

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

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

... and Jenna Rios, Sara Varela, Hannah L. Owens, Brian S. Maintner, Olivier Broennimann, Mary E. Blair, Ned Horning, Jorge Velásquez-Tibatá, Gonzalo E. Pinilla-Buitrago, Beth E. Gerstner, Peter J. Galante, Sarah I. Meenan, Valentina Grisales-Betancur, Cecina Babich-Morrow, Erica E. Johnson ...



The City College
of New York

CUNY
The City University
of New York

AMERICAN
MUSEUM OF
NATURAL
HISTORY

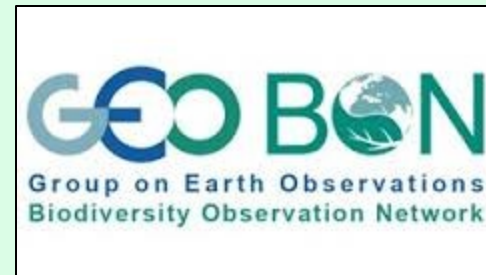
University of
Connecticut

PACE
UNIVERSITY

INSTITUTO
HUMBOLDT
COLOMBIA

and many collaborating institutions!

Range estimates needed



Applied biodiversity informatics

Anderson (2012)

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Blavatnik Awards for Young Scientists*

Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions

Robert P. Anderson^{1,2,3,4}

Agenda: Making data and modeling ready to address critical environmental issues of the 21st century

Agenda: applied biodiversity informatics

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

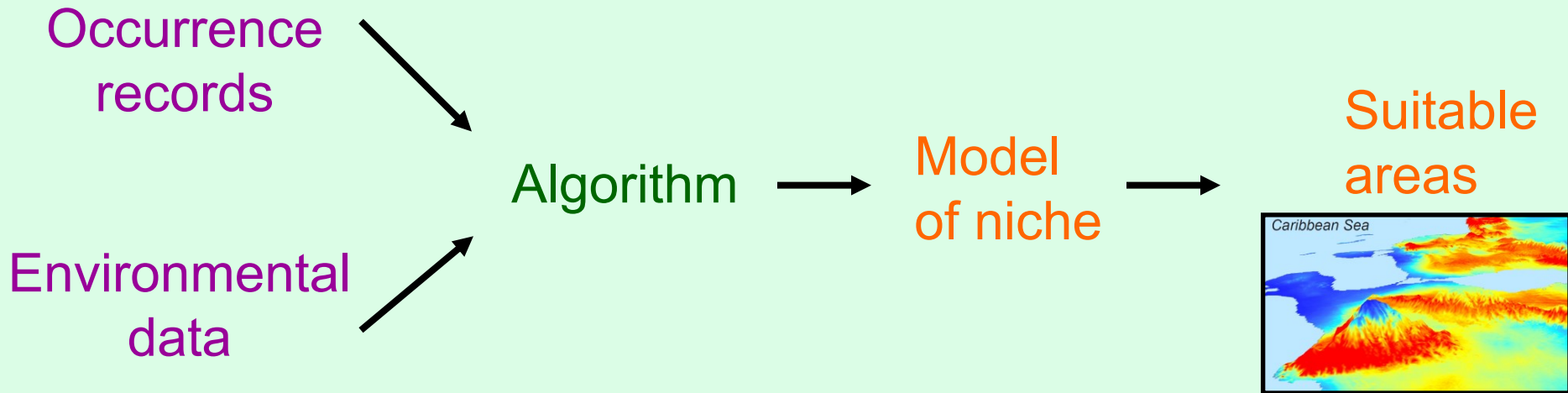
Issue: *Blavatnik Awards for Young Scientists*

Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions

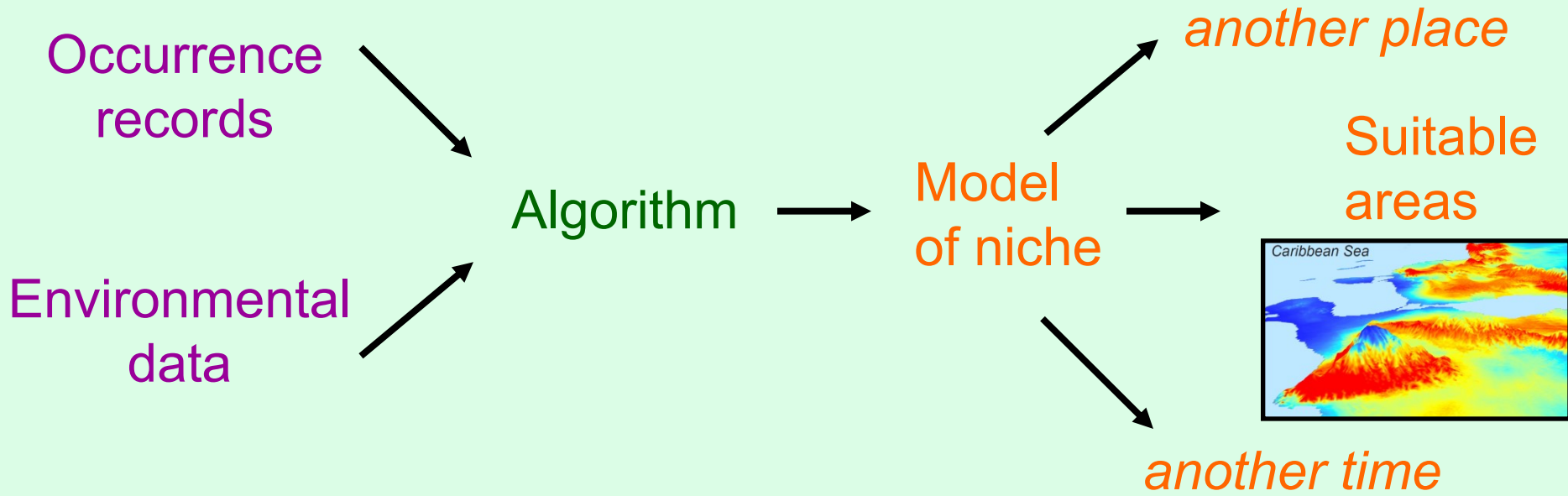
Robert P. Anderson^{1,2,3,4}

1. High-quality **data**, ready to be accessed when the particular problem presents itself
2. **Software** that achieves an appropriate balance between automation and supervision
3. **Scientists** capable of building, applying, and appraising high-quality models

Correlative modeling of species niches and ranges



Correlative modeling of species niches and ranges



Conservation, invasive species, zoonotic diseases,
climate change



Anderson (2015)

BIOGEOGRAFÍA 8

LEAD ARTICLE

**EL MODELADO DE NICHOS Y DISTRIBUCIONES:
NO ES SIMPLEMENTE “CLIC, CLIC, CLIC”**

Robert P. Anderson

BIOGEOGRAFÍA 8

**MODELING NICHES AND DISTRIBUTIONS:
IT'S NOT JUST “CLICK, CLICK, CLICK”**

Robert P. Anderson

BIOGEOGRAFÍA 8

**LA MODÉLISATION DE NICHE ET DE DISTRIBUTIONS:
CE N'EST PAS JUSTE “CLIC, CLIC, CLIC”**

Robert P. Anderson

Agenda: applied biodiversity informatics

Anderson (2012)

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Blavatnik Awards for Young Scientists*

Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions

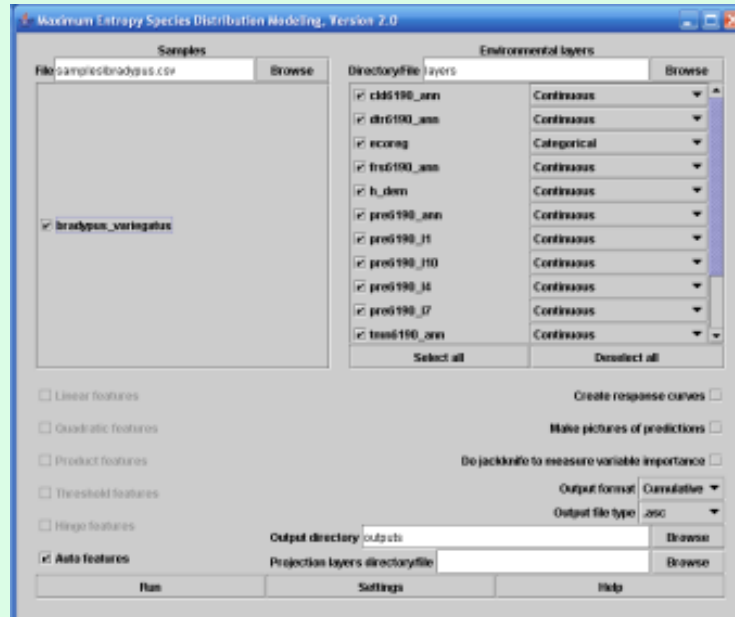
Robert P. Anderson^{1,2,3,4}

Software that achieves an appropriate balance between automation and supervision

- *automates* repetitive aspects
- *forces* user to make critical biological and conceptual decisions
- *general* with respect to algorithms used

Problems with existing analyses

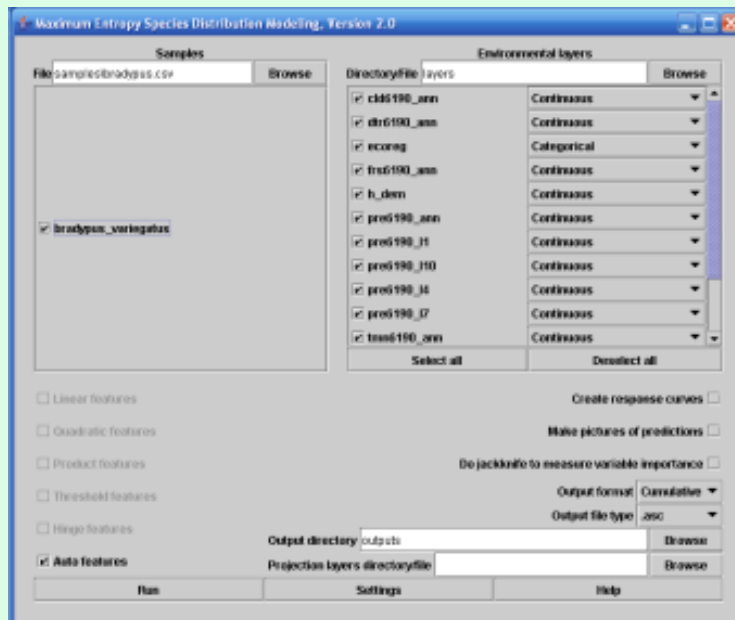
Existing GUIs



- (too) easy
- default settings
- inflexible
- infrequently updated
- strings of applications

Problems with existing analyses

Existing GUIs



- (too) easy
- default settings
- inflexible
- infrequently updated
- strings of applications

Code

```
33 results <- occ(query = "Tremarctos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
34 results.data <- results[["gbif"]]$data[["format$packageName("Tremarctos ornatus)"]]
35 occs <- remDups(results.data) # remove rows with duplicate coordinates
36 ~~~
37
38 ## Process Occurrence Data
39
40 You chose 42 of 51 total occurrence localities via polygon selection to keep in the analysis.
41
42 ~~~{r}
43 occs <- occs[c(1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, 26, 2
44 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51), ] # subset occs by selected rows
```

- error-prone
- not easily generalizable
- slow to be taken up

From R to a web app



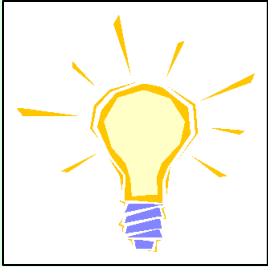
Innovative use of GBIF occurrence data

Wallace (beta v0.1): Harnessing Digital Biodiversity Data for Predictive Modeling, Fueled by R

This team has created a GUI interface for R packages related to biodiversity informatics and predictive modeling.

