Day 1: Introduction and Data Preparation

1. General Introduction
2. Introduction to the Data
3. Occurrence Records
4. Environmental Data
5. *Optional: Getting ready to use R*

1. General Introduction

1.1 Hello and Course Info

1.2 What do we mean by “Ecological Niche”?

1.3 Niche Modeling Overview

1.4 Ecological Niche Model versus Species’ Distribution Model?

1.5 Niche Modeling Applications

1.6 Exercise

1.1 Hello and Course Info

Bonjour Hallo Hola 你好 Ciao Olá

Who I am

- Post-doc in Loren Rieseberg's lab
- CSEE student/post-doc representative (renew your membership!)
- Studying species' geographic range limits and factors shaping the distribution of genetic diversity
- Study systems include amphibians and flowering plants

1.1 Hello and Course Info

Bonjour Hallo Hola 你好 Ciao Olá

My Niche

Office: 238 Biodiversity Centre
Email: lee-yaw [at] zoology.ubc.ca

1.1 Hello and Course Info

Bonjour Hallo Hola 你好 Ciao Olá

My Niche

Variables that determine presence…

…Perfect Latte
... Wifi
…large windows
...within dispersal distance from home

1.1 Hello and Course Info

Course Website

→ this is where all slides, dataset and R scripts can be found

Important!

→ Be sure to have R and all the listed libraries installed

→ Look over the R resources on the course website if you have never used R before

1.1 Hello and Course Info

What you can expect this course to deliver:

→ An overview of the steps involved in niche modeling, from data collection to using models to make predictions

→ Information about the parameters and assumptions that one has to consider when building niche models

→ Many suggestions for further reading

→ Scripts and instruction on how to generate a niche model in R using MAXENT

What this course won't do:

→ Replace the need to read up on the methods on your own to determine the best way forward for your own project

→ Provide detailed instruction on other modeling algorithms

→ Provide tutorials for all uses of niche models

→ Teach you R

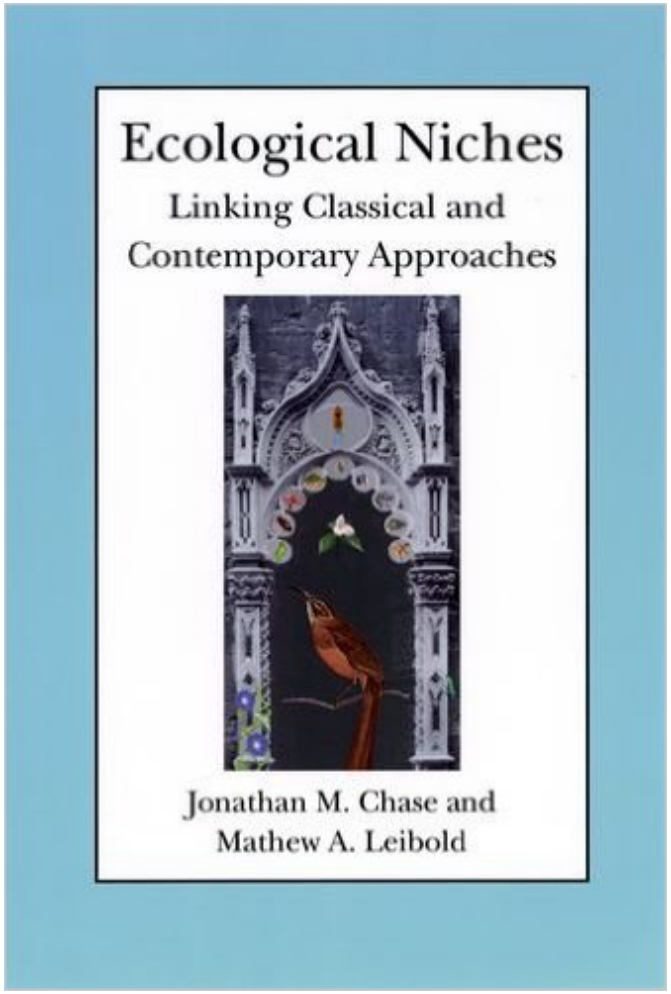
1.2 What do we mean by “niche”?

Many definitions!

Eltonian-type: the role of the organism in its biotic community (response to and effect of the species on the environment)

Grinnellian-type: habitat requirements and behaviours that allow a species to persist

