Climatic Processing Overview

Shelly Gaynor
University of Florida

FLORIDA MUSEUM

UF | UNIVERSITY of FLORIDA

iDigBio
Integrated Digitized Biocollections

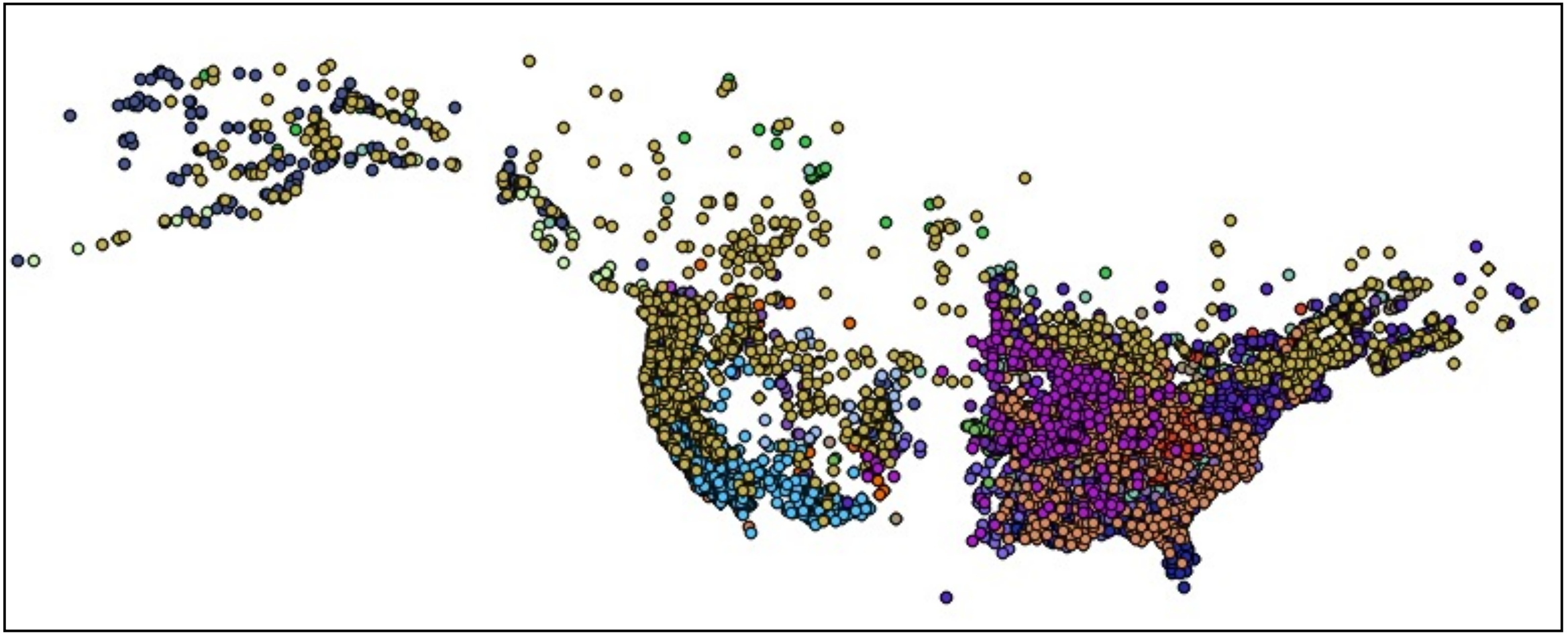
NSF

BiotaPhy

What data are we using?

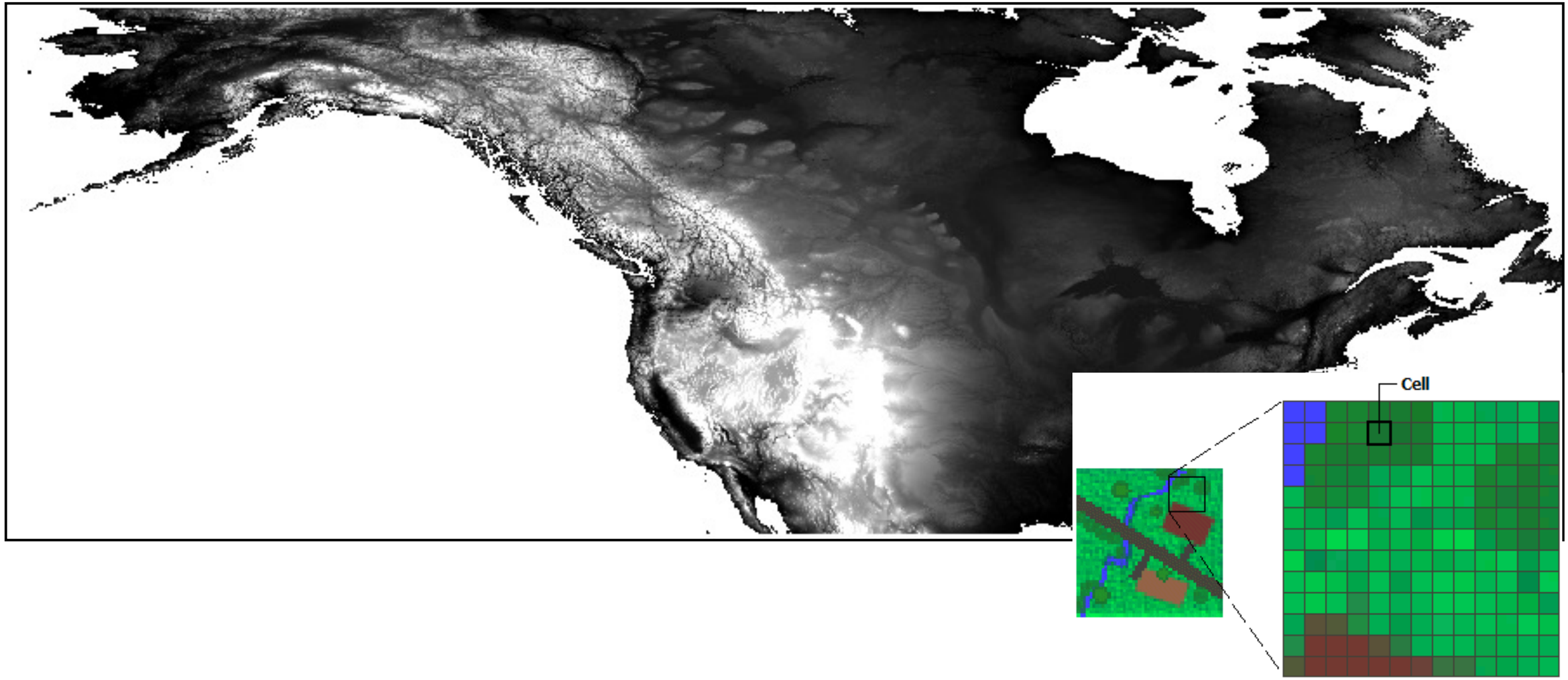
What data are we using?

1. Georeferenced occurrence records



What data are we using?

2. Gridded abiotic data layers (or Raster)



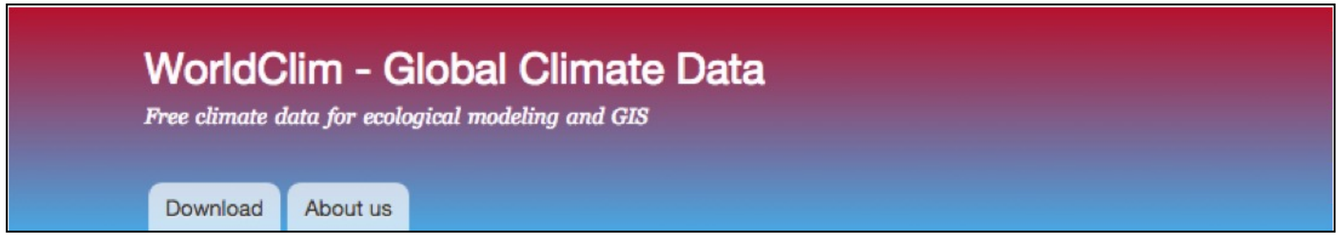
Source: ArcMap

Abiotic data layers

- What kinds?
  - Temperature, precipitation, soil, vegetation, land use
- Where?
  - Local, USA, North America, Global
- When?
  - Past, current, future

Data layer sources

- WorldClim
- PRISM
- Unified North American Soil Map
- USGS
- EPA
- NOAA
- PMIP
- AquaMaps
- And more!



