

## Creating and Optimizing Species Distribution Models: A Vignette for Wallace Ecological Modeling Application v2.0

**Author(s):** Bethany A. Johnson, Gonzalo E. Pinilla-Buitrago, Andrea Paz, Jamie M. Kass, Sarah I. Meenan, Robert P. Anderson

**Source:** Lessons in Conservation, Vol. 13, Issue 1, pp. 76-118

**Published by:** Network of Conservation Educators and Practitioners, Center for Biodiversity and Conservation, American Museum of Natural History

**Stable URL:** [ncep.amnh.org/linc](https://ncep.amnh.org/linc)

This article is featured in ***Lessons in Conservation***, the official journal of the Network of Conservation Educators and Practitioners (NCEP). NCEP is a collaborative project of the American Museum of Natural History's Center for Biodiversity and Conservation (CBC) and a number of institutions and individuals around the world. *Lessons in Conservation* is designed to introduce NCEP teaching and learning resources (or "modules") to a broad audience. NCEP modules are designed for undergraduate and professional level education. These modules—and many more on a variety of conservation topics—are available for free download at our website, [ncep.amnh.org](https://ncep.amnh.org).

**Note to educators:** access presentations, teaching notes, exercise solutions, and associated files for these modules by registering as an educator, and searching for module by title.



To learn more about NCEP, visit our website: [ncep.amnh.org](https://ncep.amnh.org).

All reproduction or distribution must provide full citation of the original work and provide a copyright notice as follows:

"Copyright 2023, by the authors of the material and the Center for Biodiversity and Conservation of the American Museum of Natural History. All rights reserved."

Illustrations obtained from the American Museum of Natural History's library: [images.library.amnh.org/digital/](https://images.library.amnh.org/digital/)

# Creating and Optimizing Species Distribution Models: A Vignette for Wallace Ecological Modeling Application v2.0

Bethany A. Johnson<sup>i</sup>, Gonzalo E. Pinilla-Buitrago<sup>i,ii</sup>, Andrea Paz<sup>iii</sup>, Jamie M. Kass<sup>iv</sup>, Sarah I. Meenan<sup>iv</sup>, Robert P. Anderson<sup>i,ii,iv</sup>

<sup>i</sup>City College of the City University of New York, Biology, New York, NY, USA; <sup>ii</sup>Graduate Center of the City University of New York, Biology, New York, NY, USA; <sup>iii</sup>Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zürich, Switzerland; <sup>iv</sup>Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi, Japan; <sup>v</sup>American Museum of Natural History, Division of Vertebrate Zoology, New York, NY, USA

## ABSTRACT

Species distribution modeling (SDM) enables ecologists and conservation scientists to make range estimates for species of concern, as well as predictions of potential range in unsampled areas and for different time periods. Although advances in model building and evaluation are widespread in the ecology and evolution literature, most cutting-edge modeling methods are inaccessible to those who cannot read and write computer code, resulting in a “barrier to use” for many who would like to apply SDM to their research or to conservation efforts. The Wallace ecological modeling application, implemented in the R programming language as the CRAN package “wallace,” provides a graphical user interface that allows any user to implement advanced SDM methods. Each Wallace module provides extensive guidance text and references key papers from the literature to help both new and experienced users learn best practices. In this exercise, we will introduce you to the Wallace software and its key features and learn how to create optimized SDMs for a species.

## PREFACE

Welcome to the Wallace EcoMod v2.0 vignette. This vignette was written specifically for Wallace v2.0, so if you are using a different version of Wallace, some text and images in this vignette may not match up exactly with the version you use. For other versions of this vignette and other vignettes, go to: <https://wallaceecomod.github.io/wallace/index.html> and select from the “Vignettes” dropdown menu.

## INTRODUCTION

Wallace is an R-based GUI (Graphical User Interface) application (R package `wallace`) for ecological modeling that currently focuses on building, evaluating, and visualizing models of species niches and distributions. We will refer to these models as species distribution models (SDMs), and we will not explain them at length here—as you read through, you will be pointed to some sources of detailed information within the application (such as the “Component Guidance” and “Module Guidance” tabs) for reference. Additionally, we highly recommend you read NCEP Synthesis “A Brief Introduction to Species Distribution Modeling for Conservation Educators and Practitioners” (available in the NCEP Module Collection at <https://ncep.amnh.org>) for additional background information.

Wallace has many qualities which we think make it a good example of next-generation scientific software: it’s 1) accessible, 2) open, 3) expandable, 4) flexible, 5) interactive, 6) instructive, and 7) reproducible.

The application, which is constructed with the R package `shiny`, features a pannable/zoomable map and dynamic plots and tables. Data for the models can be downloaded from online databases or uploaded by the user. Most results can be downloaded, including the option to save R code that can reproduce your analysis. For more details, including on SDMs, please see our initial publication<sup>1</sup> in *Methods in Ecology and Evolution* and our follow-up publication<sup>2</sup> in *Ecography*.

1. Kass, J.M., B. Vilela, M.E. Aiello-Lammens, R. Muscarella, C. Merow, and R.P. Anderson. 2018. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution* 9(4):1151–1156. <https://doi.org/10.1111/2041-210X.12945>.
2. Kass, J.M., et al. 2022. Wallace 2: A shiny app for modeling species niches and distributions redesigned to facilitate expansion via module contributions. *Ecography*, 2023(3):e06547. <https://doi.org/10.1111/ecog.06547>.

### Helpful Links

- Wallace project's main page: <https://wallaceecomod.github.io/>
- Wallace Google Group: <https://groups.google.com/forum/#!forum/wallaceecomod>
- Official email: [wallaceecomod@gmail.com](mailto:wallaceecomod@gmail.com)
- CRAN page (<https://cran.r-project.org/package=wallace>) hosts the stable version
- GitHub development page: <https://github.com/wallaceEcoMod/wallace>
- YouTube channel (<https://www.youtube.com/@WallaceEcoMod>) with various tutorial videos and recorded seminars

## 1. SETUP

### Installing the package

For the `wallace` package to work, you should be using R version 3.5.0 or later. Download for Windows (<https://cran.r-project.org/bin/windows/base/>) or Mac (<https://cran.r-project.org/bin/macosx/>).

To install and load Wallace v2.0.5, open either the R GUI or RStudio (<https://posit.co/download/rstudio-desktop/>) and run the code below. It's the only code you'll have to run to use `wallace`.

```
# Install Wallace from CRAN
install.packages("wallace")

# OR install Wallace from GitHub
install.packages("remotes")
remotes::install_github("wallaceEcoMod/wallace")

# THEN Load Wallace
library(wallace)
# AND Run Wallace
run_wallace()
```

The Wallace GUI will open in your default web browser, and the R console will be occupied while Wallace is running.

The R console displays messages regarding R-package information or any error messages if complications arise, including valuable information for troubleshooting. If you intend to ask a question in the Google Group (preferred) or by email, please include any error messages from the console.

If you'd like to use the R console while running Wallace, open another R session, or alternatively a terminal window (MacOS/Linux) or command prompt (Windows) and initialize R, then run the lines above.

An example screenshot of running in Terminal in MacOS appears below.

```

bajohnson135 — R — 66x45
Last login: Mon Apr 18 10:58:31 console
[Bethanys-MBP:~ bajohnson135$ R

R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]
> library(wallace)
Loading required package: shiny
Loading required package: leaflet
Warning message:
package 'leaflet' was built under R version 4.0.5
> run_wallace()

Listening on http://127.0.0.1:5414

```

To exit Wallace, hit 'Escape' while in the R console and close the browser window, or click the quit button in the top right corner of the GUI. Note: If you close the browser window running Wallace, *your session will be over and all progress will be lost*. See section 9. Save & Load Session for how to save your work and be able to restart your analysis.

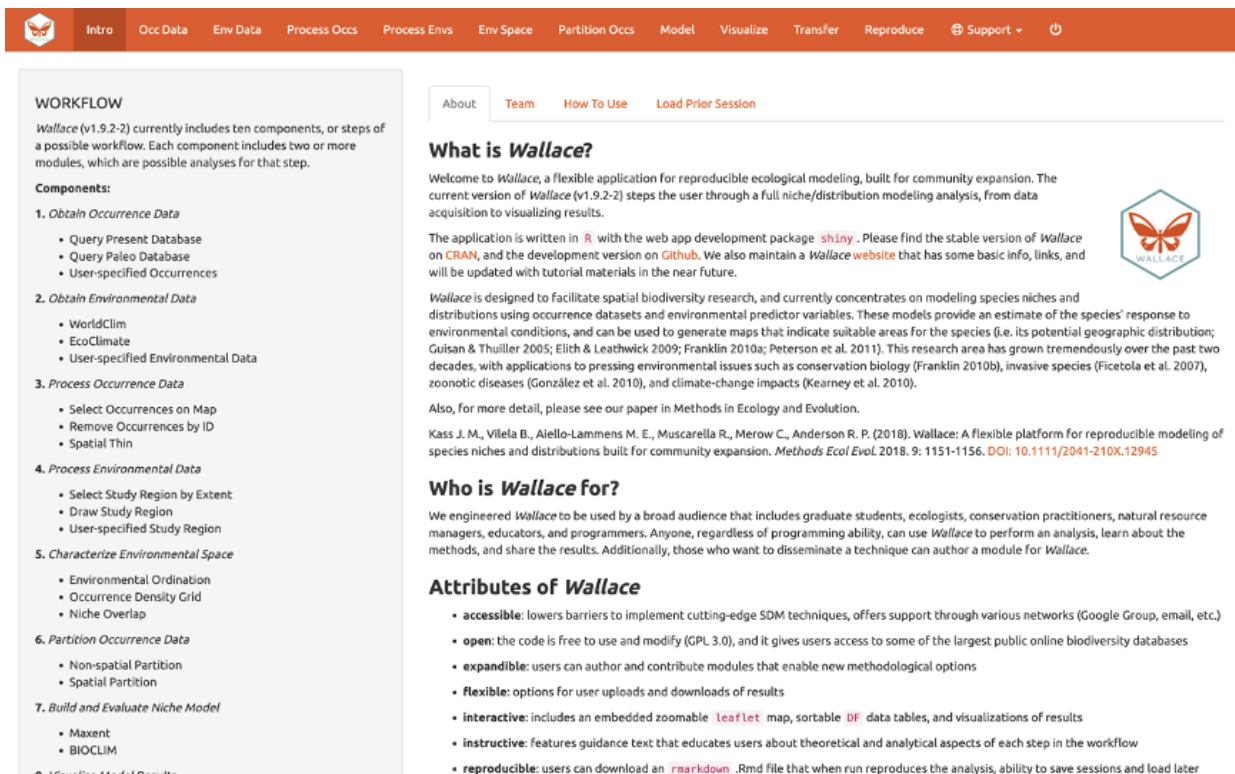
## Setting up Java version of Maxent

Wallace v2.0 includes two options to run Maxent models: maxnet and maxent.jar. The former, which is an R implementation of Maxent and fits the model leveraging the package **glmnet**, is now the default and does not require running Java (see Phillips et al. 2017). The latter, which is the original Java implementation, runs the **maxent()** function in the package **dismo**, which in turn relies on tools from the package **rJava**. When using **dismo** to run maxent.jar, the user must place the maxent.jar file in the /java directory of the **dismo** package root folder. You can download Maxent ([https://biodiversityinformatics.amnh.org/open\\_source/maxent/](https://biodiversityinformatics.amnh.org/open_source/maxent/)) and find maxent.jar, which runs Maxent, in the downloaded folder. You can find the directory path to dismo/java by running **system.file('java', package="dismo")** in the R console. Simply copy maxent.jar and paste it into the /java directory of the **dismo** package root folder. If you try to run Maxent in Wallace without the file in place, you will get a warning message in the log window and Maxent will not run. Also, if you have trouble installing **rJava** and making it work, there is a bit of troubleshooting on the Wallace Github repository README (<https://github.com/wallaceEcoMod/wallace/blob/master/README.md>) that hopefully should help.

## 2. ORIENTATION

We'll begin with an orientation of the Wallace interface. After running **run\_wallace()** in the R console, Wallace opens to the "Intro" page. The "About" tab contains background information about the program. The "Team" tab has details about the developers and collaborators who contributed to Wallace. The "How to Use" tab contains a brief user manual, which is an abridged version of this





**WORKFLOW**

Wallace (v1.9.2-2) currently includes ten components, or steps of a possible workflow. Each component includes two or more modules, which are possible analyses for that step.

**Components:**

- 1. Obtain Occurrence Data**
  - Query Present Database
  - Query Paleo Database
  - User-specified Occurrences
- 2. Obtain Environmental Data**
  - WorldClim
  - EcoClimate
  - User-specified Environmental Data
- 3. Process Occurrence Data**
  - Select Occurrences on Map
  - Remove Occurrences by ID
  - Spatial Thin
- 4. Process Environmental Data**
  - Select Study Region by Extent
  - Draw Study Region
  - User-specified Study Region
- 5. Characterize Environmental Space**
  - Environmental Ordination
  - Occurrence Density Grid
  - Niche Overlap
- 6. Partition Occurrence Data**
  - Non-spatial Partition
  - Spatial Partition
- 7. Build and Evaluate Niche Model**
  - Maxent
  - BIOCLIM
- 8. Visualize Model Results**

**What is Wallace?**

Welcome to Wallace, a flexible application for reproducible ecological modeling, built for community expansion. The current version of Wallace (v1.9.2-2) 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. DOI: 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**

- **accessible:** lowers barriers to implement cutting-edge SDM techniques, offers support through various networks (Google Group, email, etc.)
- **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
- **interactive:** includes an embedded zoomable leaflet map, sortable DF 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 rmarkdown .Rmd file that when run reproduces the analysis, ability to save sessions and load later

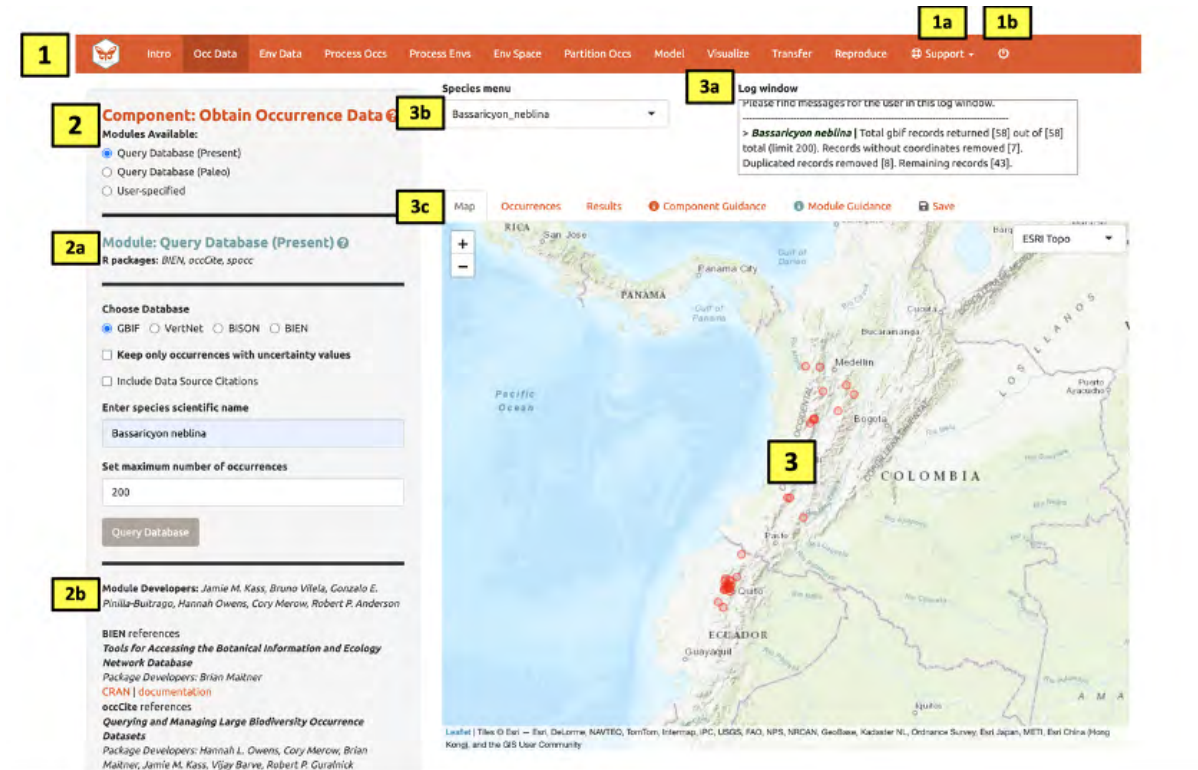
vignette without a worked example. The “Load Prior Session” tab is for loading a prior session, which we will cover later in section 9. Save & Load Session of this exercise.

At the top in the orange panel are the “Components”, which represent the steps of analysis. Each of these component tabs opens to the corresponding step. Within each component are several “Modules”, which are discrete analysis options within the components. To the left in the gray panel is the Wallace WORKFLOW, outlining the version number, components (numbered), and modules (bulleted) currently included. Click on the component tab “Occ Data”, select a module, and consult the schematic below showing the different parts of the Wallace interface.

(1) These are the components. You will be stepping sequentially through them. Wallace v2 includes a Support button (1a), which links to the Google Group, email, website, and the Github page to report issues, as well as the quit button (1b), which will end the session. (2) This is the toolbar with all the user interface controls, such as buttons, text inputs, etc. You can see that the module “Query Database (Present)” is currently selected. You’ll see that two other modules exist for this component: Query Database (Paleo) and User-specified. This last module lets you upload your own occurrence data. Try choosing it instead and notice that the toolbar changes, then click back to “Query Database (Present)”.

Both the “Component” and “Module” have question mark buttons (?) next to the title text. Clicking these will link to the respective guidance texts.

Within this toolbar, you can find the module name and the R packages it uses (2a), as well as the control panel for the selected module (2b). Modules can be contributed by other researchers and the developers; CRAN links and documentation are at the bottom. (3) The right side is the visualization space. Any functions performed will trigger a message in the log window (3a). This window will also display any error messages. Wallace v2.0.5 allows the user to load multiple species. If multiple species are loaded, toggle and select between species using the species drop-down menu (3b).



The screenshot shows the Wallace software interface. At the top, a navigation bar contains tabs: Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce, and Support. A species menu on the left lists 'Bassaricyon neblina'. Below it, the 'Component: Obtain Occurrence Data' section shows 'Query Database (Present)' as the selected module. The 'Module: Query Database (Present)' section includes options for 'Choose Database' (GBIF, VertNet, BISON, BIEN), checkboxes for 'Keep only occurrences with uncertainty values' and 'Include Data Source Citations', a text input for 'Enter species scientific name' (Bassaricyon neblina), and a 'Set maximum number of occurrences' field (200). A 'Query Database' button is at the bottom. A log window on the right shows a message: '> Bassaricyon neblina | Total gbif records returned [58] out of [58] total (limit 200). Records without coordinates removed [7]. Duplicated records removed [8]. Remaining records [43]'. A map on the right shows the distribution of occurrence records in western Colombia and Ecuador. Numbered callouts 1 through 3a and 3c highlight various interface elements.

The visualization space includes several tabs (3c), including an interactive map, occurrence records table, results window, model and component guidance text windows, and a tab for saving outputs and the current session.

At this stage of the analysis, no results exist, and you have no data yet for the table, but you can view the “Component Guidance” and “Module Guidance” text now. This text was written by the developers to prepare users for each component and module theoretically (why we should use the tools) and methodologically (what the tools do). The guidance text also references scientific papers from the literature for more detailed reading. Please get into the habit of consulting these before undertaking analyses—and discussing them with your peers—as this should give you a more solid foundation for moving forward.

The next tab in the visualization space is “Save”. At any point along the workflow, selecting “Save session” within this tab will save the progress as a .rds file. This file can be loaded back into **wallace** to resume analysis. If at any point during the vignette you need to pause, jump to section 9. Save & Load Session in this exercise to learn how to save and load your Wallace session. This tab is also where you will be able to download and save results. The session code, metadata, and package citations can be downloaded within “Component: Reproduce” (section 13. Reproduce in this exercise).

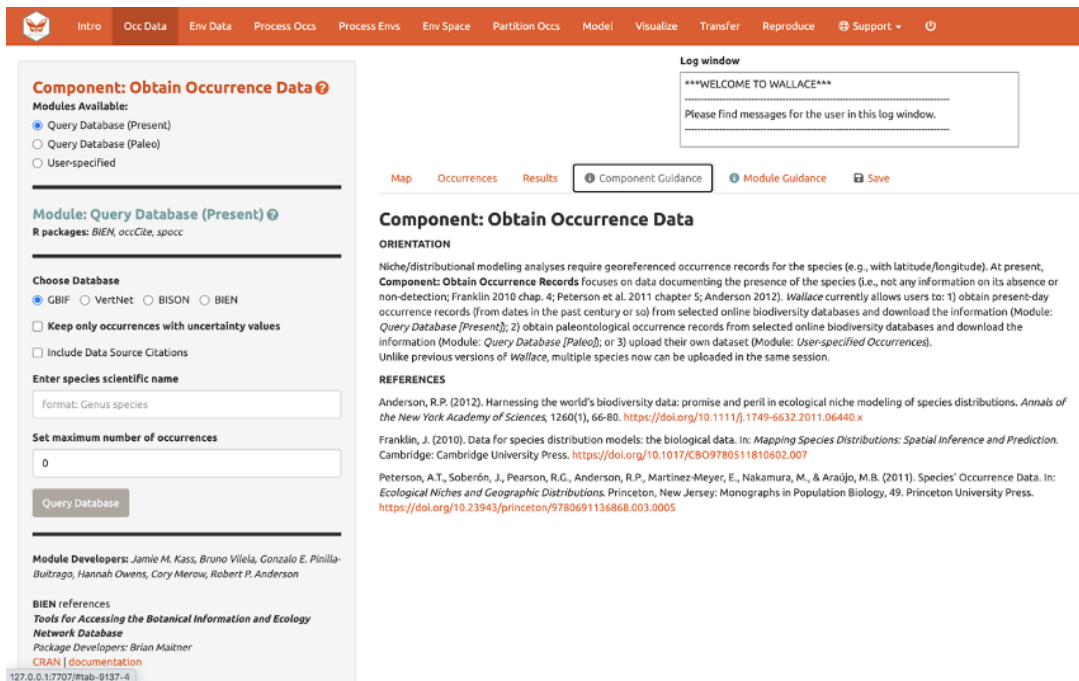
Now let’s begin our analysis.

We’ll be modeling the ranges of two mammal species of the genus *Bassaricyon*, which are members of the family Procyonidae that includes raccoons. *Bassaricyon neblina*, or the olinguito, is found in tropical montane areas of western Colombia and Ecuador in South America. The olinguito gained species status in 2013 when it was identified from existing museum specimens and is currently a species of concern listed as “Near Threatened” by the IUCN (Helgen et al. 2020). *Bassaricyon alleni*, or the eastern lowland

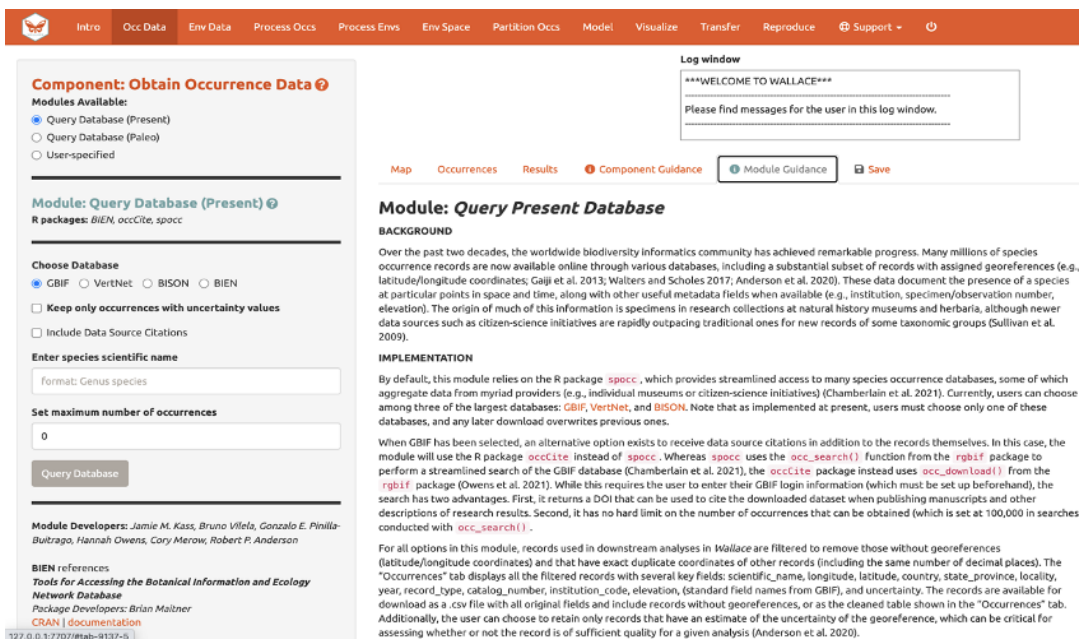
olingo, is a relative of the olinguito and has a broader range throughout northern South America; it is currently listed as “Least Concern” by the IUCN (Helgen et al. 2016).