Anderson et al. GBIF Nielsen Challe... YouTube

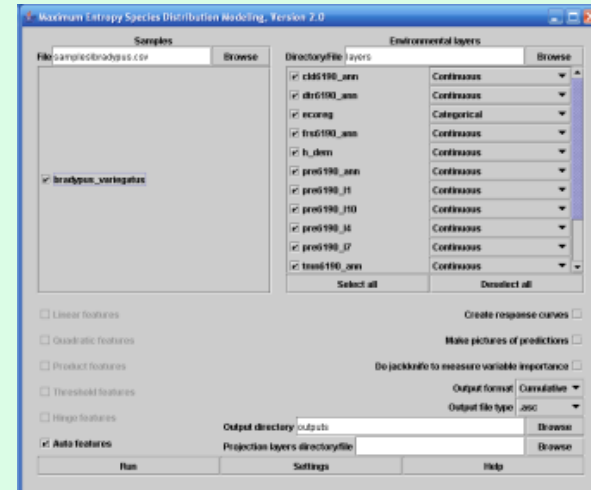




Combine code-based methods with GUI

Foreground:

Flexible, interactive,
expandable GUI








Background:

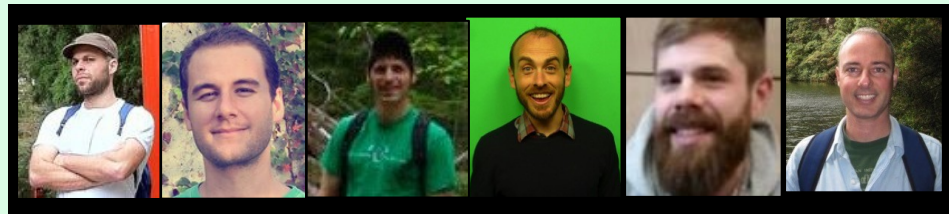
R code from new
packages: reproducible

```
42 ~ [{"r"}]
43 results <- occ(query = "Tremarctos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
44 results.data <- results[["gbif"]]$data[[formatSpName("Tremarctos ornatus")]]
45 occs <- remDups(results.data) # remove rows with duplicate coordinates
46 ~ [{"r"}]
47 ~ [{"r"}]
48 ~ [{"r"}]
49 ~ [{"r"}]
50 ~ [{"r"}]
51 ~ [{"r"}]
52 ~ [{"r"}]
53 ~ [{"r"}]
54 ~ [{"r"}]
55 ~ [{"r"}]
56 ~ [{"r"}]
57 ~ [{"r"}]
58 ~ [{"r"}]
59 ~ [{"r"}]
60 ~ [{"r"}]
61 ~ [{"r"}]
62 ~ [{"r"}]
63 ~ [{"r"}]
64 ~ [{"r"}]
65 ~ [{"r"}]
66 ~ [{"r"}]
67 ~ [{"r"}]
68 ~ [{"r"}]
69 ~ [{"r"}]
70 ~ [{"r"}]
71 ~ [{"r"}]
72 ~ [{"r"}]
73 ~ [{"r"}]
74 ~ [{"r"}]
75 ~ [{"r"}]
76 ~ [{"r"}]
77 ~ [{"r"}]
78 ~ [{"r"}]
79 ~ [{"r"}]
80 ~ [{"r"}]
81 ~ [{"r"}]
82 ~ [{"r"}]
83 ~ [{"r"}]
84 ~ [{"r"}]
85 ~ [{"r"}]
86 ~ [{"r"}]
87 ~ [{"r"}]
88 ~ [{"r"}]
89 ~ [{"r"}]
90 ~ [{"r"}]
91 ~ [{"r"}]
92 ~ [{"r"}]
93 ~ [{"r"}]
94 ~ [{"r"}]
95 ~ [{"r"}]
96 ~ [{"r"}]
97 ~ [{"r"}]
98 ~ [{"r"}]
99 ~ [{"r"}]
100 ~ [{"r"}]
```