Hutchinson-type: the set of conditions necessary for a species to persist



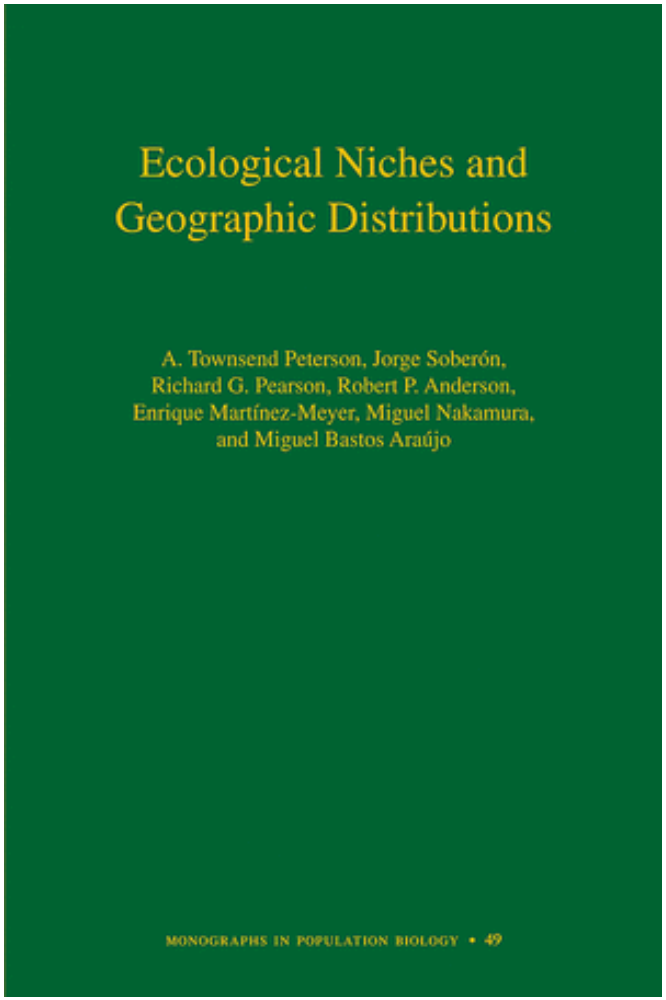
1.2 What do we mean by “niche”?

Many definitions!

Eltonian-type: the role of the organism in its biotic community (response to and effect of the species on the environment)

Grinnellian-type: habitat requirements and behaviours that allow a species to persist

Hutchinson-type: the set of conditions necessary for a species to persist



1.2 What do we mean by “niche”?

More specifically, niche models:

Treat the niche as habitat, rather than function

Deal with variables that don’t change with the density of the species (as opposed to variables that are consumed or influenced by the species)

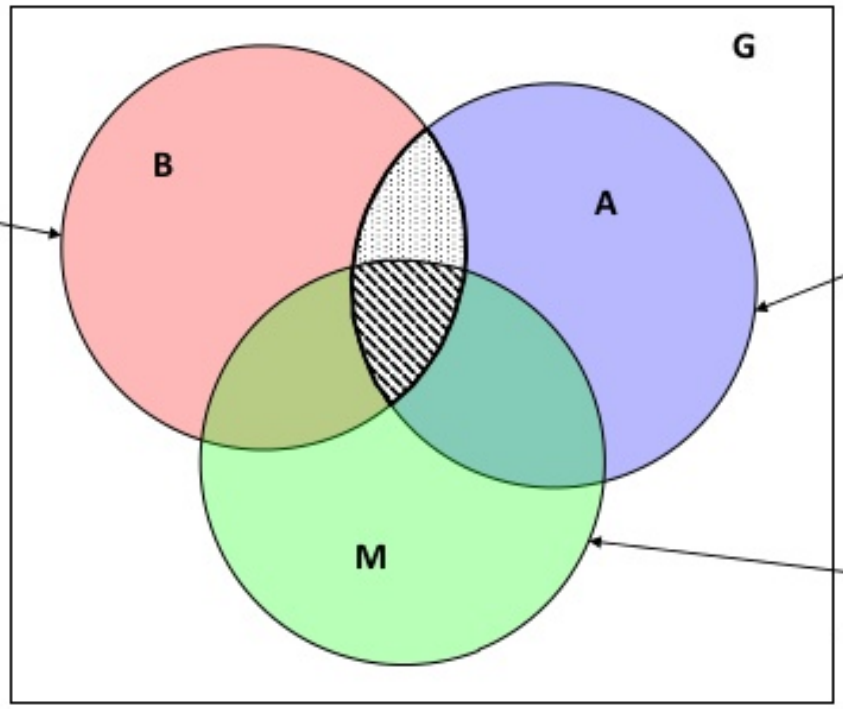
Most commonly, only directly address abiotic factors (though biotic factors may still be captured indirectly)

Model the realized or potential niche, not the fundamental niche

1.2 What do we mean by “niche”?

The BAM diagram: a useful conceptual tool

The region of geographic space where biotic interactions permit the species to persist



The region of geographic space where abiotic conditions are suitable for the species → **Available Fundamental Niche**

The region of geographic space that is/has been accessible to the species via dispersal

