



#### Species Distribution Modeling with Remote Sensing

Amber McCullum, Juan Torres-Perez, Zach Bengtsson Guest Speakers: Erica E. Johnson, Andrea Paz Velez, Mary E. Blair

Aug 17, 2021

### **Course Structure and Materials**

- Three 1.5-hour sessions on August 12, 17, & 19
- Sessions will be presented once in English 12:00-13:30 EDT
- Webinar recordings, PowerPoint presentations, and the homework assignment can be found after each session at:
  - <u>https://appliedsciences.nasa.gov/j</u> <u>oin-mission/training/english/arset-</u> <u>species-distribution-modeling-</u> <u>remote-sensing</u>
- Q&A following each lecture and/or by email at:
  - <u>amberjean.mccullum@nasa.gov</u>
  - juan.l.torresperez@nasa.gov
  - bengtsson@baeri.org

TRAINING ARSET - Species Distribution Modeling with Remote Sensing

PROGRAM AREA: ECOLOGICAL FORECASTING





Amber McCullum

Juan Torres-Pérez



Zach Bengtsson



2

#### Webinar Agenda

Part 1: Overview of Species Distribution Models (SDMs) Part 2: Using Wallace to Model Species Niches and Distributions Part 3: Additional SDM Tools and Techniques, ASP Projects, and Summary





# Using Wallace to Model Species Niches & Distributions

#### **Session 2 Overview**



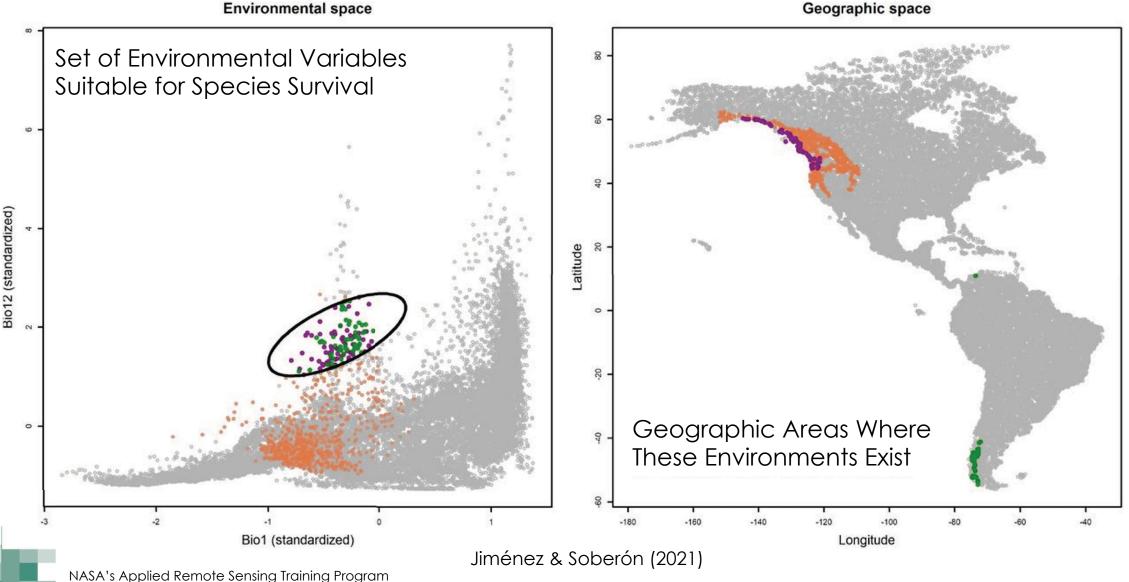
- 1. Species Distribution Models in Wallace
- 2. Wallace Walkthrough: Running a Full SDM Workflow
- 3. Redesign & Expansion: Making it Easier to Visualize Results and Add New Modules
- 4. Extensions: SDM Post-Processing for Conservation Decision-Making and Beyond





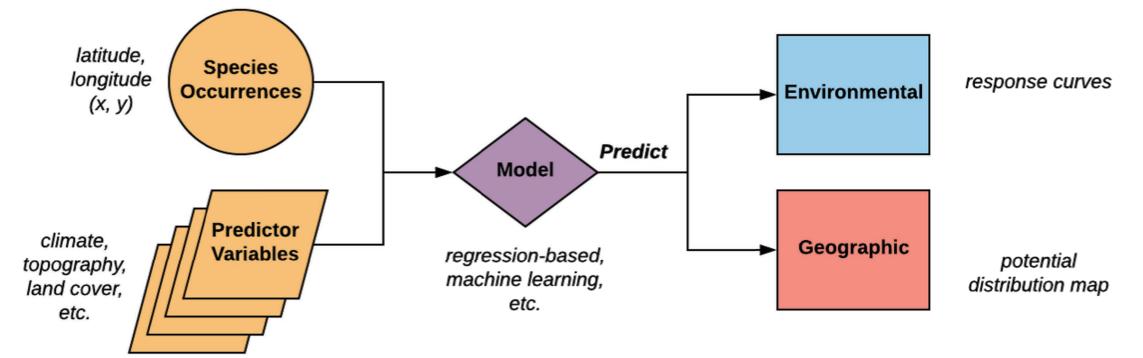
# Species Distribution Models in Wallace

# From Ecological Niches to Geographic Distributions



7

# Species Distribution Modeling Overview



Elith & Graham 2009, Phillips et al. 2017; Guisan et al. 2017



# **Needed in Species Distribution Modeling**

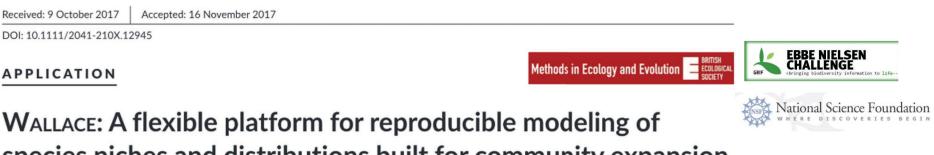
Software that achieves a balance between automation and supervision by:

- Automating repetitive tasks
- Forcing the user to make critical biological and conceptual decisions
- Being general with respect to the algorithm(s) used

# **Needed in Species Distribution Modeling**

Software that achieves a balance between automation and supervision by:

