaluation results for 20 models.
  - Map, Occs Tbl, Results, Component Guidance, Module Guidance tabs. Results tab is selected.
  - Full model and partition bin average evaluation statistics table:

	settings	Features	rm	fullAUC	avg.test.AUC	var.test.AUC	avg.diff.AUC	var.diff.AUC	avg.test.or10pct	var.test.or10pct	avg.test.orMTI
1	L_1	L	1	0.817	0.767	0.046	0.09	0.026	0.25	0.042	0.03
2	LQ_1	LQ	1	0.9	0.781	0.033	0.116	0.026	0.188	0.057	0.14
3	H_1	H	1	0.926	0.806	0.024	0.124	0.029	0.312	0.047	0.21
4	LQH_1	LQH	1	0.93	0.801	0.023	0.132	0.027	0.344	0.035	0.28
5	L_1.5	L	1.5	0.816	0.779	0.032	0.075	0.016	0.219	0.025	0.03
6	LQ_1.5	LQ	1.5	0.88	0.785	0.029	0.11	0.019	0.25	0.042	0.05
7	H_1.5	H	1.5	0.914	0.83	0.02	0.092	0.023	0.188	0.057	0.12
8	LQH_1.5	LQH	1.5	0.917	0.821	0.018	0.108	0.019	0.281	0.046	0.12
9	L_2	L	2	0.819	0.79	0.025	0.06	0.012	0.188	0.026	0.03
10	LQ_2	LQ	2	0.864	0.798	0.028	0.093	0.019	0.219	0.025	0.05
  - Individual partition bin evaluation statistics table:

	AUC_bin.1	AUC_bin.2	AUC_bin.3	AUC_bin.4	AUC.DIFF_bin.1	AUC.DIFF_bin.2	AUC.DIFF_bin.3	AUC.DIFF_bin.4	OR10_bin.1	OR10_bin.2
1	0.641	0.709	0.786	0.934	0.206	0.125	0.029	0	0.5	0.25
2	0.695	0.791	0.71	0.927	0.201	0.081	0.182	0	0.5	0

# Construcción:

## Feature Classes

L

Q

H

P

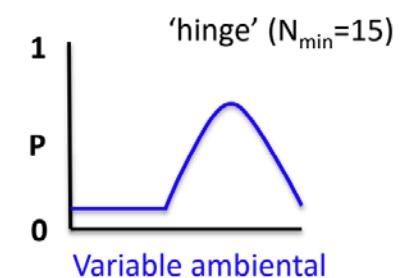
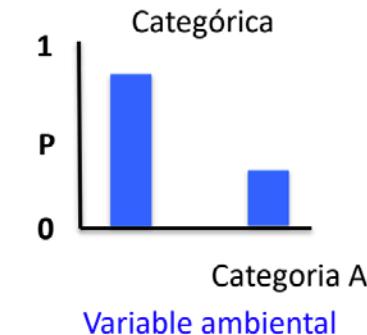
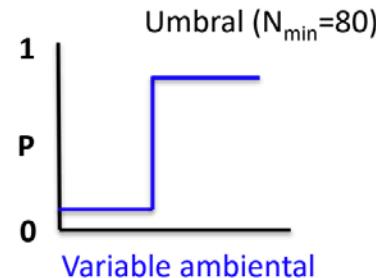
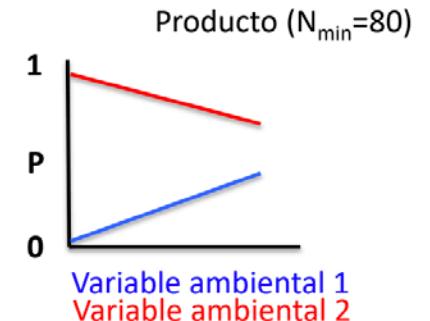
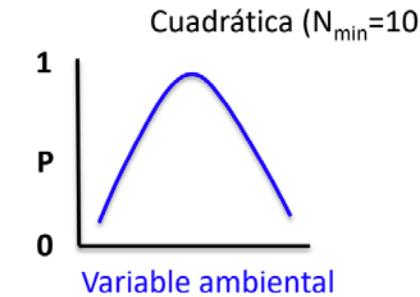
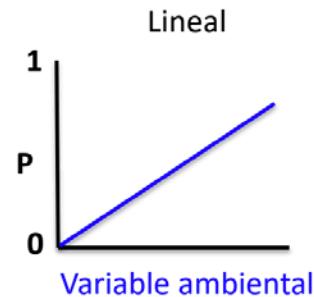
T

■

■

Relación entre ambiente (predictores) y  
'probabilidad de presencia' (estimado)

'features'



# Construcción:

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

Regularization  
Multiplier

0.5

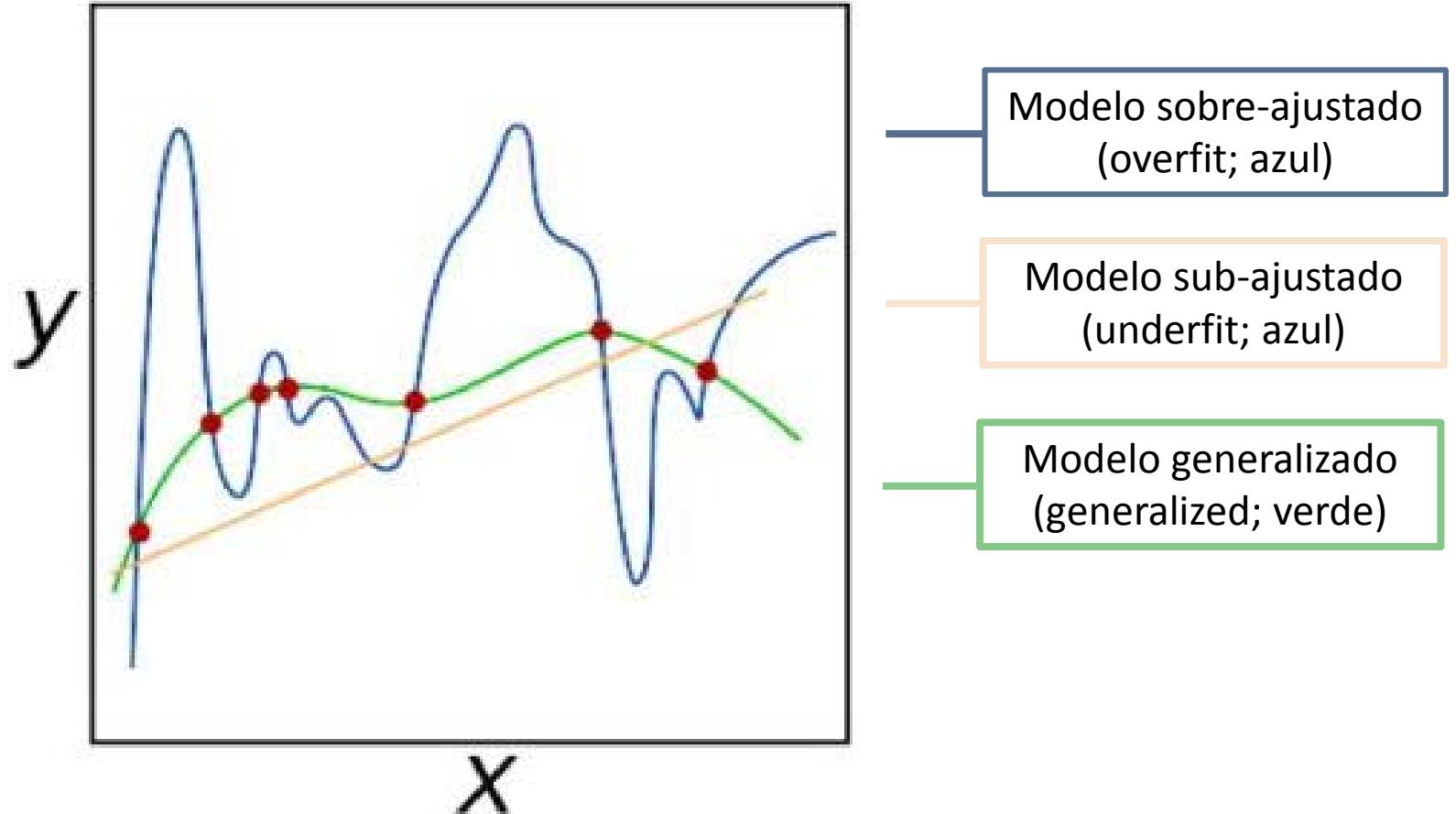
1

1.5

2

2.5

:



Modelo sobre-ajustado  
(overfit; azul)

Modelo sub-ajustado  
(underfit; azul)

Modelo generalizado  
(generalized; verde)