### 3. OBTAIN OCCURRENCE DATA

Make sure you are in the first component (“Obtain Occurrence Data”) and click to read the component guidance text. There are three modules available for obtaining occurrence data: “Query Database (Present)”, “Query Database (Paleo)”, and “User-Specified”. Choose a module and click on the module guidance text. Notice the module guidance text changes as you select among the three modules. Read through these to get a better understanding of how occurrence data is typically obtained and how Wallace implements it.



The screenshot shows the Wallace software interface. The top navigation bar includes: Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce, Support, and a power icon. The main content area is titled "Component: Obtain Occurrence Data". Under "Modules Available:", "Query Database (Present)" is selected. Below this, the "Module: Query Database (Present)" section lists R packages: BIEN, occCite, spocc. The "Choose Database" section has "GBIF" selected. There are checkboxes for "Keep only occurrences with uncertainty values" and "Include Data Source Citations". A text input field for "Enter species scientific name" is present, along with a "Set maximum number of occurrences" field set to 0. A "Query Database" button is visible. On the right, a "Log window" displays "\*\*\*WELCOME TO WALLACE\*\*\*" and "Please find messages for the user in this log window." Below the log window, tabs for "Map", "Occurrences", "Results", "Component Guidance", "Module Guidance", and "Save" are shown. The "Component: Obtain Occurrence Data" section includes an "ORIENTATION" paragraph and a "REFERENCES" list with citations.

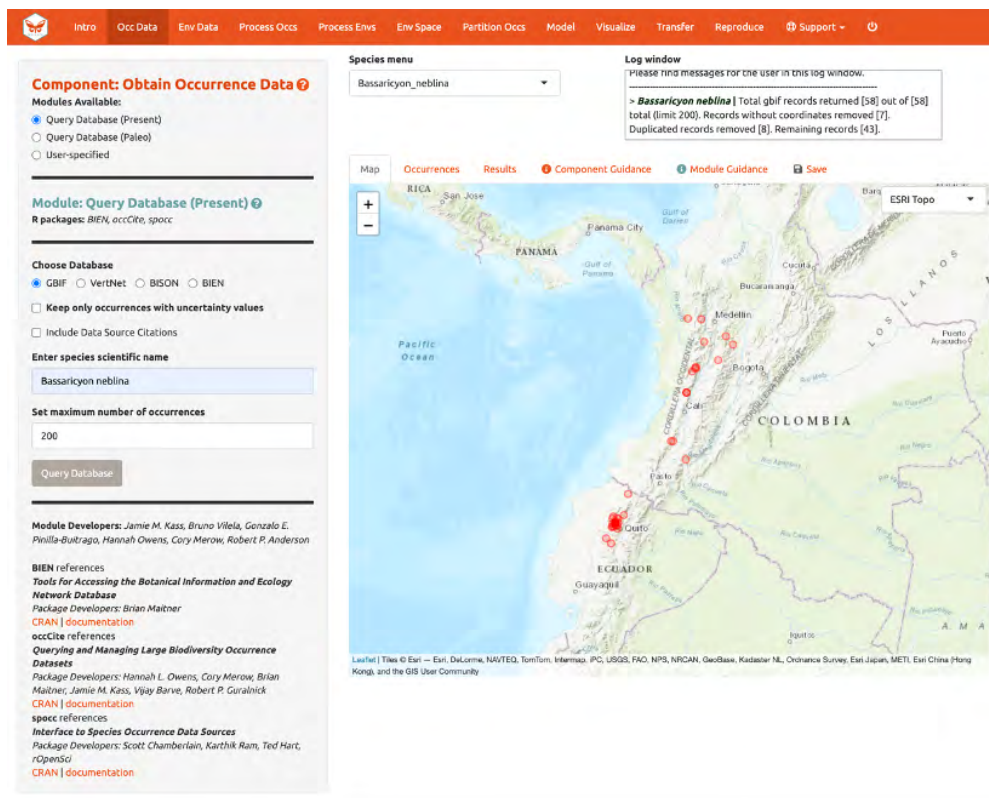


This screenshot is similar to the one above but shows the "Module: Query Present Database" section selected. The "ORIENTATION" text is replaced by "BACKGROUND" text, which discusses the progress of biodiversity informatics and the use of citizen-science initiatives. The "IMPLEMENTATION" section explains that the module relies on the R package spocc and provides streamlined access to occurrence databases. It also mentions that GBIF has been selected as an alternative option for receiving data source citations. The "REFERENCES" list is also present.

Let's proceed to get some occurrence data. We'll be using present occurrences (as opposed to those from the deep past via fossil data, etc.) and therefore use Module: "Query Database (Present)". There is a selection of databases to choose from, as well as the option to return only those occurrences that contain information on coordinate uncertainty (which can be useful to filter by later). Optional: If you have a GBIF User ID, checking the "Include Data Source" box will allow you to log in with your username and password to download a DOI for the dataset. In order for this to work, you will need to install the R-package `occCite` from RStudio or R prior to running Wallace. Since `occCite` is a suggested package, it will not install automatically like the other package dependencies. If this is the case, you will see a notification to install `occCite` in the log window. If you don't already have a GBIF account, you can create a user ID login at <https://www.gbif.org>.

Choose GBIF (the Global Biodiversity Information Facility—one of the largest storehouses for biodiversity data), keep uncertainty unchecked, type in *Bassaricyon neblina* into the scientific name box, set the maximum number of occurrences to 200, and click "Query Database".

After the download is complete, the log window will contain information on the analysis performed. Your search should return at least 58 records (numbers recorded at the time of writing), but after accounting for records without coordinate information (latitude, longitude) and removing duplicate records, at least 43 should remain. This species has relatively few records, so setting the maximum to 200 is sufficient, but for modeling with data-rich species, 200 may not be enough for adequately sampling the known range, and the maximum can be increased. **\*\*Numbers may be different as more records are added to GBIF.**



The screenshot shows the Wallace software interface. On the left, the "Component: Obtain Occurrence Data" panel is active. It displays the following settings:

- Species menu:** *Bassaricyon neblina*
- Log window:** Shows a message: "> *Bassaricyon neblina* | Total gbif records returned [58] out of [58] total (lim: 200). Records without coordinates removed [7]. Duplicated records removed [8]. Remaining records [43]."
- Module: Query Database (Present)**
- R packages:** BIEN, occCite, spocc
- Choose Database:** GBIF (selected), VertNet, BISON, BIEN
- Options:** "Keep only occurrences with uncertainty values" and "Include Data Source Citations" are unchecked.
- Enter species scientific name:** *Bassaricyon neblina*
- Set maximum number of occurrences:** 200
- Buttons:** "Query Database"

On the right, a map shows the distribution of occurrences for *Bassaricyon neblina* in Central and South America, with red dots indicating recorded locations. The map includes labels for countries like PANAMA, COLOMBIA, and ECUADOR, and cities like San Jose, Panama City, Medellin, Bogotá, and Quito. The "Occurrences" tab is selected in the top navigation bar.

Now click on the "Occurrences" tab to view more information on the records. The developers chose the fields that are displayed based on their general relevance to studies on species ranges. Note that you can download the full table with all fields.



Intro
Occ Data
Env Data
Process Occs
Process Envs
Env Space
Partition Occs
Model
Visualize
Transfer
Reproduce
Support

### Component: Obtain Occurrence Data ?

**Modules Available:**

- Query Database (Present)
- Query Database (Paleo)
- User-specified

---

**Module: Query Database (Present) ?**

R packages: BIEN, occCite, spocc

**Choose Database**

GBIF  VertNet  BISON  BIEN

Keep only occurrences with uncertainty values

Include Data Source Citations

**Enter species scientific name**

**Set maximum number of occurrences**

---

**Module Developers:** *Jamie M. Kass, Bruno Villela, Gonzalo E. Pinilla-Buitrago, Hannah Owens, Cory Merow, Robert P. Anderson*

**BIEN references**  
*Tools for Accessing the Botanical Information and Ecology Network Database*  
 Package Developers: *Brian Maitner*  
[CRAN | documentation](#)

**occCite references**  
*Querying and Managing Large Biodiversity Occurrence Datasets*  
 Package Developers: *Hannah L. Owens, Cory Merow, Brian Maitner, Jamie M. Kass, Vijay Barve, Robert P. Gurainick*  
[CRAN | documentation](#)

**spocc references**  
*Interface to Species Occurrence Data Sources*  
 Package Developers: *Scott Chamberlain, Karthik Ram, Ted Hart, rOpenSci*  
[CRAN | documentation](#)

**Species menu**

Bassaricyon\_neblina

**Log window**

Please find messages for the user in this log window.

```
> Bassaricyon neblina | Total gbif records returned [58] out of [58] total (limit: 200). Records without coordinates removed [7]. Duplicated records removed [8]. Remaining records [43].
```

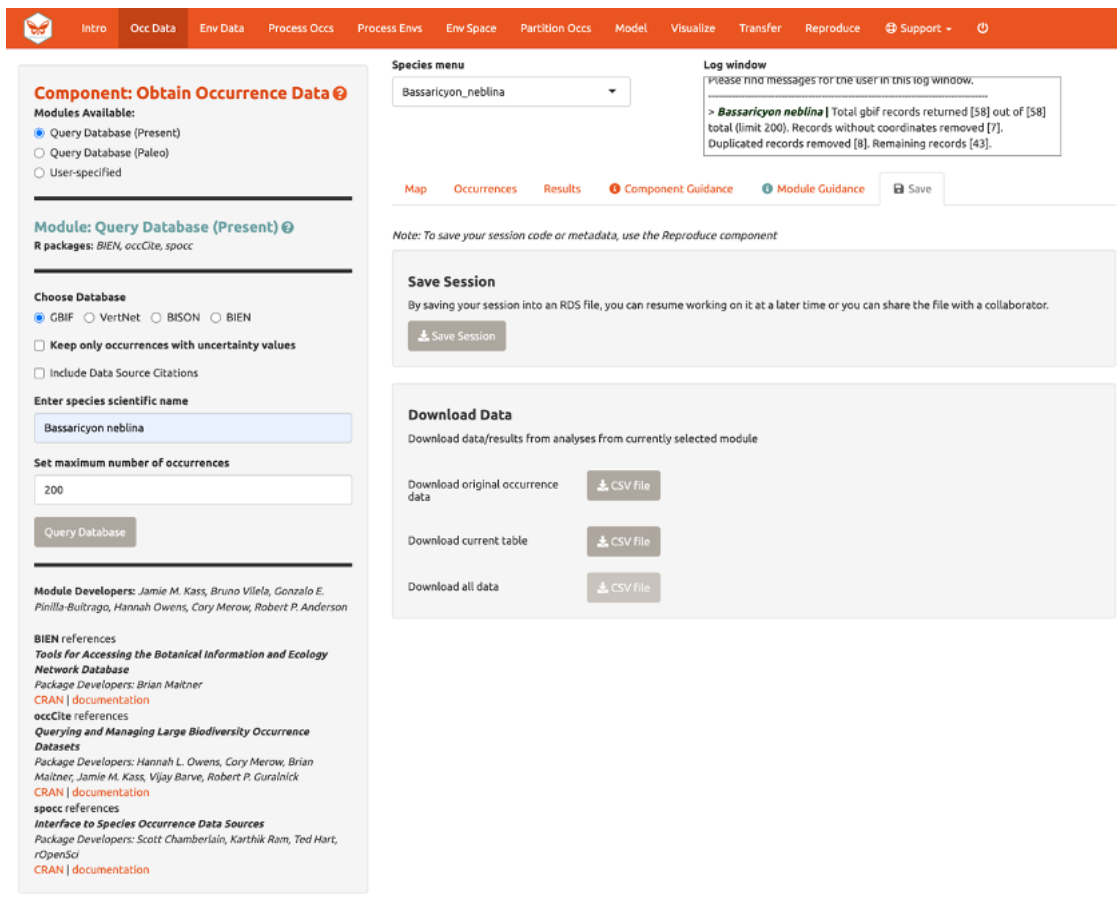
Show  entries Search:

occID	scientific_name	longitude	latitude	country	state_province	locality	year	record_type
1	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-78.61	0.11	Ecuador	Pichincha		2021	HUMAN_OBSERV
2	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-75.48	5.05	Colombia	Caldas	Puente Vizcaya, zona urbana	2021	PRESERVED_SPE
3	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-78.73	-0.14	Ecuador	Pichincha		2020	HUMAN_OBSERV
4	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-76.01	6.32	Colombia	Antioquia		2020	HUMAN_OBSERV
5	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-78.77	-0.04	Ecuador	Pichinche		2019	HUMAN_OBSERV
6	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-78.68	-0.02	Ecuador		Ecuador - Pichincha - Bellavista Cloud Forest Reserve	2019	HUMAN_OBSERV
7	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-78.78	0.11	Ecuador	Pichincha		2019	HUMAN_OBSERV
8	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-76.46	4.05	Colombia	Valle del Cauca	PNR Páramo del Duende	2018	HUMAN_OBSERV
9	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-76.46	4.05	Colombia	Valle del Cauca	PNR Páramo del Duende	2018	HUMAN_OBSERV
10	Bassaricyon neblina Helgen, Pinto, Kays, Helgen, Tsuchiya, Quinn, Wilson & Maldonado, 2013	-76.21	4.78	Colombia	Valle del Cauca	Finca La Floresta	2018	MACHINE_OBSEF

Showing 1 to 10 of 43 entries

Click the "Save" tab. The first save box allows you to download your session. It is available in all the components and modules (See section 9. Save & Load Session for more details). The download options below the Save Session box change depending on which component is selected. Here, you can get a .csv file of the records just acquired. The first option will download the original database fields for every downloaded record (before any filtering). The second option downloads the current table. The third option, "Download all data", is unavailable at this point, but that will change after we include our second species.

*Note to Chrome users: If you find the map is loading incorrectly after downloading an object, specifically the corner tile loads but the rest of the map is gray, closing the download bar at the bottom of the page should reset the map and fix the problem.*



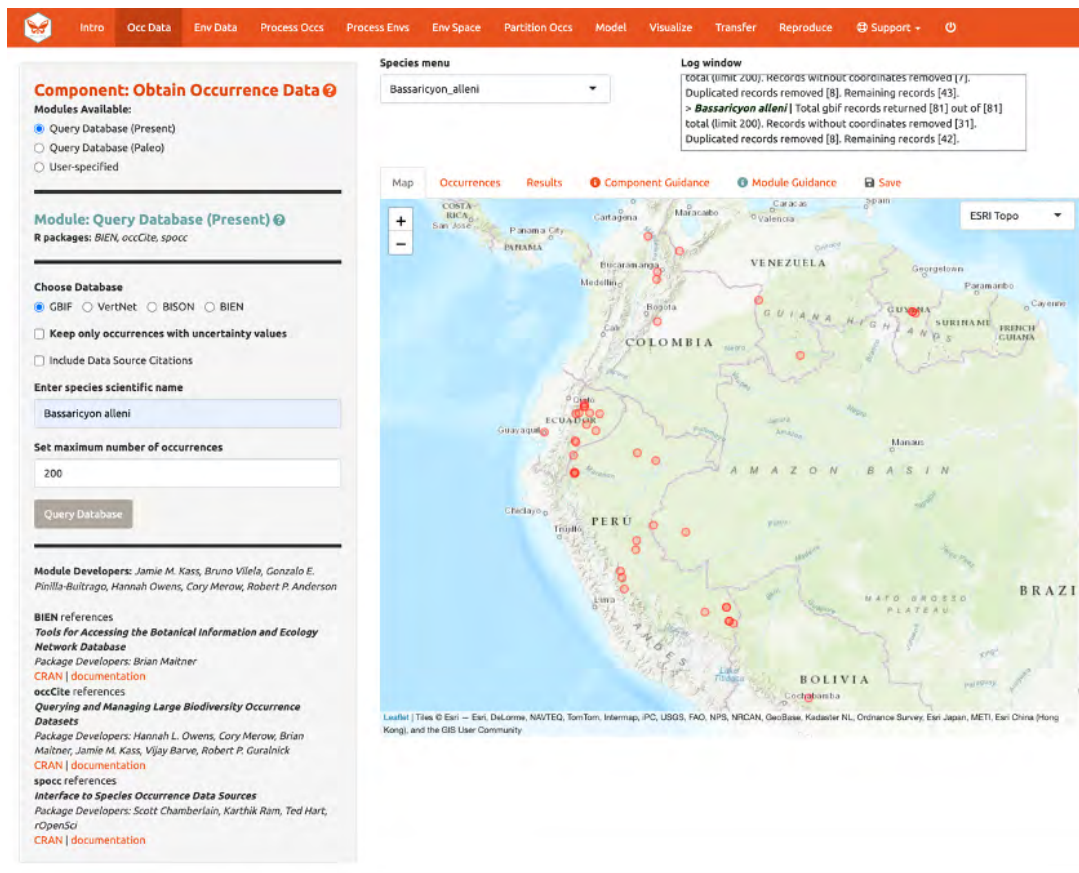
A major improvement in Wallace v2.0 from previous versions is the ability to consider multiple species (separately) in the same session. Let's add another species to model.

Aside from GBIF, you can query the Vertnet (for vertebrate data) and BIEN (for botanical data) for species occurrence records. In the second module "Query Database (Paleo)", you can query PaleobioDB databases for fossil records by selecting a time interval and species. Specific packages may have to be downloaded prior to loading Wallace to use these (e.g., BIEN and paleobioDB). Reminder: depending on the version of Wallace you are using and the availability of associated packages or databases, you may see variations between these instructions and your application options.

If you have your own occurrence data, you can import it using the third module, "User-specified". Your occurrence data file must be a .csv with the columns "scientific\_name", "longitude", and "latitude", explicitly named and in that order. It may have other columns, but those must be the first three. You also have the option to specify the delimiter and separator of your file.

We'll continue with GBIF occurrence data. Search the database for *Bassaricyon alleni* (eastern lowland olingo), keeping the max set at 200. This should return at least 81 records and after cleaning should come to at least 42 records. You might have noticed that the log window was updated, but the map remains the same. The map will not change automatically, as *Bassaricyon neblina* is still selected in the Species menu. Toggle between species to show the map for *Bassaricyon alleni*.

Click back to the "Save" tab. Notice that the third option is now available.



The screenshot shows the Wallace software interface. At the top is a navigation bar with tabs: Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce, Support, and a refresh icon. The main interface is divided into several sections:

- Component: Obtain Occurrence Data**: This panel contains options for 'Modules Available' (Query Database (Present), Query Database (Paleo), User-specified), 'Module: Query Database (Present)', 'Choose Database' (GBIF, VertNet, BISON, BIEN), and a search for the species *Bassaricyon alleni*. It also includes a 'Set maximum number of occurrences' field set to 200 and a 'Query Database' button.
- Species menu**: A dropdown menu currently showing *Bassaricyon alleni*.
- Log window**: A text box displaying the results of the query: 'Total gbif records returned [81] out of [81] total (limit 200). Records without coordinates removed [31]. Duplicated records removed [8]. Remaining records [42].'
- Map**: A map of South America showing the distribution of *Bassaricyon alleni* as red circular points. The map includes labels for countries like Colombia, Venezuela, Peru, Ecuador, and Bolivia, and major cities like Bogotá, Medellín, and Lima. A 'Species menu' dropdown is visible above the map.

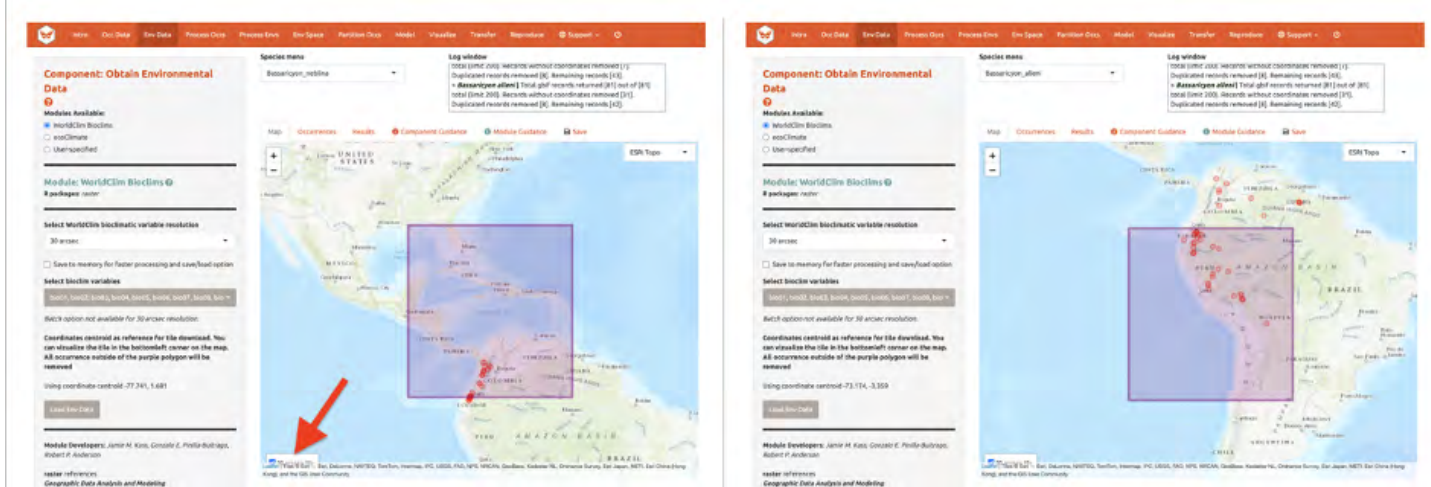
## 4. OBTAIN ENVIRONMENTAL DATA

Next, you will need to obtain environmental variables for the analysis. The values of the variables are extracted for the occurrence records, and this information is provided to the model. These data are in raster form, which simply means a large grid where each grid cell specifies a value. Rasters can be displayed as colored grids on maps (we'll see this later). Click on the component "Env Data". The first module, "WorldClim Bioclims", lets you download bioclimatic variables from WorldClim, a global climate database of interpolated climate surfaces derived from weather station data at multiple resolutions. The interpolation is better for areas with more weather stations, and more uncertainty exists in areas with fewer stations. The "bioclim variables" are summaries of temperature and precipitation that have been proposed to have general biological significance. You have the option to specify a subset of the 19 total variables to use in the analysis.

The second module, "ecoClimate", is a module included with v2 that includes paleoclimate reconstructions. It accesses climatic layers from the PMIP3 – CMIP5 projects from ecoClimate. Users can select from Atmospheric Oceanic General Circulation Models and choose a temporal scenario to use. All ecoClimate layers have a resolution of 0.5 degrees, whereas WorldClim allows resolution options of 30 arcsec, 2.5 arcmin, 5 arcmin, or 10 arcmin.

The third module, "User-specified", is for uploading your own rasters into Wallace. These can be continuous, numerical, or categorical variables to provide to the model. We'll be using WorldClim. The first time you use Wallace, these data are downloaded to a temporary folder on your hard drive; after that, they will simply be loaded from this local directory (which will be quicker than downloading from the web). You also have the option to save to memory for faster processing—this saves the data temporarily as a RasterBrick in your RAM for Wallace to access. Finer resolution

rasters will take longer to download. The finest resolution data (30 arcsec) is served in large global tiles when downloading through R with the raster package (which `wallace` uses) and a single tile that corresponds to the map center will be downloaded. Set the resolution to 30 arcsec and the latitude and longitude of the map center will be given. To visualize how well the tile will cover the occurrence points, click the “30 arcsec tile” box in the bottom left corner of the map. The points outside the tile will be excluded; you may need to zoom out to see fully.