- **Automating** repetitive tasks
- Forcing the user to make critical biological and conceptual decisions
- **Being general** with respect to the algorithm(s) used



species niches and distributions built for community expansion

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



Anderson 2012; Kass et al. 2017

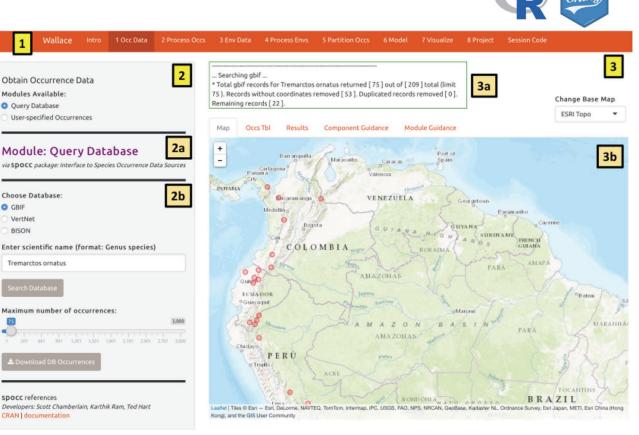


# Wallace: An Ecological Modeling Application

A **user-friendly** application for species distribution modeling that provides **guidance** towards following **best-practices** at each step.

#### Wallace is:

- Accessible
- Instructive
- Flexible
- Interactive
- Reproducible
- Expandable
- Open





11

### Accessible

- Lowers barriers to implement cutting-edge SDM techniques
- Allows users to download occurrence & environmental data from diverse sources
- Users can find support from the community through our various networks
  - Google Groups
  - GitHub

btain Occurrence Data		2
Iodules Available:		
Query Database		
User-specified Occurrences		
		22
Aodule: Query Data		20
a <b>SPOCC</b> package: Interface to Speci	es Occurrence D	ata Sources
hoose Database:		2b
GBIF		
VertNet		
BISON		
nter scientific name (format: C	enus species	)
Tremarctos ornatus		
Court Database		
Search Database		
	ac.	
Search Database laximum number of occurrence	es:	3,000
laximum number of occurrence		
laximum number of occurrence	1	×1 P
aximum number of occurrence	1	×1 P
Taximum number of occurrence	1	×1
Taximum number of occurrence	1	×1
Taximum number of occurrence	801 2,101 2,401	×1



### Instructive

- Provides educational & instructional resources addressing:
  - Concepts/Theories
  - Methods
- Guides users in following best-practices

#### Map Occs Tbl Results Component Guidance Module Guidance

#### Module: Query Database

#### BACKGROUND

Over the past two decades, the worldwide biodiversity informatics community has achie available online through various databases–including a substantial subset of records wit Peterson et al. 2015). These data document the presence of a species at particular point institution, specimen/observation number, elevation, etc.). The origin of much of this inl although newer data sources such as citizen-science initiatives are growing contributors

#### IMPLEMENTATION

The R package **spocc** provides *streamlined* access to many species occurrence databas **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 fit georeferences. The table displays all such records (and allows their download), but reco downstream components of *Wallace*. Users can download a .csv file with all the original

#### REFERENCES

Peterson, A. T., Soberón, J., & Krishtalka, L. (2015). A global perspective on decadal chall



### Flexible

- Allows users to select from multiple:
  - Data Sources
  - Analytical Tools
- Designed to fit the user's needs

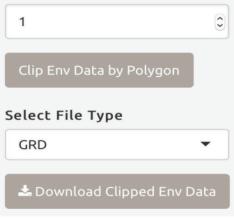
#### Module: Select **Study Region**

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 0
- User-specified polygon

#### Study region buffer distance (degree)



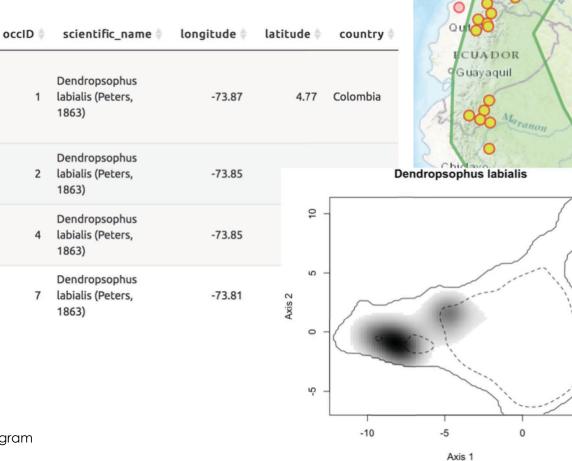


Leaflet | Tiles © Esri – Esri, DeLorme, NAVTEQ, 1 GeoBase, Kadaster NL, Ordnance Survey, Esri Ja Community, Map data: © OpenStreetMap, SRTM Source: Esri, i-cubed, USDA, USGS, AEX, GeoEy



### Interactive

- Provides an assortment of dynamic visualizations to explore data and results
  - Maps
  - Tables
  - Graphs

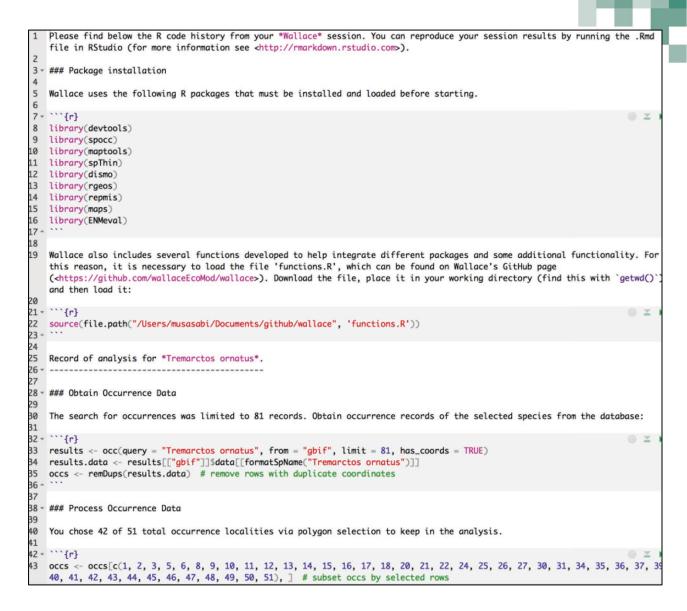


NAMA



### Reproducible

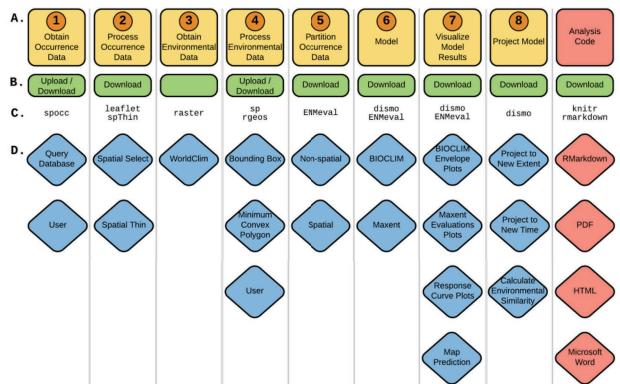
- Allows users to:
  - Download executable code
  - Replicate analyses
  - Save work and load later (no need to redo steps)