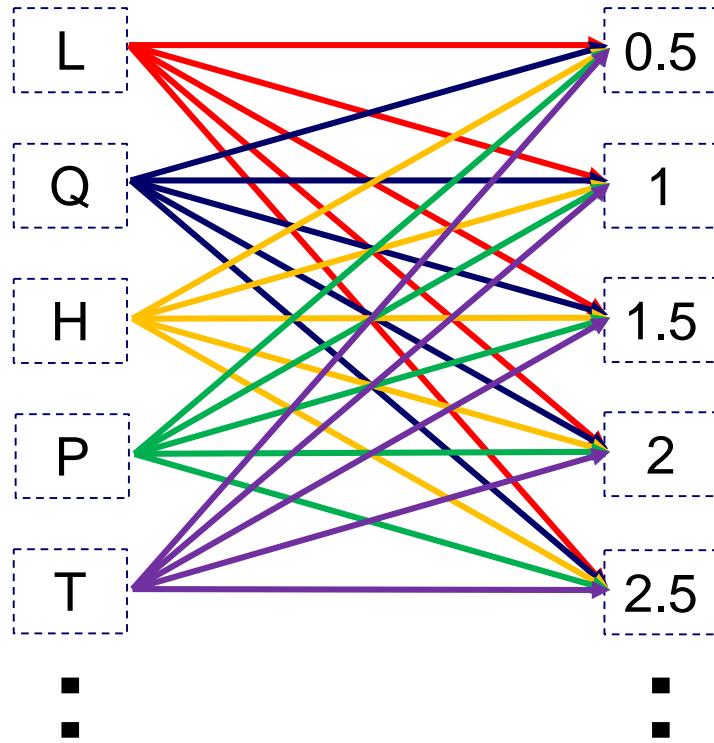
Idealmente queremos un modelo como la línea verde, que se ajuste a nuestros datos pero no demasiado ('overfit') ya que queremos que el modelo tenga capacidad predictiva.

Los RM nos permiten encontrar funciones de este tipo penalizando complejidad.

# Construcción:

Respuesta a variable ambiental

Feature Classes



Penalización a complejidad del modelo

Regularization Multiplier

	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

## Recomendación:

- Todas las clases para al menos 80 registros de ocurrencia
- L, Q y H para tamaños de muestra de 15 a 79
- L y Q para 10 a 14 registros
- Solo L para menos de 10 registros.

# Evaluación:

Distintas medidas para evaluar el desempeño:

a) Del modelo completo

b) Por partición

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

Change Base Map  
ESRI Topo

Current Env Variable  
bio02

Current model  
L\_1

Build and Evaluate Niche Model

Modules Available:  
 BIOCLIM  
 Maxent

Module: Maxent

ENMeval: Automated Runs and Evaluations of Ecological Niche Models  
dismo : Species Distribution Modeling

(NOTE: see module guidance for troubleshooting tips if you are experiencing problems.)

Select feature classes (flexibility of modeled response)  
key: L inear, Q uadratic, H inge, P roduct, T hreshold

L  LQ  H  LQH  LQHP  LQHPT

Select regularization multipliers (penalty against complexity)

Multiplier step value: 0.5

Clamp predictions?

Run

> Occurrences partitioned by block method.  
> Occurrences partitioned by block method.  
> BIOCLIM ran successfully and output evaluation results.  
> Maxent ran successfully and output evaluation results for 28 models.

Map   Occs Tbl   **Results**   Component Guidance   Module Guidance

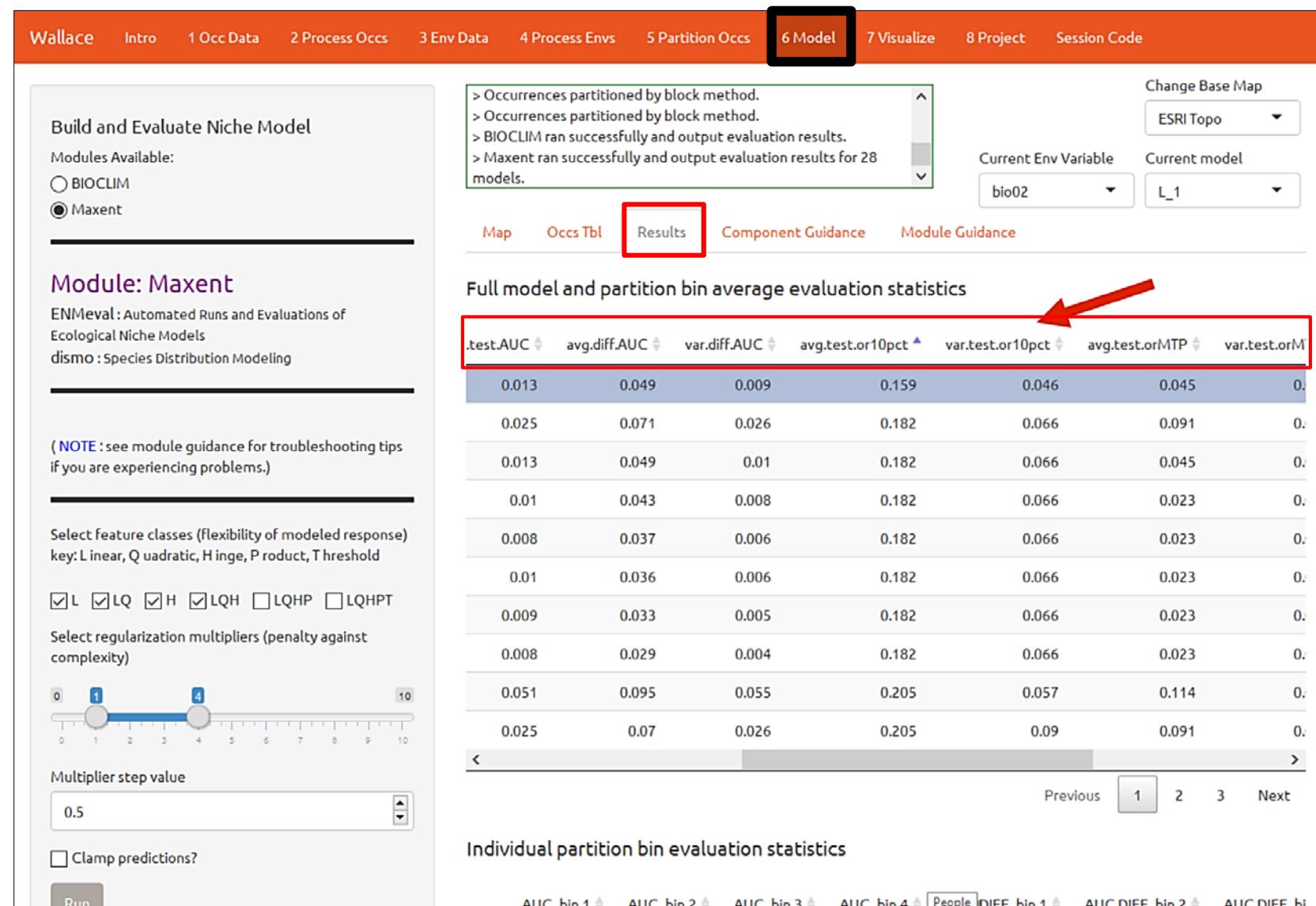
Full model and partition bin average evaluation statistics

.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.

Individual partition bin evaluation statistics

AUC bin.1 AUC bin.2 AUC bin.3 AUC bin.4 People DIFF bin.1 DIFF bin.2 DIFF bin.3 DIFF bin.4

