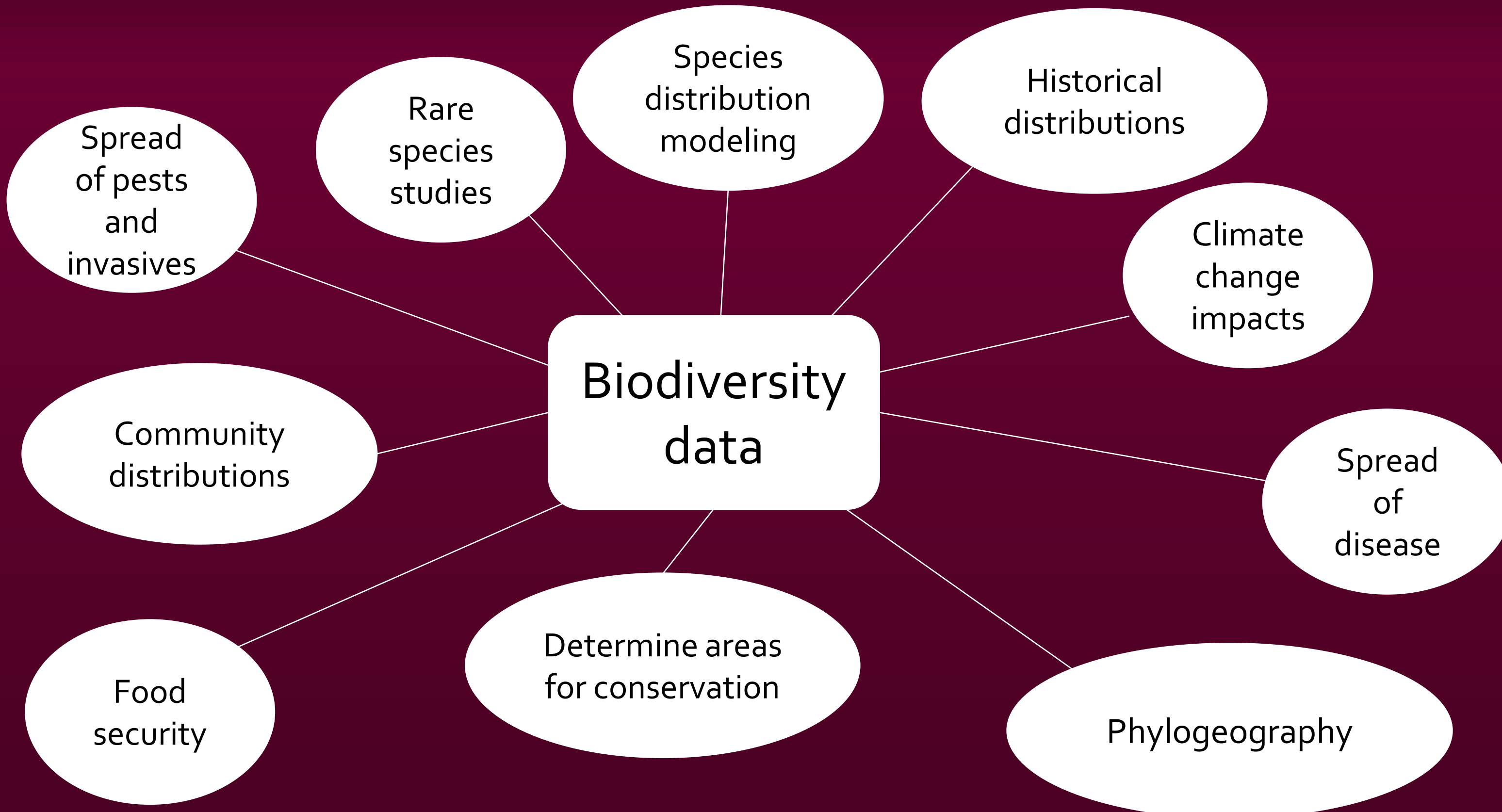
The Contribution of Small Collections: A Case Study from Fuireneae (Cyperaceae)

Heather E. Dame [1,2]*, Benjamin W. Heumann [1,3], J. Richard Carter [4], Jessica M. Bartek [4], Anna K. Monfils [1,2]

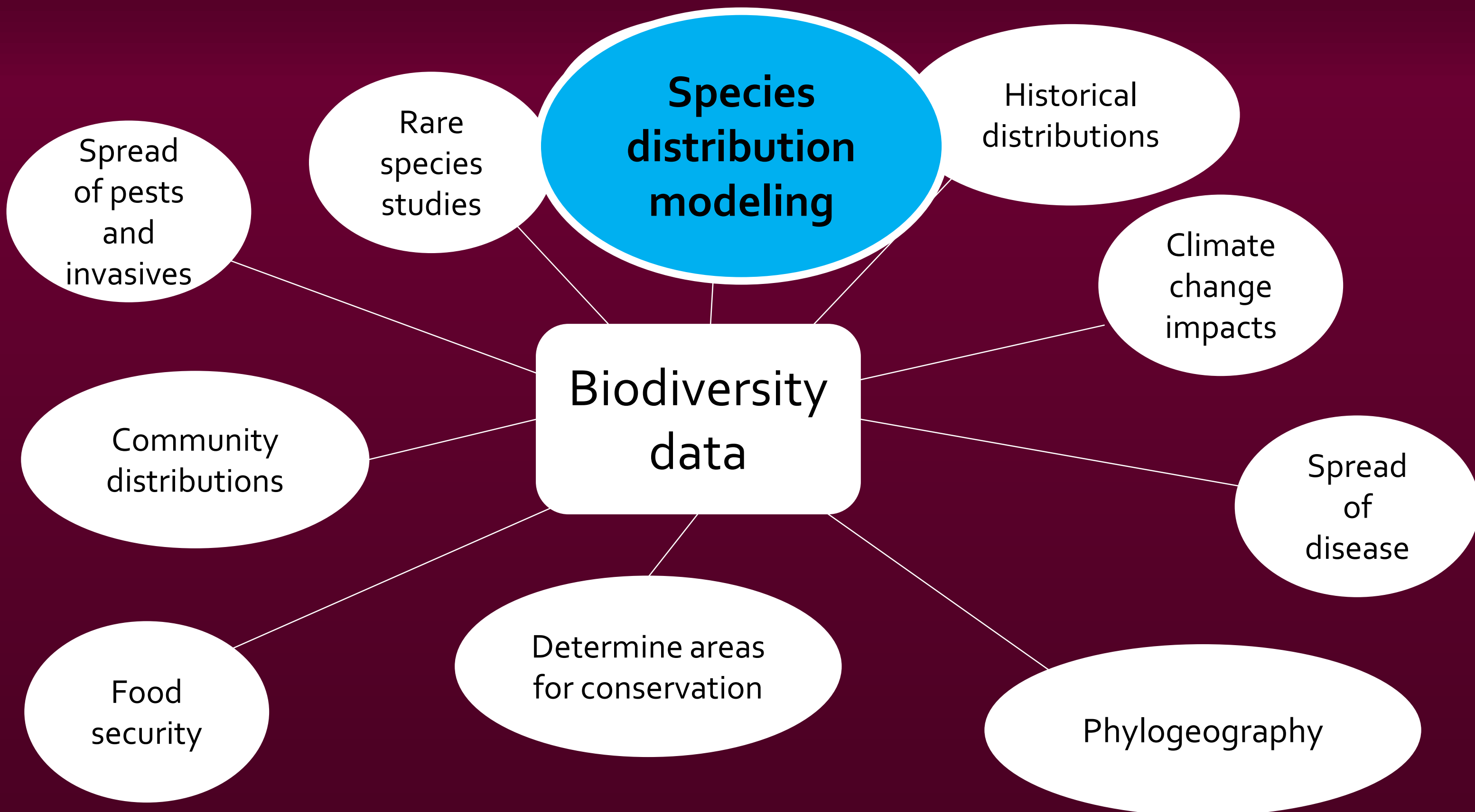
(1) Central Michigan University, Institute for Great Lakes Research, Mount Pleasant, MI
(2) Central Michigan University, Department of Biology, Mount Pleasant, MI
(3) Central Michigan University, Department of Geography, Center for Geographic Information Science, Mount Pleasant, MI
(4) Department of Biology, Valdosta State University, Valdosta, GA

CMU
CENTRAL MICHIGAN UNIVERSITY

Uses of primary biodiversity data



Uses of primary biodiversity data



Species Distribution Modeling in brief

- Allows understanding of distributions without having complete sampling of species
- Models are largely reliant on the data that is put into them

Global Biodiversity Information Facility (GBIF)

- Free, online portal for species occurrence records linked to primary biodiversity data

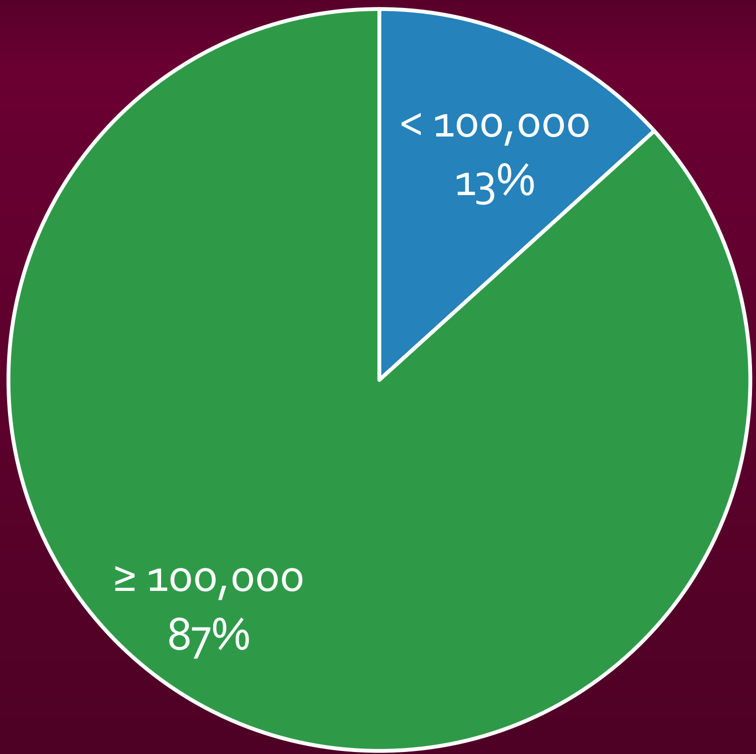
- Largest biodiversity database available:
  - >500 million records
  - >1.5 million species
  - Contains over 300 years of data collections
  - Cited in >1,300 peer-reviewed research publications

Vision: *"A world in which biodiversity information is freely and universally available for science, society and a sustainable future."*

Digitization of Small Collections

Growing appreciation for the potential contribution of small collections in the national digitization effort