### Expandable

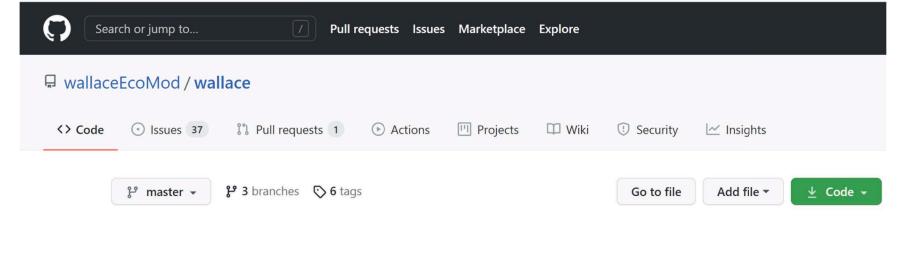
- Users can contribute new functions & analytical tools to be integrated as:
  - Components (A)
  - Modules (D)





### Open

- Code is publicly available for users to:
  - Download (CRAN, GitHub)
  - Modify & suggest enhancements
- Enables users to become contributors





# Wallace Walkthrough: Running a Full SDM Workflow

# Install and Run Wallace

- 1. Install R (version  $\geq$  v.3.5.0) and RStudio
- 2. Install the Wallace package from CRAN (v.1.1; stable)
- 3. Run Wallace in RStudio

# install the package
install.packages('wallace')
# load the package
library(wallace)
# run the app
run\_wallace()





Redesign & Expansion: Making it Easier to Visualize Results and Add New Modules

# Wallace V2 Coming Soon

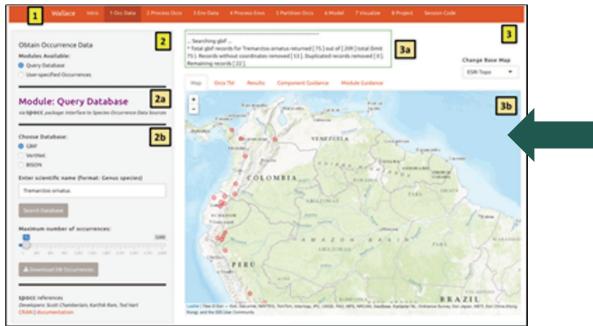
- New modules added with help from collaborators around the world
  - More data sources
  - More analytical features
- Re-designed to make **module additions easier**
- Now based on RMMS metadata (Merow et al. 2019)
- Development guided by **user feedback**:
  - Conference workshops, emails, Google group