APPLICATION

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

Jamie M. Kass^{1,2}  | Bruno Vilela³  | Matthew E. Aiello-Lammens⁴  |
Robert Muscarella⁵  | Cory Merow⁶ | Robert P. Anderson^{1,2,7} 



The City College
of New York

 The City
University
of New York

 **UFG**
UNIVERSIDADE
FEDERAL DE GOIÁS

 University of
Connecticut

 AARHUS UNIVERSITY



 Universidad
de Alcalá

 **PACE**
UNIVERSITY

Yale

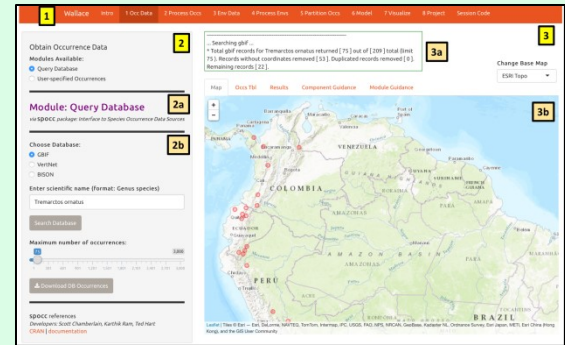
Wallace: currently under expansion ...

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

... and Jenna Rios, Sara Varela, Hannah L., Owens, Brian S. Maintner, Olivier Broennimann, Mary E. Blair, Ned Horning, Jorge Velásquez-Tibatá, Gonzalo E. Pinilla-Buitrago, Beth E. Gerstner, Peter J. Galante, Sarah I. Meenan, Valentina Grisales-Betancur, Cecina Babich-Morrow, Erica E. Johnson ...



What is *Wallace*?



Point-and-click (GUI) application that includes most steps of a niche/distribution modeling workflow

- Harnesses R packages and gives credit to their authors

R packages on the inside

0000100000
0000101000
0000101000
0100000100
0100000100
0000000001
Software notes

Ecography 38: 001–005, 2015
doi: 10.1111/ecog.01132
© 2015 The Authors. Ecography © 2015 Nordic Society Oikos
Subject Editor: Thiago Rangel. Editor-in-Chief: Miguel Aráujo. Accepted 18 November 2014

spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models

Methods in Ecology and Evolution
Methods in Ecology and Evolution 2014, 5, 1198–1205 doi: 10.1111/2041-210X.12261

APPLICATION
ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models

Robert Muscarella^{1*}, Peter J. Galante², Mariano Soley-Guardia^{2,3}, Robert A. Boria², Jamie M. Kass^{2,3}, Maria Uriarte¹ and Robert P. Anderson^{2,3,4}

¹Department of Ecology, Evolution and Environmental Biology, Columbia University, 1200 Amsterdam Ave., New York, NY 10027, USA; ²Department of Biology, City College of the City University of New York, 160 Convent Ave., New York, NY 10031, USA; ³Graduate Center of the City University of New York, 365 5th Ave., New York, NY 10016, USA; and ⁴Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West & 79th Street, New York, NY 10024, USA





... and now many other contributors

Wallace

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

2 Obtain Occurrence Data

Modules Available:

- Query Database
- User-specified Occurrences

2a Module: Query Database

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

2b Choose Database:

- GBIF
- VertNet
- BISON

Enter scientific name (format: Genus species)

Tremarctos ornatus

Search Database


Maximum number of occurrences:

75 / 3,000

Download DB Occurrences

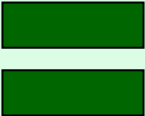
3a ... Searching gbif ...
* Total gbif records for Tremarctos ornatus returned [75] out of [209] total (limit 75). Records without coordinates removed [53]. Duplicated records removed [0]. Remaining records [22].