Herbaria specimens in the United States



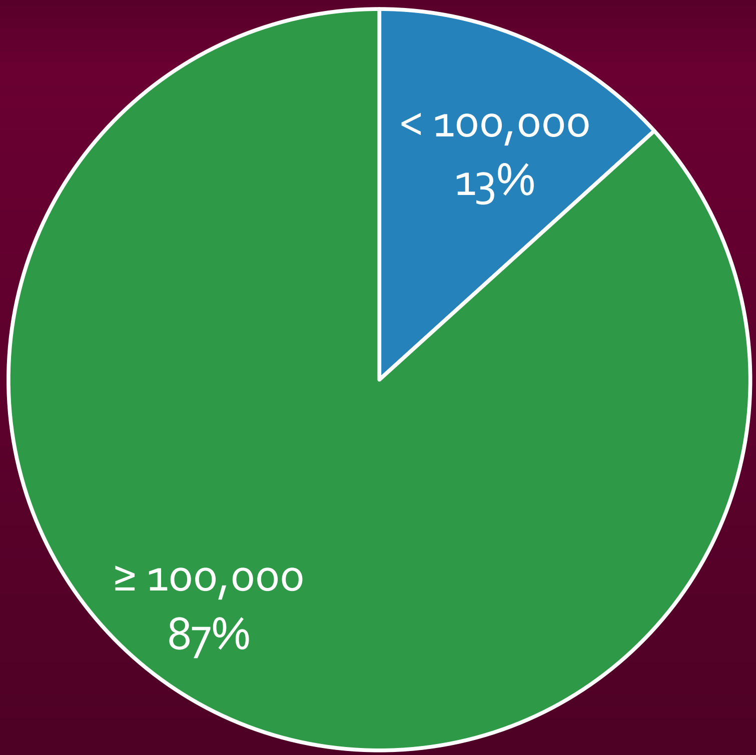
Percent specimens in size class

Small Collections

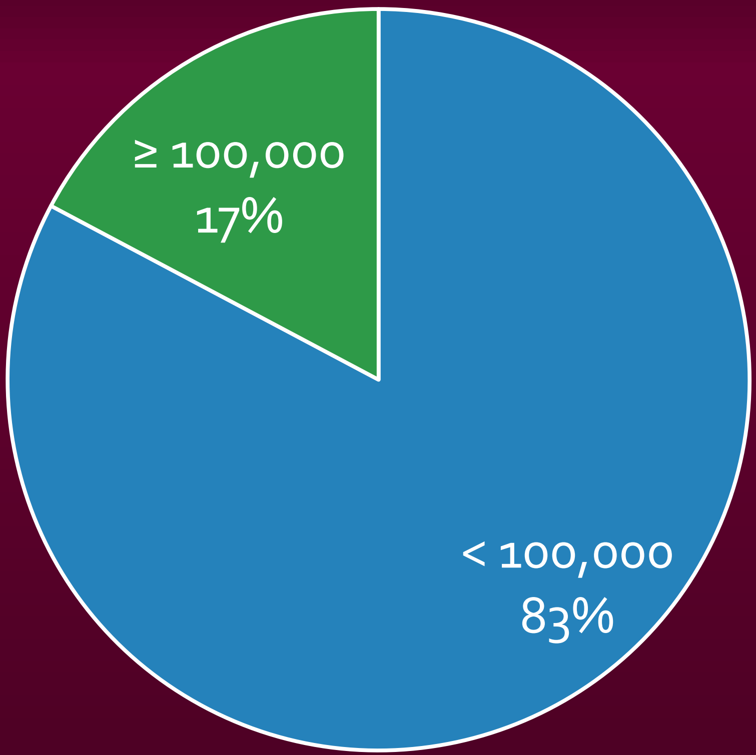
- < 100,000 specimens
- Regional collecting

Index herbariorum

Herbaria specimens in the United States



Percent specimens in size class



Percent of herbaria in size class

Index herbariorum

Research Question

What is the relative contribution of small collections to our understanding of species distribution and niche modeling?

Objectives

- Assess the predictive power of large, small, and combined collection datasets
- Evaluate the relative influence of large, small, and combined collection datasets on geographic predictions

Species Distribution Modeling using Maximum Entropy

MaxEnt (Phillips et al., 2006)

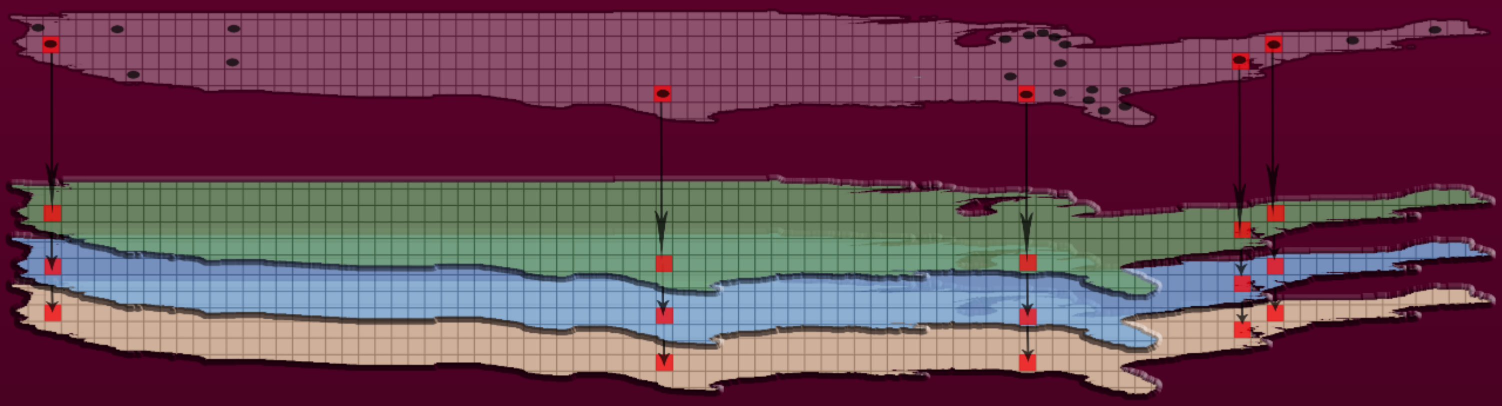
- Predict suitable habitat over a geographic space
- Presence-only modeling method
- Consistent high performance among other modeling methods

Modeling extent of the Contiguous U.S.

Consistently high quality environmental variables

MaxEnt builds a landscape showing the probability of suitable habitat

- Creates background data grid
- Correlates presence points to background grid cells
- Defines mean and variance of niche
- Builds predictions of the probability of suitable habitat



Methodology

- Obtain Species Occurrence Data
- Filter Data
- Species Distribution Model
- Model Prediction Evaluation
- Geographic Space Analysis

Methodology

- Obtain Species Occurrence Data
- Filter Data
- Species Distribution Model
- Model Prediction Evaluation
- Geographic Space Analysis

Two sources of collections data

- GBIF collections data
  - Routinely used in species distributions modeling studies
- Small regional collections collaborations
  - Central Michigan University (CMC) and Valdosta State University (VSC)

CMC HERBARIUM
Central Michigan University

VALDOSTA STATE UNIVERSITY

Fuireneae (Cyperaceae; Sedges)

- Wetland plants
- 4 genera naturally occur in the United States
  - Wide ranging
  - Narrow endemics

Methodology

- Obtain Species Occurrence Data
- Filter Data
- Species Distribution Model
- Geographic Space Analysis
- Model Prediction Evaluation

Three Independent Datasets

- GBIF Large Collections
  - Occurrences from collections >100,000
- GBIF Small Collections
  - Occurrences from collections <100,000
- CMC/VSC Collections
  - Occurrences from CMC and VSC herbaria

GBIF Large Collections

GBIF Small Collections

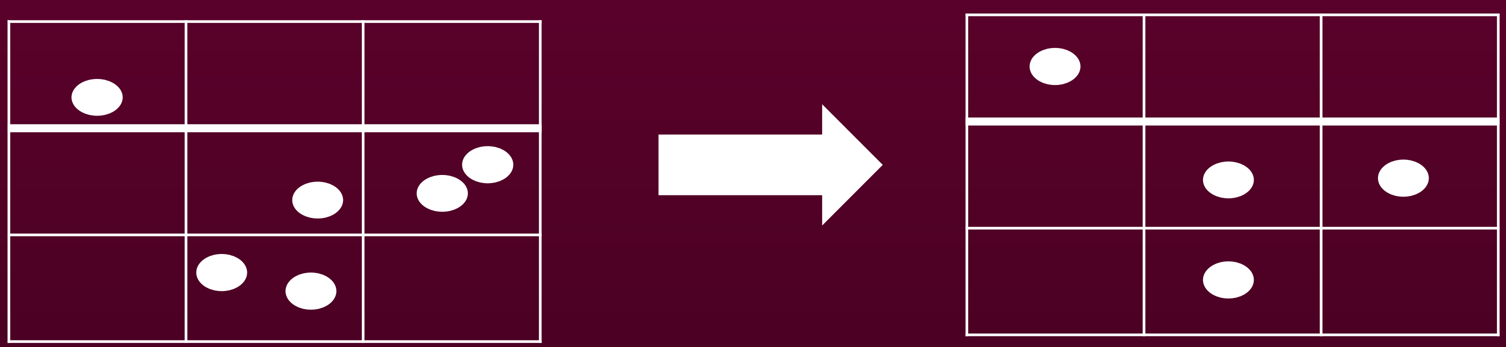
CMC/VSC Collections

Data quality

- Retained data with sufficient metadata
  - Preserved Specimens
  - No voucher duplicates
  - Georeferenced quality
    - Georeferenced CMC / VSC collections using GeoLocate (Rios & Bart, 2010)

Selecting species

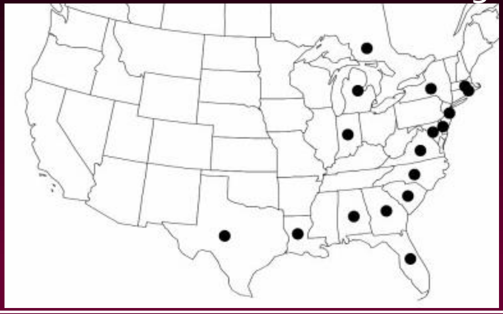
- ≥ 10 occurrence records
- Species present in both GBIF and CMC/VSC datasets
- No obligate halophytes
- Removed geographic replicates using ENMTools (Warren et al., 2010)



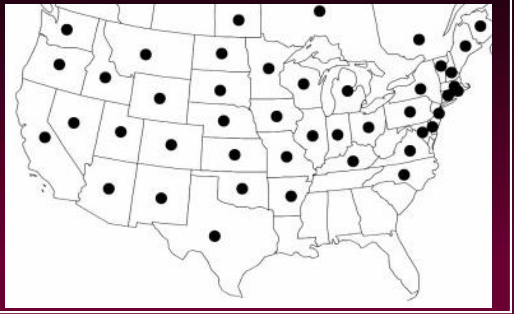
Species selected for analysis

Species	GBIF Large	GBIF Small	CMC/VSC	Total
Fuirena pumila (Torrey) Sprengel	10	n/a	22	32
Fuirena squarrosa Michx.	25	n/a	44	69
Schoenoplectiella purshiana (Fernald) Lye	45	n/a	15	60
Schoenoplectus acutus (Bigelow) Á. Löve & D. Löve	434	52	13	499
Schoenoplectus pungens (Vahl) Palla	413	32	26	471
Schoenoplectus tabernaemontani (C.C. Gmel.) Palla	352	38	29	419
Total	1270	122	149	1550

Fuirena pumila
Dwarf umbrella sedge

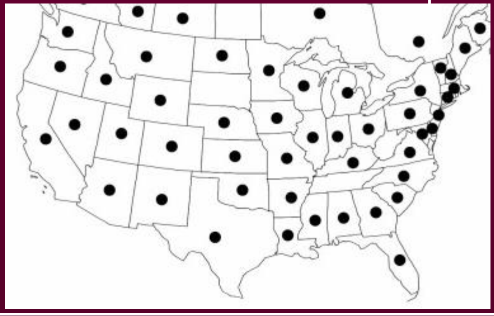


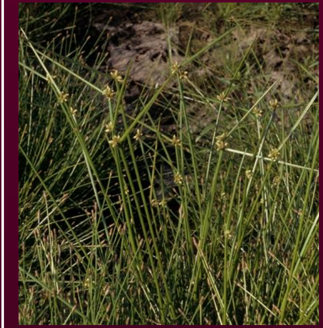
Schoenoplectus acutus
Hardstem bulrush



Fuirena squarrosa
Hairy umbrella sedge

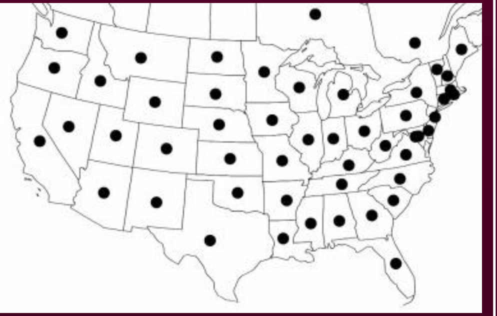
Schoenoplectus pungens
Common Threesquare