Extensions: SDM Post-Processing for Conservation Decision-Making and Beyond

# Wallace V3 Partnership for Conservation

#### Wallace



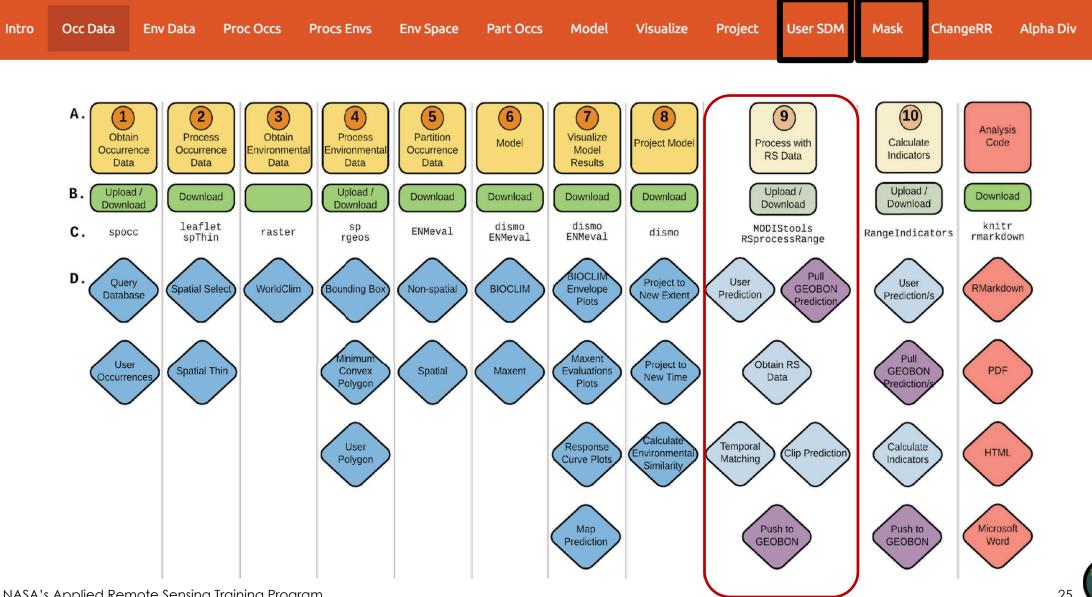


GEO BON in a Box

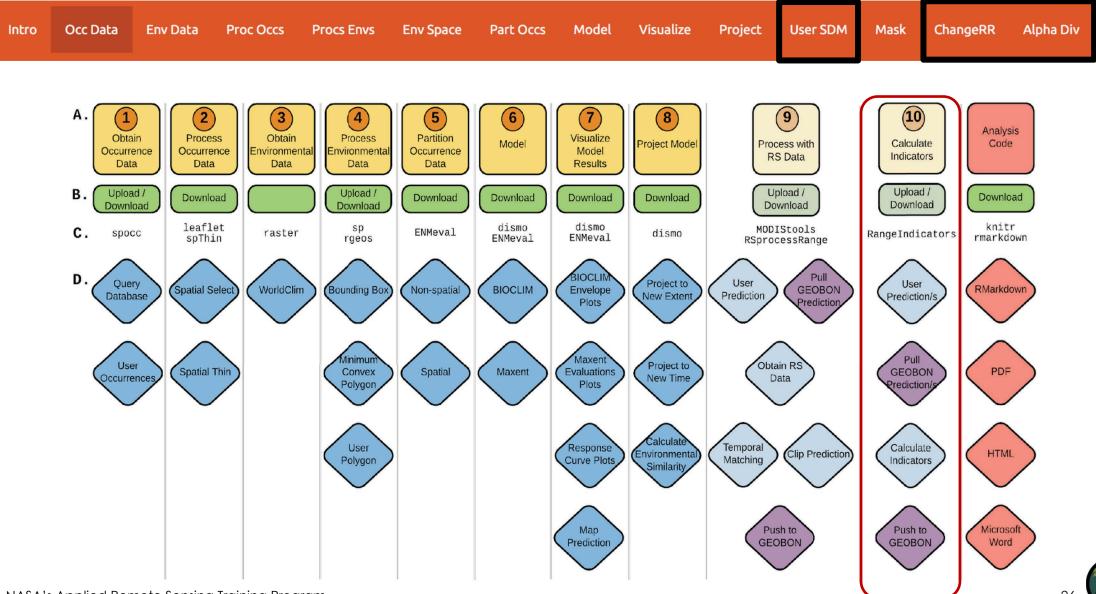
Galante et al. in prep.



# New Components Added to Wallace (V3)

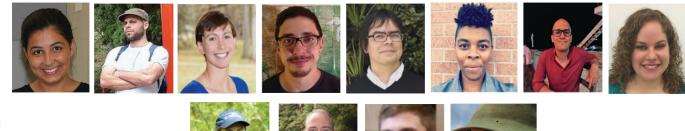


# New Components Added to Wallace (V3)

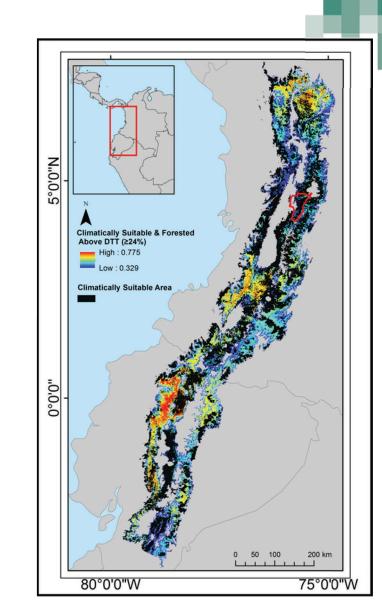


# Mask SDM (Based on maskRangeR)

- Post-processes SDMs to estimate current ranges:
  - RS Products
    - E.g., Forest Cover, Urbanization, Cultivars
  - User-Defined Polygons
    - E.g., Protected Areas, Land Cover
  - Expert-Defined Maps/Ranges







Merow et al. in rev.



NASA's Applied Remote Sensing Training Program

# Mask Example: Olinguito (Bassaricyon neblina)

- Recently described small carnivore
- Limited to high-altitude cloud forests
- Data-poor
- Needs IUCN status update given recent deforestation





Merow et al. In rev.



# Mask Example: Olinguito (Bassaricyon neblina)

#### Masked Distribution SDM **MODIS Yearly** Recent Forest Cover Occurrences 5°0'0"N Climatically Suitable & Forested Above DTT (≥24%) Suitability High : 0.775 High ow : 0.329 **Climatically Suitable Are** ..0,0<sub>°</sub>0 80°W 79°W 78°W 77°W 76°W 75°W 74°W 73°W 0 50 100 Km 11111 80°0'0"W 75°0'0"W 200 km 80°0'0"W 75°0'0"W

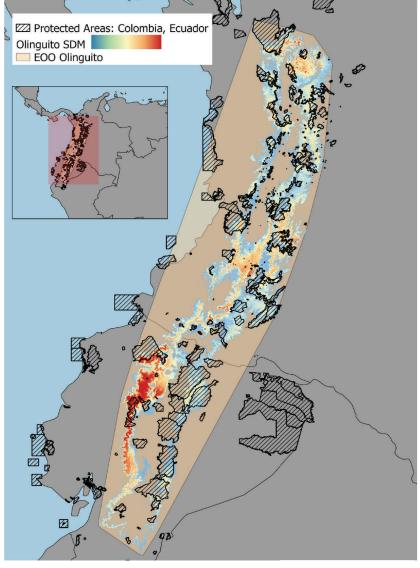


29

# Calculate Indicators (Based on changeRangeR)

- Calculate key biodiversity change indicators
  - IUCN AOO & EOO
  - Percentage of Suitable Land Cover
  - Protected Area Representativeness

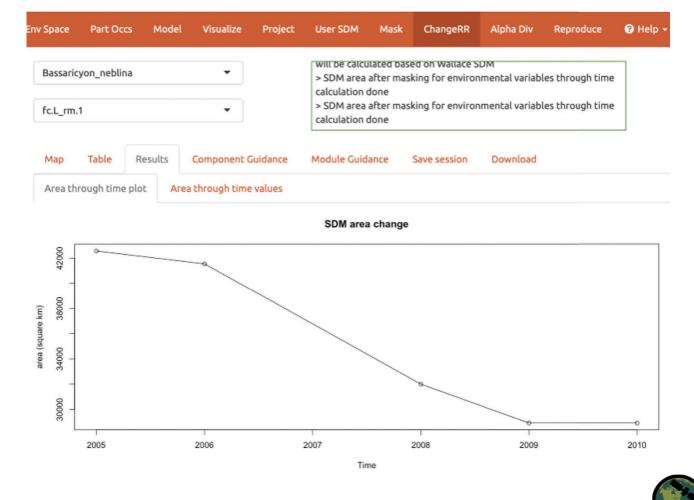






# Calculate Indicators Example: Olinguito (Bassaricyon neblina)

- Changes in Range, AOO, EOO given GIS data:
  - Changes over time/space
  - E.g., deforestation over time



#### Calculate Indicators Example: Colombian Primates Richness

- Multispecies Diversity Metrics:
  - Species Richness
  - Species Endemism
- Soon in **Wallace**:
  - Phylogenetic Diversity
  - Phylogenetic Endemism
  - Complementarity





