

A DATA-INTENSIVE ASSESSMENT OF THE SPECIES-ABUNDANCE DISTRIBUTION

by

Elita Baldrige

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Ecology

Approved:

---

Ethan P. White  
Major Professor

---

S. K. Morgan Ernest  
Committee Member

---

David N. Koons  
Committee Member

---

Karen H. Beard  
Committee Member

---

Peter B. Adler  
Committee Member

---

Mark R. McLellan  
Vice President for Research and  
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2015

UMI Number: 3700756

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3700756

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

Copyright © Elita Baldrige 2015

All Rights Reserved

## ABSTRACT

## A Data-intensive Assessment of the Species Abundance Distribution

by

Elita Baldrige, Doctor of Philosophy

Utah State University, 2013

Major Professor: Ethan White  
Department: Biology

The hollow curve species abundance distribution describes the pattern of large numbers of rare species and a small number of common species in a community. The species abundance distribution is one of the most ubiquitous patterns in nature and many models have been proposed to explain the mechanisms that generate this pattern. While there have been numerous comparisons of species abundance distribution models, most of these comparisons only use a small subset of available models, focus on a single ecosystem or taxonomic group, and fail to use the most appropriate statistical methods. This makes it difficult to draw general conclusions about which, if any, models provide the best empirical fit to species abundance distributions. I compiled data from the literature to significantly expand the available data for underrepresented taxonomic groups, and combined this with other macroecological datasets to perform comprehensive model comparisons for the species abundance distribution. A multiple model comparison showed that most available models for the species abundance distribution fit the data equivalently well across a diverse array of ecosystems and taxonomic groups. In addition, a targeted comparison of the species abundance distribution predicted by a major ecological theory, the unified neutral theory of biodiversity (neutral theory), against a non-neutral model of species

abundance, demonstrates that it is difficult to distinguish between these two classes of theory based on patterns in the species abundance distribution. In concert, these studies call into question the potential for using the species abundance distribution to infer the processes operating in ecological systems.

(104 pages)

## PUBLIC ABSTRACT

A Data-intensive Assessment of the Species Abundance Distribution

By

Elita Baldrige, Doctor of Philosophy

Utah State University, 2015

Major Professor: Ethan White

Department: Biology

One of the most commonly observed patterns in ecology is the fact that at most locations there are a large number of relatively rare species, composed of only a few individuals per species, and a small number of relatively common species. This pattern of commonness and rarity is quantified by the species abundance distribution. As one of the most commonly observed patterns in ecology, it has been studied intensively for over 100 years.

A major emphasis of this research has been developing models to try to understand the forces that generate such a general pattern. As a result, there are now dozens of different models for how this pattern might be generated, and these models are based on at least five different major categories of forces. Because many of these models appear to match ecological observations reasonably well, it is difficult to tell which model or models are most likely to be correct.

To address this issue, I compiled data on over 16,000 ecological systems. I used the best available statistical methods to compare a number of different models to observed data. While

there have been numerous comparisons of species abundance distribution models, most of these comparisons only use only a small subset of available models, focus on a single ecosystem or type of species, and fail to use the most appropriate statistical methods. My approach overcomes all of these challenges and thus provides the best opportunity to figure out which models provide the best description of real data.

Both general and detailed comparisons of this pattern of commonness and rarity suggest that even when using large amounts of data from across the world and the diversity of life, and the most current and powerful statistical methods, that it is generally not possible to distinguish among many common models of the species abundance distribution. I could exclude some models as clearly poorer descriptions of the pattern than others, but several models provided equivalently good descriptions of the data. This calls into question the potential for using the species abundance distribution to understand what processes are driving ecological systems.



Illustration by the author fondly representing the assorted datasets used in this dissertation.



## ACKNOWLEDGMENTS

I thank my committee members for their flexibility and support. This work and this student were supported by the USU Department of Biology, private funding to Morgan Ernest by Intellectual Ventures, a National Science Foundation CAREER grant and a Gordon and Betty Moore Foundation Data Driven Discovery Grant to Ethan White, as well as a USU Graduate School Dissertation Fellowship. Thank you to my family and friends for their support in helping me to finish my dissertation. I also thank Dr. Thomas Price and the excellent staff at the USU student health center, for their professionalism and quality of care that led to my ultimate diagnosis and management of the chronic illness I developed as a PhD student. And, of course, thanks to all past, present, and future Weecologists who have helped and supported me through this dissertation process, especially my advisor, Ethan White, who has worked with me extensively to provide accommodations and helped to raise awareness for chronic illness/disability. I couldn't have done it without you folks, and this is, in my case, not a mere expression of gratitude, but a statement of absolute fact. And finally, to my master's advisor, Rob Channell, thanks for introducing me to the poem "Ulysses," by Alfred, Lord Tennyson, whose final lines have kept me moving forward in spite of everything:

We are not now that strength which in old days  
Moved earth and heaven, that which we are, we are;  
One equal temper of heroic hearts,  
Made weak by time and fate, but strong in will  
To strive, to seek, to find, and not to yield. (66-70)

## CONTENTS

	Page
ABSTRACT .....	iii
PUBLIC ABSTRACT .....	v
ACKNOWLEDGMENTS .....	viii
LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
CHAPTER	
INTRODUCTION .....	1
COMMUNITY ABUNDANCE DATA OF UNDERREPRESENTED TAXONOMIC GROUPS.....	5
AN EXTENSIVE COMPARISON OF SPECIES ABUNDANCE DISTRIBUTION MODELS .....	26
EVALUATING ABUNDANCE DISTRIBUTION BASED SIGNALS OF NEUTRALITY IN TERRESTRIAL SYSTEMS.....	43
CONCLUSION.....	60
APPENDICES .....	63
A.METADATA FOR ABUNDANCE DATABASE.....	64
B.CHAPTER 3: SUPPLEMENTARY RESULTS FOR CHAPTER 3.....	73
C.REPLICATION OF ANALYSES .....	85
CURRICULUM VITAE .....	89

## LIST OF TABLES

Table	Page
2 – 1	Dates, sources, and search terms used to identify possible data sources ..... 21
2 – 2	List of variables collected for each dataset ..... 22
3 – 1	Details of datasets used to evaluate the form of the species-abundance distribution. Datasets marked as Private were obtained through data requests to the providers resulting in Memorandums of Understanding governing data use ..... 38
4 – 1	Description of total number of sites per taxa and dataset. Taxonomic groups are ordered by the total number of sites in the compiled dataset ..... 55

## LIST OF FIGURES

Figure	Page
2 - 1	Number of sites per taxon..... 23
2 - 2	Number of individuals per taxon..... 24
2 - 3	Number of sites per biogeographic region..... 25
3 - 1	Total number of wins by model for each dataset individually..... 39
3 - 2	Frequency distributions of AICc weights by model for all datasets combined. .... 40
3 - 3	Frequency distributions of log-likelihoods by model for all datasets combined. .... 41
3 - 4	One-to-one plots of the likelihoods of each model against the likelihood of the log-series. The likelihoods of the different models have almost perfect correspondence for individual distributions. .... 42
4 - 1	Map of the georeferenced portion of the data. Note that the data for reptiles, amphibians, bony fish, beetles, spiders, and butterflies are not represented. Redrawn from White et al. 2012 ..... 56
4 - 2	Preston plot of empirical data for single sites from each dataset with lines representing the Poisson lognormal and the negative binomial. .... 57
4 - 3	Log of distinct abundance values versus AICc weight of the lognormal distribution for each dataset ..... 58
4 - 4	Average AICc weight of the lognormal distribution for each dataset. .... 59
B - 1	Frequency distribution of log-likelihoods for Amphibia ..... 74
B - 2	Frequency distribution of log-likelihoods for BBS..... 75
B - 3	Frequency distribution of log-likelihoods for Coleoptera..... 76
B - 4	Frequency distribution of log-likelihoods for CBC ..... 77
B - 5	Frequency distribution of log-likelihoods for FIA..... 78
B - 6	Frequency distribution of log-likelihoods for Gentry ..... 79
B - 7	Frequency distribution of log-likelihoods for MCDB ..... 80
B - 8	Frequency distribution of log-likelihoods for NABA..... 81
B - 9	Frequency distribution of log-likelihoods for Reptilia ..... 82

B – 10	Frequency distribution of log-likelihoods for Arachnida .....	83
B – 11	Frequency distribution of log-likelihoods for Actinopterygii.....	84

## CHAPTER 1

### INTRODUCTION

Macroecology is a data-intensive approach to studying ecological patterns and processes. As the field has matured, increasingly greater amounts of data have become available to address these questions (Reichman et al. 2011). Although the macroecological research program has matured a great deal since its introduction (McGill and Nekola 2010, Keith et al. 2012), there are still a number of areas in which it can be improved (Beck et al. 2012). One of the major criticisms of macroecology is that it is biased towards analyzing data from terrestrial systems in North America (Beck et al. 2012), a legacy of the academic heritage of the scientists who developed macroecology, as well as the availability of large ecological datasets suitable for testing macroecological questions.

Another major criticism has been that the field focuses too much on pattern description, and not enough on the identification of pattern generating mechanisms (Beck et al. 2012). One of the most significant patterns in macroecology is the species abundance distribution, which describes the commonness and rarity of species in an ecological community. The form of this pattern is very general, with most communities composed of a small number of common species and a large number of rare species. Interest in this pattern has generated dozens of models attempting to characterize the form of the pattern and the processes underlying it. However, most attempts to determine which of these models provide the best fit to empirical data, and are therefore most likely to describe the processes generating the pattern, have been limited either by a combination of poor statistical methodology and/or restricted scope.

A set of best practices for testing patterns and models in macroecology has been developed as the discipline has matured (Burnham and Anderson 2002, McGill 2003, McGill et

al. 2006, White et al. 2008, 2012). Some of these best practices for testing macroecological theory include:

- Testing the generality of patterns with multiple taxonomic groups/ecosystems, both terrestrial and aquatic/marine (White et al. 2012).
- Simultaneous testing of multiple models and model predictions (McGill 2003, McGill et al. 2006).
- Use of likelihood based methods for comparing distributions (Edwards et al. 2007, White et al. 2008).

Following these best practices allows for a more rigorous assessment of the generality of patterns across taxonomic groups/ecosystems, may provide better insight into identifying pattern generating mechanisms, and improve the ability to make ecological predictions. Here I present research following best practices for comparing species abundance distributions by compiling a unique dataset for addressing macroecological questions for more ecosystems and taxonomic groups, evaluate multiple competing models, and use the best statistical methods available.

First, to address some of the concerns about the lack of data for underrepresented taxonomic groups and ecosystems, I compiled a set of data from the literature. My primary focus for data collection was to collect data for those vertebrate taxa that are not already well represented by publicly available data. I selected these taxa because their taxonomy is fairly well resolved when compared to the majority of invertebrate groups. While my primary focus was on fish, reptiles, and amphibians, I also compiled data on spiders and beetles. In addition, I also collected a small amount of bird data, which could potentially be used in comparison with the large, publicly available bird datasets to determine if the dataset affects the results.

Second, we compare five models from each of four different model families: purely statistical, branching process, population dynamics, and niche partitioning (McGill et al. 2007) with community abundance data for reptiles, amphibians, bony fish, beetles, spiders, birds, trees,

mammals, and butterflies to perform the largest test of species abundance distributions to date. We follow the current best practice recommendations for testing species abundance distribution models to determine which models provide the best fit to empirical data. Identifying which, if any, models best describe the data can help determine what pattern generating mechanisms are more likely to have direct influences on the shape of the species abundance distribution.

Third, we expand our exploration of the hollow curve species abundance distribution by performing a detailed analysis evaluating previous work on marine systems by Connolly et al. (2014) to determine if species abundance data is sufficient for identifying two general classes of ecological process, neutral vs. non-neutral processes (Connolly et al. 2014). Connolly et al. (2014) found that the majority of sites were best described by a non-neutral model of species abundance in marine ecosystems; however, this has not yet been tested for terrestrial systems. This work seeks to determine the generality of the non-neutral species abundance distribution method in terrestrial systems, and identify whether there are differences between terrestrial and marine systems relative to species abundance distributions.

## References

- Beck, J., et al. 2012. What's on the horizon for macroecology? *Ecography* 35:673–683.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, New York, USA.
- Connolly, S. R. et al. 2014. Commonness and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences*: 8524–8529.
- Edwards, A. M et al. 2007. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* 449:1044–1048.
- Keith, S. A. et al. 2012. What is macroecology? *Biology Letters* 8:904–906.
- McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. *Nature* 422:881–885.



McGill, B. J., and J. C. Nekola. 2010. Mechanisms in macroecology: a WOL or purloined letter? towards a pragmatic view of mechanism. *Oikos* 119:591–603.

McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10:995–1015.

McGill, B. J. et al. 2006. Empirical evaluation of neutral theory. *Ecology* 87:1411–1423.

Reichman, O. et al. 2011. Challenges and opportunities of open data in ecology. *Science* 331-331.

White, E. P. et al. 2008. On estimating the exponent of power-law frequency distributions. *Ecology* 89:905–912.

White, E. P. et al. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* 93:1772–1778.

## CHAPTER 2

## COMMUNITY ABUNDANCE DATA OF UNDERREPRESENTED TAXONOMIC GROUPS

## Abstract

The majority of publicly available datasets used for macroecological research have a North American terrestrial bias, and focus primarily on warm-blooded vertebrates and plants. This dataset helps to improve the availability of data suitable for macroecological questions for less frequently studied taxa. The data were compiled from the literature by focusing on less frequently studied groups, and includes seven classes of animals, amphibians, spiders, beetles, reptiles, birds, and ray finned and cartilaginous fish. The data contains data representing over 2000 species and more than 1.3 million individuals from over 700 sites including locations on all continents except Antarctica.

## Background & Summary

Increasingly large amounts of data are available for studying ecological systems (Reichman et al. 2011). One of the most common forms of ecological data is community abundance data, which is composed of counts of the number of individuals of each species occurring in a community or assemblage. These kinds of data can be used to address a broad array of questions and have become central to research in macroecology.

One major criticism of macroecology is that the majority of research has been driven by a few major datasets, primarily terrestrial North American and European birds, mammals, and plants (Beck et al. 2012). This is due, in part, to the fact that large publicly available datasets with many sites tend to focus on these taxonomic groups (e.g., USDA Forest Service 2010, Thibault et al. 2011, Pardieck et al. 2014). This makes it difficult to determine if observed patterns are general or whether they only apply to the few taxa for which large amounts of easy to analyze data is available. It also makes it difficult to perform meaningful cross-taxonomic comparisons, which can be valuable to understanding the processes driving ecological systems.

One suggestion for improving macroecology in this regard is to make better use of existing data (Beck et al. 2012). There is a great deal of community abundance data in the literature, but most include a single to a few communities, and the majority of the data requires data entry and processing to be useable in analyses. In particular, much of this data is only available in tables in the text of papers. To address this deficit in readily available data, I have compiled a dataset from the literature that combines data for multiple taxa and biogeographic regions into a single publicly available source.

This will allow researchers to make ecological comparisons for a wider range of taxa without having to gather and process the data from the literature before use. This data compilation contains abundance data for seven classes of animal, including vertebrates and invertebrates, endotherms and ectotherms, and was collected by intentionally focusing on the

collection of data for taxa that are not currently well represented in commonly used macroecological datasets. This emphasis on underrepresented taxa resulted in large amounts of data for fish, reptiles, and amphibians and reasonable amounts of data for spiders and insects (Figure 2-1, Figure 2-2). While the majority of the data is Nearctic, there is a worldwide distribution of sites (Figure 2-3), improving the representation of data outside of North America. This dataset will allow for a more robust comparison of patterns across taxa, especially when combined with existing macroecological datasets. While the primary focus of data collection was filling in the gaps for vertebrate taxa, I also collected community abundance data on other taxa incidentally. Additional metadata can be found in Appendix A.

## Methods

### Data Sources

Data were compiled from a combination of journal articles, theses, and dissertations. The taxonomic focus of the literature search was determined based on an initial search of the literature for community abundance data to get a sense of what data were available, and which underrepresented taxa were likely to yield reasonable amounts of data. After the initial search, I conducted a systematic review the literature, with fish, amphibians, and reptiles as the main focus of data collection. Data for other groups were collected on an ad hoc when they were encountered, which resulted in a reasonable amount of data for arachnids and insects (Figure 2-1).

### Data Collection

References found by the searches in Table 1 were downloaded. Each article, thesis, and dissertation was then manually scanned to determine if it met the criteria for inclusion in the database. The selection criteria included:

- Data must include quantitative abundances, preferably total number of individuals (no incidence only, i.e., presence-absence, data)
- Data must be for animal data
- Sampling and reporting must be complete (i.e., no data where only a fraction of the community/assemblage was sampled or reported)
- For vertebrate taxa: the majority of species must be fully identified to species
- For invertebrate taxa: the majority of species may did not have to be fully identified to species (due to the number of individuals per sample and the state of taxonomy for the invertebrate groups)
- Data must not be heavily summarized or processed

The following papers remained as data sources based on these criteria: Hoff and Ibara (1977), Bennett et al. (1980), Matthews and Hill (1980), Horn (1980), Jones (1981), Bultman and Uetz (1982), Grossman (1982), Dritschilo and Erwin (1982), Grossman et al. (1982), Allen (1982), Bell et al. (1984), Schlosser (1985), Moyle and Vondracek (1985), Ross et al. (1985), Matthews (1986), Stoner (1986), Bodkin (1988), Alvarez et al. (1989), Dobel et al. (1990), Yoklavich et al. (1991), Dalrymple et al. (1991), Ford et al. (1991), Taylor et al. (1993), Belize (1993), Kinsolving and Bain (1993), Reid and Whiting (1994), Petterson (1996), Busby and Parmelee (1996), Cobb and Summerhill (1996a), Cobb and Summerhill (1996b), McLendon et al. (1996), Brandt (1997), Laroche et al. (1997), Rodriguez and Lewis (1997), Schlesinger et al. (1997), Mitchell et al. (1997), Healey et al. (1997), Tejerina-Garro et al. (1998), Demynadier and Hunter (1998), How (1998), Russell et al. (1999), Conroy (1999), Cavitt (2000), Winemiller et al. (2000), Gido (2000), Silbano et al. (2000), Maxey and Richardson (2000), Kretzer and Cully (2001), Methven et al. (2001), Gelwick et al. (2001), Ferreira et al. (2001), Yahner et al. (2001), Vonesh (2001), Germaine and Wakeling (2001), Hofer and Bersier (2001), Morrison et al. (2002), Jaureguizar et al. (2002), Tongnunui et al. (2002), Ashton (2002), Read (2002), Menke

(2003), Feyrer and Healey (2003), Quinn and Kwak (2003), Petry et al. (2003), Wasonga (2003), Gainsbury and Colli (2003), Moseley et al. (2003), Chick et al. (2004), Adams et al. (2004), Vega-Cendejas and de Santillana (2004), Malavasi et al. (2004), Schifino et al. (2004), Penczak et al. (2004), How and Dell (2004), Beever and Brussard (2004), Shipman et al. (2004), Pombo et al. (2005), Bonner et al. (2005), Brannon and Rogers (2005), Leynaud and Bucher (2005), Smart et al. (2005), Thompson and Thompson (2005), Goldstein et al. (2005), Luiselli et al. (2005), Watling (2005), Castellano and Valone (2005), Wilgers and Horne (2006), Wilgers et al. (2006), Ribeiro et al. (2006), Kanowski et al. (2006), Habit et al. (2007), Fialho et al. (2007), Ford and Lancaster (2007), Inoue et al. (2008), Thomson (2008), Brodman (2008), Lindenmayer et al. (2008), Isaac et al. (2008), Thompson et al. (2008), Maltchik et al. (2008), Carvajal-Cogollo and Urbina-Cardona (2008), Ferguson et al. (2008), Fischer and Paukert (2009), Kennedy (2009), Siqueira et al. (2009), Hutchens and DePerno (2009), Michaelides and Kati (2009), Akani and Luiselli (2009), Mekonnen (2009), Mott et al. (2010), Cano and Leynaud (2010), Steen et al. (2010), Carvalho et al. (2011), Pianka and Goodyear (2011). Information on these data sources is also available as part of the dataset in the citations\_table\_abundances.csv file.

Data were hand entered into a raw data file as they came from the original source or were extracted from the original source computationally. Data were then manually checked for consistency with the original source. Species names were kept as given in the original source.

## Data Records

The data are stored in comma-separated values files using a relational database structure with three separate tables.

## Data files

1. Abundance data: Species\_abundances.csv
2. Sites data : Sites\_table\_abundances.csv
3. Reference data: Citations\_table\_abundances.csv

## Format and Storage mode

ASCII text, comma delimited, not compressed.

