

Ecological Niche Modeling Crash Course

Blaine Marchant

PhD Student, Soltis Lab

FLMNH and Biology Department

University of Florida

dbmarchant@ufl.edu



iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. All images used with permission or are free from copyright.

National Center for Digitization of Biodiversity Collections

Collaboration among FLMNH, Engineering, & FSU

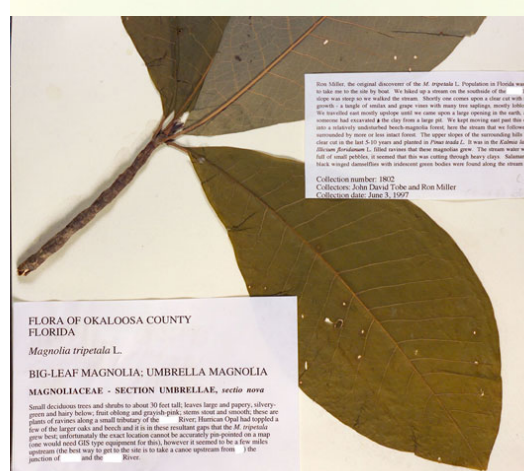
Coordinate digitization and databasing of US collections

Ingest, serve, integrate data:

Localities

Dates

Images



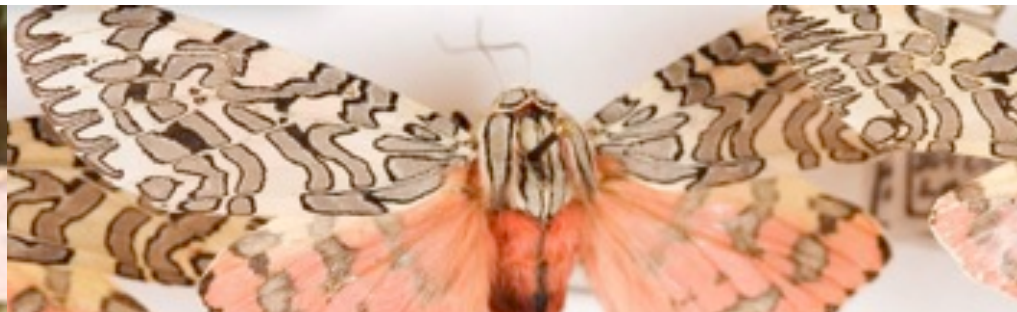
Using Museum Specimens and Computer Models in Biodiversity Studies

- Herbaria important sources of information on past and present species distributions
- Location information and environmental data
 - temperature, precipitation, soil
- Software to model the range of each species
- Project onto other areas, past & future climate conditions



'Big Data' Research in Biodiversity Science

- Monitoring shifts in biodiversity
- Tracking invasive species
- Ecological Niche Modeling, climate change
- Tracking phenological shifts
- Integration of ENM with phylogeny
- Past movements and climate change
- Landscape genetics
- Community phylogenetics/assembly

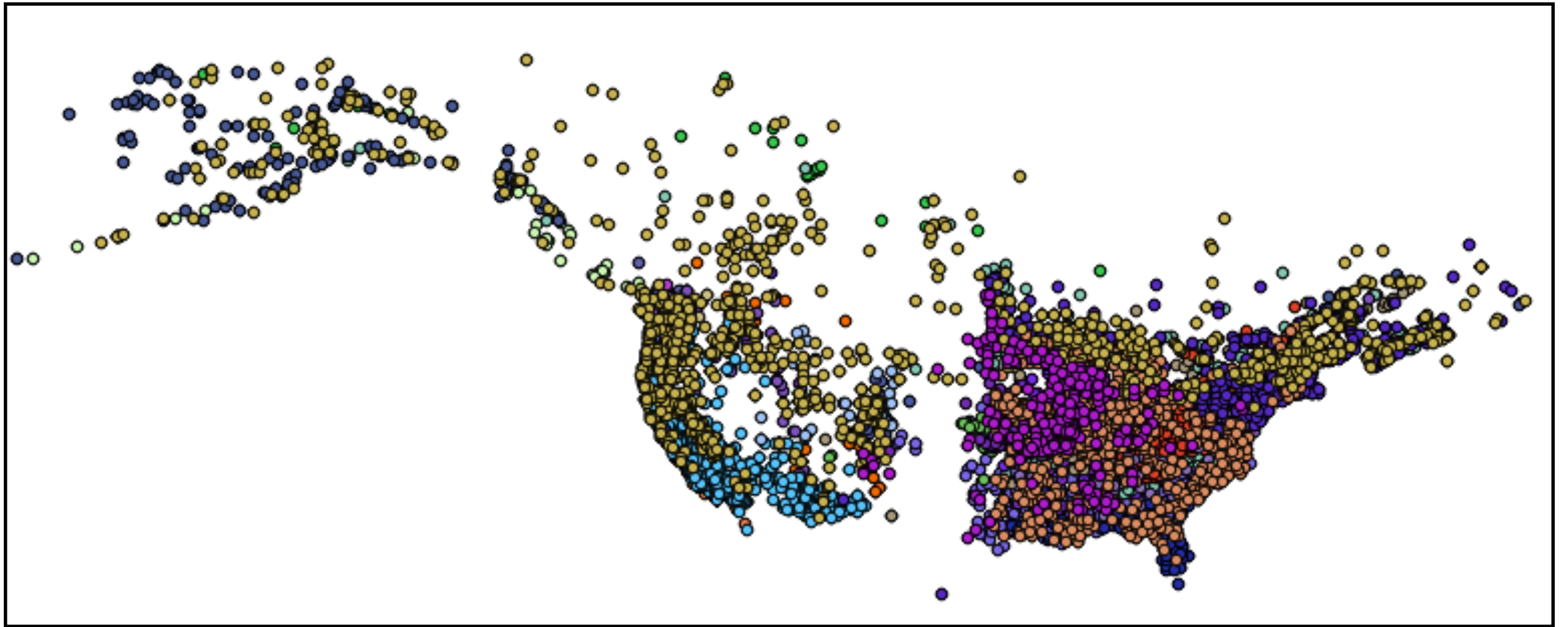


What is ecological niche modeling?

The use of associations between environmental variables and known species occurrence records to define conditions within which populations can be maintained

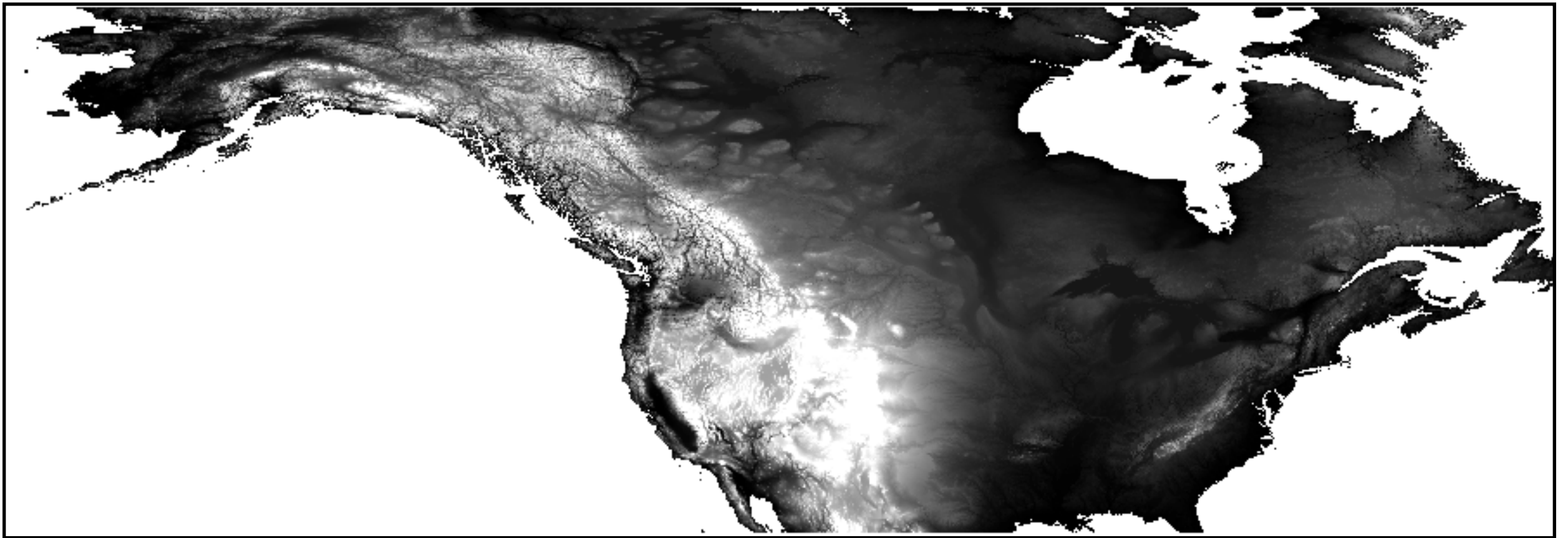
What data are we using?