#### Summary

- Wallace is a user-friendly application for species distribution modeling that provides guidance towards following best-practices at each step.
  - Accessible, Instructive, Flexible, Interactive, Reproducible, Expandable, Open
- Wallace V2 (coming soon)
  - Provides additional SDM data sources & analytical tools.
  - Facilitates module contributions (from user community)
- Wallace V3 (in development)
  - Added tools for conservation applications
    - Estimate species ranges
    - Calculate biodiversity indicators.
- Next Session: Additional SDM Tools and Techniques

- Trainers:
  - Juan Torres-Pérez: juan.l.torresperez@nasa.gov
  - Amber McCullum: <u>amberjean.mccullum@nasa.gov</u>
  - Zach Bengtsson: <u>bengtsson@baeri.org</u>
- Training Webpage:
  - <u>https://appliedsciences.nasa.gov/join-mission/training/english/arset-species-distribution-modeling-remote-sensing</u>
  - ARSET Website:
  - <u>https://appliedsciences.nasa.gov/what-we-do/capacity-building/arset</u>

Follow us on Twitter <u>@NASAARSET</u>





### Thank You!

● wallaceecomod.github.io

groups.google.com/g/wallaceecomod

NASA's Applied Remote Sensing Training Program

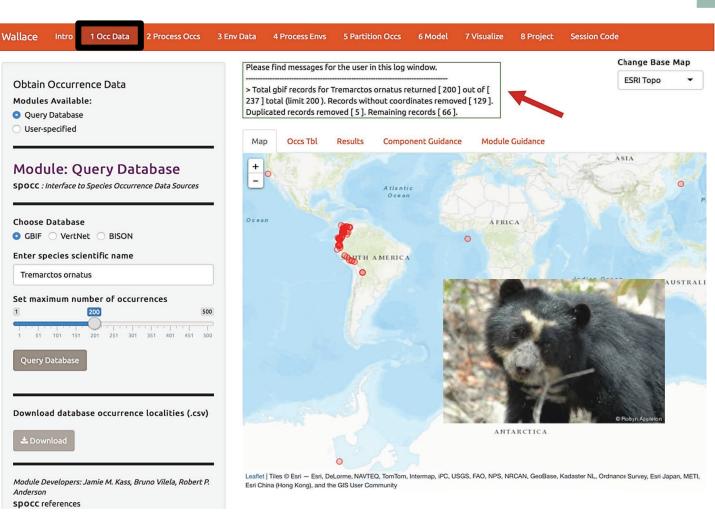




# Supplemental: Wallace Walkthrough Slides

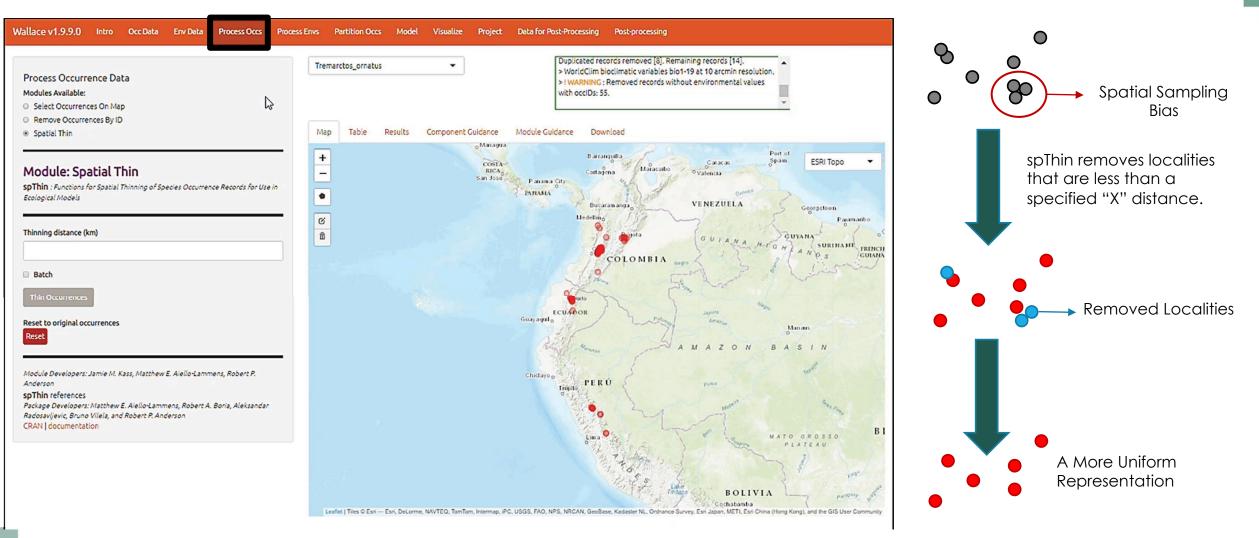
## Component 1: Obtain Presence Data (Occ Data)

- User-specified
- Download from online databases
  - GIBF
  - Ver Net





#### Component 2: Process Presence Data (Process Occ)





## Component 3: Obtain Environmental Data (Env Data)

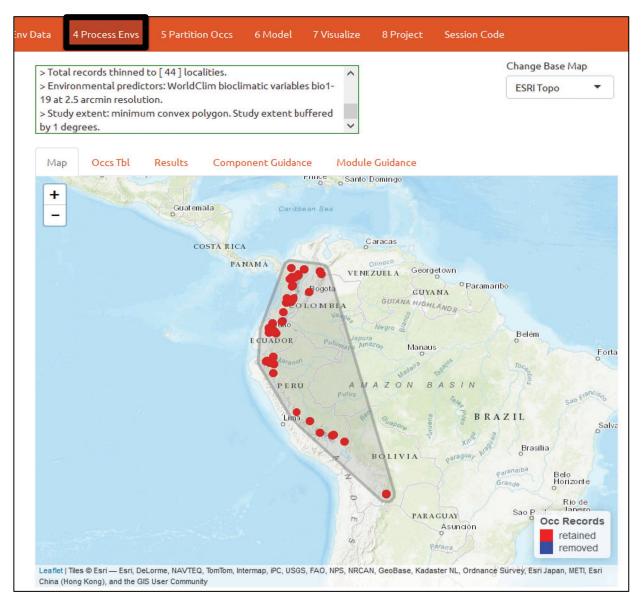
- User-specified
- Download from online databases
  - WorldClim

