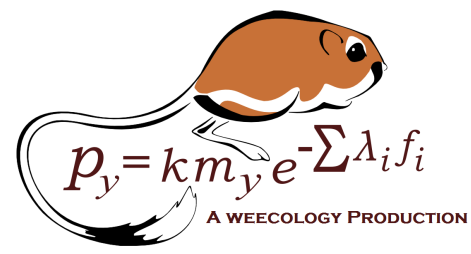
A data-intensive assessment of the species-abundance distribution.

Elita Baldridge

UtahStateUniversity
ECOLOGY CENTER



Macroecology

One approach to studying ecological patterns and processes.

- Data intensive.
- Large scales
  - Spatial
  - Temporal
  - Taxonomic
- Search for generality.

Macroecology

Criticisms of macroecology

- North American terrestrial bias.
- Lack of identification of pattern generating mechanisms.

Macroecology

Best practice recommendations

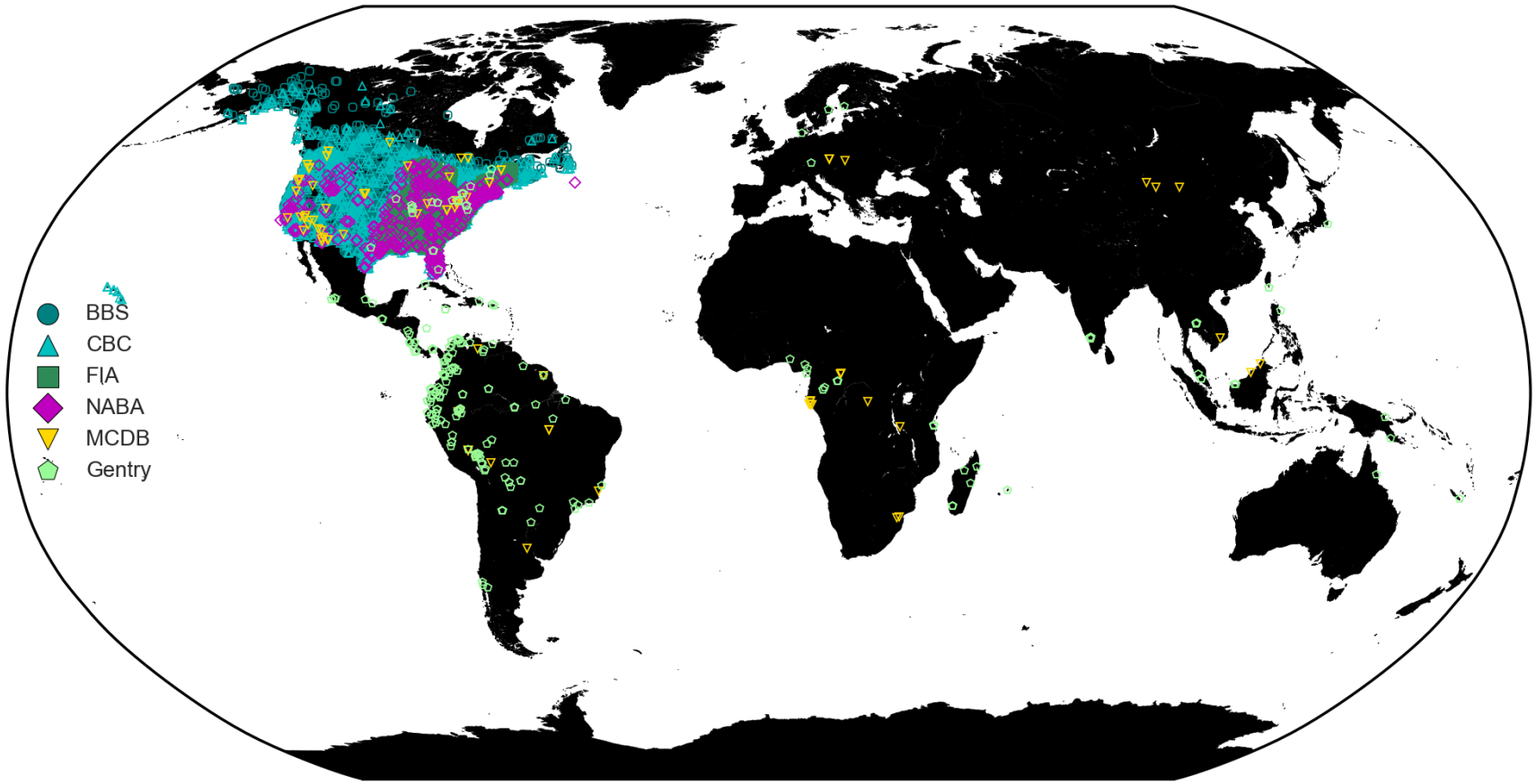
- Test patterns with multiple taxonomic groups/ecosystems.
- Simultaneous testing of competing models and model predictions with a consistent statistical approach.

The Rules of Ecoinformatics

Garbage in, garbage out.

- All data are good, not all data are appropriate.
- Fit the data to the question.

Data



Data

Major macroecological datasets

- Largely terrestrial
- Largely North American
- Many publicly available, some not.

