5-2014

Evaluating Process- and Constraint-Based Approaches for Modeling Macroecological Patterns

Xiao Xiao
Utah State University

Recommended Citation

Xiao, Xiao, "Evaluating Process- and Constraint-Based Approaches for Modeling Macroecological Patterns" (2014). *All Graduate Theses and Dissertations*. 3861.
https://digitalcommons.usu.edu/etd/3861

EVALUATING PROCESS- AND CONSTRAINT-BASED APPROACHES FOR MODELING MACROECOLOGICAL PATTERNS

by

Xiao Xiao

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

of

DOCTOR OF PHILOSOPHY

in

Ecology

Approved:

Ethan P. White Major Professor	Peter B. Adler Committee Member
S. K. Morgan Ernest Committee Member	Mevin B. Hooten Committee Member
James A. Powell Committee Member	Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2014

ABSTRACT

Evaluating Process- and Constraint-Based Approaches for Modeling Macroecological Patterns

by

Xiao Xiao, Doctor of Philosophy

Utah State University, 2014

Major Professor: Ethan P. White
Department: Biology

Ecological systems are characterized by a plethora of macroecological patterns observed across ecosystems and taxonomic groups. Explanations proposed for the patterns belong to two major categories – the process-based approaches, which characterize the structure of ecological systems by directly modeling a few key processes; and the constraint-based approaches, where macroecological patterns are viewed as emergent statistical properties that arise independent of the details of the processes.

Models that successfully reproduce the shape of a pattern do not necessarily reveal its true mechanism. Indeed, multiple models with different assumptions on mechanisms often make equivalent predictions for a single pattern. Patterns observed across systems regulated by different underlying processes are more likely to be statistical in nature. For example, here I show how one such pattern, Taylor’s Law (the power-law relationship between the mean and the variance of one or more populations), can be explained without invoking biological processes. Two inherent constraints on the system, the number of individuals and the number of groups they belong to, force most possible configurations of the system to match the empirically observed relationship.

While it is possible that some macroecological patterns are emergent statistical patterns, insights regarding mechanisms based on examining a single pattern are typically weak when there are multiple models that make similar predictions. Stronger inference is made possible by conducting tests that evaluate multiple patterns simultaneously. I evaluate the performance of two of the most comprehensive models of biodiversity and energy use: the size-structured neutral theory (SSNT; a process-based approach) and the Maximum Entropy Theory of Ecology (METE; a constraint-based approach). SSNT provides a better general characterization of empirical data, capturing the central tendency in the allocation of different-sized individuals among species (though with considerable scatter) while METE fails. The direct comparison between meaningful alternative models, instead of comparing single models to a null hypothesis, provides the strongest level of inference possible when comparing models to their empirical predictions. The results of this comparison suggest that demographic processes contain ecologically meaningful information not fully encapsulated in the current set of METE constraints, which help shape patterns of biodiversity and energy use.

(111 pages)

PUBLIC ABSTRACT

Evaluating Process- and Constraint-Based Approaches for Modeling Macroecological Patterns

by

Xiao Xiao, Doctor of Philosophy

Utah State University, 2014

Major Professor: Ethan P. White
Department: Biology

Macroecological patterns, such as the highly uneven distribution of individuals among species and the monotonic increase of species richness with area, exist across ecological systems despite major differences in the biology of different species and locations. These patterns capture the general structure of ecological communities, and allow relatively accurate predictions to be made with limited information for under-studied systems. This is particularly important given ongoing climate change and loss of biodiversity. Understanding the mechanisms behind these patterns has both scientific and practical merits.

I explore two conceptually different approaches that have been proposed as explanations for ecological patterns – the process-based approaches, which directly model key ecological processes such as birth, death, competition, and dispersal; and the constraint-based approaches, which view the patterns as the most likely state when the system is constrained in certain ways (e.g., the system has a fixed number of 100 individuals among five species, but the distribution may vary). While the process-based approaches directly link patterns to processes, the constraint-based approaches do not rely on the operation of specific processes and thus can be more broadly applied. I develop a new constraint-based approach to one of the most well established patterns in ecology, the power-law relationship between the mean and variance of a population. This pattern

has been widely observed and adopted as characterization of population stability. I find that the shape of the pattern can be well explained with two numerical constraints on the system, lending support to the idea that some macroecological patterns may not arise from specific processes but be statistical in nature instead.