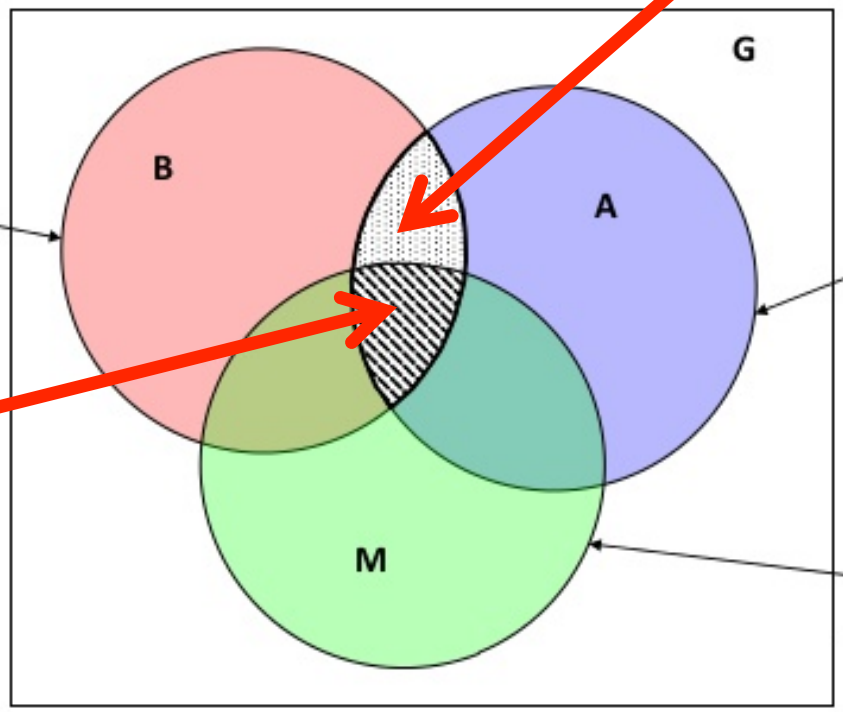
Diagram by A.T. Peterson

1.2 What do we mean by “niche”?

The BAM diagram: a useful conceptual tool

This is where the species could be if dispersal barriers were removed → **Potential Niche**

The region of geographic space where biotic interactions permit the species to persist



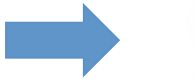
The region of geographic space where abiotic conditions are suitable for the species → **Available Fundamental Niche**

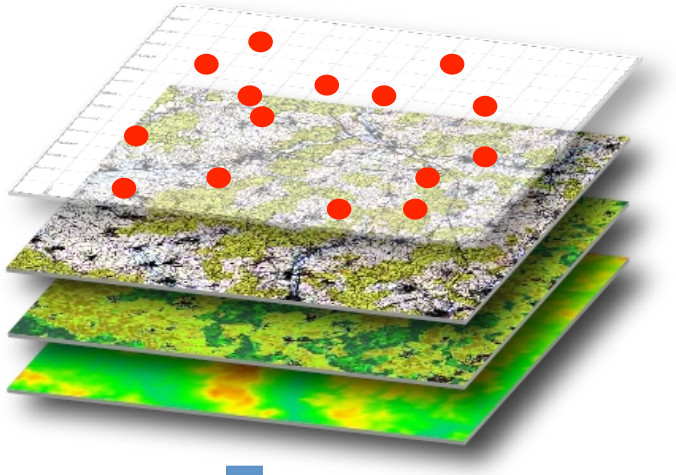
This is where the species is found… and where all our locality data are from → **Realized Niche**

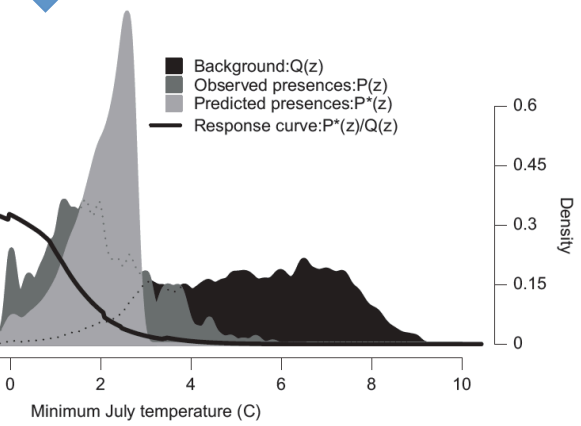
The region of geographic space that is/has been accessible to the species via dispersal

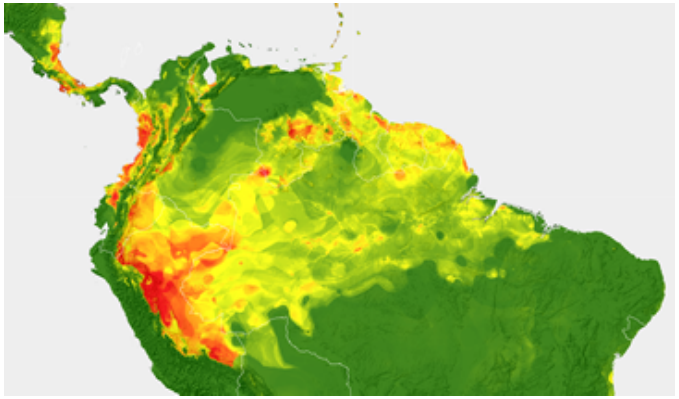
Diagram by A.T. Peterson

1.3 Niche Modeling Overview









1.4 Niche Model versus Distribution Model?

Even though the data used to construct the models are inherently distributional (e.g. *geographic* coordinates of places where the species has been observed)

"…most applications of ENM/SDM methods require the assumption that such models estimate some subset of conditions within which a species can survive and reproduce (i.e. the niche)" – Warren 2012

A personal preference: Depends on what you are trying to model and how you go about doing so!