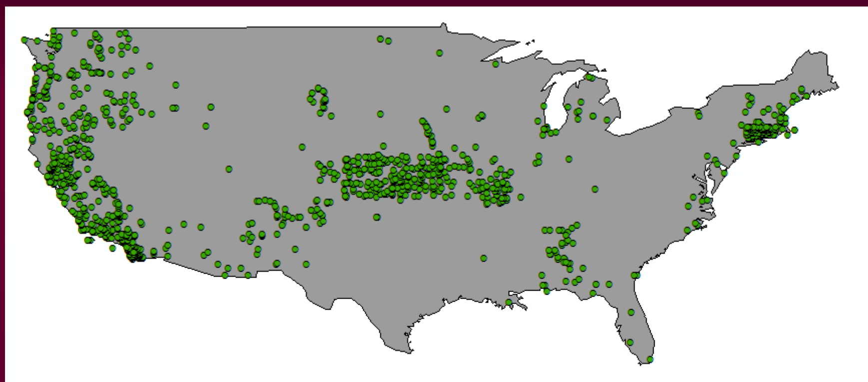
Schoenoplectiella purshiana
Weakstalk bulrush

Schoenoplectus tabernaemontani
Soft-stem bulrush

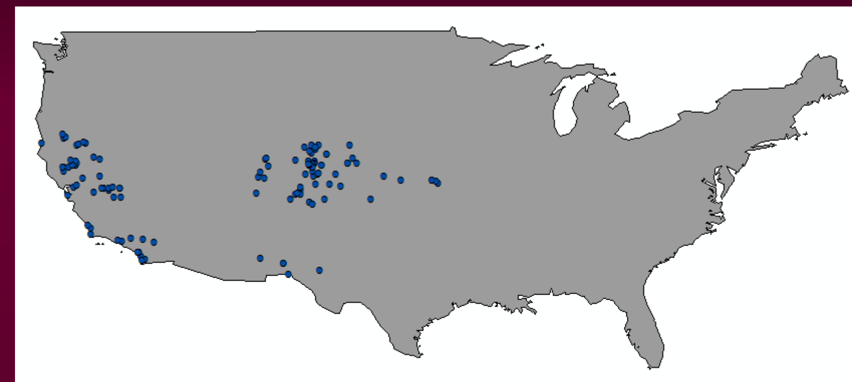


Maps: floranorthamerica.org
Photos: MichiganFlora.net and Plants.USDA.org

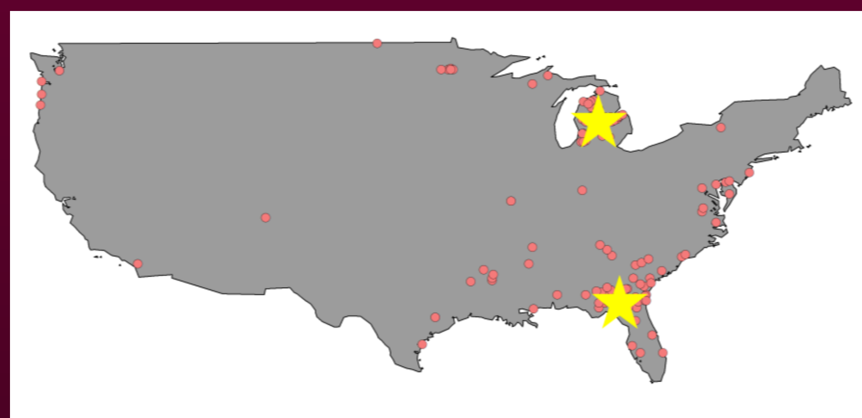
Geographic locations of occurrences



GBIF Large Dataset Occurrences



GBIF Small Dataset Occurrences



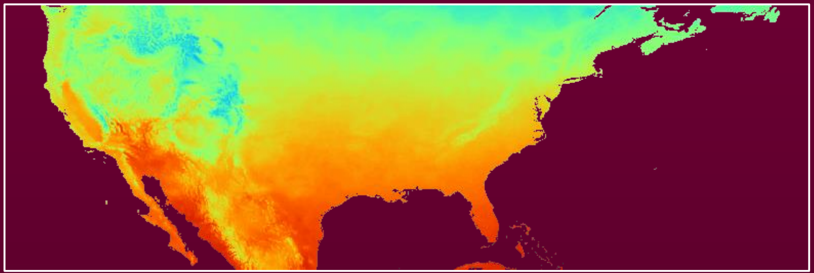
CMC/VSC Dataset Occurrences

CMC Herbarium (Michigan)
VSC Herbarium (Georgia)

Environmental Factors

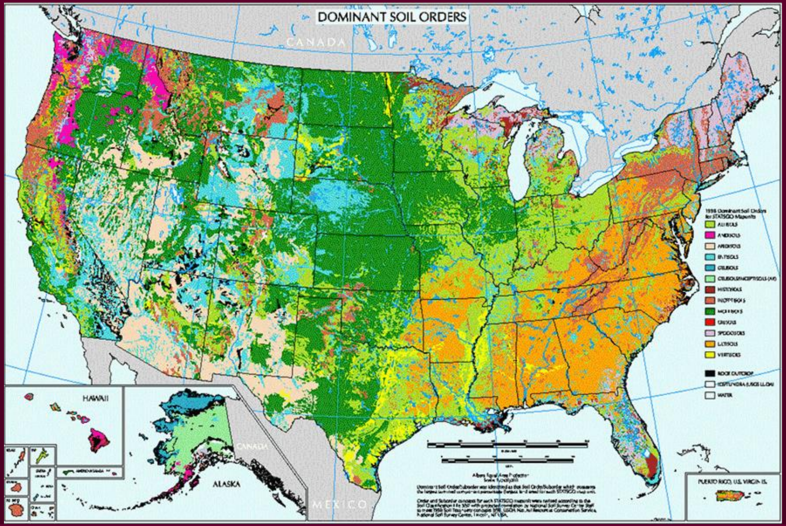
WorldClim Bioclimatic Factors(Hijmans et al. 2005)

- 7 factors



STATSGO2 Soil Factors(Soil Survey Staff)

- 7 factors



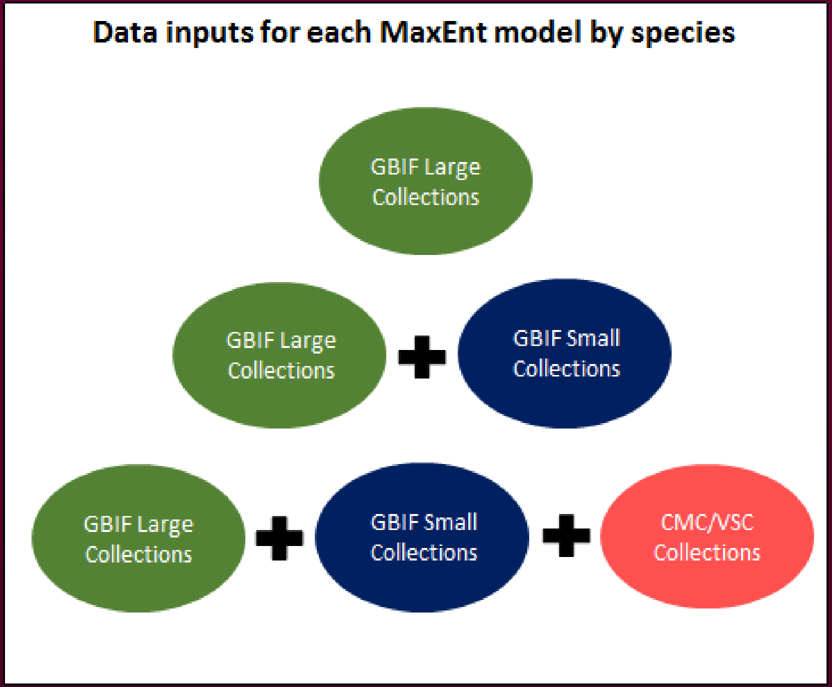
Total: 14 environmental variables

Methodology

- Obtain Species Occurrence Data
- Filter Data
- Species Distribution Model
- Model Prediction Evaluation
- Geographic Space Analysis

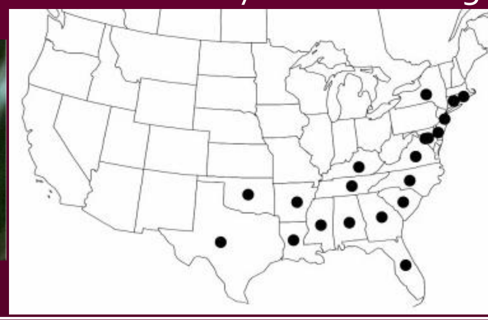
Three Additive Data Inputs

- Occurrences from GBIF Large collections
- Occurrences from both GBIF Small and Large collections
- Occurrences from all GBIF and CMC/VSC collections



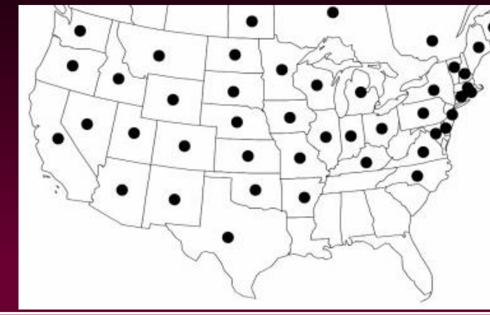
Fuirena pumila
Dwarf umbrella sedge

Fuirena squarrosa
Hairy umbrella sedge

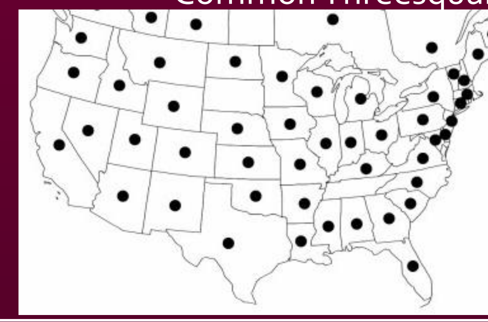


Schoenoplectiella purshiana
Weakstalk bulrush

Schoenoplectus acutus
Hardstem bulrush



Schoenoplectus pungens
Common Threesquare

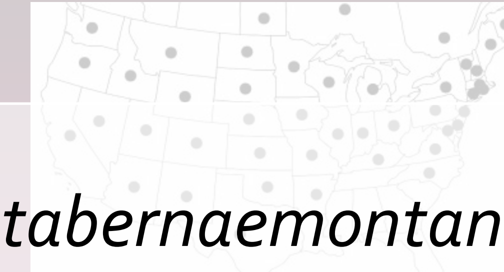


Schoenoplectus tabernaemontani
Soft-stem bulrush

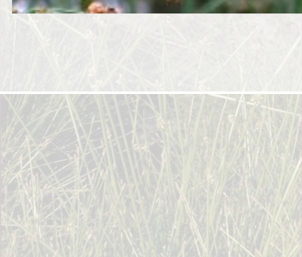
Maps: floranorthamerica.org
Photos: MichiganFlora.net and Plants.USDA.org

Fuirena pumila
Dwarf umbrella sedge

Schoenoplectus acutus
Hardstem bulrush



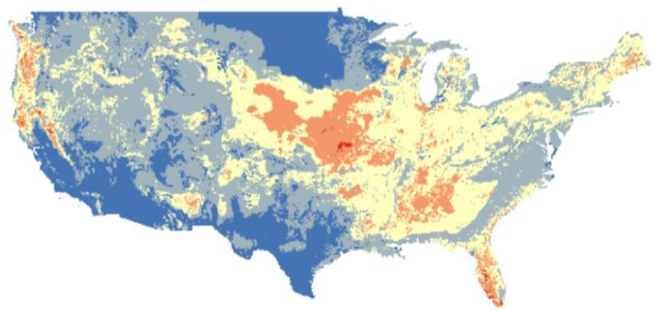
Schoenoplectus tabernaemontani



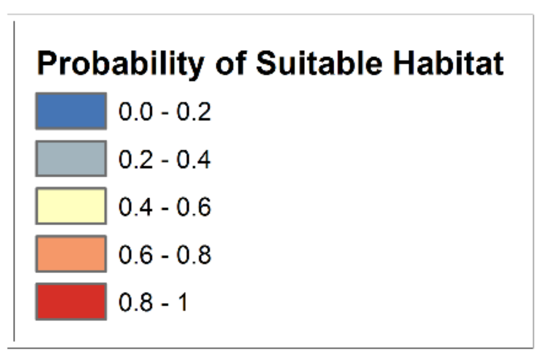
Maps: floranorthamerica.org
Photos: MichiganFlora.net and Plants.USDA.org

