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

Jamie M. Kass, Bruno Vilela, Matthew Aiello-Lammens, Robert Muscarella, and Robert P. Anderson
What is **Wallace**? And Why?
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

Agenda: Making data and modeling ready to address critical environmental issues of the 21st century
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
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
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
1. High-quality data, ready to be accessed when the particular problem presents itself

Continued large-scale, across-the-board databasing and georeferencing initiatives

High priority taxonomic projects realized in parallel
2. **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 the algorithm/s used
2. 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 the algorithm/s used

“*umbrella tools that can be employed with any particular modeling algorithm*”
Agenda: applied biodiversity informatics

3. **Scientists** capable of building, applying, and appraising high-quality models

- clear literature
- workshops
- graduate courses
Correlative modeling of species niches and distributions

Occurrence records → Algorithm → Model of niche → Suitable areas

Environmental data
LEAD ARTICLE

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

Robert P. Anderson

INTRODUCTION

I offer this short essay with the idea of emphasizing that, despite being a relatively new field still under development, ecological niche modeling (and its applications for studies of most important tools in biogeography and related fields (Guisan & Thuiller 2005, Wiens & Graham 2005, Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011), but I feel that it is not clear to many researchers just how tight the
Recent R packages: spThin and ENMeval

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

Matthew E. Aiello-Lammens, Robert A. Boria, Aleksandar Radosavljevic, Bruno Vilela and Robert P. Anderson

Methods in Ecology and Evolution

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

Robert Muscarella1,2, Peter J. Galante3, Mariano Soley-Guardia4,5, Robert A. Boria5, Jamie M. Kass5,6, Maria Uriarte1 and Robert P. Anderson2,3,4

1Department of Ecology, Evolutionary Biology and Environmental Studies, Columbia University, 1300 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; 4Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West & 79th Street, New York, NY 10024, USA

Matthew Aiello-Lammens

Robert Anderson, Robert Boria, Peter Galante, Jamie Kass, Aleksandar Radosavljevic, Mariano Soley-Guardia, Maria Uriarte, Bruno Vilela

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

Methods in Ecology and Evolution

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

Robert Muscarella1, Peter J. Galante2, Mariano Soley-Guardia2,3, Robert A. Borja2, Jamie M. Kass2,3, Maria Uriarte1 and Robert P. Anderson4,5,6

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
4Department of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West & 79th Street, New York, NY 10024, USA
Creativity, discernment, “discussion” and FUN
Wallace

Module: Query Database

Choose Database:
- GBIF
- VertNet
- BISON

Enter scientific name (format: Genus species)
Tremarctos ornatus

Maximum number of occurrences:
- 1
- 3,000

Download DB Occurrences

SPOCC references
Developer: Scott Chamberlain, Kaarthik Ram, Ted Hart
CRAN | documentation
What is *Wallace* beta 0.6?

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

- Harnesses R packages and gives credit to their authors
Qualities of Wallace

*Wallace’s code is free and OPEN*

(& users can download data from online databases.)
Qualities of Wallace

Wallace provides **GUIDANCE** that addresses conceptual and methodological issues.
Qualities of Wallace

Wallace is **FLEXIBLE** by providing multiple options, and allowing user inputs and downloads for most components.
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.
Wallace is **EXPANDABLE**!

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

**A.**
1. Obtain Occurrence Data
2. Process Occurrence Data
3. Obtain Environmental Data
4. Process Environmental Data
5. Partition Occurrence Data
6. Model
7. Visualize Model Results
8. Project Model

**B.**
- Upload / Download
- Download
- Upload / Download
- Download
- Download
- Download
- Download

**C.**
- spocc
- leaflet
- spThin
- raster
- sprgeos
- ENMeval
- dismo
- ENMeval
- dismo

**D.**
- Query Database
- Spatial Select
- WorldClim
- Bounding Box
- Non-spatial
- BIOCLIM

- Minimum Convex Polygon
- Spatial
- Maxent
- Maxent Evaluations Plots

- User
- Spatial Thin

- User
- Response Curve Plots
- Map Prediction

**Analysis Code**
- dismo
- knitr
- rmarkdown

**Other Tools**
- BIOCLIM Envelope Plots
- Project to New Extent
- PDF
- HTML
- Microsoft Word
Wallace is expandable
Wallace is expandable
Wallace is expandable
Future directions

1. First *full release soon* (software note in review)

2. Ongoing NSF funding to respond to immediate user needs (*Google group and Wallace e-mail*)

3. Seeking more funding, *with Cory Merow* (to work with external partners to add new modules)
THANK YOU

… please help us make *Wallace* better … and grow

DEB-1119915 and DBI-1650241

http://wallaceecomod.github.io
http://www.andersonlab.ccny.cuny.edu