I further examine the performance of the process- and constraint-based approaches for patterns of biodiversity and energy use, which are among the most essential as well as most well-studied aspects of community structure. Candidate models from both categories are able to partially capture the patterns across 60 globally distributed forest communities, however the process-based model is shown to provide a better general characterization of community structure than the constraint-base model in all communities. Thus the constraint-based approaches in their current forms do not fully encapsulate the effect of processes, which also contribute to the shape of the macroecological patterns of biodiversity and body size in addition to the constraints.

ACKNOWLEDGMENTS

I wish to thank my committee members, Peter Adler, Morgan Ernest, Mevin Hooten, and Jim Powell, for their support through my doctoral study. Special thanks go to my major professor and mentor, Ethan White, who has been most generous with his time, and provided tremendous help and guidance both in science and in many other aspects of academic life. Thank you all past and present members of the Weecology group, Ethan White, Morgan Ernest, Ken Locey, Glenda Yenni, Sarah Supp, Kate Thibault, Elita Baldrige, Zack Brym, Dan McGlinn, Ben Morris, Erica Christensen, and Kristina Reimer, for creating together a stimulating and supportive environment. I will always remember the good times we shared.

I also would like to acknowledge my collaborators, Ethan White, Ken Locey, Dan McGlinn, and James O’Dwyer, as well as many others in projects not included in my dissertation. I have learned a lot from each of you during our collaborations, and have greatly enjoyed the inspiring experience. My research is made possible by the work of numerous scientists, citizen scientists and volunteers who have contributed to data collection and curation for the various Long Term Ecological Research sites, the Center for Tropical Forest Science, PlotNet, SALVIAS, and many other institutes and organizations, as well as other researchers who have made their data publicly available. Chapter 2 of my dissertation was originally inspired by Joel Cohen. Chapter 3 has benefitted from the helpful comments from Ken Locey, John Harte, Erica Newman, Justin Kitzes, Sean Connolly, and two anonymous reviewers in the review process of the American Naturalist.

During my study at Utah State University, I have been financially supported by the university’s Diversity Fellowship, teaching assistantships from the Department of Biology, and research assistantships through a CAREER grant from the U.S. National Science Foundation awarded to Ethan White (DEB 0953694). The Ecology Center, the School of Graduate Studies,

and the Center for Women and Gender have generously provided additional support for travel. I am also grateful for the help from the staff of the Department of Biology, the Ecology Center, and the Office of Global Engagement at USU, who have made my life as an international graduate student as smooth as it could possibly be.

Last but certainly not least, I would like to express my gratitude for my friends and family, both near and far, for their support and encouragement through the years. Most of all I would like to thank my mom, who has never failed to support me with her understanding and unconditional love, despite being thousands of miles away.

Xiao Xiao

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Ecological systems are intrinsically complex (Darwin 1859). Individuals belonging to different taxonomic groups with varying life history traits go through demographic processes such as birth and death while interacting with each other through competition, predation, and mutualism. They are also influence by the environment through processes such as environmental filtering, environmental stochasticity, and the influence of the environment on demography and species interactions. Despite the different configurations different systems have in both their components and their associated processes, general patterns that are consistent across systems commonly occur (Brown 1995). These macroecological patterns, such as the highly uneven distribution of individuals among species (the species abundance distribution; Fisher et al. 1943; McGill et al. 2007) and the increase of species richness with area (the species-area relationship; Arrhenius 1921; Rosenzweig 1995), serve as general characterizations of community structure that can be universally applied across taxonomic groups and ecosystems.

Understanding the mechanisms underlying macroecological patterns is desirable both for advancing scientific knowledge (Brown 1999; Lawton 1999) and for extrapolating predictions of community structure to previously unobserved scenarios (McGill and Nekola 2010). Explanations proposed for these patterns belong primarily to two conceptual categories: one based on processes and the other on constraints.

Process-based approaches attempt to identify the few processes that are key for a given pattern, among the numerous processes simultaneously operating in a system. General patterns would result from processes that operate in a consistent way across systems. Examples in this category include the Theory of Island Biogeography (MacArthur and Wilson 1967), where species richness on islands is explained by the equilibrium between species dispersal and extinction, and the neutral theory (Hubbell 2001), where community-level patterns of biodiversity arise from individual-level processes of birth, death, speciation, and dispersal.