1.5 Niche Modeling Applications

Predicting effects of climate change

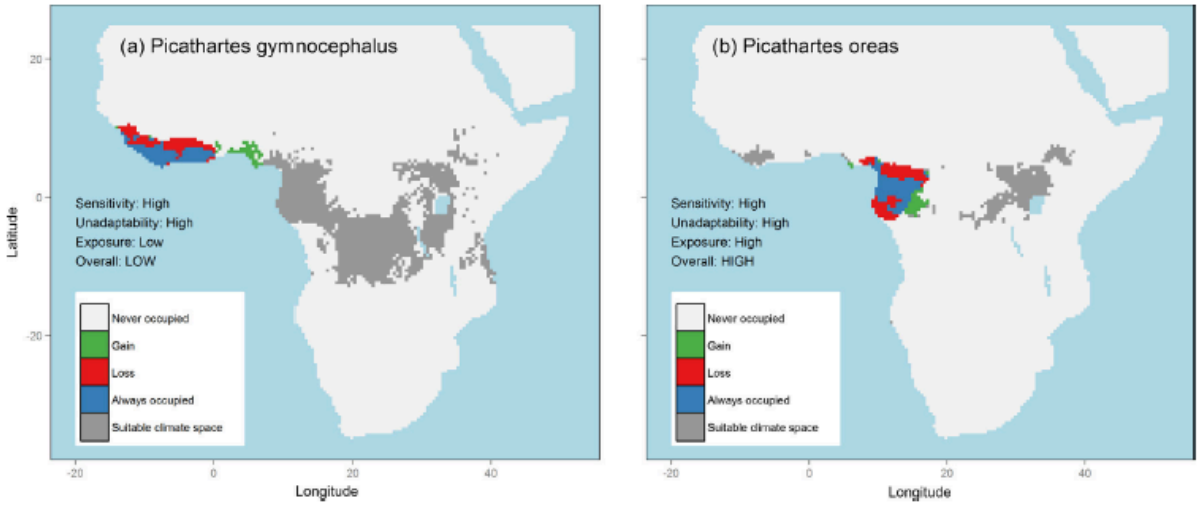


Fig. 1. Climate change vulnerability over the current century as assessed by simple SDMs and corresponding dynamic dispersal models; the latter considering changing climate suitability from SDMs along with dispersal-related traits (mean dispersal distance and generation length) from TVAs. Text within maps gives the TVA scores for sensitivity, low adaptive capacity, exposure and overall vulnerability for each species (from Foden et al., 2013). Figures show changes in simulated areas of occurrence (from SDMs) for two African bird species of current conservation concern: (a) White-necked Picathartes, *Picathartes gymnocephalus* and (b), Grey-necked Picathartes, *Picathartes oreas* between the present and 2100. Blue shading indicates climate suitability in both periods, green indicates areas becoming newly suitable by 2100 and within the species' dispersal capability over that period. Red indicates areas becoming unsuitable over the period, and grey indicates areas that are climatically suitable by 2100 (from the simple SDM) but beyond the species' dispersal capability during the period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Integrating climate change vulnerability assessments from species distribution models and trait-based approaches

S.G. Willis [a,*], W. Foden [b], D.J. Baker [a], E. Belle [c], N.D. Burgess [c,d], J.A. Carr [e], N. Doswald [c], R.A. Garcia [f,g,h], A. Hartley [i], C. Hof [j], T. Newbold [c], C. Rahbek [h], R.J. Smith [k], P. Visconti [l], B.E. Young [m], S.H.M. Butchart [n]

1.5 Niche Modeling Applications

Inferring historically suitable habitat (e.g. potential refugial areas)

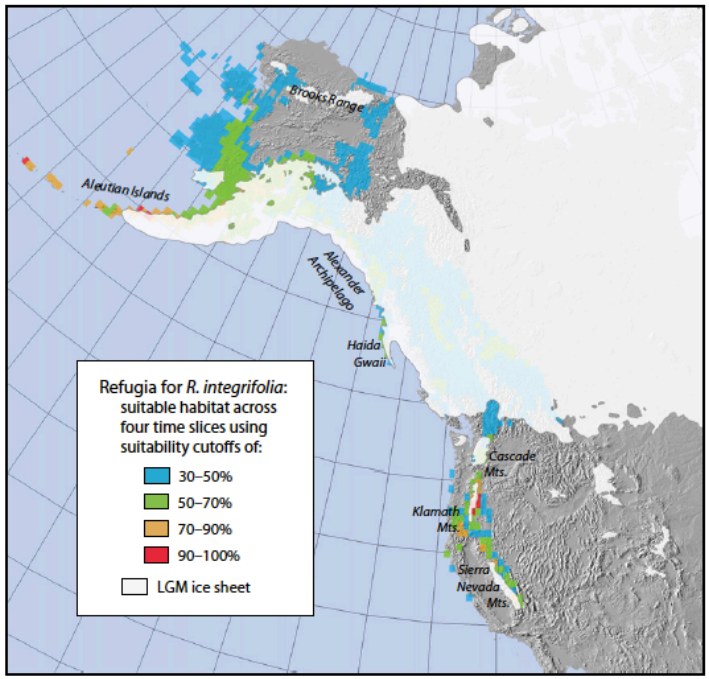


Figure 3 Consensus refugial model indicating areas of suitable habitat for *Rhodiola integrifolia* across four time slices: current conditions, mid-Holocene, Last Glacial Maximum (LGM) and Last Interglacial. Major features are labelled. Ice sheet layer is from Dyke *et al.*, 2003. Map projection: North America Albers Equal Area Conic.