rWBclimate
World

prism
World

weathercan
Canada

getCRUCLdata
World

clifro
New Zealand

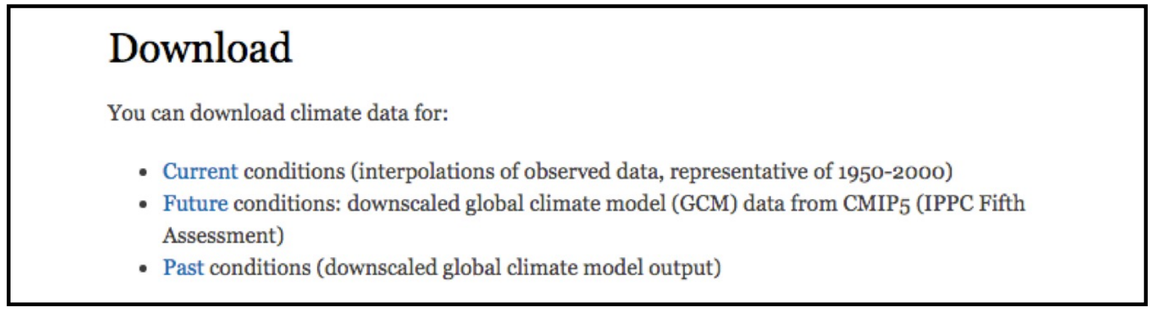
chirps
World

World
Current & Future & Past

WorldClim

www.worldclim.org

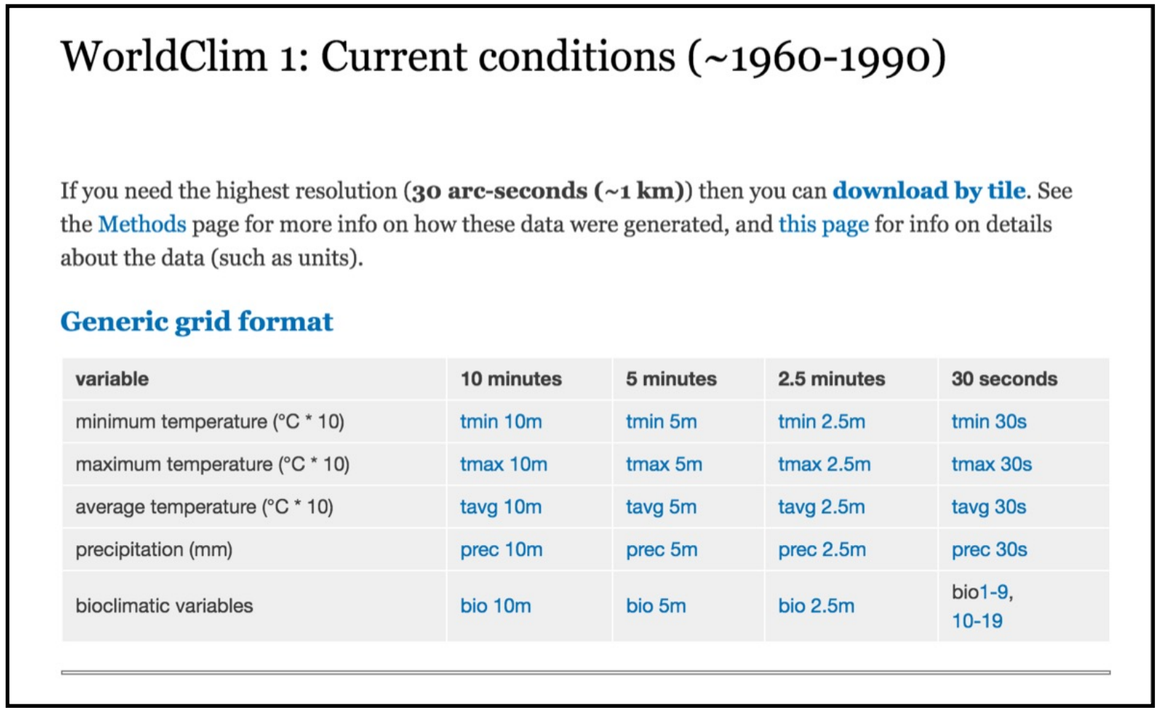
- Global climate data
- Multiple resolutions
- 19 Bioclimatic variables derived from monthly temperature and rainfall values
- Current conditions:
  - averaged over past 50 years
- Future conditions:
  - climate projections from a number of global climate models with different greenhouse gas concentration trajectories
- Past conditions:
  - Mid-Holocene, Last Glacial Maximum, Last Inter-Glacial
- WorldClim 2 now available, has same 19 + new variable layers



Download

You can download climate data for:

- Current conditions (interpolations of observed data, representative of 1950-2000)
- Future conditions: downscaled global climate model (GCM) data from CMIP5 (IPPC Fifth Assessment)
- Past conditions (downscaled global climate model output)



WorldClim 1: Current conditions (~1960-1990)

If you need the highest resolution (**30 arc-seconds (~1 km)**) then you can **download by tile**. See the Methods page for more info on how these data were generated, and this page for info on details about the data (such as units).

Generic grid format

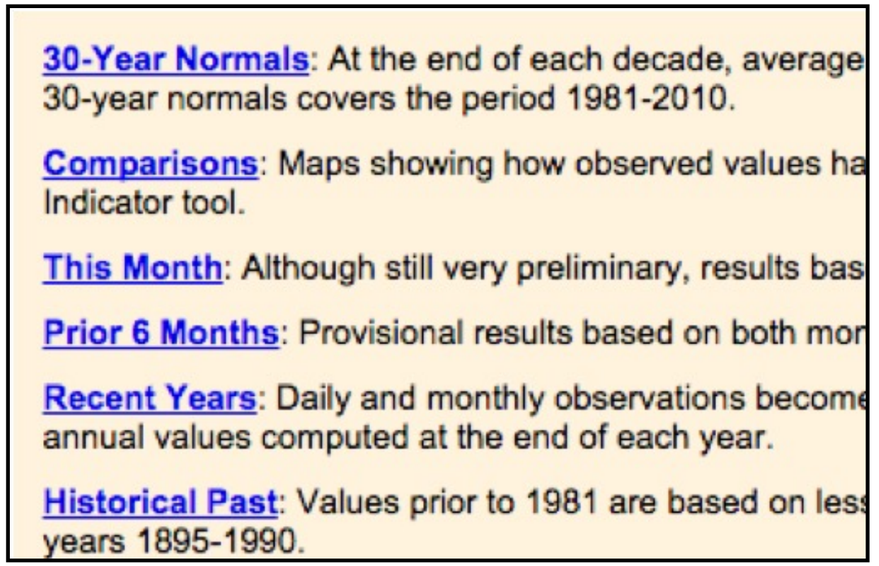
variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C * 10)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C * 10)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C * 10)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
bioclimatic variables	bio 10m	bio 5m	bio 2.5m	bio1-9, 10-19

Regional
Current & Past

PRISM

www.prism.oregonstate.edu

- Climate data for the US
- More precise than WorldClim
- Data available from 1895 to present
- Lots of data and tools to explore



30-Year Normals: At the end of each decade, average 30-year normals covers the period 1981-2010.

Comparisons: Maps showing how observed values ha Indicator tool.

This Month: Although still very preliminary, results bas

Prior 6 Months: Provisional results based on both mor

Recent Years: Daily and monthly observations become annual values computed at the end of each year.

Historical Past: Values prior to 1981 are based on less years 1895-1990.

High-Resolution Data Available

The native grid resolution of the PRISM datasets is 800m, but they have been filtered to 4km resolution for easier downloading and manipulation on this website. The 800m versions of all PRISM datasets, which contain 25x more information, are available to users for a fee. Details on availability, pricing, and ordering are found here.

Soil Sources

- **SoilGrids2.0: https://soilgrids.org**
  - Entire globe
  - Eight characteristics of topsoil and subsoil
  - 2.5 arc second resolution

SoilGrid Download

SoilGrid v2.0 Download
July 2020
ML Gaynor
With lots of help by: Mike Belitz (Github: Mbelitz)