Constraint-based approaches do not directly model the biological processes, but instead view macroecological patterns as emergent statistical properties arising from specific ways that systems are constrained (e.g., Dewar and Porté 2008; Harte 2011; Locey and White 2013). The constraints are often descriptive statistics of the system (e.g., species richness, total abundance, etc.), and biological processes are assumed to act only indirectly on the patterns through their effects on the constraints. Free from assumptions about specific processes and their associated parameter values, constraint-based approaches can potentially be applied to any system, making them particularly suitable as explanations for patterns that are observed across systems governed by different processes.

I developed a new constraint-based approach to one of the most general patterns in ecology, Taylor's Law (Taylor 1961; Taylor and Woiwod 1980), which describes the power-law relationship between the mean and the variance of a population across space and time. This pattern has been widely documented both in ecological (e.g., Taylor et al. 1978; Taylor and Woiwod 1980) and non-ecological systems (e.g., Anderson and May 1988; Azevedo and Leroi 2001; de Menezes and Barabási 2004), strongly suggesting that its true mechanism cannot be processes specific to any field of study. The constraint-based approach thus serves as a more parsimonious explanation for the pattern than the numerous process-based models that have been proposed.

I also compared process-based and constraint-based theories for patterns of biodiversity and body size; two of the essential aspects of community structure. These patterns are some of the most well-studied of all macroecological patterns, and there are numerous existing models proposed to explain them (e.g., see McGill et al. 2007 for a review of over 20 models for the SAD). Since examination of a single pattern can rarely help distinguish models that make similar or identical predictions, I employed the stronger inference that can be gained by evaluating multiple predictions of a model simultaneously (McGill 2003; McGill et al. 2006).

I evaluated one constraint-based model, the Maximum Entropy Theory of Ecology (METE; Harte 2011), and one process-based model, the size-structured neutral theory (SSNT; O'Dwyer et al. 2009), to determine their ability to characterize ecological community structure. Both models attempt to capture the species abundance distribution as well as distributions of body size both within and across species, making them the most powerful models among the constraint- and process-based approaches for patterns of diversity and body size. By comparing the performance of METE and SSNT using data from 60 forest communities, my study represents the first attempt to scrutinize models of macroecological patterns by directly comparing two meaningful models simultaneously on multiple predictions using data from a diverse array of ecosystems. Such an evaluation not only assesses the ability of each theory to characterize community structure, but also provides a direct evaluation of the current status of the process- and constraint-based approaches.

References

Anderson, R. M., and R. M. May. 1988. Epidemiological parameters of HIV transmission. Nature 333:514–519.

Arrhenius, O. 1921. Species and area. Journal of Ecology 9:95–99.

Azevedo, R. B., and A. M. Leroi. 2001. A power law for cells. Proceedings of the National Academy of Sciences of the United States of America 98:5699–5704.

Brown, J. H. 1995. Macroecology. University Of Chicago Press, Chicago.

Brown, J. H. 1999. Macroecology: Progress and prospect. Oikos 87:3–14.

Darwin, C. 1859. On the Origin of Species. John Murray, London.

de Menezes, M. A., and A.-L. Barabási. 2004. Fluctuations in network dynamics. Physical Review Letters 92:028701–028704.

Dewar, R. C., and A. Porté. 2008. Statistical mechanics unifies different ecological patterns. Journal of Theoretical Biology 251:389–403.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 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 12:42–58.

Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. Oxford University Press, Oxford.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton.

Lawton, J. H. 1999. Are there general laws in ecology? Oikos 84:177–192.

Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. Ecology Letters 16:1177–1185.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton.

McGill, B. J. 2003. Strong and weak tests of macroecological theory. Oikos 102:679–685.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, 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.

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.

O’Dwyer, J. P., J. K. Lake, A. Ostling, V. M. Savage, and J. L. Green. 2009. An integrative framework for stochastic, size-structured community assembly. Proceedings of the National Academy of Sciences of the United States of America 106:6170–6175.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge.

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature 189:732–735.

Taylor, L. R., and I. P. Woiwod. 1980. Temporal stability as a density-dependent species characteristic. Journal of Animal Ecology 49:209–224.