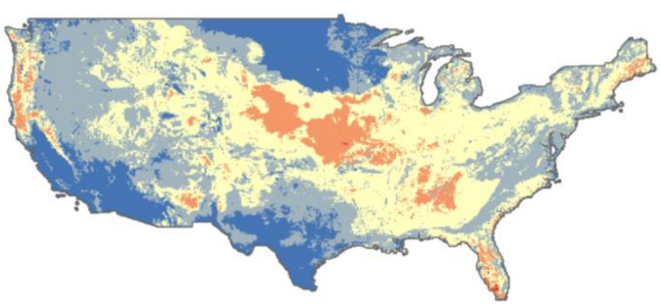
MaxEnt maps displaying the probability of suitable habitat

Schoenoplectus tabernaemontani

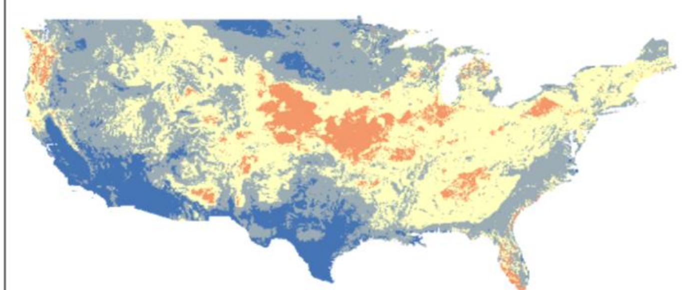


Model built from GBIF Large Collections





Model built from GBIF Small & Large Collections

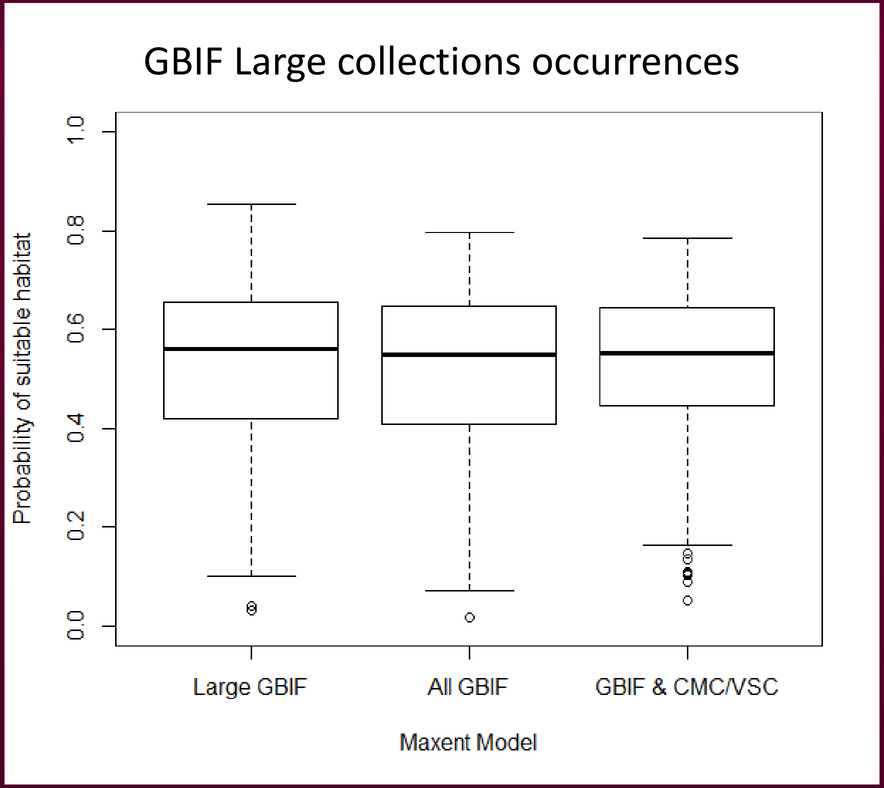


Model built from all GBIF Collections & CMC/VSC

Methodology

- Obtain Species Occurrence Data
- Filter Data
- Species Distribution Model
- Model Prediction Evaluation
- Geographic Space Analysis

Comparing model results displayed significant differences between extracted probabilities of suitable habitat



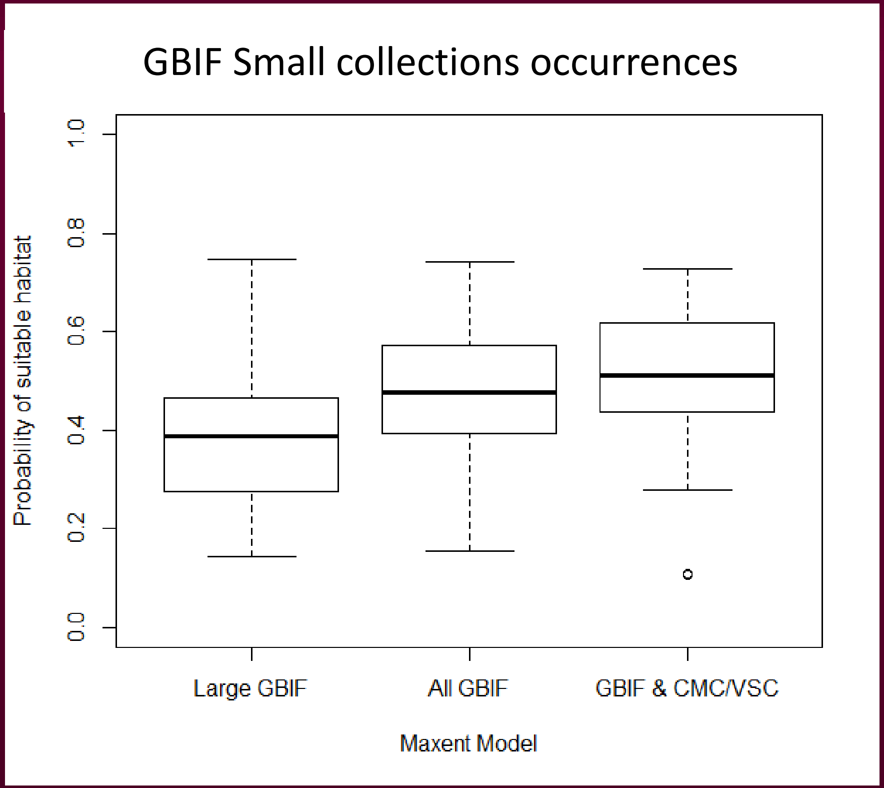
S. tabernaemontani

GBIF Large Collections

Friedman test results

$p < 0.05$

Comparing model results displayed significant differences between extracted probabilities of suitable habitat



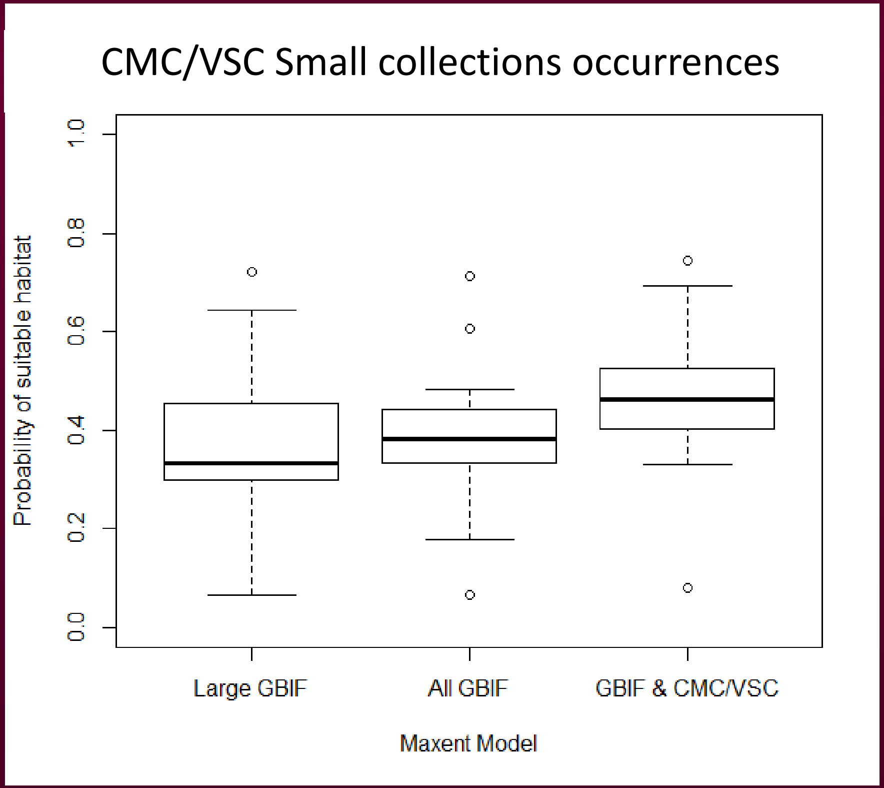
S. tabernaemontani

GBIF Small Collections

Friedman test results

$p < 0.005$

Comparing model results displayed significant differences between extracted probabilities of suitable habitat



S. tabernaemontani

CMC/VSC Collections

Friedman test results

$p < 0.005$

Methodology

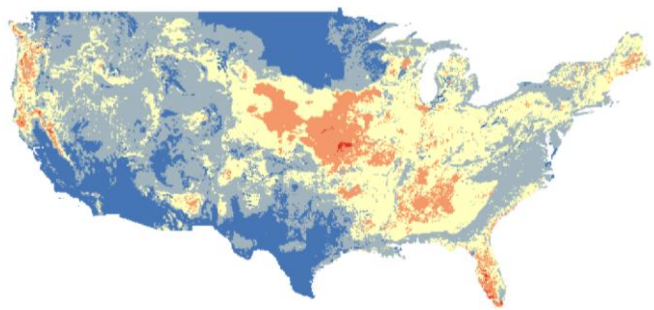
- Obtain Species Occurrence Data
- Filter Data
- Species Distribution Model
- Model Prediction Evaluation
- Geographic Space Analysis

MaxEnt maps displaying the probability of suitable habitat

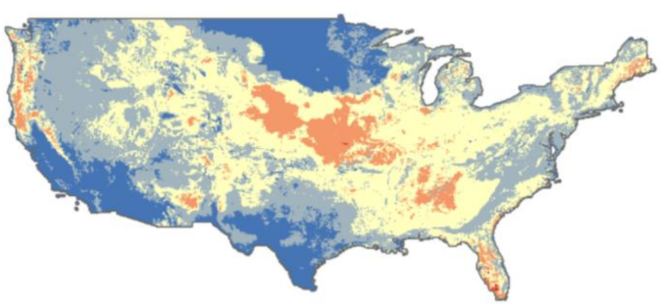
Schoenoplectus tabernaemontani

Probability of Suitable Habitat

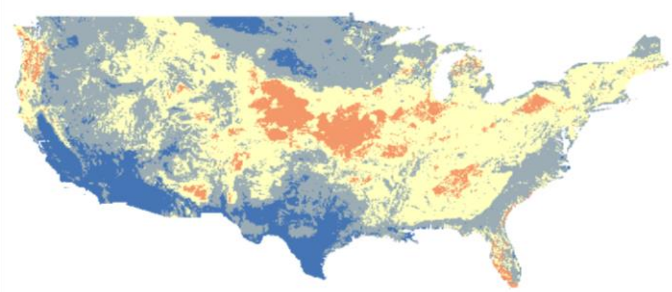
- 0.0 - 0.2
- 0.2 - 0.4
- 0.4 - 0.6
- 0.6 - 0.8
- 0.8 - 1