Wallace Intro 1 Occ Data 2 Process Occs	3 Env Data 4 Process Envs 5 Partition Occs 6 Model 7 Visualize 8 Project Session (
Obtain Environmental Data Modules Available: WorldClim Bioclims User-specified	<ul> <li>&gt; Removed occurrence with ID = 159. Updated data has n = 56 records.</li> <li>&gt; Total records thinned to [44] localities.</li> <li>&gt; Environmental predictors: WorldClim bioclimatic variables bio1-19 at 2.5 arcmin resolution.</li> <li>Map Occs Tbl Results Component Guidance Module Guidance</li> </ul>
Module: WorldClim Bioclims raster : Geographic Data Analysis and Modeling Select WorldClim bioclimatic variable resolution	<pre>class : RasterStack dimensions : 3600, 8640, 31104000, 19 (nrow, ncol, ncell, nlayers) resolution : 0.04166667, 0.04166667 (x, y) extent : -180, 180, -60, 90 (xmin, xmax, ymin, ymax) coord. ref. : +proj=longlat +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0 names : bio01, bio02, bio03, bio04, bio05, bio06, bio07, bio08, bio09, biol3, bio14, bio15,</pre>
2.5 arcmin	min values : -278, 9, 8, 64, -86, -559, 53, -278, -501, 0, 0, 0, max values : 319, 213, 96, 22704, 489, 258, 725, 376, 365, 2437, 697, 265,
Select bio1 bio2 bio3 bio4 bio5 bio6 bio7 bio8 bio9 bio10 vbio11 vbio12 bio13 vbio14 vbio15 bio16 vbio17 vbio18 vbio19	
Using map center coordinates as reference for tile download. Using map center -75.512, 4.356	
Load Env Data	



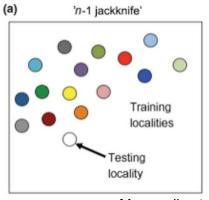
## Component 4: Process Environmental Data (Process Envs)

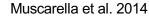
- Select Study Region
  - User-Specified
  - Bounding Box
  - Minimum Convex Polygon
  - Point-Buffers
- Sample Background Points



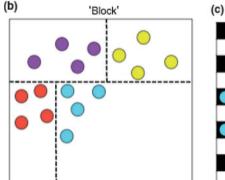
## Component 5: Partition Occurrence Data (Partition Occs)

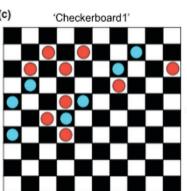
• Non-Spatial (Random)

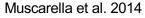


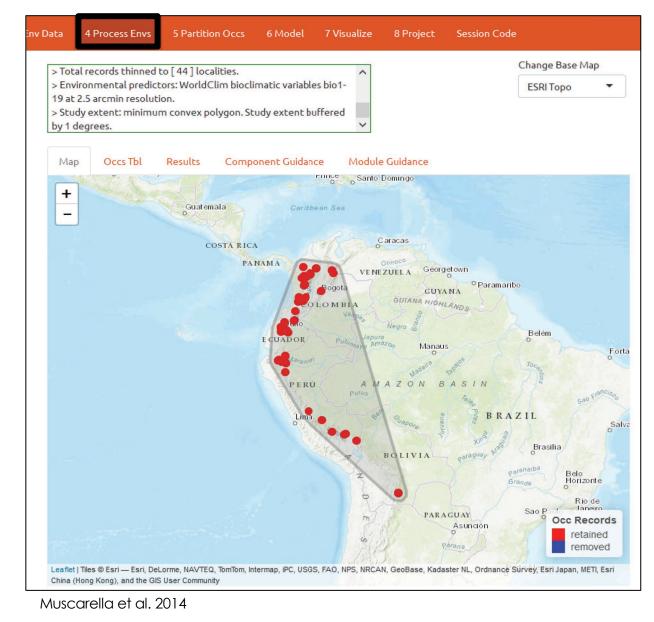


• Spatial









NASA's Applied Remote Sensing Training Program

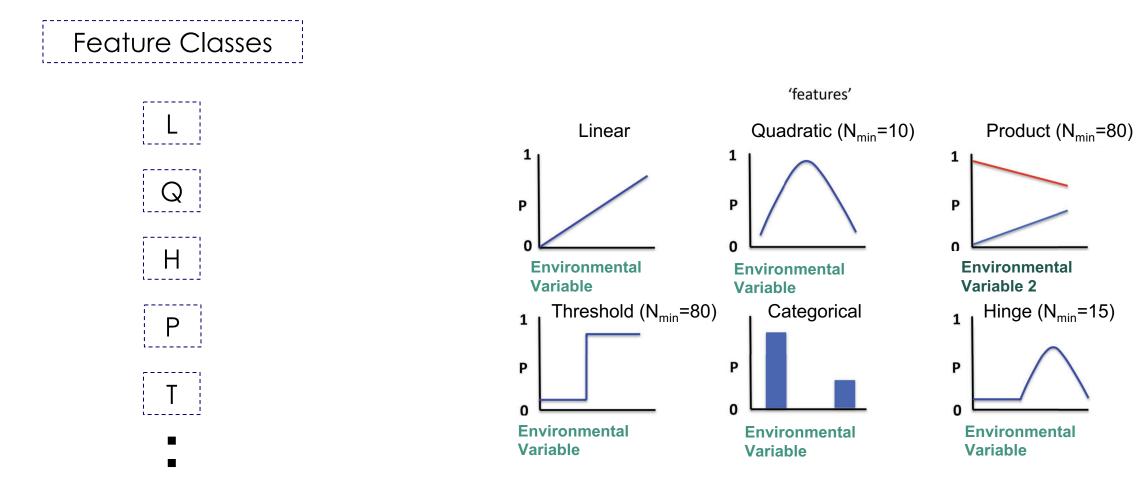
## Component 6: Model Building & Evaluation (Model)

- Modeling Algorithms:
  - Bioclim
  - Maxent
    - maxent.jar
    - Maxnet
- Model Tuning:
  - Test parameter combinations
  - Find best-performing model
  - Reduce complexity