## Header information

1. Class, Family, Genus, Species, Relative\_abundance, Abundance, Site\_ID, Citation\_ID
2. Site\_ID, Collection\_Year, End\_Collection, Citation\_ID, Site\_Name, Biogeographic\_region, Site\_notes
3. Citation\_ID, Authors, Yr, Title, Journal, Issue, Pages

## Special characters/fields

Blanks indicate no data: no special characters used.

## Technical Validation

Data have undergone manual quality and assurance checking. Data were entered directly from the source material into the raw data file and values were double checked on entry. Validation of proper downloading and importing of the data can be determined using the following information.

#### Abundance table

1. Number of records, not including header row = 22142
2. Sum of Relative\_abundance = 10797.37352
3. Sum of Abundance = 1320592
4. Number of distinct values in species = 1953
5. Number of distinct values in genus = 1262
6. md5 checksum for file = 225508ec2acc8cadd230b5e80446504e

#### Sites table

1. Number of records, not including header row = 706
2. Number of distinct values in collection\_year = 48
3. Number of distinct values in biogeographic\_region = 6
4. Sum of collection\_year = 1378306
5. md5 checksum for file = 9935391079863726d24a9204ea68149d

#### References table

1. Number of records, not including header row = 116
2. Sum of yr = 231916
3. Number of distinct values in journal = 83
4. md5 checksum for file = e42838ee418a44e9e5d33ff99bf96ebb

#### Usage Notes

This is compiled data from a variety of literature sources. Within a study, methods of data collection are the same. However, among studies, even within the same taxonomic grouping, methods of collection, capture success, etc. vary substantially. Because of the methodological



variation present in compiled data, it is more appropriate to treat each site individually, rather than aggregating sites across studies for doing things like looking for geographic patterns.

Aggregating data across sites can lead to false signals in species richness, abundance, etc. that are due to methodological rather than biological/ecological differences. In addition, some sites also have recorded absences (zeros); in cases where zeros should not be included, data queries should be written accordingly.

The data can be easily downloaded and installed into a variety of database management and programming environments using the EcoData Retriever (Morris and White 2013).

## References

- Adams, S. B., et al. 2004. Spatial and temporal patterns in fish assemblages of upper coastal plain streams, Mississippi, USA. *Hydrobiologia* 528:45–61.
- Akani, G. C., and L. M. Luiselli. 2009. Aspects of community ecology of amphibians and reptiles at Bonny Island (Nigeria), an area of priority relevance for petrochemical industry. *African Journal of Ecology* 48:939–948.
- Allen, L. G. 1982. Seasonal abundance, composition, and productivity of the littoral fish assemblage in upper Newport Bay, California. *Fishery Bulletin* 80:769–790.
- Alvarez, S., et al. 1989. Structure and composition of two lizard communities of the cape region, Baja California Sur, Mexico. *Bulletin of the Maryland Herpetological Society* 25:40–48.
- Ashton, D.T. 2002. A comparison of abundance and assemblage of lotic amphibians in late-seral and second-growth redwood forests in Humboldt county, California. Humboldt State University.
- Beck, J., et al. 2012. What's on the horizon for macroecology? *Ecography* 35:673–683.
- Beever, E. A., and P. F. Brussard. 2004. Community and landscape-level responses of reptiles and small mammals to feral-horse grazing in the Great Basin. *Journal of Arid Environments* 59:271–297.
- Belize, C. A. 1993. The fish community of a shallow tropical lagoon in Belize, Central America. *Estuaries* 16:198–215.
- Bell, J. D., et al. 1984. Structure of a fish community in a temperate tidal mangrove creek in Botany Bay, New South Wales. *Australian Journal of Marine Freshwater Restoration* 35:33–46.
- Bennett, S. H., et al. 1980. Terrestrial activity, abundance, and diversity of

- amphibians in differently managed forest types. *American Midland Naturalist* 103:412–416.
- Bodkin, J. 1988. Effects of kelp forest removal on associated fish assemblages in central California. *Journal of Experimental Marine Biology and Ecology* 117:227–238.
- Bonner, T. H. et al. 2005. Temporal assessment of a west Texas stream fish assemblage. *The Southwestern Naturalist* 50:74–78.
- Brandt, A. 1997. Abundance, diversity and community patterns of epibenthic- and benthic-boundary layer peracarid crustaceans at 75N off east Greenland. *Polar Biology* 17:159–174.
- Brannon, M. P., and S. Rogers. 2005. Effects of canopy thinning by hemlock woolly adelgids on the local abundance of terrestrial salamanders. *Journal of the North Carolina Academy of Sciences* 121:151–156.
- Brodman, R. 2008. A 14-year study of amphibian populations and metacommunities. *Herpetological Conservation and Biology* 4:106–119.
- Bultman, T. L., and G. W. Uetz. 1982. Abundance and community structure of forest floor spiders following litter manipulation. *Oecologia* 55:34–41.
- Busby, W. H., and J. R. Parmelee. 1996. Historical changes in a herpetofaunal assemblage in the flint hills of Kansas. *American Midland Naturalist* 135:81–91.
- Cano, P. D., and G. C. Leynaud. 2010. Effects of fire and cattle grazing on amphibians and lizards in northeastern Argentina (humid Chaco). *European Journal of Wildlife Research* 56:411–420.
- Carvajal-Cogollo, J. E., and J. N. Urbina-Cardona. 2008. Patrones de diversidad y composición de reptiles en fragmentos de bosque seco tropical en Córdoba, Colombia. *Tropical Conservation Science* 1:397–416.
- Carvalho, J. C., et al. 2011. Biogeographic patterns of spiders in coastal dunes along a gradient of mediterraneity creosote bush ecotone in the Chihuahuan desert. *Biodiversity Conservation*.
- Castellano, M. J., and T. J. Valone. 2005. Effects of livestock removal and perennial grass recovery on the lizards of a desertified arid grassland. *Journal of Arid Environments* 66:87–95.
- Cavitt, J. F. 2000. Fire and a tallgrass prairie reptile community: effects on relative abundance and seasonal activity. *Journal of Herpetology* 34:12–20.
- Chick, J. H., et al. 2004. Spatial scale and abundance patterns of large fish communities in freshwater marshes of the Florida everglades. *Wetlands* 24:652–664.
- Cobb, V. A., and J. A. Summerhill. 1996. A one-year study of the species diversity and relative abundance of snakes and lizards in the Jack Mountains region of Hot Spring County, Arkansas. *Proceedings Arkansas Academy of Science* 50:120–127.
- Conroy, S. 1999. Lizard assemblage response to a forest ecotone in northeastern Australia: a synecological approach. *Journal of Herpetology* 33:409–419.

- Dalrymple, G. H., et al. 1991. Patterns of species diversity of snake community assemblages, with data on two everglades snake assemblages. *Copeia* 2:517–521.
- Demynadier, P. G., and M. L. Hunter Jr. 1998. Effects of silvicultural edges on the distribution and abundance of amphibians in maine. *Conservation Biology* 12:340–352.
- Dobel, H.G. et al. 1990. Spider (Araneae) community structure in an intertidal salt marsh: effects of vegetation structure and tidal flooding. *Environmental Entomology* 19:1356–1370.
- Dritschilo, W., and T.L. Erwin. 1982. Responses in abundance and diversity of cornfield carabid communities to differences in farm practices. *Ecology* 63:900–904.
- Ferguson, A. W., et al. 2008. Herpetofaunal inventory of Camp Mabry, Austin, Texas: community composition in an urban landscape. *Texas Journal of Science* 60:123–136.
- Ferreira, C. E. L., et al. 2001. Community structure of fishes and habitat complexity on a tropical rocky shore. *Environmental Biology of Fishes* 61:353–369.
- Feyrer, F., and M. P. Healey. 2003. Fish community structure and environmental correlates in the highly altered southern Sacramento-San Joaquin delta. *Environmental Biology of Fishes* 66:123–132.
- Fialho, A. P., et al. 2007. Fish assemblage structure in tributaries of the Meia Ponte river, Goias, Brazil. *Neotropical Ichthyology* 5:53–60.
- Fischer, J. R., and C. P. Paukert. 2009. Effects of sampling effort, assemblage similarity, and habitat heterogeneity on estimates of species richness and relative abundance of stream fishes. *Canadian Journal of Fisheries Aquatic Science* 66:277–209.
- Ford, N.B., et al. 1991. Species diversity and seasonal abundance of snakes in a mixed pine-hardwood forest of eastern Texas. *The Southwestern Naturalist* 36:171–177.
- Ford, N.B., and D. Lancaster. 2007. The species-abundance distribution of snakes in a bottomland hardwood forest of the southern United States. *Journal of Herpetology* 41:385–393.
- Gainsbury, A. M., and G. R. Colli. 2003. Lizard assemblages from natural cerrado enclaves in southwestern Amazonia: the role of stochastic extinctions and isolation. *Biotropica* 35:503–519.
- Gelwick, F. P., et al. 2001. Fish assemblage structure in relation to environmental variation in a Texas gulf coastal wetland. *Estuaries* 24:285–296.
- Germaine, S. S., and B. F. Wakeling. 2001. Lizard species distributions and habitat occupation along an urban gradient in Tucson, Arizona, USA. *Biological Conservation* 97:229–237.
- Gido, K. B., et al. 2000. Long-term changes in a reservoir fish assemblage: stability in an unpredictable environment. *Ecological Applications* 10:1517–1529.

- Goldstein, M. I., et al. 2005. Spatiotemporal responses of reptiles and amphibians to timber harvest treatments. *Journal of Wildlife Management* 69:525–539.
- Grossman, G. D. 1982. Dynamics and organization of a rocky intertidal fish assemblage: the persistence and resilience of taxocene structure. *The American Naturalist* 119:611–637.
- Grossman, G. D., et al. 1982. Stochasticity in structural and functional characteristics of an Indiana stream fish assemblage: a test of community theory. *The American Naturalist* 120:423–454.
- Habit, E., et al. 2007. Spatio-temporal distribution patterns and conservation of fish assemblages in a Chilean coastal river. *Biodiversity Conservation* 16:3179–3191.
- Healey, M., et al. 1997. Amphibian communities associated with billabong habitats on the Murrumbidgee floodplain, Australia. *Australian Journal of Ecology* 22:270–278.
- Hofer, U., and L. Bersier. 2001. Herpetofaunal diversity and abundance in tropical upland forest of Cameroon and Panama. *Biotropica* 33:142–152.
- Hoff, J.G, and R. M. Ibara. 1977. Factors affecting the seasonal abundance, composition, and diversity of fishes in a southeastern New England estuary. *Estuarine and Coastal Marine Science* 5:665–678.
- Horn, M. H. 1980. Diel and seasonal variation in abundance and diversity of shallow-water fish populations in Morro Bay, California. *Fishery Bulletin* 78:759–770.
- How, R. A. 1998. Long-term sampling of a herpetofaunal assemblage on an isolated urban bushland remnant, Bold Park, Perth. *Journal of the Royal Society of Western Australia* 81:143–148.
- How, R.A., and J. Dell. 2004. Reptile assemblage of the Abydos plain, north-eastern Pilbara, western Australia. *Journal of the Royal Society of Western Australia* 87:85–95.
- Hutchens, S., and C. DePerno. 2009. Measuring species diversity to determine land-use effects on reptile and amphibian assemblages. *Amphibia-Reptilia* 30:81–88.
- Inoue, T., et al. 2008. Surf zone fishes in an exposed sandy beach at Sanrimatsubara, Japan: does fish assemblage structure differ among microhabitats? *Estuarine, Coastal, and Shelf Science* 77:1–11.
- Isaac, J. L., et al. 2008. Resistance and resilience: quantifying relative extinction risk in a diverse assemblage of Australian tropical rainforest vertebrates. *Diversity and Distributions* 2:280–288.
- Jaureguizar, A. J., et al. 2002. Fish assemblage and environmental patterns in the Rio de la Plata estuary. *Estuarine Coastal and Shelf Science* 56:921–933.
- Jones, K. B. 1981. Effects of grazing on lizard abundance and diversity in western Arizona. *The Southwestern Naturalist* 26:107–115.

- Kanowski, J. J., et al. 2006. Factors affecting the use of reforested sites by reptiles in cleared rainforest landscapes in tropical and subtropical Australia. *Restoration Ecology* 14:67–76.
- Kennedy, T. B. 2009. Aquatic community organization in a diverse floodplain river fish fauna of the southeastern United States. University of Alabama. Thesis.
- Kinsolving, A.D., and M. B. Bain. 1993. Fish assemblage recovery along a riverine disturbance gradient. *Ecological Applications* 3:531–544.
- Kretzer, J. E., and J.F. Cully Jr. 2001. Effects of black-tailed prairie dogs on reptiles and amphibians in Kansas shortgrass prairie. *The Southwest Naturalist* 46:171–177.
- Laroche, J., et al. 1997. Temporal patterns in a fish assemblage of a semiarid mangrove zone in Madagascar. *Journal of Fish Biology* 51:4–20.
- Leynaud, G. C., and E. H. Bucher. 2005. Restoration of degraded chaco woodlands: effects on reptile assemblages. *Forest Ecology and Management* 213:384–390.
- Lindenmayer, D. B., et al. 2008. How predictable are reptile response to wildfire? *Oikos* 117:1086–1097.
- Luiselli, L., et al. 2005. Relationships between body size, population abundance and niche characteristics in the communities of snakes from three habitats in southern Nigeria. *Journal of the Zoological Society of London* 265:207–213.
- Malavasi, S., et al. 2004. Fish assemblages of Venice lagoon shallow water: an analysis based on species, families, and functional guilds. *Journal of Marine Systems* 51:19–31.
- Maltchik, L., et al. 2008. Dynamics of the terrestrial amphibian assemblage in a flooded riparian forest fragment in a neotropical region in the south of Brazil. *Brazilian Journal of Biology* 68:763–769.
- Matthews, W. J. 1986. Fish faunal structure in an Ozark stream: stability, persistence, and a catastrophic flood. *Copeia* 1986:388–397.
- Matthews, W. J., and L. G. Hill. 1980. Habitat partitioning in the fish community of a southwestern river. *The Southwestern Naturalist* 25:51–66.
- Maxey, K. A., and J. Richardson. 2000. Abundance and movements of terrestrial salamanders in second-growth forests of southwestern British Columbia. *Proceedings of a conference on the biology and management of species and habitats at risk, Kamloops, B.C., 15-19 Feb., 1999.* 1:295–302.
- McLendon, J. P., et al. 1996. Reptile and amphibian characterization of a thermally disturbed braided stream undergoing restoration near Aiken, SC. *Proceedings of the 23rd Annual Conference on Ecosystems Restoration and Creation*:228–241.
- Mekonnen, A. 2009. Reptile survey and some ecological studies of two sympatric lizard species,

- Tarentola annularis* and *Mabuya striata* from Awash National Park, Ethiopia. Addis Abba University. Dissertation
- Menke, S. B. 2003. Lizard community structure across a grassland- creosote bush ecotone in the Chihuahuan desert. *Canadian Journal of Zoology* 81:1829–1838.
- Methven, D. A., et al. 2001. The fish assemblage of a Newfoundland estuary: Diel, monthly and annual variation. *Estuarine, Coastal and Shelf Science* 52:669–687.
- Michaelides, G., and V. Kati. 2009. Diversity patterns and conservation management of the lizard community in a Mediterranean reserve (Cyprus). *Journal of Biological Research- Thessaloniki* 12:211–220.
- Mitchell, J. C., et al. 1997. Factors influencing amphibian and small mammal assemblages in central Appalachian forests. *Forest Ecology and Management* 96:65–76.
- Morris, B. D., and E. P. White. 2013. The EcoData retriever: improving access to existing ecological data. *PLOS One* 8:e65848.
- Morrison, M. A., et al. 2002. Diurnal and tidal variation in the abundance of the fish fauna of a temperate tidal mudflat. *Estuarine, Coastal and Shelf Science* 54:793–807.
- Moseley, K. R., et al. 2003. Effects of prescribed fire on herpetofauna in bottomland hardwood forests. *Southeastern Naturalist* 2:475–486.
- Mott, B., et al. 2010. Tropical reptiles in pine forests: assemblage response to plantations and plantation management by burning. *Forest Ecology and Management* 259:916–925.
- Moyle, P. B., and B. Vondracek. 1985. Persistence and structure of the fish assemblage in a small California stream. *Ecology* 66:1–13.
- Pardieck, K. L. et al. 2014. North American breeding bird survey dataset 1966 - 2013, version 2013.0. U.S. Geological Survey, Patuxent Wildlife Research Center.
- Penczak, T., et al. 2004. Fish assemblage changes relative to environmental factors and time in the Warta river, Poland, and its oxbow lakes. *Journal of Fish Biology* 64:483–501.
- Petry, A. C., et al. 2003. Spatial variation of the fish assemblage structure from the upper Rio Parana floodplain, Brazil, in a dry year. *Acta Limnologica Brasiliensia* 15:1–13.
- Petterson, R. B. 1996. Effects of forestry on the abundance and diversity of arboreal spiders in the boreal spruce forest. *Ecography* 19:221–228.
- Pianka, E. R., and S. E. Goodyear. 2011. Lizard response to wildfire in arid interior Australia: long-term experimental data and commonalities with other studies. *Austral Ecology*.
- Pombo, L., et al. 2005. Environmental influences on fish assemblage distribution of an estuarine coastal lagoon, Rio de Aveiro (Portugal). *Scientia Marina* 69:143–159.

- Quinn, J. W., and T. J. Kwak. 2003. Fish assemblage changes in an Ozark river after impoundment: a long-term perspective. *Transactions of the American Fisheries Society* 132:110–119.
- Read, J. 2002. Experimental trial of Australian arid zone reptiles as early warning indicators of overgrazing by cattle. *Austral Biology* 27:55–66.
- Reichman, O., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. *Science* 331:331.
- Reid, J. A., and R. M. Whiting Jr. 1994. Herpetofauna of pitcher plant bogs and adjacent forests in eastern Texas. *Proceedings of the Annual Conference of Southeastern Association of Fish and Wildlife Agencies* 48:411–421.
- Ribeiro, J., et al. 2006. Seasonal, tidal, and diurnal changes in fish assemblages in the Ria Formosa lagoon (Portugal). *Estuarine, Coastal, and Shelf Science* 67:461–474.
- Rodriguez, M.A., and W. M. Lewis. 1997. Structure of fish assemblages along environmental gradients in floodplain lakes of Orinoco river. *Ecological Monographs* 67:109–128.
- Ross, S. T., et al. 1985. Persistence of stream fish assemblages: effects of environmental change. *The American Naturalist* 126:24–40.
- Russell, K. R., et al. 1999. Amphibian and reptile communities associated with beaver (*Castor canadensis*) ponds and unimpounded streams in the piedmont of South Carolina. *Journal of Freshwater Ecology* 14:149–158.
- Schifino, L. C., et al. 2004. Fish community composition, seasonality, and abundance in Fortaleza lagoon, Cidreira. *Brazilian Archives of Biology and Technology* 47:755–763.
- Schlesinger, C. A., et al. 1997. Fire studies in mallee (eucalyptus spp.) communities of western New South Wales: reptile and beetle populations in sites of differing fire history. *The Rangeland Journal* 19:190–205.
- Schlosser, I. J. 1985. Flow regime, juvenile abundance, and the assemblage structure of stream fishes. *Ecology* 66:1484–1490.
- Shipman, P. A., et al. 2004. Reptile communities under diverse forest management in the Ouachita mountains, Arkansas. *Ouachita and Ozark Mountains symposium: ecosystem management research*. U.S. Department of Agriculture, Forest Service, Southern Research Station. U.S. Department of Agriculture, Forest Service, Southern Research Station. Gen. Tech. Rep. SRS-74:174–182.
- Silbano, R.A.M., et al. 2000. Spatial and temporal patterns of diversity and distribution of the upper Jurua river fish community (Brazilian Amazon). *Environmental Biology of Fishes* 57:25–35.
- Siqueira, C.C., et al. 2009. Density and richness of leaf litter frogs (Amphibia: Anura) of an Atlantic rainforest area in the Serra dos Orgaos, Rio de Janeiro state, Brazil. *Zoologia* 26:97–102.