Model built from GBIF Large Collections



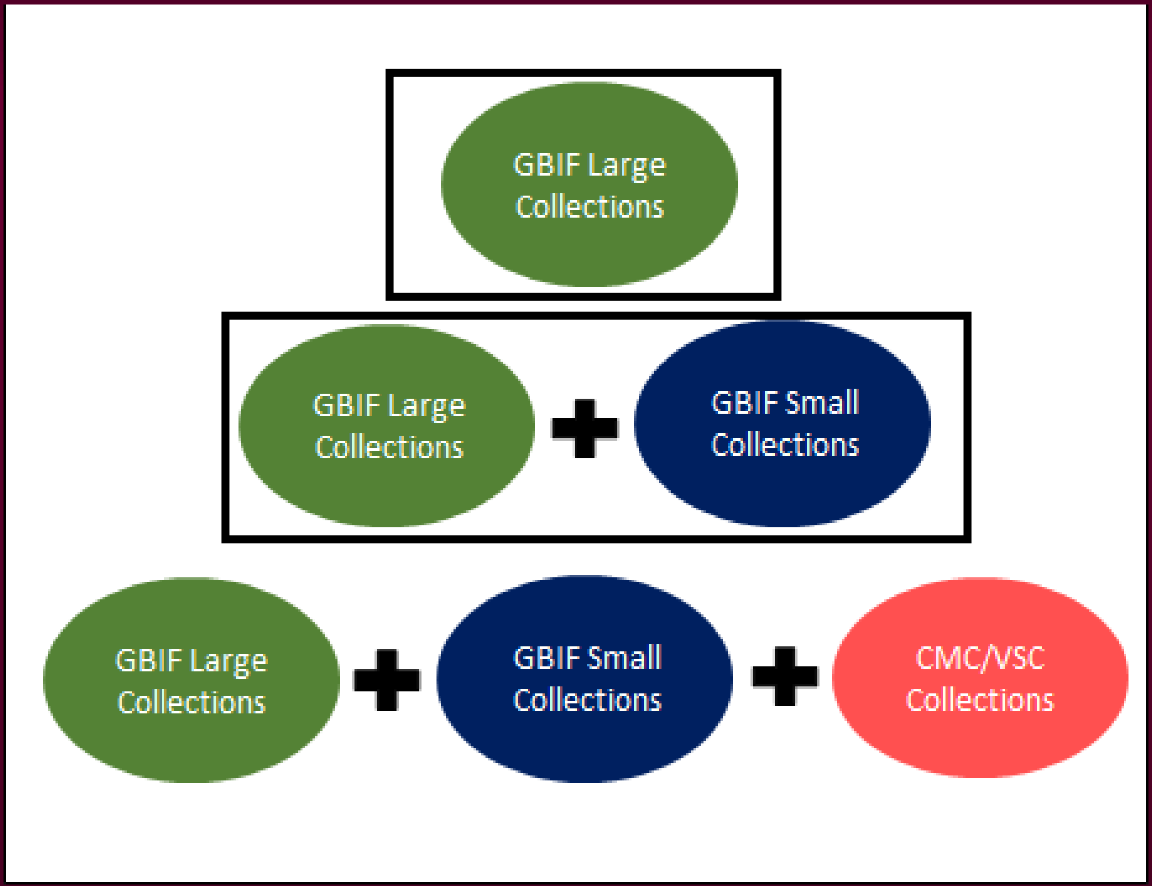
Model built from GBIF Small & Large Collections



Model built from all GBIF Collections & CMC/VSC

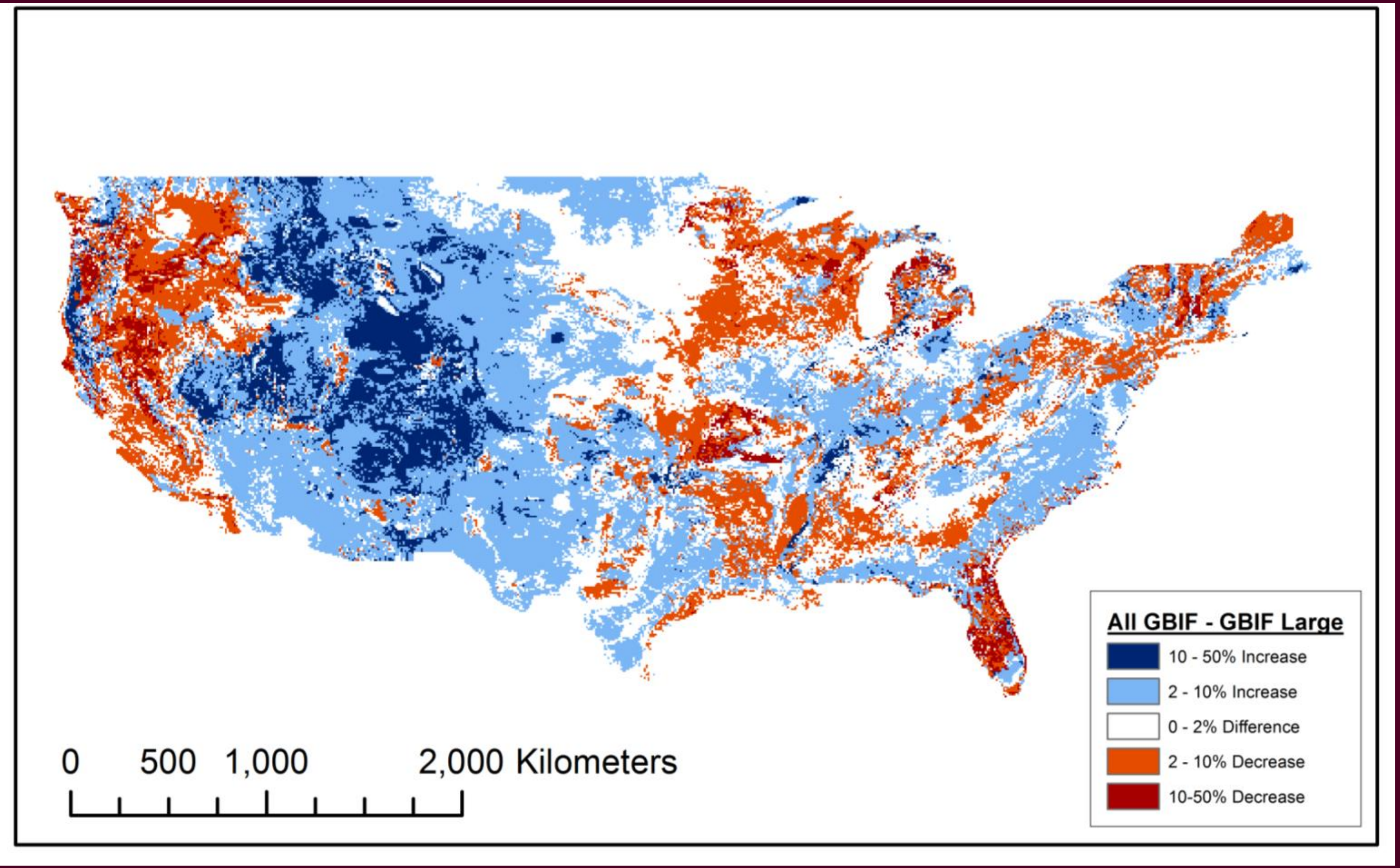
CMU CENTRAL MICHIGAN UNIVERSITY

Comparing geographic predictions between each model

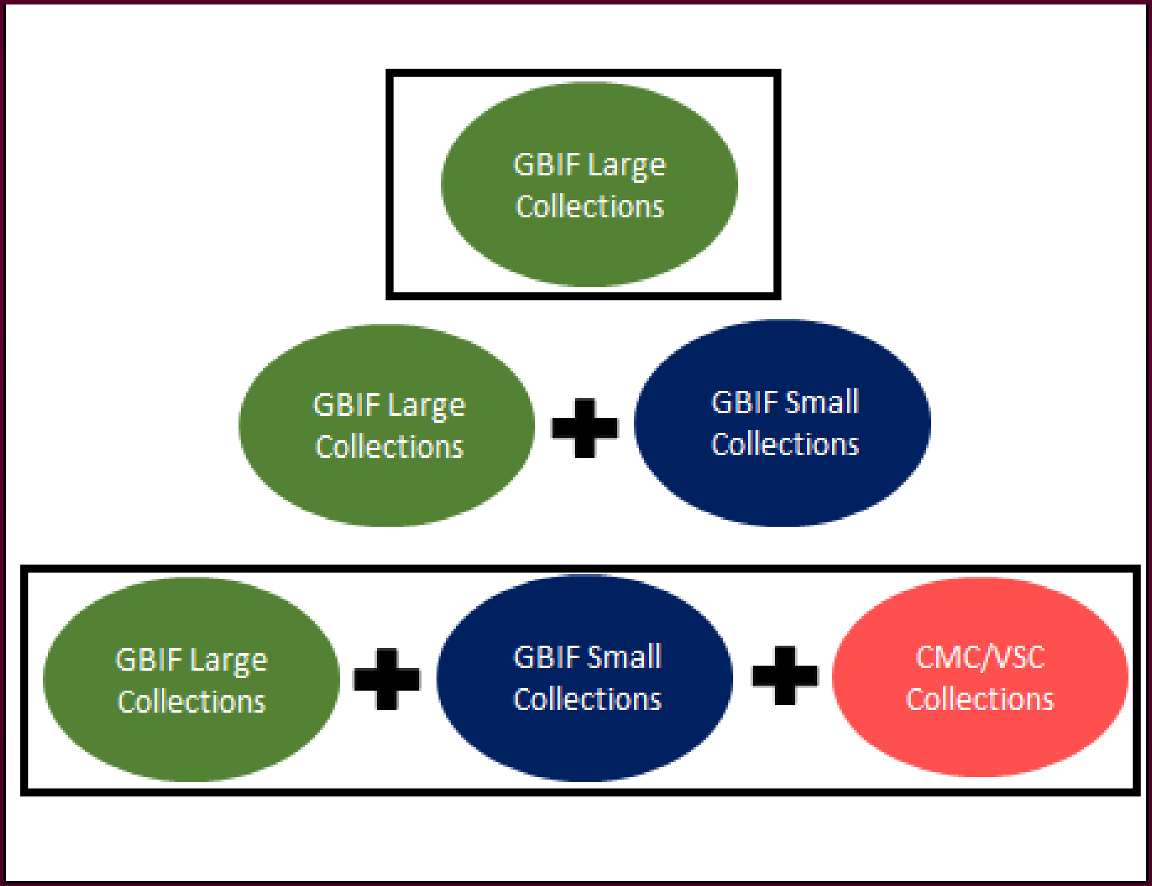


Differences when occurrences from GBIF small collections are added to GBIF large collection based models