Set working directory

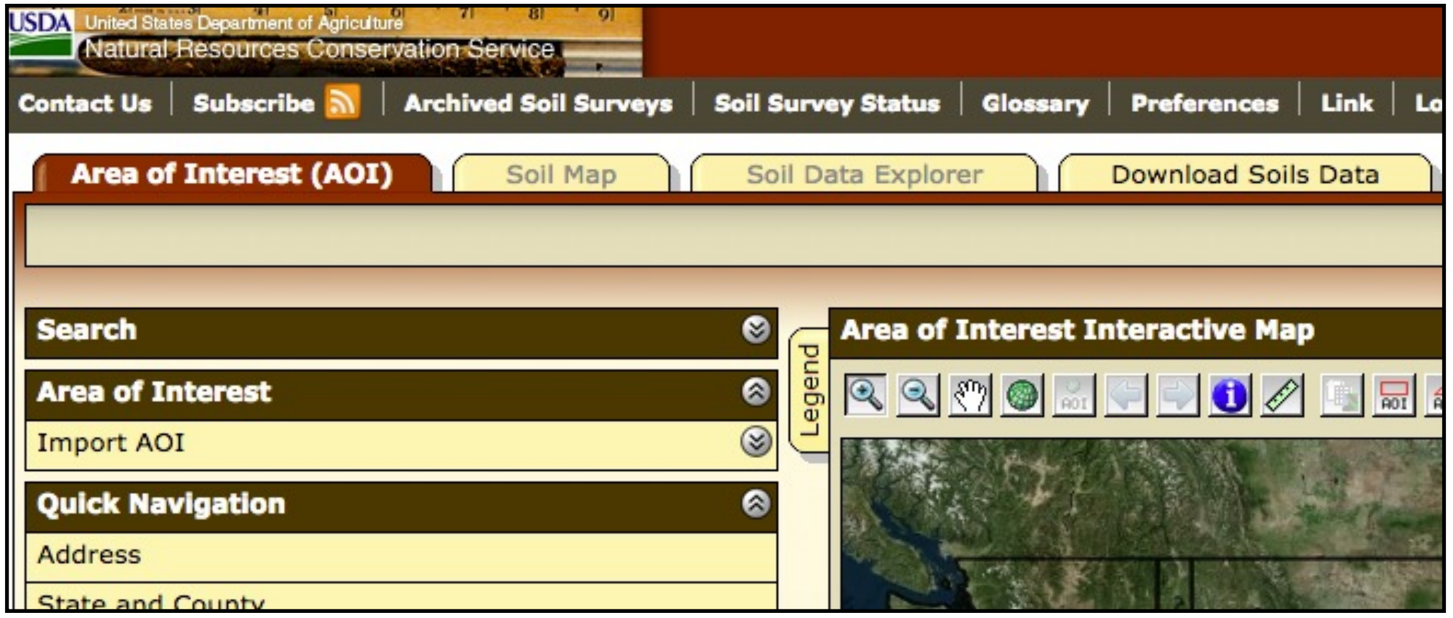
Set the working directory, here I just checked to see where I was and made sure this is where I wanted to save the files.

]: pwd

mgaynor1/long-winded-scripts/SoilGridDownload/

Soil Sources

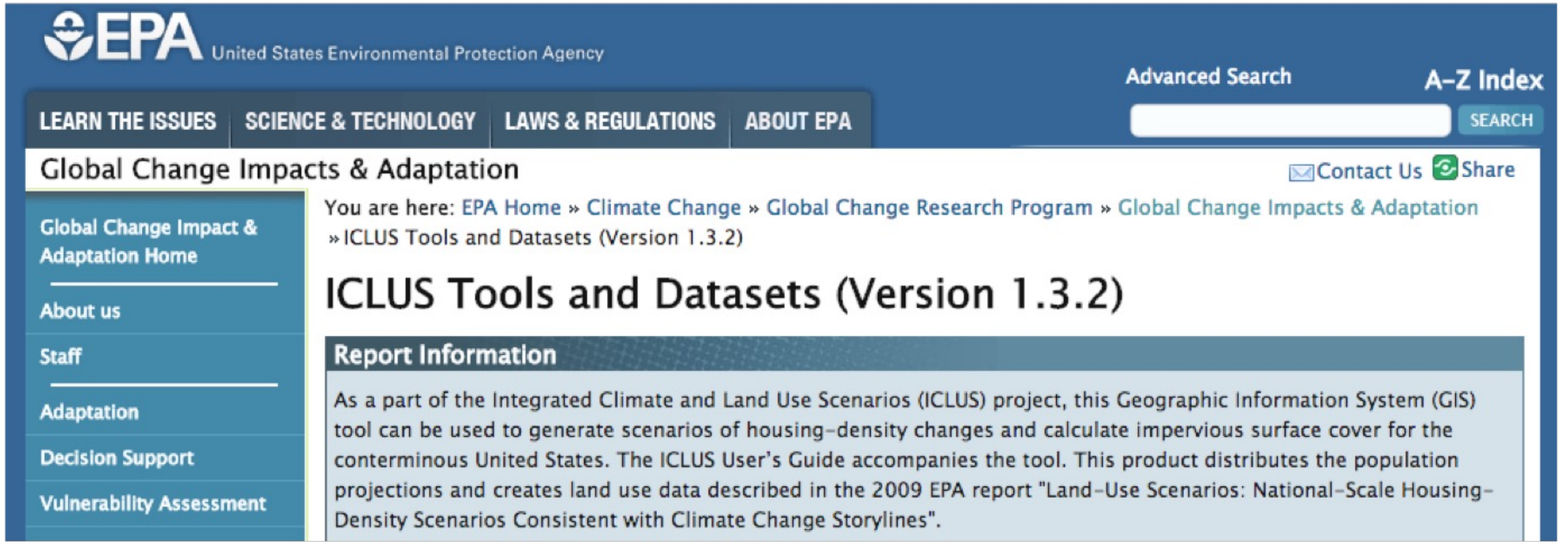
- **SoilGrids2.0: https://soilgrids.org**
  - Entire globe
  - Eight characteristics of topsoil and subsoil
  - 2.5 arc second resolution
- **USGS: http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm**
  - **US only**
  - **Need to define area of interest**
  - **Big files and tough to get large areas**



EPA

http://cfpub.epa.gov/ncea/global/recordisplay.cfm?deid=257306

- Land use
  - County population projections
  - Housing density projections
  - Percent impervious surface projections



Other resources

Paleoclimate data

- NOAA (NCEI):https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets
- PMIP: https://pmip.lsce.ipsl.fr/
- PaleoClim: http://paleoclim.org/

Aquatic Environments

- AquaMaps: http://www.aquamaps.org/main/envt_data.php

Hourly or Daily data:

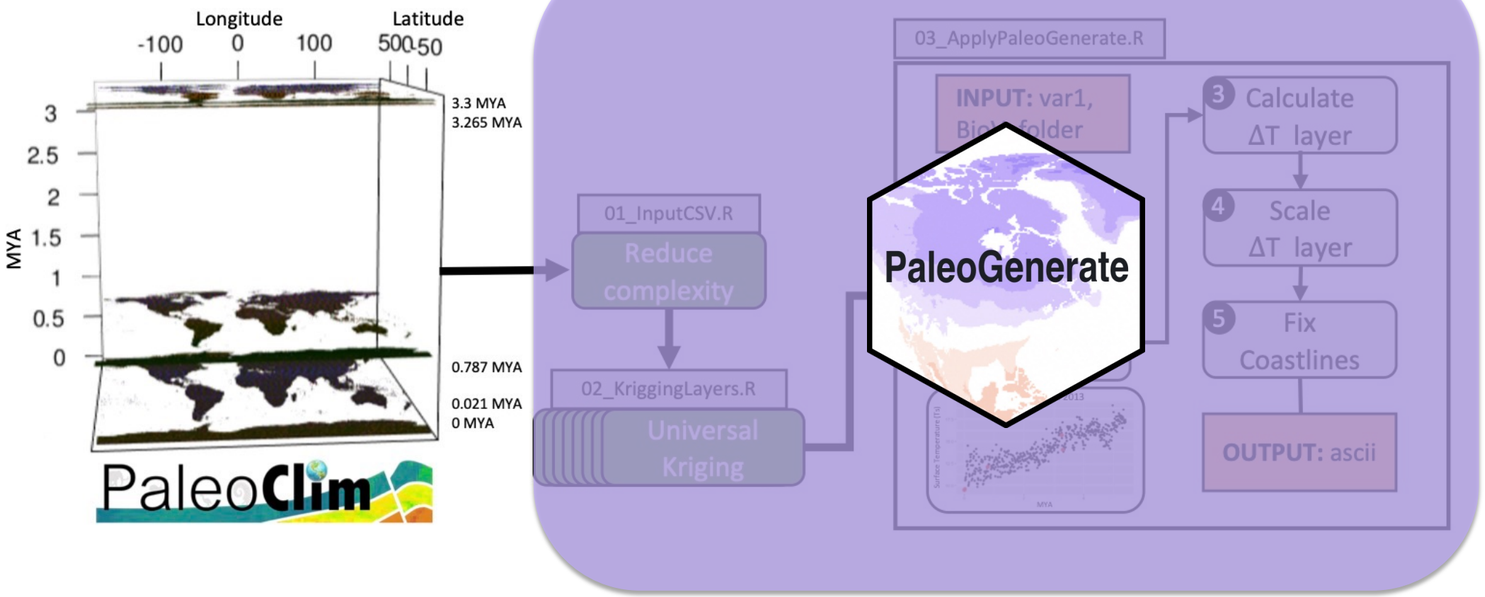
- OGIMET: ogimet.com
- Daymet: https://daymet.ornl.gov/