Previous   1   2   3   Next



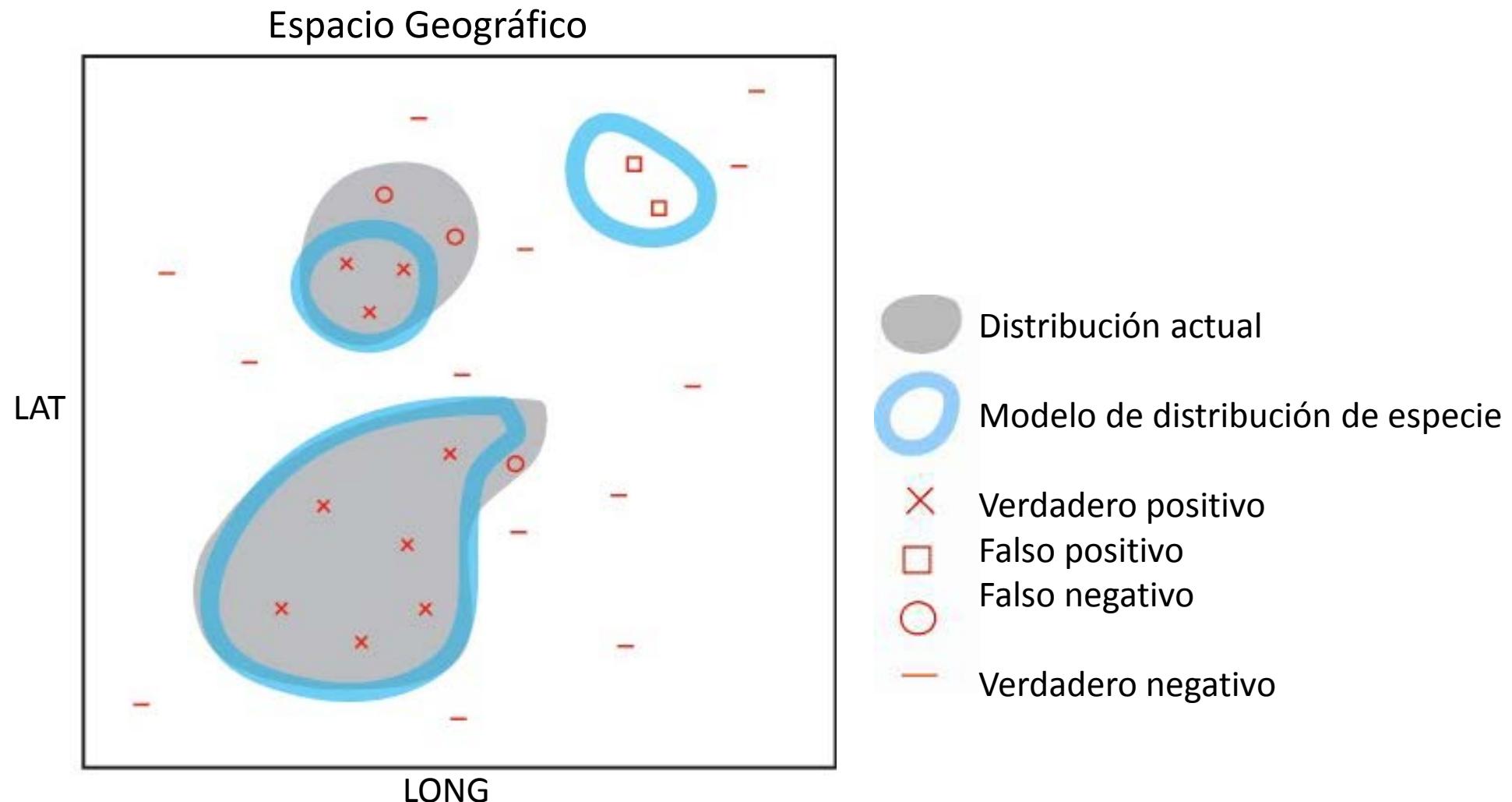
# Evaluación:

AUC	orMTP/10pct	AICc	Nparam
(Área bajo la curva)  Mide la capacidad de un modelo para diferenciar entre presencias verdaderas (sensibilidad; eje <b>y</b> ) vs. presencias falsas (especificidad; eje <b>x</b> ).  AUC ≈ 1 → buen modelo	(Tasa de omisión)  Mide el porcentaje de puntos de presencia que no son correctamente predichas como tal.  Si OR es grande → modelo esta “sobre-ajustando”	(Criterio de información Akaike para muestras pequeñas)  Mide la calidad relativa de un modelo para un conjunto de datos → ‘trade-off’ entre ajuste y complejidad.  Si AICc más bajo → mejor modelo	(Número de parámetros)  Mientras mas parámetros mas complejo será el modelo.  Nparam < N variables → mejor modelo

Por lo general queremos un modelo con:

- AUC cercano a 1
- OR menor a 20%
- AICc bajo
- Pocos parámetros

# Cuatro tipo de resultados posibles cuando evaluamos un modelo de distribución



# Componente 7: Visualización de resultados (Visualize)

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

Change Base Map  
ESRI Topo

Current Env Variable: bio01   Current model: H\_1

Visualize Model Results

Module Available:

Maxent Evaluation Plots

Plot Response Curves

Map Prediction

Module: Maxent Evaluation Plots

ENMeval: Automated Runs and Evaluations of Ecological Niche Models  
dismo: Species Distribution Modeling

Select evaluation statistic:  
average OR 10%

Download Maxent evaluation plot (.png)

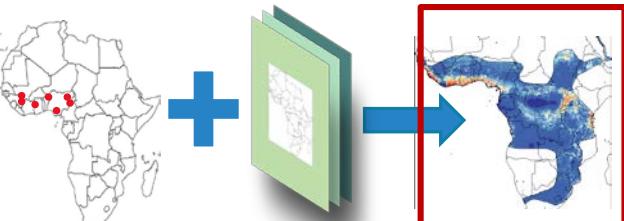
Module Developers: Jamie M. Kass, Robert Muscarella, Bruno Vilela, Robert P. Anderson  
ENMeval references  
Package Developers: Robert Muscarella, Peter J. Galante, Mariano Soley-Guardia, Robert A. Boria, Jamie M. Kass, Maria Uriarte, Robert P. Anderson  
[CRAN documentation](#) | [software note](#)  
dismo references  
Package Developers: Robert J. Hijmans, Steven Phillips, John Leathwick, Jane Elith

> Occurrences partitioned by block method.  
> Occurrences partitioned by block method.  
> BIOCLIM ran successfully and output evaluation results.  
> Maxent ran successfully and output evaluation results for 28 models.

Map   Ocs Tbl   Results   Component Guidance   Module Guidance

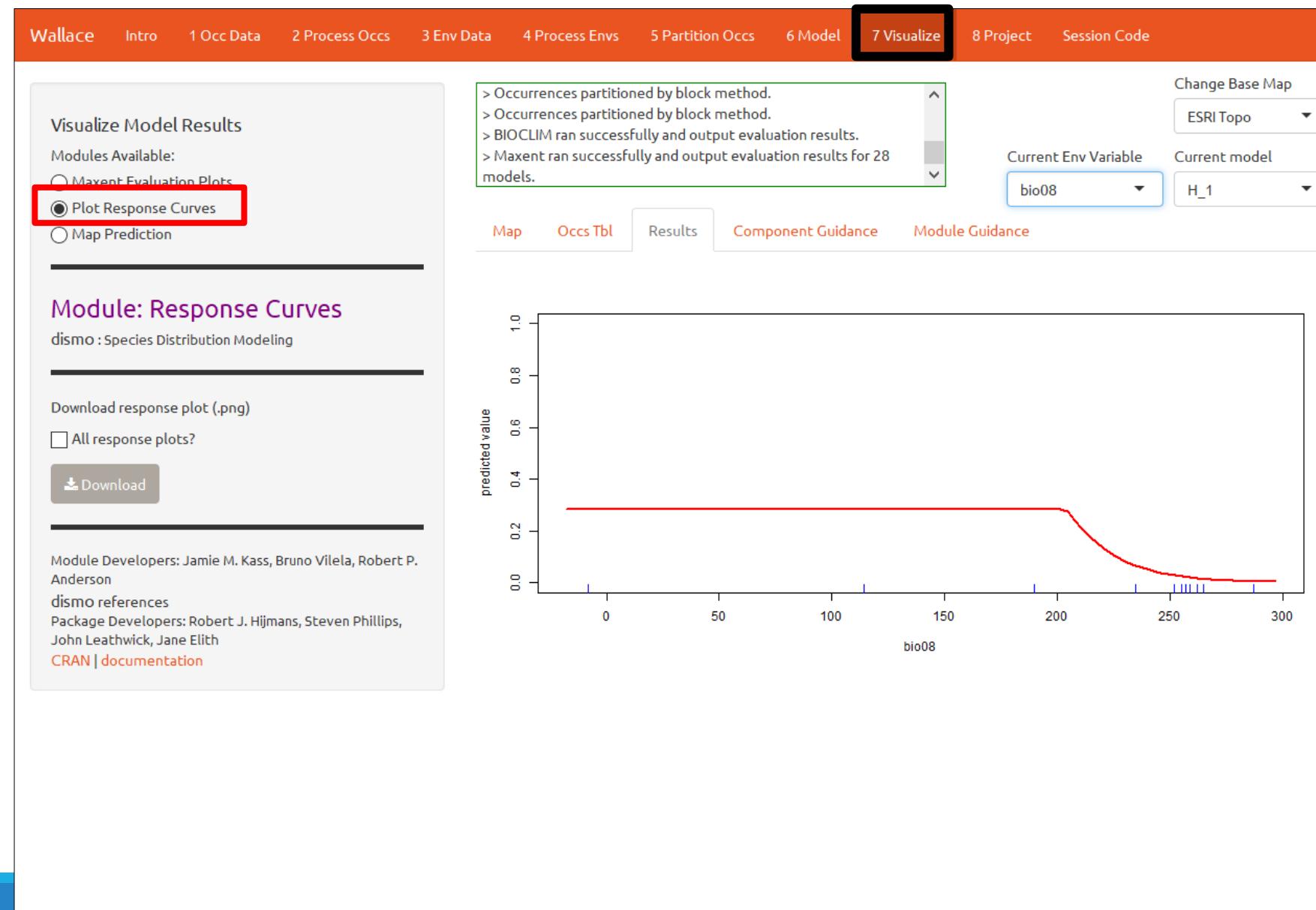
Regularization Multiplier	L	H	LQ	LQH
1.0	0.30	0.15	0.25	0.20
1.5	0.20	0.15	0.20	0.18
2.0	0.22	0.20	0.18	0.18
2.5	0.25	0.20	0.20	0.18
3.0	0.22	0.20	0.22	0.18
3.5	0.23	0.20	0.22	0.18
4.0	0.20	0.20	0.20	0.18