What data are we using?



1. Georeferenced occurrence records

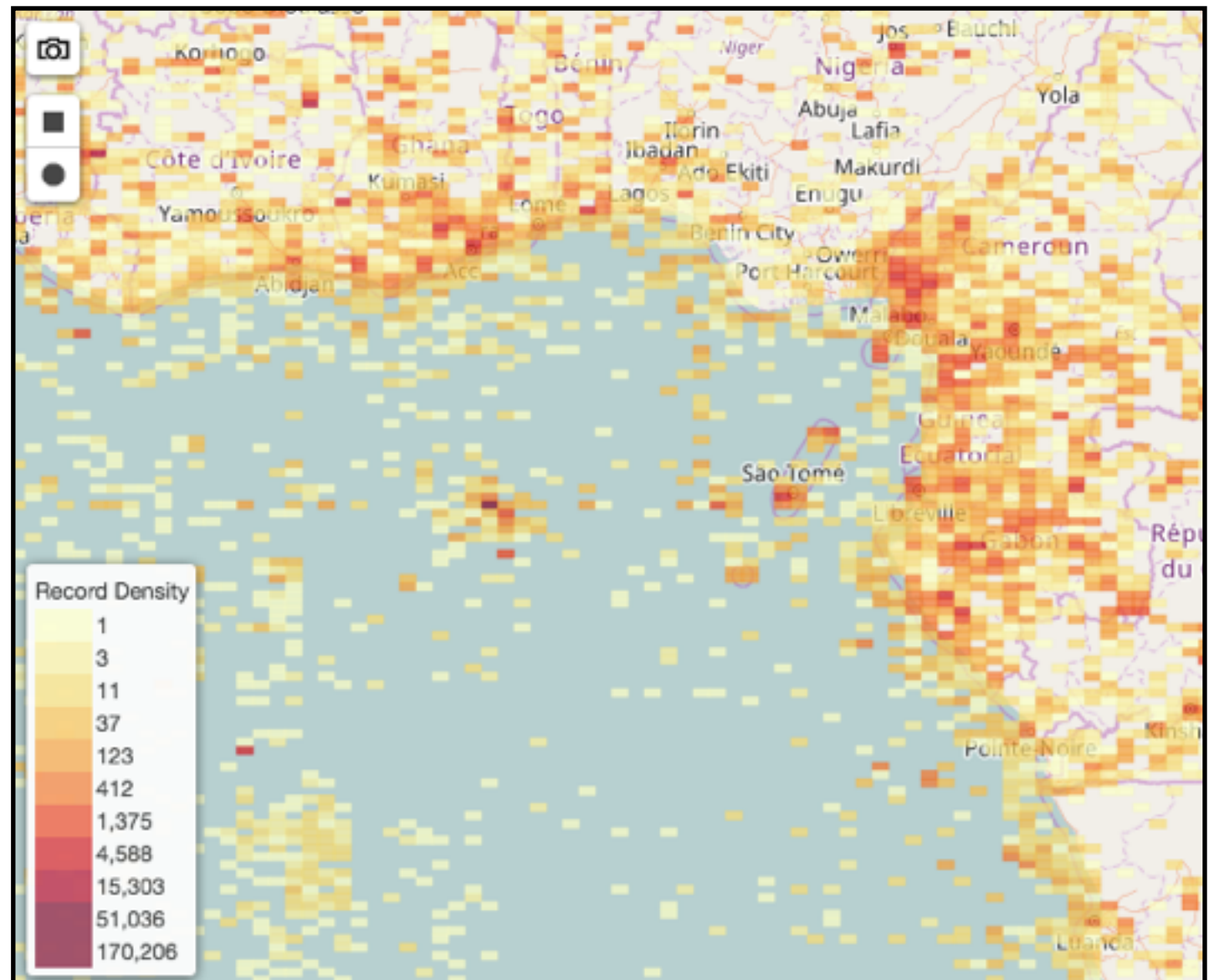
What data are we using?



2. Gridded abiotic data layers

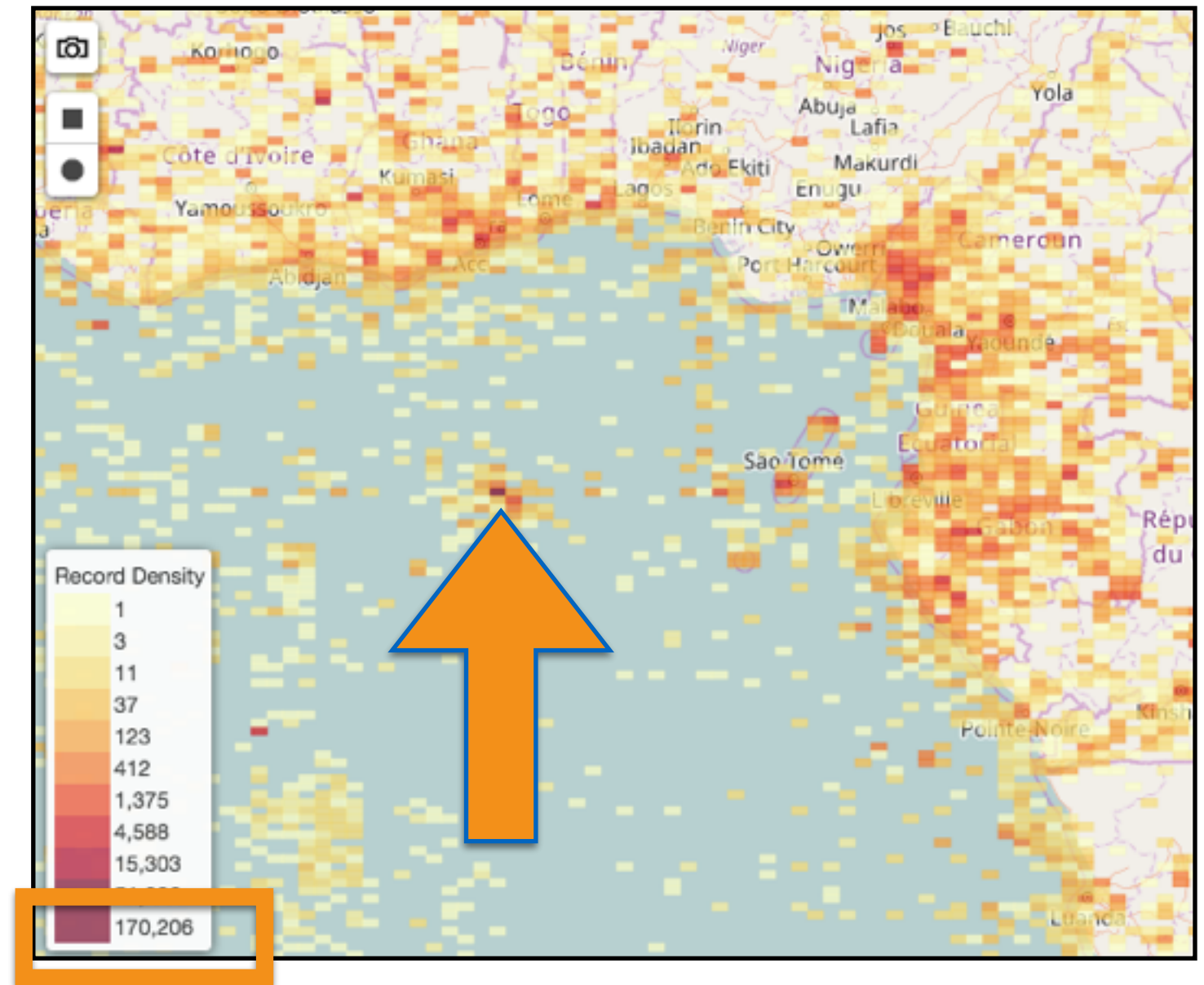
Occurrence data cleaning

- Check taxon names
- Round up the latitude and longitude to desired coarseness
- Remove points that are not precise enough
- Remove bad coordinates
 - Most common coordinate: 0.00, 0.00
 - Cultivated zones, botanical gardens, etc
- Remove duplicates
- Remove coordinates outside desired range