Schoenoplectus tabernaemontani

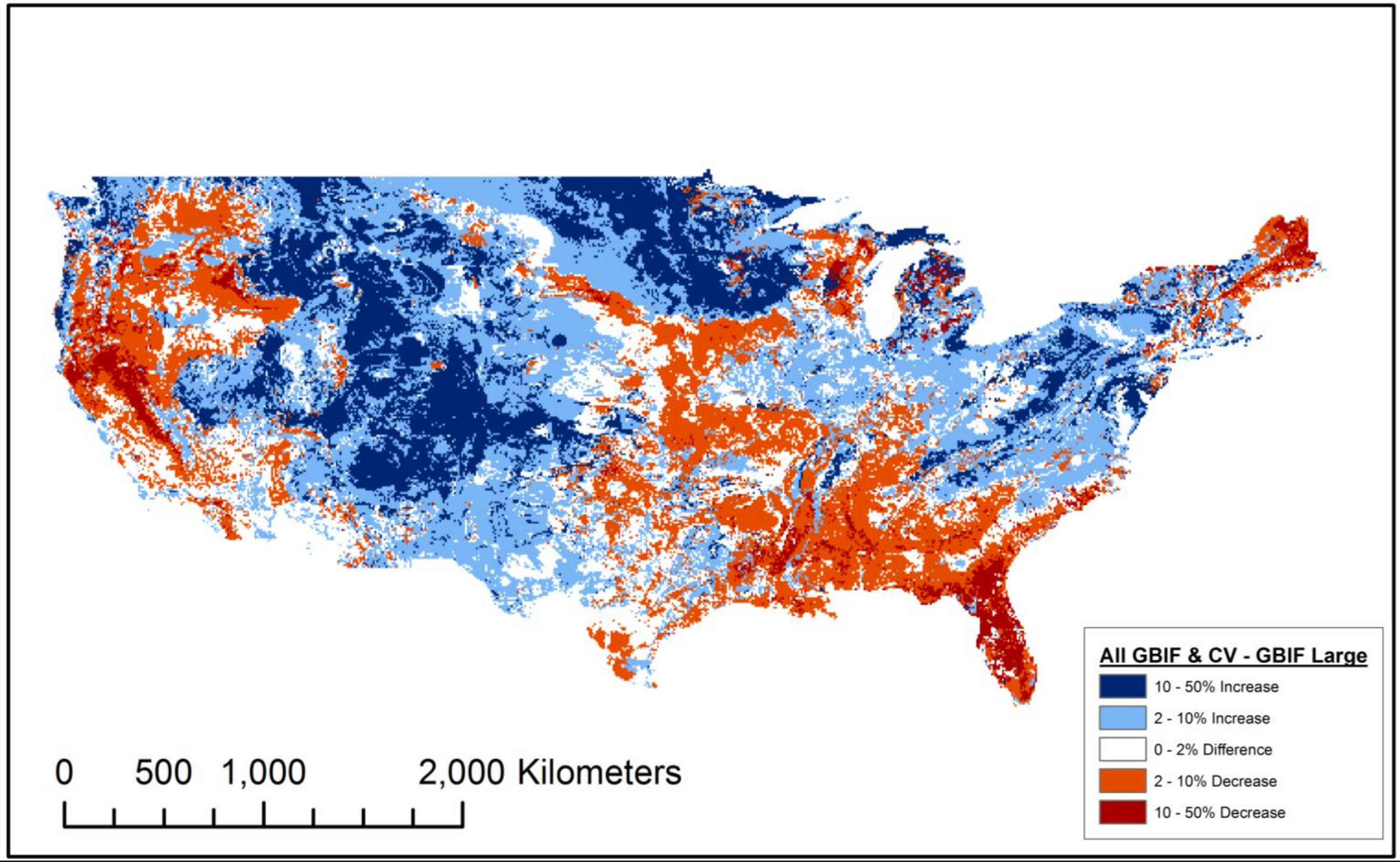


Comparing geographic predictions between each model

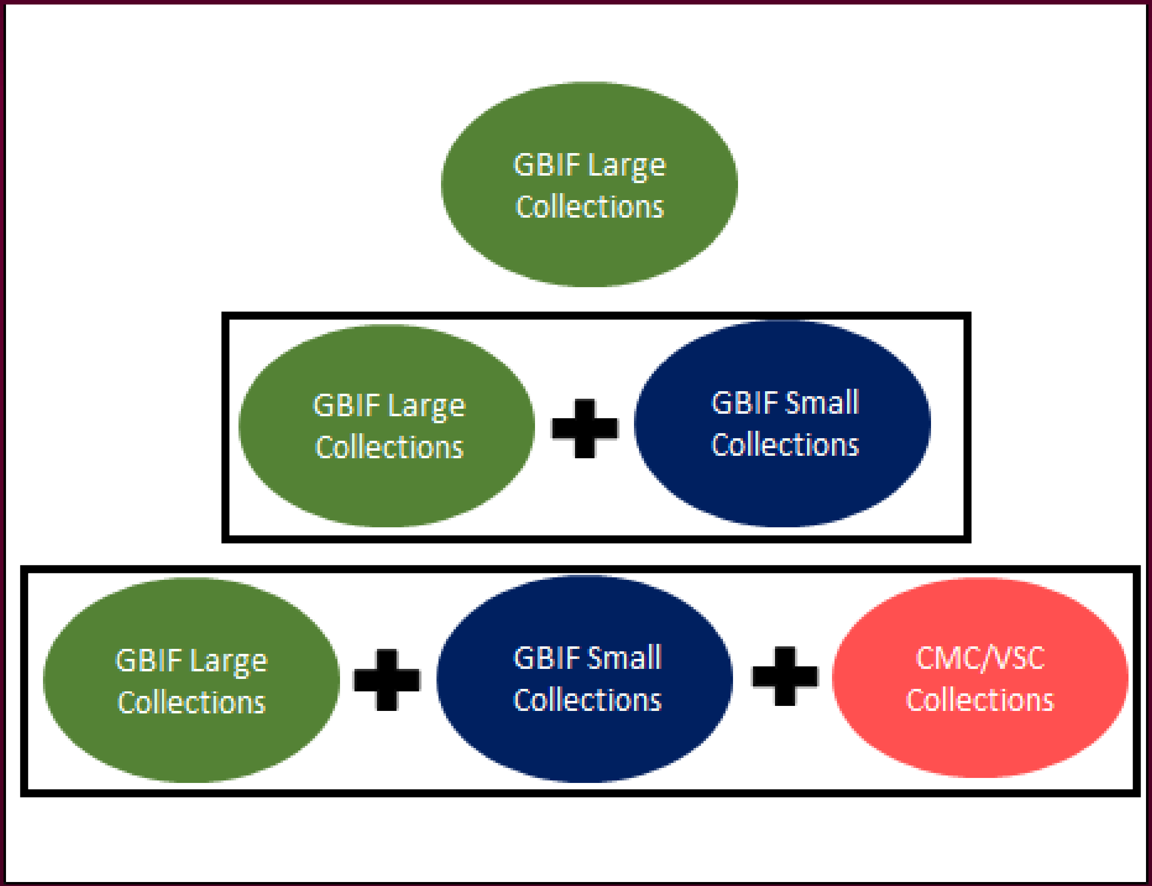


Differences when occurrences from all small collections are added to large collection based models

Schoenoplectus tabernaemontani

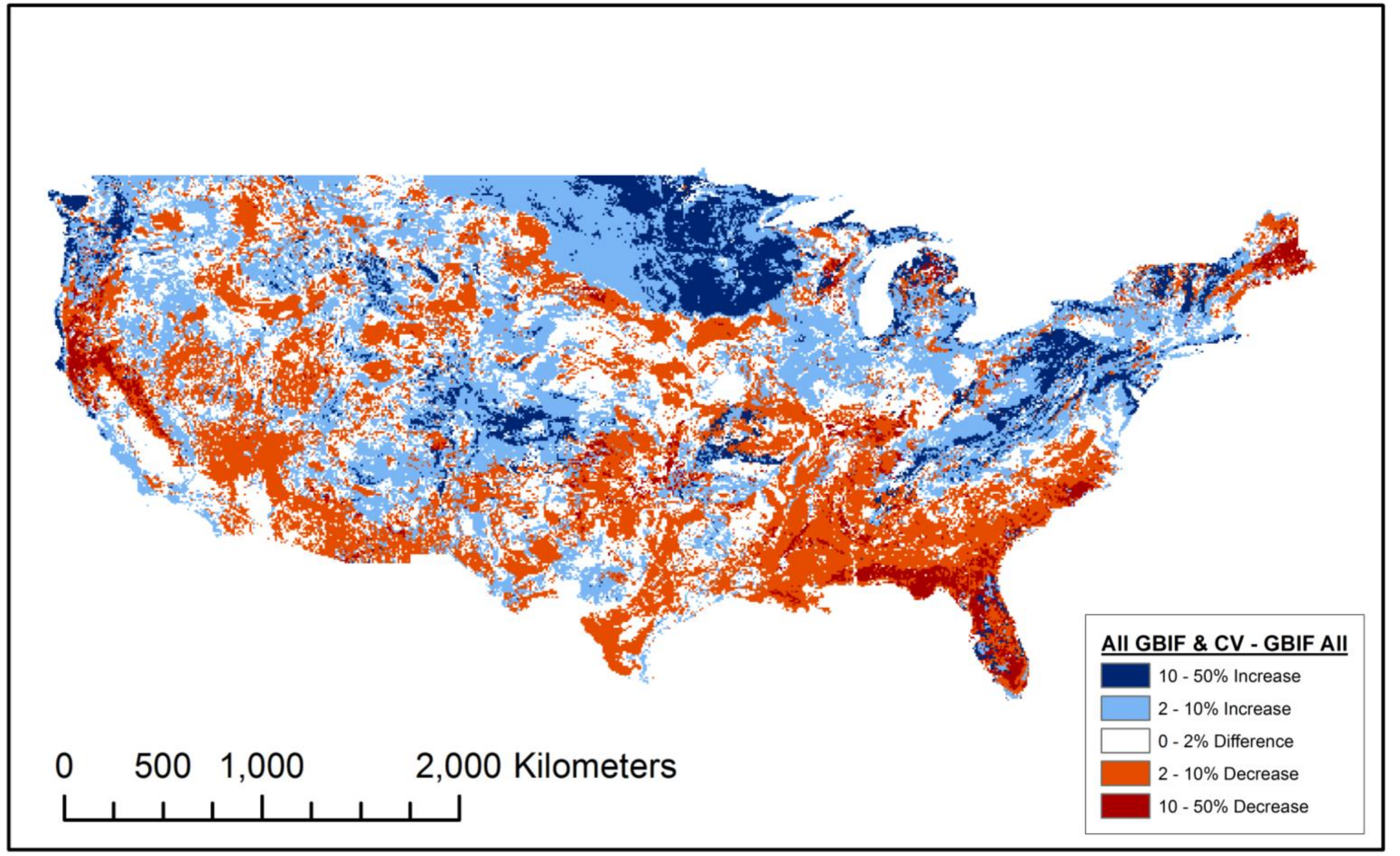


Comparing geographic predictions between each model

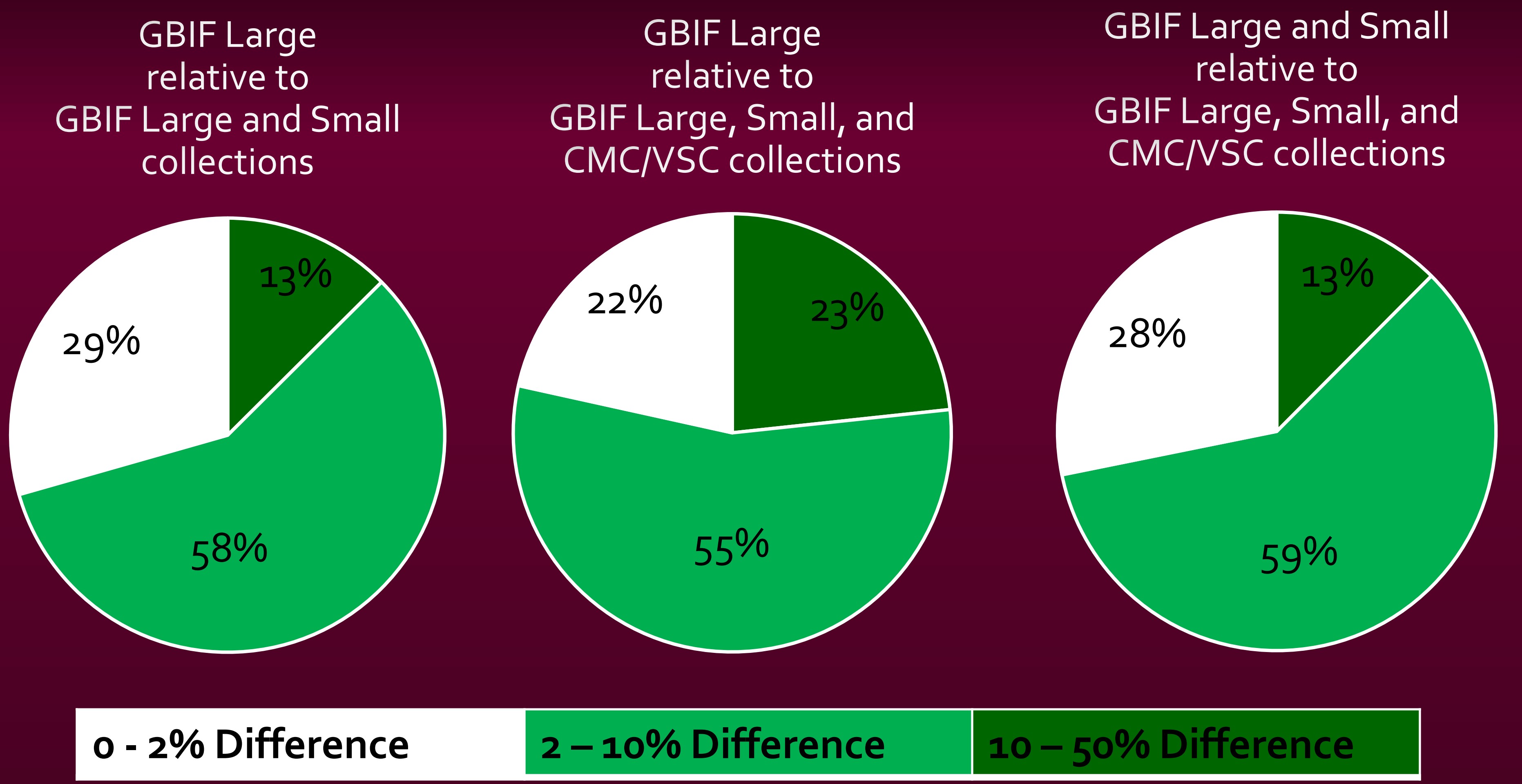


Differences when occurrences from CMC/VSC collections are added to GBIF large and small collections based models

Schoenoplectus tabernaemontani



Comparison of models predicted probability based on different datasets



Small collections are strong contributors to models of species distribution and niche models

- Models inclusive of small collections data resulted in statistically significant increases in occurrence predictions

- Models inclusive of small collections data resulted in a 23% major (10-50%) change in geographic predictions

Small is Big!