- Smart, R. et al. 2005. Lizards and landscapes: integrating field surveys and interviews to assess the impact of human disturbance on lizard assemblages and selected reptiles in a savanna in South Africa. *Biological Conservation* 122:23–31.
- Steen, D. A., et al. 2010. Effects of forest management on amphibians and reptiles: generalist species obscure trends among native forest associates. *Open Environmental Sciences* 4:24–30.
- Stoner, A. W. 1986. Community structure of the demersal fish species of Laguna Joyuda, Puerto Rico. *Estuaries* 9:142–152.
- Taylor, C. M., et al. 1993. Fish species-environment and abundance relationships in a Great Plains river system. *Ecography* 16:16–23.
- Tejerina-Garro, F. L., et al. 1998. Fish community structure in relation to environmental variation in floodplain lakes of the Araguaia river, Amazon basin. *Environmental Biology of Fishes* 51:399–410.
- Thibault, K. M. et al. 2011. Species composition and abundance of mammalian communities: ecological archives e092-201. *Ecology* 92:2316–2316.
- Thompson, S. A., and G. G. Thompson. 2005. Temporal variations in reptile assemblages in the goldfields of western Australia. *Journal of the Royal Society of Western Australia* 88:25–36.
- Thompson, S. A., et al. 2008. Rehabilitation index for evaluating restoration of terrestrial ecosystems using the reptile assemblage as the bio-indicator. *Ecological Indicators* 8:530–549.
- Thomson, S. K. 2008. The influence of livestock watering ponds (dugouts) on native stream fishes, especially the endangered Topeka Shiner (*Notropis topeka*). South Dakota State University. Thesis.
- Tongnunui, P. et al. 2002. Fish fauna of the Sikao Creek mangrove estuary, Trang, Thailand. *Fisheries Science* 68:10–17.
- USDA Forest Service. 2010. Forest inventory and analysis National Core field guide (phase 2 and 3). version 4.0. USDA Forest Service, Forest Inventory; Analysis.
- Vega-Cendejas, M. E., and M. Hernandez de Santillana. 2004. Fish community structure and dynamics in a coastal hypersaline lagoon: Rio Lagartos, Yucatán, Mexico. *Estuarine, Coastal, and Shelf Science* 60:285–299.
- Vonesh, J. R. 2001. Patterns of richness and abundance in a tropical African leaf-litter herpetofauna. *Biotropica* 33:502–510.
- Wasonga, D. V. 2003. The diversity and abundance of amphibians in Meru National Park, Kenya. Addis Ababa University. Dissertation.



Watling, J. I. 2005. Edaphically-biased distributions of amphibians and reptiles in a lowland tropical rainforest. *Studies on Neotropical Fauna and Environment* 40:15–21.

Wilgers, D. J., and E. A. Horne. 2006. Effects of different burn regimes on tallgrass prairie herpetofaunal species diversity and community composition in the Flint Hills, Kansas. *Journal of Herpetology* 40:73–84.

Wilgers, D. J., et al. 2006. Effects of rangeland management on community dynamics of the herpetofauna of the tallgrass prairie. *Herpetologica* 62:378–388.

Winemiller, K. O., et al. 2000. Fish assemblage structure in relation to environmental variation among brazos river oxbow lakes. *Transactions of the American Fisheries Society* 129:451–468.

Yahner, R. H., et al. 2001. Response of amphibian and reptile populations to vegetation maintenance of an electric transmission line right-of-way. *Journal of Arboriculture* 27:215–221.

Yoklavich, M. M. et al. 1991. Temporal and spatial pattern in abundance and diversity of fish assemblages in Elkhorn Slough, California. *Estuaries* 14:465–480.

Table 2-1. Dates, source, and search terms used to identify possible data sources.

Search Parameters	Search engine	Date Accessed
community* abundance in Biology, Life Sciences, etc.	Google Scholar	29 Nov 2010
fish assemblage abundance, fish community* abundance in Biology, Life Sciences, etc.	Google Scholar	14 Feb 2011
fish community* abundance, fish assemblage abundance	ProQuest UMI Dissertations & Theses	15 Feb 2011
reptile assemblage abundance, reptile community* abundance in Biology, Life Sciences, etc.	Google Scholar	20 Aug 2011
reptile community* abundance, reptile assemblage abundance	ProQuest UMI Dissertations & Theses	21 Aug 2011
amphibian assemblage abundance, amphibian community* abundance in Biology, Life Sciences, etc.	Google Scholar	7 Oct 2011
amphibian community* abundance, amphibian assemblage abundance	ProQuest UMI Dissertations & Theses	7 Oct 2011

Table 2-2. List of variables collected for each dataset

Variable name	Variable definition	Units	Storage type	Value Range
Class	Taxonomic class of species	N/A	Character	N/A
Family	Taxonomic family of species	N/A	Character	N/A
Genus	Taxonomic genus of species	N/A	Character	N/A
Species	Specific epithet of species	N/A	Character	N/A
Relative_abundance	Relative abundance of species	N/A	Double	0 - 309
Abundance	Abundance of species	N/A	Integer	0-181726
Collection_Year	Start of collecting	N/A	Integer	1952-2008
End_Collection	End of collecting	N/A	Integer	1977-2009
Site_Name	Name/description of site	N/A	Character	N/A
Biogeographic_region	Biogeographic region	N/A	Character	N/A
Site_notes	Additional site information	N/A	Character	N/A

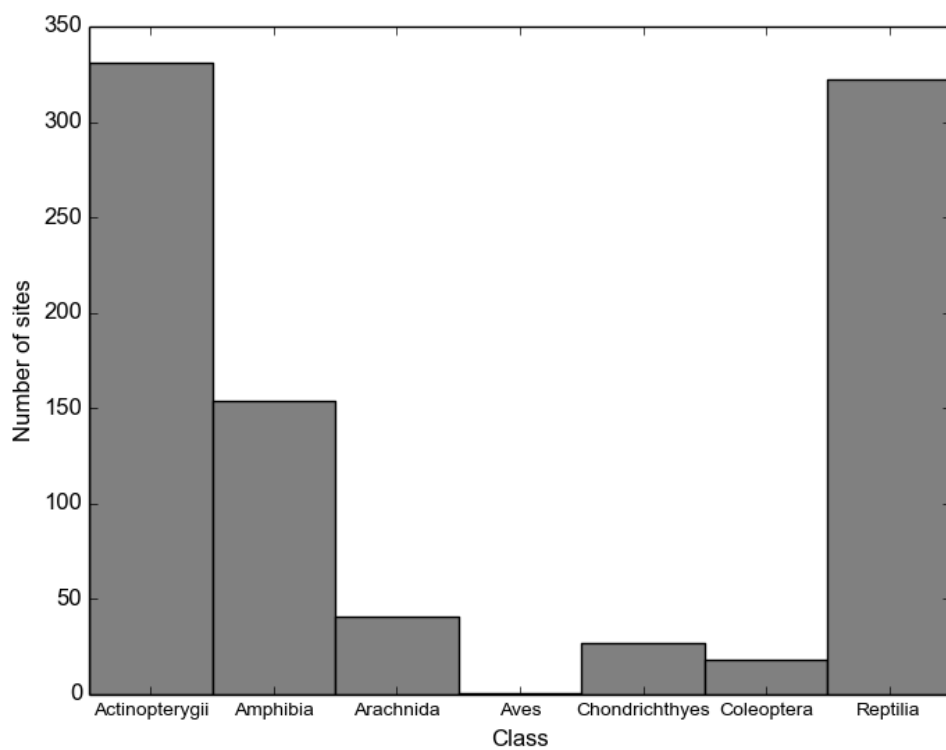


Figure 2 - 1 Number of sites per taxon

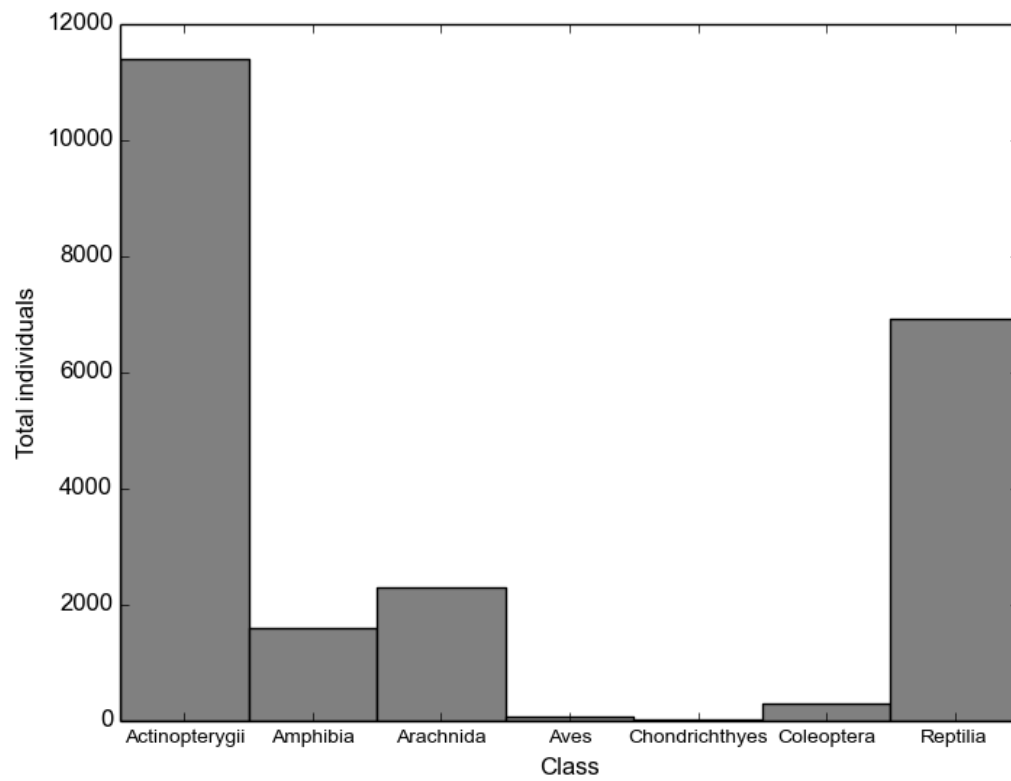


Figure 2 – 2 Number of individuals per taxon.

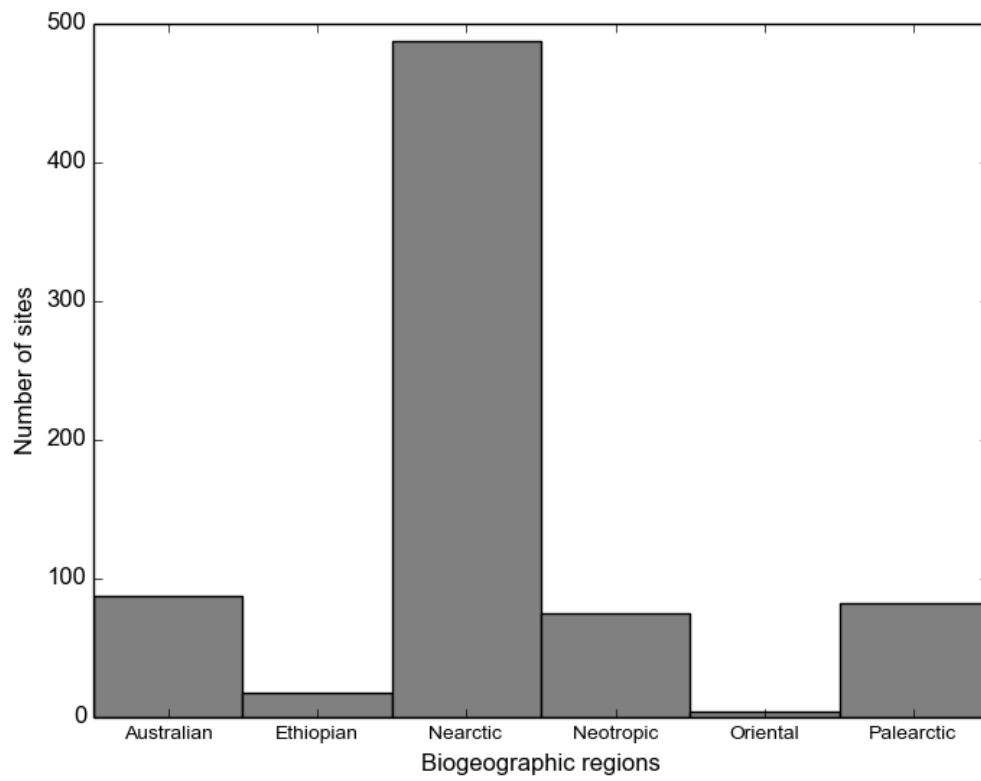


Figure 2 – 3 Number of sites per biogeographic region.

## CHAPTER 3

## AN EXTENSIVE COMPARISON OF SPECIES ABUNDANCE DISTRIBUTION MODELS

## Introduction

The species abundance distribution (SAD) describes the full distribution of commonness and rarity in ecological systems. It is one of the most fundamental and ubiquitous patterns in ecology, and exhibits a consistent general form with many rare species and few abundant species occurring within a community. This general shape is often referred to as a hollow curve distribution.

The SAD is one of the most widely studied patterns in ecology, leading to a proliferation of models that attempt to characterize the shape of the distribution and identify potential mechanisms for the pattern (see McGill et al. 2007 for a recent review of SADs). These models range from arbitrary distributions that are chosen based on providing a good fit to the data (Fisher et al. 1943), to distributions chosen based on combinatorics and the most likely state of the system (Frank 2011, Harte 2011, Locey and White 2013), to models based on ecological process (Tokeshi 1993, Hubbell 2001, Volkov et al. 2003).

Which model or models provide the best fit to the data, and the resulting implications for the processes structuring ecological systems, has been an active area of research (e.g., McGill 2003, Volkov et al. 2003, Ulrich et al. 2010, White et al. 2012, Connolly et al. 2014). However, most comparisons of the different models: 1) use only a small subset of available models (typically two; e.g., McGill 2003, Volkov et al. 2003, White et al. 2012, Connolly et al. 2014); 2) focus on a single ecosystem or taxonomic group (e.g., McGill 2003, Volkov et al. 2003); or 3) fail to use the most appropriate statistical methods (e.g., Ulrich et al. 2010). This makes it difficult to draw general conclusions about which, if any, models provide the best empirical fit to species abundance distributions.

Here, we evaluate the performance of five of the most widely used models for the species abundance distribution. We evaluate their performance using likelihood based model selection on data from 16,218 communities, from nine taxonomic groups. This includes data from terrestrial, aquatic, and marine ecosystems representing roughly 50 million individual organisms in total.

## Materials and Methods

### Data

We compiled data from citizen science projects, government surveys, and literature mining to produce a dataset with 16,209 communities, from nine taxonomic groups, representing nearly 50 million individual terrestrial, aquatic, and marine organisms (Table 3-1). Data for trees, birds, butterflies and mammals was compiled by White et al. (2012) from six data sources: the US Forest Service Forest Inventory and Analysis (FIA; USDA Forest Service 2010), the North American Butterfly Associations North American Butterfly Count (NABC; North American Butterfly Association 2009), the Mammal Community Database (MCDB; Thibault et al. 2011), Alwyn Gentry's Forest Transect Data Set (Gentry; Phillips and Miller 2002), the Audubon Society Christmas Bird Count (CBC; NAS 2002), and the US Geological Survey's North American Breeding Bird Survey (BBS; Pardieck et al. 2014). The publicly available datasets (FIA, MCDB, Gentry, and BBS) were acquired using the EcoData Retriever (Morris and White 2013). Details of the treatment of these datasets can be found in Appendix A of White et al. (2012). Data on Actinopterygii, Reptilia, Coleoptera, Arachnida, and Amphibia were mined from literature by Baldrige (see details in Chapter 2 of this dissertation). All abundances in the compiled datasets were counts of individuals.



## Models

The majority of species-abundance distributions (SADs) are constructed using counts of individuals (for discussion of alternative approaches, see McGill et al. 2007). As such, the data are discrete and as a result, the most appropriate models are discrete distributions. Therefore, we used only abundance data based on individual counts and used only discrete distributions that have been used as models for SADs.

McGill et al. (2007) classified models into five different families: purely statistical, branching process, population dynamics, niche partitioning, and spatial distribution of individuals. We evaluated models from each of the separate families, excluding the spatial distribution family, which requires spatially explicit data. Specifically, we evaluated the log-series, the Poisson log-normal, the negative binomial, the geometric series, and the Zipf distributions. All distributions were defined to have support defined by the positive integers (i.e., they are capable of having non-zero probability at values from 1 to infinity). We excluded models from analysis that do not have explicit likelihoods (e.g., some niche partitioning models; Sugihara 1980, Tokeshi 1993) so that we could use the likelihood based methods for fitting and evaluating distributions (see Analysis).

The log-series is one of the first distributions used to describe the SAD, being derived as a purely statistical distribution by Fisher et al. (1943). It has since been derived as the result of both ecological processes, the metacommunity SAD for ecological neutral theory (Hubbell 2001, Volkov et al. 2003), and several different maximum entropy models (Pueyo et al. 2007, Harte et al. 2008).

The lognormal is one of the most commonly used distributions for describing the SAD (McGill 2003) and has been derived as a null form of the distribution resulting from the central limit theorem (May 1975), population dynamics (Engen and Lande 1996), and niche partitioning

(Sugihara 1980). We use the Poisson lognormal because it is a discrete form of the distribution appropriate for fitting discrete abundance data (Bulmer 1974).

The negative-binomial (which can be derived as a mixture of the Poisson and Gamma distributions) provides a good characterization of the SAD predictions for several different ecological neutral models for the purposes of model selection (Connolly et al. 2014). We use it to represent neutral models as a class.

The geometric series was one of the first distributions derived as a model of the SAD and was derived based on niche partitioning (Motomura 1932).

The Zipf (or power law) distribution was derived based on branching processes and was one of the best fitting distributions in a recent meta-analysis of SADs (Ulrich et al. 2010).

## Analysis

Following current best practices for fitting distributions to data and evaluating their fit, we used maximum likelihood estimation to fit models to the data (Clark et al. 1999, Newman 2005, White et al. 2008) and likelihood based model selection to compare the fits of the different models (Burnham and Anderson 2002, Edwards et al. 2007). These general best practices have recently been affirmed as best practices for species abundance distributions (Connolly et al. 2014, Matthews and Whittaker 2014).

For model comparison, we used corrected Akaike Information Criterion (AICc) weights to compare the fits of models while correcting for differences in the number of parameters and appropriately handling the small sample sizes (i.e., numbers of species) in some communities (Burnham and Anderson 2002). The Poisson log-normal and the negative binomial each have two fitted parameters, while the log-series, geometric series, and Zipf distributions have one fitted parameter each. The model with the greatest AICc weight in each community was considered to

be the best fitting model for that community. We also assessed the full distribution of AICc weights to evaluate the similarity of the fits of the different models.

In addition to evaluating AICc of each model, we also examined the log-likelihood values of the models directly. We did this to assess the fit of the model while ignoring corrections for the number of parameters and the influence of similarities to other models in the set of candidate models. Model fitting, log-likelihood, and AICc calculations were performed using the macroecotools Python package at <https://github.com/weecology/macroecotools>. All of the code and the majority of the data necessary to replicate these analyses is available at <https://github.com/weecology/sad-comparison>. The CBC datasets and NABA datasets are not publicly available and therefore are not included.

The negative-binomial distribution failed to converge for 1444 sites in FIA (13.9%), 5 sites in Gentry (2.3%), 3 sites in Reptilia (2.2%), and 1 site in NABA (0.25%). For these sites likelihoods and AICc weights were calculated for only those models that successfully converged.

## Results

Across all datasets, the log-series had the lowest value of AICc, indicating the best fit to the data, in the greatest proportion of datasets (42.9%). The geometric series also performed well based on AICc, providing the best fit in 33.7% of the datasets. The Poisson lognormal and negative binomial distributions provided the best fit in 8.8% and 8.5% of the datasets respectively, and the Zipf distribution had the fewest cases of the lowest AICc with 6.1% of datasets (Figure 3-1).

Evaluating the best fitting distributions within individual datasets and taxonomic groups, the log-series was the most frequent best fitting model for all datasets except FIA (Figure 3-2). For the FIA data the geometric series provided the most frequent best fit to the data, and the strong performance of the geometric series in the FIA data is the cause of its strong performance

when all of the data are analyzed together. The relative performance of the other models varies among datasets and taxonomic groups. The negative binomial performed well in the bird datasets (BBS and CBC), but was almost never the best fitting model for plants (FIA and Gentry), Coleoptera, Arachnida, or Reptilia. The Poisson lognormal performed well for the bird datasets and the Gentry tree data, but almost never won in the FIA and Coleoptera datasets (Figure 3-2). The Zipf distribution performed well for Arachnida, but was never the best fitting model for the bird datasets.

The full distribution of AICc weights shows separation among models (Figure 3-3). On average, the Zipf and geometric distributions perform poorly, with the primary mode of the weight distribution occurring near zero (Figure 3-3). However, the geometric distribution also exhibits better performance for a subset of communities, with a secondary mode near 0.5. This mode is driven by the FIA data. The negative binomial and the Poisson lognormal distributions have peaks around 0.1, with the Poisson lognormal also having a small peak close to 1.0 indicating that in a small number of cases it provides a fit that is clearly superior to that of the other distributions (Figure 3-3). The logseries performs the best overall, with a large mode spanning AICc values from 0.3 to 0.5, and secondary mode from 0.6-0.7 (Figure 3-3).