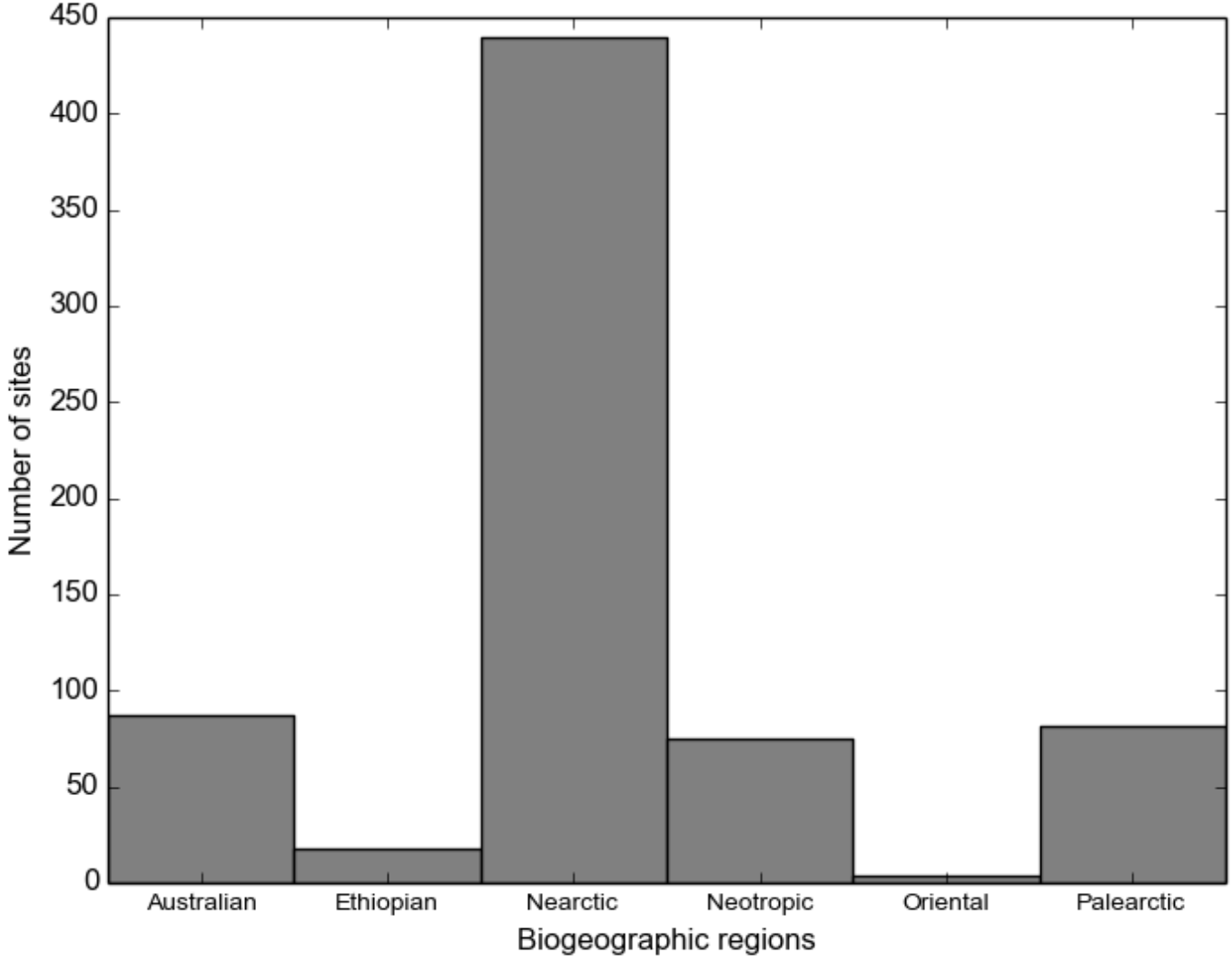
Lots of data in the literature.

Data

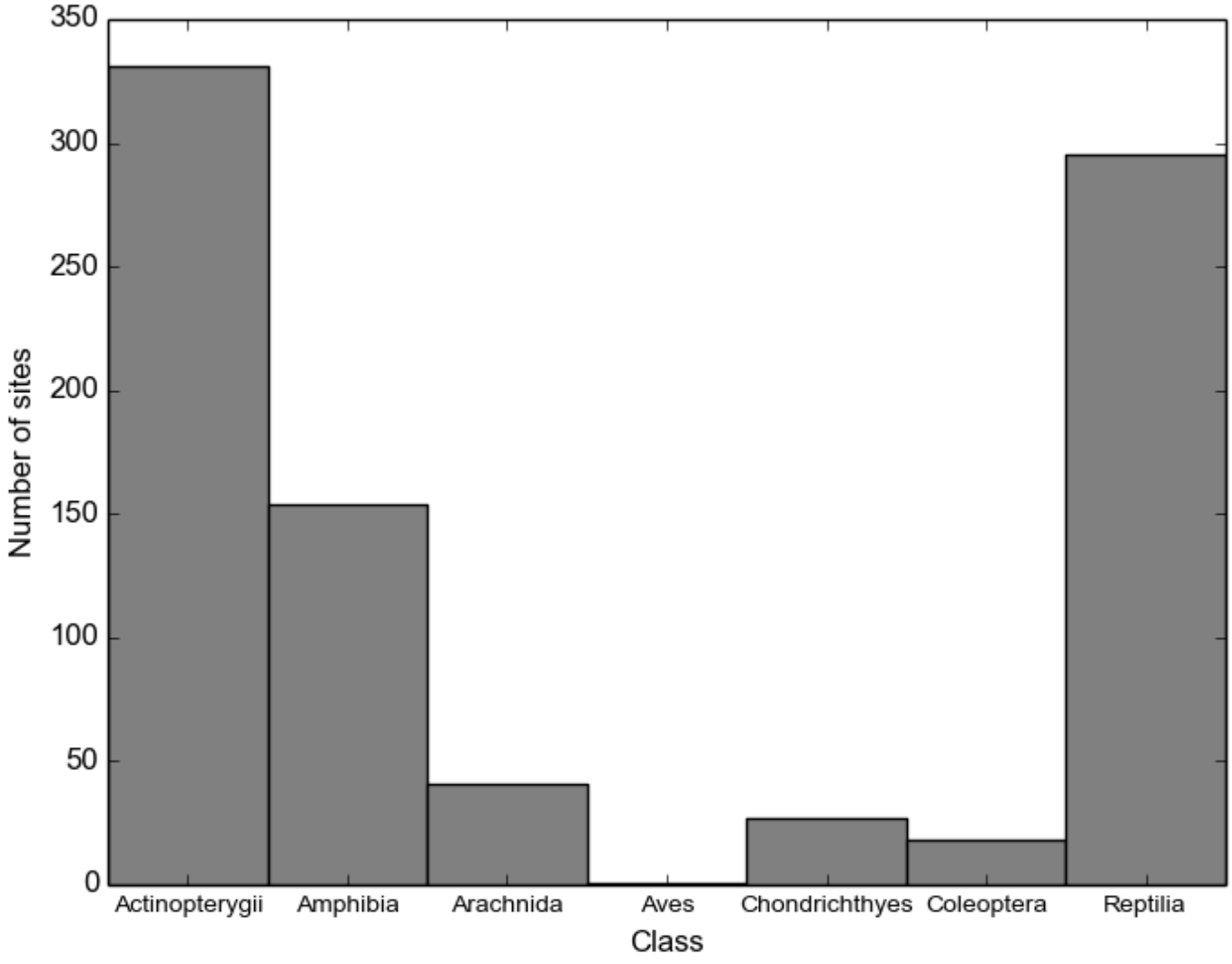
Variable name	Variable definitions
Class	Taxonomic class of species
Family	Taxonomic family of species
Genus	Taxonomic genus of species
Species	Specific epithet of species
Relative_abundance	Relative abundance of species
Abundance	Abundance of species
Collection_Year	Start of collecting
End_Collection	End of collecting
Site_Name	Name/description of site
Biogeographic_region	Biogeographic region
Site_notes	Additional site information