Syst. Biol. XX(XX):XX–XX, 2023

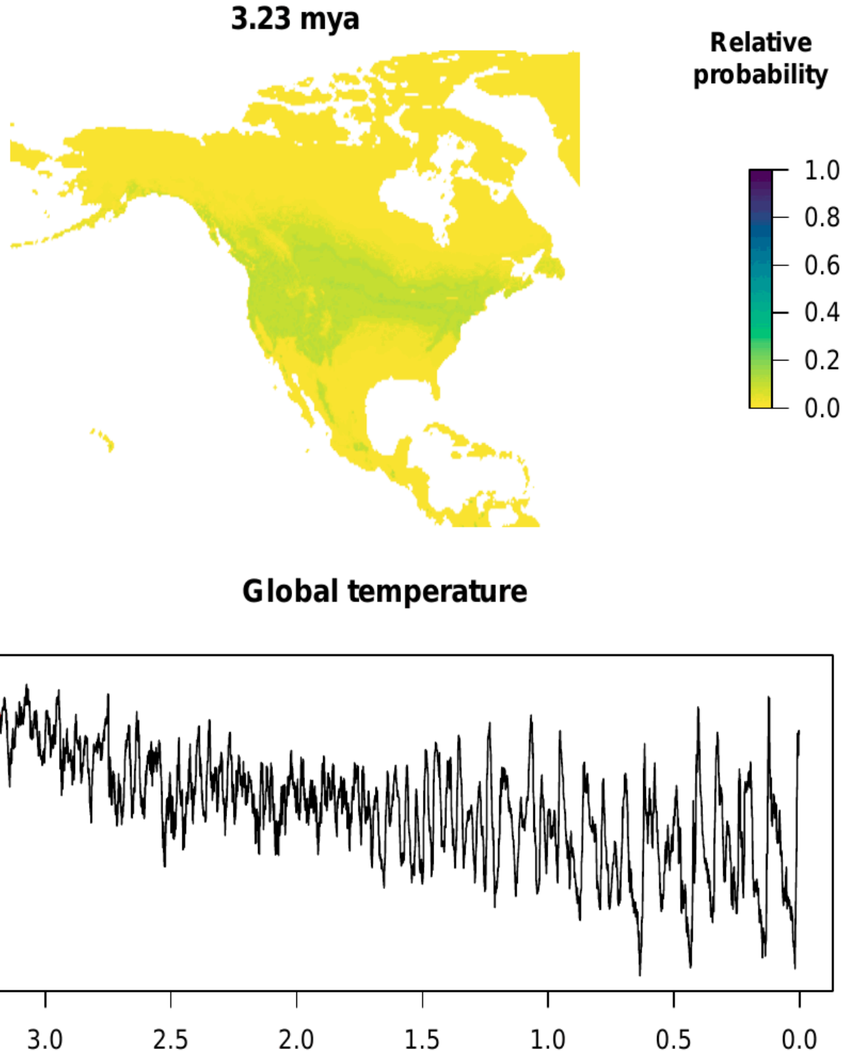
https://doi.org/10.1093/sysbio/syad018
Advance Access Publication April 19, 2023

Identifying Climatic Drivers of Hybridization with a New Ancestral Niche Reconstruction Method

Ryan A. Folk[1,*], Michelle L. Gaynor[2,3], Nicholas J. Engle-Wrye[1], Brian C. O'Meara[4], Pamela S. Soltis[2,5,6], Douglas E. Soltis[2,3,5,6], Robert P. Guralnick[2,6], Stephen A. Smith[7], Charles J. Grady[8] and Yudai Okuyama[9]

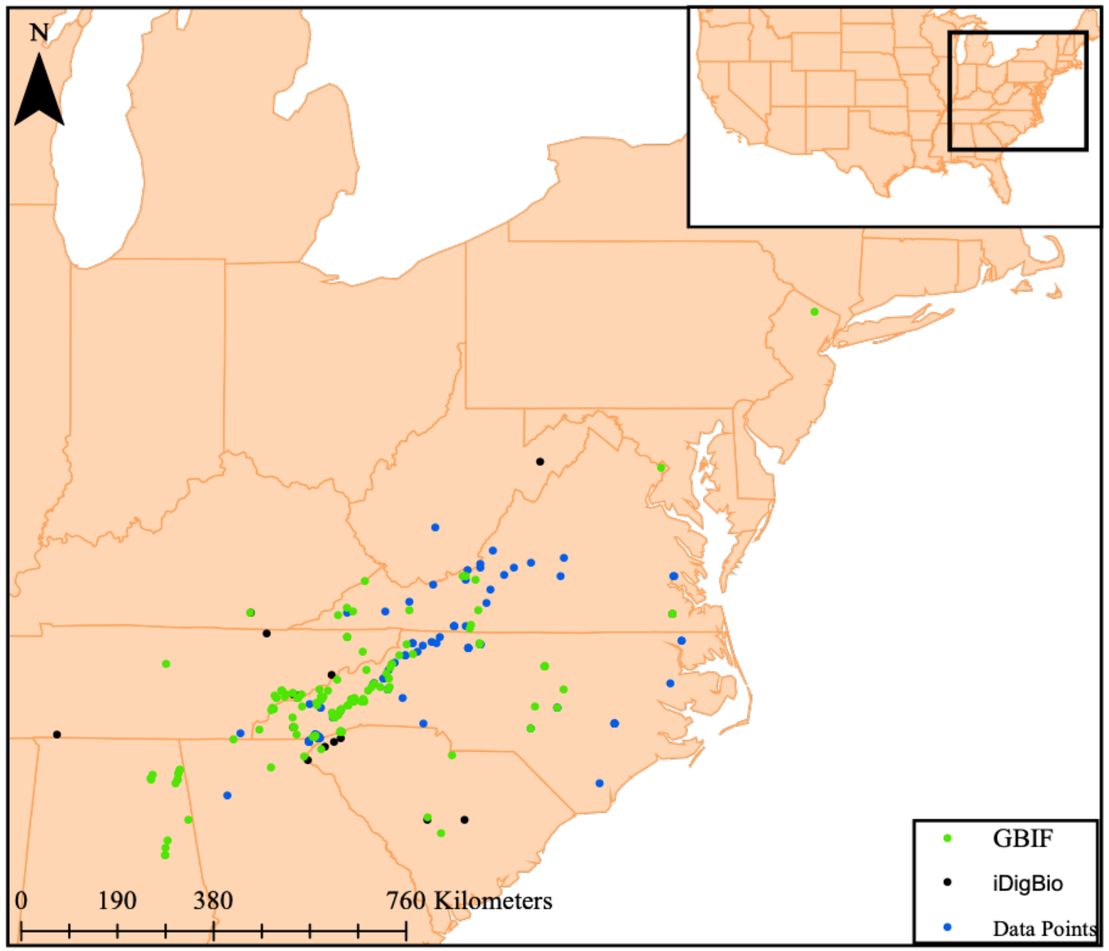


mgaynor1/paleogenerate



Folk & Gaynor et al. 2023. Systematic Biology.