While the AICc weights show separation among models, these values include a correction for the number of parameters and are also influenced by the similarity between models. Therefore, we also compared the negative log-likelihoods of the different models to determine whether or not their absolute fits differed. Frequency distributions of log-likelihoods show almost complete overlap among models (Figure 3-4) and one-to-one plots of the likelihoods of each model against the likelihood of the log-series show that the likelihoods of the different models correspond almost perfectly for individual distributions (Figure 3-5). This indicates that all models fit the data equivalently and that differences in AICc weights resulted primarily from differences in the number of parameters and differences in how similar different models in the set

of models were (i.e., if three identically fitting models are included in the analysis none of them can have a AICc weight  $> 0.34$ ). Additional supplemental figures can be found in Appendix B.

## Discussion

Our extensive comparison of different models for the species abundance distribution (SAD) using rigorous statistical methods demonstrates that most existing models provide equivalently good absolute fits to empirical data. As a result, the models with the fewest parameters perform better in AIC-based model selection because these approaches penalize model complexity. Because the log-series provides equivalent likelihoods to the other distributions, has a single fitted parameter, is easy to fit to empirical data, and is the best overall model using standard model selection, it provides a good naive model for fitting SADs.

The similar absolute fits of these five commonly used distributions emphasizes the challenges of inferring the processes operating in ecological systems from the form of the abundance distribution. It is already well established that models based on different processes can yield equivalent models of the SAD, i.e., they predict distributions of exactly the same form (Cohen 1968). It is also possible for the same biological explanations to result in different forms of the species abundance distribution depending on community conditions (Hughes 1986). Our results support the idea that even when models do differ in their precise mathematical predictions that they are often not distinguishable enough to identify potential mechanisms with any degree of certainty (Volkov et al. 2005). In other words, it is difficult to distinguish among the different distributions used to characterize the SAD, let alone the processes that generate the form of a particular distribution.

In cases where it is desirable to infer process based on macroecological patterns like the SAD, compare the predictions of different models using multiple macroecological patterns simultaneously is likely to be more effective (McGill 2003). It has also been suggested that

examining second-order effects, such as the scale-dependence of macroecological patterns (Blonder et al. 2014) or how the parameters of the distribution change across gradients (Mac Nally et al. 2014), can provide better inference about process from these kinds of pattern.

A previous analysis of ~500 SADs comparing three models, concluded that the form of the distribution varied consistently between fully censused communities, best fit by the lognormal, and incompletely sampled communities, best fit by the Zipf and logseries (Ulrich et al. 2010). The most completely sampled data in our analysis is arguably the forest inventories (Gentry, FIA), because these inventories count all trees above a certain stem diameter and detection of trees is straightforward so they are unlikely to be missed. The lognormal model is not the best fitting model in either of these datasets. The methods used by Ulrich et al. (2010) involve the use of binning and fitting models to rank abundance plots, which deviates from the best practices (Matthews and Whittaker 2014) used in this paper. A comparison of these two studies with equivalent methods will be necessary to resolve the discrepancies with respect to the influence of sampling on the observed form of the SAD.

In some cases, linking ecological patterns to particular sets of processes is not the goal. In particular, ecological patterns can be used for prediction in the absence of any link to process. For example, the species-area relationship, which characterizes how the number of species observed changes with spatial scale, is often used to make predictions for how many species will occur at larger and smaller scales than those observed. This is done without a strong link between biological processes and the empirical pattern. The SAD has been similarly used by White et al. (2012) who used the log-series to make predictions for the number of rare species occurring in a community. These predictions are independent of the processes generating the log-series. Given the equivalent fit of the five different distributions observed in this study, it is likely that any choice of distribution would have yielded equivalently strong predictions. In fact, patterns that

not strongly contingent on the operation of specific processes can be applied to prediction more broadly, because it is not necessary to understand the detailed biology of the system to use them.

It is interesting to consider why so many different models for the SAD yield similar predictions and fits to empirical data. Frank (2009, 2011) suggests that general patterns do not result from specific processes, but from the fact that there are many possible ways in which that pattern can be generated. For the SAD, it has been shown that of the possible forms of the SAD (the “feasible set”) most have similar general shapes (Locey and White 2013). This suggests that most data and most model predictions will have similar forms because most possible forms are similar. Maximum entropy based predictions for the SAD similarly suggest that the observed SAD should be the most likely possible form based on the random assignment of abundances to species under some basic constraints (Pueyo et al. 2007, Harte et al. 2008, Harte 2011, White et al. 2012). The fact that we observed equivalent log-likelihoods across five different models from a diverse array of ecosystems and taxonomic groups, that are likely influenced by a diverse array of processes, supports the idea that the detailed processes operating in ecological systems are not having direct and meaningful influences on the SAD (White et al. 2012, but see Mac Nally et al. 2014).

## References

Blonder, B., L. Sloat, B. J. Enquist, and B. McGill. 2014. Separating macroecological pattern and process: comparing ecological, economic, and geological systems. *PloS One* 9:e112850.

Bulmer, M. 1974. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*:101–110.

Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer. New York, New York, USA.

Clark, R., S. Cox, and G. Laslett. 1999. Generalizations of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats. *Geophysical Journal International* 136:357–372.

- Cohen, J. E. 1968. Alternate derivations of a species-abundance relation. *American Naturalist*:165–172.
- Connolly, S. R. et al. 2014. Commonness and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences*: 8524-8529.
- Edwards, A. M. et al. 2007. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* 449:1044–1048.
- Engen, S., and R. Lande. 1996. Population dynamic models generating species abundance distributions of the gamma type. *Journal of Theoretical Biology* 178:325–331.
- Fisher, R. A. et al. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*:42–58.
- Frank, S. A. 2009. The common patterns of nature. *Journal of Evolutionary Biology* 22:1563–1585.
- Frank, S. A. 2011. Measurement scale in maximum entropy models of species abundance. *Journal of Evolutionary Biology* 24:485–496.
- Harte, J. 2011. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford University Press.
- Harte, J., T. et al. 2008. Maximum entropy and the state-variable approach to macroecology. *Ecology* 89:2700–2711.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography (mPB-32)*. Princeton University Press. Princeton, NJ, USA.
- Hughes, R. 1986. Theories and models of species abundance. *American Naturalist*: 879–899.
- Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. *Ecology Letters* 16:1177–1185.
- Mac Nally, R. et al. 2014. The control of rank-abundance distributions by a competitive despotic species. *Oecologia* 176:849–857.
- Matthews, T. J., and R. J. Whittaker. 2014. Fitting and comparing competing models of the species abundance distribution: assessment and prospect. *Frontiers of Biogeography* 6.
- May, R. M. 1975. Patterns of species abundance and diversity. *Ecology and evolution of communities*:81–120.
- McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. *Nature* 422:881–885.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10:995–1015.



- Morris, B. D., and E. P. White. 2013. The EcoData Retriever: improving access to existing ecological data. *PLOS One* 8:e65848.
- Motomura, I. 1932. On the statistical treatment of communities. *Zoological Magazine* 44:379–383.
- National Audubon Society. 2002. The Christmas Bird Count historical results. National Audubon Society, New York, New York, USA.
- Newman, M. E. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46:323–351.
- North American Butterfly Association. 2009. NABA butterfly counts: 2009 report. NABA, Morristown, New Jersey, USA.
- Pardieck, K. L. et al. 2014. North American breeding bird survey dataset 1966 - 2013, version 2013.0. U.S. Geological Survey, Patuxent Wildlife Research Center.
- Phillips, O., and J. S. Miller. 2002. Global patterns of plant diversity: Alwyn H. Gentry's forest transect data set. Missouri Botanical Garden Press, St., Louis, Missouri, USA.
- Pueyo, S. et al. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology Letters* 10:1017–1028.
- Sugihara, G. 1980. Minimal community structure: an explanation of species abundance patterns. *American Naturalist*:770–787.
- Thibault, K. M. et al. 2011. Species composition and abundance of mammalian communities: ecological archives e092-201. *Ecology* 92:2316–2316.
- Tokeshi, M. 1993. Species abundance patterns and community structure. *Advances in Ecological Research* 24:111–186.
- Ulrich, W. et al. 2010. A meta-analysis of species–abundance distributions. *Oikos* 119:1149–1155.
- USDA Forest Service. 2010. Forest inventory and analysis national core field guide (phase 2 and 3).version 4.0. USDA Forest Service, Forest Inventory; Analysis.
- Volkov, I. et al. 2005. Density dependence explains tree species abundance and diversity in tropical forests. *Nature* 438:658–661.
- Volkov, I. et al. 2003. Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037.
- White, E. P. et al. 2008. On estimating the exponent of power-law frequency distributions. *Ecology* 89:905–912.

White, E. P. et al. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* 93:1772–1778.

Table 3-1. Details of datasets used to evaluate the form of the species-abundance distribution.

Datasets marked as Private were obtained through data requests to the providers resulting in Memorandums of Understanding governing data use.

Dataset	Dataset code	Availability	Total sites	Citation
Breeding Bird Survey	BBS	Public	2769	Pardieck et al. (2014)
Christmas Bird Count	CBC	Private	1999	NAS (2002).
Alwyn Gentry's Forest Transects	Gentry	Public	10355	Phillips and Miller (2002)
Forest Inventory Analysis	FIA	Public	220	USDA Forest Service (2010)
Mammal Community Database	MCDB	Public	103	Thibault et al. (2011)
North American Butterfly Count	NABA	Private	400	North American Butterfly Association
Actinopterygii	Actinopterygii	Public	161	this dissertation
Reptilia	Reptilia	Public	129	this dissertation
Amphibia	Amphibia	Public	43	this dissertation
Coleoptera	Coleoptera	Public	5	this dissertation
Arachnida	Arachnida	Public	25	this dissertation

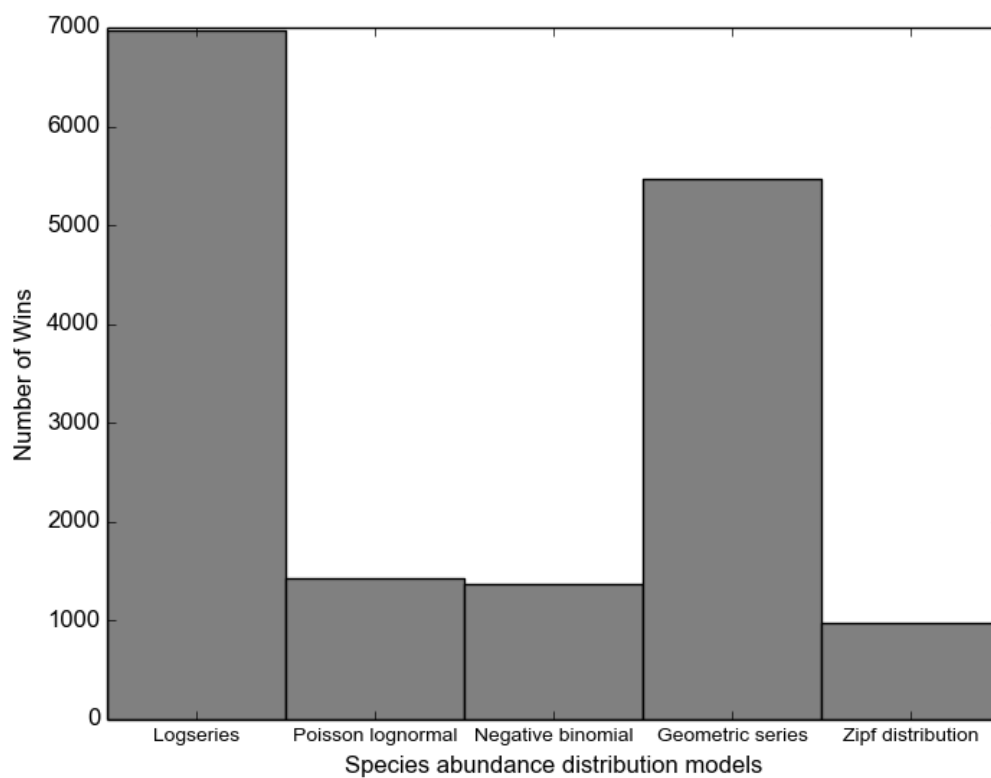


Figure 3 – 1 Total number of wins by model for all datasets combined.

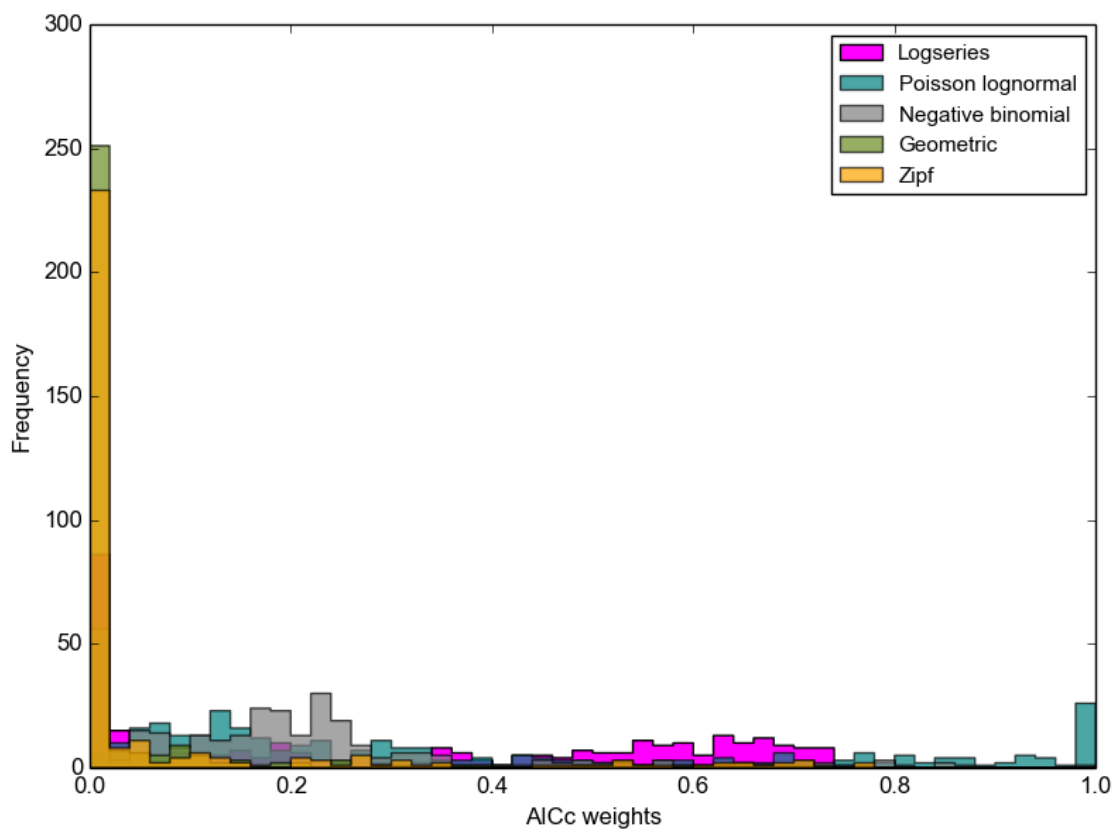


Figure 3 – 2 Frequency distributions of AICc weights by model for all datasets combined.

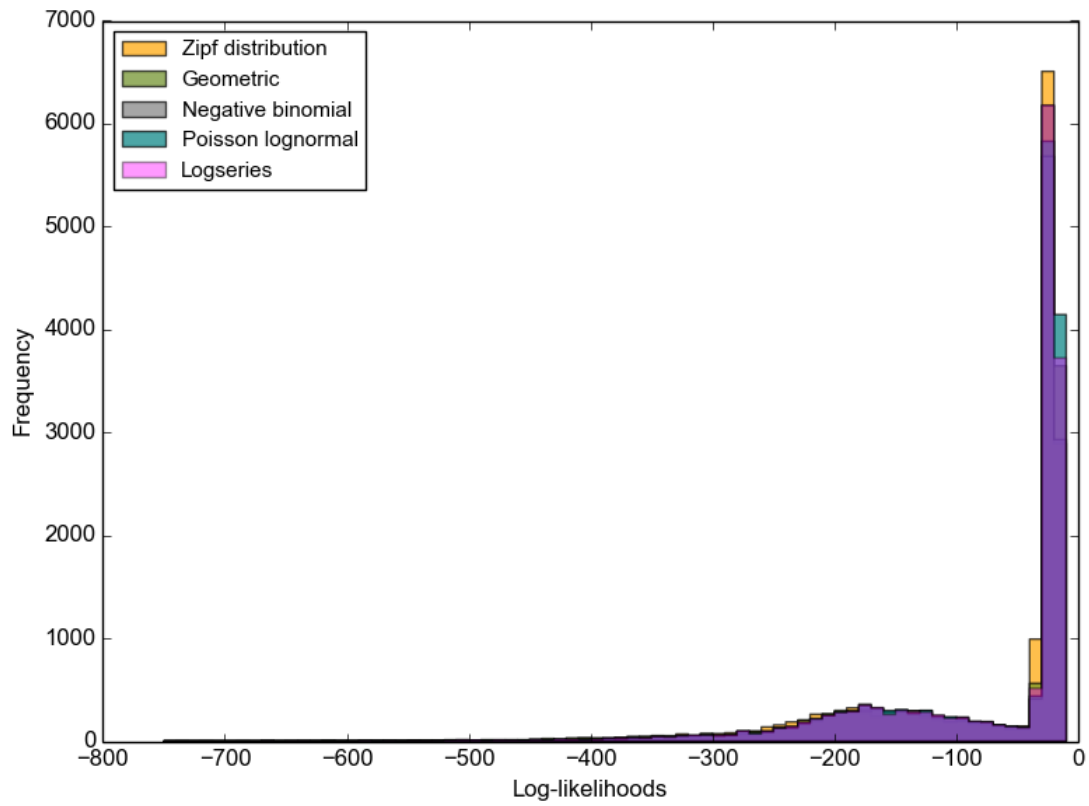


Figure 3 – 3 Frequency distributions of log-likelihoods by model for all datasets combined.

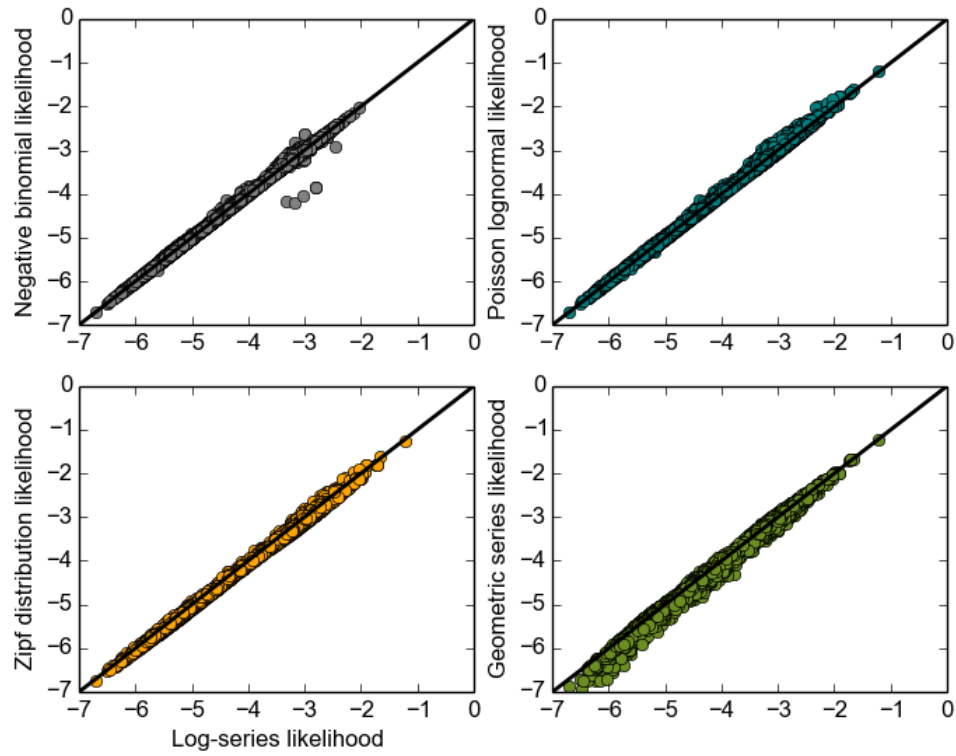


Figure 3 – 4 One-to-one plots of the likelihoods of each model against the likelihood of the log-series. The likelihoods of the different models have almost perfect correspondence for individual distributions.

CHAPTER 4  
EVALUATING ABUNDANCE DISTRIBUTION BASED SIGNALS OF NEUTRALITY IN  
TERRESTRIAL SYSTEMS

### Introduction

One of the fundamental goals of ecology is understanding what processes are important in structuring ecological communities. One of the major areas of debate surrounding this goal is whether simple neutral models that ignore differences between species can explain many of the empirical patterns observed in ecological systems (McGill et al. 2006, Rosindell et al. 2012, Matthews and Whittaker 2014). While there are multiple formulations of neutral theory, all models are based on the assumption that species and individuals are ecologically and demographically equivalent to one another, meaning that stochastic variation in birth, death, immigration, and speciation drives differences in a broad array of ecological patterns including the species abundance distribution, the species-area relationship, and the distance decay of similarity (Rosindell et al. 2011).