Although you could download the (very big) 30 arcsec global raster from the WorldClim website and load it into Wallace (preferably after cropping it with GIS software or in R), we will instead choose the 2.5 arcmin bioclimatic variable resolution that Wallace serves in a global extent to cover all our occurrence points, and we will keep all 19 bioclimatic variables checked. Note that the selections made will be performed only for the species selected in the Species Menu box, unless the “Batch” box is checked. The “Batch” button will perform the analysis you’ve set up in the module for all the species you have uploaded. You’ll notice this option in many of the modules. If you want to perform individualized analyses for each species (in this case, different environmental variables), leave “Batch” unchecked. Note: The batch option is not available for 30 arcsec resolution since different tiles may need to be accessed.

Check “Batch” and click “Load Env Data”. Notice the progress bar in the bottom-right corner.

After the rasters have loaded, the “Results” tab will display summary information about them (e.g., resolution, extent, cell number, etc.). In addition to downloading the rasters, Wallace will also remove any occurrence points with no environmental values (i.e., points that did not overlap with grid cells with data in the rasters).

You can download your environmental variables from within the Download Data section of the “Save” tab.



## 5. PROCESS OCCURRENCE DATA

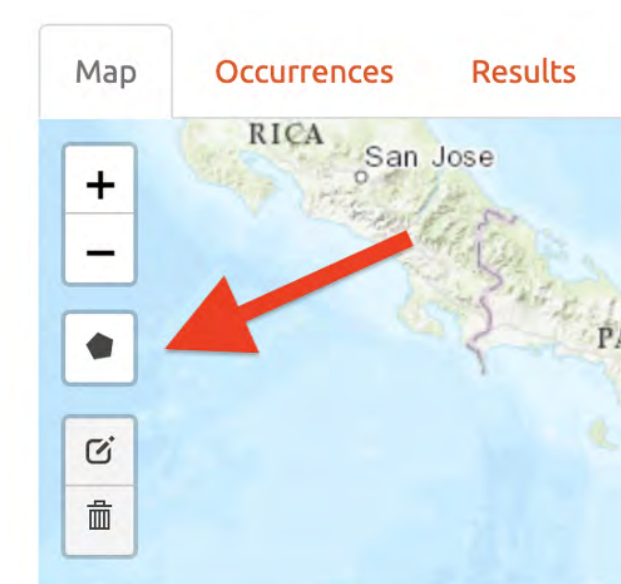
The next component, “Process Occs”, gives you access to some data-cleaning tools. The data you retrieved from GBIF are raw, and there will almost always be some erroneous points. Some basic knowledge of the species’ range can help us remove the most obvious errors. For databases like GBIF that accumulate lots of data from various sources, there are inevitably some dubious localities. For example, coordinates might specify a museum location instead of those associated with the specimen,

or the latitude and longitude might be inverted. In order to eliminate these obviously erroneous records, select only the points you want to keep for analysis with the module “Select Occurrences On Map”. Alternatively, you can also remove specific occurrences by ID with the module “Remove Occurrences by ID”. Even after removing problematic points, those you have left may be clustered due to sampling bias. This often leads to artifactually inflated spatial autocorrelation, which can bias the environmental signal for the occurrence data that the model will attempt to fit. For example, there might be clustering of points near cities because the data are mostly from citizen scientists who live in or near them. Or, the points can cluster around roads because the field biologists who took the data were either making observations while driving or gained access to sites from them. The last module, “Spatial Thin” will help reduce the effects of sampling bias. Unlike other components, in Process Occs the modules are not exclusive, and all can be used in any order.

Make sure *Bassaricyon alleni* is in the species menu.

We will practice using the first two modules with this species. In the first module, we will use the polygon-drawing tool to select occurrences. The polygon drawing tool is useful to draw extents and will be seen in other modules later on as well.

Click on the polygon icon on the map toolbar.



This opens the drawing tool. Click to begin drawing—each click connects to the last with a line. Draw a shape around South America, omitting the record in Bolivia. If you make a mistake in drawing, you can click “Delete last point” or “Cancel” to start over. To finish drawing, click again on the first point you made, or click “Finish” in the drawing tool. This finalizes the shape to use in analysis. Now click “Select Occurrences” and you will see the point in Bolivia disappear. To remove the blue shaded polygon, click on the trashcan icon on the map toolbar and hit “Clear All”. If you are displeased or have made an error, the red “Reset” button in the module interface will revert back to the original points. Since we arbitrarily removed the record in Bolivia, click reset to return to our original dataset.

We will now remove it again, this time using the second module, “Remove Occurrences by ID”. With the pointer, click on the record in Bolivia. Information on the record will pop up, starting with the OccID.

In this case it is OccID #18 (it may be a different number for you). Other information from the attribute table will be available. For example, the record has no information (NA) on the institution code, state/province, or basis. Since we know the OccID number, we can find the full information associated in the Occurrences tab. Click there and find the record. Here we can see it is a preserved specimen from the Museum of Southwestern Biology (MSB). Go back to the map. Enter "18" for the ID to be removed and "Remove Occurrence". You will see the point disappear again. Click reset to get it back again.



We are now left with 35 points for *Bassaricyon alleni* and 21 for *Bassaricyon neblina* (your numbers may be different). You can zoom in to see what the function did. Red points were retained and blue ones were removed. Download the processed occurrence datasets as a .csv file by clicking on the button in the “Save” tab. Reminder: the data downloaded are only for the species currently in the species menu.

## 6. PROCESS ENVIRONMENTAL DATA

Now we will need to choose the study extent for modeling. This will define the region from which “background” points are drawn for model fitting. Background points are meant to sample the environments in the total area available to the study species. Methods like Maxent are known as presence-background techniques because they compare the predictor variable values at background points to those at the occurrence points (as opposed to presence-absence techniques, which require absence data). In making decisions about the study extent, we want to avoid areas the species has historically been unable to move to—for example, regions beyond a barrier like a mountain range or large river that the species cannot cross. Including these areas may send a false signal to the model that these areas are not environmentally suitable. Like every other step of the analysis, please see the relevant guidance text for more details.

You can explore the different options for delineating the study extent here. Each module has two steps: 1) choosing the shape of the background extent, and 2) sampling the background points. To begin, go to the module “Select Study Region”. Under “Step 1”, try out the different options and see how each one draws the background shape. Try increasing and decreasing the buffer to see how the shape is affected. Now set the species to *B. neblina* and use “Select study region” to a minimum convex polygon with a 0.7° buffer distance. Then switch to *B. alleni* and use point buffers with a 0.7° buffer.

Alternatively, you can draw your own polygon (using the same polygon drawing tool we tested in “Component: Process occs”). If you have a file specifying the background extent, you can upload it with the “User-specified Study Region” module. This module can accept a shapefile (must include .shp, .shx, and .dbf files) or a .csv file of polygon vertex coordinates with the field order: longitude, latitude. Note that the polygon you draw or shape you upload needs to contain all the occurrence points.

Next, complete “Step 2”, which both clips the rasters by the study extent and samples the background points. Set the number of background points to 10,000 (larger samples can be appropriate for larger extents or those with finer resolution; see component guidance text), check “Batch”, and click the “Sample” button.

You may find that requesting 10,000 background points exceeds the number of grid cells in the background extent. The available number of points will be given in the log window, and that amount can be used instead of 10,000.

A .zip file of the clipped rasters (e.g., the environmental data clipped to the extent of the background you just created) is available to download in the “Save” tab. Make sure to toggle the species to download the file for each one.



Intro
Occ Data
Env Data
Process Occs
Process Envs
Env Space
Partition Occs
Model
Visualize
Transfer
Reproduce
Support

### Component: Process Environmental Data

**Modules Available:**

- Select Study Region
- Draw Study Region
- User-specified Study Region

---

**Module: Select Study Region by Extent**

R packages: *sp*, *rgeos*

**Step 1: Choose Background Extent**

**Background Extents:**

- bounding box
- minimum convex polygon
- point buffers

**Study region buffer distance (degree)**

Batch

Select

---

**Step 2: Sample Background Points**

Mask predictor rasters by background extent and sample background points

**No. of background points**

Batch

Sample

---

Reset background

---

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

**sp references**

**Classes and Methods for Spatial Data**

Package Developers: *Edzer Pebesma, Roger Bivand, Barry Rowlingson, Virgilio Gomez-Rubio, Robert Hijmans, Michael Sumner, Don MacQueen, Jim Lemon, Finn Lindgren, Josh O'Brien, Joseph O'Rourke*

[CRAN | documentation](#)

**rgeos references**

**Interface to Geometry Engine - Open Source (GEOS)**

Package Developers: *Roger Bivand, Colin Rundel, Edzer Pebesma, Rainer Stuetz, Karl Ove Hufthammer, Patrick Giraudoux, Martin Davis, Sandro Santilli*

[CRAN | documentation](#)

**Species menu**

Bassaricyon\_neblina



**Log window**

To use:

- > **Bassaricyon allenii** | Total records thinned (10 km) to 35 localities
- > **Bassaricyon neblina** | Total records thinned (10 km) to 22 localities
- > **Bassaricyon neblina** | Study extent: minimum convex polygon. Buffered by 0.7 degrees.

Map Occurrences Results Component Guidance Module Guidance Save

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


Intro Occ Data Env Data Process Occs Process Envs Env Space Partition Occs Model Visualize Transfer Reproduce Support + 

### Component: Process Environmental Data

**Modules Available:**

- Select Study Region
- Draw Study Region
- User-specified Study Region

---

**Module: Select Study Region by Extent**

**R packages:** sp, rgeos

---

**Step 1: Choose Background Extent**

**Background Extents:**

- bounding box
- minimum convex polygon
- point buffers

**Study region buffer distance (degree)**

Batch

**Select**

---

**Step 2: Sample Background Points**

**Mask predictor rasters by background extent and sample background points**

**No. of background points**

Batch

**Sample**

---

**Reset** background

---

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

**sp references**

**Classes and Methods for Spatial Data**

*Package Developers: Edzer Pebesma, Roger Bivand, Barry Rowlingson, Virgilio Gomez-Rubio, Robert Hijmans, Michael Sumner, Don MacQueen, Jim Lemon, Finn Lindgren, Josh O'Brien, Joseph O'Rourke*

[CRAN | documentation](#)

**rgeos references**

**Interface to Geometry Engine - Open Source (GEOS)**

*Package Developers: Roger Bivand, Colin Rundel, Edzer Pebesma, Rainer Stuetz, Karl Ove Hufthammer, Patrick Giraudoux, Martin Davis, Sandro Santilli*

[CRAN | documentation](#)

**Species menu**


▼
Bassaricyon\_allyi

**Log window**

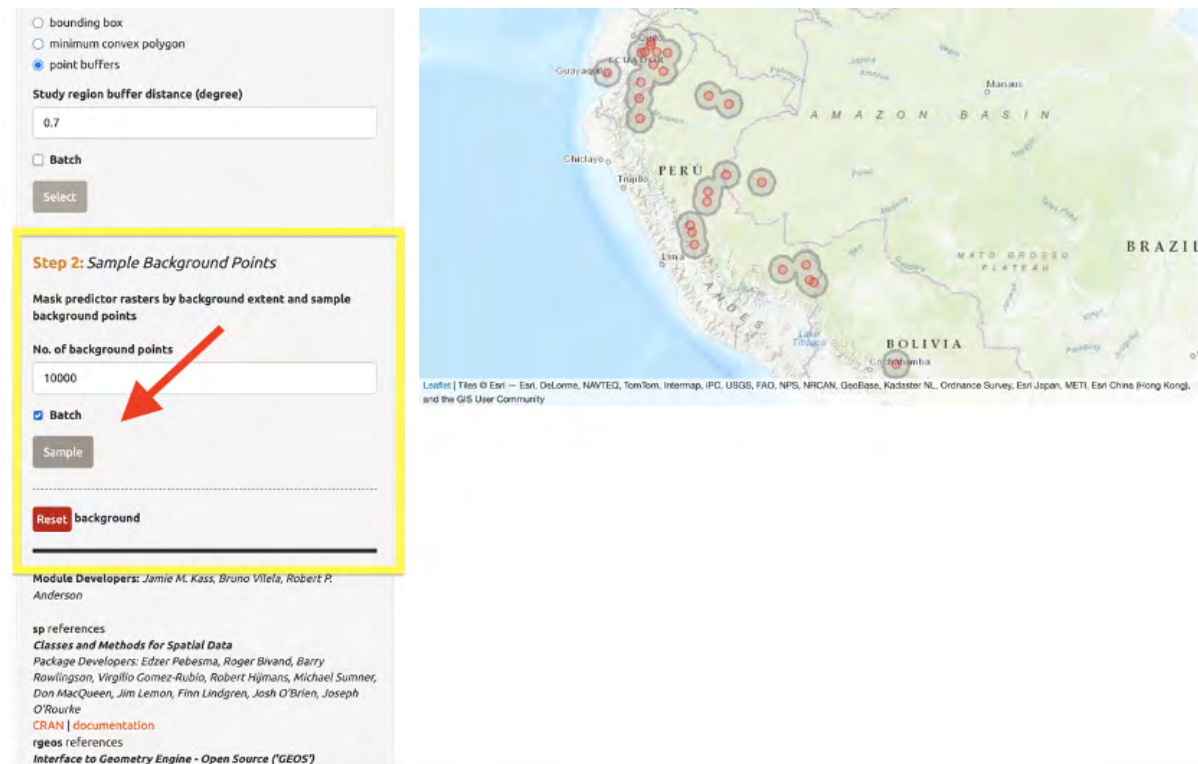
```

> Bassaricyon neblina | Total records thinned (10 km) to 22 localities
> Bassaricyon neblina | Study extent: minimum convex polygon.
  Buffered by 0.7 degrees.
> Bassaricyon allyi | Study extent: buffered points. Buffered by 0.7
  degrees.
                    
```

Map Occurrences Results 1 Component Guidance i Module Guidance S Save



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



bounding box  
 minimum convex polygon  
 point buffers

Study region buffer distance (degree)

0.7

Batch

Select

**Step 2: Sample Background Points**

Mask predictor rasters by background extent and sample background points

No. of background points

10000

Batch

Sample

Reset background

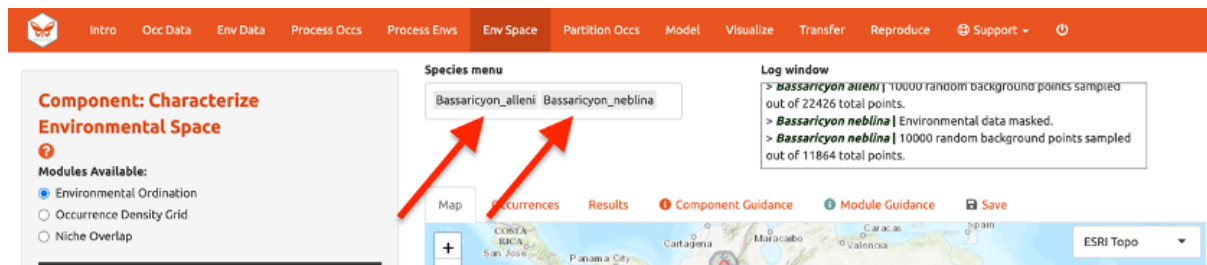
Module Developers: Jamie M. Kass, Bruno Vitela, Robert P. Anderson

sp references  
**Classes and Methods for Spatial Data**  
 Package Developers: Edzer Pebesma, Roger Bivand, Barry Rowlington, Virgilio Gomez-Rubio, Robert Hijmans, Michael Sumner, Don MacQueen, Jim Lemon, Finn Lindgren, Josh O'Brien, Joseph O'Rourke  
[CRAN | documentation](#)  
 rgeos references  
 Interface to Geometry Engine - Open Source ("GEOS")

## 7. CHARACTERIZE ENVIRONMENTAL SPACE

“Component: Characterize Environmental Space” contains multi-species analyses and is optional. Unlike some other components which let you perform the modules in any order, the modules within “Characterize Environmental Space” are sequential and thus need to be performed consecutively (i.e., you can’t get an “Occurrence Density Grid” without first performing an “Environmental Ordination”).

Before we begin the Module: “Environmental Ordination” analysis, you need to select two species to work with. If you had more than two species uploaded, select two from the species menu drop-down. Since we only have two uploaded, click in the species menu box and select the second species. Both names will appear in the box simultaneously—this functionality is currently only available for the “Characterize Environmental Space” component.



Intro Occ Data Env Data Process Occs Process Envs **Env Space** Partition Occs Model Visualize Transfer Reproduce Support

**Component: Characterize Environmental Space**

Modules Available:

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

Species menu

Bassaricyon\_alleni Bassaricyon\_neblina

Log window

```

> Bassaricyon alleni | 10000 random background points sampled out of 22426 total points.
> Bassaricyon neblina | Environmental data masked.
> Bassaricyon neblina | 10000 random background points sampled out of 11864 total points.
    
```

Module: “Environmental Ordination” is for conducting an ordination approach called Principal Component Analysis (PCA), that maximizes the variation contained in the predictor variables into fewer ones. To perform a PCA, select the variables available for both species by checking/unchecking the bioclimatic variables. Choose between “Occurrences Only” or “Occurrences & Background” for the plot selection and set the x- and y-axis components. The PCA Scatter Plot appears in the Results tab.

**Component: Characterize Environmental Space**

**Modules Available:**

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

**Module: Environmental Ordination**

R packages: ade4

Select variables available for both species

wc2.1\_2.5m\_bio\_1, wc2.1\_2.5m\_bio\_2, wc2.1\_2.5m\_bio\_3

Plot selection: Occurrences + Background

X-axis Component: 1

Y-axis Component: 2

Run

Module Developers: Olivier Broennimann, Gonzalo E. Pinilla-Buitrago, Jamie M. Kass

ade4 references  
 Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences  
 Package Developers: Stéphane Dray <stephane.dray@univ-lyon1.fr>, Anne-Béatrice Dufour <anne-beatrice.dufour@univ-lyon1.fr>, and Jean Thioulouse <jean.thioulouse@univ-lyon1.fr>, with contributions from Thibaut Jombart, Sandrine Pavoine, Jean R. Lobry, Sébastien Ollier, Daniel Borcard, Pierre Legendre, Stéphanie Bougeard and Aurélie Siberchicot. Based on earlier work by Daniel Cheset.  
 CRAN | documentation

Species menu: Bassaricyon\_alleni, Bassaricyon\_neblina

Log window

```
> bassaricyon neblina | environmental data masked.
> Bassaricyon neblina | 10000 random background points sampled out of 13007 total points.
> Bassaricyon alleni and Bassaricyon neblina | Principal component analysis.
```

Map Occurrences Results **Component Guidance** Module Guidance Save

PCA scatter plot PCA correlation circle PCA screeplot PCA results summary

You can also view the PCA correlation circle, PCA screeplot, and the PCA results summary. For more information on these statistics and how to evaluate the results, consult the module guidance text.

**Component: Characterize Environmental Space**

**Modules Available:**

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

**Module: Environmental Ordination**

R packages: ade4

Select variables available for both species

wc2.1\_2.5m\_bio\_1, wc2.1\_2.5m\_bio\_2, wc2.1\_2.5m\_bio\_3

Plot selection: Occurrences + Background

X-axis Component: 1

Y-axis Component: 2

Run

Module Developers: Olivier Broennimann, Gonzalo E. Pinilla-Buitrago, Jamie M. Kass

ade4 references  
 Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences  
 Package Developers: Stéphane Dray <stephane.dray@univ-lyon1.fr>, Anne-Béatrice Dufour <anne-beatrice.dufour@univ-lyon1.fr>, and Jean Thioulouse <jean.thioulouse@univ-lyon1.fr>, with contributions from Thibaut Jombart, Sandrine Pavoine, Jean R. Lobry, Sébastien Ollier, Daniel Borcard, Pierre Legendre, Stéphanie Bougeard and Aurélie Siberchicot. Based on earlier work by Daniel Cheset.  
 CRAN | documentation

Species menu: Bassaricyon\_alleni, Bassaricyon\_neblina

Log window

```
> bassaricyon neblina | environmental data masked.
> Bassaricyon neblina | 10000 random background points sampled out of 13007 total points.
> Bassaricyon alleni and Bassaricyon neblina | Principal component analysis.
```

Map Occurrences Results **Component Guidance** Module Guidance Save

PCA scatter plot PCA correlation circle **PCA screeplot** PCA results summary



Intro Occ Data Env Data Process Occs Process Envs Env Space Partition Occs Model Visualize Transfer Reproduce Support +

### Component: Characterize Environmental Space

**Modules Available:**

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

**Module: Environmental Ordination**

R packages: *ade4*

Select variables available for both species

wc2.1\_2.5m\_bio\_1, wc2.1\_2.5m\_bio\_2, wc2.1\_2.5m\_bio\_3

Plot selection: Occurrences + Background

X-axis Component: 1

Y-axis Component: 2

Run

**Module Developers:** Olivier Broennimann, Gonzalo E. Pinilla-Buitrago, Jamie M. Kass

**ade4 references**  
*Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences*  
 Package Developers: Stéphane Dray <stephane.dray@univ-lyon1.fr>, Anne-Béatrice Dufour <anne-beatrice.dufour@univ-lyon1.fr>, and Jean Thioulouse <jean.thioulouse@univ-lyon1.fr>, with contributions from Thibaut Jombart, Sandrine Pavoine, Jean R. Lobry, Sébastien Ollier, Daniel Borcard, Pierre Legendre, Stéphanie Bougeard and Aurélie Siberchicot. Based on earlier work by Daniel Chessel.  
[CRAN](#) | [documentation](#)

**Species menu**

Bassaricyon\_alleni Bassaricyon\_neblina

**Log window**

```
> bassaricyon_neblina | Environmental data masked.
> Bassaricyon_neblina | 10000 random background points sampled out of 13007 total points.
> Bassaricyon_alleni and Bassaricyon_neblina | Principal component analysis.
```

Map Occurrences Results **Component Guidance** Module Guidance Save

PCA scatter plot PCA correlation circle PCA screeplot **PCA results summary**

Intro Occ Data Env Data Process Occs Process Envs Env Space Partition Occs Model Visualize Transfer Reproduce Support +

### Component: Characterize Environmental Space

**Modules Available:**

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

**Module: Environmental Ordination**

R packages: *ade4*

Select variables available for both species

wc2.1\_2.5m\_bio\_1, wc2.1\_2.5m\_bio\_2, wc2.1\_2.5m\_bio\_3

Plot selection: Occurrences + Background

X-axis Component: 1

Y-axis Component: 2

Run

**Module Developers:** Olivier Broennimann, Gonzalo E. Pinilla-Buitrago, Jamie M. Kass

**ade4 references**  
*Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences*  
 Package Developers: Stéphane Dray <stephane.dray@univ-lyon1.fr>, Anne-Béatrice Dufour <anne-beatrice.dufour@univ-lyon1.fr>, and Jean Thioulouse <jean.thioulouse@univ-lyon1.fr>, with contributions from Thibaut Jombart, Sandrine Pavoine, Jean R. Lobry, Sébastien Ollier, Daniel Borcard, Pierre Legendre, Stéphanie Bougeard and Aurélie Siberchicot. Based on earlier work by Daniel Chessel.  
[CRAN](#) | [documentation](#)

**Species menu**

Bassaricyon\_alleni Bassaricyon\_neblina

**Log window**

```
> bassaricyon_neblina | Environmental data masked.
> Bassaricyon_neblina | 10000 random background points sampled out of 13007 total points.
> Bassaricyon_alleni and Bassaricyon_neblina | Principal component analysis.
```

Map Occurrences Results **Component Guidance** Module Guidance Save

PCA scatter plot PCA correlation circle **PCA screeplot** PCA results summary

Variance explained:

PC1(%)	PC2(%)	PC3(%)	PC4(%)	PC5(%)	PC6(%)	PC7(%)	PC8(%)	PC9(%)
49.88	26.94	0.81	6.84	2.80	2.25	0.66	0.40	0.25
PC10(%)	PC11(%)	PC12(%)	PC13(%)	PC14(%)	PC15(%)	PC16(%)	PC17(%)	PC18(%)
0.05	0.05	0.02	0.02	0.01	0.01	0.00	0.00	0.00