Taylor, L. R., I. P. Woiwod, and J. N. Perry. 1978. The density-dependence of spatial behaviour and the rarity of randomness. Journal of Animal Ecology 47:383–406.

CHAPTER 2

A PROCESS-INDEPENDENT EXPLANATION FOR THE GENERAL FORM OF TAYLOR’S LAW *

Abstract

Taylor’s Law (TL) describes the scaling relationship between the mean and variance of one or more populations as a power-law. TL is widely observed in ecological systems across space and time with exponents varying largely between 1 and 2. Many ecological explanations have been proposed for TL but it is also commonly observed outside ecology. We propose that TL arises from the constraining influence of two primary variables: the number of individuals and the number of censuses or sites. We show that most possible configurations of individuals among censuses or sites produce the power-law form of TL with exponents between 1 and 2. This “feasible set” approach suggests that TL is a statistical pattern driven by two constraints; providing an *a priori* explanation for this ubiquitous pattern. However, the exact form of any specific mean-variance relationship cannot be predicted by the constraints, suggesting that TL may still contain ecological information.

*Coauthored by: Xiao, X., K. J. Locey, and E. P. White.

Introduction

One of the most basic goals of ecology is to understand how ecological systems change across scales. Scaling relationships such as Kleiber's Law (Kleiber 1932; Brown et al. 2004) and the species-area relationship (Rosenzweig 1995) have been intensively studied for decades and serve as both useful tools for extrapolation and empirical targets for ecological theories. One of the most general scaling relationships is Taylor's Law (TL; Taylor 1961). TL proposes that the relationship between the variance (s^2) and the mean density (m) of one or more populations is a power-law, which can be expressed mathematically as $s^2 = am^b$, where a and b are constants. The exponent b, which measures the magnitude of fluctuation with respect to the mean, is generally bounded between 1 and 2 (Taylor and Woiwod 1982). TL has been confirmed as an adequate description of population fluctuations both spatially (Taylor 1961; Taylor and Taylor 1977; Taylor et al. 1978; He and Gaston 2003; Kaltz et al. 2012) and temporally (Taylor and Woiwod 1980; Anderson et al. 1982; Perry 1994) across thousands of studies from a diverse array of taxonomic groups (Taylor et al. 1978; Taylor and Woiwod 1980; Taylor et al. 1983), making it one of the most widely-documented patterns in ecology.

A number of different models based on distinct ecological processes have been proposed to explain Taylor's Law. Explanations for the spatial TL include density-dependent population growth (Perry 1994), density-independent population growth (Cohen et al. 2013), random walks of individuals in space (Hanski 1980), and simultaneous attraction and repulsion among conspecific individuals (Taylor 1981a; Taylor 1981b). Similarly, the temporal TL has been argued to arise from environmental and demographic stochasticity (Ballantyne 2005; Ballantyne and Kerkhoff 2007), interspecific competition (Kilpatrick and Ives 2003), and even sampling error (Kalyuzhny et al. 2014). However, similar power-law mean-variance relationships with exponents between 1 and 2 have recently been documented in a number of non-ecological systems, ranging from the distribution of genes on a chromosome (Kendal 2003) and the number of cells in individuals (Azevedo and Leroi 2001) to fluctuations in the stock market (Eisler and Kertész 2006) and traffic flow (de Menezes and Barabási 2004) (see Eisler et al. 2008 for a review). This suggests that the mechanism underlying the pattern may not be specific to ecological systems but may instead be purely statistical.

An alternative explanation to process-based models is that TL describes the mean-variance scaling relationship of most possible states of a system where individuals are divided among groups (such as censuses or plots). This constraint-based view has been explored for another common pattern, the species-abundance distribution, using the set of all possible configurations of the pattern given some constraints (the feasible set; Locey and White 2013) to determine if the general shape of the pattern simply reflects that of the majority of possible outcomes. This kind of reasoning would suggest that TL is not generated by any particular set of processes, but emerges because many different combinations of processes result in the same general pattern (Harte 2011; White et al. 2012; Frank 2014).