Table : List of variables collected.

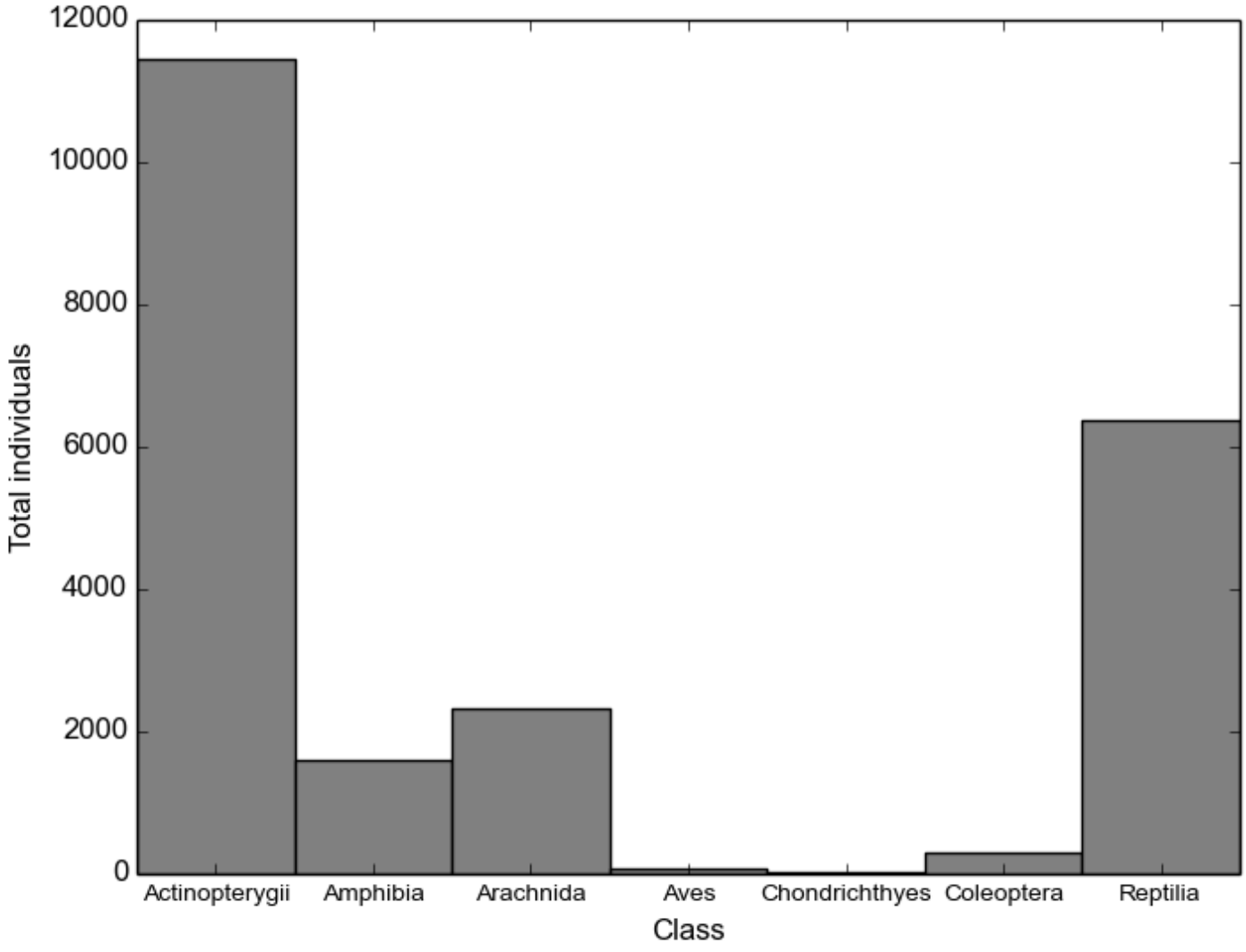
Data


450
400
350
300
250
200
150
100
50
0
Number of sites
Australian
Ethiopian
Nearctic
Neotropic
Oriental
Palearctic
Biogeographic regions