Early evaluations of neutral theory were based, in part, on comparing the fit of empirical species abundance distributions to the neutral prediction (e.g., Hubbell 2001, McGill 2003, Volkov et al. 2003). However, further evaluations of neutral theory suggested that comparisons based on the species abundance distribution were not sufficient for rigorous tests of neutrality (Volkov et al. 2005, 2006, McGill et al. 2006). This idea is further supported by work suggesting that species abundance distributions may contain little information about the detailed processes operating in ecological system more generally (Pielou 1975, White et al. 2012, Locey and White 2013). In contrast, recent work by Connolly et al. (2014) suggests that comparisons of species abundance distributions may be sufficient for evaluating whether or not neutral processes are dominant or whether other processes are important in structuring communities. Building on work



by Pueyo (2006), Connolly et al. (2014) were able to demonstrate that simulated neutral communities were typically better fit by negative-binomial distributions (referred to as Poisson gamma distributions by Connolly et al. 2014) than by Poisson lognormal distributions. They then performed the same analysis on over 1000 marine communities, and showed that the empirical communities were better fit by the lognormal (Connolly et al. 2014). This suggests that, at least in marine environments, the shape of the species abundance distribution can be used to exclude neutral processes as the sole determinant of community structure. By focusing on the detailed fits of alternative models, this approach takes advantage of “second-order effects,” which have been proposed to provide an avenue for inferring ecological process based on patterns of community structure (Blonder et al. 2014).

While this approach has been well tested within marine communities, it has not yet been used in terrestrial systems. Here, we use Connolly et al.’s (2014) method to assess potential patterns of neutrality across a broad range of ecosystems and taxonomic groups. We tested this approach for vertebrate, invertebrate and plant communities in primarily terrestrial ecosystems. In total, we used abundance data from 16,218 communities from across to globe to determine whether we observe patterns that are more consistent with neutrality or non-neutrality.

## Methods

### Data

We compiled data from nine distinct taxonomic groups and include birds, mammals, reptiles, amphibians, beetles, spiders, butterflies, trees, and bony fish from 16,209 distinct communities over all major biogeographic regions (Table 4-1, Figure 4-1). This dataset is a combination of the data compiled by White et al. (2012) and the data described in Chapter 2 of this dissertation. The majority of the data are publicly available and were accessed through the EcoData Retriever (Morris and White 2013). These data included the US Geological Survey’s

North American Breeding Bird Survey (BBS; Pardieck et al. 2014), Mammal Community Database (MCDB; Thibault et al. 2011), US Forest Service Forest Inventory and Analysis (FIA; USDA Forest Service 2010), and Gentry's Forest Transect Data Set (Gentry; Phillips and Miller 2002), and the data from Chapter 2. The North American Butterfly Association count data (NABC; North American Butterfly Association 2009) and the Audubon Society Christmas Bird Count (CBC; NAS 2002) are not publicly available and were obtained through Memorandums of Understanding with their respective organizations. The total number of sites per taxa and dataset is presented in Table 4-1. The locations of all of the georeferenced data are presented in Figure 4-1. Note that the data for reptiles, amphibians, bony fish, beetles, spiders, and butterflies are not represented, due to a lack of detailed location data.

## Analysis

Following Connolly et al. (2014), we used maximum likelihood methods for fitting and evaluating species abundance distributions models to data (the currently accepted best practice) (White et al. 2008, Connolly et al. 2014, Matthews and Whittaker 2014). This yielded fits of each distribution to each of the 16,000 communities in the dataset (Figure 4-2). Connolly et al. (2014) used Akaike Information Criterion (AIC) weights to compare the fits of the negative binomial and Poisson lognormal distributions to the empirical data. We modified this approach slightly by using weights calculated from the corrected Akaike Information Criterion (AICc) values, because AICc is more robust to small sample sizes (Burnham and Anderson 2002), which was a consideration for some communities. Model weights were calculated relative to the Poisson lognormal, meaning that weights near zero support the negative-binomial as the better fitting model while weights near one support the Poisson lognormal as the better fitting model.

Following the approach of Connolly et al. (2014), we looked at the relationship between the AICc weight and the number of distinct abundance values in the dataset. Connolly et al.

(2014) have argued that sites with more distinct abundance values provide greater power for differentiating among the fit of different models of the abundance distribution. Therefore, if the Poisson lognormal is superior to the negative-binomial distribution, the prediction is that AICc weights should approach one as the number of distinct abundance values increase. In Connolly et al.'s analysis, distinct abundance values greater than ~15 yielded AICc weights consistently above 0.8. Data were first analyzed at the level of the individual site. We also evaluated the patterns of the average AICc and number of distinct abundance values for each dataset as a whole. These approaches differ somewhat from those of Connolly et al., in that: 1) do not present individual site level results; and 2) the structure of our data is different from Connolly et al.'s in that there are not natural spatial groupings, and thus grouping at different spatial scales is less natural. Therefore, we have only analyzed the patterns at the site and whole dataset levels.

## Results

The site level results show a large amount of scatter in the values of AICc both within and among datasets (Figure 4-3). Values in all datasets with reasonably large numbers of data points range from near zero to near 1. There is no consistent directional trend in AICc weight as a function of the number of distinct abundance values. In datasets where there is some directional trend in AICc with the number of distinct abundance values (Reptilia, BBS, FIA), the trend tends to be towards zero (i.e., a better fit for the negative binomial distribution) as opposed to the trend toward one (i.e., a better fit for the Poisson lognormal) observed by Connolly et al. (2014) in marine systems (Figure 4-3).

Averaging the AICc weights and distinct abundance values across all sites in a dataset yielded similar results, with all average AICc values between 0.35 and 0.7, and no notable trend in average AICc as a function of the number of distinct abundance values (Figure 4-4).

## Discussion

The use of the species abundance distribution as a tool for identifying the processes operating in ecological systems has been widely questioned (Volkov et al. 2005, 2006, McGill et al. 2006, Al Hammal et al. 2015), which makes recent results showing that it is possible to use the SAD to evaluate whether neutral processes are the dominant structuring process in ecological systems exciting (Connolly et al. 2014). In contrast to Connolly et al.'s results, which suggest that marine systems are demonstrably non-neutral in the structure of their SADs, our analysis suggests that terrestrial systems overall cannot be clearly defined as either neutral or non-neutral based on this type of analysis. Our results were consistent with our broad comparison of five different species abundance distribution models, which showed that it is difficult to identify a clear winning model (see details in Chapter 3 of this dissertation). They are also consistent with a number of studies that have suggested that it should be difficult to identify underlying processes from the form of the abundance distribution alone (Pielou 1975, Volkov et al. 2005, 2006, McGill et al. 2006, White et al. 2012, Locey and White 2013, Al Hammal et al. 2015). However, most of these studies focused on either terrestrial data or models originally based on terrestrial ecosystems. This suggests that there may be important differences between marine and terrestrial systems with regards to the processes operating in these systems and/or the ability to make inferences about these processes based on patterns like the SAD.

In combination, our results and those of Connolly et al. (2014) suggest that while marine systems are generally approximated by non-neutral dynamics, terrestrial systems show more variability between neutral and non-neutral dynamics. Several studies have noted that both patterns and processes may vary between marine and terrestrial systems (Webb 2012, Horne et al. 2015). While macroecological patterns have not been studied as extensively in marine systems as in terrestrial systems, marine and terrestrial systems do tend to exhibit many of the same general macroecological patterns (Webb 2012). However, Webb (2012) points out that while the same

general patterns may occur, the processes generating those patterns may be different. This could lead to subtle differences in the details of the patterns ('second order effects'), which Blonder et al. (2014) noted as the most promising avenue for identifying process using macroecological patterns.

One key difference between terrestrial and marine systems is the way in which these systems have been differently impacted by anthropogenic activities. There has been a difference in the historical intensity and patterns of resource extraction in marine systems (Goudie 2013). Humans, being terrestrial, have been able to exploit terrestrial and coastal systems for a long period of time (Grayson 2001, Mannino and Thomas 2002). Only recently in human history has technology advanced to intensively exploit non-coastal marine systems (resulting in such effects as fisheries collapses) (Jackson et al. 2001, Crain et al. 2008). One of the major differences is that in marine systems, anthropogenic efforts to extract resources have focused primarily on wild populations of consumers, while in terrestrial systems these efforts have focused more on land use for domesticated producers and consumers (Goudie 2013). These differences lead to direct influences on marine species, but indirect effects in terrestrial systems (e.g., through land use changes) (marine, Jackson et al. 2001, Tittensor et al. 2009; terrestrial, Haberl et al. 2007). This legacy of marine exploitation and over-exploitation is a distinctly non-neutral influence on the structure of marine species abundance distributions that has the potential to produce a strong non-neutral signal in the SAD. In fact, there is an area of research using abundance distributions in marine systems to identify disturbed systems (Gray et al. 1979, Patil and Taillie 1982, Warwick 1986, Magurran 2013). There are several additional potential explanations for the difference in results between our study and the Connolly et al. 2014 paper, some non-biological (spatial structuring, sampling intensity), others related to biological/ecological differences in the data.

Other significant differences also exist between terrestrial and marine systems that could result in the differences we observed. For example, marine and aquatic ecosystems can exhibit an inverted biomass pyramid when compared to terrestrial systems, dependant on the temporal scale of analysis (Trebilco et al. 2013). If species-abundance distributions are structured based on a currency other than the number of individuals, e.g., biomass or resource use (Thibault et al. 2004, Connolly et al. 2005, McGill et al. 2007, Morlon et al. 2009), then this difference in the relationship between biomass and abundance could create a difference between marine and terrestrial systems. Another potential explanation comes from the core-occasional/core-transient species concept, in which core species, which are both common and temporally persistent, demonstrate a different shape of the species abundance distribution than transient species, which are rare and temporally variable (Magurran and Henderson 2003, Ulrich and Zalewski 2006, Magurran 2007). Differences in proportions of core and transient species occurring in terrestrial and marine systems could drive a difference in the general form of the abundance distribution, and greater variation in the proportions of core vs. transient species in terrestrial systems could drive the higher variation in the results. It is know that significant variation in the proportion of core vs. transient species exists in bird communities (Coyle et al. 2013), but there is little information on how variable these proportions are in marine systems.

While the vast majority of the data that we tested was terrestrial (approximately 99%), our data did include 161 fish communities. Of these, the majority were freshwater, rather than marine. However, we observed the same general pattern of results for both freshwater and marine fish communities as we did for the other taxa. This leaves open the possibility that some non-biological difference between the two data compilations is driving the differences in the results.

One potential non-biological explanation for the difference between our results and Connolly et al.'s (2014) results is a difference in the spatial structure of the data: the data from

Connolly et al. (2014) is structured in natural spatial groupings, whereas the data that we used in this study is not. In this study, many of the sites are widely dispersed, or are not regularly dispersed over the landscape. These differences in spatial grouping may lead to results that are more consistent due to spatial similarity than our widely dispersed sites.

Another potential non-biological explanation is related to sampling intensity. It is possible that the way in which marine communities are sampled is different from sampling of terrestrial communities, resulting in differing intensity of sampling that produce different patterns. However, the diversity of data we used covers a broad range of sampling intensities, from complete censuses (completely sampled trees above some minimum size cutoff; Forest Inventory and Analysis and Gentry), to well sampled but incomplete surveys (e.g., Christmas Bird Count), to incompletely sampled and incomplete taxonomic resolution (e.g. spiders and beetles).

In general, because of the diversity of data sources and types in our data compilation, it seems unlikely that the differences are due to the non-biological sampling differences rather than biological differences. Our compilation includes data collected at scales from a few square meters (e.g., invertebrate surveys) to 10s of square meters (Forest Inventory and Analysis) to 10s of hectares (Christmas Bird Count). As noted above, they also include samples ranging from nearly complete surveys (trees) to communities where sampling of the local community is expected to be fairly sparse (invertebrates). Sampling also spans a broad array of general approaches including the use of traps, visual observations, identification by sound, and other approaches. The consistency of these results across diverse datasets makes it unlikely that any particular sampling approach/issue could generate the observed results.

Intermediate to the sampling and biological explanations for the observed difference between marine and terrestrial systems is differences in the core constraints on the observed abundance distribution. A variety of approaches for modeling species abundance distributions

suggest that species richness (S) and the total number of individuals (N) are important inputs that constrain the shape of the empirical pattern (Harte et al. 2008, Frank 2011, Harte 2011, White et al. 2012, Locey and White 2013). Consistent differences in the ratio of S/N for terrestrial vs. marine communities could provide another potential explanation. Further research needs to be done to determine if there is a difference in S/N ratios between the terrestrial data used in this study and the marine data used in Connolly et al. (2014).

An increasing number of studies, including this one, suggest that there may be meaningful differences between marine and terrestrial systems in macroecological patterns (Webb 2012, Horne et al. 2015). While the ‘first order’ shape of these patterns may appear consistent between marine and terrestrial systems, there may be notable ‘second order’ differences (Blonder et al. 2014) related to differences in the processes driving the pattern (Webb 2012). This highlights the need for greater integration between the traditionally isolated analyses of marine and terrestrial systems to help understand differences in the processes driving these systems and the patterns that result (Beck et al. 2012, Webb 2012).

## References

- Al Hammal, O., D. Alonso, R. S. Etienne, S. J. Cornell, and S. Allesina. 2015. When can species abundance data reveal non-neutrality? *PLOS Computational Biology* 11-11.
- Beck, J., L. et al. 2012. What’s on the horizon for macroecology? *Ecography* 1235:673–683.
- Blonder, B., L. et al. 2014. Separating macroecological pattern and process: comparing ecological, economic, and geological systems. *PLOS One* 9:e112850.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, New York, USA.
- Connolly, S. R. et al. 2005. Community structure of corals and reef fishes at multiple scales. *Science* 309:1363–1365.
- Connolly, S. R. et al. 2014. Commonness and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences*: 8524-8529.



- Coyle, J. R. et al. 2013. Opposing mechanisms drive richness patterns of core and transient bird species. *The American Naturalist* 181:E83–E90.
- Crain, C. M., et al. 2008. Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters* 11:1304–1315.
- Frank, S. A. 2011. Measurement scale in maximum entropy models of species abundance. *Journal of Evolutionary Biology* 24:485–496.
- Goudie, A. S. 2013. *The human impact on the natural environment: past, present, and future*. John Wiley & Sons, Indianapolis, IN, USA.
- Gray, J., M. et al. 1979. Pollution-induced changes in populations [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences* 286:545–561.
- Grayson, D. K. 2001. The archaeological record of human impacts on animal populations. *Journal of World Prehistory* 15:1–68.
- Haberl, H. et al. 2007. Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proceedings of the National Academy of Sciences* 104:12942–12947.
- Harte, J. 2011. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford University Press, Oxford, UK.
- Harte, J. et al., 2008. Maximum entropy and the state-variable approach to macroecology. *Ecology* 89:2700–2711.
- Horne, C. R. et al. 2015. Temperature-size responses match latitudinal-size clines in arthropods, revealing critical differences between aquatic and terrestrial species. *Ecology Letters* 18:327–335.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ, USA.
- Jackson, J. B. et al. 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science* 293:629–637.
- Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. *Ecology letters* 16:1177–1185.
- Magurran, A. E. 2007. Species abundance distributions over time. *Ecology Letters* 10:347–354.
- Magurran, A. E. 2013. *Measuring biological diversity*. John Wiley & Sons, Indianapolis, IN, USA.
- Magurran, A. E., and P. A. Henderson. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature* 422:714–716.

- Mannino, A.M. and K.D. Thomas 2002. Depletion of a resource? The impact of prehistoric human foraging on intertidal mollusk communities and its significance for human settlement, mobility and dispersal. *World Archaeology* 33: 452–474.
- Matthews, T. J., and R. J. Whittaker. 2014. Fitting and comparing competing models of the species abundance distribution: assessment and prospect. *Frontiers of Biogeography* 6-6.
- McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. *Nature* 422:881–885.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology letters* 10:995–1015.
- McGill, B. J., B. A. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. *Ecology* 87:1411–1423.
- Morlon, H. et al. 2009. Taking species abundance distributions beyond individuals. *Ecology Letters* 12:488–501.
- Morris, B. D., and E. P. White. 2013. The ecoData retriever: improving access to existing ecological data. *PLOS ONE* 8:e65848.
- National Audubon Society. 2002. The Christmas bird count historical results. National Audubon Society, New York, New York, USA.
- North American Butterfly Association. 2009. NABA butterfly counts: 2009 report. NABA, Morristown, New Jersey, USA.
- Pardieck, K. L. et al. 2014. North american breeding bird survey dataset 1966 - 2013, version 2013.0. U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD, USA.
- Patil, G., and C. Taillie. 1982. Diversity as a concept and its measurement. *Journal of the American Statistical Association* 77:548–561.
- Phillips, O., and J. S. Miller. 2002. Global patterns of plant diversity: Alwyn H. Gentry's forest transect data set. Missouri Botanical Garden Press St., Louis, Missouri, USA.
- Pielou, E. 1975. *Ecological diversity*. Wiley, New York, New York, USA.
- Pueyo, S. 2006. Diversity: between neutrality and structure. *Oikos* 112:392–405.
- Rosindell, J. et al. 2011. The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution* 26:340–348.
- Rosindell, J. et al. 2012. The case for ecological neutral theory. *Trends in Ecology & Evolution* 27:203–208.
- Thibault, K. M. et al. 2011. Species composition and abundance of mammalian communities: Ecological Archives e092-201. *Ecology* 92:2316–2316.

- Thibault, K. M. et al. 2004. Temporal dynamics in the structure and composition of a desert rodent community. *Ecology* 85:2649–2655.
- Tittensor, D. P. et al. 2009. Macroecological changes in exploited marine systems. Pages 310-337 in *Marine Macroecology*, editors J. D. Witman and K. Roy. University of Chicago Press, Chicago, IL, USA.
- Trebilco, R. et al. 2013. Ecosystem ecology: size-based constraints on the pyramids of life. *Trends in Ecology & Evolution* 28:423–431.
- Ulrich, W., and M. Zalewski. 2006. Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos* 114:338–348.
- USDA Forest Service. 2010. Forest inventory and analysis national core field guide (phase 2 and 3). version 4.0. USDA Forest Service, Forest Inventory; Analysis.
- Volkov, I. et al. 2005. Density dependence explains tree species abundance and diversity in tropical forests. *Nature* 438:658–661.
- Volkov, I. et al. 2006. Theoretical biology: comparing models of species abundance (reply). *Nature* 441:E1–E2.
- Volkov, I., et al.. 2003. Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037.
- Warwick, R. 1986. A new method for detecting pollution effects on marine macrobenthic communities. *Marine Biology* 92:557–562.
- Webb, T. J. 2012. Marine and terrestrial ecology: unifying concepts, revealing differences. *Trends in Ecology & Evolution* 27:535–541.
- White, E. P. et al. 2008. On estimating the exponent of power-law frequency distributions. *Ecology* 89:905–912.
- White, E. P. et al. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* 93:1772–1778.

Table 4-1 Description of total number of sites per taxa and dataset. Taxonomic groups are ordered by the total number of sites in the compiled dataset.

Taxa	Dataset(s)	Total sites
Trees	FIA, Gentry	10575
Birds	BBS, CBC	4768
Butterflies	NABC	400
Reptiles	Reptilia	138
Bony fish	Actinopterygii	161
Mammals	MCDB	103
Amphibians	Amphibia	43
Spiders	Arachnida	25
Beetles	Coleoptera	5

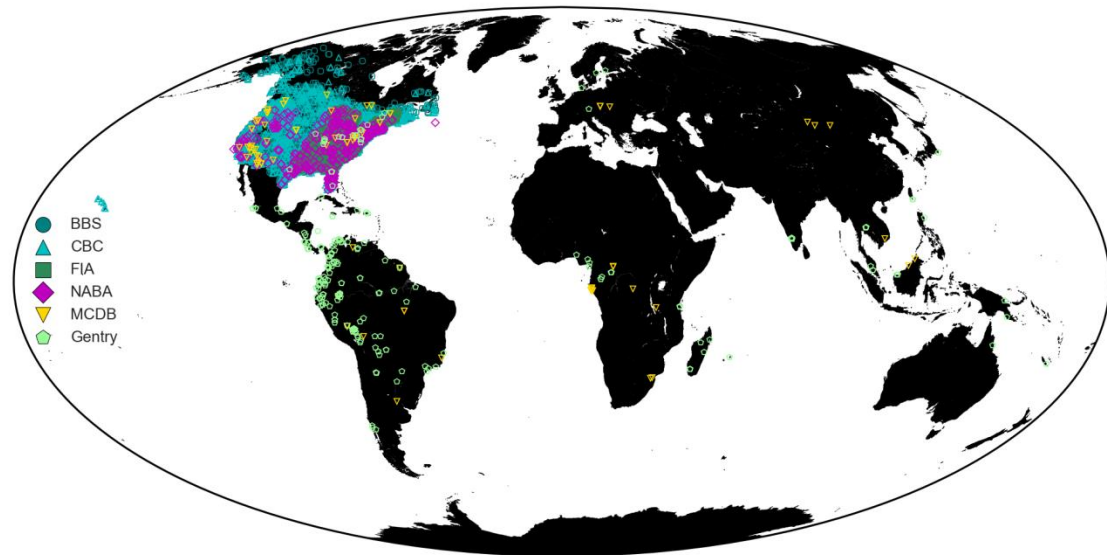


Figure 4 – 1 Map of the georeferenced portion of the data. Note that the data for reptiles, amphibians, bony fish, beetles, spiders, and butterflies are not represented. Redrawn from White et al. 2012.