To investigate this possibility for TL, we begin by recognizing that the form of TL is necessarily influenced by two values: the total number (quantity) of individuals (Q), and the number of groups (N; e.g., plots in spatial TL, censuses in temporal TL) among which those individuals are distributed, for each point in the mean-variance relationship. Given these constraints we ask whether most possible mean-variance relationships take a roughly power-law form. We adopt the concept of the feasible set (Haegeman and Loreau 2008; Locey and White 2013), which provides a general context under which the observed patterns can be examined. By comparing the empirical TLs to mean-variance relationships created from randomly sampling the set of all possible relationships constrained by Q and N, we find that both the form of the power-law and the exponent b between 1 and 2 are expected to occur for most possible configurations of the system; though the exact shape of each individual relationship cannot be accurately characterized without additional information.

Methods

1. The feasible set approach

The feasible set is the set of all possible configurations of a system (Haegeman and Loreau 2008; Locey and White 2013). This concept can be applied with different sets of constraints and configurations (e.g. ordered vectors of labeled individuals, unordered vectors of unlabeled individuals, etc.). Since each pair of (m_i, s_i^2) in a TL relationship results from distributing Q_i individuals into N_i groups (i.e., plots in the spatial TL and censuses in the temporal TL), we adopted (Q_i, N_i) as a minimal set of constraints that naturally defines the system. It follows that m_i is fully determined by Q_i and N_i as $m_i \equiv Q_i / N_i$. In contrast, the value of s_i^2 depends on how the individuals are distributed, with a minimal value of 0 (assuming that individuals can be evenly distributed among groups) and a maximal value of Q_i^2 / N_i (where all individuals are aggregated in a single group with zeroes everywhere else).

We used two different combinatorial approaches to define configurations that may give rise to TL – integer partitions, where each configuration is a unique set of unordered non-negative integers (expressed in non-increasing order), and integer compositions, where each configuration is a unique (i.e., ordered) vector of non-negative integers (Bona 2006). Technically, these configurations are referred to as “weak” because they allow zeros (Severs and White 2010). For example, the feasible set of integer partitions for $Q_i = 4$ and $N_i = 2$ is: (4, 0), (3, 1), (2, 2). Here, differently ordered configurations having the same integer values, e.g., (4, 0) and (0, 4), represent the same integer partition (4, 0) but different integer compositions. The difference between partitions and compositions is analogous to the difference between combinations and permutations, where each combination (unordered) can potentially arise from many permutations (ordered).

The relationship between partitions and compositions is also equivalent to that between microstates and macrostates in statistical mechanics and the application of the Maximum Entropy Principle (MaxEnt; Jaynes 2003) in ecology (e.g., Shipley et al. 2006; Dewar and Porté 2008; Harte 2011), where the most likely macrostate of a system is inferred as that with the greatest number of microstates. Unlike MaxEnt, the feasible set attempts to understand the full distribution of macrostates, rather than choosing the macrostate that is the most common (Locey and White 2013). However, as with predictions based on a maximum entropy framework, where the form of the prediction changes greatly with the decisions made about state variables, constraints, and prior distributions (Haegeman and Etienne 2010), the appearance of the feasible set (i.e. distribution of statistical features within it) will depend on the configuration used.

Using integer compositions as configurations to explore TL is equivalent to shifting the weights of the partitions, i.e. making some macrostates more likely to arise due to the differences in the number of compositions (microstates) each partition (macrostate) has. Rearranging the order of numbers within a configuration (e.g. partition, composition) does not change its variance, meaning that the variance of a composition is always equal to the variance of its corresponding partition. In the above example of $Q_i = 4$ and $N_i = 2$, the partition (4, 0) has the same frequency as (2, 2) in the feasible set of partitions, but twice the frequency in the feasible set of compositions where (0, 4) is also a unique configuration that corresponds to the partition (4, 0). Therefore the two applications of the feasible set approach may yield different results given the same set of constraints (Q_i, N_i). We examined the application of both partitions and compositions to TL because it is unclear whether it is best to focus solely on the macroscopic property of interest (i.e., variance), consider the number of different ways each macrostate can arise, or examine both to obtain a more complete context for understanding the constraining influence of Q and N.

2. Data

To explore whether the feasible set could generate realistic empirical patterns, we compiled TL relationships from the literature by surveying all papers that cited Taylor (1961) on Google Scholar to which we had access. We collected all empirical relationships that directly reported the values of (m_i, s_i^2) and (Q_i, N_i) pairs, or contained enough information for these values to be calculated. Due to the limited number of available non-ecological TL relationships, we focused exclusively on ecological TLs describing spatial or temporal fluctuations of population abundances. This approach to data compilation resulted in an imbalance between the spatial and the temporal TLs, with 90 spatial relationships and only four temporal ones. To offset this imbalance, we added a compilation of community-level time-series data (Yenni 2013) to our analysis, boosting the number of temporal TLs to 113 in total.