Data


350
300
250
200
150
100
50
0
Number of sites
Actinopterygii
Amphibia
Arachnida
Aves
Chondrichthyes
Coleoptera
Reptilia
Class

Data


12000
10000
8000
6000
4000
2000
0
Total individuals
Actinopterygii
Amphibia
Arachnida
Aves
Chondrichthyes
Coleoptera
Reptilia
Class

Data

The final community abundance database is publicly available and importable through the EcoData Retriever. (http://www.ecodataretriever.org/)

Commonness & rarity

The species abundance distribution:

- Describes the distribution of commonness & rarity of species.
- One of the most fundamental and ubiquitous patterns in ecology.
- Exhibits a hollow curve distribution.
  - Many rare species.
  - Few common species.
- Many forms of the species abundance distribution (SAD).

Forms of the SAD

Model classes:

- Purely statistical
- Branching process
- Population dynamics
- Niche partitioning
- Maximum entropy
- Feasible set/combinatorics

SAD Comparisons

Most comparisons of the different models:

- Use only a small subset of available models (typically two).
- Focus on a single ecosystem or taxonomic group
- Fail to use the most appropriate statistical methods.

SAD Comparisons

Selected five models from four classes for comparison.

Model class	Form of the distribution
Purely statistical	Logseries, Poisson lognormal
Branching process	Zipf
Population dynamics	Negative binomial
Niche partitioning	Geometric

Table : After B.J. McGill et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology letters 10: 995-1015.

SAD Comparisons

Analysis:

- Model fitting with maximum likelihood estimation.
- Likelihood based model selection to compare the fits of the different models.
- Model comparison with corrected Aikaike Information Criterion (AICc) weights.

SAD Comparisons

Computational tools:

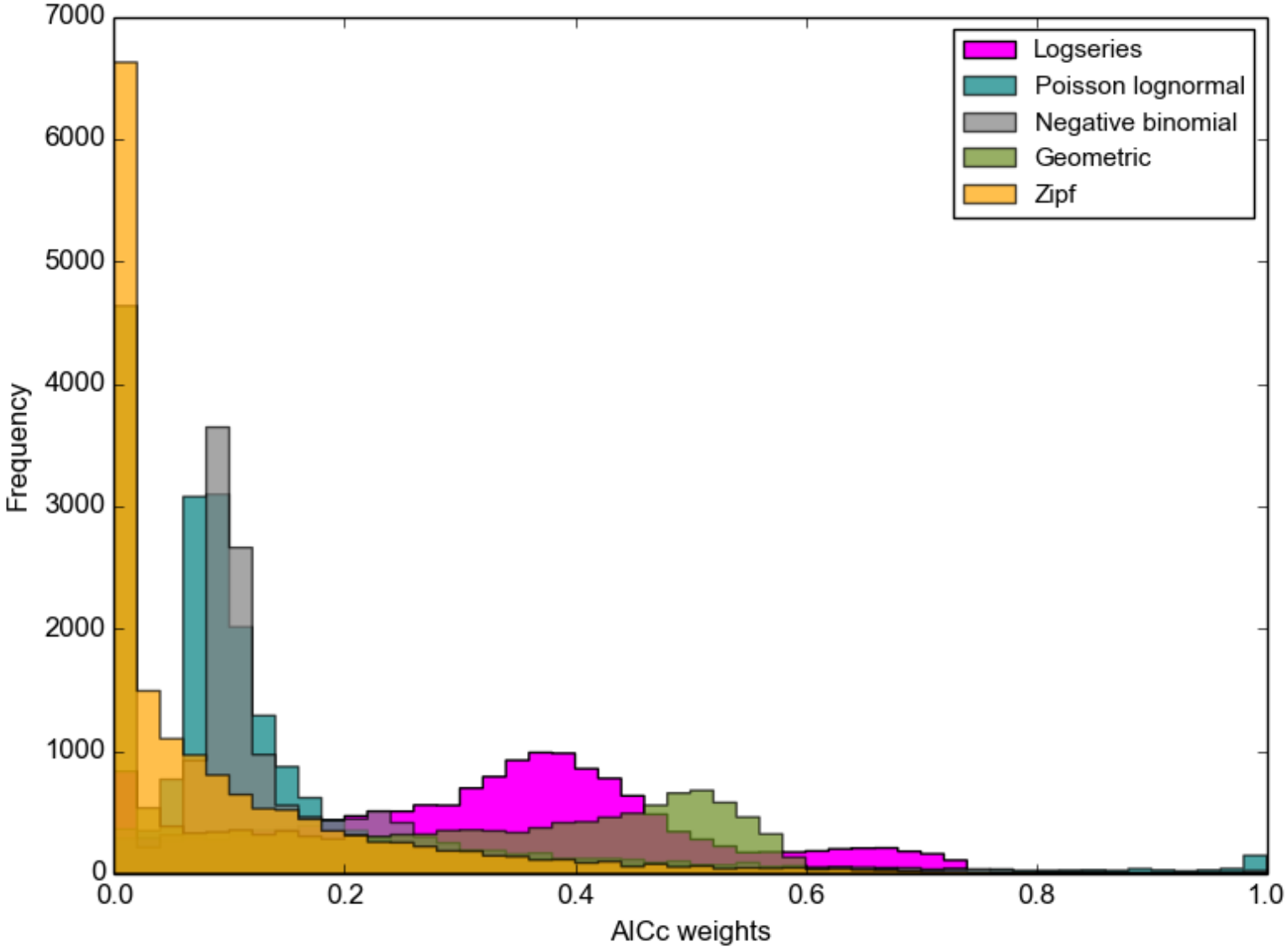
- Model fitting, log-likelihood, and AICc calculations performed with macroecotools Python package.
  (https://github.com/weecology/macroecotools)
- All of the analysis code and the majority of the data is publicly available.
  (https://github.com/weecology/sad-comparison)