3b Map Occs Tbl Results Component Guidance Module Guidance



3 Change Base Map
ESRI Topo

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



Qualities of *Wallace*

Wallace's code is free
and **OPEN**

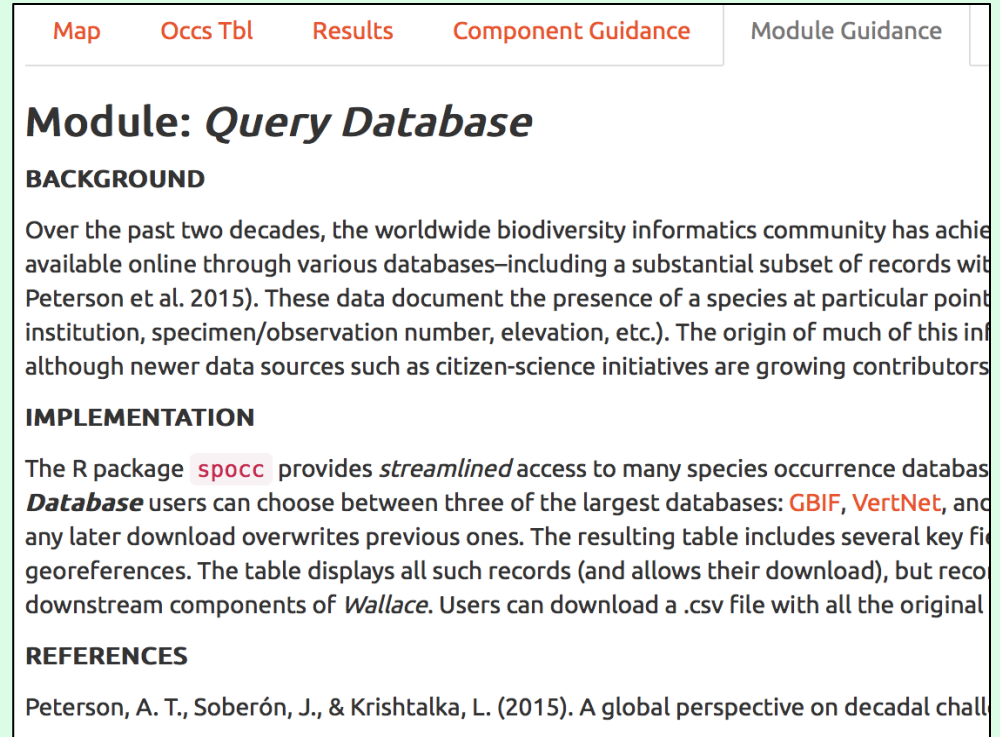
(& users can download
data from online
databases.)

The screenshot displays the Wallace web application interface. At the top, a navigation bar includes a yellow box with the number '1' and the following links: Wallace, Intro, 1 Occ Data, and 2 Process Occs. The main content area is titled 'Obtain Occurrence Data' with a yellow box containing the number '2'. Below this, the 'Modules Available:' section shows two radio buttons: 'Query Database' (selected) and 'User-specified Occurrences'. A horizontal line separates this from the 'Module: Query Database' section, which has a yellow box with '2a' and the text 'via spocc package: Interface to Species Occurrence Data Sources'. Another horizontal line leads to the 'Choose Database:' section, which has a yellow box with '2b' and three radio buttons: 'GBIF' (selected), 'VertNet', and 'BISON'. Below this is a text input field containing 'Tremarctos ornatus' and a 'Search Database' button. Further down is a slider for 'Maximum number of occurrences:' with a value of '75' and a maximum of '3,000'. A 'Download DB Occurrences' button is located below the slider. At the bottom, there is a section for 'spocc references' with the text 'Developers: Scott Chamberlain, Karthik Ram, Ted Hart' and a link to 'CRAN | documentation'.