Occurrence data cleaning

- Check taxon names
- Round up the latitude and longitude to desired coarseness
- Remove points that are not precise enough
- Remove bad coordinates
 - Most common coordinate: 0.00, 0.00
 - Cultivated zones, botanical gardens, etc
- Remove duplicates
- Remove coordinates outside desired range



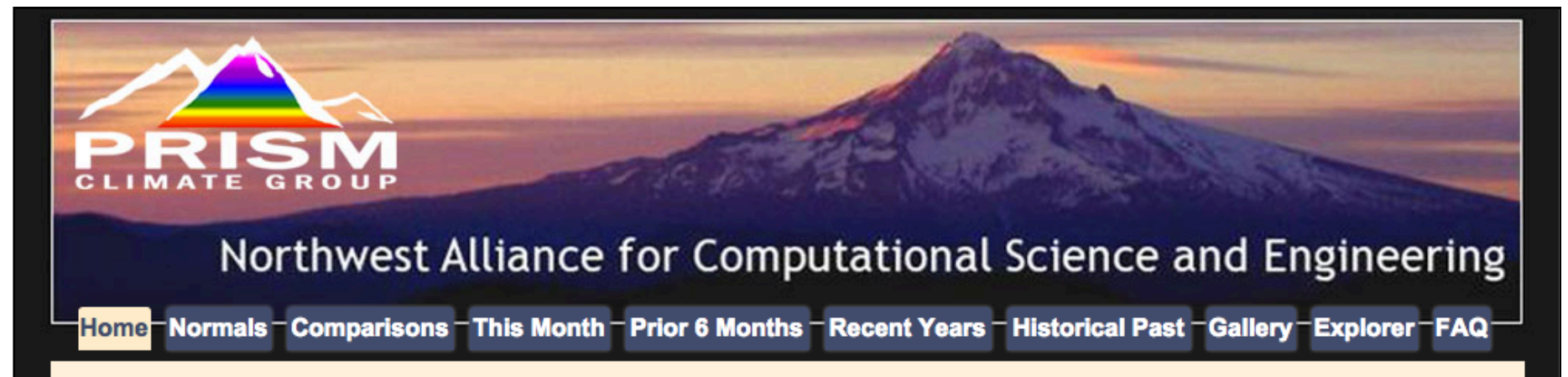
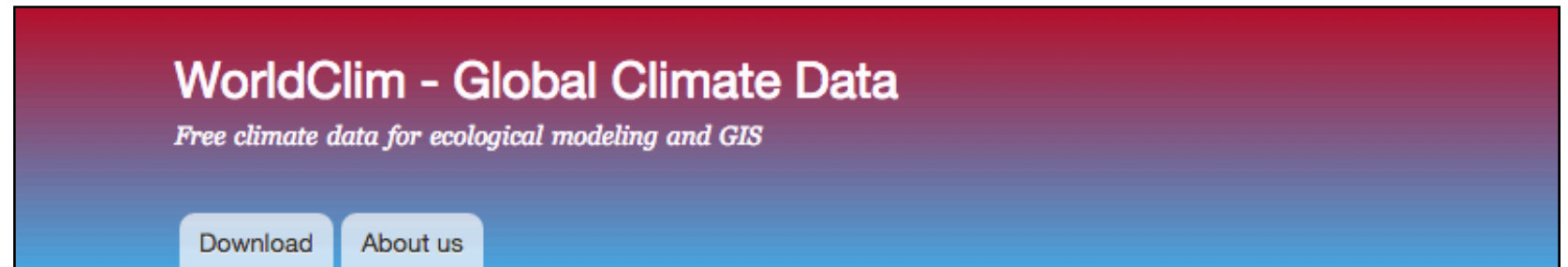
www.idigbio.org

Abiotic data layers

- **What kinds?**
 - Temperature, precipitation, soil, 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



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

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

PRISM

www.prism.oregonstate.edu

- Climate data for the US
- More precise than WorldClim
 - Larger files, more work
- 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 have changed over time. Indicator tool.

This Month: Although still very preliminary, results based on current month's data.

Prior 6 Months: Provisional results based on both monthly and annual data.

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

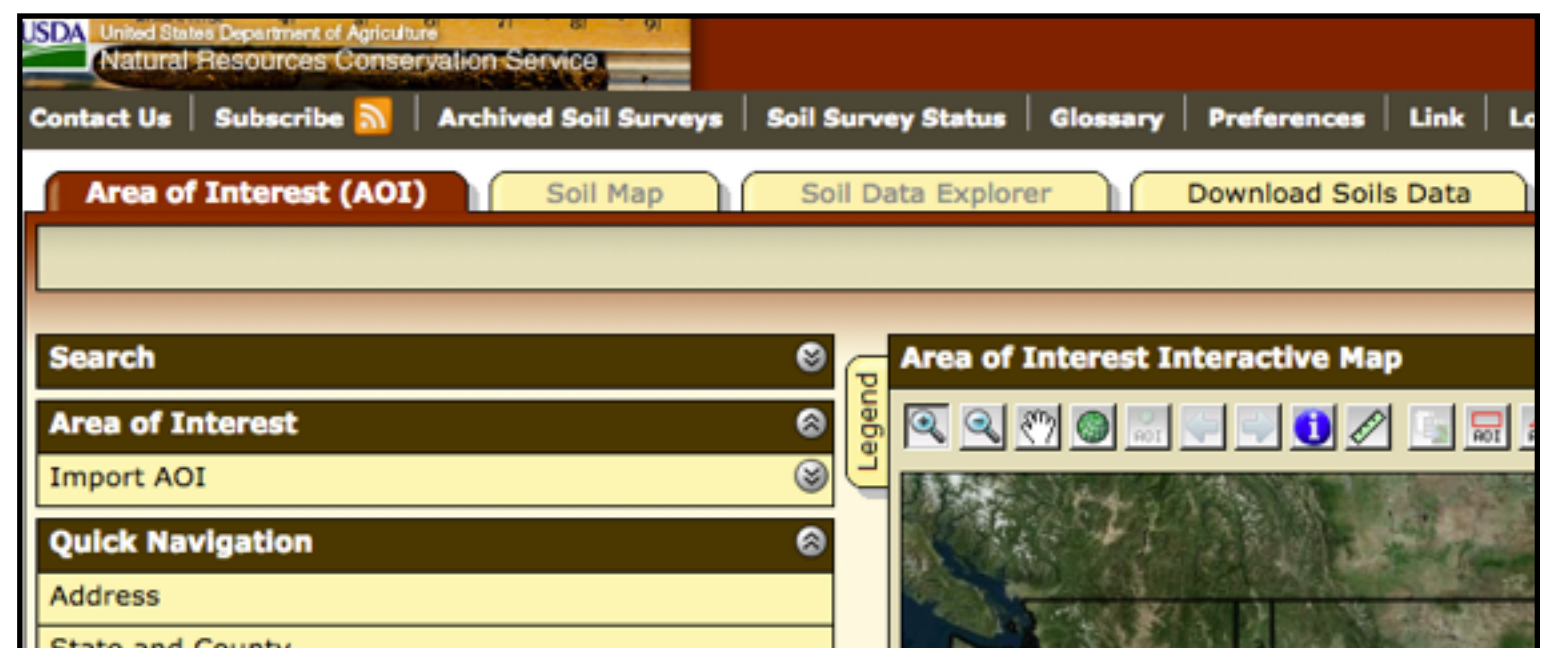
Historical Past: Values prior to 1981 are based on less than 30 years 1895-1990.