SAD Comparisons

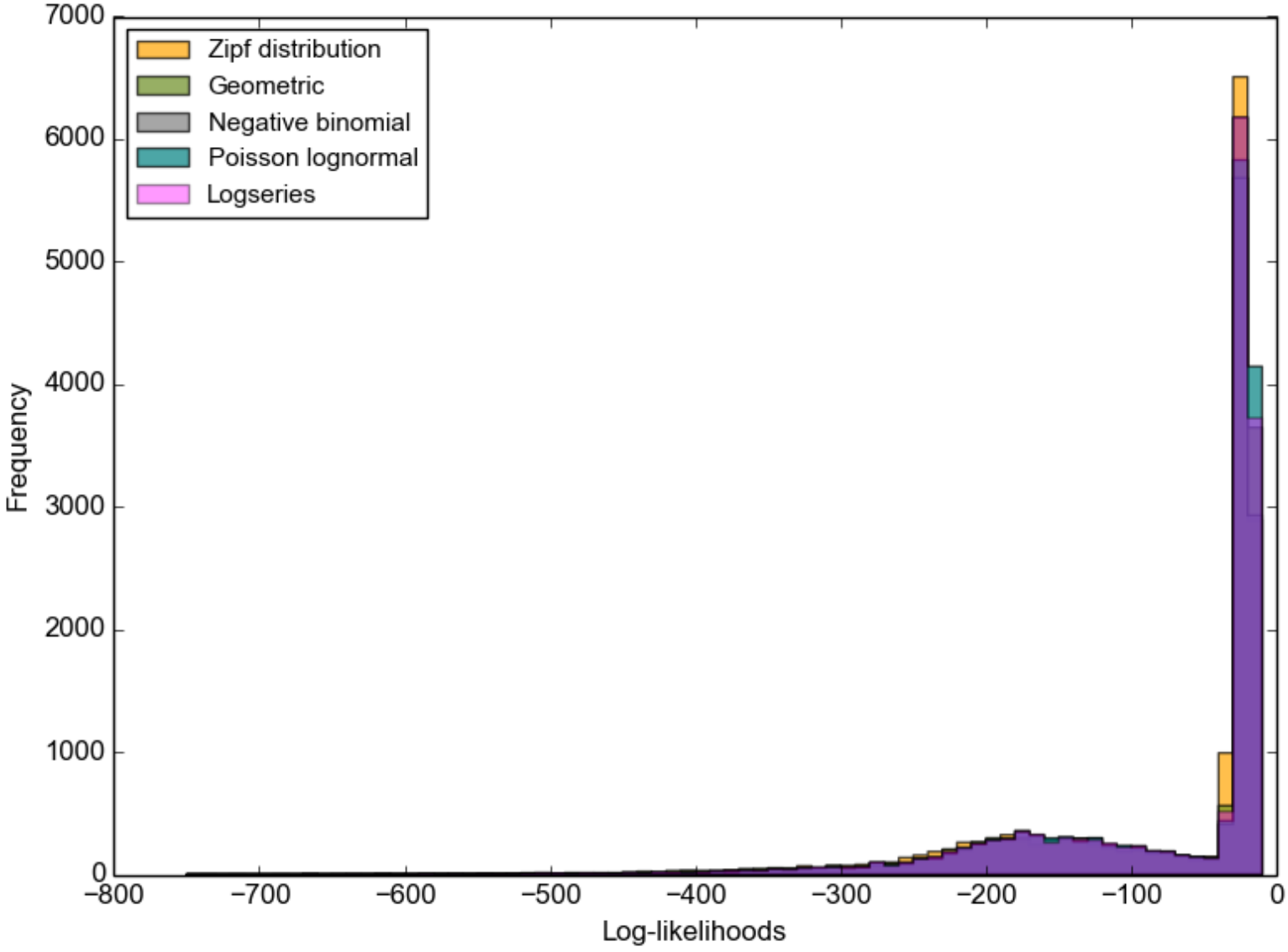
Dataset	Dataset code	Availability	Total sites
Breeding Bird Survey	BBS	Public	2769
Christmas Bird Count	CBC	Private	1999
Gentry's Forest Transects	Gentry	Public	10355
Forest Inventory Analysis	FIA	Public	220
Mammal Community Database	MCDB	Public	103
N. American Butterfly Count	NABA	Private	400
Actinopterygii, this dissertation	Actinopterygii	Public	161
Reptilia, this dissertation	Reptilia	Public	138
Amphibia, this dissertation	Amphibia	Public	43
Coleoptera, this dissertation	Coleoptera	Public	5
Arachnida, this dissertation	Arachnida	Public	25

Table : Datasets used for species-abundance distribution comparisons. Datasets marked as Private were obtained through data requests to the providers resulting in Memorandums of Understanding governing data use.