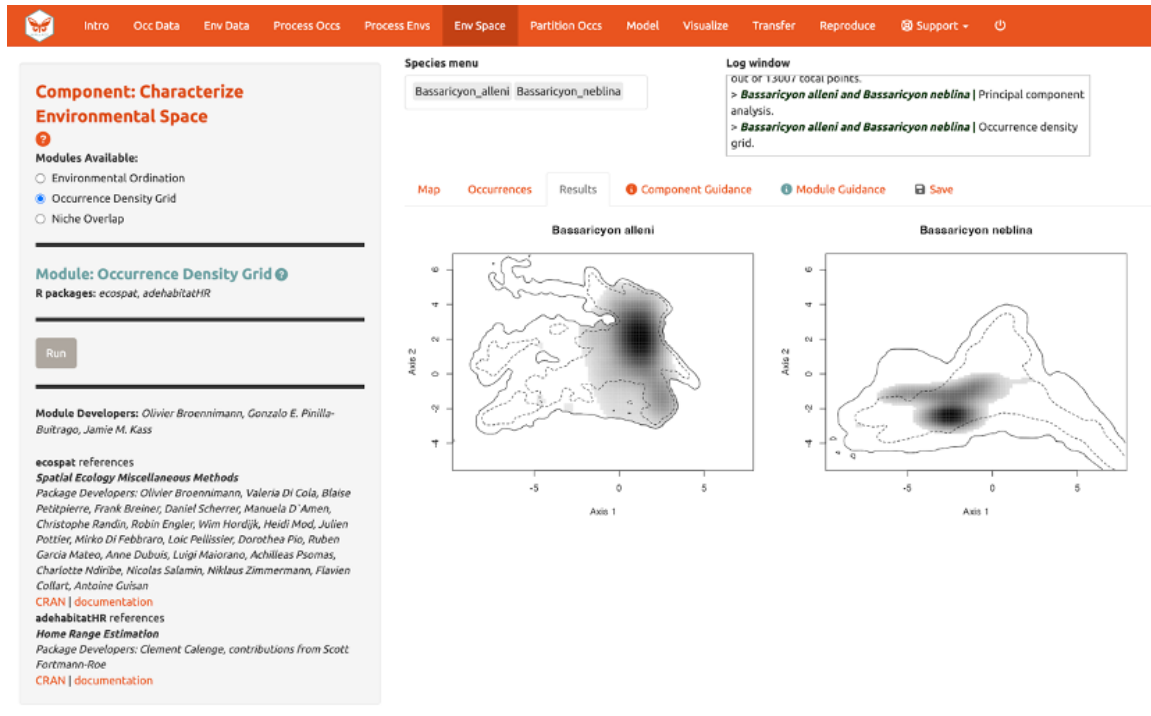
Loadings:

	PC1	PC2	PC3	PC4
wc2.1_2.5m_bio_1	0.284258443	0.19228950	0.15441379	-0.013411705
wc2.1_2.5m_bio_2	-0.895531527	0.23926931	-0.8598021	-0.036101029
wc2.1_2.5m_bio_3	0.887238083	-0.34804009	0.38845585	-0.348852553
wc2.1_2.5m_bio_4	0.014532886	0.38854288	0.45584595	0.02287801
wc2.1_2.5m_bio_5	0.267545888	0.23662115	0.12688777	-0.067239754
wc2.1_2.5m_bio_6	0.294403789	0.12828828	0.21633843	0.073378191
wc2.1_2.5m_bio_7	-0.878312288	0.33529789	-0.26547638	-0.425476863
wc2.1_2.5m_bio_8	0.281183969	0.28276138	0.12654324	-0.080863626
wc2.1_2.5m_bio_9	0.286958887	0.17160882	0.18966579	-0.006847377
wc2.1_2.5m_bio_10	0.281251683	0.28668646	0.12943825	-0.004431633
wc2.1_2.5m_bio_11	0.286851187	0.17154885	0.19271618	-0.007116209
wc2.1_2.5m_bio_12	0.272958584	-0.19371912	-0.21813986	-0.365384658
wc2.1_2.5m_bio_13	0.275289232	-0.09587888	0.28514865	0.117988331
wc2.1_2.5m_bio_14	0.219653637	-0.27196439	-0.11485288	-0.249617344
wc2.1_2.5m_bio_15	-0.850881798	0.29645420	-0.17183384	0.482989220
wc2.1_2.5m_bio_16	0.275881684	-0.09263420	-0.38542793	0.112445221
wc2.1_2.5m_bio_17	0.227278991	-0.27588788	-0.18599367	-0.224832747
wc2.1_2.5m_bio_18	0.215884388	-0.16138861	-0.38914845	0.066615995
wc2.1_2.5m_bio_19	0.228969583	-0.21857561	-0.06937515	-0.081194791

PC5 PC6 PC7 PC8

wc2.1_2.5m_bio_1	0.851882784	0.088143398	-0.08318115	-0.818915163
wc2.1_2.5m_bio_2	-0.141528265	-0.421685477	0.809855859	-0.841348853
wc2.1_2.5m_bio_3	-0.863279232	-0.488757106	0.282743643	-0.542183984
wc2.1_2.5m_bio_4	0.872715669	0.484787997	0.361423448	-0.568666536
wc2.1_2.5m_bio_5	-0.813593842	0.004299888	-0.009038647	-0.883432153
wc2.1_2.5m_bio_6	0.884883821	0.039537438	0.817938738	-0.835784793
wc2.1_2.5m_bio_7	-0.856893287	-0.185798843	-0.881432323	0.897157927
wc2.1_2.5m_bio_8	0.124541177	-0.015869348	0.026421419	-0.029593888
wc2.1_2.5m_bio_9	-0.843887117	0.842646596	0.002139871	0.818668329
wc2.1_2.5m_bio_10	0.848936398	0.838380865	0.832064234	-0.823675865
wc2.1_2.5m_bio_11	0.848691561	-0.025218170	0.019687538	0.837642584
wc2.1_2.5m_bio_12	-0.893375351	-0.833787161	-0.137780174	0.888966897
wc2.1_2.5m_bio_13	-0.196754823	-0.175184533	-0.286699694	-0.045250581
wc2.1_2.5m_bio_14	0.177347893	0.111865268	0.419045573	0.338986516
wc2.1_2.5m_bio_15	-0.218815338	-0.492811438	0.527188515	0.278079579
wc2.1_2.5m_bio_16	-0.178631699	-0.158267284	-0.28863242	-0.838863491
wc2.1_2.5m_bio_17	0.128256325	0.098733755	0.334856892	0.275927254

Next, run an “Occurrence Density Grid”. This calculates and plots which part of the environmental space is occupied more densely by the species and the availability of environmental conditions present within the background extent. Darker areas represent higher occurrence density. Areas within solid lines represent all environmental conditions available in the background extent, and areas within dashed lines represent the 50% most frequent ones.



**Component: Characterize Environmental Space**

**Modules Available:**

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

**Module: Occurrence Density Grid**

R packages: *ecospat, adehabitatHR*

Run

**Module Developers:** Olivier Broennimann, Gonzalo E. Pinilla-Buitrago, Jamie M. Kass

**ecospat references**

**Spatial Ecology Miscellaneous Methods**

Package Developers: Olivier Broennimann, Valeria Di Cola, Blaise Petitpierre, Frank Breiner, Daniel Scherrer, Manuela D'Amen, Christophe Randin, Robin Engler, Wim Hordijk, Heidi Mod, Julien Pottier, Mirko Di Febbraro, Loic Pellissier, Dorothea Pio, Ruben Garcia Mateo, Anne Dubuis, Luigi Maiorano, Achilles Psomas, Charlotte Ndiribe, Nicolas Salamin, Niklaus Zimmermann, Flavien Collart, Antoine Guban

CRAN | documentation

**adehabitatHR references**

**Home Range Estimation**

Package Developers: Clement Calenge, contributions from Scott Fortmann-Roe

CRAN | documentation

**Species menu**

Bassaricyon\_alleni Bassaricyon\_neblina

**Log window**

OUT OF 13007 TOTAL POINTS.

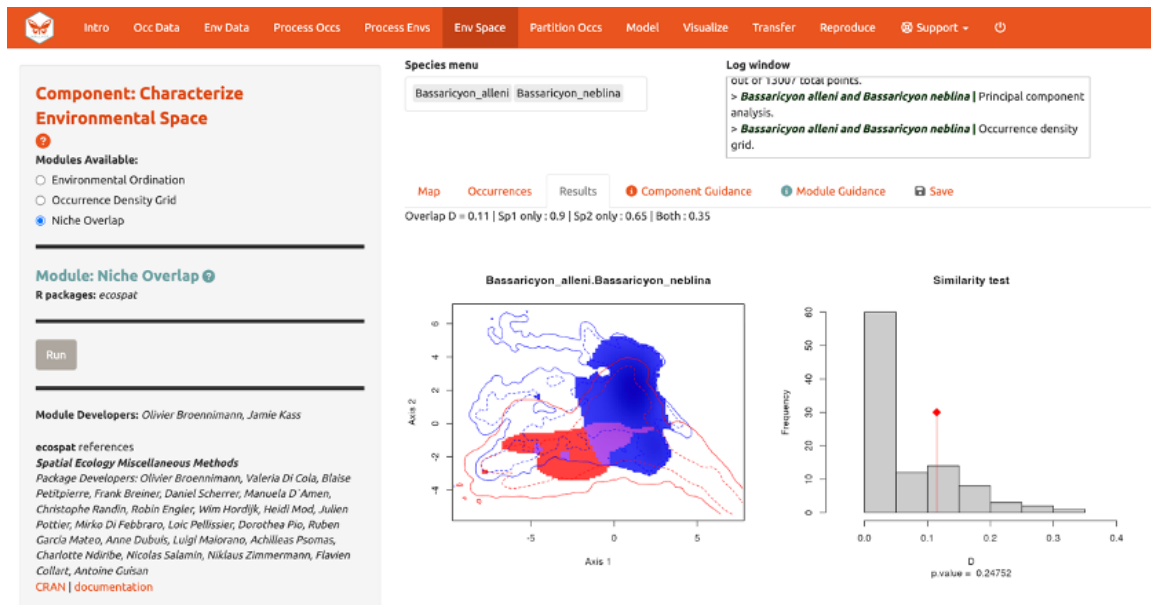
- > **Bassaricyon alleni and Bassaricyon neblina** | Principal component analysis.
- > **Bassaricyon alleni and Bassaricyon neblina** | Occurrence density grid.

Map Occurrences Results **Component Guidance** Module Guidance Save

**Bassaricyon alleni**

**Bassaricyon neblina**

And calculate “Niche Overlap” ...



**Component: Characterize Environmental Space**

**Modules Available:**

- Environmental Ordination
- Occurrence Density Grid
- Niche Overlap

**Module: Niche Overlap**

R packages: *ecospat*

Run

**Module Developers:** Olivier Broennimann, Jamie Kass

**ecospat references**

**Spatial Ecology Miscellaneous Methods**

Package Developers: Olivier Broennimann, Valeria Di Cola, Blaise Petitpierre, Frank Breiner, Daniel Scherrer, Manuela D'Amen, Christophe Randin, Robin Engler, Wim Hordijk, Heidi Mod, Julien Pottier, Mirko Di Febbraro, Loic Pellissier, Dorothea Pio, Ruben Garcia Mateo, Anne Dubuis, Luigi Maiorano, Achilles Psomas, Charlotte Ndiribe, Nicolas Salamin, Niklaus Zimmermann, Flavien Collart, Antoine Guban

CRAN | documentation

**Species menu**

Bassaricyon\_alleni Bassaricyon\_neblina

**Log window**

OUT OF 13007 TOTAL POINTS.

- > **Bassaricyon alleni and Bassaricyon neblina** | Principal component analysis.
- > **Bassaricyon alleni and Bassaricyon neblina** | Occurrence density grid.

Map Occurrences Results **Component Guidance** Module Guidance Save

Overlap D = 0.11 | Sp1 only: 0.9 | Sp2 only: 0.65 | Both: 0.35

**Bassaricyon\_alleni.Bassaricyon\_neblina**

**Similarity test**

Frequency

D

p-value = 0.24752

The niche overlap quantification is based on the occurrence and background densities in the available environmental space estimated in Module: “Occurrence Density Grid”. The overlap is quantified using Schoener’s D metric. The environmental conditions covered only by the niche of species 1 are shown in blue, the environmental conditions covered only by the niche of species 2

are shown in red, and the environmental conditions covered by both species, or the niche overlap, is shown in purple. In the Similarity Test, if the observed overlap (red line) is higher than 95% of the simulated overlaps ( $p$ -value  $< 0.05$ ), we can consider the two species to be more similar than random, which is not what we see here. Again, consult the module guidance texts for more help to understand the analyses and help on evaluating the results.

Download the PCA results (.zip), density grid (.png), and overlap plot (.png) from the “Save” tab.

## 8. PARTITION OCCURRENCES

We have not built any distribution models yet, but before we do, we will make decisions on how to partition our data for evaluation. In order to determine the strength of the model’s predictive ability, you theoretically need independent data to test it. When no independent datasets exist, one solution is to partition your data into subsets that we assume are independent of each other, then sequentially build a model on all the subsets but one and evaluate the performance of this model on the left-out subset. This is known as  $k$ -fold cross-validation (where “ $k$ ” is the total number of subsets, or “folds”), which is quite prevalent in statistics, especially the fields of machine learning and data science. After this sequential model-building exercise is complete, Wallace averages the model performance statistics over all the iterations and then builds a model using *all* the data.

There is a whole literature on how to best partition data for evaluating models. One option is to simply partition randomly, but with spatial data we run the risk that the groups are not spatially independent of each other. The jackknife method (“leave-one-out”) is recommended for species with small sample sizes and has previously been used for modeling *Bassaricyon neblina* (Gerstner et al. 2018) but may have long computational times.

Another option is to partition spatially—for example, by drawing lines on a map to divide the data. Spatial partitioning with  $k$ -fold cross-validation forces the model to predict into regions that are distant from those used to train the model (note that Wallace also excludes background points from regions corresponding to the withheld partition). For *Bassaricyon alleni*, environmental conditions in Colombia and Ecuador may differ considerably from conditions in Bolivia. If the model has accurate predictions on average on withheld spatially partitioned data, it likely has good transferability, which means it can transfer well to the new values of predictor variables (because distant areas are usually more environmentally different than close ones). As always, please refer to the guidance text for more details on all the types of partitioning offered in Wallace. The first image on the next page is an example of jackknife ( $k = n$ ), which assigns each point to its own partition group, so the number of bins equals the number of occurrences.

The second image on the next page is an example of spatial blocking, which assigns each point to one of four spatially separate partition groups.

We’ll use this last method now for faster computation, but it is recommended to review the guidance text and other literature—and talk to your peers!—to make an informed decision on partition methods.

Partition both species using Module: “Spatial Partition Block” ( $k = 4$ ) option.

Intro Occ Data Env Data Process Occs Process Envs Env Space Partition Occs Model Visualize Transfer Reproduce Support

**Component: Partition Occurrence Data**

Modules Available:

- Non-spatial Partition
- Spatial Partition

---

**Module: Non-spatial Partition**

R packages: ENMeval

---

Options Available:

Jackknife (k = n)

Batch

Partition

---

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

ENMeval references  
*Automated Tuning and Evaluations of Ecological Niche Models*  
 Package Developers: Jamie M. Kass, Robert Muscarella, Peter J. Galante, Corentin Bohl, Gonzalo E. Buitrago-Pinilla, Robert A. Boria, Mariano Soley-Guardia, Robert P. Anderson  
[CRAN | documentation](#)

Species menu

Bassaricyon\_alleni

Log window

```
> Bassaricyon alleni and Bassaricyon neblina | Principal component analysis.
> Bassaricyon alleni and Bassaricyon neblina | Occurrence density grid.
> Bassaricyon alleni | Occurrences partitioned by jackknife method.
```

Map Occurrences Results Component Guidance Module Guidance Save

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

Intro Occ Data Env Data Process Occs Process Envs Env Space Partition Occs Model Visualize Transfer Reproduce Support

**Component: Partition Occurrence Data**

Modules Available:

- Non-spatial Partition
- Spatial Partition

---

**Module: Spatial Partition**

R packages: ENMeval

---

Options Available:

Block (k = 4)

Batch

Partition

---

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

ENMeval references  
*Automated Tuning and Evaluations of Ecological Niche Models*  
 Package Developers: Jamie M. Kass, Robert Muscarella, Peter J. Galante, Corentin Bohl, Gonzalo E. Buitrago-Pinilla, Robert A. Boria, Mariano Soley-Guardia, Robert P. Anderson  
[CRAN | documentation](#)

Species menu

Bassaricyon\_neblina

Log window

```
analysis.
> Bassaricyon alleni and Bassaricyon neblina | Occurrence density grid.
> Bassaricyon alleni | Occurrences partitioned by jackknife method.
> Bassaricyon neblina | Occurrences partitioned by block method.
```

Map Occurrences Results Component Guidance Module Guidance Save

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



## 9. SAVE & LOAD SESSION

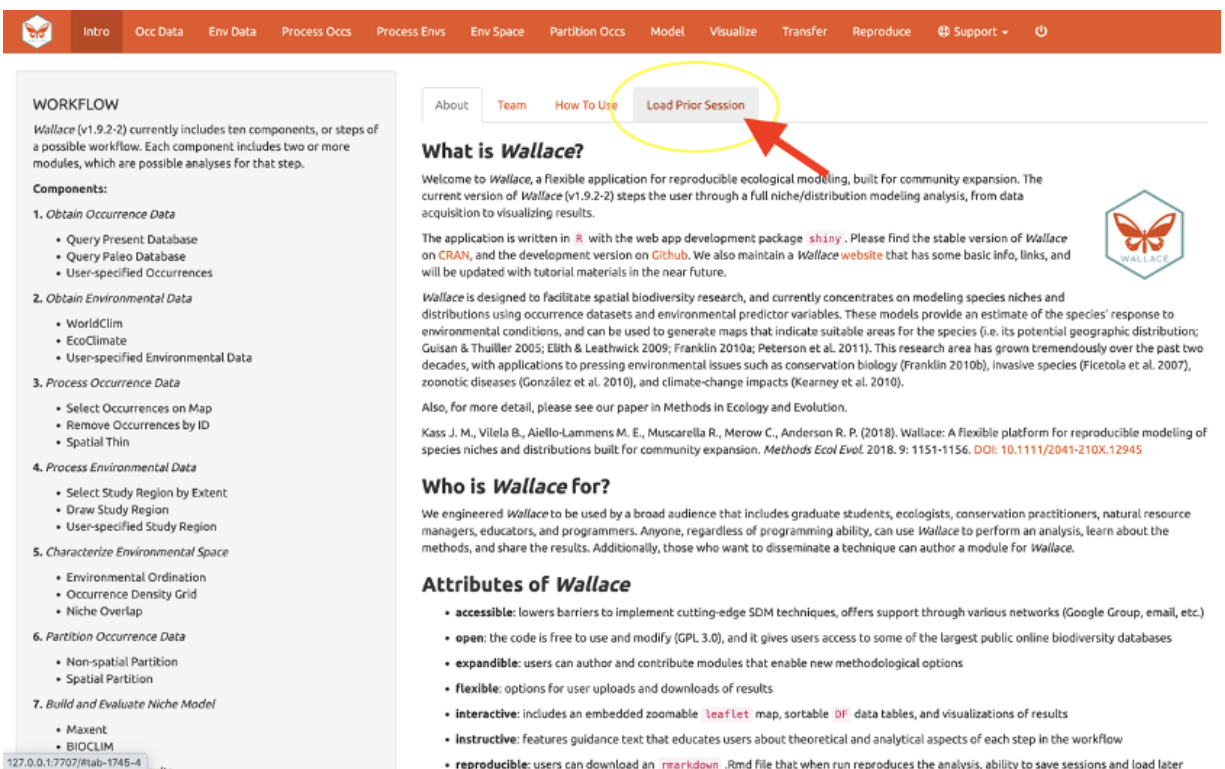
Before we go into “Modeling”, let’s explore one of the great features of Wallace v2, which is the ability to stop and save your progress to be continued later. If you want to skip this step (and risk losing everything if an error occurs except the data or results you have downloaded), you can move on to “Model”.

Click “Save Session” within the “Save” tab. This tab is available from any of the Components. This will save your progress as an RDS (.rds) file, a file type used to save R objects. After it is saved, you can hit the stop sign in the upper right corner or close the browser window and exit R/RStudio. *Note: if the Wallace session is closed before saving results and/or the session, all work will be lost.*

When you are ready to resume, load `wallace` again.

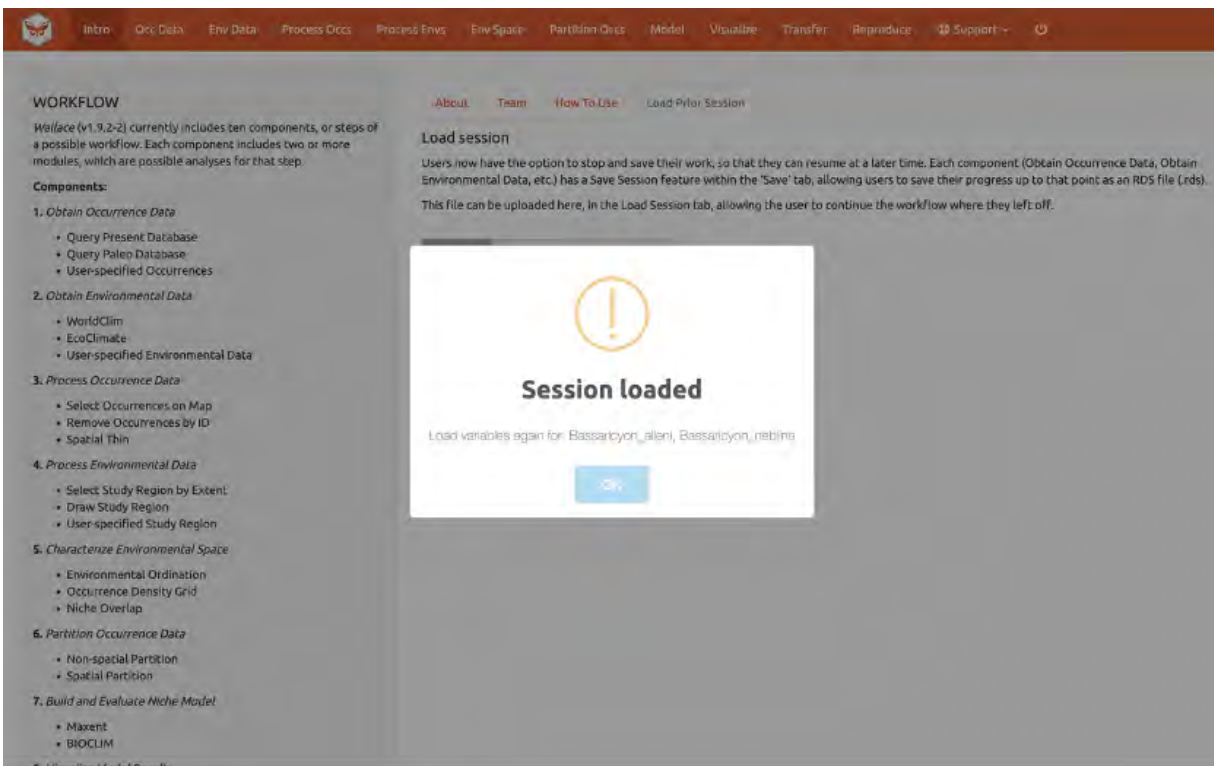
```
library(wallace)
run_wallace()
```

In the “Intro” component, use the “Load Prior Session” tab to import your .rds session file.



The screenshot shows the Wallace v2 web application interface. The top navigation bar is orange and contains tabs: Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce, Support. The 'Intro' component is active, showing a 'WORKFLOW' sidebar on the left with 7 steps. The main content area has a 'Load Prior Session' tab highlighted with a red circle and a red arrow pointing to it. Below the tabs is a 'What is Wallace?' section with introductory text and a logo. Further down is a 'Who is Wallace for?' section and an 'Attributes of Wallace' section with a bulleted list of features.

A box will pop up—it looks like other Wallace warning messages, but in this case it is indicating the session is loaded. It may be necessary to reload your variables, using “Occ Data” and “Env Data” as previously carried out. You can now carry on with the previous analysis.