Permite visualizar el desempeño del modelo según nuestras distintas medidas de evaluación



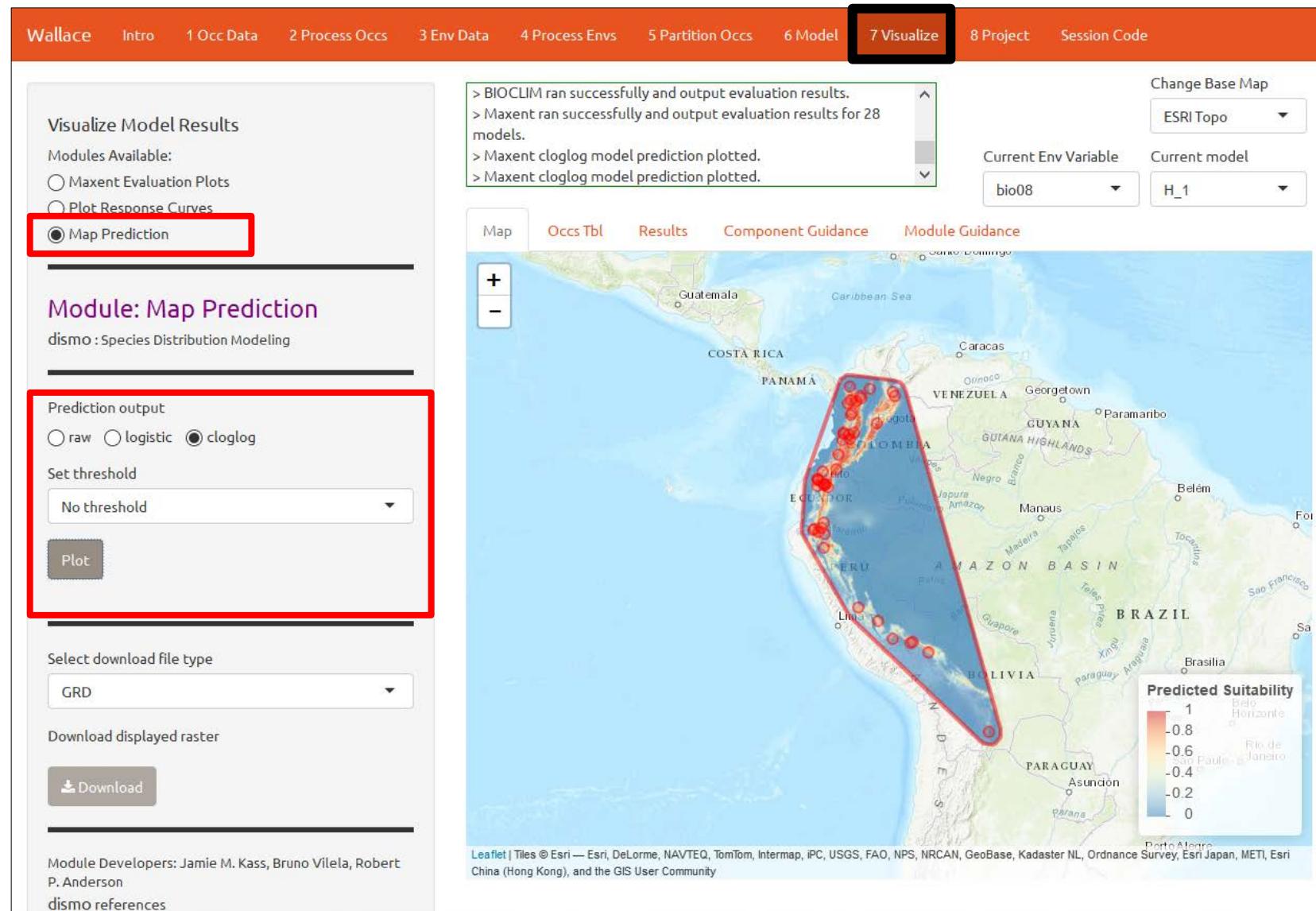
# Componente 7: Visualización de resultados (Visualize)

Curvas de respuesta para cada variable ambiental segun el modelo escogido



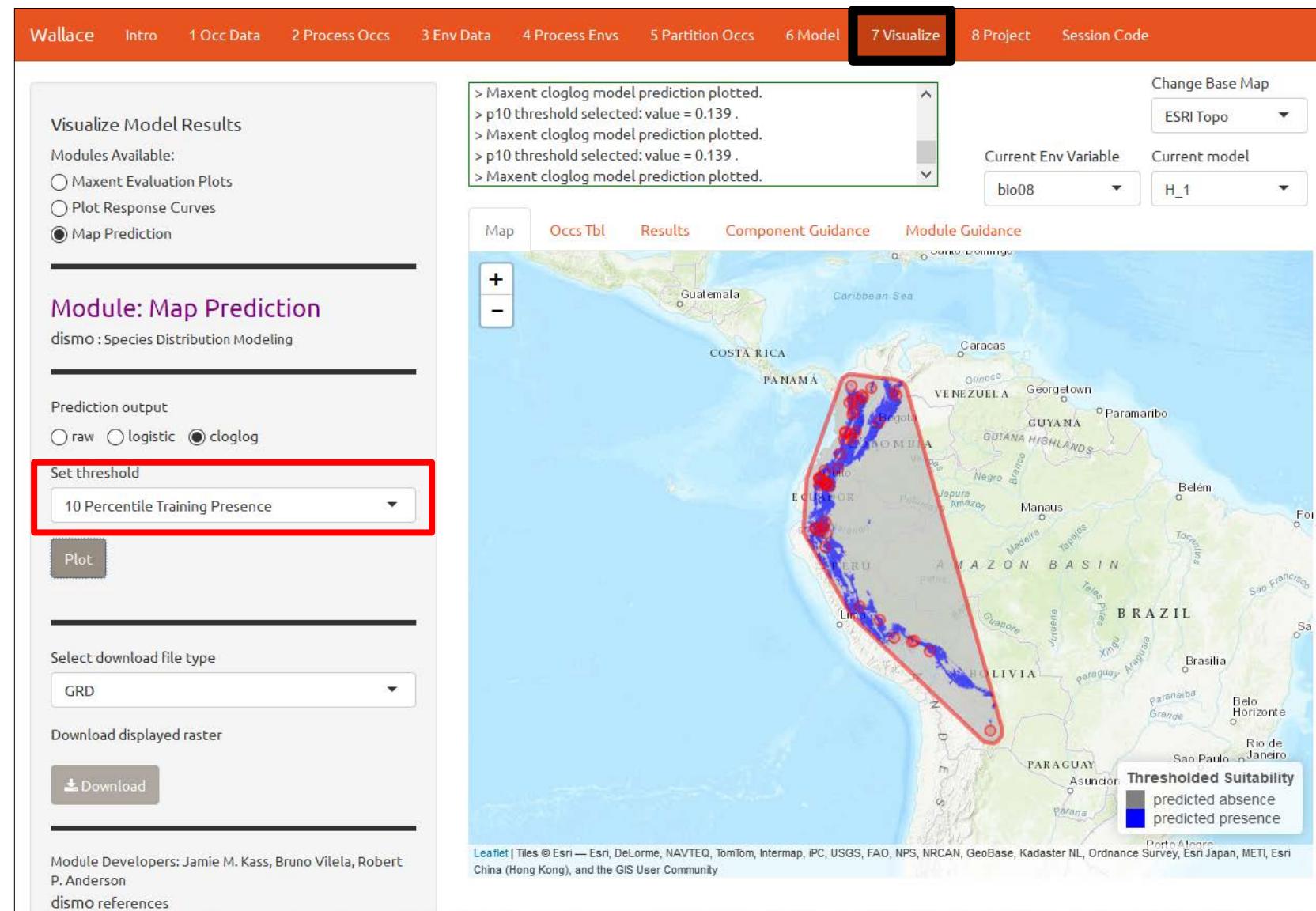
# Componente 7: Visualización de resultados (Visualize)

Mapas de nuestras predicciones de distribución geográfica (continuas)