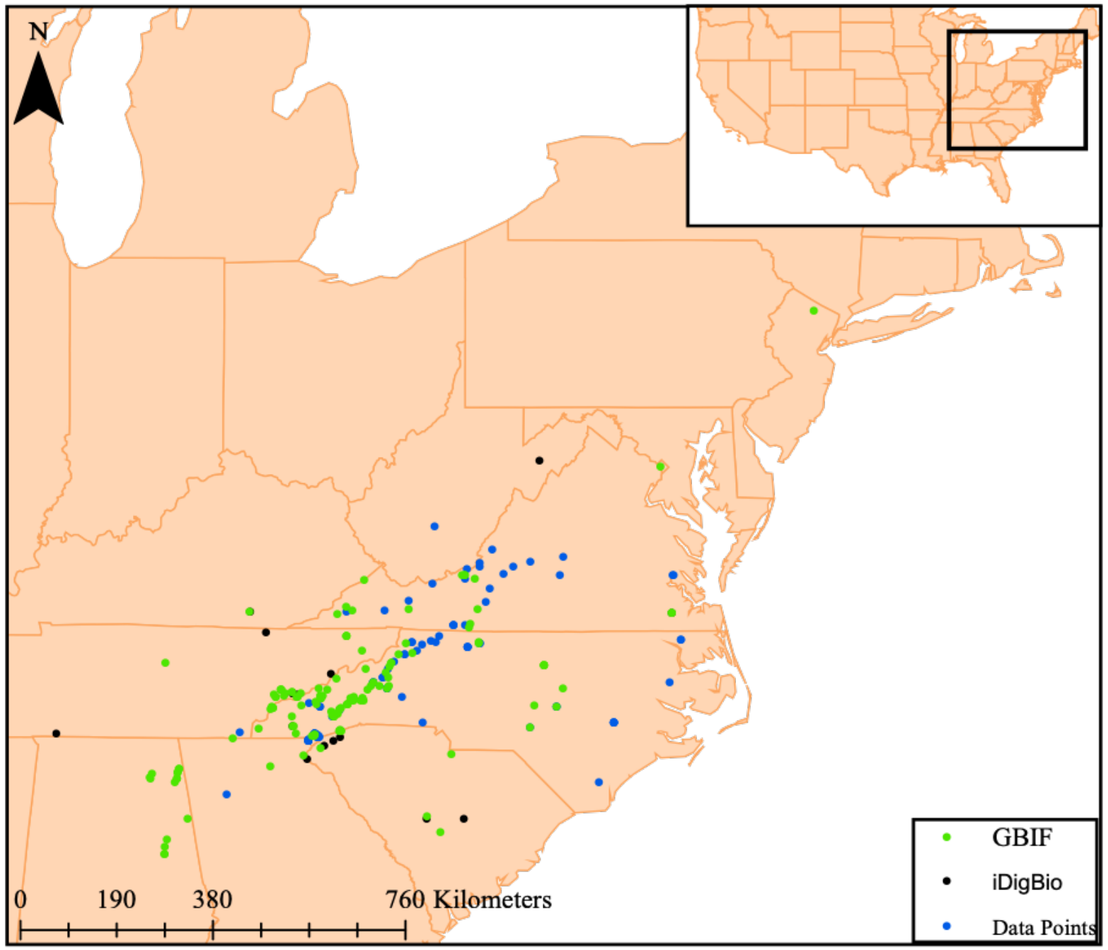
Cleaning Climatic Layers



Gaynor et al. 2018

Cleaning Climatic Layers

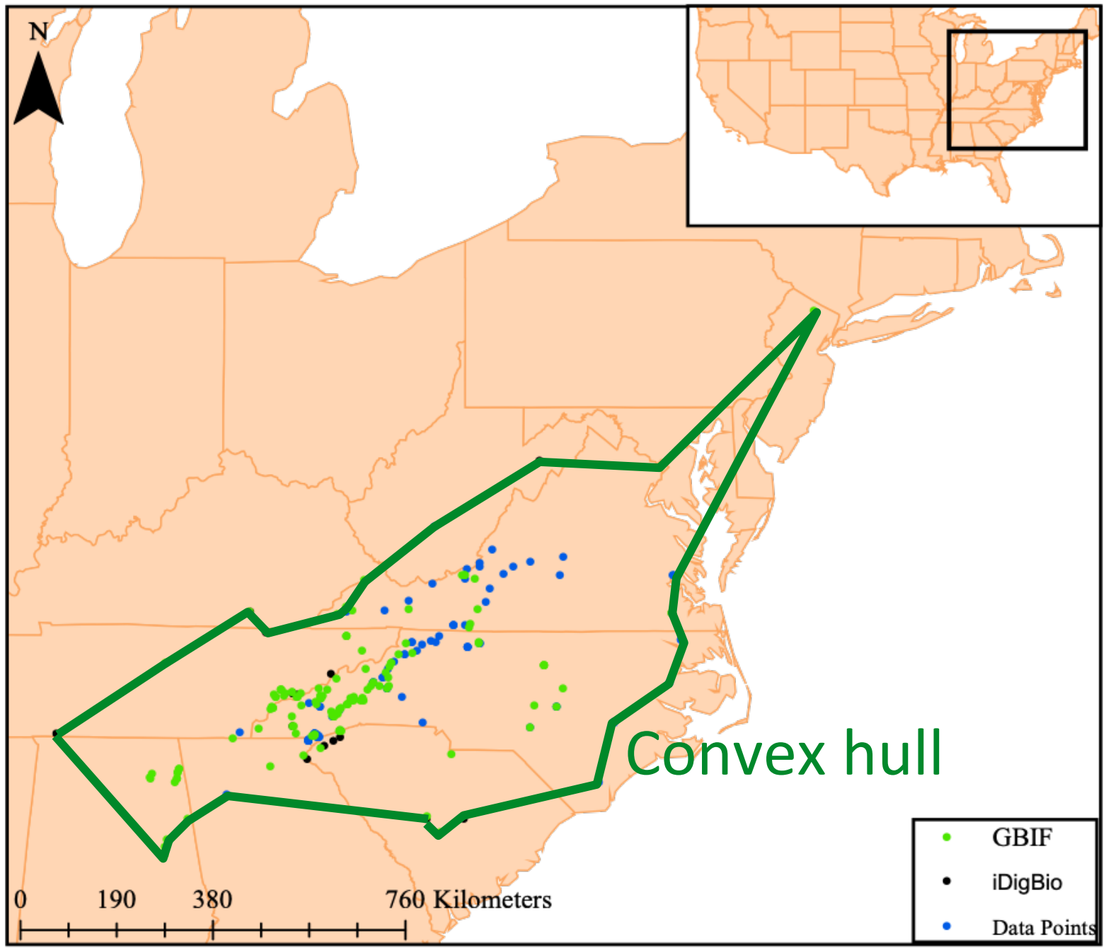
1. Obtain gridded abiotic data layers
2. Crop layers to the shared accessible area
3. Select layers for ecological niche modeling
4. Create species-specific training layers



Gaynor et al. 2018

Cleaning Climatic Layers

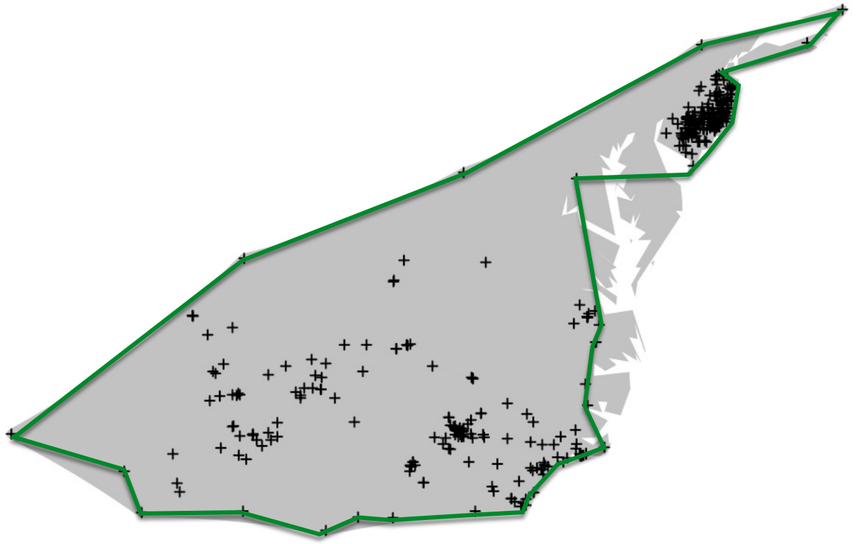
1. Obtain gridded abiotic data layers
2. Crop layers to the shared accessible area



Gaynor et al. 2018

2. Crop layers to the shared accessible area

- Accessible area can be defined based on:
  - species ecology
  - dispersal ability
  - geographical barriers