All relationships were put through additional screening before analysis. We removed (m_i, s_i^2) pairs where the corresponding N_i was less than 3 to ensure that the variance s_i^2 was properly defined among at least three numbers, as well as pairs where the corresponding Q_i was less than 5 so that the shape of TL would not be distorted by these zero-inflated, over-constrained configurations (Taylor and Woiwod 1982). We then excluded those relationships with less than five pairs of (m_i, s_i^2) remaining, leaving 73 spatial TLs and 106 temporal TLs. Due to the computationally intensive nature of the algorithm of generating partitions (Locey and White 2013), we further dropped TL relationships that contained any (Q_i, N_i) pairs that individually would take 2 hours or more to analyze (see 3. Analyses below). Overall our study encompassed 115 TL relationships, where 45 were spatial and 70 were temporal (see Appendix C for detailed characteristics of each relationship).

3. Analyses

In order to examine whether most possible mean-variance relationships exhibit TL like behavior, we need to define a range of scenarios to explore. To avoid selecting arbitrary, and potentially unrealistic, values of Q_i and N_i (and the distributions of those values within individual datasets), we followed Xiao et al. (2011) and used empirical data to define the range of values explored. We constructed the feasible set of partitions and the feasible set of compositions for each pair of (Q_i, N_i) in each empirical TL relationship that we compiled. Given the large size of the feasible set for large values of (Q_i, N_i) (Locey and White 2013), we drew 1000 random configurations from the feasible set in each case as a representation of the full set using the algorithms from Locey and McGlinn (2013). Each sampled configuration yielded one s_{ij}^2 (with j ranging from 1 to 1000), leading to 1000 s_{ij}^2's for each originally observed s_i^2, as well as 1000 sets of (m_i, s_{ij}^2) pairs for each empirical relationship. For each of these sets generated from the feasible set, we determined whether the relationship between m_i's and s_{ij}^2's was well characterized by a power-law with an exponent between 1 and 2 using OLS regression on log-transformed data, which is the standard approach to assessing TL relationships. We evaluated the significance and the value of the slope (which is an estimate of the exponent b), goodness-of-fit of the regression line (quantified with R^2), and deviation from the power-law (quantified as curvature on logarithmic scale, or the significance of a quadratic term fit to the relationship). We examined the 95% (2.5% to 97.5%) quantile of the 1000 s_{ij}^2's for each (Q_i, N_i) pair, as well as the full distribution of s_{ij}^2 values for a few randomly selected (Q_i, N_i) pairs, to see if the variance for configurations within the feasible set was aggregated with most configurations having similar variance values.

We then compared the mean-variance relationships obtained from sampling the feasible set to the empirical TL relationships with the same vectors of Q_i and N_i. We fit power-law relationships to each set of (m_i, s_i^2) (i.e., the empirical mean-variance relationship) using OLS regression on log-transformed data, and compared the fitted relationship to those obtained from fitting each set of (m_i, s_{ij}^2), both in goodness-of-fit and in the estimated value of the exponent b, which is often taken to be characteristic of a given species or system. Finally, we compared each empirical s_i^2 to its 1000 corresponding s_{ij}^2's from the feasible set. All comparative analyses were conducted separately for samples from the feasible set of partitions and those from the feasible set of compositions.

Results

Mean-variance relationships generated with randomly sampled configurations from the feasible set are well characterized by power-laws, with an average R^2 of 0.88 for relationships from the feasible set of partitions and 0.93 for those from the feasible set of compositions, though the latter show significant curvature (i.e., significant quadratic terms) in a non-negligible proportion of cases (27.8%, Table 2 – 1). The exponent b estimated for the simulated relationships largely falls between 1 and 2 (85.1% for partitions, 75.0% for compositions), consistent with results from empirical studies (e.g., Taylor and Woiwod 1982). See Table 2 – 1 for a summary of the results, and Fig. 2 – 1 for the full distributions of the statistics.