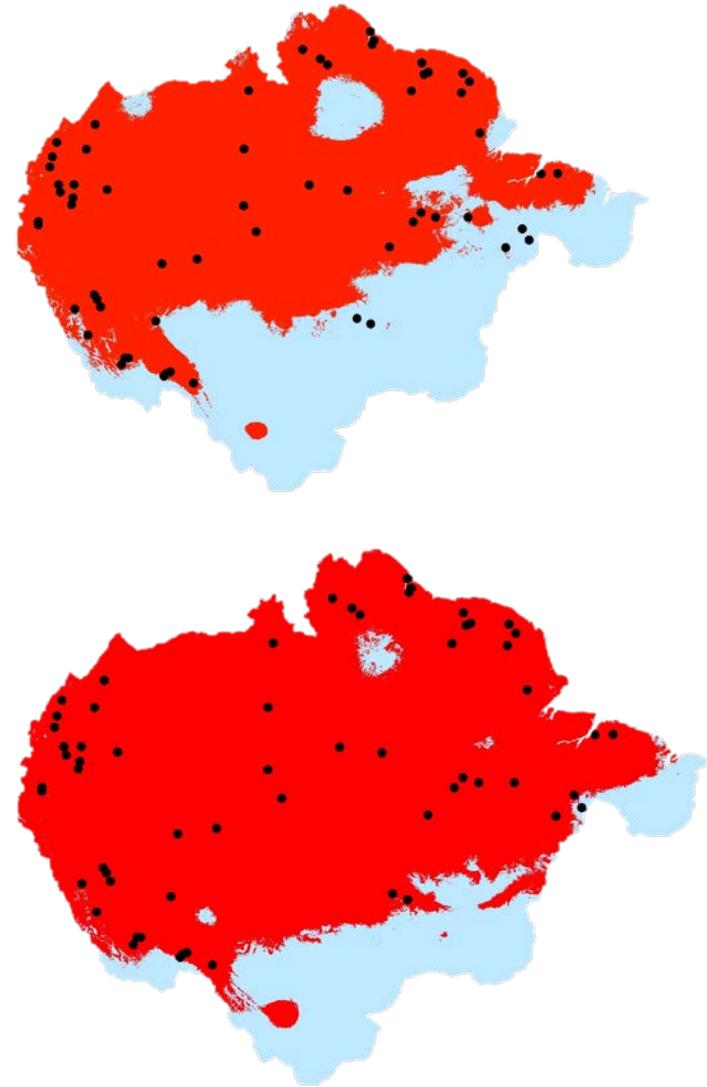
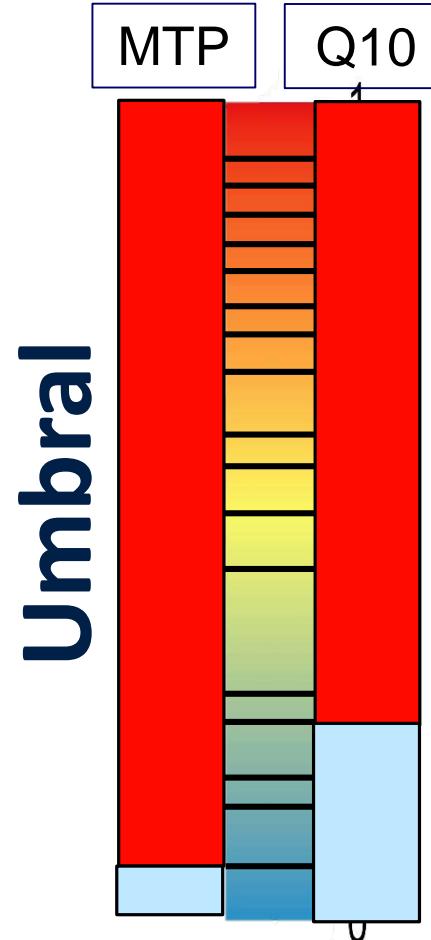
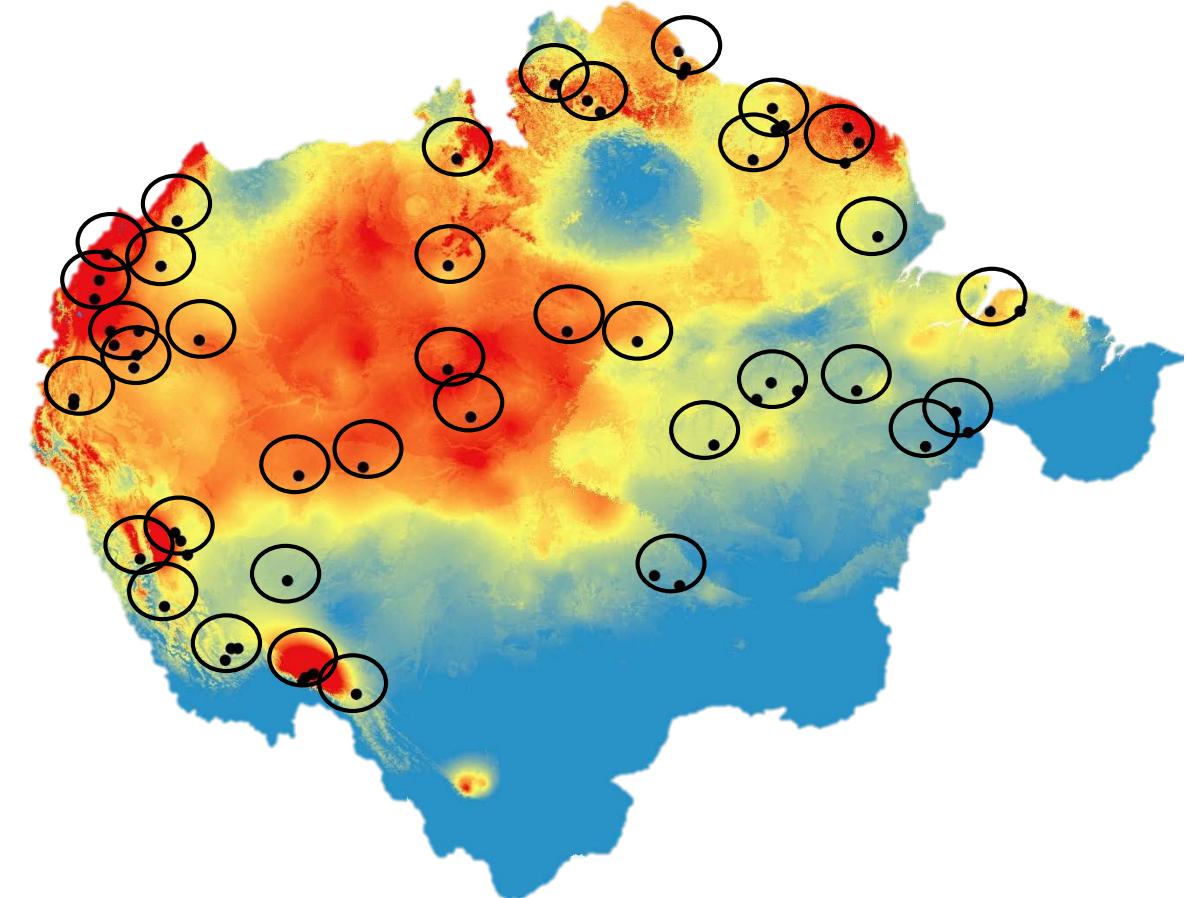


# Módulo 7: Visualización de resultados (Visualize)

Map binarios de presencia/ausencia segun las predicciones del modelo tras definir un umbral