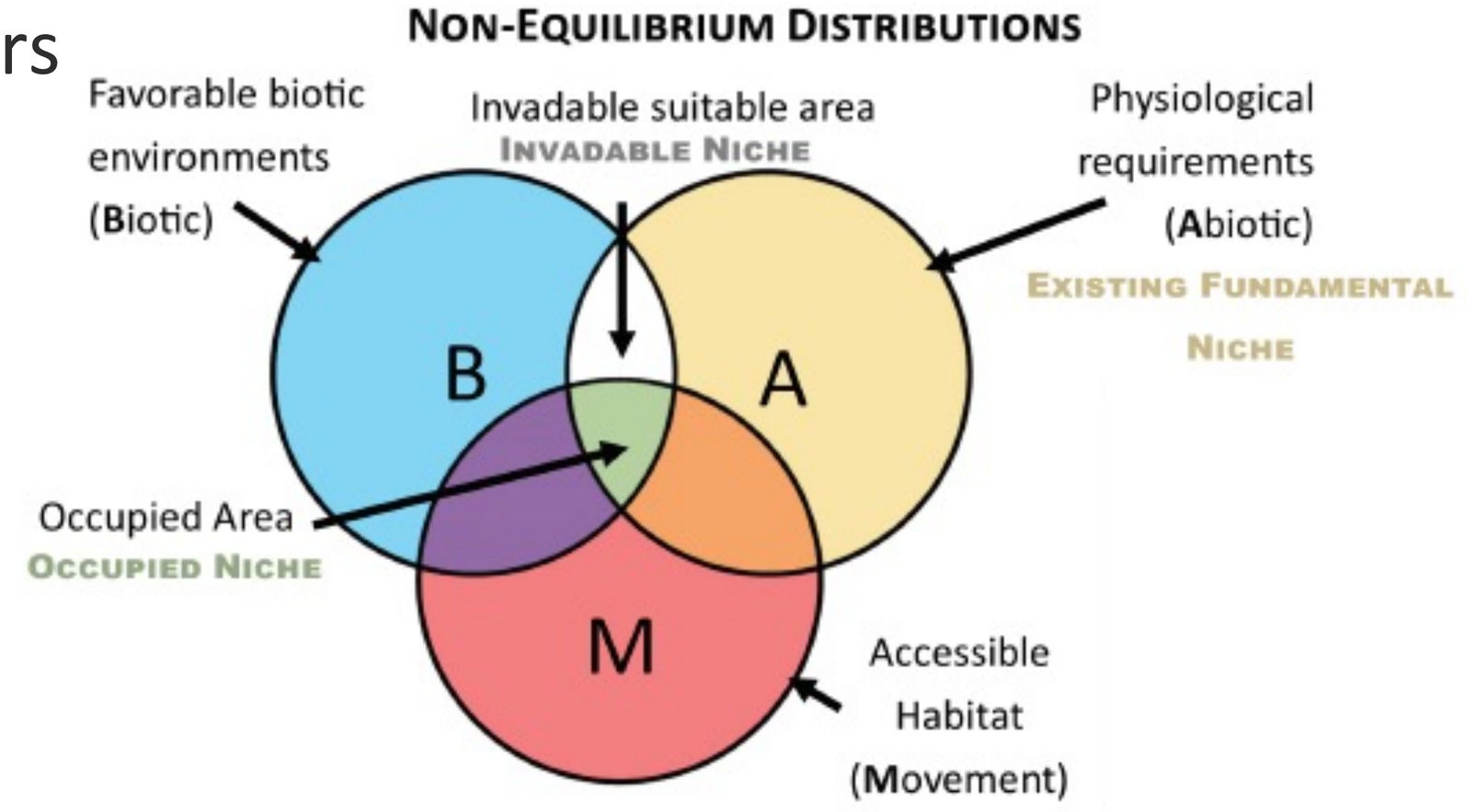


alpha-hull: piecewise series of linear simple curves

convex-hull is a type of alpha-hull

2. Crop layers to the shared accessible area

- Accessible area can be defined based on:
  - species ecology
  - dispersal ability
  - geographical barriers



Brown and Carnaval. 2019. A tale of two niche: methods, concepts, and evolution. Frontiers of Biogeography.

2. Crop layers to the shared accessible area

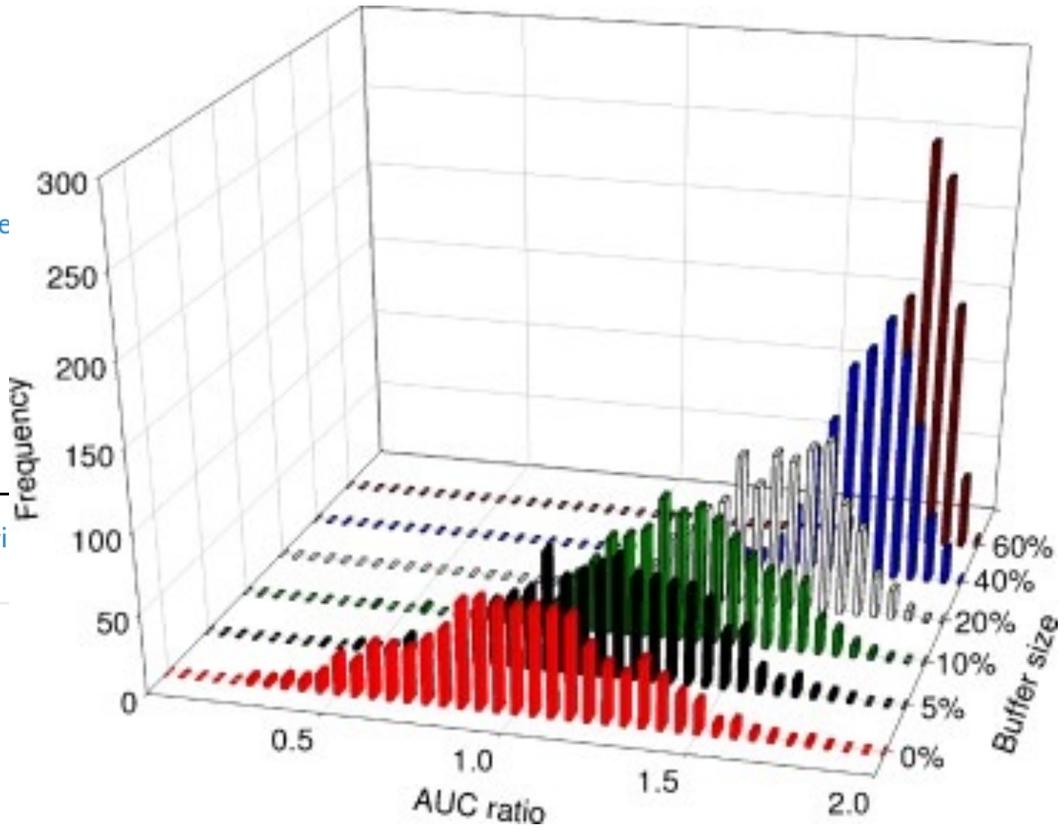
The crucial role of the accessible area in ecological niche modeling and species distribution modeling

Narayani Barve [a], Vijay Barve [a, 1], Alberto Jiménez-Valverde [a, 1], Andrés Lira-Noriega [a], Sean P. Mahe
Townsend Peterson [a], Jorge Soberón [a], Fabricio Villalobos [b]

Show more

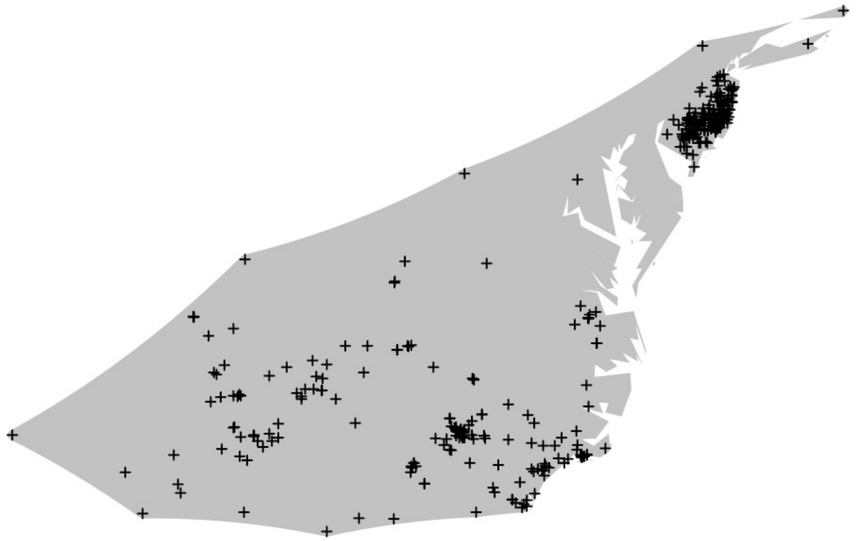
+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ecolmodel.2011.02.011 Get ri