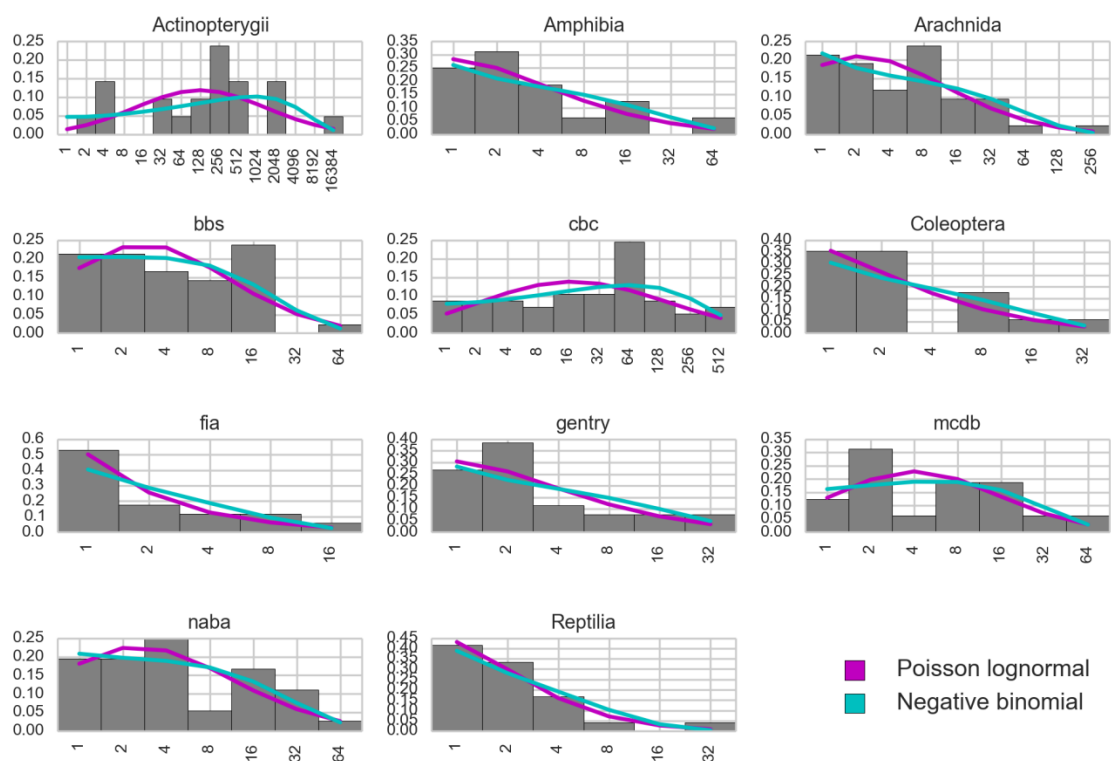


Figure 4 – 2 Preston plot of empirical data for single sites from each dataset with lines representing the Poisson lognormal and the negative binomial.

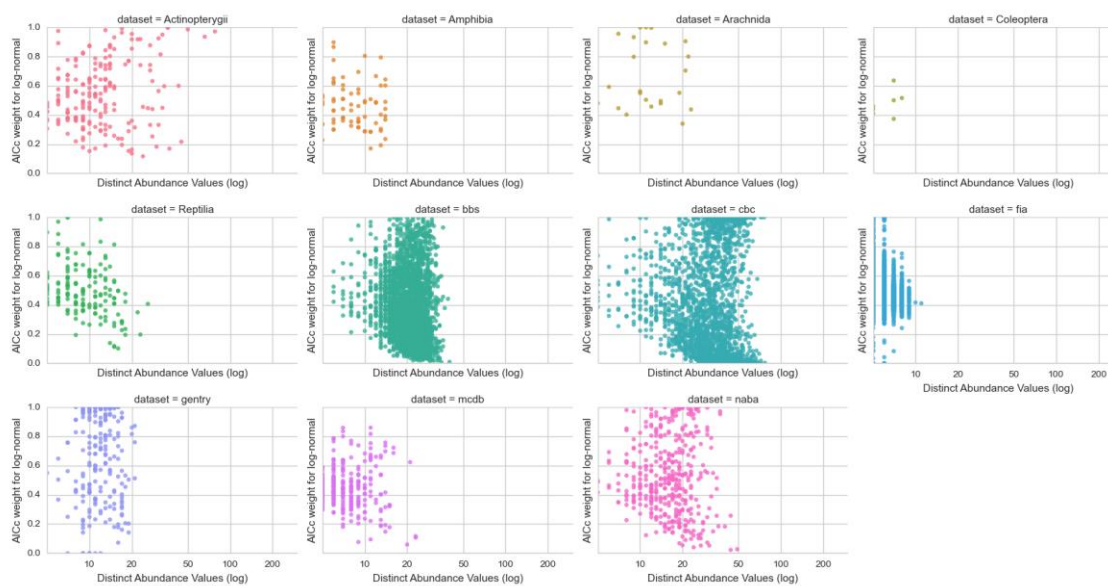


Figure 4 – 3 Log of distinct abundance values versus AICc weight of the lognormal distribution for each dataset.

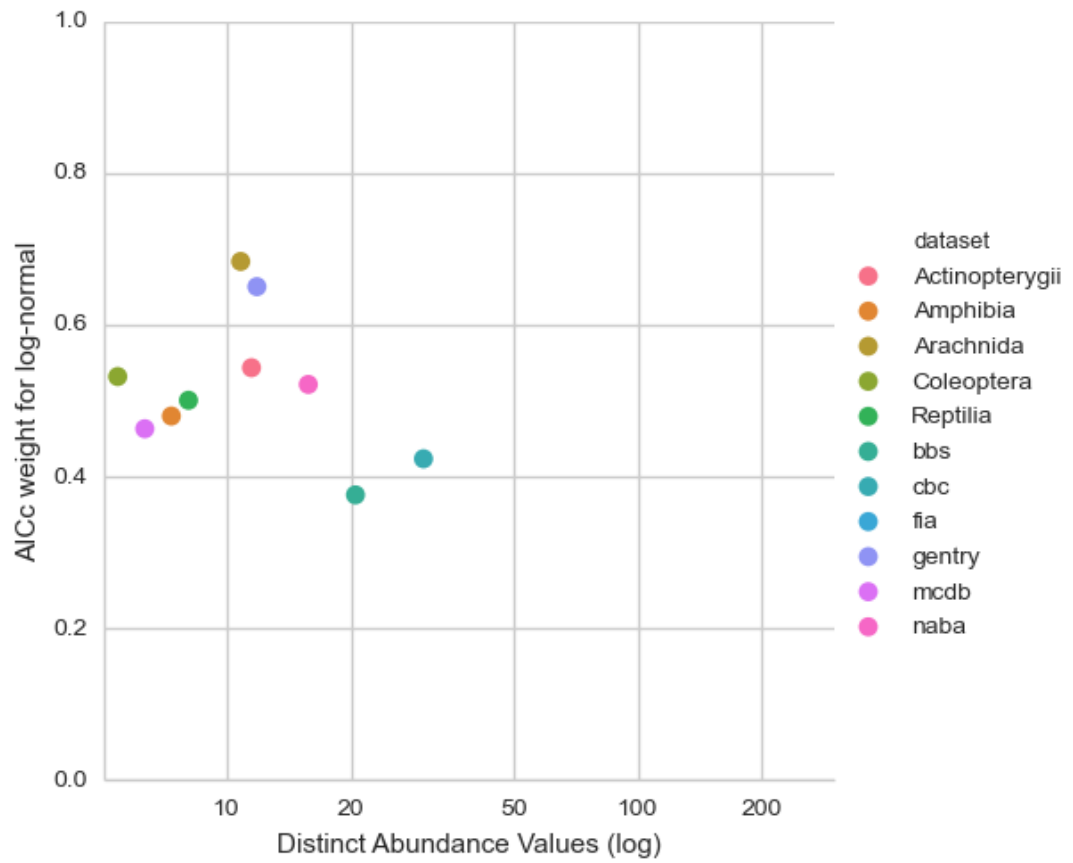


Figure 4 – 4 Average AICc weight of the lognormal distribution for each dataset.



## CHAPTER 5

### CONCLUSION

The species abundance distribution is one of the oldest and most well studied patterns in ecology (McGill et al. 2007). Despite the extensive study of this pattern, it remains an open question as to whether the pattern contains enough information to allow the operation of biological processes to be inferred from its shape (McGill et al. 2007, Harte 2011, White et al. 2012, Locey and White 2013, Connolly et al. 2014). I have compiled the largest species abundance distribution dataset ever evaluated and used it to both compare a suite of different models for the SAD and to evaluate whether a particular process can be detected as dominant in structuring a large set of SADs.

In general, I found that it is difficult to infer process from species abundance distributions alone. Part of the difficulty in identifying pattern generating mechanisms from species abundance distributions is due to the fact that multiple mechanisms have been proposed for each formulation of the species abundance distribution. In other words, it is possible for different processes to yield exactly equivalent models (Pielou 1975, McGill et al. 2007). Subsuming some of these differences into broad categories such as neutral or non-neutral (*sensu* Connolly et al. 2014) may make it possible to draw inferences on general categories of models. However, my results suggest that it may be difficult to distinguish among even these broad categories of models and their associated distributions, at least in terrestrial systems. In combination with the results of Connolly et al. (2014), this suggests that, in terrestrial systems, there may not be one single suite of processes that have equal importance in all communities, i.e., non-neutral processes may be more important in some communities, but not in others.

Using the largest compilation of species-abundance distributions ever assembled helps to mitigate a potentially important confounding factor when trying to identify pattern generating

mechanisms: non-biological variation among samples (sampling intensity, spatial scale, etc.) vs. biological differences. Applying data from different taxonomic groups and from different geographic regions assists in removing some of the uncertainty relating to non-biological data differences through covering a range of sampling intensities and scales of collection. When results are consistent across datasets using very different sampling approaches, as they were in this study, it provides confidence that methodological differences were not crucial in determining the results (White et al. 2012). Thus, the agreement in results among these different datasets strongly suggests that biological differences exist between marine vs. terrestrial in the dominance of non-neutral processes. Without a breadth of data in both my study and that of Connolly et al. (2014), it would have been difficult to differentiate biological from non-biological differences in the systems.

A traditional approach to scientific research is to identify general patterns, then pattern generating mechanisms (process), then use those processes to generate predictions. While this has been the traditional approach, pattern to process to prediction, identification of process may not be necessary for prediction in some cases. It may be possible to generate robust ecological predictions from general patterns without specific knowledge of the pattern generating mechanisms (Harte et al. 2008, Frank 2009, 2014, Harte 2011, Locey and White 2013). If true, this would mean that process and prediction may be two separate research goals (McGill and Nekola 2010, Perretti et al. 2013).

One criticism of the macroecological approach is that it attempts to replace natural history and field experimentation with entirely computational and observational approaches to experimentation. The increasing amount of data and computational power creates new opportunities to address major ecological questions in a new way. However, while computational techniques and large datasets are a powerful tool for ecology, they are not a panacea. Rather, the macroecological research program addresses different questions from a different perspective than

those of traditional ecology, and thus both approaches are vital to continue in the search for pattern generating mechanisms in ecology. Information to fully reproduce the analyses presented in this dissertation is available in Appendix C.

## References

- Connolly, S. R. et al. 2014. Commonness and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences*:8524–8529.
- Frank, S. A. 2009. The common patterns of nature. *Journal of Evolutionary Biology* 22:1563–585.
- Frank, S. A. 2014. Generative models versus underlying symmetries to explain biological pattern. *Journal of evolutionary biology* 27:1172–1178.
- Harte, J. 2011. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford University Press, Oxford, UK.
- Harte, J., et al. 2008. Maximum entropy and the state-variable approach to macroecology. *Ecology* 89:2700–2711.
- Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. *Ecology Letters* 16:1177–1185.
- McGill, B. J., and J. C. Nekola. 2010. Mechanisms in macroecology: aWOL or purloined letter? towards a pragmatic view of mechanism. *Oikos* 119:591–603.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* 10:995–1015.
- Perretti, C. T., et al. 2013. Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. *Proceedings of the National Academy of Sciences* 110:5253–5257.
- Pielou, E. 1975. *Ecological diversity*. Wiley, New York, New York, USA.
- White, E. P., et al. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* 93:1772–1778.

APPENDICES

APPENDIX A:  
METADATA FOR ABUNDANCE DATABASE

## INTRODUCTION

This dataset was developed to provide a source of abundance data for groups that do not have extensive compilations of abundance data.

There are several caveats to the use of this database. Abundance has been recorded as the raw abundance or the relative abundance, depending on what was available from the original source. Abundance is the total number of individuals captured, relative abundance is the total number of individuals captured for a single species/ total number of individuals of all species.

## METADATA CLASS I. DATA SET DESCRIPTIONS

### A. Data set identity:

Title: MiscAbundance

### B. Data set identification code:

1. Abundance data: Species\_abundances.csv
2. Sites data file : Sites\_table\_abundances.csv
3. Reference file: Citations\_table\_abundances.csv

### C. Data set description

Principal Investigators:

Elita Baldrige

Department of Biology and the Ecology Center

Utah State University

Logan, UT 84322

United States of America

D. Key words:

Abundance, communities, reptiles, spiders, beetles, fish, amphibians, birds

CLASS II. RESEARCH ORIGINS DESCRIPTORS

A. Overall project description

Identity:

Understanding drivers of species abundance.

Originators:

Elita Baldrige

Period of study:

2010 - 2012

Objectives:

To compile information on community abundance for vertebrate and invertebrate groups for which there were not already compiled abundance databases.

Abstract:

Same as above. This database is not a subset of a larger study.

Source(s) of funding:

E. B. has been supported by Utah State University.

## B. Specific subproject description

### Site description:

Sources contained data collected from a variety of different locations, resulting in a data set with a global extent.

### Experimental or sampling design:

Data were obtained from the scientific literature.

### Research Methods:

#### Field/Laboratory:

Not applicable

### Data Sources:

Data were compiled from the scientific literature. References for data sources are presented in a separate file, citations\_table\_abundances.csv, because of the extensive reference list involved with this project.

### Data Collection

Data were hand entered into a raw data file as they came from the original source or extracted from the original source computationally. All the data initially collected were not included in the final summary, because they were not deemed suitable for inclusion in the final database.

The data have not been updated taxonomically; the species names were kept as given in the original source.



Variables:

Abundance: total number of individuals captured

Relative\_abundance: total number of individuals captured for a single species/  
total number of individuals of all species.

Project Personnel:

Elita Baldrige

### CLASS III. DATA SET STATUS AND ACCESSIBILITY

#### A. Status

Latest Update:

February 2015 for the final format of all files

Latest Archive date:

February 2015

Metadata status:

Metadata is current.

Data verification:

Data quality has been checked as outline in class V, section B, below.

#### B. Accessibility

Storage location and medium:

The data are presented here in the following files:

Abundance data: Species\_abundances.csv

Reference file: Citations\_table\_abundances.csv

Sites data file : Sites\_table\_abundances.csv

Contact person:

Elita Baldrige, [elita.baldrige@weecology.org](mailto:elita.baldrige@weecology.org)

Copyright restrictions:

None.

Proprietary restrictions:

None.

Costs:

None. The authors believe that data should be freely available for use.

#### CLASS IV. DATA STRUCTURAL DESCRIPTORS

##### A. Data Set File

Identity:

1. Abundance data: Species\_abundances.csv
2. Sites data file : Sites\_table\_abundances.csv
3. Reference file: Citations\_table\_abundances.csv

Size:

1. 22143 records, including header row.
2. 707 records, including header row.
3. 117 records, including header row.

Format and Storage mode:

ASCII text, comma delimited, not compressed.

## Header information:

1. Class, Family, Genus, Species, Relative\_abundance, Abundance, Site\_ID, Citation\_ID
2. Site\_ID, Collection\_Year, End\_Collection, Citation\_ID, Site\_Name, Biogeographic\_region, Site\_notes
3. Citation\_ID, Authors, Yr, Title, Journal, Issue, Pages

## Alphanumeric attributes:

Mixed

## Special characters/fields:

Blanks indicate no data: no special characters used.

## Authentication procedures:

1. Sum of Relative\_abundance = 10797.37352
2. Sum of Abundance = 1320592

## B. Variable definitions

Variable name	Variable definition	Units	Storage type	Range of values	Missing value codes
Class	Taxonomic class of species	N/A	Character	N/A	
Family	Taxonomic family of species	N/A	Character	N/A	
Genus	Taxonomic genus of species	N/A	Character	N/A	
Species	Specific epithet of species	N/A	Character	N/A	
Relative_abundance	Relative abundance of species	N/A	Double	0 - 309	
Abundance	Abundance of species	N/A	Integer	0-181726	

Collection_Year	Start of collecting	N/A	Integer	1952-2008
End_Collection	End of collecting	N/A	Integer	1977-2009
Site_Name	Name/description of site	N/A	Character	N/A
Biogeographic_region	Biogeographic region	N/A	Character	N/A
Site_notes	Additional site information	N/A	Character	N/A

#### CLASS V. SUPPLEMENTAL DESCRIPTORS

##### A. Data acquisition

Data were compiled from the literature without the use of data forms.

##### B. Quality assurance/quality control procedures

Data have undergone initial quality and assurance checking. Data were entered directly from the source material into the raw data file and values were double checked on entry.

##### C. Related material:

Not applicable.

##### D. Computer programs and data processing algorithms:

Available for direct import to database through EcoData Retriever ([ecodataretriever.org](http://ecodataretriever.org))

##### E. Archiving:

Not applicable.

F. Publications and results:

Reference list for data set is available in Citations\_table\_abundance.csv. Data sources for each variable for each species are indicated by Citation\_ID in the files

Species\_abundances.csv and Sites\_table\_abundances.csv. The full reference list is not presented here due to the length of the list.

G. History of data set usage:

E. Baldrige 2015. A data-intensive assessment of the species abundance distribution. Dissertation, Utah State University.