Diversity and Distributions, (Diversity Distrib.) (2013) **19**, 1480–1495

BIODIVERSITY RESEARCH

Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions

Brenna R. Forester[1*†], Eric G. DeChaine[2] and Andrew G. Bunn[1]

1.5 Niche Modeling Applications

Studying invasion dynamics

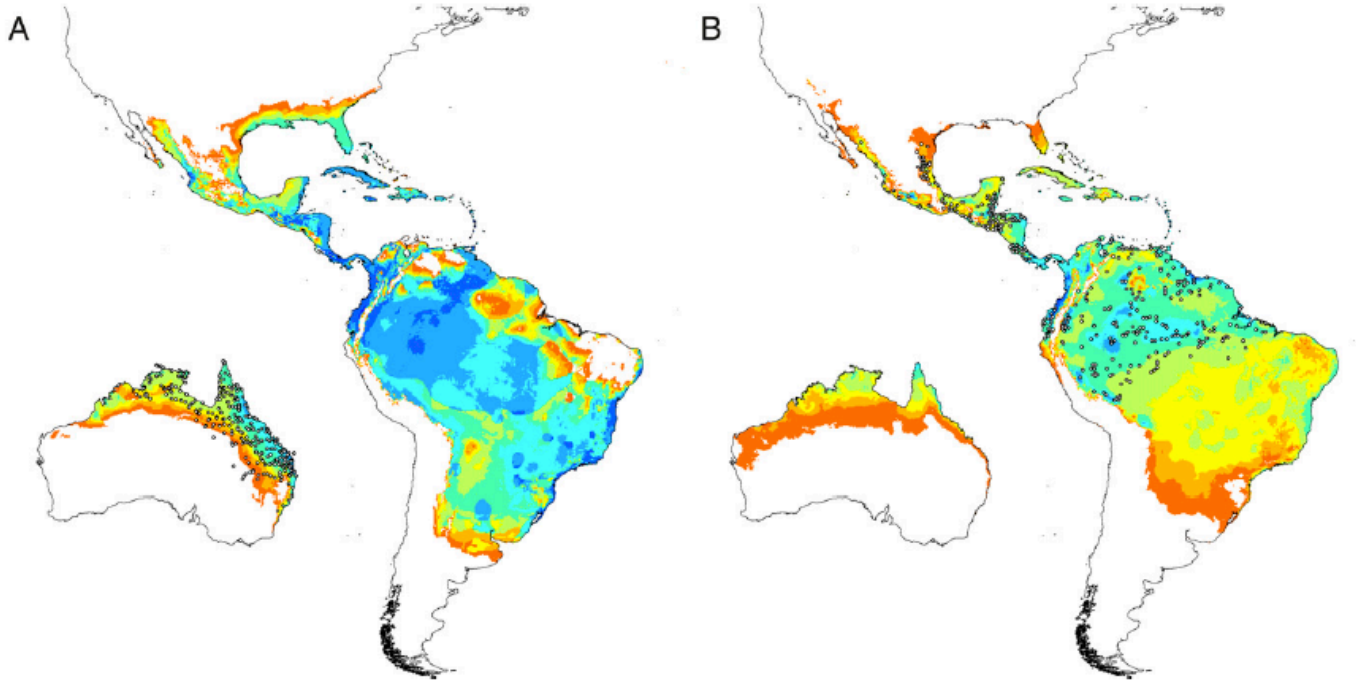


Fig. 2. Predictions of correlative Maxent models calibrated using data from either the Australian (*A*) or native (*B*) realized niche of *R. marina*. Predictions are depicted in 10% suitability classes ranging from white to orange to yellow to green to blue. White dots represent occurrence records of *R. marina* and have been thinned to improve visibility.

Realized niche shift during a global biological invasion

Reid Tingley[a,1], Marcelo Vallinoto[b,c], Fernando Sequeira[c], and Michael R. Kearney[d]

1.6 Exercise (15 minutes)

Exercise D1.1

1) Introduce yourself to the class:
   - Name and Lab
   - What system are you working on?
   - What are you using niche models for?
   - Random bit of information about you that Google has probably already figured out using it's fancy algorithms

2) Class shuffle: Sit with people working on same taxonomic group (at least for the next couple of exercises)

2. Introduction to the Tutorial Case Study

- The American Bullfrog (*Rana catesbeiana* or *Lithobates catesbeiana*)
- Native to Eastern North America (as far west as Oklahoma and Kansas)
- Inhabits large, permanent bodies of water (lakes, ponds, etc.)
- Breeds in late May, early June
- Preys upon anything it can get in its mouth
- Predated upon by herons and other large birds, otters, some snakes

2. Introduction to the Tutorial Case Study

- Has been introduced to 40 countries and four continents
- Invasive in western North America, including the Vancouver area
- Predator of native animals
- Outcompetes native amphibian species
- Carries *Batrachochytrium dendrobatidis*, a fungus that has decimated many amphibian populations

2. Introduction to the Tutorial Case Study

Our Mission:

Use niche models to identify areas of Western North America that are environmentally suitable for the species (i.e. “at risk”).

NOTE: This is a bit of a “choose your own adventure”. You have several modeling decisions to make along the way and I don’t expect everyone to necessarily make the same decisions…or produce the same final model.