## 10. MODEL

We are now ready to build a distribution model. Wallace v2.0 provides two algorithm options; Maxent and BIOCLIM. For this vignette, we'll use Maxent, a machine learning method that can fit a range of functions to patterns in the data, from simple (i.e., straight lines) to complex (i.e., curvy or with lines that can change direction; these can get jagged if complexity is not controlled). For more details on Maxent, please consult the Maxent website and guidance text.

Maxent is available to run through `maxnet` package or through Java with the `maxent.jar` option. In the interest of time and to avoid Java-related issues, let's choose the following modeling options:

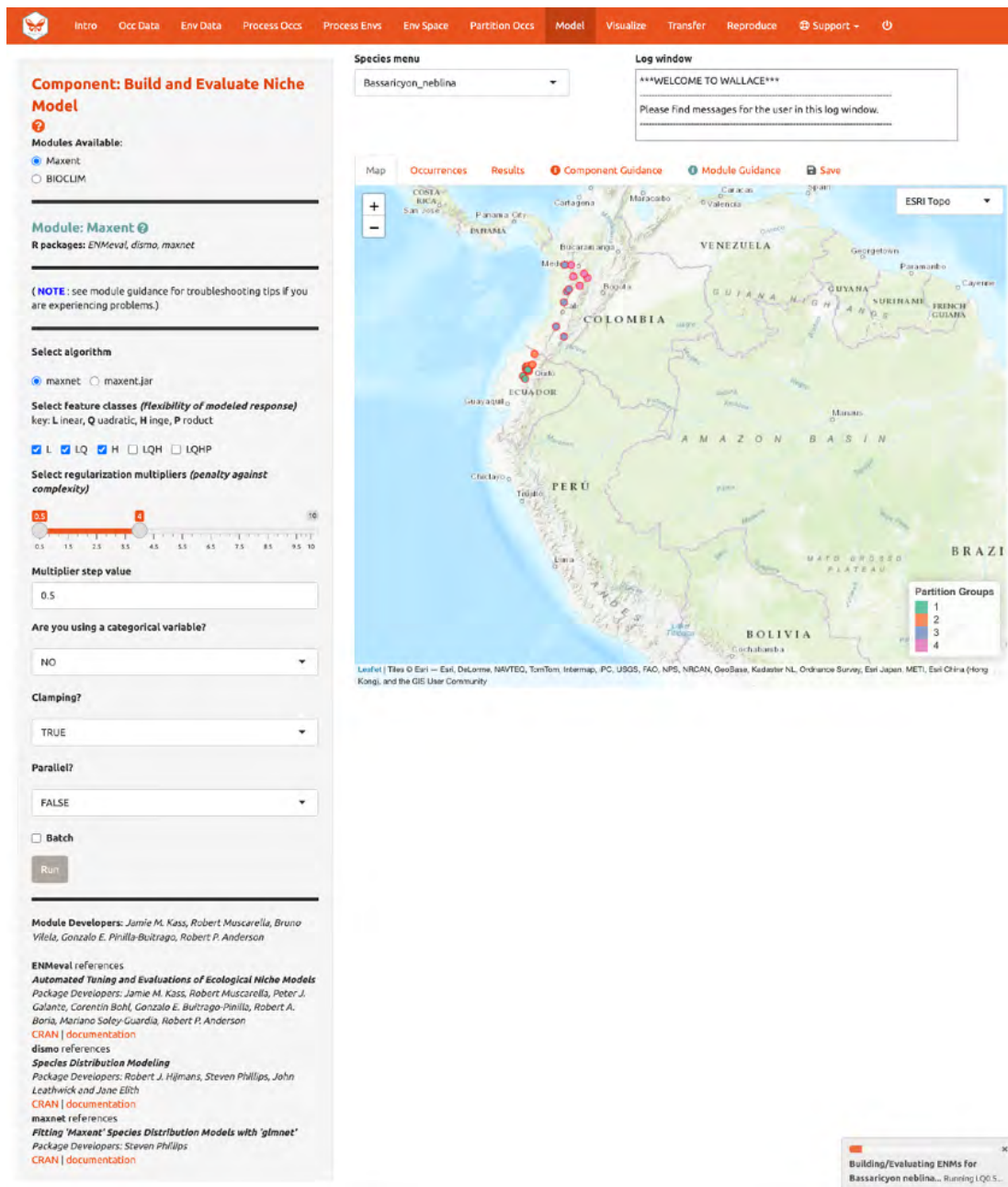
- Choose `maxnet`
- Select L, LQ, and H feature classes. These are the shapes that can be fit to the data:
  - L = Linear, e.g.  $\text{temp} + \text{precip}$
  - Q = Quadratic, e.g.  $\text{temp}^2 + \text{precip}^2$
  - H = Hinge, e.g. piecewise linear functions, like splines (think of a series of lines that are connected together)
- Select regularization multipliers between 0.5 and 4 with a step value of 0.5.
  - Regularization is a penalty on model complexity.
  - Higher values = smoother, less complex models. Basically, all predictor variable coefficients are shrunk progressively until some reach 0, when they drop out of the model. Only those variables with the greatest predictive contribution remain in the model.
- Keep "NO" selected for categorical variables. This option is to indicate if any of your predictor variables are categorical, like soil or vegetation classes.
  - Had you loaded categorical variables, you would check here and then indicate which of the rasters is categorical.

- Set Clamping? to "TRUE". This will clamp the model predictions (i.e., keep the environmental values more extreme than those present in the background data to within the bounds of the background data).
- If you set Parallel? to "TRUE", you can indicate the number of cores for parallel processing.

We will construct a model for *Bassaricyon neblina*, but note that the Batch feature can be checked to run these selections for all species you have uploaded.

Make sure *Bassaricyon neblina* is selected in the species menu and Batch is unchecked before hitting "Run".

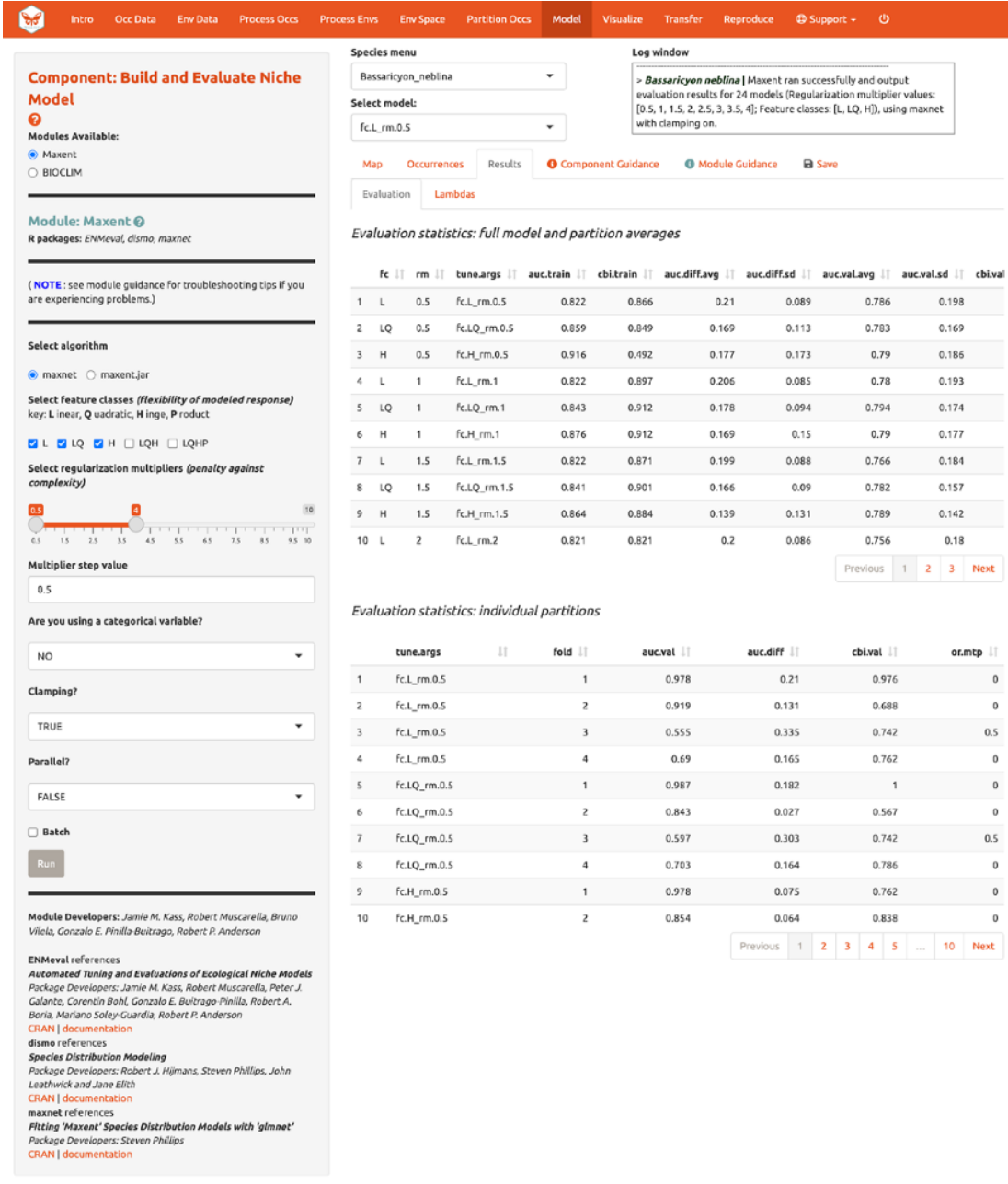
The 3 feature class combinations (L, LQ, H) x 8 regularization multipliers (0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4) = 24 candidate models. The hinge feature class (H) will enable substantial complexity in the response, so it takes a bit longer to run than the simpler models.



The screenshot displays the Wallace EXERCISE web interface for building and evaluating a niche model for *Bassaricyon neblina*. The interface is divided into several sections:

- Component: Build and Evaluate Niche Model**: This section contains configuration options for the model.
  - Modules Available**: Maxent (selected) and BIOCLIM.
  - Module: Maxent**: R packages: ENMeval, dismo, maxnet.
  - Select algorithm**: maxnet (selected) and maxent.jar.
  - Select feature classes**: L (checked), LQ (checked), H (checked), LQH (unchecked), LQHP (unchecked).
  - Select regularization multipliers**: A slider is set to 0.5.
  - Multiplier step value**: 0.5.
  - Are you using a categorical variable?**: NO.
  - Clamping?**: TRUE.
  - Parallel?**: FALSE.
  - Batch**: Unchecked.
  - Run**: A button to execute the model.
- Species menu**: A dropdown menu showing *Bassaricyon neblina*.
- Log window**: A text area displaying "\*\*\*WELCOME TO WALLACE\*\*\*" and a prompt to find messages for the user.
- Map**: A map of South America showing the distribution of *Bassaricyon neblina* with red dots. The map includes a legend for Partition Groups (1, 2, 3, 4) and a scale bar.
- Module Developers**: A list of developers for ENMeval, dismo, and maxnet.
- ENMeval references**: A link to the ENMeval documentation.
- dismo references**: A link to the dismo documentation.
- maxnet references**: A link to the maxnet documentation.

The results appear in two tables of evaluation statistics, allowing comparison of the different models you just built. The first table shows the statistics for the full model and partition averages. There should be 24 rows: one for each of the feature class / regularization multiplier combinations. In the first table, statistics from the models built from the 4 occurrence data partition groups (one withheld for each iteration) are averaged. In the second table, the partition group statistics averaged in the first table are shown, and thus it contains 96 rows (each of the 4 folds for each of the 24 models).



**Component: Build and Evaluate Niche Model**

Modules Available:  
 Maxent  
 BIOCLIM

Module: Maxent  
 R packages: ENMeval, dismo, maxnet

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

Select algorithm  
 maxnet  maxent.jar

Select feature classes (flexibility of modeled response)  
 key: L linear, Q quadratic, H inge, P product  
 L  Q  H  LQH  LQHP

Select regularization multipliers (penalty against complexity)  
 Multiplier step value: 0.5

Are you using a categorical variable?  
 NO  YES

Clamping?  
 TRUE  FALSE

Parallel?  
 TRUE  FALSE

Batch

Module Developers: Jamie M. Kass, Robert Muscarella, Bruno Vilela, Gonzalo E. Pinilla-Buitrago, Robert P. Anderson

ENMeval references  
 Automated Tuning and Evaluations of Ecological Niche Models  
 Package Developers: Jamie M. Kass, Robert Muscarella, Peter J. Galante, Corentin Bohl, Gonzalo E. Buitrago-Pinilla, Robert A. Boria, Mariano Soley-Guardia, Robert P. Anderson  
[CRAN | documentation](#)

dismo references  
 Species Distribution Modeling  
 Package Developers: Robert J. Hijmans, Steven Phillips, John Leathwick and Jane Elith  
[CRAN | documentation](#)

maxnet references  
 Fitting Maxent Species Distribution Models with 'jmaxnet'  
 Package Developers: Steven Phillips  
[CRAN | documentation](#)

Species menu: Bassaricyon neblina  
 Select model: fc.L\_rm.0.5

Log window  
 > Bassaricyon neblina | Maxent ran successfully and output evaluation results for 24 models (Regularization multiplier values: [0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4]; Feature classes: [L, Q, H]), using maxent with clamping on.

Map Occurrences Results Component Guidance Module Guidance Save

Evaluation Lambdas

Evaluation statistics: full model and partition averages

	fc	rm	tune.args	auc.train	cbItrain	auc.diff.avg	auc.diff.sd	auc.val.avg	auc.val.sd	cbIval
1	L	0.5	fc.L_rm.0.5	0.822	0.866	0.21	0.089	0.786	0.198	
2	LQ	0.5	fc.LQ_rm.0.5	0.859	0.849	0.169	0.113	0.783	0.169	
3	H	0.5	fc.H_rm.0.5	0.916	0.492	0.177	0.173	0.79	0.186	
4	L	1	fc.L_rm.1	0.822	0.897	0.206	0.085	0.78	0.193	
5	LQ	1	fc.LQ_rm.1	0.843	0.912	0.178	0.094	0.794	0.174	
6	H	1	fc.H_rm.1	0.876	0.912	0.169	0.15	0.79	0.177	
7	L	1.5	fc.L_rm.1.5	0.822	0.871	0.199	0.088	0.766	0.184	
8	LQ	1.5	fc.LQ_rm.1.5	0.841	0.901	0.166	0.09	0.782	0.157	
9	H	1.5	fc.H_rm.1.5	0.864	0.884	0.139	0.131	0.789	0.142	
10	L	2	fc.L_rm.2	0.821	0.821	0.2	0.086	0.756	0.18	

Previous 1 2 3 Next

Evaluation statistics: individual partitions

	tune.args	fold	auc.val	auc.diff	cbIval	or.mtp
1	fc.L_rm.0.5	1	0.978	0.21	0.976	0
2	fc.L_rm.0.5	2	0.919	0.131	0.688	0
3	fc.L_rm.0.5	3	0.555	0.335	0.742	0.5
4	fc.L_rm.0.5	4	0.69	0.165	0.762	0
5	fc.LQ_rm.0.5	1	0.987	0.182	1	0
6	fc.LQ_rm.0.5	2	0.843	0.027	0.567	0
7	fc.LQ_rm.0.5	3	0.597	0.303	0.742	0.5
8	fc.LQ_rm.0.5	4	0.703	0.164	0.786	0
9	fc.H_rm.0.5	1	0.978	0.075	0.762	0
10	fc.H_rm.0.5	2	0.854	0.064	0.838	0

Previous 1 2 3 4 5 ... 10 Next

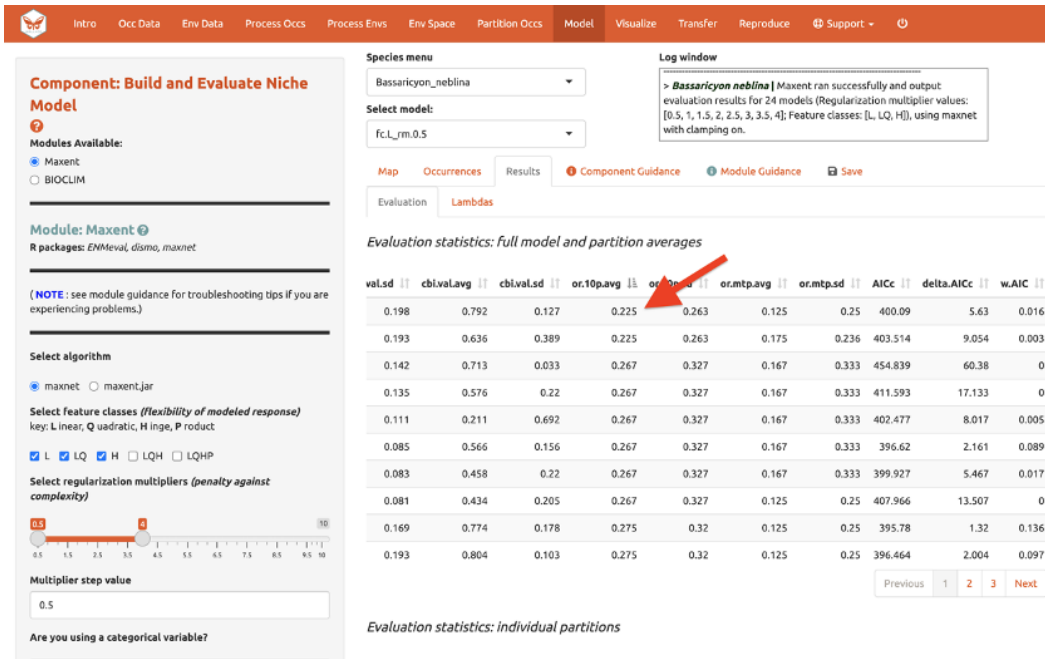
How do we choose the “best” model?

There is a mountain of literature about this, and there is really no single answer for all datasets. The model performance statistics AUC (Area Under the Curve), OR (Omission Rate), and CBI (Continuous Boyce Index) were calculated and averaged across our partitions, and AICc (corrected Akaike information criterion) was instead calculated using the model prediction of the full background extent (and all of the thinned occurrence points). Although AICc does not incorporate the cross-validation



results, it does explicitly penalize model complexity—hence, models with more parameters tend to have a worse AICc score. It's really up to the user to decide, and the guidance text has some references which should help you learn more.

The evaluation metrics table can be sorted. First, we will prioritize models that omitted few occurrence points in the predicted area during cross-validation. Sort the results table in ascending order by "or.10p.avg", or the average omission rate when applying a 10-percentile training presence threshold to the (withheld) validation data (see guidance text for details). As we would prefer a model that does not omit many withheld occurrences when it makes a range prediction, we are prioritizing low values of "or.10p.avg". Note: your values may be different than shown here.



Species menu: Bassaricyon neblina

Select model: fc\_L\_m.0.5

Log window: > Bassaricyon neblina | Maxent ran successfully and output evaluation results for 24 models (Regularization multiplier values: [0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4]; Feature classes: [L, LQ, H]), using maxent with clamping on.

Component: Build and Evaluate Niche Model

Modules Available: Maxent (selected), BIOCLIM

Module: Maxent

R packages: ENMeval, dismo, maxnet

Select algorithm: maxnet (selected), maxent.jar

Select feature classes (flexibility of modeled response): L, LQ, H, LQH, LQHP (all selected)

Select regularization multipliers (penalty against complexity): 0.5 (selected)

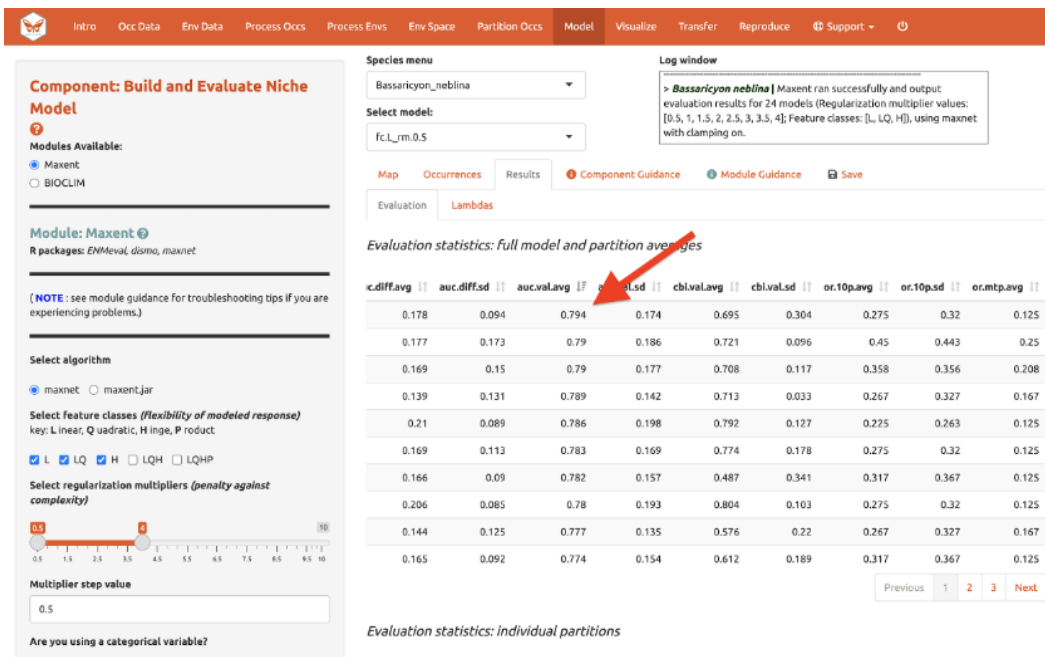
Are you using a categorical variable?

Evaluation statistics: full model and partition averages

val.sds	cbivalavg	cbival.sds	or.10p.avg	or.10p.sds	or.mtp.avg	or.mtp.sds	AICc	delta.AICc	w.AICc
0.198	0.792	0.127	0.225	0.263	0.125	0.25	400.09	5.63	0.016
0.193	0.636	0.389	0.225	0.263	0.175	0.236	403.514	9.054	0.003
0.142	0.713	0.033	0.267	0.327	0.167	0.333	454.839	60.38	0
0.135	0.576	0.22	0.267	0.327	0.167	0.333	411.593	17.133	0
0.111	0.211	0.692	0.267	0.327	0.167	0.333	402.477	8.017	0.005
0.085	0.566	0.156	0.267	0.327	0.167	0.333	396.62	2.161	0.089
0.083	0.458	0.22	0.267	0.327	0.167	0.333	399.927	5.467	0.017
0.081	0.434	0.205	0.267	0.327	0.125	0.25	407.966	13.507	0
0.169	0.774	0.178	0.275	0.32	0.125	0.25	395.78	1.32	0.136
0.193	0.804	0.103	0.275	0.32	0.125	0.25	396.464	2.004	0.097

Evaluation statistics: individual partitions

Let's also look at average validation AUC values (where higher values are better).



Species menu: Bassaricyon neblina

Select model: fc\_L\_m.0.5

Log window: > Bassaricyon neblina | Maxent ran successfully and output evaluation results for 24 models (Regularization multiplier values: [0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4]; Feature classes: [L, LQ, H]), using maxent with clamping on.