Gallery of State Maps: Prepared map images for each state.

Data Explorer: analyze and download time-series data.

Soil Sources

- **UNASM:** http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1242
 - All of North America
 - Eight characteristics of topsoil and subsoil
 - Poor 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

The screenshot shows the EPA website's "Global Change Impacts & Adaptation" section. The header includes the EPA logo and navigation links: "LEARN THE ISSUES", "SCIENCE & TECHNOLOGY", "LAWS & REGULATIONS", and "ABOUT EPA". A search bar and "Advanced Search" link are on the right. The main content area is titled "Global Change Impacts & Adaptation" and includes a breadcrumb trail: "You are here: EPA Home » Climate Change » Global Change Research Program » Global Change Impacts & Adaptation » ICLUS Tools and Datasets (Version 1.3.2)". The page features a sidebar with links to "Global Change Impact & Adaptation Home", "About us", "Staff", "Adaptation", "Decision Support", and "Vulnerability Assessment". The main content area is titled "ICLUS Tools and Datasets (Version 1.3.2)" and includes a "Report Information" section. The report information states: "As a part of the Integrated Climate and Land Use Scenarios (ICLUS) project, this Geographic Information System (GIS) tool can be used to generate scenarios of housing-density changes and calculate impervious surface cover for the conterminous United States. The ICLUS User's Guide accompanies the tool. This product distributes the population projections and creates land use data described in the 2009 EPA report 'Land-Use Scenarios: National-Scale Housing-Density Scenarios Consistent with Climate Change Storylines'."

EPA United States Environmental Protection Agency

Advanced Search A-Z Index

LEARN THE ISSUES SCIENCE & TECHNOLOGY LAWS & REGULATIONS ABOUT EPA

Global Change Impacts & Adaptation

Contact Us Share

You are here: EPA Home » Climate Change » Global Change Research Program » Global Change Impacts & Adaptation » ICLUS Tools and Datasets (Version 1.3.2)

ICLUS Tools and Datasets (Version 1.3.2)

Report Information

As a part of the Integrated Climate and Land Use Scenarios (ICLUS) project, this Geographic Information System (GIS) tool can be used to generate scenarios of housing-density changes and calculate impervious surface cover for the conterminous United States. The ICLUS User's Guide accompanies the tool. This product distributes the population projections and creates land use data described in the 2009 EPA report "Land-Use Scenarios: National-Scale Housing-Density Scenarios Consistent with Climate Change Storylines".

Global Change Impact & Adaptation Home

About us

Staff

Adaptation

Decision Support

Vulnerability Assessment

Other resources

- **Paleoclimate data**

- NOAA (NCEI): <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets>
- PMIP: <https://pmip.lsce.ipsl.fr/>

- **Aquatic Environments**

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

Layer Processing

- Need layers in ASCII format
- Clip layers to fit your desired range
- Pairwise correlation analysis to check for highly correlated layers
- Selection of uncorrelated layers



ENMTools User Manual v1.0

Dan Warren, Rich Glor, and Michael Turelli

danwarren@ucdavis.edu

Selection of Uncorrelated Layers

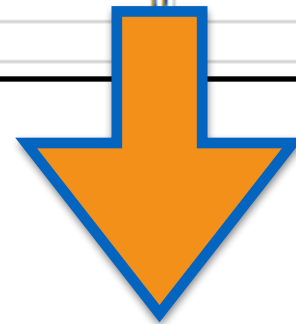
[illegible]

Selection of Uncorrelated Layers

[illegible]

Selection of Uncorrelated Layers

	A	B	C	D	E	F	G	H	I	J	K	L
1	SPECIES	bio19both.asc	altboth.asc	bio1both.asc	bio2both.asc	bio3both.asc	bio4both.asc	bio5both.asc	bio6both.asc	bio7both.asc	bio8both.asc	bio9both.asc
2	bio19both.asc	0	0.132755377	0.026638513	-0.281692623	0.093117381	-0.432234031	-0.199170708	0.214383818	-0.529137555	-0.368313351	0.215975034
3	altboth.asc	0	0	-0.104846869	0.237596592	0.216811600	-0.211514891	-0.133818881	-0.024154968	-0.104346951	-0.398638201	0.111669616
4	bio1both.asc	0	0	0	0.638140412	0.850911736	-0.673848442	0.917375342	0.940454877	0.441682251	0.544030509	0.899571693
5	bio2both.asc	0	0	0	0	0.745598092	-0.178361761	0.801328536	0.429648228	0.199313237	0.446468815	0.526264858
6	bio3both.asc	0	0	0	0	0	-0.748117897	0.745592303	0.840175143	0.472202343	0.317461982	0.839385272
7	bio4both.asc	0	0	0	0	0	0	-0.3538374	-0.867363748	0.92389995	0.013354098	-0.775161836
8	bio5both.asc	0	0	0	0	0	0	0	0.743791458	-0.001300507	0.63601491	-0.763586548
9	bio6both.asc	0	0	0	0	0	0	0	0	-0.712753715	0.352473797	0.922499686
10	bio7both.asc	0	0	0	0	0	0	0	0	0	0.141079536	-0.576336331
11	bio8both.asc	0	0	0	0	0	0	0	0	0	0	0.208183943
12	bio9both.asc	0	0	0	0	0	0	0	0	0	0	0

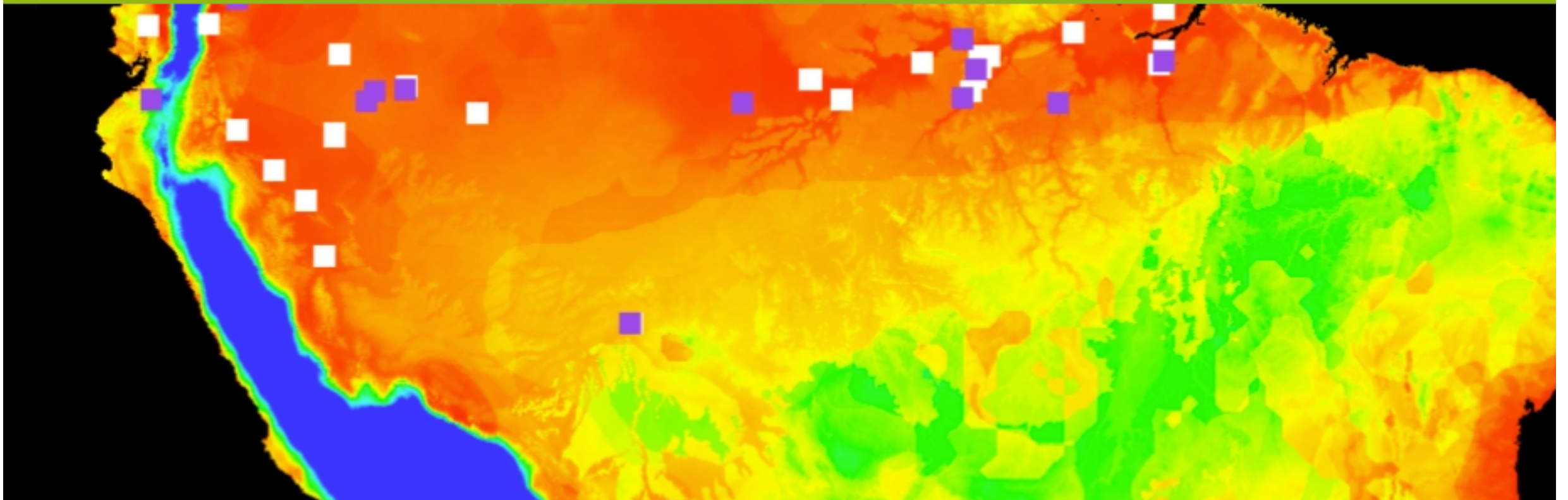


	A	B	C	D	E	F	G	H
1	SPECIES	bio19both.asc	altboth.asc	bio1both.asc	bio2both.asc	bio4both.asc	bio7both.asc	bio8both.asc
2	bio19both.asc	0	0.132755377	0.026638513	-0.281692623	-0.432234031	-0.529137555	-0.368313351
3	altboth.asc	0	0	-0.104846869	0.237596592	-0.211514891	-0.104346951	-0.398638201
4	bio1both.asc	0	0	0	0.638140412	-0.673848442	-0.441682251	0.544030509
5	bio2both.asc	0	0	0	0	-0.178361761	0.199313237	0.446468815
6	bio4both.asc	0	0	0	0	0	0.92389995	-0.013354098
7	bio7both.asc	0	0	0	0	0	0	0.141079536
8	bio8both.asc	0	0	0	0	0	0	0
9								

www.worldclim.org

Ecological Niche Modeling

Maxent software for modeling species niches and distributions



https://biodiversityinformatics.amnh.org/open_source/maxent/

Maximum Entropy Species Distribution Modeling, Version 3.3.3k

Samples

File:

☒ Asclepias_curtissii

☐ Asimina_obovata

☐ Pinus_palustris

☒ Linear features

☒ Quadratic features

☒ Product features

☒ Threshold features

☒ Hinge features

☒ Auto features

Environmental layers

Directory/File:

<input checked="" type="checkbox"/> alt	Continuous
<input type="checkbox"/> bio1	Continuous
<input checked="" type="checkbox"/> bio10	Continuous
<input type="checkbox"/> bio11	Continuous
<input checked="" type="checkbox"/> bio12	Continuous
<input type="checkbox"/> bio13	Continuous
<input checked="" type="checkbox"/> bio14	Continuous
<input type="checkbox"/> bio15	Continuous
<input checked="" type="checkbox"/> bio16	Continuous
<input type="checkbox"/> bio17	Continuous

Create response curves ☒

Make pictures of predictions ☒

Do jackknife to measure variable importance ☒

Output format:

Output file type:

Output directory:

Projection layers directory/file:

Maximum Entropy Parameters

Basic | Advanced | Experimental

☒ Random seed
☒ Give visual warnings
☒ Show tooltips
☐ Ask before overwriting
☐ Skip if output exists
☒ Remove duplicate presence records
☒ Write clamp grid when projecting
☒ Do MESS analysis when projecting

Random test percentage

Regularization multiplier

Max number of background points

Replicates

Replicated run type

Test sample file



Maximum Entropy Parameters

Basic Advanced Experimental

- ☒ Add samples to background
- ☐ Add all samples to background
- ☒ Write plot data
- ☒ Extrapolate
- ☒ Do clamping
- ☒ Write output grids
- ☒ Write plots
- ☐ Append summary results to maxentResults.csv file
- ☒ Cache ascii files

Maximum iterations 5000

Convergence threshold 0.00001

Adjust sample radius 0

Log file maxent.log

Default prevalence 0.5

Apply threshold rule

Bias file

Maximum Entropy Parameters

Basic | Advanced | Experimental

- ☒ Logscale raw/cumulative pictures
- ☐ Per species results
- ☒ Write background predictions
- ☒ Show exponent in response curves
- ☐ Fade by clamping
- ☐ Verbose
- ☐ Use samples with some missing data

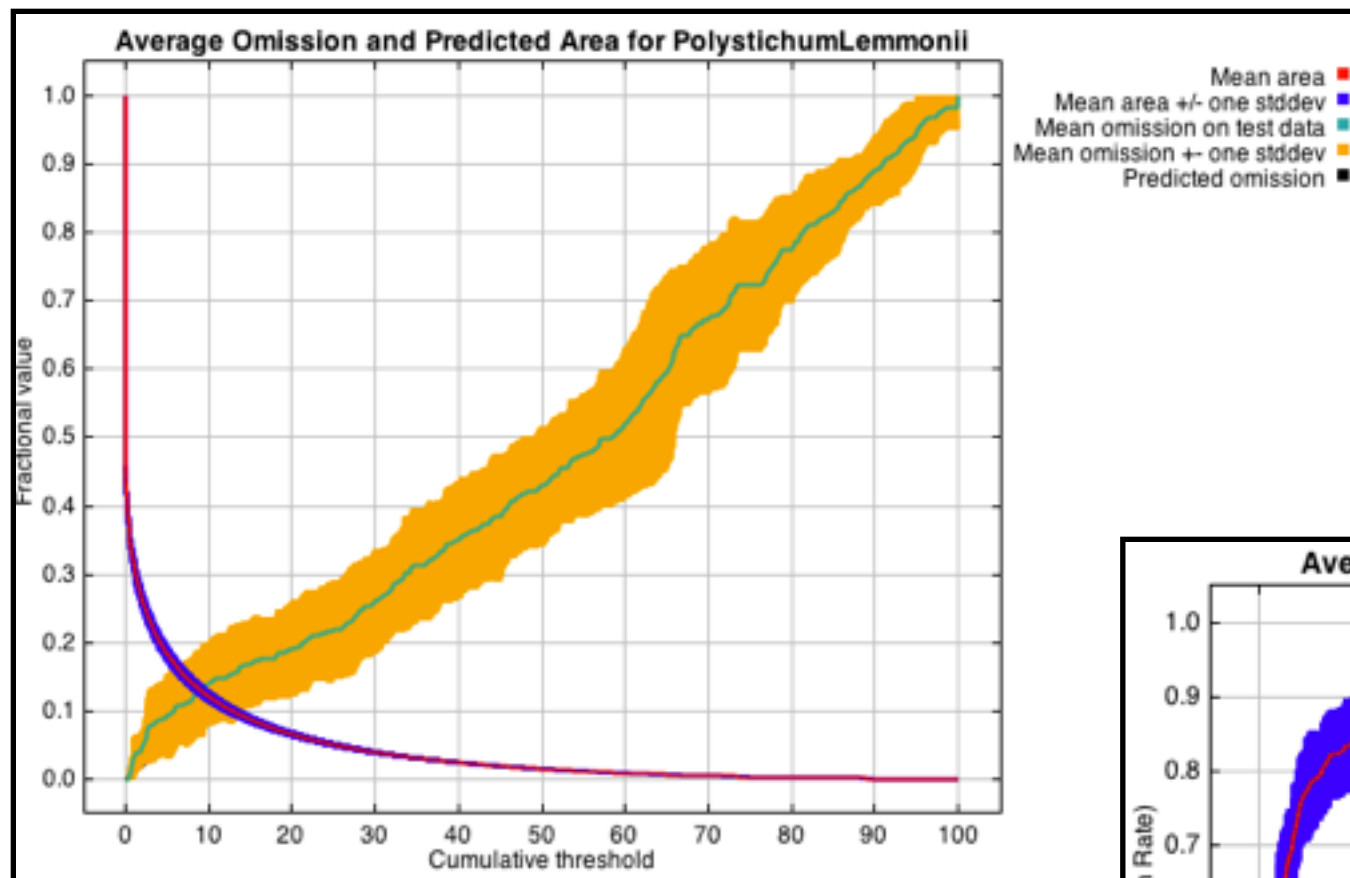
Threads	1
Lq to lqp threshold	80
Linear to lq threshold	10
Hinge threshold	15
Beta threshold	-1
Beta categorical	-1
Beta lqp	-1
Beta hinge	-1
Default nodata value	-9999

RUN!

Now what?