3. Occurrence Records

3.1 Collecting Occurrence Data

3.2 Exercise

3.3 Issues with Repository Data

3.4 Georeferencing Tools

3.5 Exercise and Break

3.1 Collecting Occurrence Records

Ideally…

Project-specific survey data

No sampling bias in geographic or environmental space (relative to study extent)

Includes absences and accounts for detectability

Independent dataset for model evaluation

3.1 Collecting Occurrence Records

But usually we have to surrender our ideals…

~~Project-specific survey data~~

→ Data from secondary sources, collected opportunistically or for other purposes

~~No sampling bias in geographic or environmental space (relative to study extent)~~

→ Lots of sampling bias

~~Includes absences and accounts for detectability~~

→ Presence-only data most common

~~Independent dataset for model evaluation~~

→ Almost never done

3.1 Collecting Occurrence Records

But usually we have to surrender our ideals…

~~Project-specific survey data~~

→ Data from secondary sources, collected opportunistically or for other purposes

~~No sampling bias in geographic or environmental space (relative to study extent)~~

→Lots of sampling bias

~~Include absences and accounts for detectability~~

→ Presence-only data most common

~~Independent survey for model evaluation~~

→ Almost never done

Lingo:
"Presence-Only" Niche Modeling

3.2 Exercise (30 minutes)

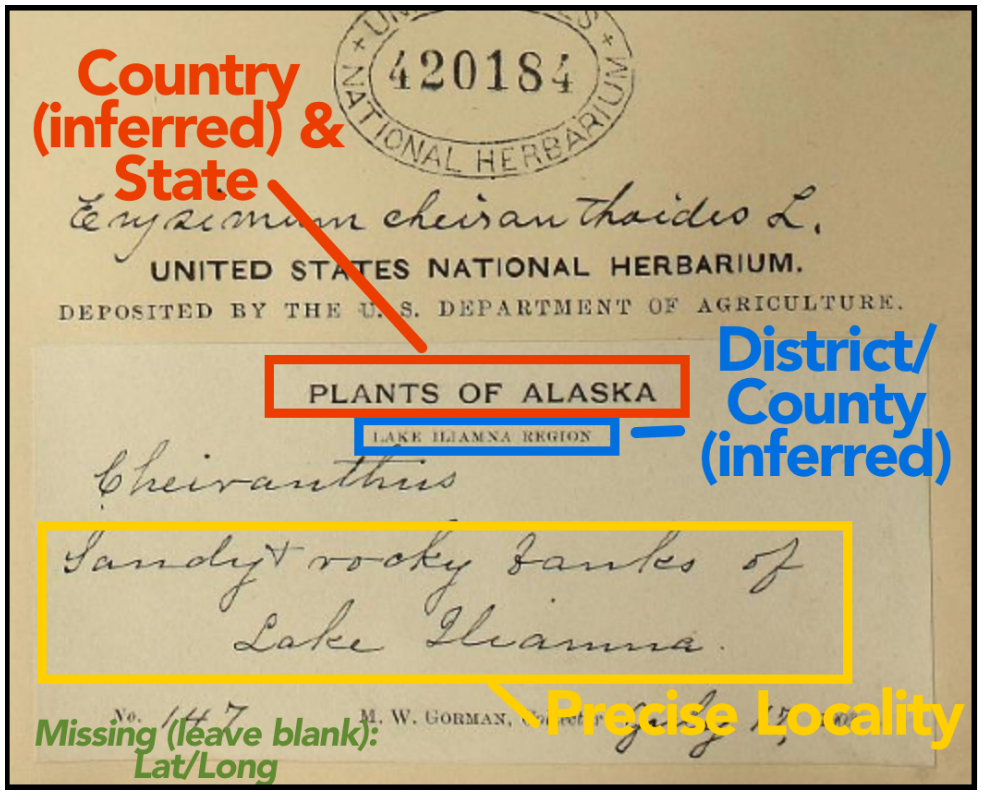
Exercise D1.2

1) Working alone or in a small group, conduct an internet search and/or survey the relevant literature to find potential sources of occurrence records for the species/taxonomic group that you work on

2) When you have a few ideas about where to look, come see me and we'll add them to the class resource page

3.3 Issues with repository data

a) Incomplete Information



3.3 Issues with repository data

b) Sampling Bias

How probable is it that all parts of this landscape were surveyed with equal frequency?