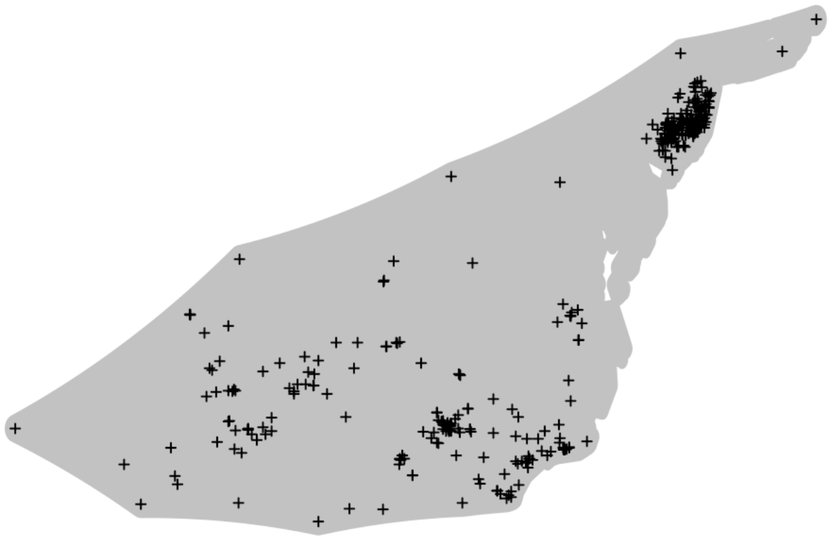


More area, AUC increased.
higher model significance as buffer size increased

2. Crop layers to the shared accessible area



alpha-hull



alpha-hull + buffer

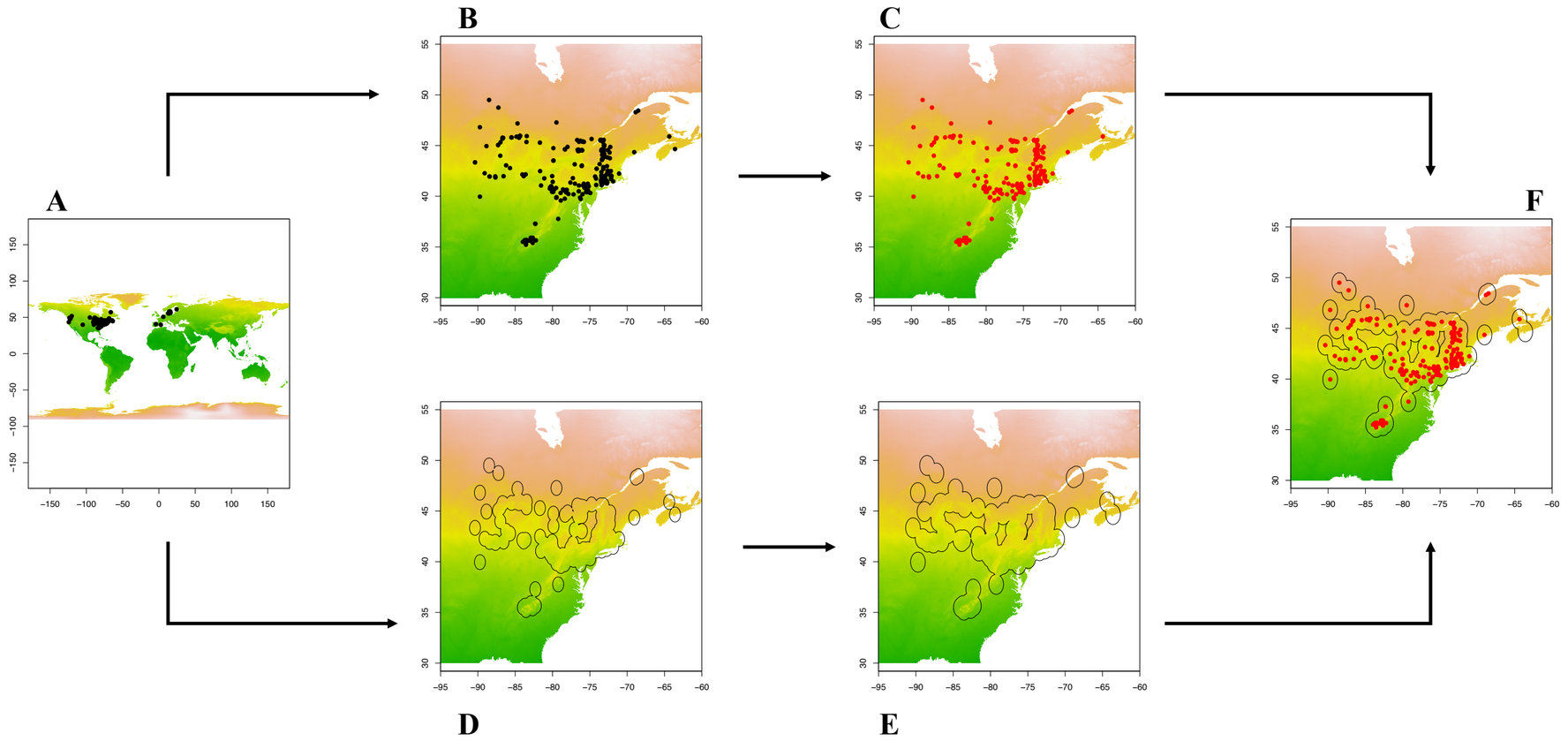
2. Crop layers to the shared accessible area

Defining the buffer:

- Identify clear distributional boundaries and ecoregions using the World Wildlife Fund Terrestrial Ecoregions (see Rautsaw et al. 2022.)
- Buffer size species-specific and based on the dispersal potential (see Naranjo et al. 2022)
- 80^{th} quantile of the max distance between points (here)

2. Crop layers to the shared accessible area

- Point-based buffers!



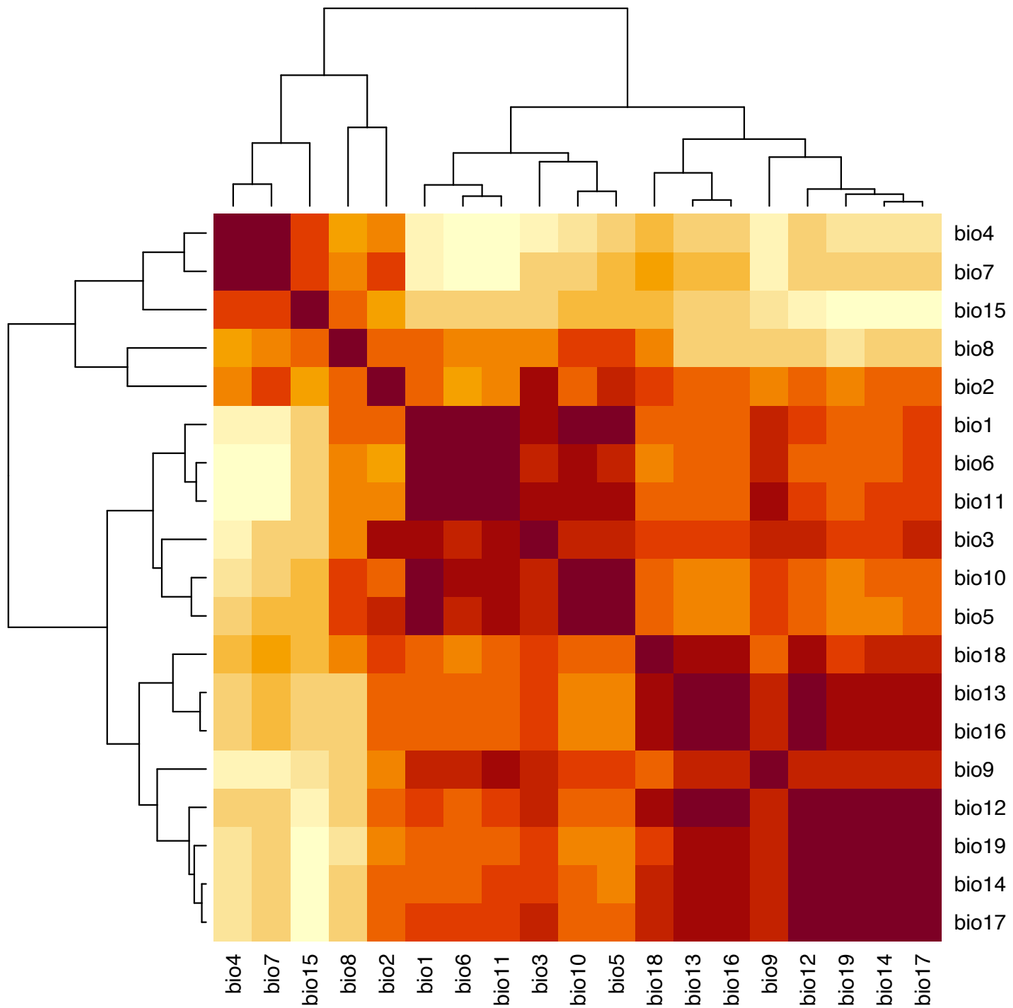
Many options for defining accessible area

Melton et al. 2022. Global Ecology and Biogeography.

3. Select layers for ecological niche modeling

PCA	PC1 + PC2 = ~95%	Reduces variables to axis
Pearson correlation	Removes variables > |0.80|	Removes correlated variables
Variable Inflation Factor	Removes variables > 10	Removes collinear variables

3. Select layers for ecological niche modeling



Layers used in ENM development

bio2
bio3
bio5
bio7
bio8
bio9
bio15
bio18
bio19

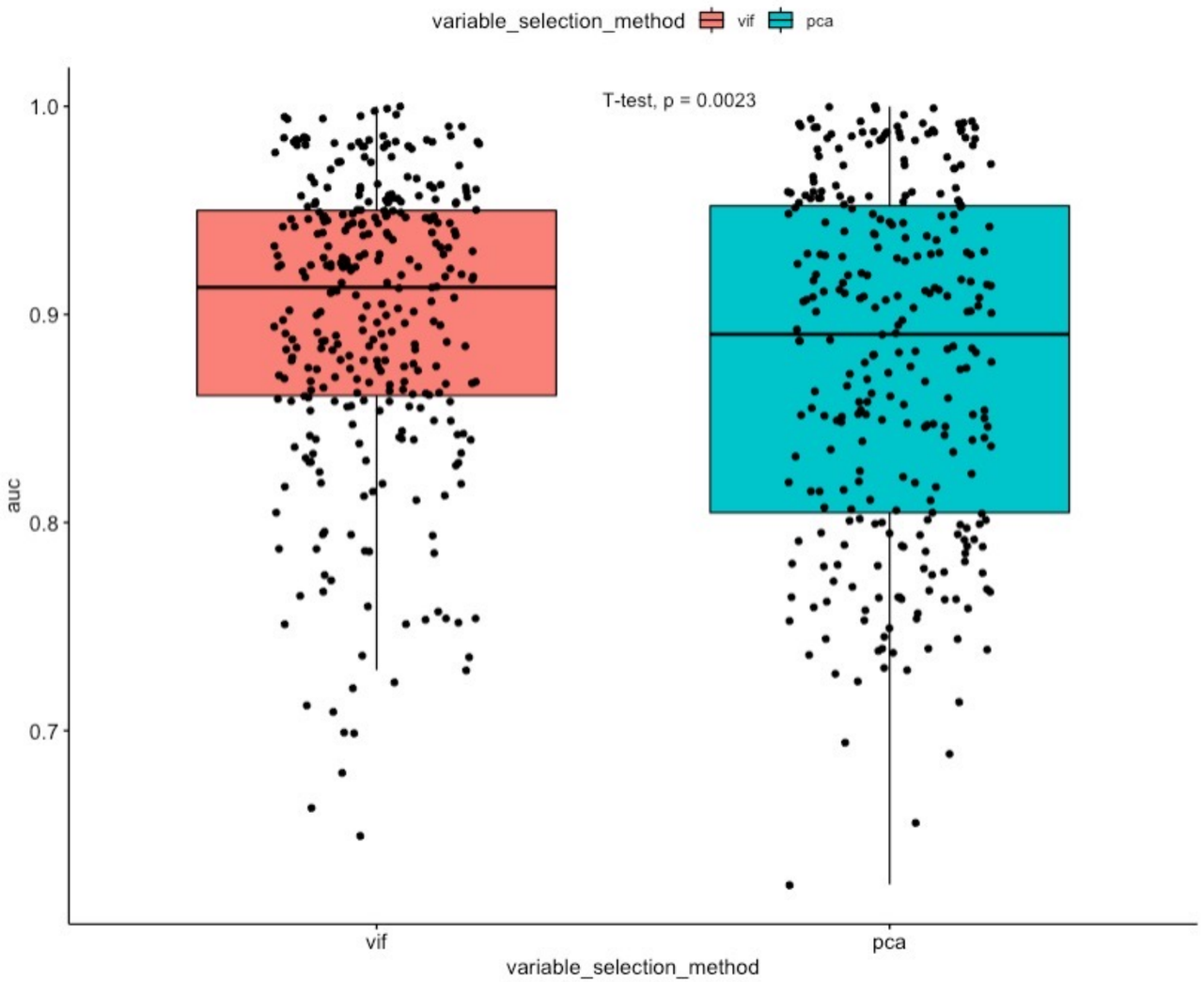
- A correlation test is performed on all 19 BioClim layers.
- Removes variables > |0.80|