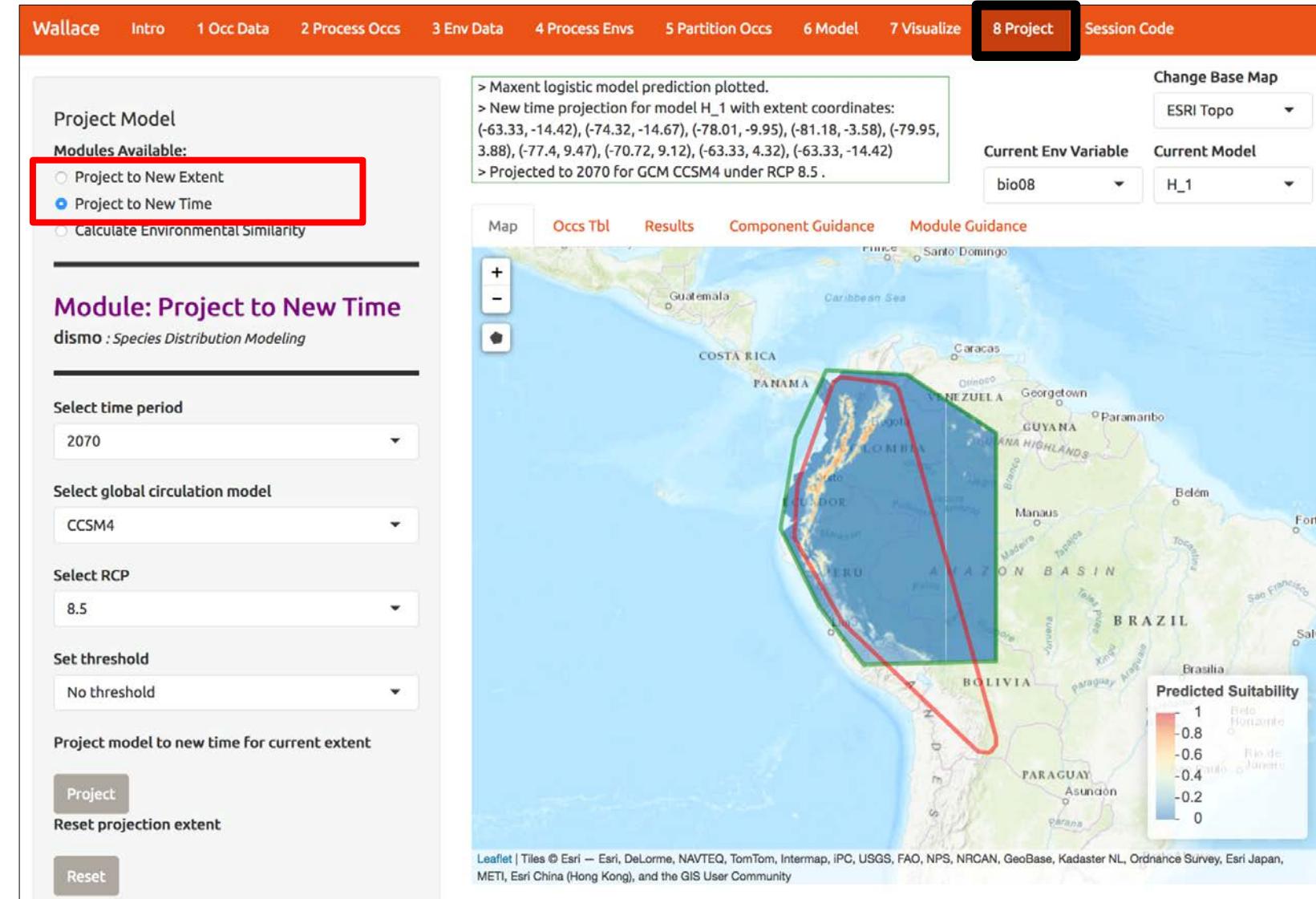
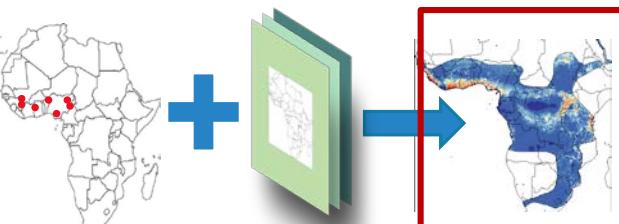
## Seleccionando umbrales



# Componente 8: Proyecciones (Proyect)

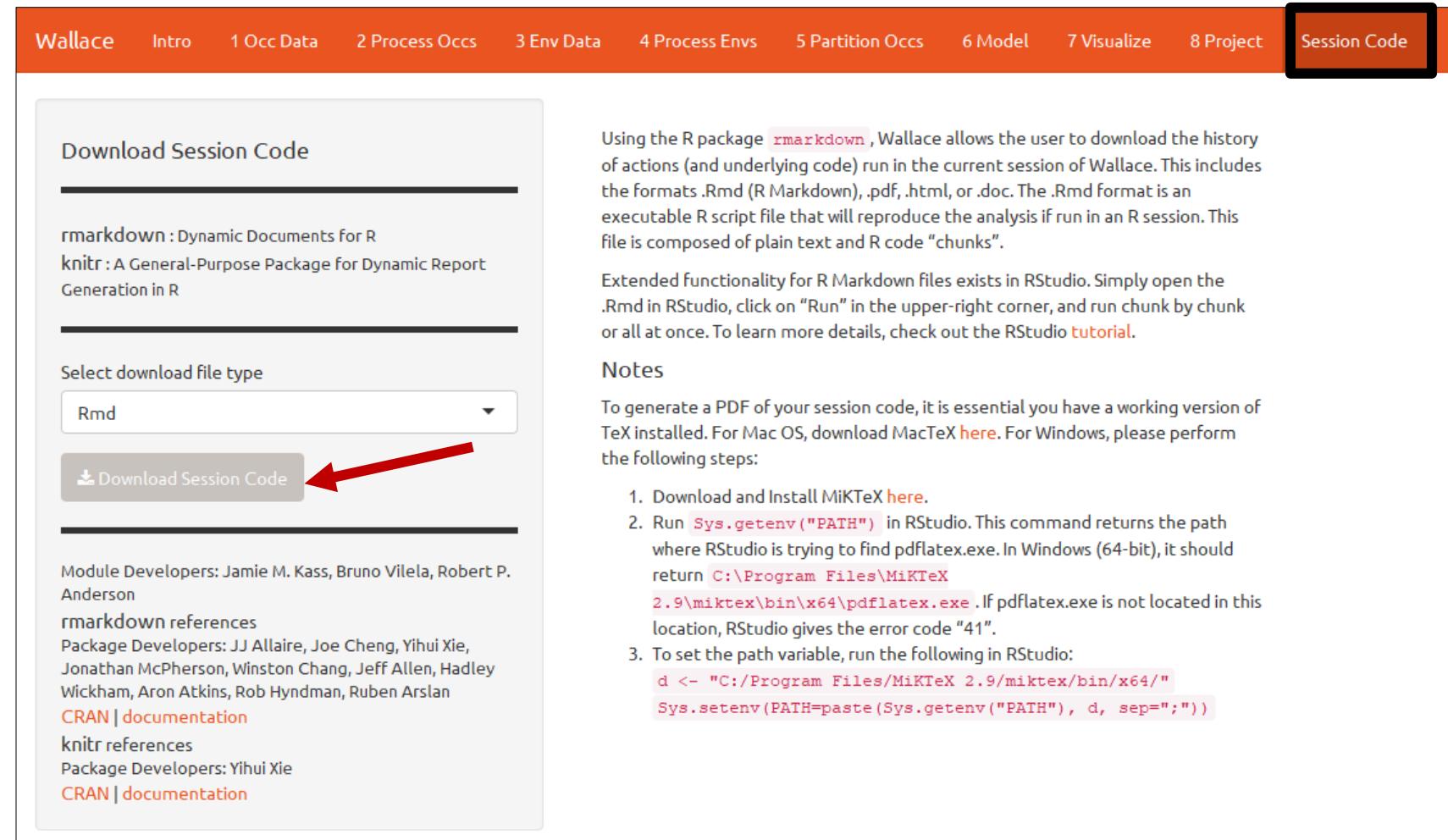
Permite la proyección de nuestro modelo a:

- a) Otras áreas geográficas
- b) Otros marcos temporales (futuro/pasado)



# Descargar el código R (Session Code)

Permite descargar el código de R de su sesión de modealdo facilitando la reproducibilidad del análisis.



The screenshot shows the Wallace software interface with the 'Session Code' tab selected. The main content area displays information about the rmarkdown and knitr packages, followed by a section for selecting the download file type ('Rmd') and a prominent 'Download Session Code' button, which is highlighted with a red arrow. To the right, there is explanatory text about the rmarkdown package and instructions for generating PDFs using RStudio, along with a 'Notes' section and steps for setting up MiKTeX.

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**

**Download Session Code**

---

rmarkdown : Dynamic Documents for R  
knitr : A General-Purpose Package for Dynamic Report Generation in R

---

Select download file type

Rmd

**Download Session Code** 

---

Module Developers: Jamie M. Kass, Bruno Vilela, Robert P. Anderson  
rmarkdown references  
Package Developers: JJ Allaire, Joe Cheng, Yihui Xie, Jonathan McPherson, Winston Chang, Jeff Allen, Hadley Wickham, Aron Atkins, Rob Hyndman, Ruben Arslan  
[CRAN | documentation](#)  
knitr references  
Package Developers: Yihui Xie  
[CRAN | documentation](#)

Using the R package `rmarkdown`, Wallace allows the user to download the history of actions (and underlying code) run in the current session of Wallace. This includes the formats .Rmd (R Markdown), .pdf, .html, or .doc. The .Rmd format is an executable R script file that will reproduce the analysis if run in an R session. This file is composed of plain text and R code "chunks".

Extended functionality for R Markdown files exists in RStudio. Simply open the .Rmd in RStudio, click on "Run" in the upper-right corner, and run chunk by chunk or all at once. To learn more details, check out the RStudio [tutorial](#).