- Small collections significantly refine species distribution models

- These collections may represent a small 13% of national specimens, but they are critical to building our understanding of habitats and biodiversity

> *Remember: there are no small parts, only small actors*
>
> – Constantin Stanislavski

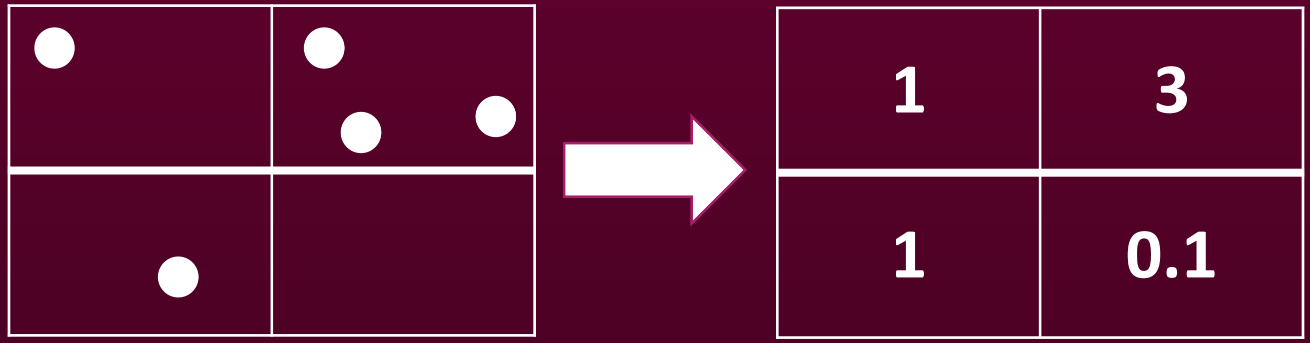
Questions?

CENTRAL MICHIGAN UNIVERSITY

CMU
CENTRAL MICHIGAN UNIVERSITY

Accounting for sampling bias

- Bias file
  - Quantity of sampling across background
  - Samples background data from weighted cells



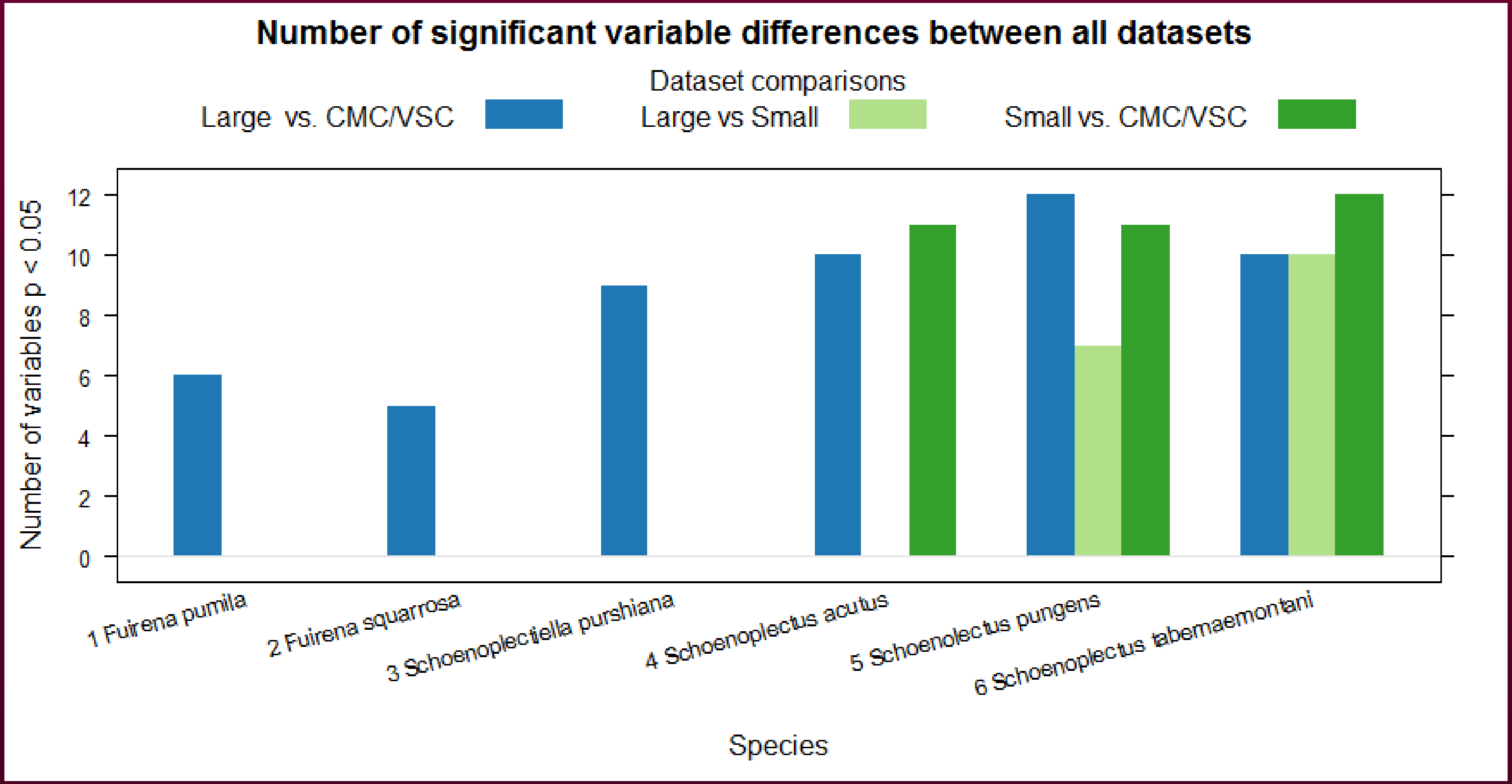
Niche comparison

- Extracted environmental data at each occurrence point for 3 independent datasets

- Compared each variable's set of values among datasets using the Kolmogorov-Smirnov test.
  - *p* <0.05 = a single dataset alone does not contain the "true" realized niche of a species.