Melton et al. 2022. Global Ecology and Biogeography.

3. Select layers for Ecological Niche Modeling

PCA vs VIF with permutation

- **AUC was significantly higher in models generated using the VIF method**
- Based on 109 species in the subfamily Crotalinae (Viperidae)
- PCA with 75% collinearity cutoff, each variable contribution had to be >5%.
- VIF < 10, permutation importance > 1%



Rautsaw et al. 2022. Scientific Data.

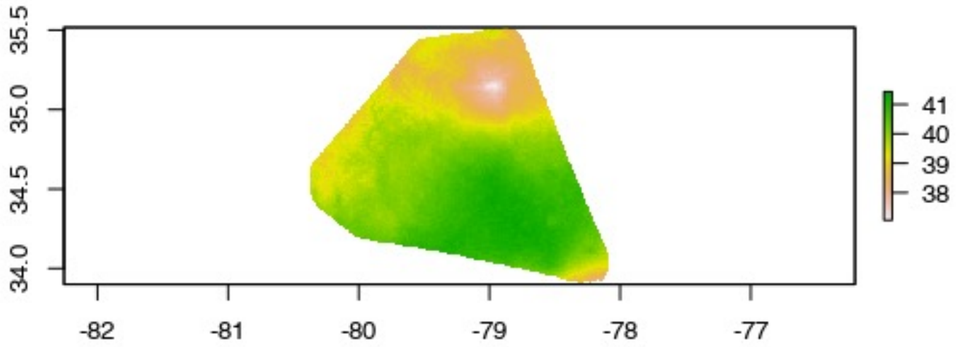
3. Select layers for ecological niche modeling

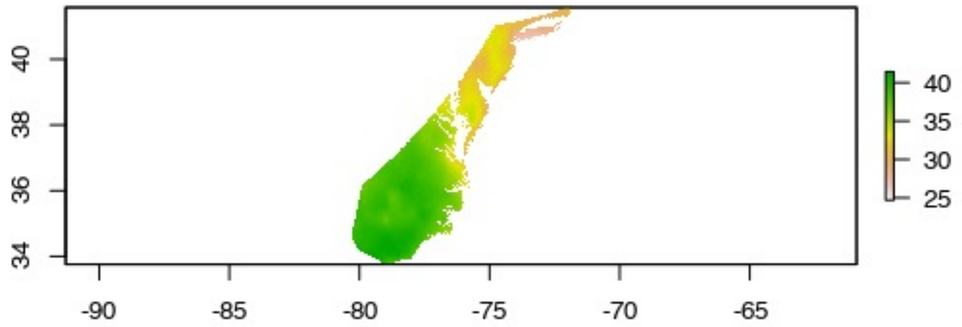
Demo:

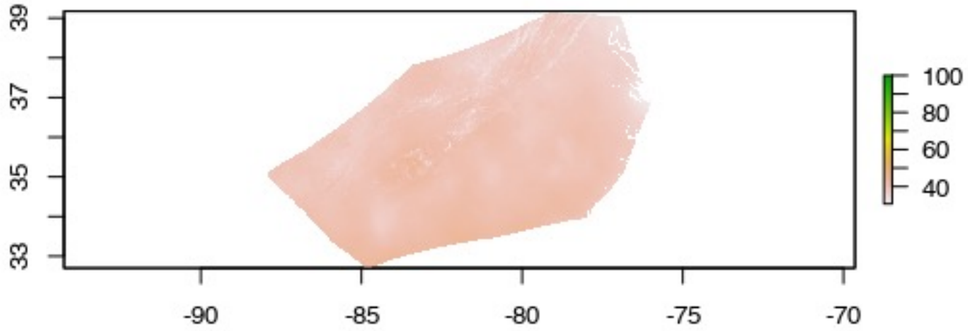
1. Pearson Correlation
    - Removes variables > |0.80|
2. VIF and permutation importance
    - Run MaxEnt and identify the permutation importance to remove any variables with 0% permutation importance.
    - Then calculated VIF, remove the variable with the highest VIF and lowest permutation importance. Repeat until all VIF values are less than 10.

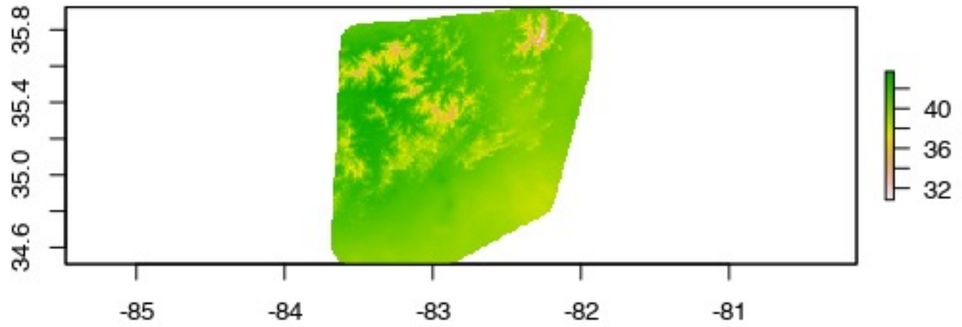
4. Create species-specific training layers

- Crop and buffer selected layers for each species!









Layer Processing

- Need layers in ASCII format
- Clip layers to fit your desired range
  - Convex hull + Buffer
- Remove correlated layers based on a Pearson Correlation.

Layer Processing

- For an intro to QGIS:
  - Github: RhettRautsaw/GIS_Tutorial

CLIMATE LAYER PROCESSING

QGIS provides a much better understanding of the processes happening with this step, but the R script streamlines a largely repetitive process. We will demo both options.

(A) Manual – QGIS - Optional

This activity was made by Rhett Rautsaw. Files for this activity can be found in *"Demo/Manual/Climate_Layer_Processing/"* folder.

QGIS version has to be 3.16; if not, this will not work

QGIS (version: 3.16)

1. Open QGIS.
2. Drag the layers (.tif files) found in the *"data/climate_processing/bioclim/"* folder into QGIS. They should automatically appear. The box on the left lists the different layers not the layer is displayed.
3. Add occurrence records from text-delimited file (Layer Menu > Add Layer > Add Delimited Text Layer…). Nagivate to *"data/cleaning_demo/maxent_ready/diapensiaceae_maxentready_20210625.csv"*. X field is "longitude" and Y field is "latitude". Make sure the CRS is

Layer Processing

- Need layers in ASCII format
- Clip layers to fit your desired range
  - Dynamic alpha hull
    - + 80th quartile buffer

04_ClimateProcessing.R

Layer Processing

- Need layers in ASCII format
- Clip layers to fit your desired range
  - Dynamic alpha hull
    - + 80th quartile buffer
- Remove layers with collinearity based on variable inflation factor and permutation importance.

04_ClimateProcessing.R

Layer Processing

- Need layers in ASCII format
- Clip layers to fit your desired range
  - Dynamic alpha hull
    - + 80th quartile buffer
- Remove layers with collinearity based on variable inflation factor and permutation importance.
- Define **shared projection layers** and **training layers**.

04_ClimateProcessing.R

Layer Processing

04_ClimateProcessing.R