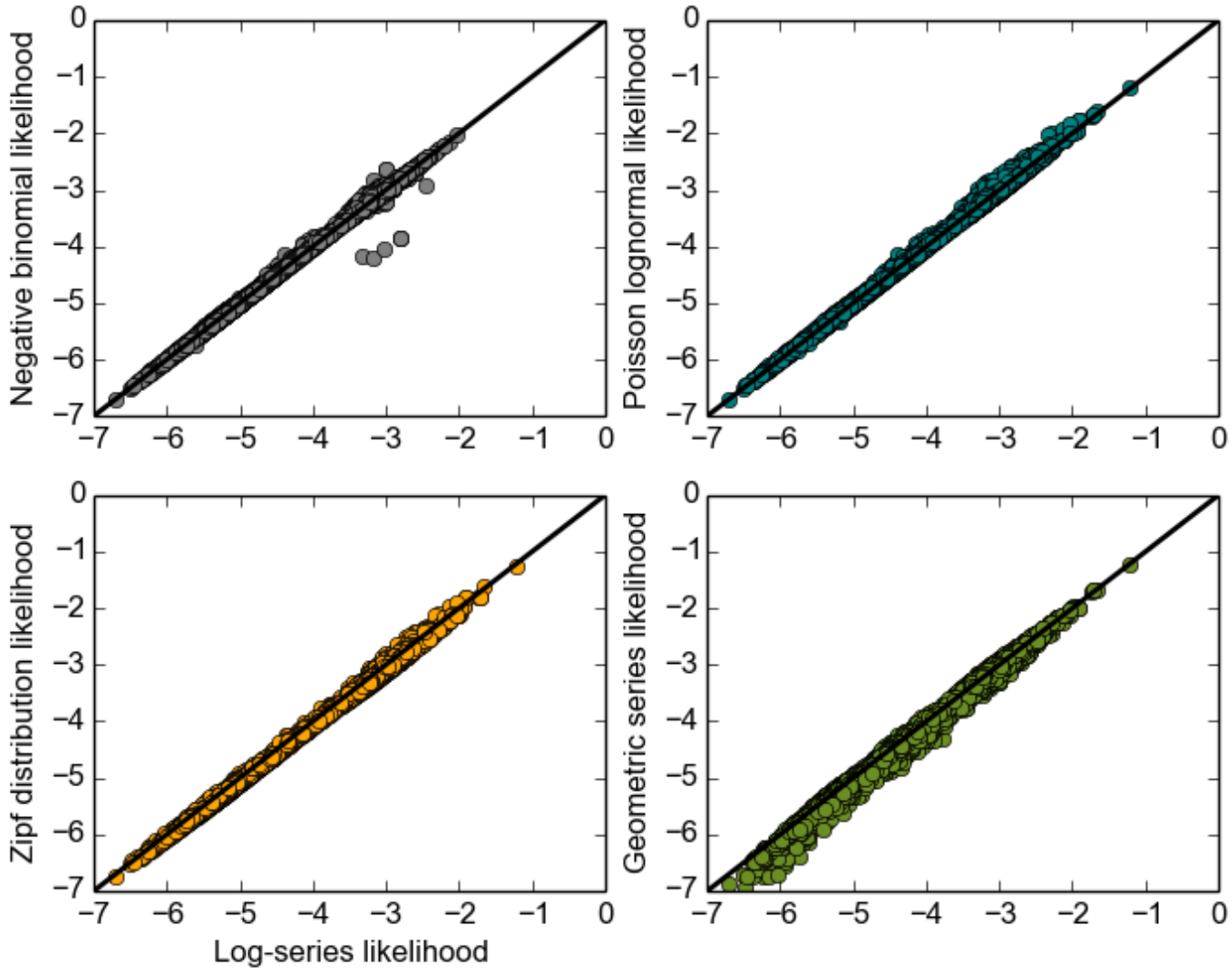
SAD Comparisons



SAD Comparisons



SAD Comparisons



SAD Comparisons

Existing models provide equivalently good absolute fits to empirical data.

- Models with fewer parameters perform better in AIC-based model selection.
- Logseries provides a good naive model for fitting SADs.
    - Produces equivalent likelihoods.
    - Has a single fitted parameter.
    - Easy to fit to empirical data.
    - Best overall model.

SAD Comparisons

Identifying pattern generating mechanisms:

- Compare predictions of different models using multiple macroecological patterns simultaneously.
- Examine scale dependence of pattern.

However, identification of mechanism may not be necessary for prediction.

Neutral Analysis

The unified neutral theory of biodiversity:

- Multiple formulations.
  - Species and individuals are ecologically and demographically equivalent.
  - Stochastic variation in birth, death, immigration, & speciation results in species abundance differences.

Neutral Analysis

Early tests of neutral theory based on comparing the fit of empirical species abundance distributions to the neutral prediction.

Later tests suggested species abundance comparisons were insufficient for a rigorous test of neutrality.

Neutral Analysis

Connolly et al. 2014 identified non-neutral species abundance distributions in marine communities.

- Compared model fits of a non-neutral distribution (Poisson lognormal) to a neutral distribution (negative binomial distribution).