Qualities of *Wallace*

Wallace provides
GUIDANCE

that addresses
conceptual and
methodological
issues.



The screenshot shows the Wallace software interface. At the top, there is a navigation bar with tabs for 'Map', 'Occs Tbl', 'Results', 'Component Guidance', and 'Module Guidance'. The 'Component Guidance' tab is currently selected. Below the navigation bar, the page title is 'Module: Query Database'. The main content area is divided into sections: 'BACKGROUND', 'IMPLEMENTATION', and 'REFERENCES'. The 'BACKGROUND' section discusses the worldwide biodiversity informatics community and the availability of species occurrence data through various databases. The 'IMPLEMENTATION' section describes the R package 'spocc' and its streamlined access to species occurrence databases, mentioning 'GBIF', 'VertNet', and 'any later download overwrites previous ones'. The 'REFERENCES' section lists a citation by Peterson, A. T., Soberón, J., & Krishtalka, L. (2015).

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

Module: Query Database

BACKGROUND

Over the past two decades, the worldwide biodiversity informatics community has achieved significant progress in making species occurrence data available online through various databases—including a substantial subset of records with georeferences (Peterson et al. 2015). These data document the presence of a species at particular points in time and space (e.g., institution, specimen/observation number, elevation, etc.). The origin of much of this information is in museum collections, although newer data sources such as citizen-science initiatives are growing contributors.

IMPLEMENTATION

The R package `spocc` provides *streamlined* access to many species occurrence databases. **Database** users can choose between three of the largest databases: **GBIF**, **VertNet**, and **any later download overwrites previous ones**. The resulting table includes several key fields, including georeferences. The table displays all such records (and allows their download), but records are not yet integrated into the downstream components of *Wallace*. Users can download a .csv file with all the original data.

REFERENCES

Peterson, A. T., Soberón, J., & Krishtalka, L. (2015). A global perspective on decadal challenges in biodiversity informatics.

Qualities of *Wallace*

Wallace is **FLEXIBLE**

by providing multiple options, and allowing user inputs and downloads for most components.

Module: Select Study Region

via `sp` and `rgeos` packages: *Title Classes and Methods for Spatial Data | Interface to Geometry Engine - Open Source (GEOS)*

Background Extents:

- Bounding box
- Minimum convex polygon
- User-specified polygon

Study region buffer distance (degree)


1

Clip Env Data by Polygon

Select File Type

GRD

Download Clipped Env Data



Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, Community, Map data: © OpenStreetMap, SRTM Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye

Qualities of *Wallace*

Wallace features

INTERACTIVE

maps, tables, and graphs to explore data and model predictions.