**Notes**

To generate a PDF of your session code, it is essential you have a working version of TeX installed. For Mac OS, download MacTeX [here](#). For Windows, please perform the following steps:

1. Download and Install MiKTeX [here](#).
2. Run `Sys.getenv("PATH")` in RStudio. This command returns the path where RStudio is trying to find pdflatex.exe. In Windows (64-bit), it should return `C:\Program Files\MiKTeX 2.9\miktex\bin\x64\pdflatex.exe`. If pdflatex.exe is not located in this location, RStudio gives the error code "41".
3. To set the path variable, run the following in RStudio:  

```
d <- "C:/Program Files/MiKTeX 2.9/miktex/bin/x64/"
Sys.setenv(PATH=paste(Sys.getenv("PATH"), d, sep=";"))
```

Se puede importar y correr en R!!

The screenshot shows the RStudio interface with the following details:

- Title Bar:** RStudio
- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help
- Toolbar:** Go to file/function, Addins
- Code Editor:** di.RMD x  
The code editor displays R code from a session named "wallace".
  - Line 1: Please find below the R code history from your "wallace" v1.0.4 session.
  - Line 2: You can reproduce your session results by running this R Markdown file in Rstudio.
  - Line 3: Each code block is called a "chunk", and you can run them either one-by-one or all at once by choosing an option in the "Run" menu at the top-right corner of the "Source" pane in Rstudio.
  - Line 4: For more detailed information see <<http://rmarkdown.rstudio.com>>.
  - Line 5: **## Package installation**
  - Line 6: wallace uses the following R packages that must be installed and loaded before starting.
  - Line 7: 

```
library(spocc)
library(sptbin)
library(dismo)
library(rgeos)
library(ENMeval)
```
  - Line 8: **Record of analysis for \*Tremarctos ornatus\*.**
  - Line 9: **## obtain occurrence Data**
  - Line 10: The search for occurrences was limited to 100 records. Obtain occurrence records of the selected species from the gbif database.
  - Line 11: 

```
# query selected database for occurrence records
results <- spocc::occ(query = "Tremarctos ornatus", from = "gbif", limit = 100, has_coords = TRUE)
# retrieve data table from spocc object
results.data <- results[["gbif"]]\$data[[formatSpName("Tremarctos ornatus")]]
```
  - Line 12: # remove rows with duplicate coordinates
  - Line 13: occs.dups <- duplicated(results.data[c('longitude', 'latitude')])
  - Line 14: occs <- results.data[-occs.dups,]
  - Line 15: # make sure latitude and longitude are numeric (sometimes they are characters)
occs\$latitude <- as.numeric(occs\$latitude)
occs\$longitude <- as.numeric(occs\$longitude)
  - Line 16: # give all records a unique ID
occs\$occID <- row.names(occs)
  - Line 17: # extract occurrence coordinates
occs.xy <- occs[c('longitude', 'latitude')]
  - Line 18: **## Process Occurrence Data**
  - Line 19: The following code recreates the polygon used to select occurrences to keep in the analysis.
- Status Bar:** 1:1 (Top Level) ▾ R Markdown ▾

# Contenido



4. Aplicaciones potenciales de modelos de distribución de especies en biología de la conservación

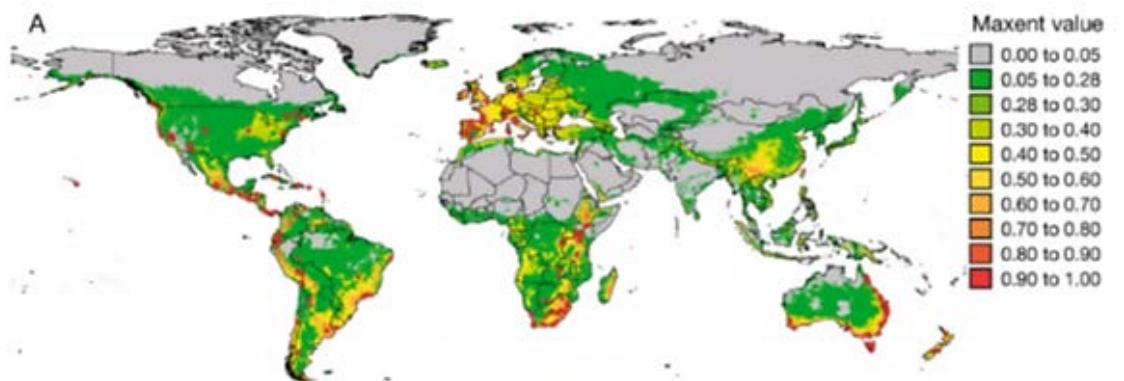
# Cambios en el rango de distribución tras cambio climático



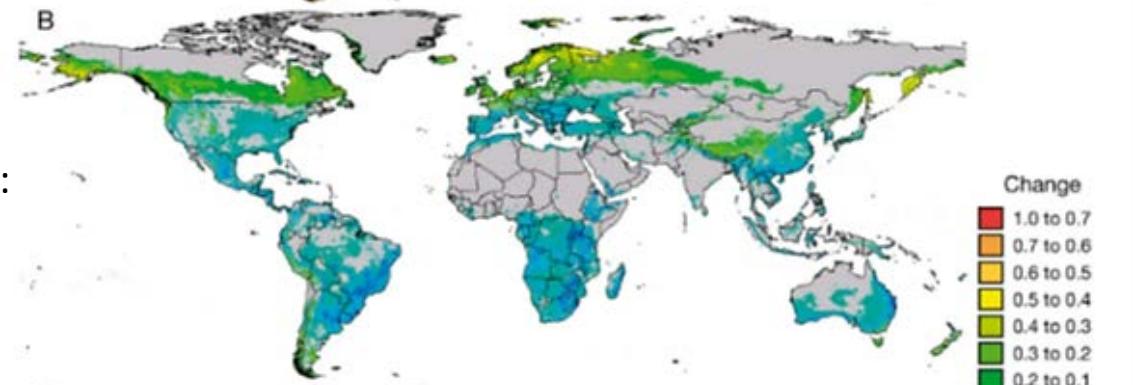
## Quitridiomicosis

- Enfermedad causada por el hongo *Batrachochytrium dendrobatidis* (Bd)
- Riesgo para la extinción de anfibios

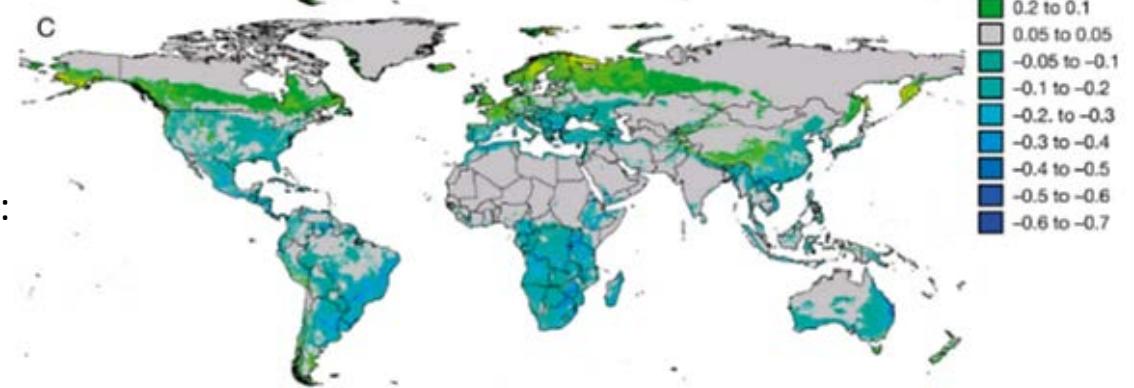
Distribucion potencial:  
Presente



Cambio proyectado:  
Año 2080  
Escenario IPCC A2a



Cambio proyectado:  
Año 2080  
Escenario IPCC B2a



(Rödder et al. 2010)