Component: Build and Evaluate Niche Model

Modules Available: Maxent (selected), BIOCLIM

Module: Maxent

R packages: ENMeval, dismo, maxnet

Select algorithm: maxnet (selected), maxent.jar

Select feature classes (flexibility of modeled response): L, LQ, H, LQH, LQHP (all selected)

Select regularization multipliers (penalty against complexity): 0.5 (selected)

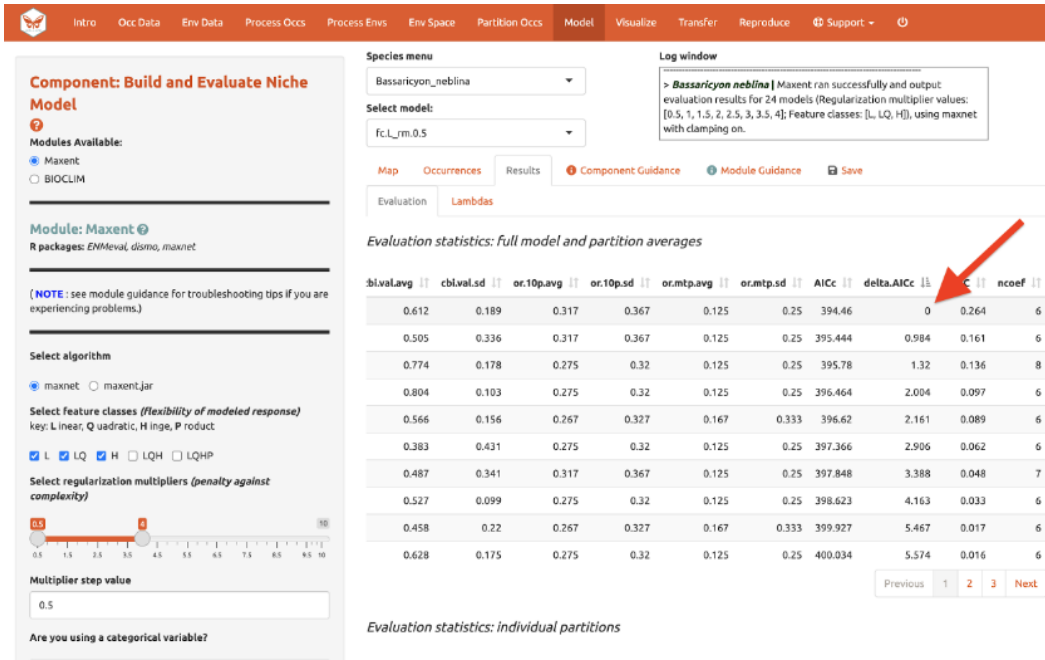
Are you using a categorical variable?

Evaluation statistics: full model and partition averages

c.diff.avg	auc.diff.sds	auc.val.avg	auc.val.sds	cbivalavg	cbival.sds	or.10p.avg	or.10p.sds	or.mtp.avg
0.178	0.094	0.794	0.174	0.695	0.304	0.275	0.32	0.125
0.177	0.173	0.79	0.186	0.721	0.096	0.45	0.443	0.25
0.169	0.15	0.79	0.177	0.708	0.117	0.358	0.356	0.208
0.139	0.131	0.789	0.142	0.713	0.033	0.267	0.327	0.167
0.21	0.089	0.786	0.198	0.792	0.127	0.225	0.263	0.125
0.169	0.113	0.783	0.169	0.774	0.178	0.275	0.32	0.125
0.166	0.09	0.782	0.157	0.487	0.341	0.317	0.367	0.125
0.206	0.085	0.78	0.193	0.804	0.103	0.275	0.32	0.125
0.144	0.125	0.777	0.135	0.576	0.22	0.267	0.327	0.167
0.165	0.092	0.774	0.154	0.612	0.189	0.317	0.367	0.125

Evaluation statistics: individual partitions

And AICc (where lower values are better)...



**Component: Build and Evaluate Niche Model**

Species menu: Bassaricyon neblina

Select model: fcLQ\_rm.0.5

Log window: > Bassaricyon neblina | Maxent ran successfully and output evaluation results for 24 models (Regularization multiplier values: [0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4]; Feature classes: [L, LQ, H]), using maxnet with clamping on.

Map Occurrences Results **Component Guidance** Module Guidance Save

Evaluation Lambdas

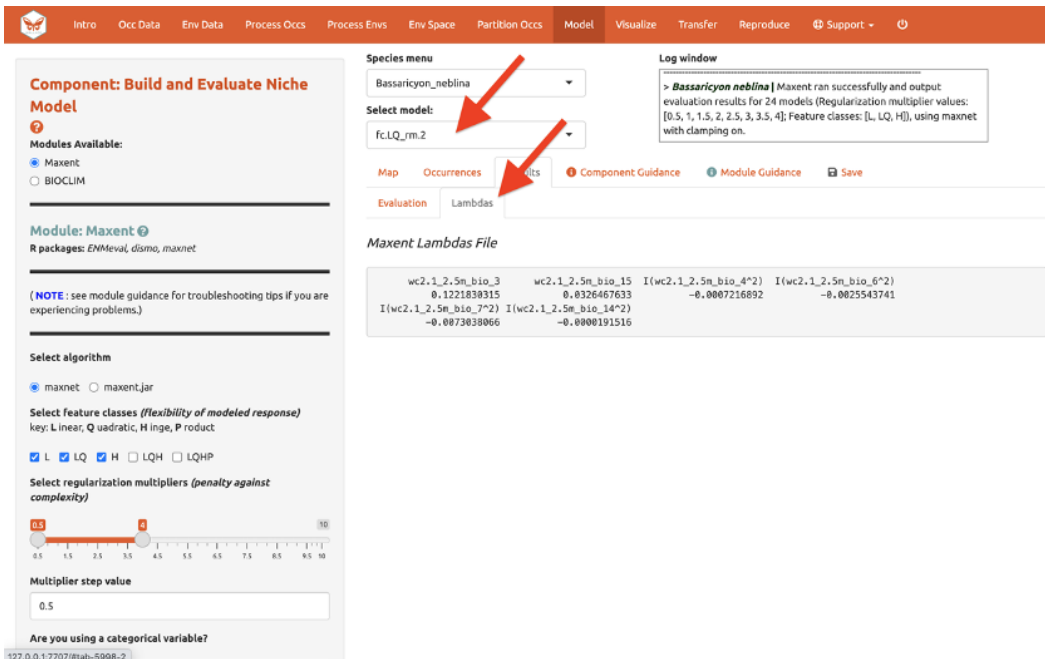
Evaluation statistics: Full model and partition averages

lambda.avg	cbi.val.sd	or.10p.avg	or.10p.sd	or.mtp.avg	or.mtp.sd	AICc	delta.AICc	lambda	lambda.C	ncoef
0.612	0.189	0.317	0.367	0.125	0.25	394.46	0	0.264	6	
0.505	0.336	0.317	0.367	0.125	0.25	395.444	0.984	0.161	6	
0.774	0.178	0.275	0.32	0.125	0.25	395.78	1.32	0.136	8	
0.804	0.103	0.275	0.32	0.125	0.25	396.464	2.004	0.097	6	
0.566	0.156	0.267	0.327	0.167	0.333	396.62	2.161	0.089	6	
0.383	0.431	0.275	0.32	0.125	0.25	397.366	2.906	0.062	6	
0.487	0.341	0.317	0.367	0.125	0.25	397.848	3.388	0.048	7	
0.527	0.099	0.275	0.32	0.125	0.25	398.623	4.163	0.033	6	
0.458	0.22	0.267	0.327	0.167	0.333	399.927	5.467	0.017	6	
0.628	0.175	0.275	0.32	0.125	0.25	400.034	5.574	0.016	6	

Evaluation statistics: Individual partitions

In our example, if we had chosen the model with the lowest AICc score, we would have ended up with LQ\_2. Note: your values may be different.

Next to the “Evaluation” results, you can access the Maxent “Lambdas” file (which describe the weights for feature classes of each variable) for each of the models (changing the candidate model in the drop-down box changes the output).



**Component: Build and Evaluate Niche Model**

Species menu: Bassaricyon neblina

Select model: fcLQ\_rm.2

Log window: > Bassaricyon neblina | Maxent ran successfully and output evaluation results for 24 models (Regularization multiplier values: [0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4]; Feature classes: [L, LQ, H]), using maxnet with clamping on.

Map Occurrences Results **Component Guidance** Module Guidance Save

Evaluation Lambdas

Maxent Lambdas File

```

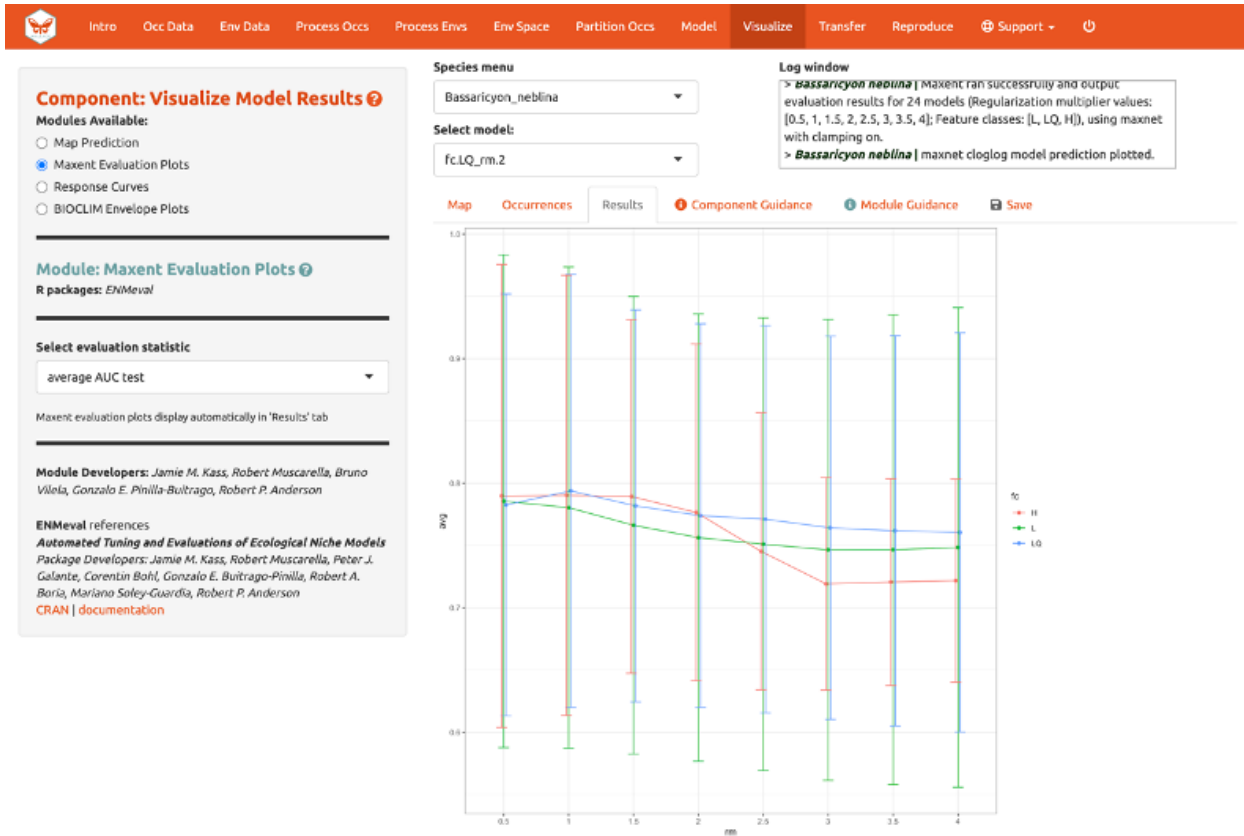
wc2-1_2-5m_bio_3    wc2-1_2-5m_bio_15    I(wc2-1_2-5m_bio_4^2)    I(wc2-1_2-5m_bio_6^2)
0.1221838315      0.8326467633        -0.0007216892          -0.0025543741
I(wc2-1_2-5m_bio_7^2)    I(wc2-1_2-5m_bio_14^2)
-0.0073038066      -0.0000191516
    
```

Use the “Save” tab to download the evaluation tables.

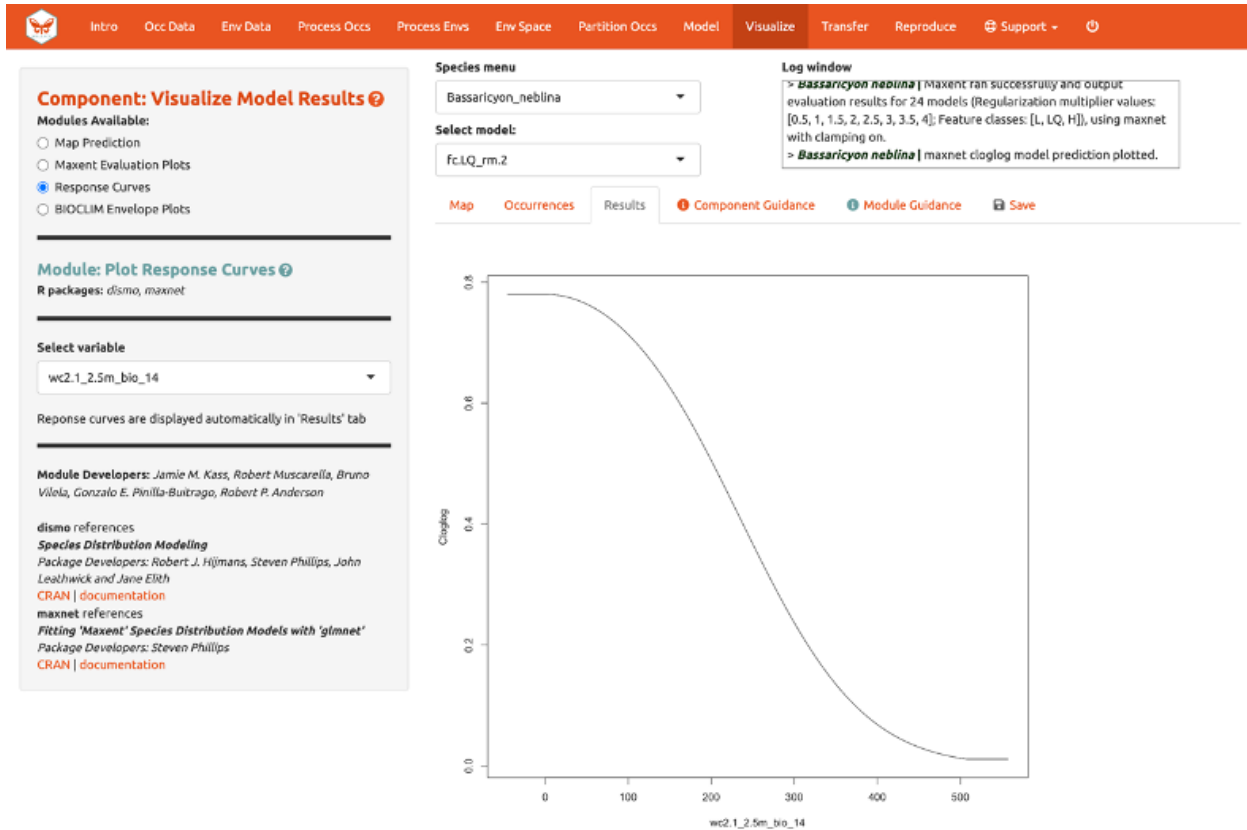
## 11. VISUALIZE

There are four modules for Visualization. We'll save the first, "Map Prediction", for last. We'll skip the fourth module, "BIOCLIM Envelope Plot", since we used Maxent instead of BIOCLIM.

The module "Maxent Evaluation Plots", enables users to evaluate the performance statistics across models. Graphs appear in the Results tab. Below, see how the feature class and regularization multiplier selections affect average validation AUC values.

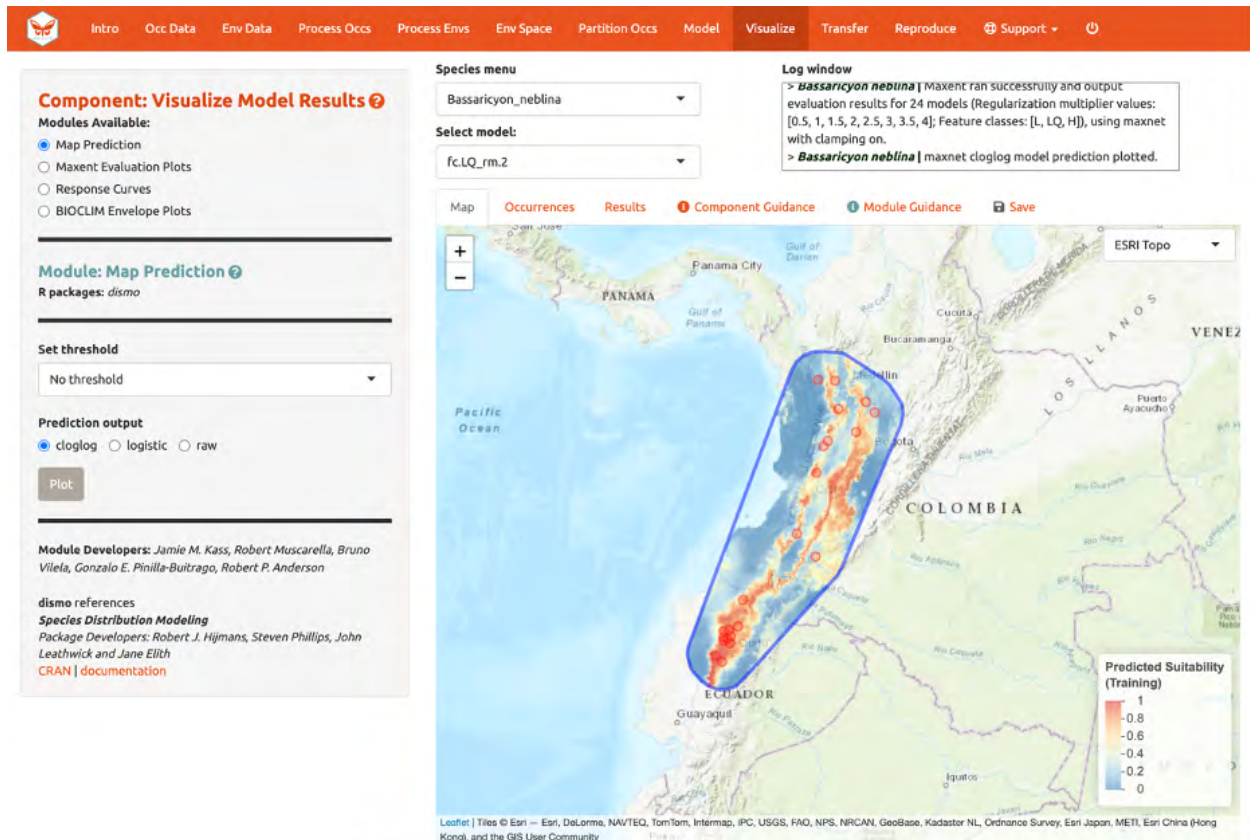


We should also examine the “Response Curves,” which show how the predicted suitability (y-axis) changes based on different values of each variable (x-axis). For these curves, the marginal response of one variable is shown while the other variables are held at their mean values. If you want to see the results for a particular model, select it by using the dropdown menu below the species box. Below is a response curve for model LQ\_2 for the mean precipitation of the driest month (bio14).

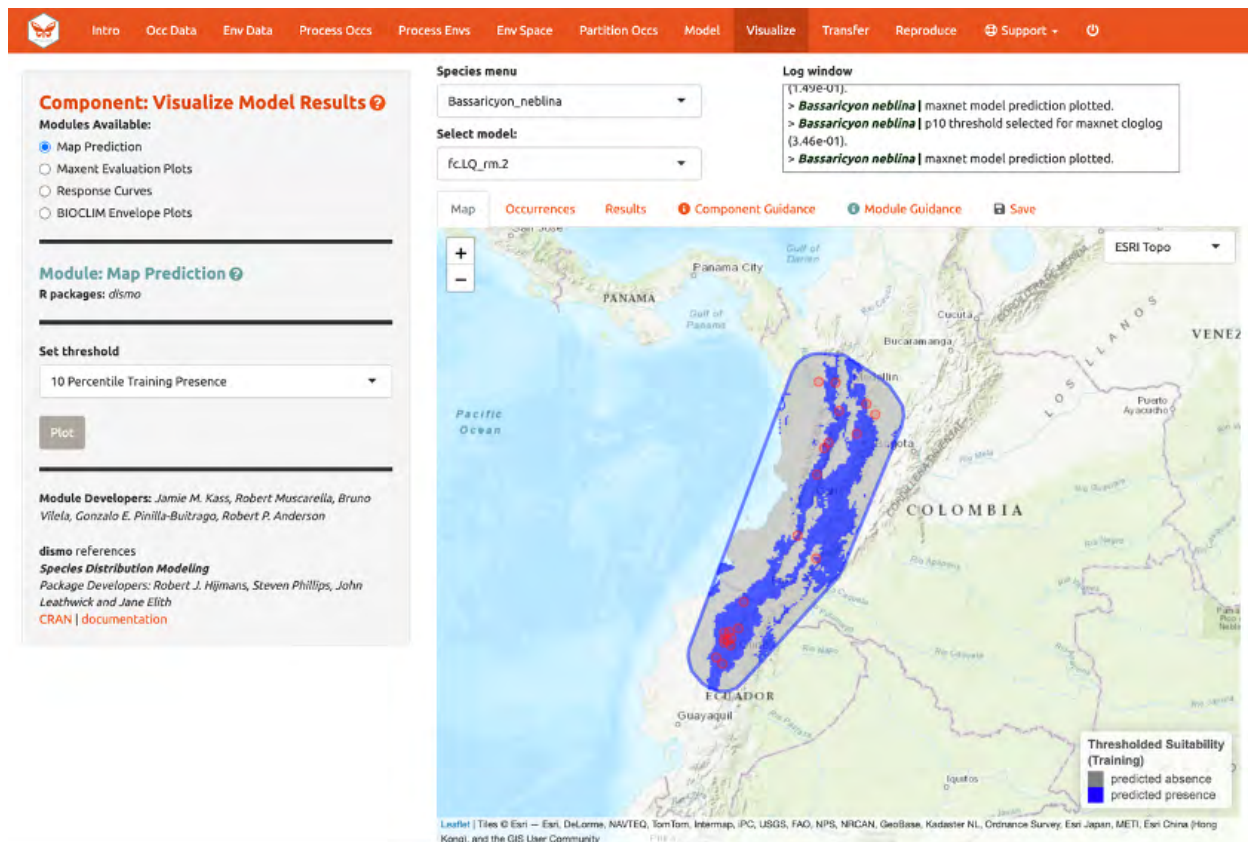




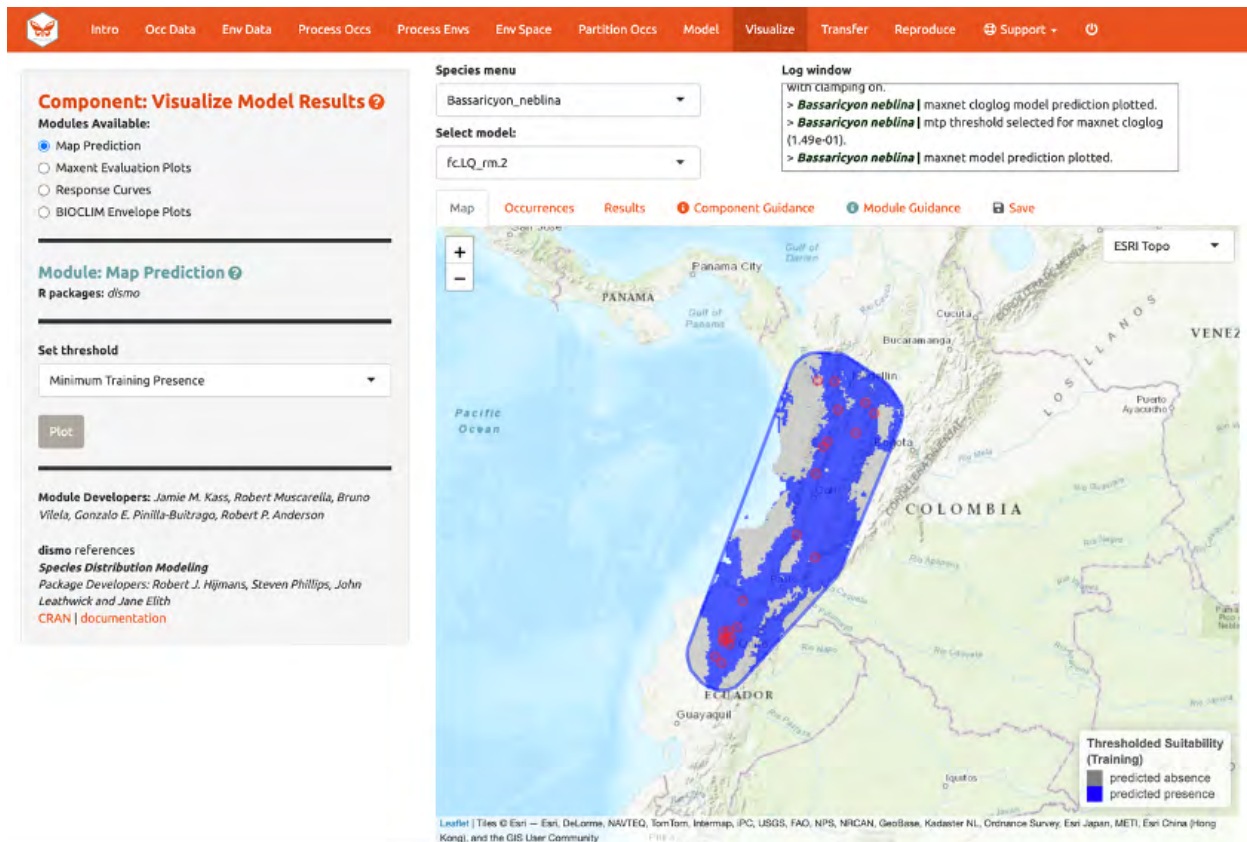
Of course, you can also visualize model predictions on the map. Predictions of suitability can be continuous (a range of values from low to high) or binary (thresholded to two values: 0, unsuitable and 1, suitable). We are visualizing predictions made with the “cloglog” transformation, which converts the raw Maxent output (relative occurrence rate) to a probabilistic scale between 0 and 1 to approximate probability of presence (given key assumptions). Please see the module guidance for information about Maxent model output scalings and thresholding rules. Here is the mapped prediction for model LQ\_2, no threshold, in cloglog output.