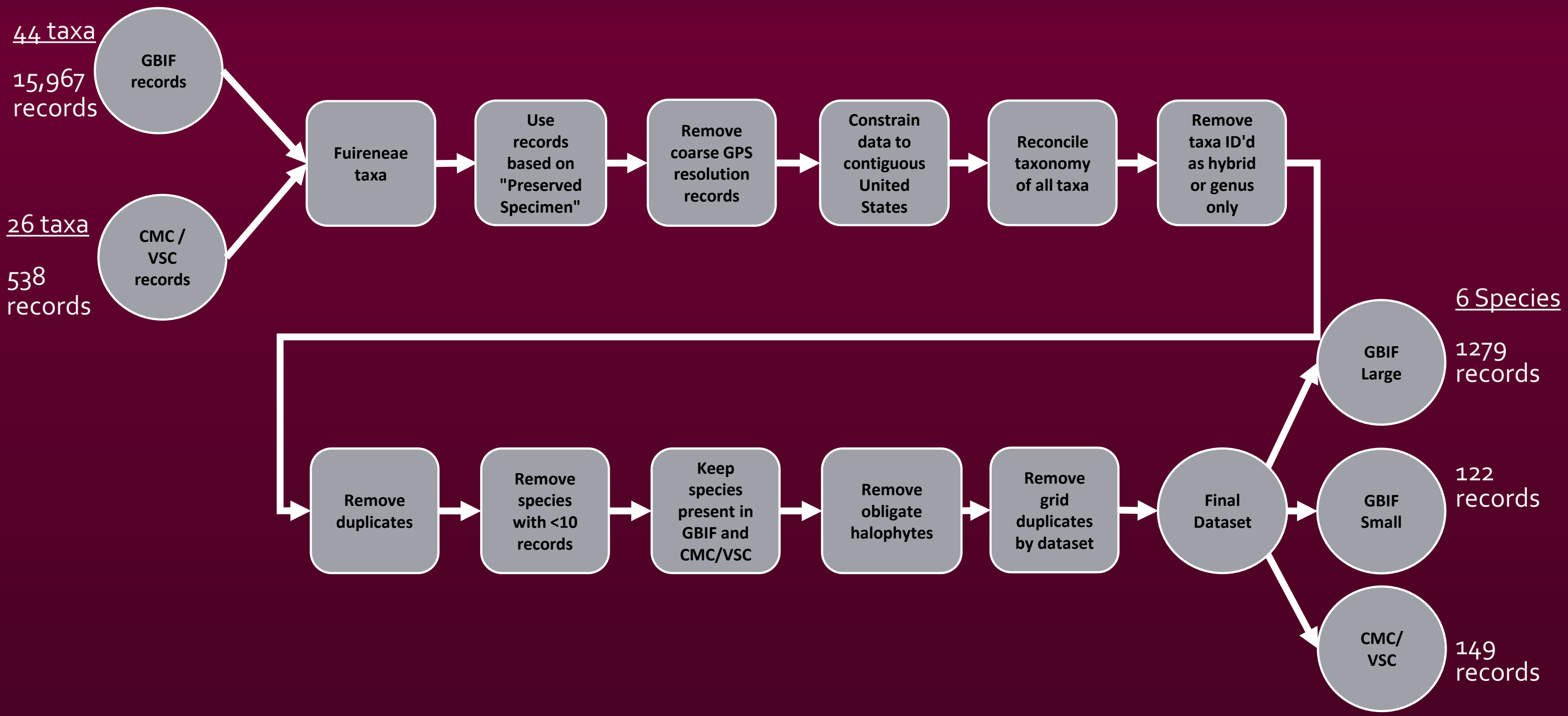
CMU
CENTRAL MICHIGAN UNIVERSITY

Results: Niche comparison



CMU
CENTRAL MICHIGAN UNIVERSITY

Filtering occurrence records

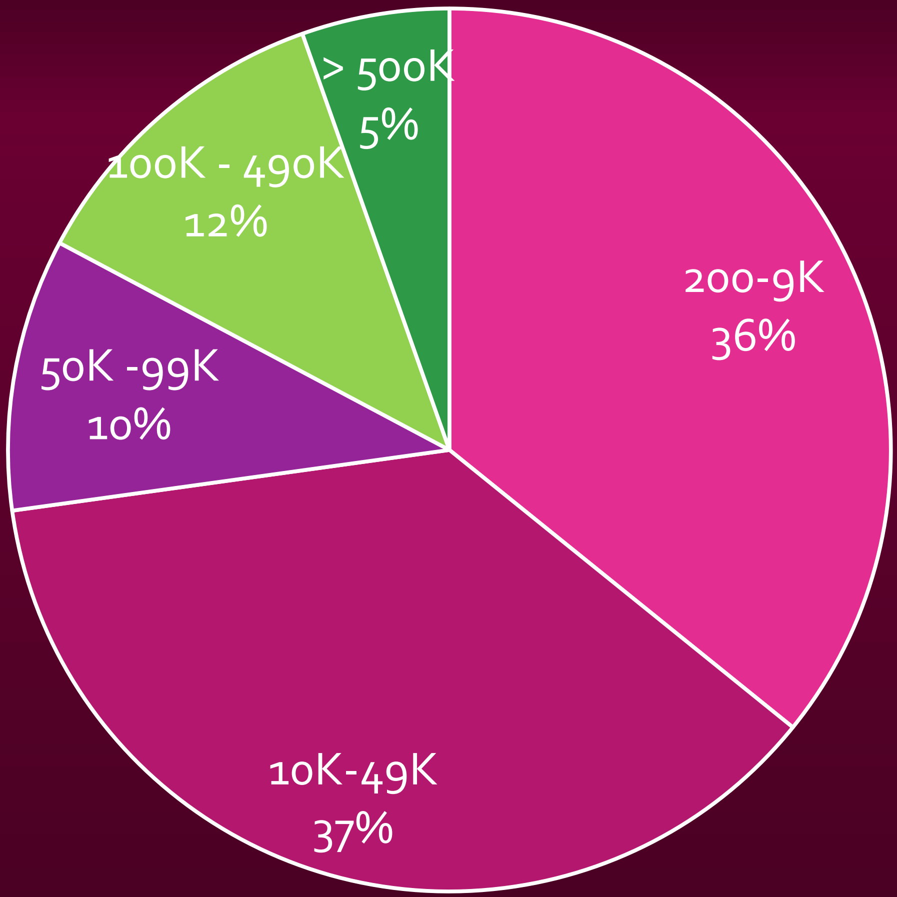


CMU
CENTRAL MICHIGAN UNIVERSITY

Significant differences between predicted distributions are present

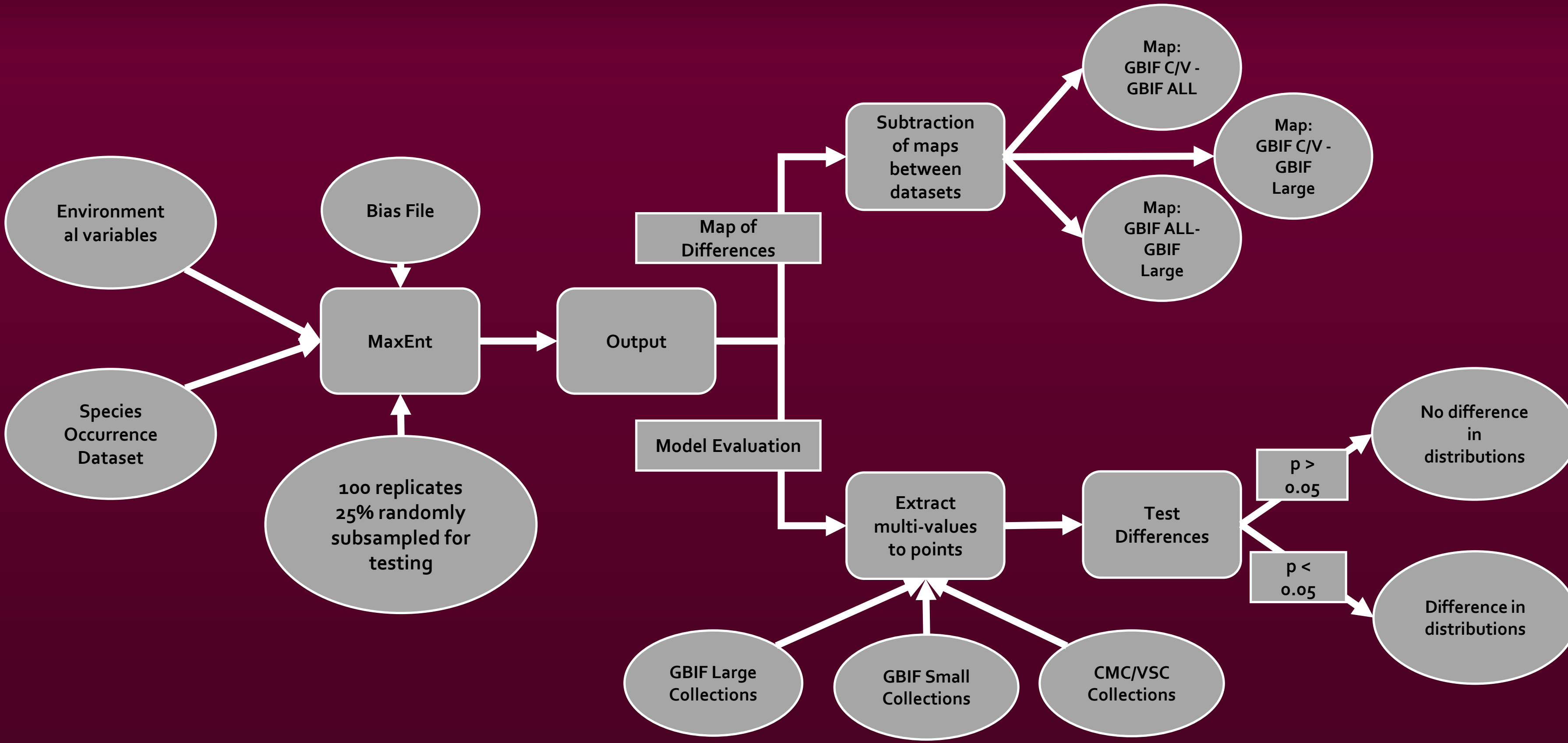
Species	Occurrence Dataset	Wilcox *P* - value	Friedman *P* - value
Fuirena pumila	**Large**	< 0.05	-
	CMC/VSC	< 0.005	-
Fuirena squarrosa	Large	> 0.05	-
	CMC/VSC	< 0.005	-
Schoenoplectiella purshiana	**Large**	< 0.005	-
	CMC/VSC	< 0.005	-
Schoenoplectus acutus	**Large**	-	< 0.05
	Small	-	> 0.05
	CMC/VSC	-	> 0.05
Schoenoplectus pungens	**Large**	-	< 0.05
	Small	-	< 0.05
	CMC/VSC	-	< 0.005
Schoenoplectus tabernaemontani	**Large**	-	< 0.05
	Small	**-**	< 0.005
	CMC/VSC	**-**	< 0.005

Number of Herbaria in each size class

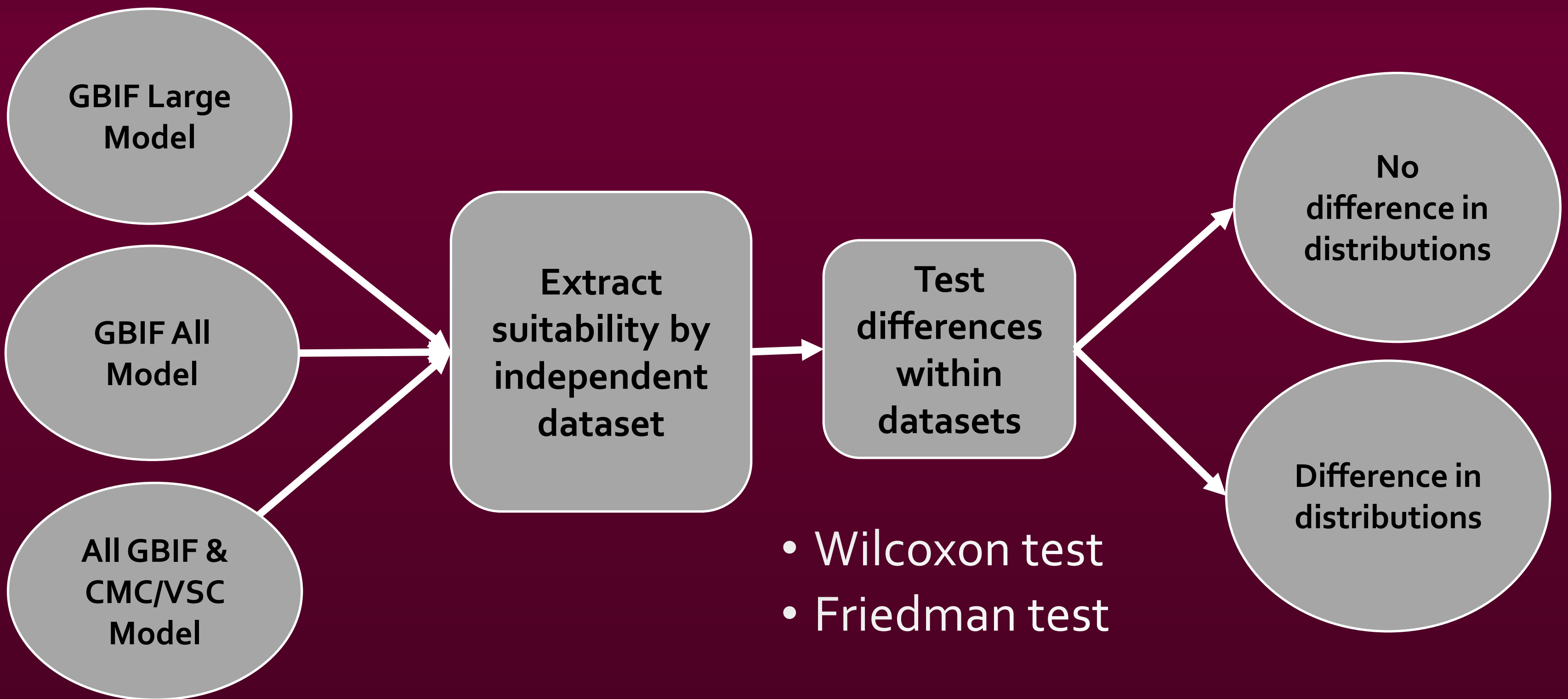


CMU
CENTRAL MICHIGAN UNIVERSITY

Species Distribution Modeling



Assessing likelihood of suitable habitat by dataset



- Wilcoxon test
- Friedman test

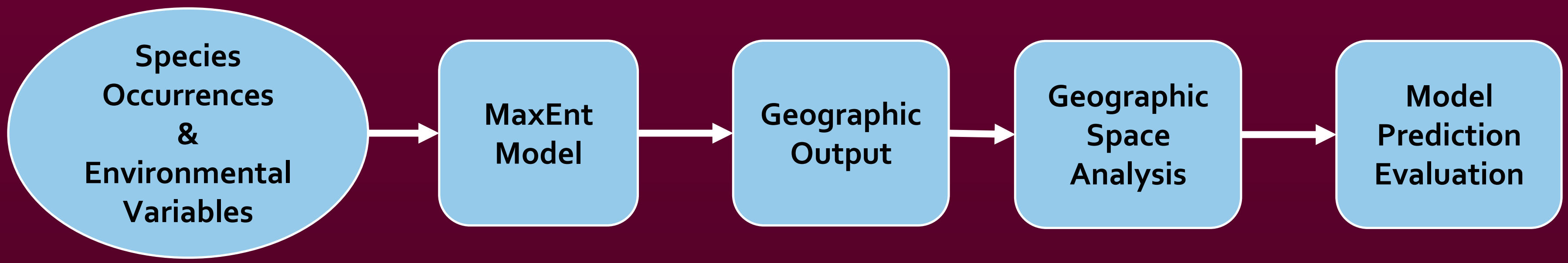
Species Distribution Modeling in brief

- Allows understanding of distributions without having complete sampling of species

- Modeling studies:
  - Habitat suitability modeling (ex. Abdi, 2013; Ballesteros-Mejia et al. 2013)
  - Historical speciation patterns (ex. Liu et al. 2013)
  - Invasive species potential distributions (ex.Gallardo et al 2013)
  - Environmental variable impacts (ex. Oriega & Obero, 2013)
  - Distributions under climate change (ex. Jueterbock et al., 2013; Kriticos et al., 2013)

- Models are reliant on the data that is put into them.

Species Distribution Modeling



References

- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.
- Soil Survey Staff, Natural Resources Conservation Service U.S.D. of A. Web Soil Survey: STATSGO2.

CMU
CENTRAL MICHIGAN UNIVERSITY