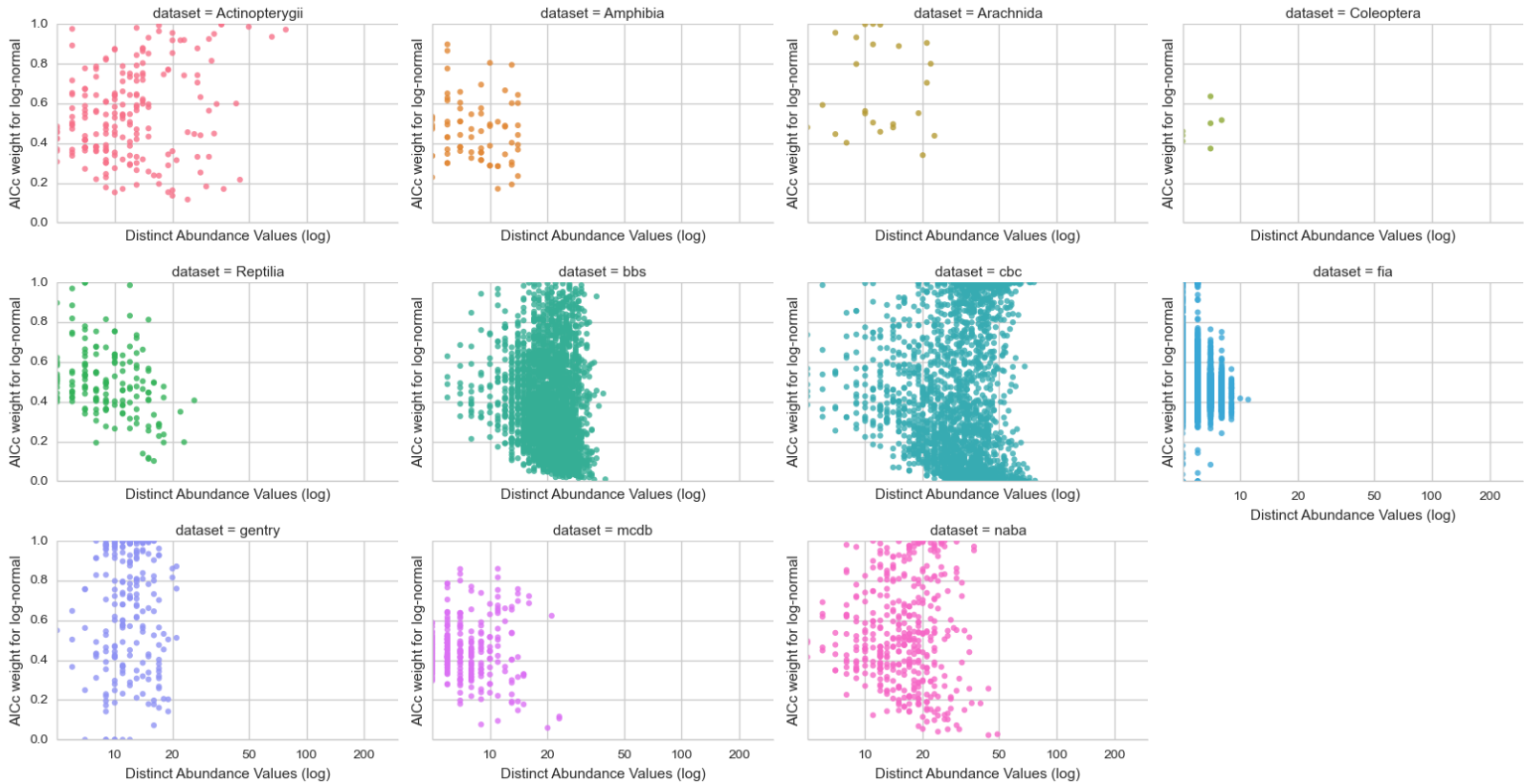
May be a robust method for identifying communities that exhibit non-neutrality.

S.R. Connolly et al. 2014. Commonness and rarity in the marine biosphere. PNAS 111: 8524-8529.