Wallace Intro 1 Occ Data 2 Process Occs 3 Env Data 4 Process Envs	5 Partition	Occs 6 Mc	odel 7 Visua	lize	8 Project	Session Code		
Build and Evaluate Niche Model Modules Available: O BIOCLIM Maxent	> Oc > Oc	currences par currences par exent ran succe els.	titioned by bloc titioned by bloc titioned by bloc essfully and out bl Results	k metho k metho put eval	d. d.	~	idance	
Module: Maxent	Full	model and	l partition b	in avei	age evalua	ation statistics		
ENMeval : Automated Runs and Evaluations of Ecological Niche Models dismo : Species Distribution Modeling		settings	features 🖨	rm∳	Full.AUC 🔷	avg.test.AUC 🛊	var.test.AUC 🖗	avg.diff.AUC
	1	L_1	L	1	0.817	0.767	0.046	0.09
(NOTE : see module guidance for troubleshooting tips if you are experiencing	2	LQ_1	LQ	1	0.9	0.781	0.033	0.09 0.114 0.124 0.134 0.07 0.14 0.097 0.104 0.097 0.104 0.097
problems.)	3	H_1	н	1	0.926	0.806	0.024	0.124
Select feature classes (flexibility of modeled response)	4	LQH_1	LQH	1	0.93	0.801	0.023	0.132
key: Linear, Quadratic, Hinge, Product, Threshold	5	L_1.5	L	1.5	0.816	0.779	0.032	0.075
⊠r ⊠ró ⊠h ⊠róh □róhb □róhbi	6	LQ_1.5	LQ	1.5	0.88	0.785	0.029	0.11
Select regularization multipliers (penalty against complexity)	7	H_1.5	н	1.5	0.914	0.83	0.02	0.092
0 0 0 0 10	8	LQH_1.5	LQH	1.5	0.917	0.821	0.018	0.108
	9	L_2	L	2	0.819	0.79	0.025	0.06
Multiplier step value	10	LQ_2	LQ	2	0.864	0.798	0.028	0.093
0.5	<							
Clamp predictions?	Indi	vidual part	ition bin ev	aluatic	n statistic	5		
Run		AUC_bin.1	AUC_bin.	2 🕴 🗛	UC_bin.3 🛊	AUC_bin.4 AL	JC.DIFF_bin.1 🛊	AUC.DIFF_bin.2
	1	0.64	1 0.7	09	0.786	0.934	0.206	0.12
Module Developers: Jamie M. Kass, Robert Muscarella, Bruno Vilela, Robert P. Anderson	2	0.69	5 0.7	91	0.71	0.927	0.201	0.08

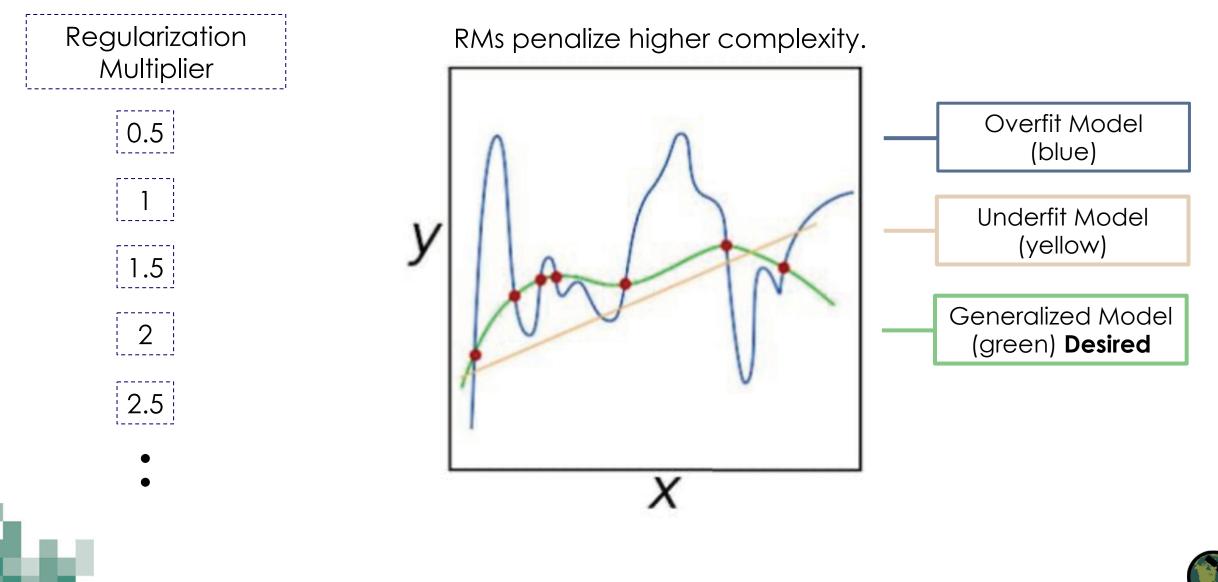


## **Component 6: Model Building Background**



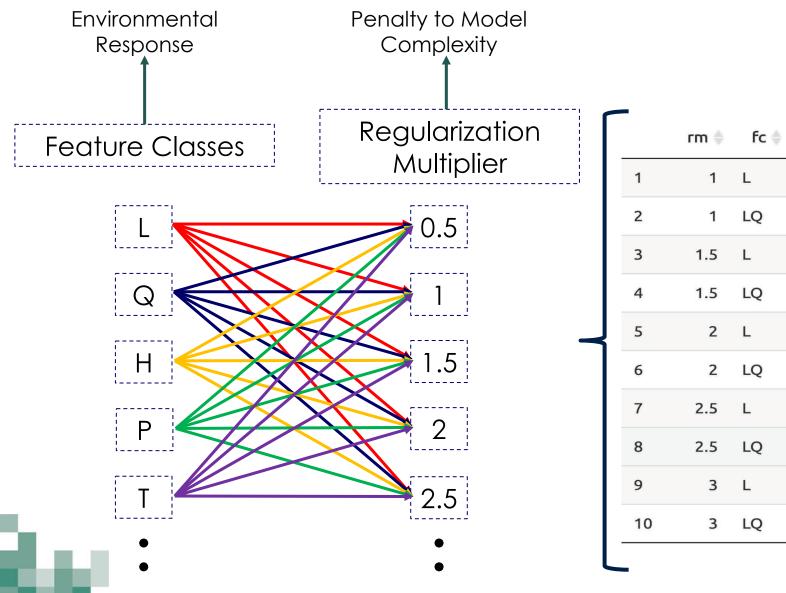


## **Component 6: Model Building Background**



44

# **Component 6: Model Tuning**



#### Recommendations:

- >80 occurrence records: All Feature Classes
- 15 79 occurrence records: L, Q, H
- 10 14 occurrence records: L and Q
- < 10 occurrence records: L only