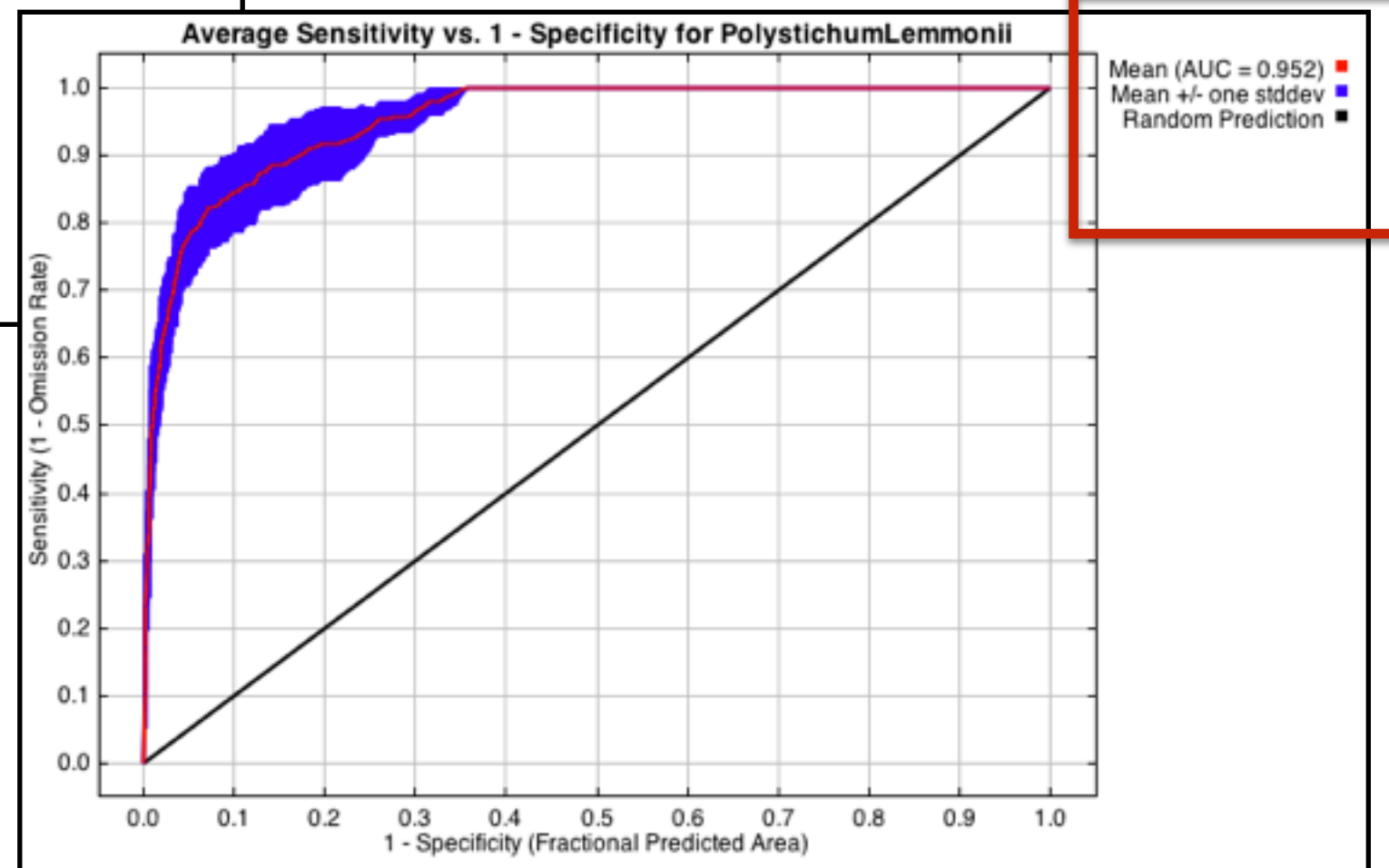
- Maxent outputs a .html with all of the stats and maps
 - Model statistics: training vs test data, AUC scores
 - Maps: suitability scores, binary thresholds, projections
 - Variable performance: contribution to model, response curves

Model Statistics



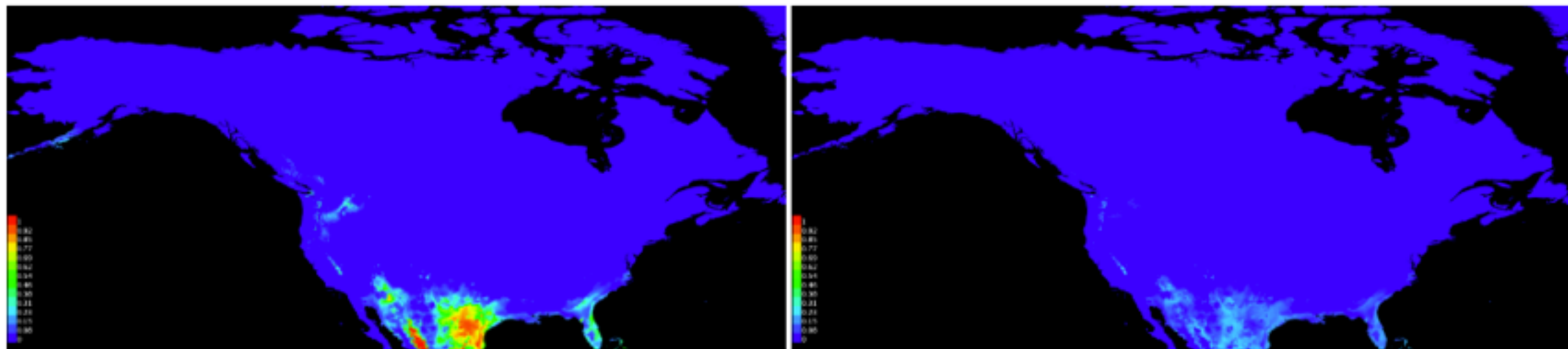
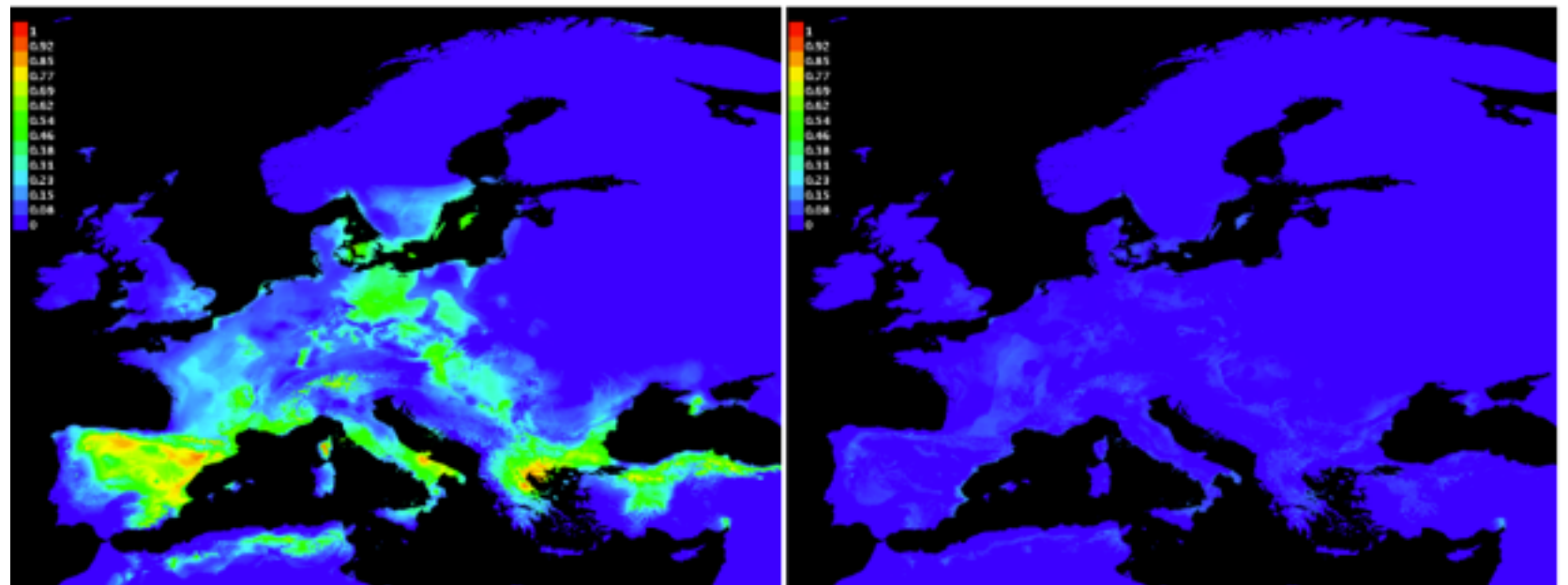
Omission rate should be close to the predicted omission