Below is the mapped prediction of the same model, this time with the threshold set to the 10-percentile training presence value (the occurrence suitability value we used to calculate omission rates above to help us select models). Some of the occurrence points will fall outside the blue regions that represent suitable areas for *Bassaricyon neblina*. For the 10-percentile training presence value, as it represents not the lowest predicted suitability, but the value greater than the 10% lowest, the expected omission would be 0.1 (i.e., 10% omitted).



Try mapping the prediction with the threshold set to the less strict “Minimum Training Presence” and notice the difference. You can also threshold by a quantile of training presences that are omitted. Try setting the quantile to different values and notice the change in prediction.



You may have noticed the batch option is not available for this component. Users need to select optimal models relative to each species, therefore predictions can only be mapped individually. You can download your Maxent or BIOCLIM evaluation plots, response curves, and map predictions from the “Save” tab. Note that this will download the current plot. For instance, if you wanted to download the continuous prediction, you’ll have to plot again, since we last plotted the threshold map.

## 12. MODEL TRANSFER

Next, you can transfer the model to new locations and past/future climate scenarios. “Transferring” simply means making predictions with the selected model using new environmental values (i.e., those not used for model building) and getting suitability predictions for new variable ranges. Note: This can also be referred to as “projecting” a model, but do not confuse this with the GIS term typically used for changing the coordinate reference system of a map.

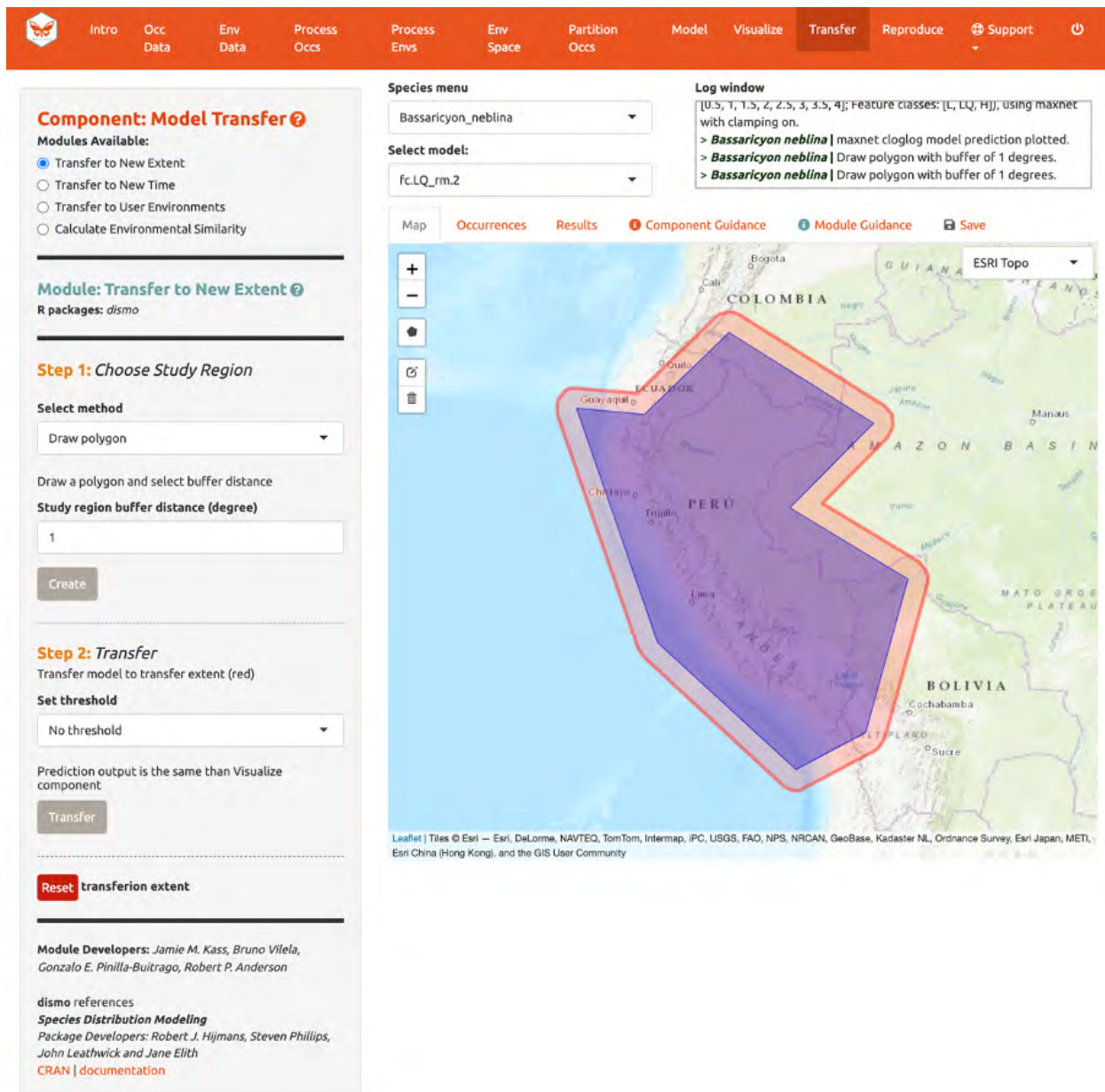
This is potentially confusing because the cross-validation step we used also transferred to new conditions. The spatial cross-validation step iteratively forced models to predict to new areas (and thus likely new environments), and the evaluation statistics summarized the ability of the particular model settings to result in models that transfer accurately. However, the final model that we used to make the predictions we are visualizing was built with all the data (it did not exclude any partition groups or the geographic areas they correspond to). So the variable ranges associated with all of the background points in our dataset were used in the model-building process.

We are now taking this model and transferring it to variable ranges that might not have been used in



model-building (i.e., not represented in the training data). Thus, these environmental values for different places and times could be completely new to our model, even potentially so different that we may be uncertain in the accuracy of our prediction. This is because although the modeled variable responses remain the same, predictions for variable values more extreme than the training data can result in unexpected suitability predictions. For this reason, clamping is often used to constrain model transfers (see below). Please see the guidance text for more orientation regarding these “non-analog conditions”.

Let's begin with “Transfer to New Extent” and see if Peru has suitable areas for the olinguito. In Step 1, use the polygon drawing tool to draw around Peru with a 1-degree buffer and hit “Create”. Alternatively, you can upload a shapefile or .csv file with records for vertices with fields “longitude, latitude” to use as a study region.



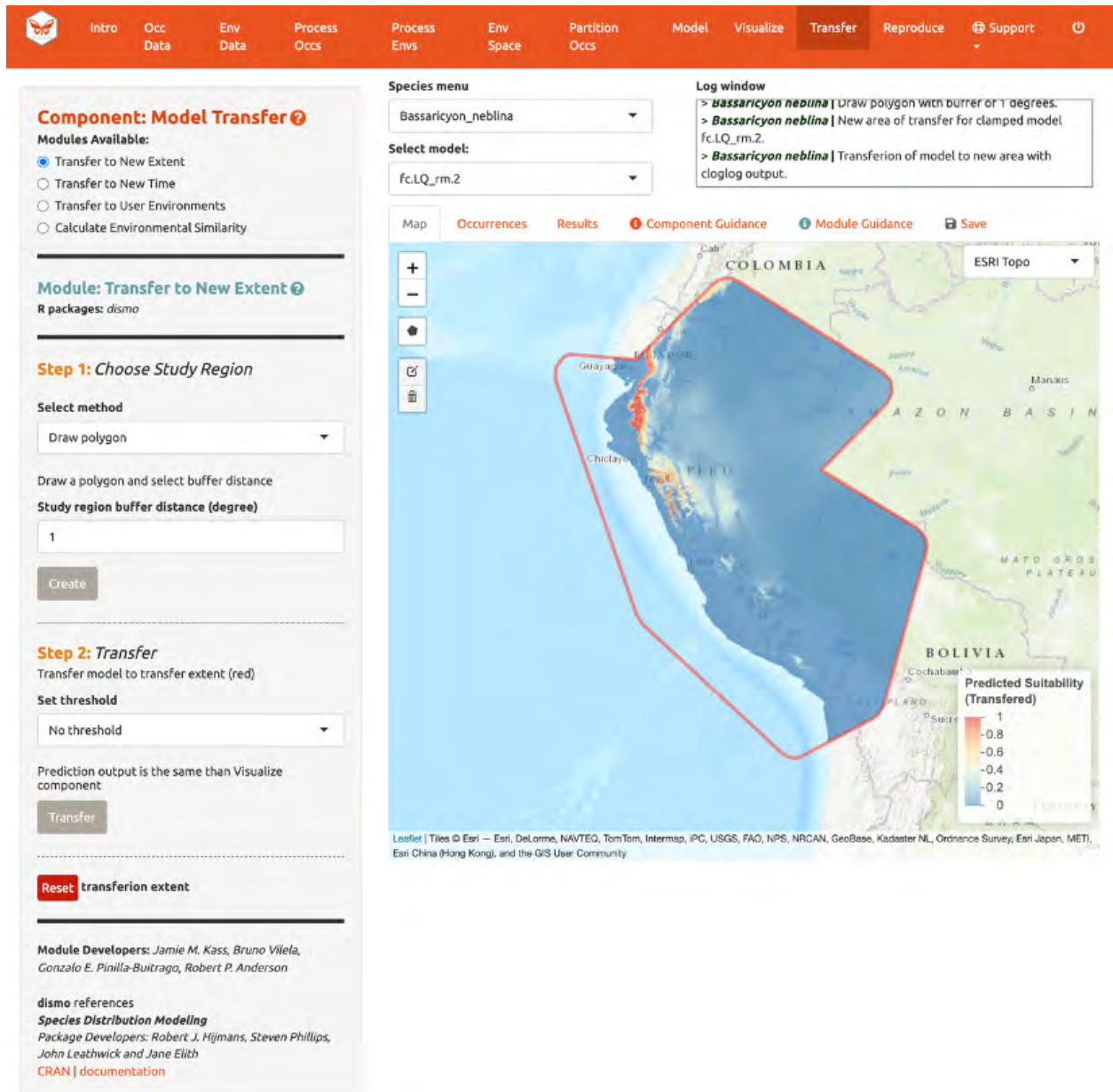
The screenshot displays the MaxNet web interface for the 'Transfer to New Extent' module. The top navigation bar includes options like 'Intro', 'Occ Data', 'Env Data', 'Process Occs', 'Process Envs', 'Env Space', 'Partition Occs', 'Model', 'Visualize', 'Transfer', 'Reproduce', and 'Support'. The main interface is divided into several sections:

- Component: Model Transfer**: Shows 'Modules Available' with radio buttons for 'Transfer to New Extent' (selected), 'Transfer to New Time', 'Transfer to User Environments', and 'Calculate Environmental Similarity'.
- Module: Transfer to New Extent**: Lists 'R packages: dismo'.
- Step 1: Choose Study Region**: Includes a 'Select method' dropdown set to 'Draw polygon', a 'Study region buffer distance (degree)' input field set to '1', and a 'Create' button.
- Step 2: Transfer**: Includes a 'Set threshold' dropdown set to 'No threshold' and a 'Transfer' button.
- Species menu**: Shows 'Bassaricyon neblina' selected.
- Select model**: Shows 'fc.LQ\_rm.2' selected.
- Log window**: Displays the following text:
 

```
[U.S, T, 1.5, Z, 2.5, 3, 3.5, 4]; Feature classes: [L, LQ, H]; using maxnet with clamping on.
> Bassaricyon neblina | maxnet cloglog model prediction plotted.
> Bassaricyon neblina | Draw polygon with buffer of 1 degrees.
> Bassaricyon neblina | Draw polygon with buffer of 1 degrees.
```
- Map**: Shows a map of South America with a purple polygon representing the study region around Peru, outlined in red. The map includes labels for 'PERU', 'COLOMBIA', 'GUAYANA FRANCESA', 'GUAYANA', 'AMAZON BASIN', 'MARIKUPA', 'MATO GROSSO PLATEAU', and 'BOLIVIA'.

In Step 2, choose a threshold to make a binary prediction or “No Threshold” for a continuous one and click “Transfer”. Here, we see very low suitability for most of Peru for the olinguito.





**Component: Model Transfer**

**Modules Available:**

- Transfer to New Extent
- Transfer to New Time
- Transfer to User Environments
- Calculate Environmental Similarity

**Module: Transfer to New Extent**

R packages: *dismo*

**Step 1: Choose Study Region**

Select method

Draw polygon

Draw a polygon and select buffer distance

Study region buffer distance (degree)

1

Create

**Step 2: Transfer**

Transfer model to transfer extent (red)

Set threshold

No threshold

Prediction output is the same than Visualize component

Transfer

**Reset** transfer extent

**Module Developers:** Jamie M. Kass, Bruno Vilela, Gonzalo E. Pinilla-Buitrago, Robert P. Anderson

**dismo references**

**Species Distribution Modeling**

Package Developers: Robert J. Hijmans, Steven Phillips, John Leathwick and Jane Elith

[CRAN | documentation](#)

**Species menu**

Bassaricyon\_neblina

**Select model:**

fc.LQ\_rm.2

**Log window**

```
> Bassaricyon neblina | Draw polygon with buffer of 1 degrees.
> Bassaricyon neblina | New area of transfer for clamped model fc.LQ_rm.2.
> Bassaricyon neblina | Transfer of model to new area with cloglog output.
```

Map Occurrences Results Component Guidance Module Guidance Save

ESRI Topo

Predicted Suitability (Transferred)

1  
-0.8  
-0.6  
-0.4  
-0.2  
0

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

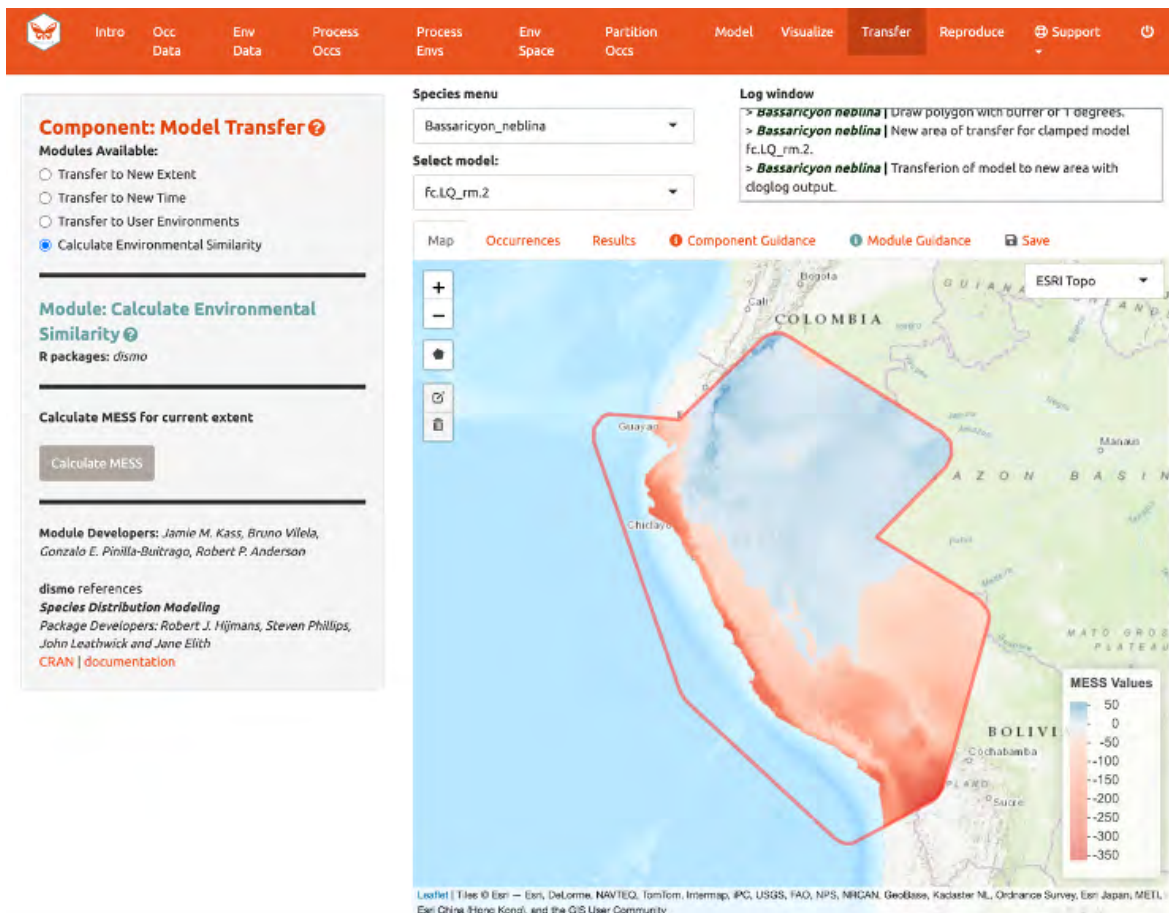
Note: to remove the outline of the polygon from the prediction, click the Trashcan icon and “Clear all”.

If you initially used WorldClim or ecoClimate as environmental variables, you can use “Transfer to New Time”. In Step 1, there are three options to choose a study region; to draw a polygon, use the same extent, or upload a polygon. In Step 2, you have the choice of WorldClim or Ecoclimate for source variables. The choice depends on your initial selection of environmental variables in “Component: Env Data”. For WorldClim, select a time period, a global circulation model, a representative concentration pathway (RCP), and a threshold. Notice also that there are several global circulation models (GCMs) to choose from—these all represent different efforts to model future climate. Not all GCMs have raster data for each RCP. See the module guidance text for more on RCPs and GCMs. Note: some databases have phased out RCPs for Shared Socioeconomic Pathways (SSPs), so be advised that some literature might use SSP terminology instead of RCP. For ecoClimate, you can select a Atmospheric Oceanic General Circulation Model (AOGCM), temporal scenario, and threshold.

The third module, “Transfer to User Environments”, gives users the option to project their model to their own uploaded environmental data. The first step is the same as before (select the study region), but in the second step users can upload single-format rasters (.tif, .asc) to use as new data for model projection. The rasters must have the same extent and resolution (cell size), and the names of the files must correspond to the environmental variables used in modeling. To assist, there is a message “Your files must be named as: ...” indicating the correct file names to use.

We will skip the “Transfer to New Time” and “Transfer to User Environments” and move on to to “Calculate Environmental Similarity”.

When transferring a model, there may be areas within our new ranges of values that have high uncertainty because they are very different from the values used in model-building. In order to visualize where these areas are, we can use the fourth module, “Calculate Environmental Similarity”, to plot a MESS map. MESS stands for (M)ultivariate (E)nviroental (S)imilarity (S)urface, and the map shows a continuous scale of environmental difference from the training data used for model-building, where increasing positive values mean more similar (blue), and decreasing negative values mean more different (red); please see the module guidance text for details. We can see that future climate values at high elevation are more similar to our training data, whereas those at lower elevations towards the coast are very different in some places. We may therefore interpret that predicted suitability in these areas has higher uncertainty.



The screenshot displays the software interface for the "Calculate Environmental Similarity" module. The top navigation bar includes options like Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce, Support, and a power icon. The "Transfer" tab is active.

**Component: Model Transfer**  
**Modules Available:**  
 Transfer to New Extent  
 Transfer to New Time  
 Transfer to User Environments  
 Calculate Environmental Similarity

**Module: Calculate Environmental Similarity**  
 R packages: *dismo*

**Calculate MESS for current extent**

**Module Developers:** Jamie M. Kass, Bruno Vilela, Gonzalo E. Pinilla-Buitrago, Robert P. Anderson

**dismo references**  
**Species Distribution Modeling**  
 Package Developers: Robert J. Hijmans, Steven Phillips, John Leathwick and Jane Elith  
[CRAN | documentation](#)

**Species menu:** Bassaricyon\_neblina  
**Select model:** fc.LQ\_rm.2

**Log window:**  
 > *Bassaricyon neblina* | Draw polygon with buffer of 1 degrees.  
 > *Bassaricyon neblina* | New area of transfer for clamped model fc.LQ\_rm.2.  
 > *Bassaricyon neblina* | Transfer of model to new area with cloglog output.

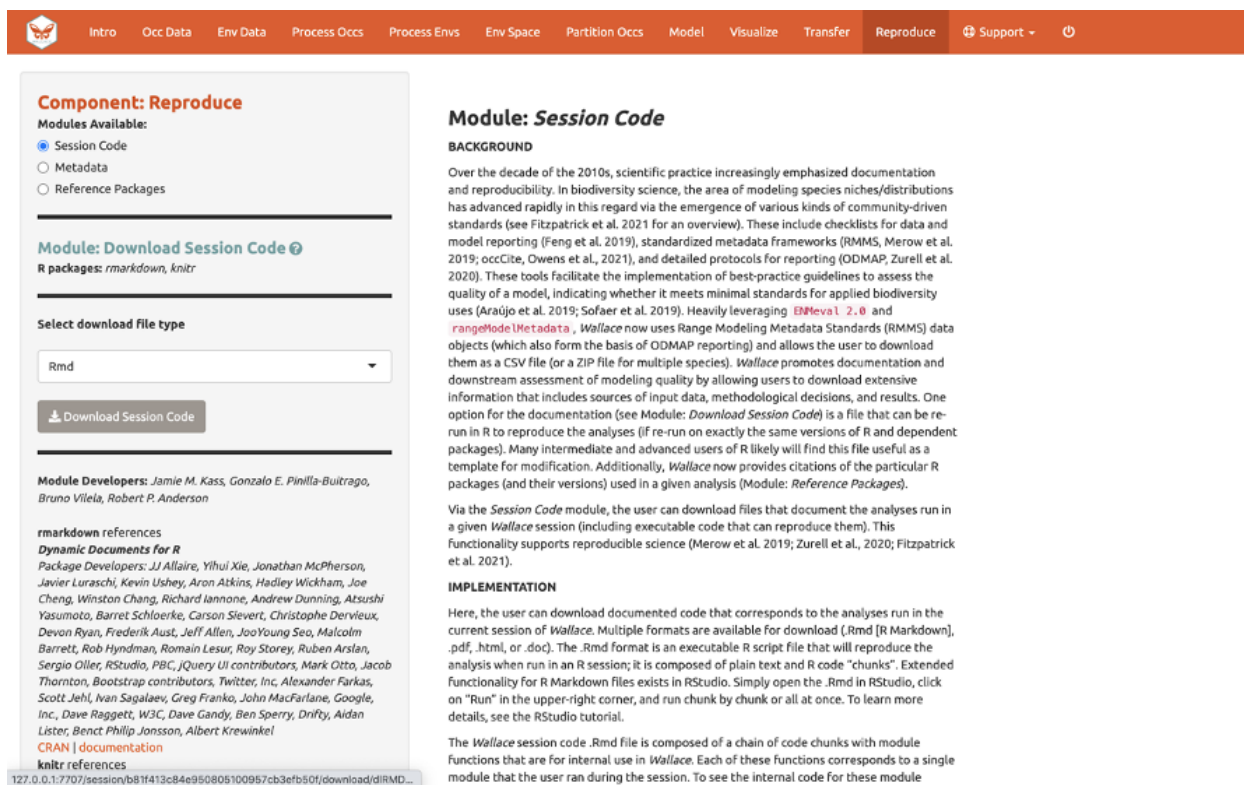
The map shows a geographical area with a red outline. A legend titled "MESS Values" indicates a scale from 50 (blue) to -350 (red). The map includes a scale bar and a "Map" button.