3.3 Issues with repository data

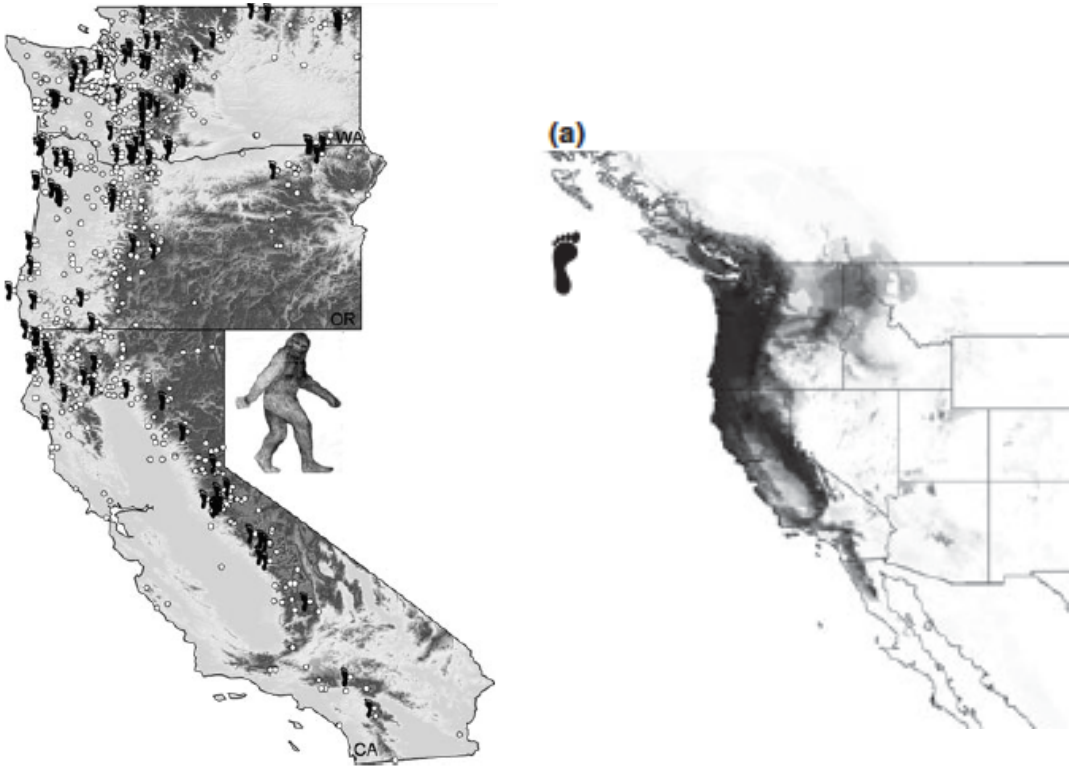
c) Erroneous taxonomy

Journal of Biogeography (J. Biogeogr.) (2009) **36**, 1623–1627

GUEST EDITORIAL

Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling

J. D. Lozier[1]*, P. Aniello[2] and M. J. Hickerson[3]



3.3 Issues with repository data

c) Erroneous taxonomy

Journal of Biogeography (J. Biogeogr.) (2009) **36**, 1623–1627

GUEST EDITORIAL

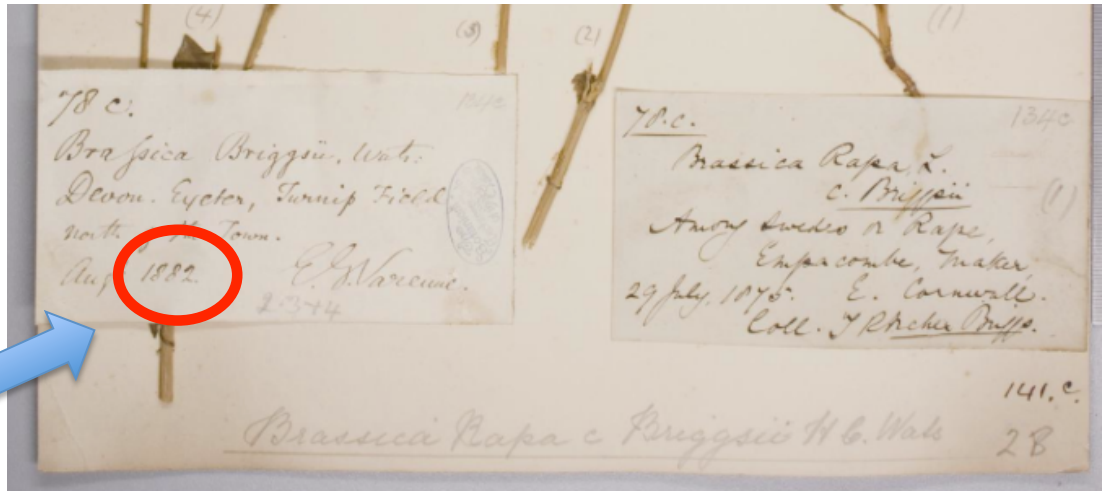
Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling

J. D. Lozier[1]*, P. Aniello[2] and M. J. Hickerson[3]

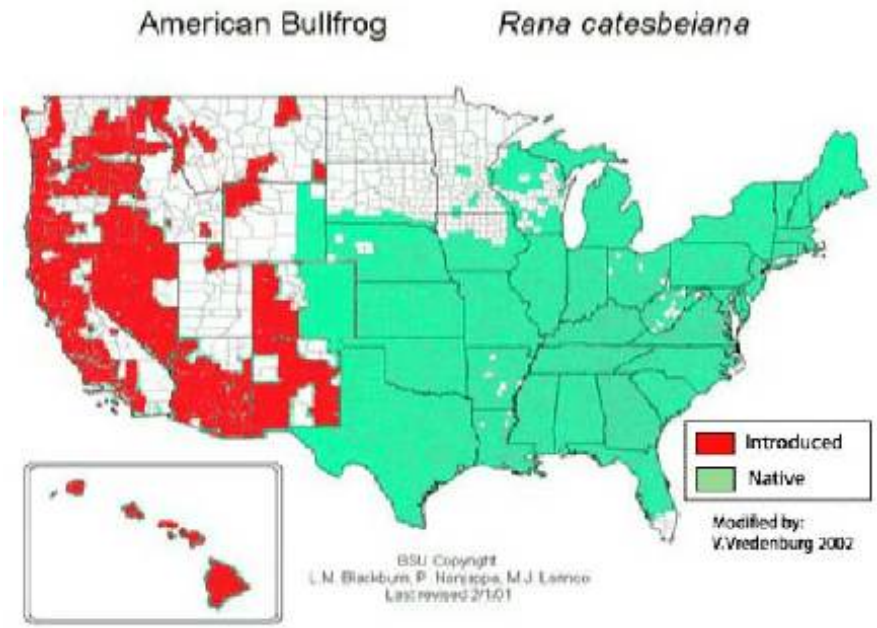
Note: In addition to errors in taxonomy, subspecies may not be correctly identified

3.3 Issues with repository data

d) Relevance



Record from 1882: include or exclude?



Record from introduced range: include or exclude?

What are you trying to model?