H. Data set update history:

Review history:

Not applicable

Questions and comments from secondary users:

Not applicable

APPENDIX B:

CHAPTER 3: SUPPLEMENTARY RESULTS FOR CHAPTER 3

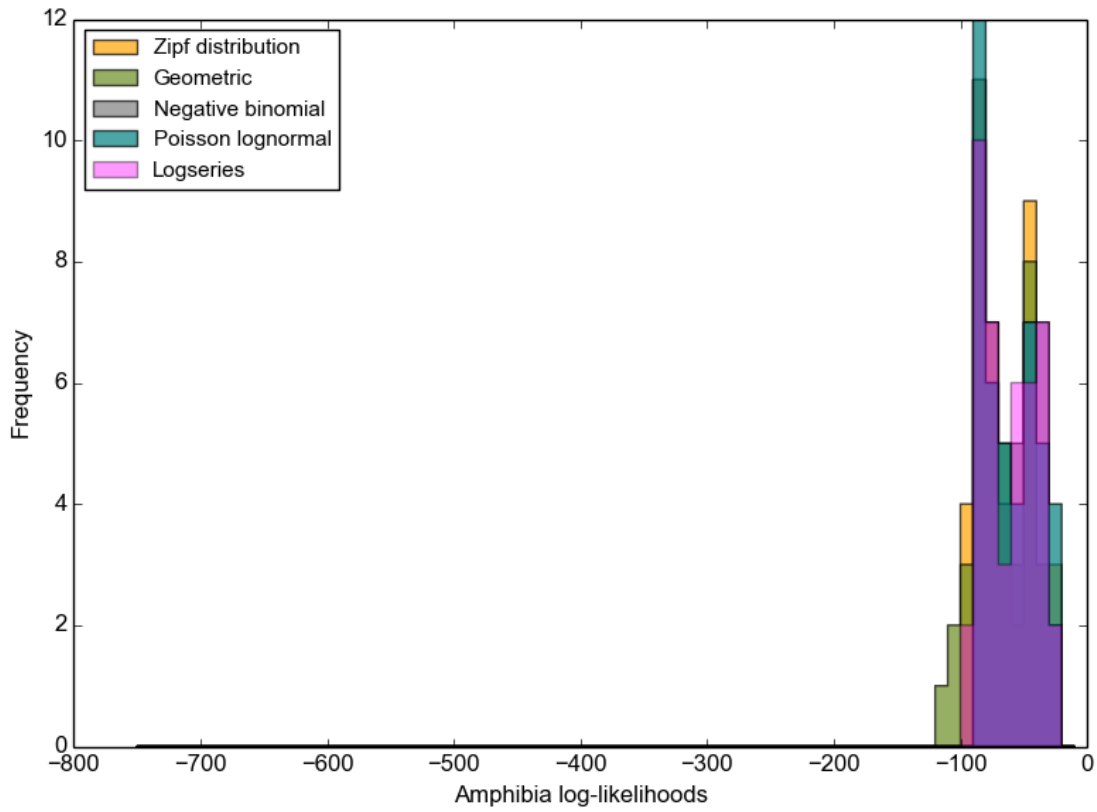


Figure B – 1 Frequency distribution of log-likelihoods for Amphibia

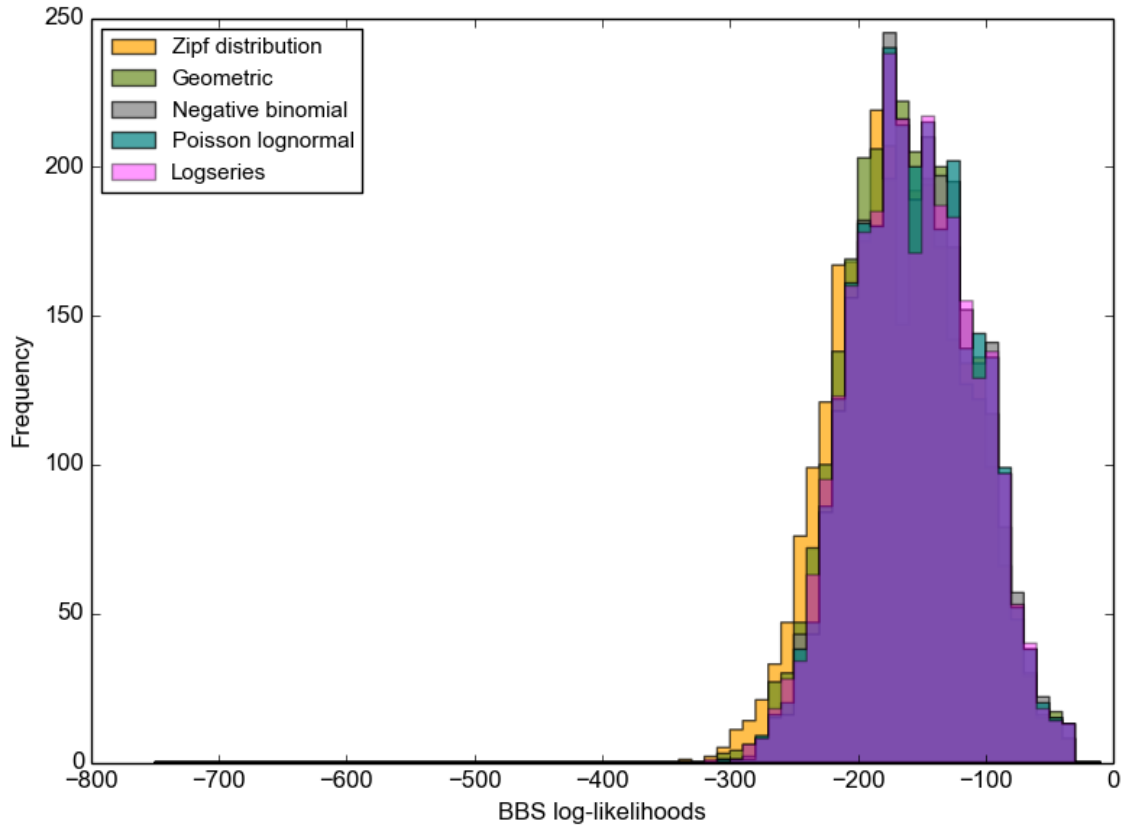


Figure B – 2 Frequency distribution of log -likelihoods for BBS



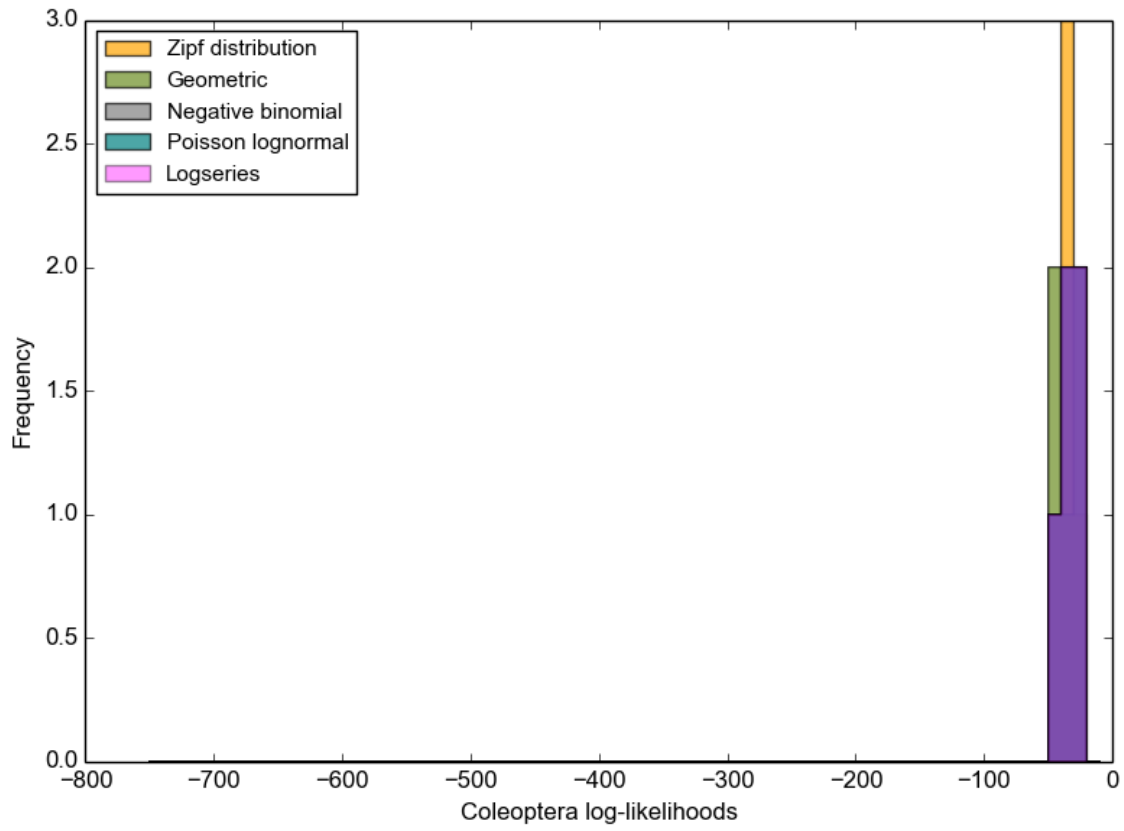


Figure B – 3 Frequency distribution of log -likelihoods for Coleoptera.

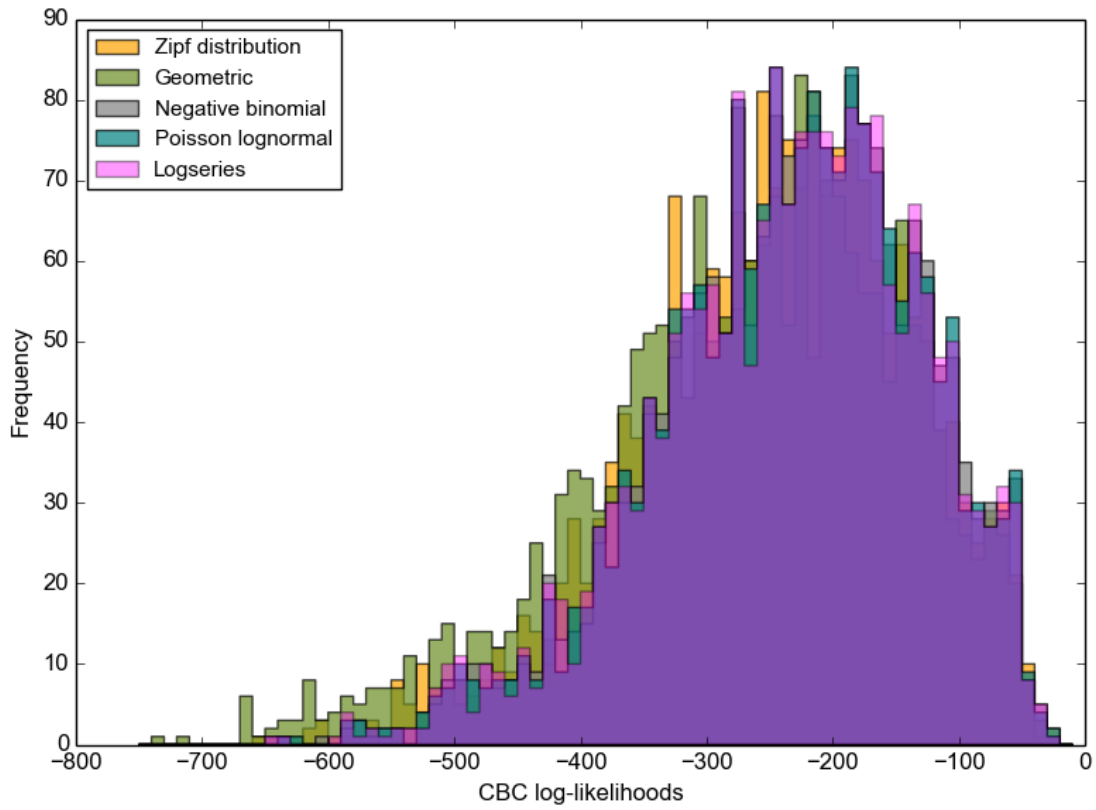


Figure B – 4 Frequency distribution of log -likelihoods for CBC.

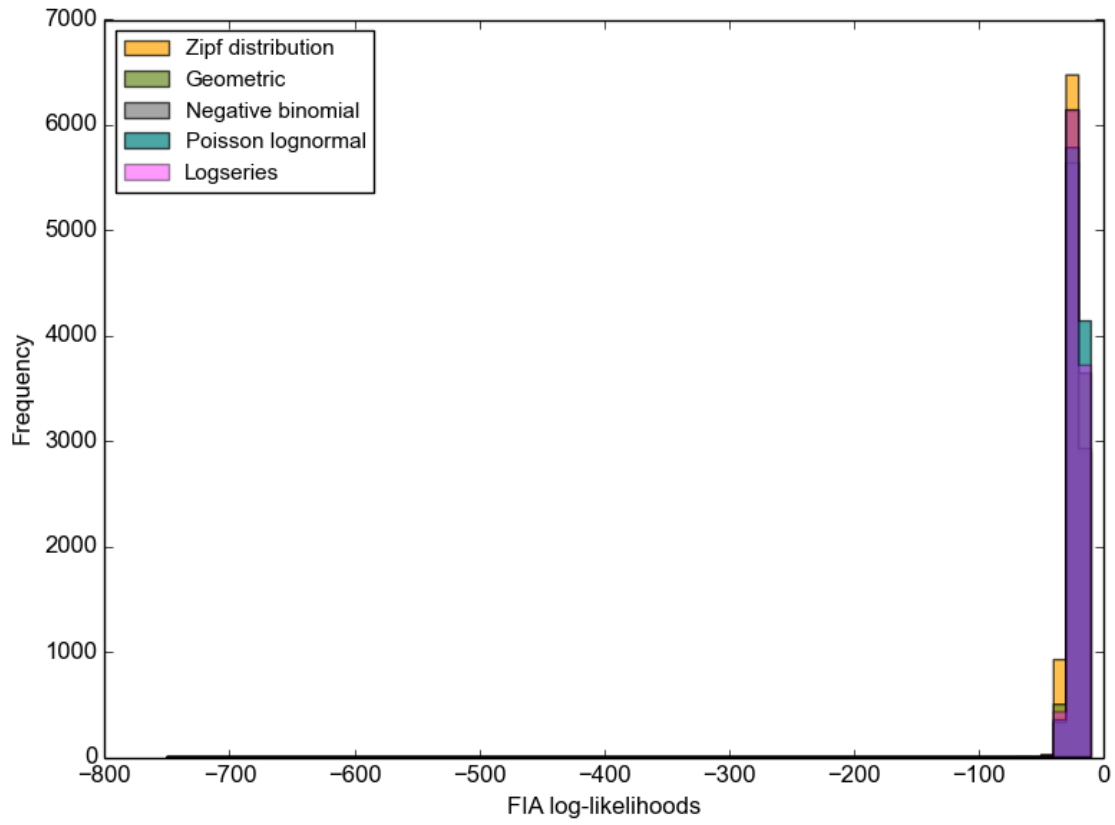


Figure B – 5 Frequency distribution of log -likelihoods for FIA.

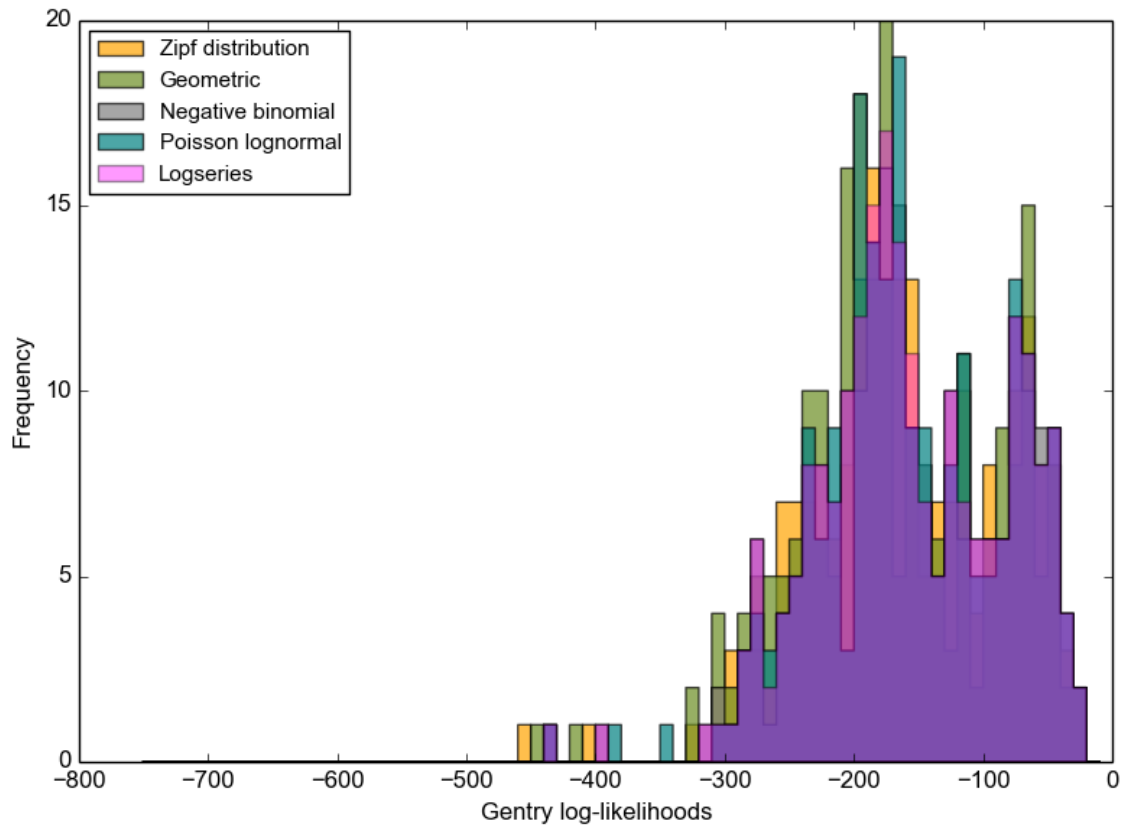


Figure B – 6 Frequency distribution of log -likelihoods for Gentry.

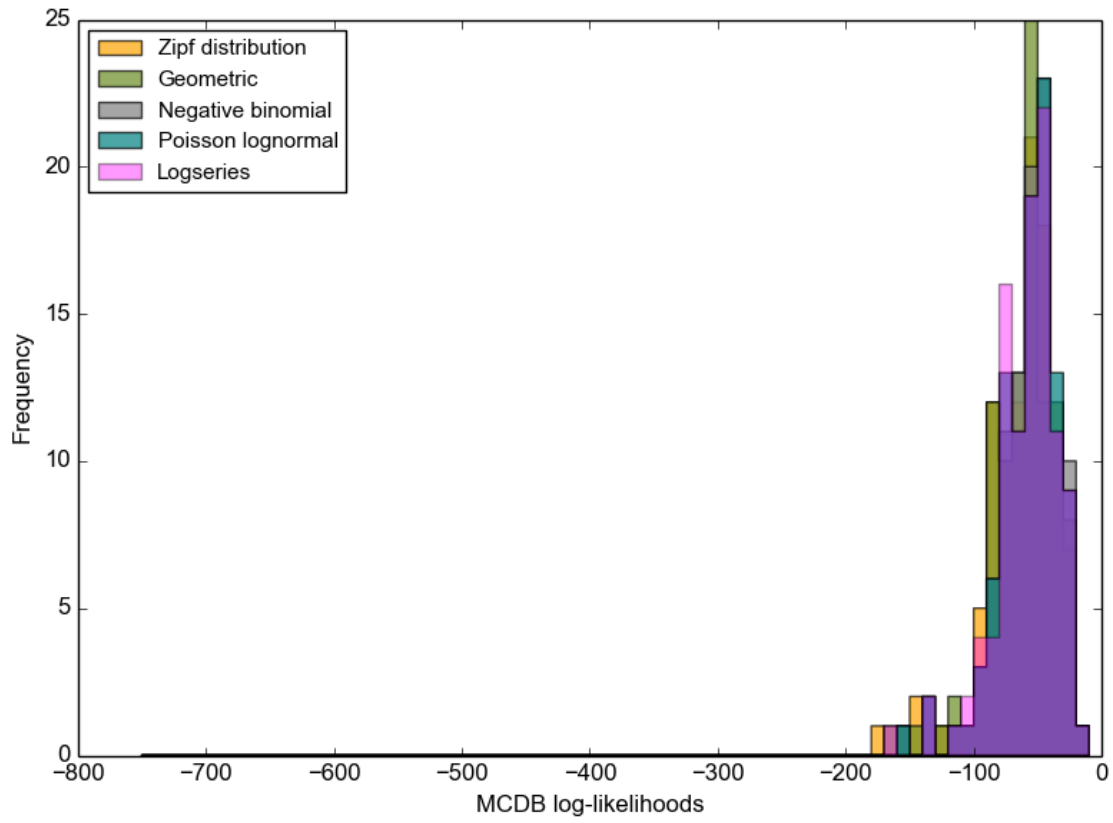


Figure B – 7 Frequency distribution of log -likelihoods for MCDB.

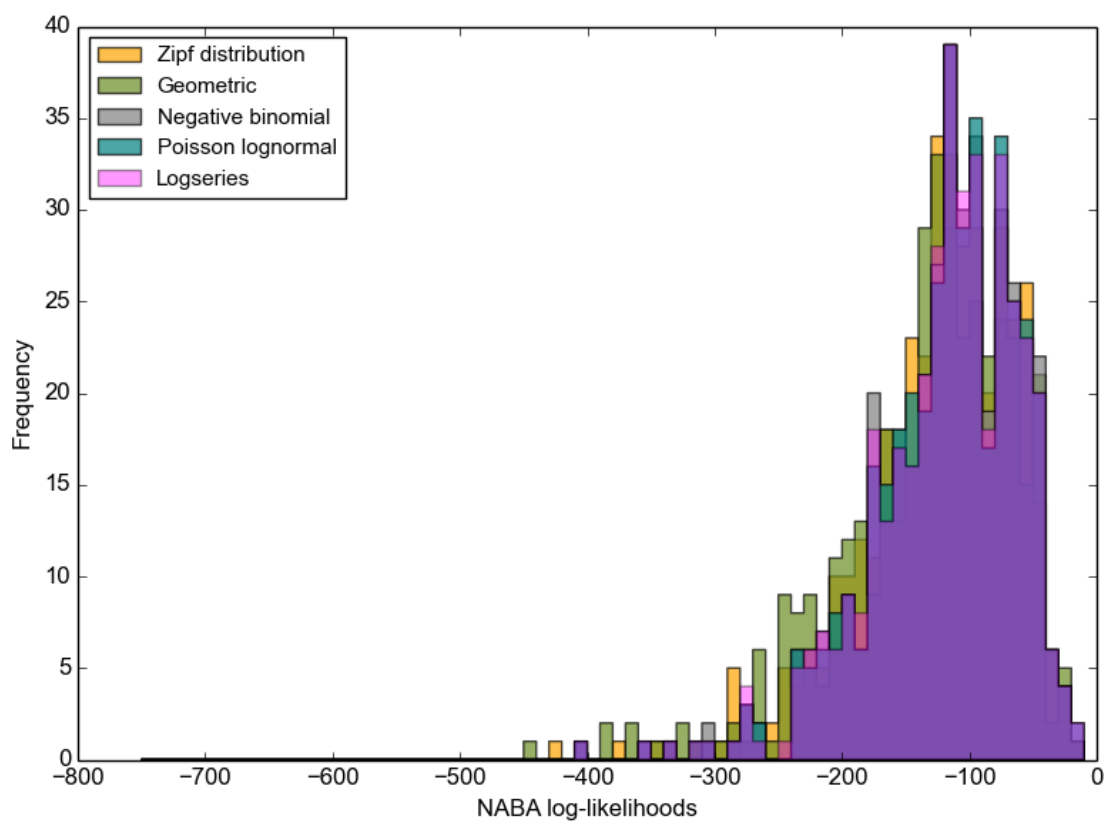


Figure B – 8 Frequency distribution of log -likelihoods for NABA.