AUC should be over 0.8



Maps

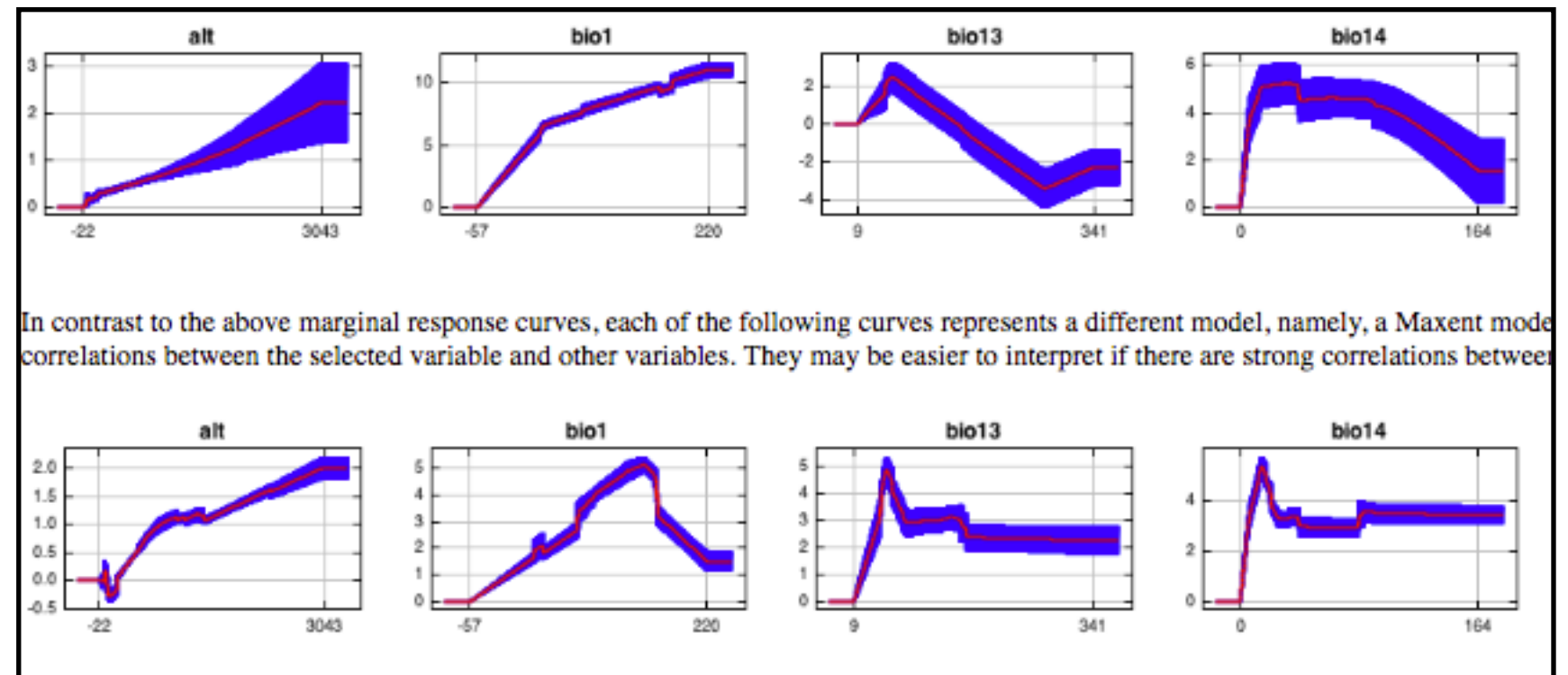
Suitability heat maps
of the point-wise mean
and standard deviation
of the replicates



You can also project those models onto other regions or time
periods with projections

Variable Performance

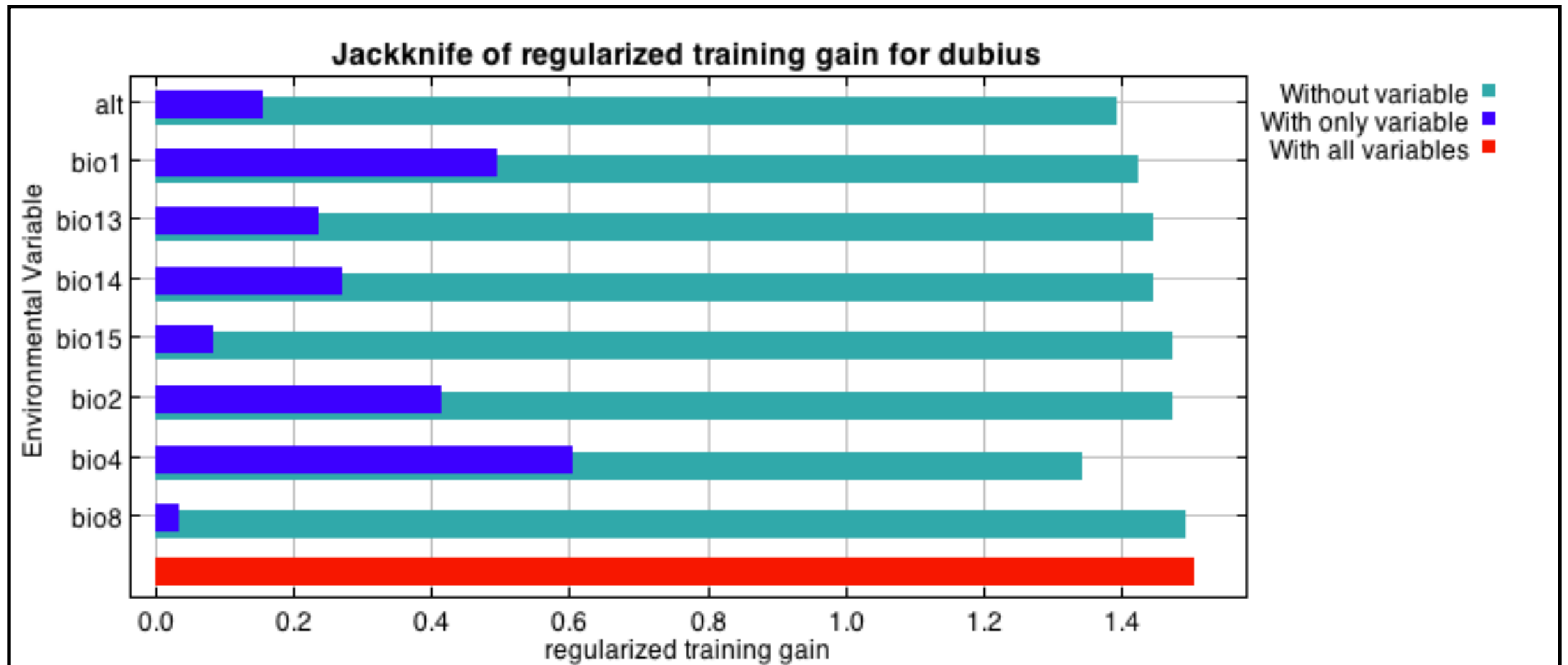
Response Curves



Variable	Percent contribution	Permutation importance
bio4	41.2	43.8
bio14	17.3	10.8
bio2	11.7	1.9
bio1	10.2	20.2
bio15	8.9	4
bio13	6.7	9.3
alt	3.1	9.2
bio8	1	0.9