3.4 Georeferencing

When there aren't exact coordinates associated with an occurrence record, we can try to estimate them from the description of the locality…

This is called georeferencing

We may also want to georeference sites to verify that coordinates are correct or to get estimates of error in the coordinates

3.4 Georeferencing

Time for a Demo!

3.4 Georeferencing

Example 1

Lucus County, Ohio: 0.1 mi. N. of preserve office, Kitty Todd Preserve, NW., jct. of Eber & Old State Line Rds,

Example 2

Cass Co. Minnesota: Near Dry Sand Lake Wildlife Management Area; T135N R32W NE1/4 of NE1/4 of Sec 8.

3.4 Georeferencing

Some Online Tools

Google Earth with Township-Section-Range extension:

https://www.google.com/earth/
http://www.earthpoint.us/Townships.aspx

Geolocate:
http://www.museum.tulane.edu/geolocate/

Google Maps Geocoding:
https://developers.google.com/maps/documentation/geocoding/intro

A tool for estimating precision:
http://manisnet.org/gci2.html

3.5 Exercise and Break (90 minutes)

Exercise D1.3

As a class:

1) Everyone download and open the “bullfrogLoc_unfiltered.csv” file in Excel and examine the file
2) What challenges do we face cleaning these data? What *a priori* decisions do we want to make as a group for filtering records?

In small groups:

3) Georeference records 950 and 1133 and bring me your estimated coordinates.
4) Clean your assigned portion of the remainder of the locality database. Filter according to the decisions made as a group. If you have records missing lat/long info, georeference at least 10 of these.
5) Save your subset as new csv “bullfrogLocs_clean_subset.csv”

4. Environmental Data

4.1 Collecting environmental data

4.2 Exercise

4.3 Environmental data considerations

4.4 Exercise

4.1 Collecting Environmental Data

Ideally…

All relevant variables for the species included

Site-specific values or data collected at spatial scales representative of what individuals at each locality actually experience

Collected from time period consistent with when locality data were collected

Variation across years incorporated

4.1 Collecting Environmental Data

But usually we have to surrender our ideals…

~~All relevant variables for the species included~~
→ Can't know all variables relevant to species
→ Many variables not represented as GIS layers

~~Site-specific values or data collected at spatial scales representative of what individuals at each locality actually experience~~
→ Data often at very coarse resolution

~~Collected from time period consistent with when locality data were collected~~
→ Data often averaged across large chunk of time

~~Variation across years incorporated~~
→ Not usually done

4.2 Exercise (30 minutes)

Exercise D1.4

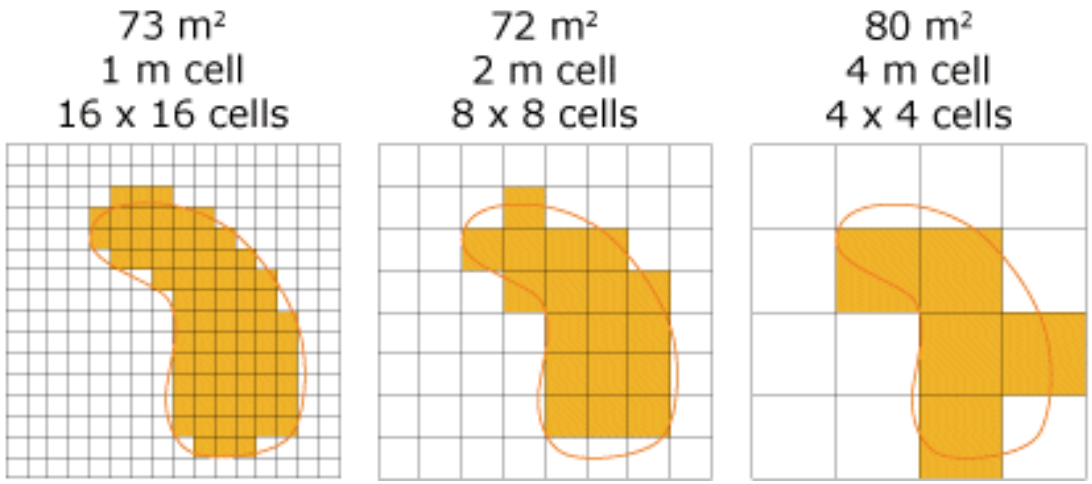
1) What types of environmental variables are you interested for your specific project? Working alone or in a small group, conduct an internet search to see if you can find GIS datasets representing these variables.

2) For each variable, note the region covered (e.g. county, country, continent)

3) When you have a list of data sources, come see me and we'll add them to the class resource page

4.3 Environmental Data Considerations

Resolution…
Governs information content and
scale of model application



4.3 Environmental Data Considerations

Missing Variables…
Governs what our models speak to and don't speak to

4.4 Exercise (40 minutes)

Exercise D1.5

1) Download the tutorial environmental layers. Open them in R using the raster package (instructions on website)

2) What do you notice about the layers? Any problems with them?

3) Follow the tutorial to get rasters ready for tomorrow's analysis.

Homework Checklist

1) Finish processing the locality data assigned to you

2) Finish preparing the environmental layers for analysis

   …which involves making sure you have R and the necessary libraries installed (stay for the next topic if you need help with R stuff)

Tomorrow we will:

→ Further preparing the data for our frog niche model

→ Learn how to build a niche model using MAXENT

5. (Optional) Getting ready to use R

Time for a Demo!