## **Component 6: Model Evaluation**

- Modeling Evaluation Metrics
  - AUC
  - Omission Rate
  - AICc
  - N Parameters
- Applicable To:
  - Entire Model
  - Each Partition

> Occurren > BIOCLIM	nces partitioned by Inces partitioned by Iran successfully an Iran successfully and	block method. d output evaluatio		Current Env Varia		
Map Full mod	Occs Tbl Resu		evaluation statis	bio02 ule Guidance	• L_1	
.test.AUC 🍦	avg.diff.AUC 🔷	var.diff.AUC 崇	avg.test.or10pct 🔺	var.test.or10pct 🖨 a	vg.test.orMTP	var.test
0.013	0.049	0.009	0.159	0.046	0.045	
0.025	0.071	0.026	0.182	0.066	0.091	
0.013	0.049	0.01	0.182	0.066	0.045	
0.01	0.043	0.008	0.182	0.066	0.023	
0.008	0.037	0.006	0.182	0.066	0.023	
0.01	0.036	0.006	0.182	0.066	0.023	
0.009	0.033	0.005	0.182	0.066	0.023	
0.008	0.029	0.004	0.182	0.066	0.023	
0.051	0.095	0.055	0.205	0.057	0.114	
0.025	0.07	0.026	0.205	0.09	0.091	
٢				Previou	s 1 2	3 Ne

## **Component 6: Model Evaluation**

	_		

AUC	orMTP/10pct	AICc	Nparam
Measures a model's ability to distinguish true & false presences	Measures the % of true presence points predicted as absences	Measures the relative quality of a model by balancing fit & complexity	More parameters increase model complexity

Generally, we want models with:

- High AUC
  - Low Omission Rate

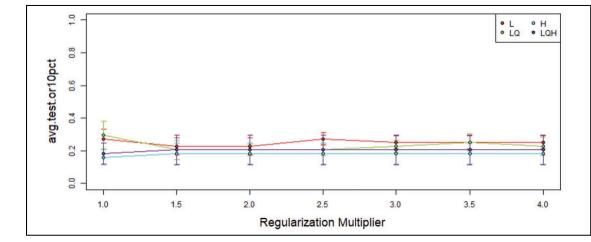
- Low AICc
- Fewer Parameters

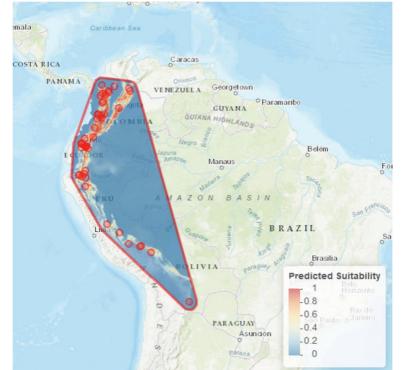


•

## Component 7: Visualizing Results (Visualize)

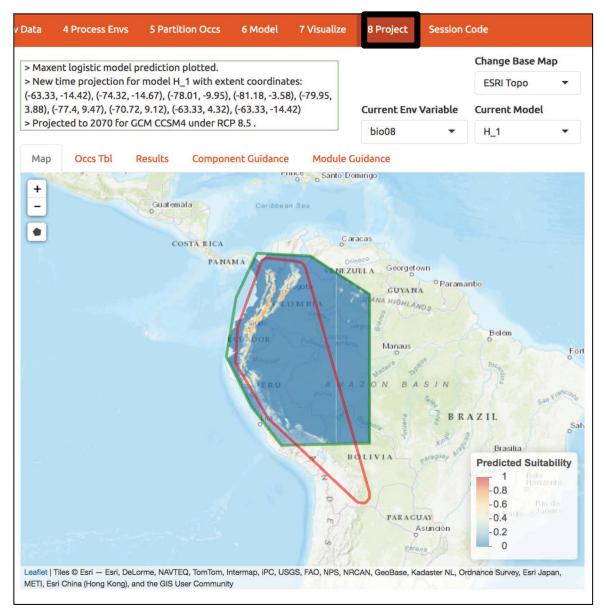
- Visualize results as:
  - Model Evaluation Plots
  - Response Curves
  - Prediction Maps
    - Continuous
    - Binary





## Component 8: Spatial & Temporal Projections (Project)

- Allows us to project models to:
  - New Geographic Extents
  - Different Time Periods
    - Future
    - Past



## Component 9: Download the R Code (Session Code)

- Download the R code.
- Replicate analyses in R.

allace Intro	1 Occ Data	2 Process Occs	3 Env			
Download Sess	ion Code					
rmarkdown : Dyna knitr : A General-Pu Generation in R						
Select download fil Rmd	e type	•				
📩 Download Sess	ion Code					
Module Developers Anderson rmarkdown refere Package Developer	ences	file in RStud 2 3 - ### Package i 4 5 Wallace uses	dio (for more information installation	<pre>see <http: pre="" rmarkdown.<=""></http:></pre>	sion. You can reproduce you rstudio.com>). and loaded before starting	
Wickham, Aron Atkins, Rob Hyndman, R 8 CRAN   documentation 9 knitr references 11 Package Developers: Yihui Xie 12 CRAN   documentation 13 15 16	<ul> <li>R 8 library(devto</li> <li>9 library(spoce</li> <li>10 library(spoce</li> <li>11 library(spThi</li> <li>12 library(dismo</li> <li>13 library(regesi</li> <li>14 library(regeini</li> <li>15 library(maps)</li> </ul>	c) bools) in) b) s) b)				
		18 19 Wallace also this reason, ( <https: git<br="">and then load</https:>	it is necessary to load t thub.com/wallaceEcoMod/wal	the file 'functions.R',	tegrate different packages o which can be found on Walla ile, place it in your workin	ace's GitHub page
		20 21 - ```{r} 22	path("/Users/musasabi/Docu	uments/github/wallace",	'functions.R'))	
		25 Record of ana 26	alysis for *Tremarctos orn	natus*.		





#### Thank You!



NASA's Applied Remote Sensing Training Program