Variable Contribution

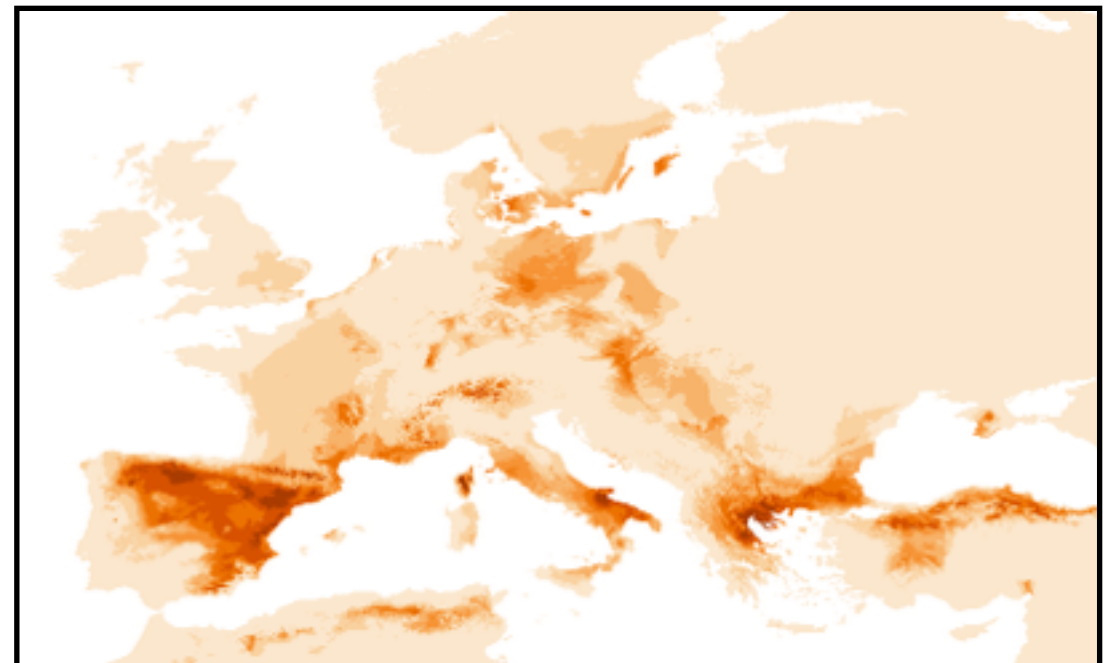
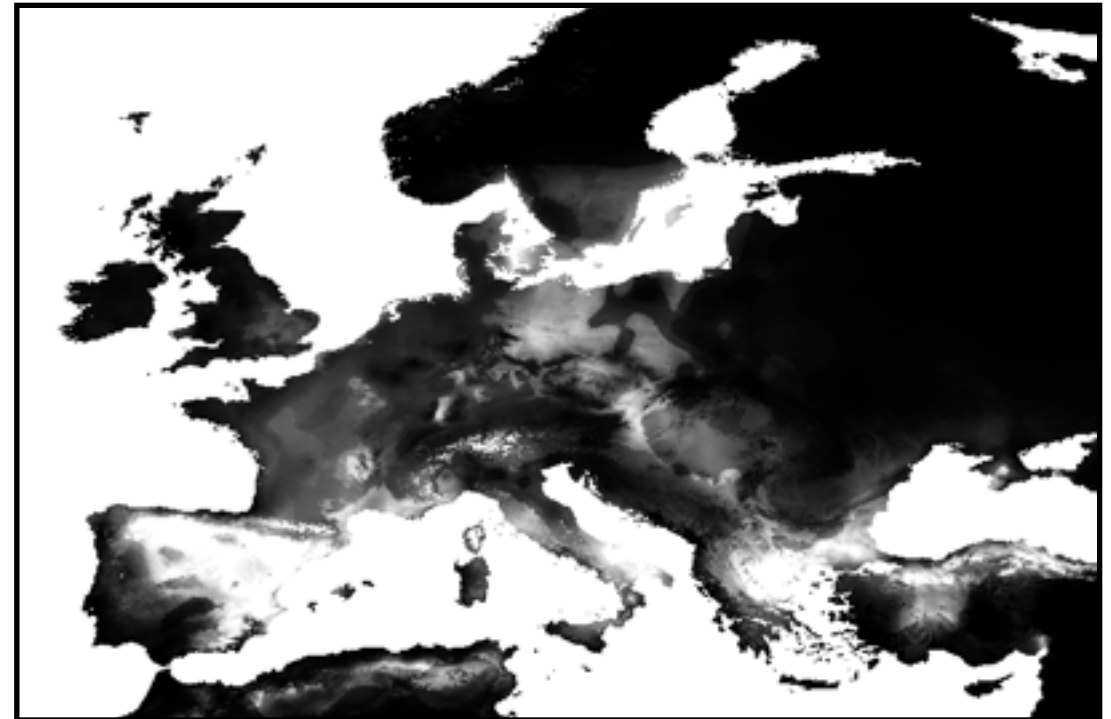
Variable Performance



Shows importance of variables

Other files

- Suitability ASCIIs for the average, maximum, median, minimum, and standard deviation
 - Also for the projections, if included
- Maxent log
- MaxentResults.csv
- Results for the replicates



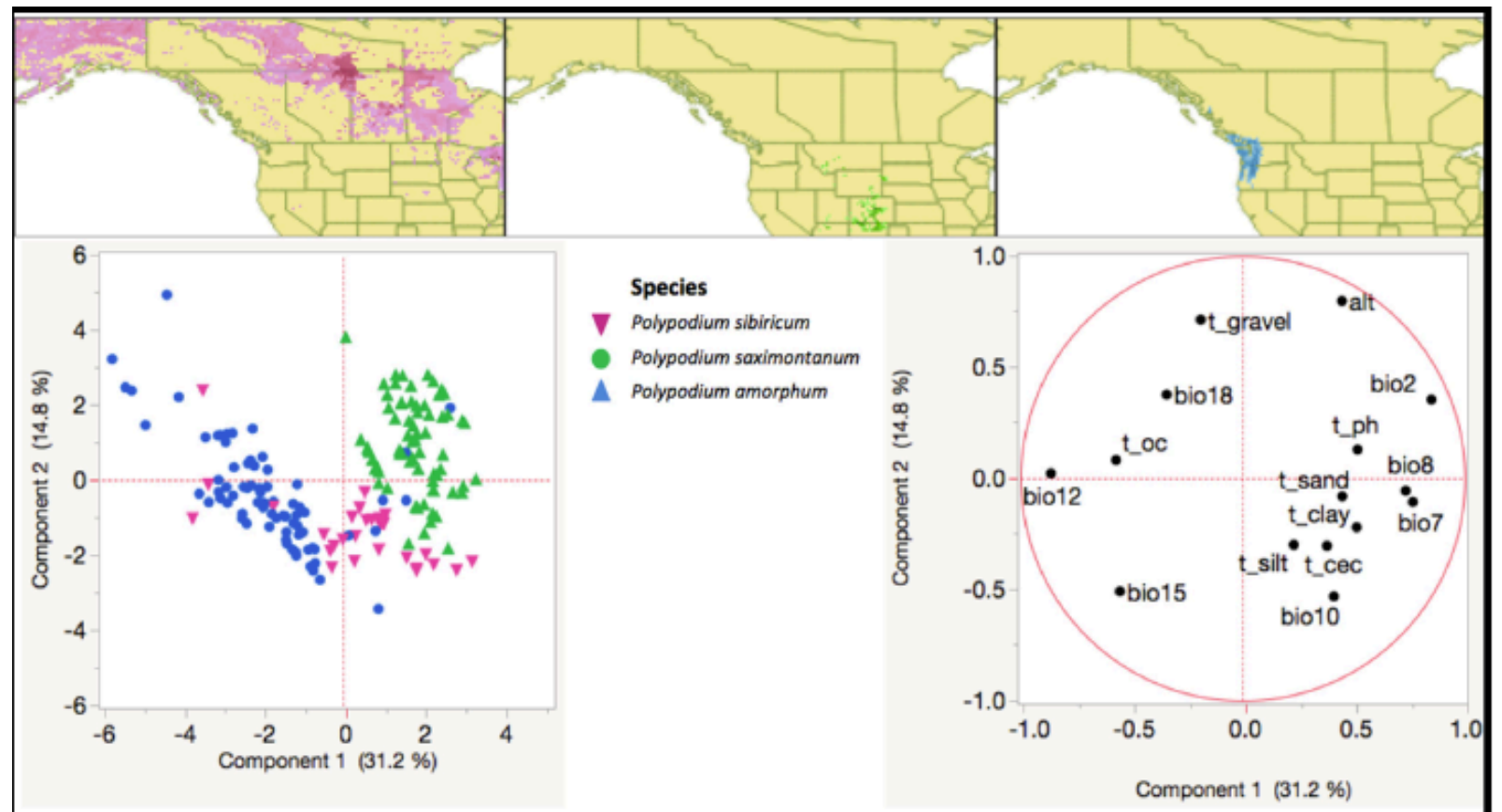
Additional Analyses

- Niche overlap
- Niche breadth
- Range overlap
- Point-sampling and PCA

ENMTools User Manual v1.0

Dan Warren, Rich Glor, and Michael Turelli

danwarren@ucdavis.edu



ENM Resource Page

<https://www.idigbio.org/content/using-digitized-collections-based-data-research-free-hands-crash-course-ecological-niche>



Video Tutorials
R Scripts
Example Datasets

Link to Sococo Online Office Hours
Tuesdays at 10am