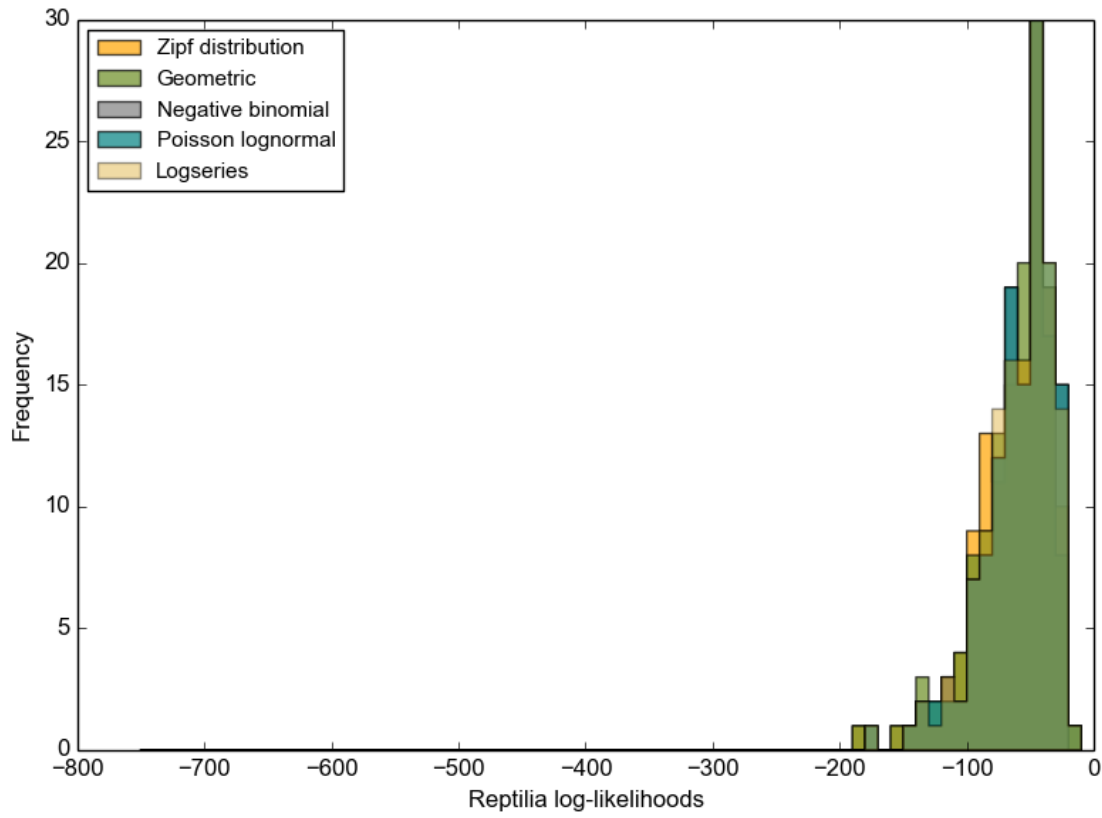


Figure B – 9 Frequency distribution of log -likelihoods for Reptilia.

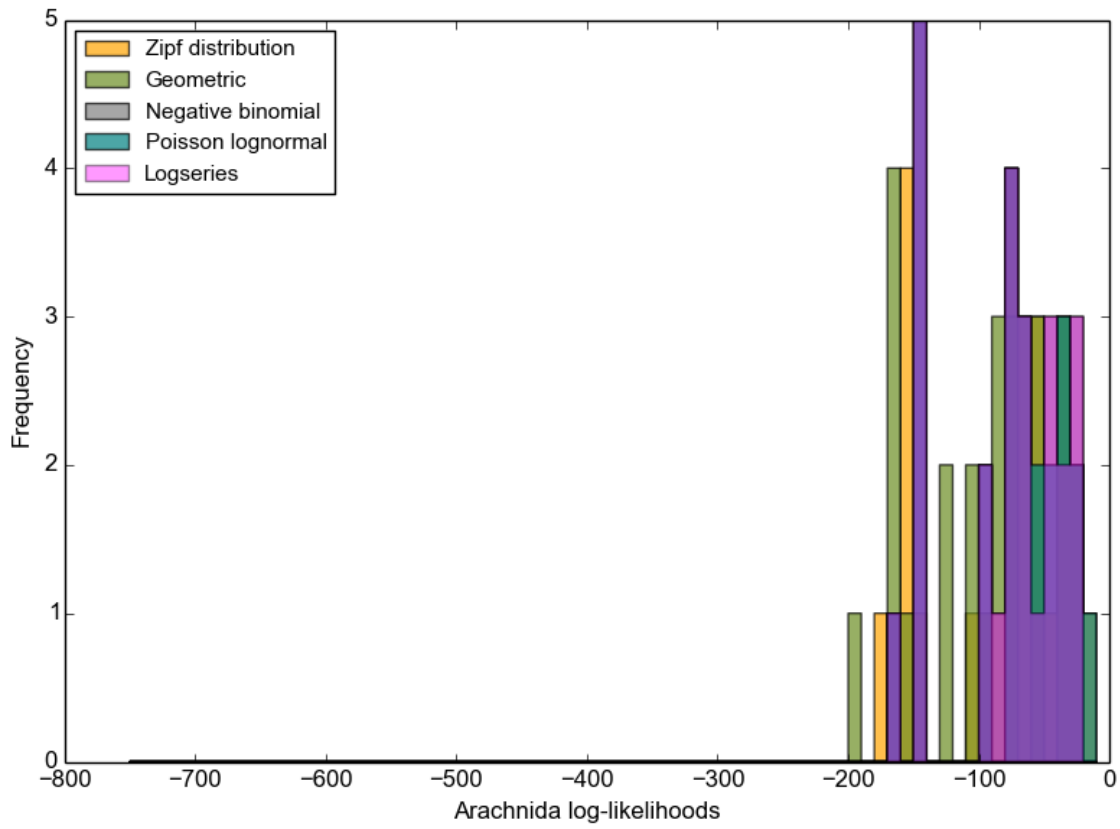


Figure B – 10 Frequency distribution of log -likelihoods for Arachnida



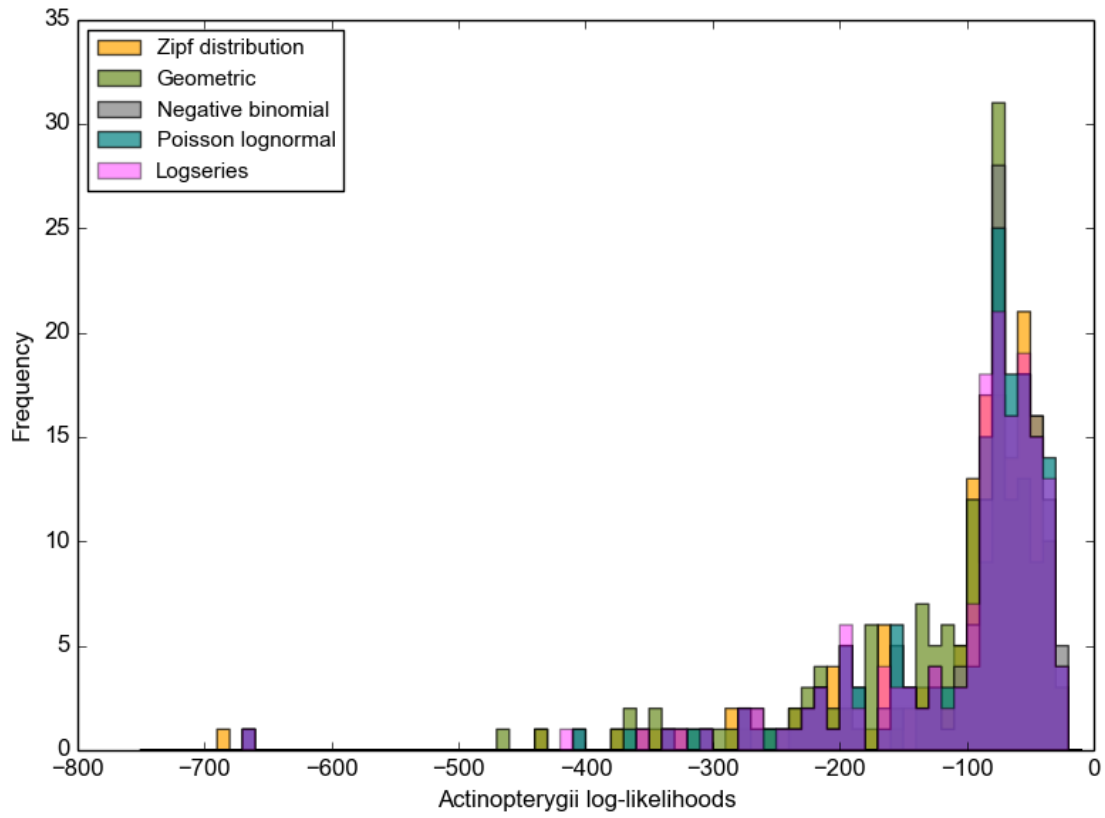


Figure B – 11 Frequency distribution of log -likelihoods for Actinopterygii

APPENDIX C:  
REPLICATION OF ANALYSES

Code for this dissertation has been archived following best practices.

<https://zenodo.org/record/16422>

<https://github.com/embaldrige/sad-comparison/archive/dissertation.zip>

<https://github.com/embaldrige/sad-comparison/archive/dissertation.tar.gz>

GitHub repository: sad-comparison (<https://github.com/weecology/sad-comparison>)

Repository for comparisons among species abundance distribution (SAD) models.

Data:

Community data the same as used in White EP, Thibault KM, Xiao X. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*. 93(8):1772-1778. Further details and scripts for data extraction and processing can be found in the GitHub repository for that paper at <https://github.com/weecology/white-et-al-2012-ecology>. Additional community data for Actinopterygii, Reptilia, Coleoptera, Arachnida, and Amphibia were mined from the literature and are publicly available for import through the EcoData Retriever (<https://github.com/weecology/retriever>) or on figshare (Baldridge, Elita (2013): Community abundance data. figshare. <http://dx.doi.org/10.6084/m9.figshare.769251>).

Python dependencies:

METE: <https://github.com/weecology/METE.git>

macroecotools: <https://github.com/weecology/macroecotools.git>

matplotlib

basemap for matplotlib: <http://matplotlib.org/basemap/users/installing.html#installation> numpy

scipy

pandas

seaborn

Installation:

The METE module and the macroecotools module can be installed from the command line (with appropriate permissions)

```
git clone https://github.com/weecology/METE.git
```

```
cd METE
```

```
python setup.py install (sudo python setup.py install on Linux)
```

```
cd ..
```

```
git clone https://github.com/weecology/macroecotools.git
```

```
cd macroecotools
```

```
python setup.py install (with sudo for Linux)
```

SAD models tested and packages used:

Maximum Entropy Theory of Ecology (METE) (METE)

Logseries (macroecotools/macroeco\_distributions)

Poisson lognormal (macroecotools/macroeco\_distributions)

Negative binomial (macroecotools/macroeco\_distributions)

Geometric series (macroecotools/macroeco\_distributions)

Neutral theory:

Because neutral theory predicts the negative binomial distribution at the local scale (Connolly et al. 2014. Commonness and rarity in the marine biosphere. PNAS 111: 8524-8529.

<http://www.pnas.org/content/111/23/8524.abstract>), we used the prediction for the negative

binomial distribution (macroecotools/macroecodistributions) instead of fitting the neutral theory model directly. The AICc for neutral theory was calculated with the appropriate number of parameters for neutral theory.

Reproducing the workflow and analyses in this repository:

Data extraction:

`misc-data-query.py` to extract the Baldrige 2013 data used in addition to the White 2012 data (Data were first imported into an sqlite database with the EcoData Retriever).

After data extraction:

To generate the results and figures from “An extensive comparison of species distribution models”:

Run:

`sad-comparisons.py` to perform the analyses

`sad-process-db.py` to create a database from the analysis results from `sad-comparisons.py`

`sad-comparison-graphs.py` to generate the figures

To generate the results and figures from “Evaluating abundance distribution based signals of neutrality in terrestrial systems”:

Run:

`sad_neutral_analysis.py` to perform the analysis and generate the figures.

## CURRICULUM VITAE

Elita Baldrige  
(April 2015)

## EDUCATION

Utah State University, Logan UT; in progress: 2010-present  
Ph.D. (Doctoral candidate), Ecology Dissertation topic: A data-intensive assessment of the species-abundance distribution.  
<https://github.com/weecology/sad-comparison>  
Advisor: Dr. Ethan White

Fort Hays State University, Hays, KS; 2010  
M.S., Biology  
Thesis title: Nested subsets, scale, and the distribution of abundance: A macroecological approach.  
Advisor: Dr. Rob Channell

Kansas State University, Manhattan, KS; 2007  
B.S., Wildlife Ecology and Management

## PROFESSIONAL EXPERIENCE

Fall 2012- Fall 2014, Spring 2015  
Graduate teaching assistant  
Utah State University, Logan UT.

Summer 2011- 2012  
Graduate research assistant  
Utah State University, Logan UT.

Fall 2010-Spring 2011, Fall 2014  
Graduate teaching assistant  
Utah State University, Logan UT.

Fall 2009-Spring 2010 Fleharty Fellow  
Fort Hays State University, Hays, KS.

Summer 2009  
Kansas Wetlands Education Center assistant  
Kansas Wetlands Education Center, Cheyenne Bottoms, KS.

Fall 2008-Spring 2009  
Graduate teaching assistant  
Fort Hays State University, Hays KS

## SUMMARY OF COURSES TAUGHT

Utah State University, Logan, UT

Introductory Biology Laboratory, Biology 1610 and 1620, one credit hour (Fall 2010, 2012, 2014; Spring 2011, 2013)

Taught introductory biology course for biology majors, covering evolution, basic genetics, diversity of life, and ecology.

Fort Hays State University, Hays, KS

Ecology Laboratory, one credit hour (Spring 2009)

Taught ecology laboratory for majors, covering diversity indices, life tables, foraging, distributions, and mark/recapture.

Microbiology for Allied Health Laboratory, one credit hour (Fall 2008, Spring 2009)

Assisted with laboratory preparation for non-majors course, making microbiological media and laboratory preparation for non-majors course, making microbiological media, culturing bacteria, as well as assisting students with laboratory activities.

Laboratory Experiences in Biology, one credit hour (Fall 2008)

Taught introductory biology course for non-majors, covering evolution, basic genetics, diversity of life, and ecology.

## RESEARCH

### RESEARCH INTERESTS

Commonness and rarity

Community ecology

Macroecology

Ecoinformatics

### PUBLICATION AND RESEARCH PRODUCT IMPACTS:

Impact Story: <http://weecology.org/people/elitabaldridge>

Google Scholar: <http://scholar.google.com/citations?user=BHY3EXUAAAAJ>

### PUBLICATIONS

[OA]: The published paper is open access (or at least free to read)

[OA version]: Link to an open or free version of the paper if the published version is not open access.

White, E.P, E. Baldridge+, Z.T. Brym+, K.J. Locey+, D.J. McGlinn+, S.R. Supp+. 2013. Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*. 6(2): 1–10.

[OA] <http://library.queensu.ca/ojs/index.php/IEE/article/view/4608/4898>

(+: Authors contributed equally to the work.)

E. Baldridge. 2010. Nested subsets, scale, and the distribution of abundance: A macroecological approach. Master's thesis, Fort Hays State University.

[OA version][http://files.figshare.com/1701415/Nested\\_Subsets\\_Master\\_s\\_Thesis.pdf](http://files.figshare.com/1701415/Nested_Subsets_Master_s_Thesis.pdf)

#### PRESENTATIONS

Ecological Society of America 2013 Meeting, Minneapolis MN  
Elita Baldrige & Ethan White, organizers. Ignite session "Constraints in Ecology".

Ecological Society of America 2011 Meeting, Austin TX  
Elita Baldrige & Rob Channell. Testing the assumptions of the nested subset pattern.

Southwestern Association of Naturalists 2010 Meeting, Junction TX  
Elita Baldrige & Rob Channell. Nested subset analysis: Examining geographic scale and abundance.

Kansas Ornithological Society 2009 Meeting, Hutchison KS  
Elita Baldrige & Rob Channell. Effects of environmental conditions and distance from the geographic range center on the abundance of avian species.

#### POSTERS

Gordon Research Conference 2014 "Unifying Ecology Across Scales", Boston MA  
Nathan Myrhvold, Elita Baldrige, Benjamin Chan, Dan Freeman, and Morgan Ernest.  
An Amniote Life History Database to Perform Comparative Analyses with Birds,  
Mammals, and Reptiles. Presented by Morgan Ernest  
[http://files.figshare.com/1587021/Ernest\\_Gordon\\_2014.pptx](http://files.figshare.com/1587021/Ernest_Gordon_2014.pptx)

Ecological Society of America 2012 Meeting, Portland OR  
Elita Baldrige, Nathan Myrhvold, and Morgan Ernest. Macroecological life-history trait  
database for birds, mammals, and reptiles.  
[http://files.figshare.com/98378/BaldrigeESA2012\\_poster.pdf](http://files.figshare.com/98378/BaldrigeESA2012_poster.pdf)

Research & Creative Activities Week 2010, Fort Hays State University, Hays KS  
Elita Baldrige. Niche position, range position, and the distribution of abundance.

Research & Creative Activities Week 2009, Fort Hays State University, Hays KS  
Elita Baldrige. Nested Subsets: A pattern of community assembly.

#### SCIENCE COMMUNICATION

Guest post, Jabberwocky Ecology,<http://jabberwocky.weecology.org/> Elita Baldrige.  
How technology can help scientists with chronic illnesses (or Technology FTW!)  
Jabberwocky Ecology, 13 October 2014.  
<http://jabberwocky.weecology.org/2014/10/13/how-technology-can-help-scientists-with-chronic-illnesses-or-technology-ftw/>

Guest post, Jabberwocky Ecology,<http://jabberwocky.weecology.org/> Elita Baldrige. I  
am a graduate student. I have fibromyalgia. Jabberwocky Ecology, 13 January 2014.  
<http://jabberwocky.weecology.org/2014/01/13/i-am-a-graduate-student-i-have-fibromyalgia/>



Guest post, PhD(isabled), <http://phdisabled.wordpress.com/> Elita Baldrige. Developing a mystery illness in grad school. PhD(isabled), 5 November 2013.

<http://phdisabled.wordpress.com/2013/11/05/developing-a-mystery-illness-in-grad-school/>

Guest post, Jabberwocky Ecology, <http://jabberwocky.weecology.org/> Elita Baldrige. Four basic skills areas for a macroecologist. Jabberwocky Ecology, 19 August 2013.

<http://jabberwocky.weecology.org/2013/08/19/four-basic-skill-areas-for-a-macroecologist-guest-post/>

## DATA

MiscAbundanceDB: Community abundance data for a variety of mostly vertebrate taxa.

Available on figshare at the following links:

Metadata: <http://dx.doi.org/10.6084/m9.figshare.95841>

Sites table: <http://dx.doi.org/10.6084/m9.figshare.95842x>

Main data table: <http://dx.doi.org/10.6084/m9.figshare.95843x>

Citations table: <http://dx.doi.org/10.6084/m9.figshare.95844>

## CODE

GitHub repositories at <https://github.com/embaldrige>

## AWARDS & HONORS

Fleharty Fellowship, Department of Biological Sciences, Fort Hays State University, \$7,000

## SKILLS

### DATABASE

SQL- SQLite & postgres

### PROGRAMMING

Python (primary language)

Git

Markdown

Bash

Visual Basic

R