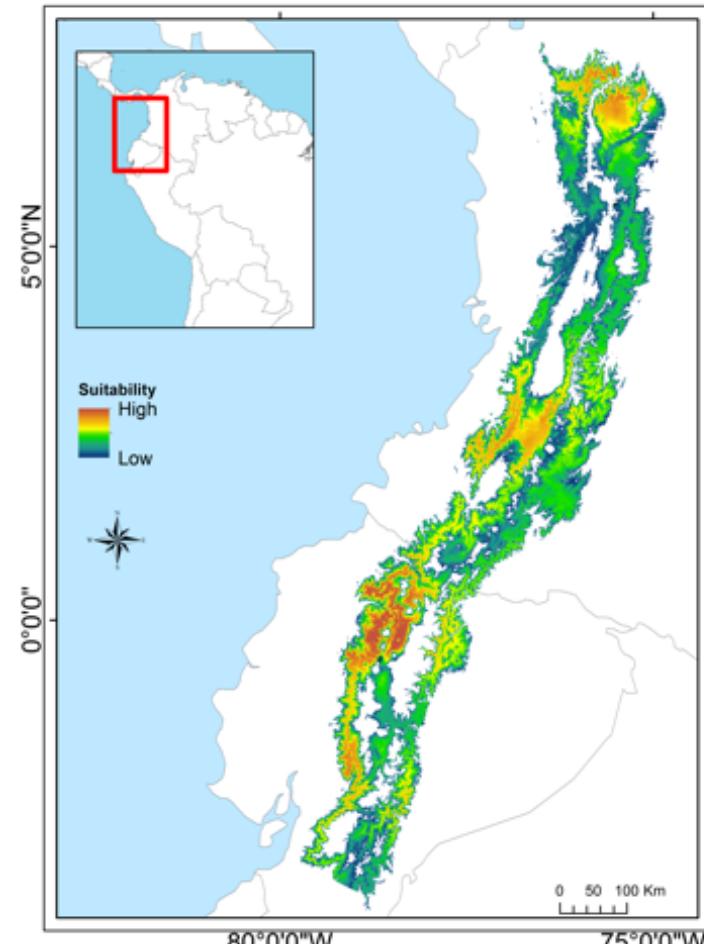
# Evaluación de especies en riesgo



## Olinguito:

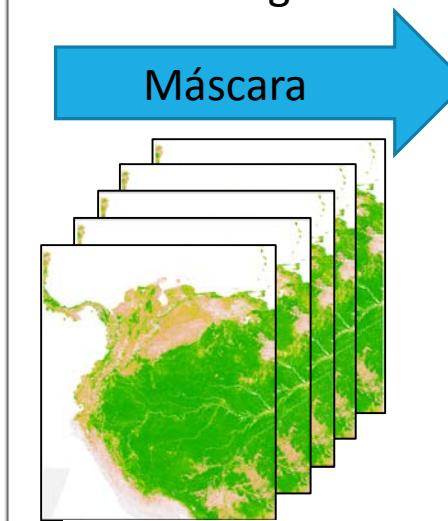
- Restringido a zonas boscosas
- Afectado por deforestación

Distribución potencial (solo clima)  
¿Refleja la realidad?



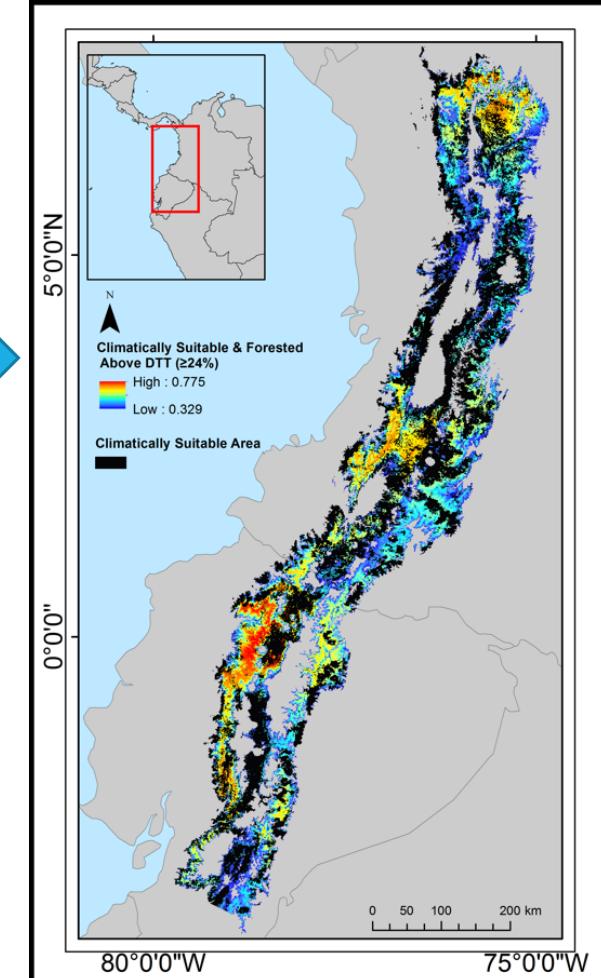
Nuevo paquete:  
maskRangeR

Máscara

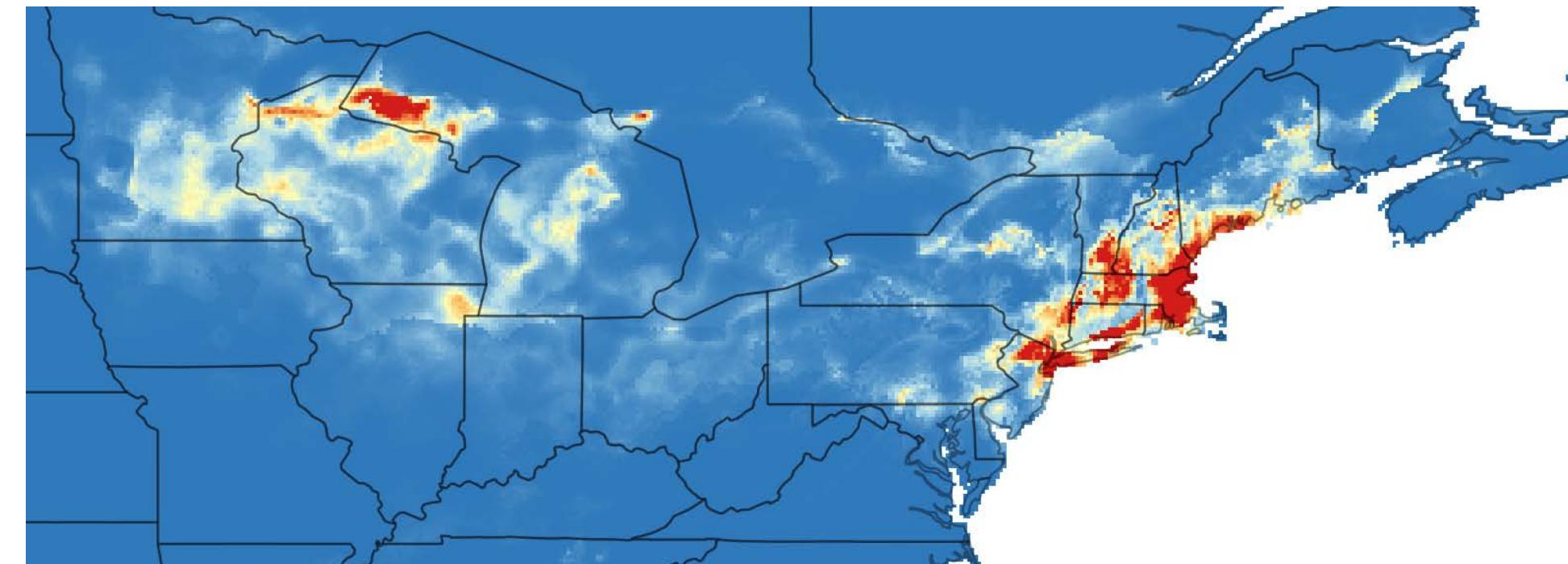


Cobertura de bosques anual

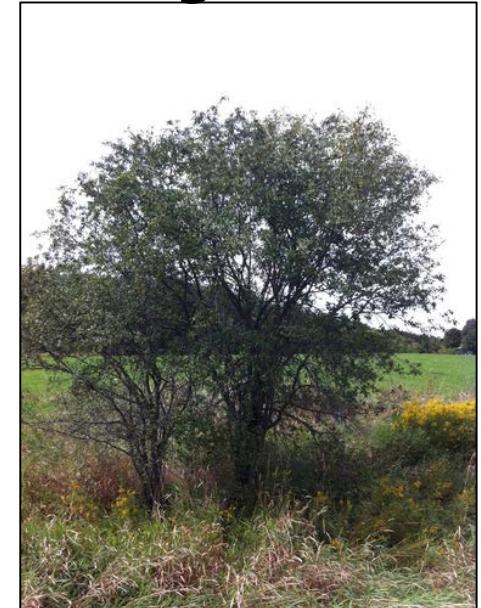
Distribución corregida  
(remueve zonas deforestadas)



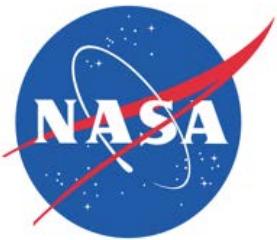
# Riesgo de propagación de especies invasoras



*Frangula alnus*



GRACIAS!



GEO GROUP ON  
EARTH OBSERVATIONS



GC Digital Initiatives

THE GRADUATE  
CENTER  
CITY UNIVERSITY  
OF NEW YORK



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

<http://www.andersonlab.ccny.cuny.edu>

**CUNY**  
The City  
University  
of  
New York

The **City College**  
of New York

**PACE**  
UNIVERSITY

AMERICAN  
MUSEUM  
OF  
NATURAL  
HISTORY



# Próximamente....

WALLACE v2.0.0

- Resolver errores de v1.6.0
- Agregar análisis multi-especies
- Integrar nuevos componentes/módulos

Nuevos componentes y módulos