Qualities of *Wallace*

Wallace is
REPRODUCIBLE

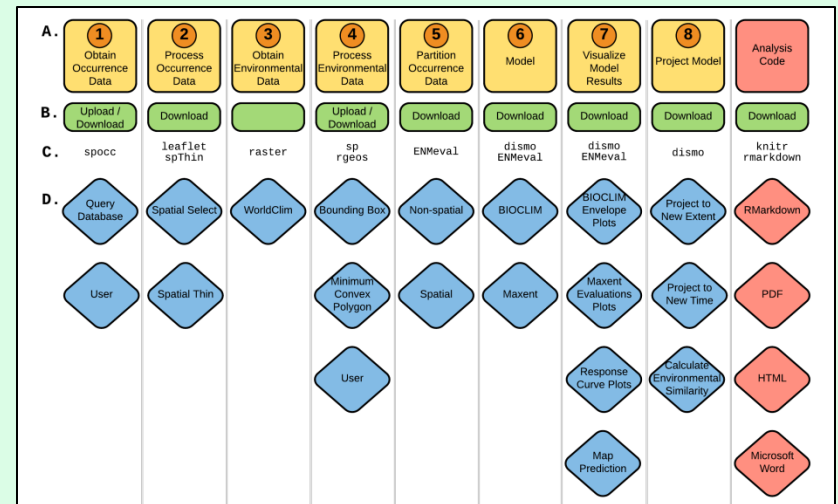
by providing
executable code for
documenting and
rerunning the
analysis.

```
1 Please find below the R code history from your *Wallace* session. You can reproduce your session
file in RStudio (for more information see <http://rmarkdown.rstudio.com>).
2
3 ## Package installation
4
5 Wallace uses the following R packages that must be installed and loaded before starting.
6
7 {r}
8 library(devtools)
9 library(spocc)
10 library(maptools)
11 library(spThin)
12 library(dismo)
13 library(rgeos)
14 library(repmis)
15 library(maps)
16 library(ENMeval)
17 {r}
18
19 Wallace also includes several functions developed to help integrate different packages and some of
this reason, it is necessary to load the file 'functions.R', which can be found on Wallace's GitHub
(<https://github.com/wallaceEcoMod/wallace>). Download the file, place it in your working directory
and then load it:
20
21 {r}
22 source(file.path("/Users/musasabi/Documents/github/wallace", 'functions.R'))
23 {r}
24
25 Record of analysis for *Tremarctos ornatus*.
26 -----
27
28 ## Obtain Occurrence Data
29
30 The search for occurrences was limited to 81 records. Obtain occurrence records of the selected s
31
32 {r}
33 results <- occ(query = "Tremarctos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
34 results.data <- results[["gbif"]]$data[[formatSpName("Tremarctos ornatus")]]
35 occs <- remDups(results.data) # remove rows with duplicate coordinates
36 {r}
37
38 ## Process Occurrence Data
39
40 You chose 42 of 51 total occurrence localities via polygon selection to keep in the analysis.
41
42 {r}
43 occs <- occs[c(1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, 26, 2
40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51), ] # subset occs by selected rows
```

Qualities of *Wallace*

Wallace is **EXPANDABLE!**

by featuring different methodological options (modules) that researchers can contribute to advance functionality.



Wallace: software for niche/distribution modeling

Qualities:

open
instructive
flexible
interactive
reproducible
expandable

*broadening
human diversity*

Female and minority-student coder-biologists



External partners, undergraduates, and interns/fellows:

Sara Varela, Hannah L. Owens, Cecina Babich-Morrow,
Jenna Rios, Sarah I. Meenan, Valentina Grisales-Betancur

Installing *Wallace*

- Install R and (optionally) Rstudio
- For Maxent java version, install maxent.jar (https://biodiversityinformatics.amnh.org/open_source/maxent/) and appropriate Java version for your system, then place maxent.jar in `dismo` folder
- Install the `wallace` package from CRAN

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

- Troubleshooting on Github page (<https://github.com/wallaceEcoMod/wallace>)

Wallace today

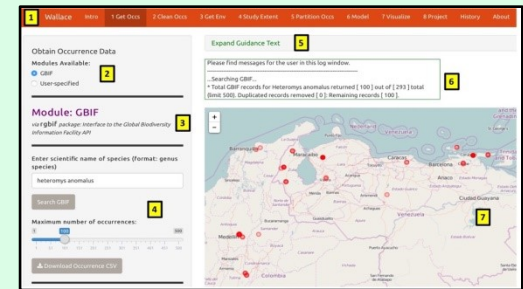
1. Full Release 1.0.6 (Kass et al. 2018, Methods in Ecology & Evolution software note; Google group; Wallace e-mail)

2. Ongoing NSF funding to work with external partners to add new modules

3. Ongoing NASA funding, led by Mary Blair (AMNH), to develop new R packages and add them to Wallace and interface with BioModelos



Mary E. Blair



NSF DEB-1119915
& DBI-1650241



NSF DBI-1661510



NASA 80NSSC18K0406

Acknowledgments

Anderson lab
(chronosequence)



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

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