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

## 13. REPRODUCE

A major advantage of Wallace is reproducibility. The first option within this component is downloading code to run the analysis. While we were using Wallace, R code has been running in the background, evident from the messages printed to the R console. This option allows you to download a simplified version of this code in the form of a condensed and annotated R script. This script serves as documentation for the analysis and can be shared. It can also be run to reproduce the analysis, or edited to change aspects of it. The script can be downloaded as several file types, but the R Markdown format (.Rmd), which is a convenient format for combining R code and notation text, can be run directly in R. For .pdf downloads, the software TeX is necessary to be installed on your system. Please see the text on this page for more details.

To download the script, select "Rmd" and click "Download".



The screenshot shows the Wallace web interface. At the top is a navigation bar with tabs: Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce (selected), and Support. Below the navigation bar, the 'Component: Reproduce' section is active. It shows 'Modules Available:' with radio buttons for 'Session Code' (selected), 'Metadata', and 'Reference Packages'. Underneath, the 'Module: Download Session Code' section is visible, showing 'R packages: rmarkdown, knitr'. A 'Select download file type' dropdown menu is set to 'Rmd', and a 'Download Session Code' button is present. To the right, the 'Module: Session Code' section contains a 'BACKGROUND' section with text about scientific practice and reproducibility, an 'IMPLEMENTATION' section, and a list of 'Module Developers'. At the bottom left of the interface, a URL is partially visible: '127.0.0.1:7707/session/b81f413c84e950805100957cb3efb50f/download/dirMD...'.

Now, you should have an .Rmd file that contains your complete analysis. Modules from Wallace are indicated as headers denoted by `###`.



```

1 Wallace Session 2021-03-12
2 =====
3
4 Please find below the R code history from your *Wallace* v1.9.0 session.
5
6 You can reproduce your session results by running this R Markdown file
7 in RStudio.
8
9 Each code block is called a “chunk”, and you can run them either
10 one-by-one or all at once by choosing an option in the “Run” menu at the
11 top-right corner of the “Source” pane in RStudio.
12
13 For more detailed information see <http://rmarkdown.rstudio.com>.
14
15 ### Package installation
16
17 Wallace uses the following R packages that must be installed and loaded
18 before starting.
19
20 {r}
21 library(spocc)
22 library(spThin)
23 library(dismo)
24 library(rgeos)
25 library(ENMeval)
26 library(wallace)
27 {}
28
29 The *Wallace* session code .Rmd file is composed of a chain of module
30 functions that are internal to *Wallace*. Each of these functions
31 corresponds to a single module that the user ran during the session. To
32 see the internal code for these module functions, click on the links in
33 the .Rmd file. Users are encouraged to write custom code in the .Rmd
34 directly to modify their analysis, and even modify the module function
35 code to further customize. To see the source code for any module
36 function, just type its name into the R console and press Return.
37
38 {r}
39 # example:
40 # just type the function name and press Return to see its source code
41 # paste this code into a new script to edit it
42 occs_queryDb
43 {}
44
45 Your analyses are below.

```

You might want to open a new R window and try running some of this code. Remember that later sections of code may depend on things that were done earlier, so they may not all run if you skip ahead. Note that any “Env Space” analysis will appear at the end of the file. Also remember that if you close your Wallace session you’ll lose your progress in the web browser (but your .Rmd will be unaffected). If you use RStudio, you can open this Rmd and click “knit” to compile your workflow into a shareable html document.

You can also download the Metadata. Wallace generates and provides a variety of metadata objects that facilitate documentation and reproducibility by recording the user’s methodological decisions (e.g., parameter settings) and stores them in a Range Model Metadata Standards object. This will download as a zip and contain a .csv file for each species.



Intro Occ Data Env Data Process Occs Process Envs Env Space Partition Occs Model Visualize Transfer Reproduce Support
⌵

**Component: Reproduce**

Modules Available:

- Session Code
- Metadata
- Reference Packages

---

**Module: Download Metadata**

R packages: `rangeModelMetadata`

---

**Download metadata CSV file**

Download Metadata

---

**Module Developers:** *Jamie M. Kass, Gonzalo E. Pinilla-Buitrago, Cory Merow.*

**rangeModelMetadata references**  
*Provides Templates for Metadata Files Associated with Species Range Models*

**Package Developers:** *Cory Merow, Brian Maltner, Hannah Owens, Jamie Kass, Brian Enquist, Rob Guralnik, Damaris Zurell, Christian Koenig*

[CRAN | documentation](#)

**Module: Metadata**

**BACKGROUND**

Over the decade of the 2010s, scientific practice increasingly emphasized documentation and reproducibility. In biodiversity science, the area of modeling species niches/distributions has advanced rapidly in this regard via the emergence of various kinds of community-driven standards (see Fitzpatrick et al. 2021 for an overview). These include checklists for data and model reporting (Feng et al. 2019), standardized metadata frameworks (RMMS, Merow et al. 2019; occCite, Owens et al. 2021), and detailed protocols for reporting (ODMAP, Zurell et al. 2020). These tools facilitate the implementation of best-practice guidelines to assess the quality of a model, indicating whether it meets minimal standards for applied biodiversity uses (Araújo et al. 2019; Sofer et al. 2019). Heavily leveraging **ENMeval 2.0** and **rangeModelMetadata**, Wallace now uses Range Modeling Metadata Standards (RMMS) data objects (which also form the basis of ODMAP reporting) and allows the user to download them as a CSV file (or a ZIP file for multiple species). Wallace promotes documentation and downstream assessment of modeling quality by allowing users to download extensive information that includes sources of input data, methodological decisions, and results. One option for the documentation (see Module: Download Session Code) is a file that can be re-run in R to reproduce the analyses (if re-run on exactly the same versions of R and dependent packages). Many intermediate and advanced users of R likely will find this file useful as a template for modification. Additionally, Wallace now provides citations of the particular R packages (and their versions) used in a given analysis (Module: Reference Packages).

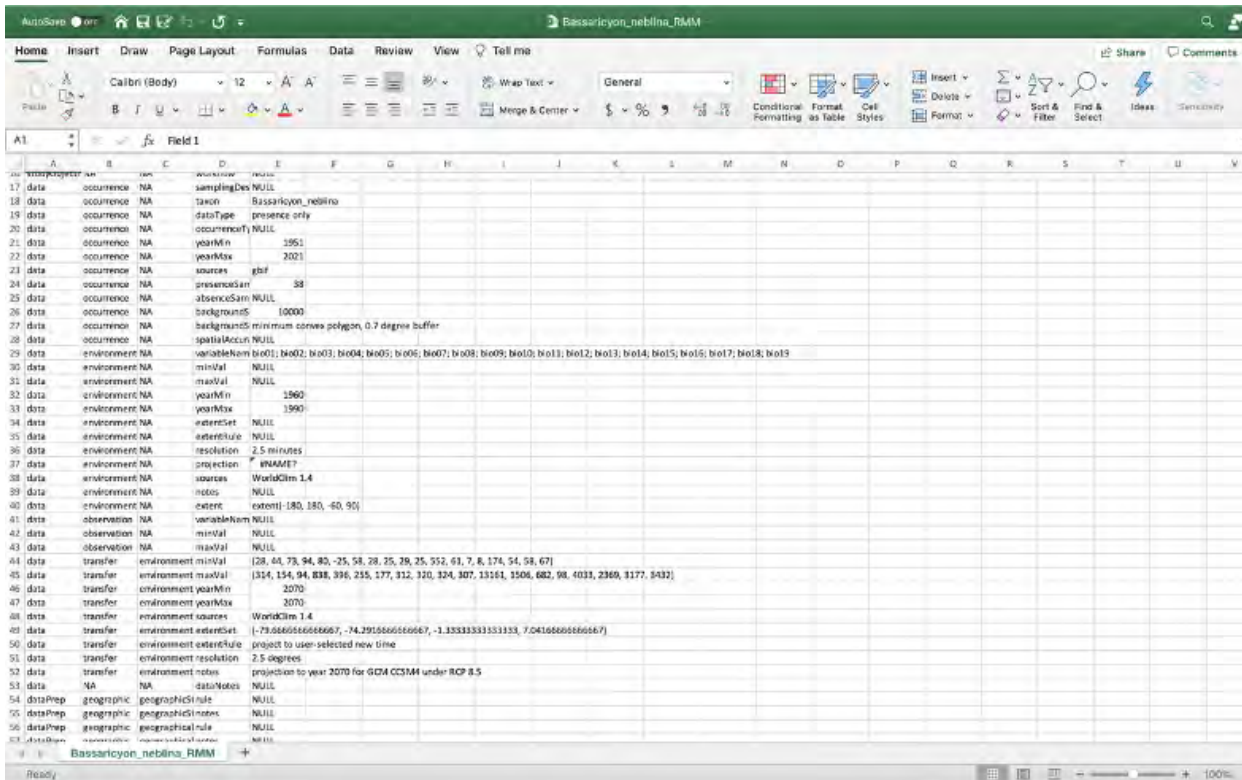
Wallace generates and provides a variety of metadata objects that facilitate documentation and reproducibility. **ENMeval 2.0.0** (Kass et al. 2021) now generates or populates standard metadata for model-building and evaluations, via the Range Model Metadata Standards (RMMS; Merow et al. 2021). Harnessing this functionality, Wallace now records the user's methodological decisions (e.g., parameter settings) and stores them in a RMMS object. This object is available for download at any stage of the analysis.

**IMPLEMENTATION**

Metadata from the Wallace analyses can be downloaded as a zip file via the **rangeModelMetadata** package (Merow et al. 2021). The RMMS includes eight fields; authorship, studyObjective, data, dataPrep (data preparation), model, prediction, assessment, and code; to catalog the modeling process. As a Wallace analysis progresses, these fields autofill to maintain data. (Merow et al. 2019).

**REFERENCES**

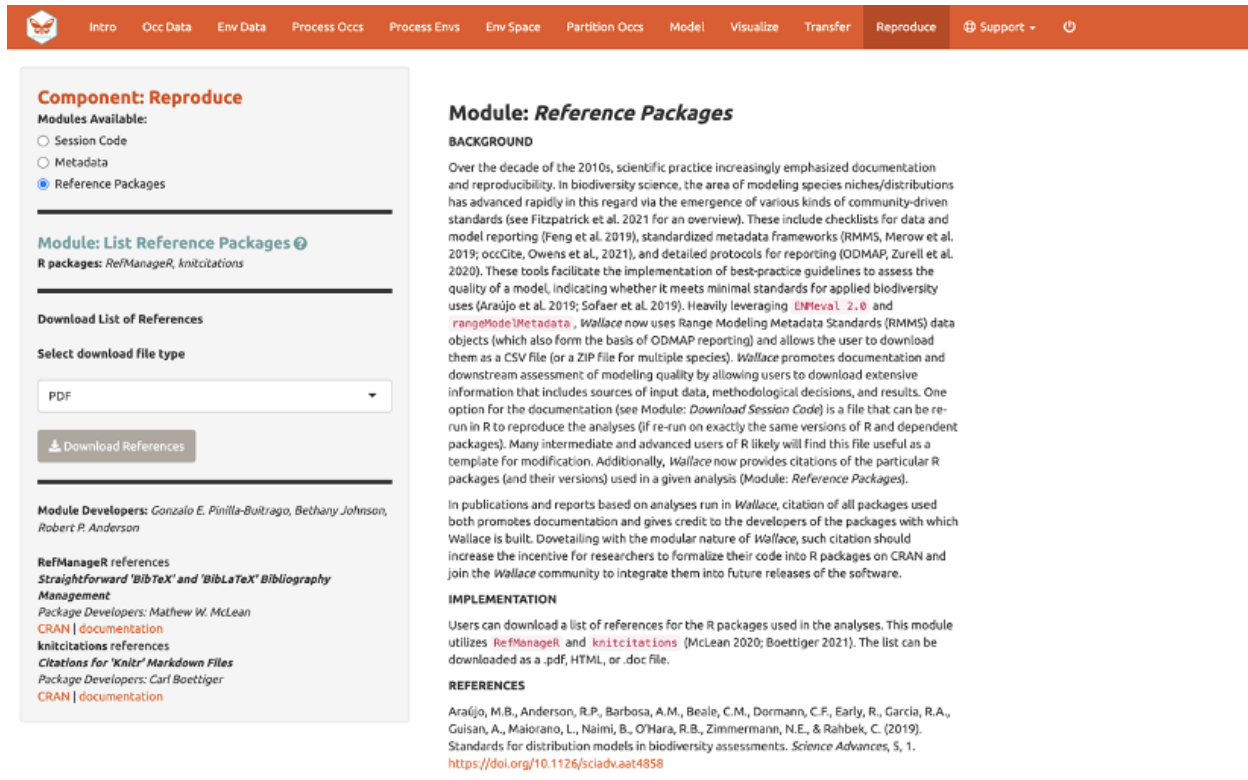
Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Mairano, L., Naimi, R., O'Hara, R.B., Zimmermann, N.E. & Rabibek, C. (2019).



Field	Value	Notes
taxon	Bassaricyon neblina	
yearMin	1951	
yearMax	2021	
sources	gbif	
presenceSam	38	
absenceSam		
background	10000	
backgroundUnits	minimum convex polygon, 0.7 degree buffer	
spatialOccur		
variableNm	bio1; bio2; bio3; bio4; bio5; bio6; bio7; bio8; bio9; bio10; bio11; bio12; bio13; bio14; bio15; bio16; bio17; bio18; bio19	
minVal		
maxVal		
yearMin	1960	
yearMax	1990	
extent		
extentUnit		
resolution	2.5 minutes	
projection	WGS84	
sources	WorldClim 1.4	
notes		
extent	extent: 180, 180, -60, 90	
minVal		
maxVal		
yearMin	2070	
yearMax	2070	
sources	WorldClim 1.4	
extent	extent: 180, 180, -60, 90	
resolution	2.5 degrees	
projection	projection to year: 2070 for GCM CCSM under RCP 8.5	
notes		
geographicNotes		
geographicRules		
geographicData		

The last module available in the “Reproduce” component is “Reference packages”. Here, you can download the citations for all the R-packages used in the analysis. To give people credit for the underlying packages that make Wallace possible (and to document your analyses properly), it is critical to cite the packages and their version number. Remember, Wallace is modular and aims to facilitate access to and use of many R packages being produced by the biogeography research community. Please promote developer credit and reproducible science by citing packages... and think about

making one of your own R packages and adding it to a future version of Wallace someday!



The screenshot shows the Wallace v2 web interface. The top navigation bar includes: Intro, Occ Data, Env Data, Process Occs, Process Envs, Env Space, Partition Occs, Model, Visualize, Transfer, Reproduce, Support, and a power icon. The main content area is divided into two columns.

**Component: Reproduce**  
 Modules Available:  
 Session Code  
 Metadata  
 Reference Packages

**Module: List Reference Packages**  
 R packages: *RefManagerR*, *knitcitations*

**Download List of References**  
 Select download file type  
 PDF  
 Download References

**Module Developers:** Gonzalo E. Pinilla-Buitrago, Bethany Johnson, Robert P. Anderson

**RefManagerR references**  
*Straightforward 'BibTeX' and 'BibLaTeX' Bibliography Management*  
 Package Developers: Mathew W. McLean  
[CRAN | documentation](#)

**knitcitations references**  
*Citations for 'Knitr' Markdown Files*  
 Package Developers: Carl Boettiger  
[CRAN | documentation](#)

**Module: Reference Packages**  
**BACKGROUND**  
 Over the decade of the 2010s, scientific practice increasingly emphasized documentation and reproducibility. In biodiversity science, the area of modeling species niches/distributions has advanced rapidly in this regard via the emergence of various kinds of community-driven standards (see Fitzpatrick et al. 2021 for an overview). These include checklists for data and model reporting (Feng et al. 2019), standardized metadata frameworks (RMMS, Merow et al. 2019; occCite, Owens et al., 2021), and detailed protocols for reporting (ODMAP, Zurell et al. 2020). These tools facilitate the implementation of best-practice guidelines to assess the quality of a model, indicating whether it meets minimal standards for applied biodiversity uses (Araújo et al. 2019; Sofaer et al. 2019). Heavily leveraging [ENMeval 2.0](#) and [rangeMod!Metadata](#), Wallace now uses Range Modeling Metadata Standards (RMMS) data objects (which also form the basis of ODMAP reporting) and allows the user to download them as a CSV file (or a ZIP file for multiple species). Wallace promotes documentation and downstream assessment of modeling quality by allowing users to download extensive information that includes sources of input data, methodological decisions, and results. One option for the documentation (see Module: *Download Session Code*) is a file that can be re-run in R to reproduce the analyses (if re-run on exactly the same versions of R and dependent packages). Many intermediate and advanced users of R likely will find this file useful as a template for modification. Additionally, Wallace now provides citations of the particular R packages (and their versions) used in a given analysis (*Module: Reference Packages*).

In publications and reports based on analyses run in Wallace, citation of all packages used both promotes documentation and gives credit to the developers of the packages with which Wallace is built. Dovetailing with the modular nature of Wallace, such citation should increase the incentive for researchers to formalize their code into R packages on CRAN and join the Wallace community to integrate them into future releases of the software.

**IMPLEMENTATION**  
 Users can download a list of references for the R packages used in the analyses. This module utilizes *RefManagerR* and *knitcitations* (McLean 2020; Boettiger 2021). The list can be downloaded as a .pdf, HTML, or .doc file.

**REFERENCES**  
 Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5, 1. <https://doi.org/10.1126/sciadv.aat4858>



The screenshot shows the Wallace v2 web interface displaying a list of references. The top navigation bar includes: REFERENCES, 1 / 1, 100%, and icons for search, refresh, and settings.

**REFERENCES**

Allaire, J.J., Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2022. *Rmarkdown: Dynamic Documents for r* (version R package 2.13). <https://github.com/rstudio/rmarkdown>.

Boettiger, Carl. 2021. *Knitcitations: Citations for 'Knitr' Markdown Files* (version R package 1.0.12). <https://CRAN.R-project.org/package=knitcitations>.

Hijmans, Robert J. 2022. *Raster: Geographic Data Analysis and Modeling* (version R package 3.5-15). <https://CRAN.R-project.org/package=raster>.

Hijmans, Robert J., Steven Phillips, John Leathwick, and June Elith. 2021. *Dismo: Species Distribution Modeling* (version R package 1.3-5). <https://CRAN.R-project.org/package=dismo>.

Kass, Jamie M., Robert Mascarella, Peter J. Gahante, Coertuis Bohl, Gonzalo E. Buitrago-Piñilla, Robert A. Boris, Mariano Soley-Guardia, and Robert P. Anderson. 2022. *ENMeval: Automated Tuning and Evaluations of Ecological Niche Models* (version R package 2.0.3).

Kass, Jamie M., Gonzalo E. Pinilla-Buitrago, Andrea Paz, Bruno Vilés, Matthew E. Aiello-Lammens, Robert Mascarella, Cory Mecon, Robert P. Anderson, and Eran Aitala. 2022. *Wallace: A Modular Platform for Reproducible Modeling of Species Niches and Distributions* (version R package 1.9.2.2). <https://wallaceEcoMod.github.io>.

Phillips, Steven. 2021. *Maxnet: Fitting 'Maxent' Species Distribution Models with 'Glmnet'* (version R package 0.1.4). <https://CRAN.R-project.org/package=maxnet>.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman: Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.

———. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r* (version R package 1.28). <https://yihui.org/knitr/>.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown/>.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>.

## CONCLUSION

Thank you for following the Wallace v2 vignette. We hope you learned more about the updated application, its features, and modeling of species distributions and niches in general. We hate to be repetitive, but we highly encourage you to read the guidance text, follow up on the recommended publications, and hopefully let them lead you to other relevant publications that can inform you further. Also, remember to discuss these topics with your peers.

We encourage you to join the Wallace Google Group: <https://groups.google.com/g/wallaceecomod>. We'd love to hear your thoughts, opinions, or suggestions on how to make Wallace better for all users. Members can post to the community and be updated on any future announcements.

If you find a bug in the software, it can be reported on the GitHub issues page: <https://github.com/wallaceEcoMod/wallace/issues>, or using the bug reporting form: <https://forms.gle/UffoBCQnMZjWYjfk9>.

We are currently working with various partners on exciting additions, so stay tuned for future versions of Wallace.

## ACKNOWLEDGMENTS

Wallace was recognized as a finalist for the 2015 Ebbe Nielsen Challenge of the Global Biodiversity Information Facility (GBIF), and received prize funding.

This material is based upon work supported by the National Science Foundation under Grant Numbers DBI-1661510 (RPA; Robert P. Anderson), DBI-1650241 (RPA), DEB-1119915 (RPA), DEB-1046328 (MEA; Matthew E. Aiello-Lammens), DBI-1401312 (RM; Robert Muscarella), and funding from the National Aeronautics and Space Administration grant 80NSSC18K0406 (MEB; Mary E. Blair). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or NASA.

## RESOURCES

- Wallace website: <https://wallaceecomod.github.io/>
- ENM2020 W19T2 Online open access Ecological Niche Modeling Course by A.T. Peterson, summary of modeling, includes Walkthrough of Wallace V1: <https://www.youtube.com/watch?v=kWNyNd2X1uo&t=1226s>
- Learn more about Olingos and the Olinguito: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3760134/>
- Gerstner et al. 2018. Revised distributional estimates for the recently discovered olinguito (*Bassaricyon neblina*), with comments on natural and taxonomic history. <https://doi.org/10.1093/jmammal/gyy012>.
- Aiello-Lammens, M.E., R.A. Boria, A. Radosavljevic, B. Vilela, and R.P. Anderson. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38(5):541–545. <https://doi.org/10.1111/ecog.01132>.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25(15):1965–1978. <https://doi.org/10.1002/joc.1276>.
- Lima-Ribeiro, M.S., S. Varela, J. González-Hernández, G. Oliveira, J.A.F. Diniz-Filho, and L.C. Terribile. 2015. ecoClimate: A database of climate data from multiple models for past, present, and future for Macroecologists and Biogeographers. *Biodiversity Informatics* 10:1–21. Available from: <https://www.ecoclimate.org/> (accessed December 2, 2023).
- Kass, J., R. Muscarella, P.J. Galante, C.L. Bohl, G.E. Pinilla-Buitrago, R.A. Boria, M. Soley-Guardia, and R.P. Anderson. 2021. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution*, 12(9):1602–1608. <https://doi.org/10.1111/2041-210X.13628>.

- Merow, C., M.J. Smith, and J.A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography* 36(10):1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.

## REFERENCES

- GBIF.org (2022). GBIF Home Page. Available from: <https://www.gbif.org> (accessed April 2022).
- Gerstner, B.E., J.M. Kass, R. Kays, K.M. Helgen, and R.P. Anderson. 2018. Revised distributional estimates for the recently discovered olinguito (*Bassaricyon neblina*), with comments on natural and taxonomic history. *Journal of Mammalogy* 99(2,3):321–332. <https://doi.org/10.1093/jmammal/gyy012>.
- Helgen, K., R. Kays, C. Pinto, and J. Schipper. 2016. *Bassaricyon alleni*. The IUCN Red List of Threatened Species 2016:e.T48637566A45215534. Available from: <https://www.iucnredlist.org/species/48637566/45215534> (accessed December 2, 2023).
- Helgen, K., R. Kays, C. Pinto, J. Schipper, and J.F. González-Maya. 2020. *Bassaricyon neblina* (amended version of 2016 assessment). The IUCN Red List of Threatened Species 2020:e.T48637280A166523067. Available from: <https://www.iucnredlist.org/species/48637280/166523067> (accessed December 2, 2023).
- Phillips, S.J., R.P. Anderson, M. Dudík, R.E. Schapire, and M.E. Blair. 2017. Opening the black box: An open-source release of Maxent. *Ecography* 40(7):887–893. <https://doi.org/10.1111/ecog.03049>.

NCEP materials are meant to be modifiable for each educator's specific classroom or training needs; adaptable Microsoft Word versions of modules are available for download through the NCEP module collection at <https://ncep.amnh.org> along with any available accompanying data files, appendices, presentations, teaching notes, and exercise solutions.