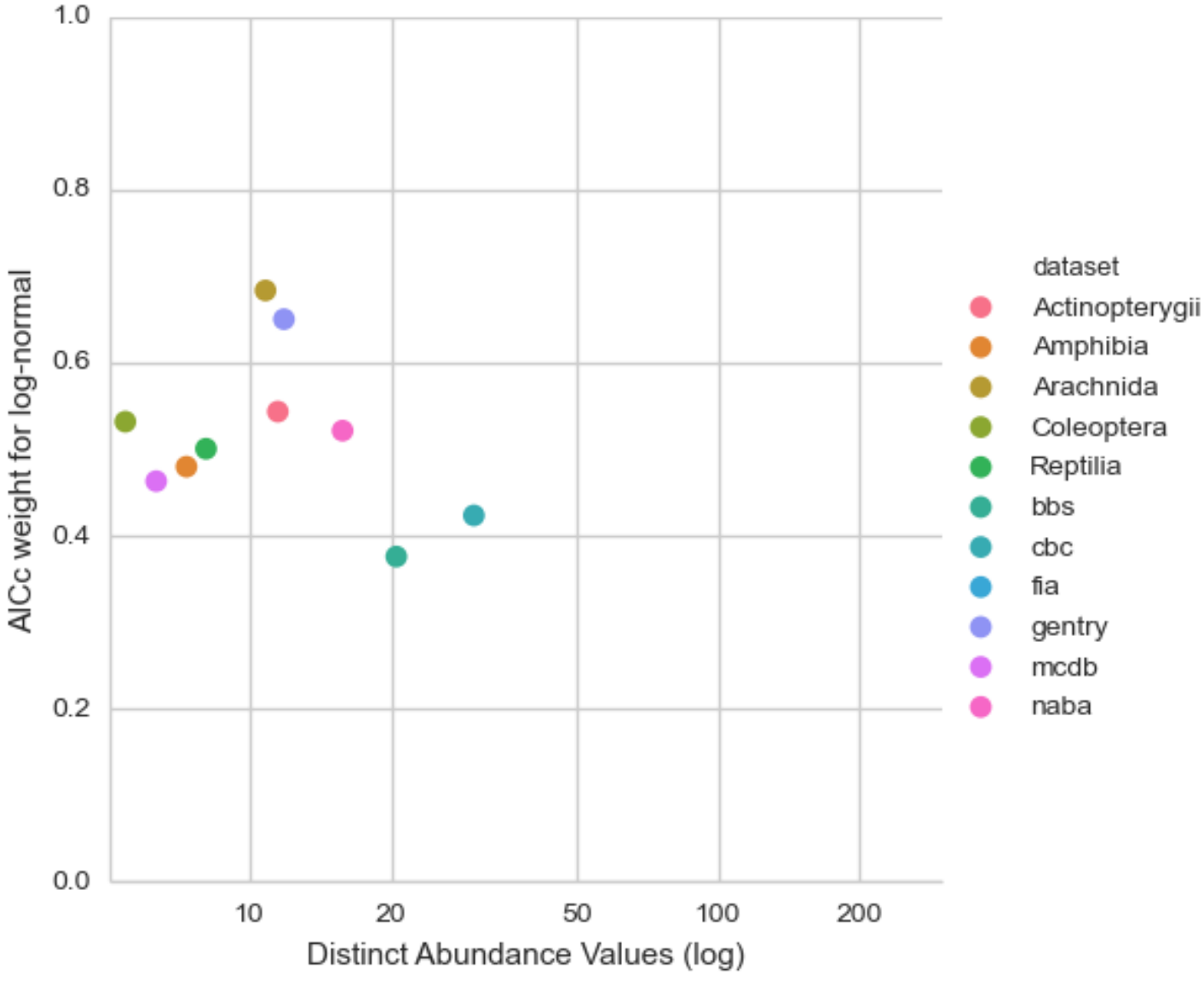
Neutral Analysis

Used the same data and model fitting approach.
Compared a non-neutral model (Poisson lognormal) to a neutral model (negative binomial).

Neutral Analysis



Neutral Analysis



Neutral analysis

Difficult to identify a clear winning model.

- Results consistent with our species abundance distribution model comparisons.
- Results different from Connolly et al. 2014.
    - Non-neutral model outperforms the neutral model in marine systems.
    - Our results suggest marine systems more generally approximated by non-neutral dynamics; terrestrial systems more variable between neutral and non-neutral dynamics.

Conclusions

Challenging to infer process from species abundance distributions alone.

- Multiple mechanisms proposed for each SAD formulation.
- Broad model categorization (i.e. neutral or non-neutral) may be more productive.
- May not be one single suite of processes that dominates.

Conclusions

Challenges in identifying mechanism among datasets.

- Biological vs. non-biological differences (spatial structuring, sampling intensity).
- Diverse data removes uncertainty about non-biological pattern generating mechanisms.
- Even with a great deal of data, identifying mechanism is still challenging.

Conclusions

Predictive macroecology

- Traditional approach is pattern to process to prediction.
- May be possible to generate robust ecological predictions from general patterns.
- Process and prediction may be two separate research goals.

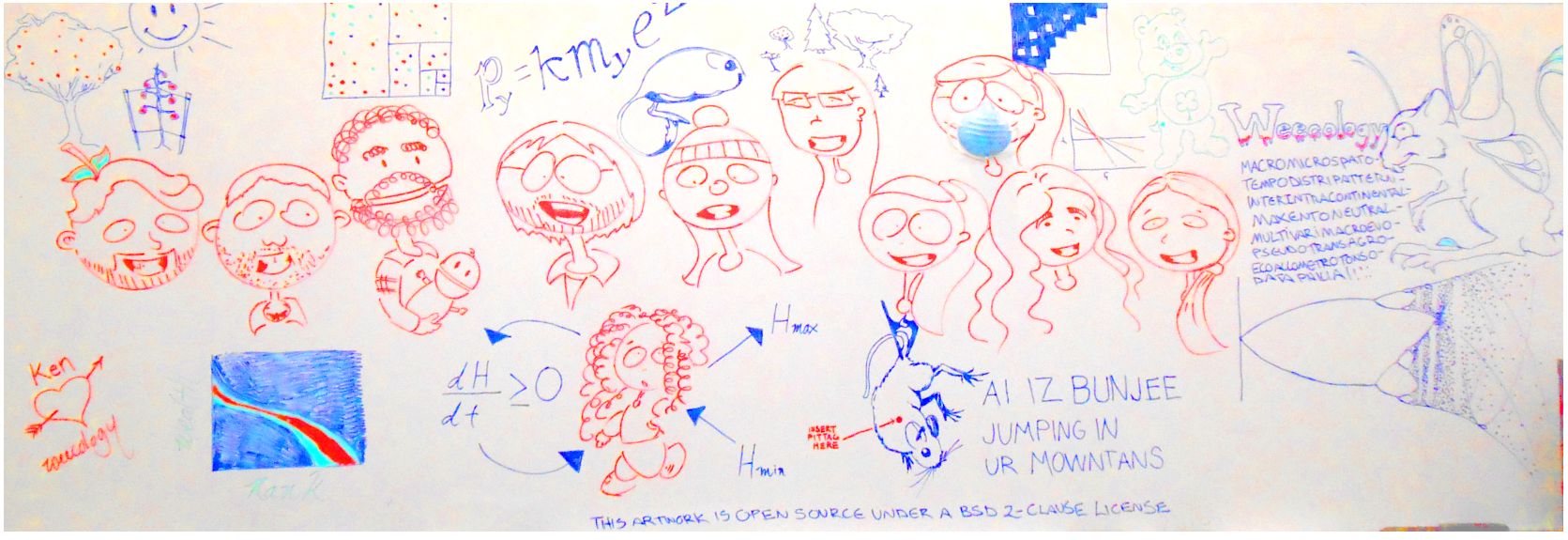
Acknowledgements

Funding sources:

- USU Department of Biology
- Intellectual Ventures private funding to Morgan Ernest
- National Science Foundation CAREER Grant to Ethan White
- Gordon & Betty Moore Foundation's Data-Driven Discovery Initiative Grant to Ethan White.
- USU Graduate School Dissertation Fellowship

Acknowledgements

Weecologists past, present, & future



(especially Xiao Xiao & Ken Locey (creator of the whiteboard))

Acknowledgements

Dr. Thomas Price & USU Student Health Center.

The Flint Hills of Kansas.

Tea, electric blankets, & heating pads.

A very supportive husband & family.

Accessibility

This dissertation brought to you by:

Disability accommodations

- Remote access & participation.
- Computational tools & tricks.
  - Version control (GitHub).
  - Publicly available data.
  - Programming skills (data manipulation & analysis).