Examination of values of s_{ij}^2's for individual (Q_i, N_i) pairs shows that considerable variation exist for at least some combinations of Q_i and N_i, leading to distributions of s_{ij}^2's that are dispersed without a sharp peak where most configurations are aggregated (Appendix D, Fig. D – 1, top panel). This is particularly true for partitions. Consequently, the exponent b varies fairly broadly among the mean-variance relationships constructed for the same set of (Q_i, N_i) pairs (Fig. D – 1, bottom panel). Therefore, while most possible mean-variance relationships are well characterized by power-laws, the precise form of the power-law for a particular set of (Q_i, N_i) is not strongly constrained by the feasible set.

Most of the empirical datasets in our compilation have a significant power-law relationship between the mean and the variance, with little curvature, high R^2, and b largely between 1 and 2 (Table 2 – 1). Comparison between empirical TLs and those constructed from the feasible set shows general agreement both in fit (p-values and R^2) and in the values of b (Table 2 – 1, Fig. 2 – 1). Specifically, relationships constructed from the feasible set of partitions provide a surprisingly good match to empirical observations in the distribution of all statistics, while those from the feasible set of compositions exhibit deviations with higher proportions of significant curvature and exponents shifted towards higher values (Fig. 2 – 1). On the other hand, comparison between individual empirical s_i^2's and those calculated for configurations in the feasible set show that a high percentage of s_i^2's fall outside of the 95% quantile of the feasible set values for both spatial TLs (43.3% for partitions, 73.0% for compositions) and temporal TLs (26.5% for partitions, 41.9% for compositions) (Fig. 2 – 2, top panels). Similarly, the exponent *b* estimated for empirical TLs falls outside of the 95% quantile of those estimated from the feasible set in 40.0% of spatial TLs and 28.6% of temporal TLs for partitions (73.3% and 60.0% for compositions; Fig. 2 – 2, bottom panels). Consistent with the results for the distribution of variances and *b* values within the feasible set, these results suggest that neither feasible set approach is able to accurately characterize the variance of individual observations, or the exact shape of individual mean-variance relationships.

Discussion

Taylor's Law, or the power-law relationship between the mean and the variance of one or more populations, is a general pattern that has been widely observed in both ecological and non-ecological systems (Eisler et al. 2008). While numerous processes have been proposed as explanations for this pattern (e.g., Taylor 1981a; Kilpatrick and Ives 2003; Ballantyne and Kerkhoff 2007; Cohen et al. 2013), our study shows that TL can arise from numerical constraints on the system without explicitly or implicitly invoking processes. By sampling the set of all possible mean-variance scaling relationships we find that the form of the power-law relationship is exceptionally robust (Table 2 – 1, Fig. 2 – 1) with distributions of the exponent *b* closely matching those estimated from the empirical relationships (Fig. 2 – 1D), despite the relatively high observable variation for variance values among the configurations (Fig. 2 – 2, Fig. D – 1).

Thus the feasible set provides a general explanation for the ubiquity of TL in nature and helps explain why the pattern can be produced by models based on different underlying processes.

Our study is not the first to suggest that TL may be statistical in nature. By applying the Maximum Entropy Principle, Fronczak and Fronczak (2010) proposed that TL is the most likely macrostate associated with the largest number of microstates. However, their approach was later criticized for its reliance on physical quantities such as free energy and an external field, which lack analogues in biological systems (Kendal and Jørgensen 2011). Kendal and Jørgensen (2011) suggested instead that TL is associated with the Tweedie distribution family (Tweedie 1984), which by definition is characterized by a strict power-law relationship between the variance and the mean. While it has been argued that many statistical systems converge to distributions in the Tweedie family as limiting cases (Tweedie convergence theorem; Jørgensen et al. 1994), it is not clear how such convergence is achieved in nature.

Our approach shows that most of the possible macrostates for dividing individuals among groups result in TL-like mean-variance scaling relationships. This approach is unique in that it is built upon the concept of the feasible set, which can be unambiguously defined and applied to any system. With the feasible set approach, Locey and White (2013) showed that the majority of the configurations constrained by total abundance N and total species richness S conform to a hollow-curve similar in shape to empirically observed species-abundance distributions (SAD). Our study shows similar results for a second general pattern, TL, illustrating again that patterns can arise simply as the aggregated central tendency within a feasible set. This further emphasizes the importance of examining empirical patterns in terms of how primary constraints determine observable variation.

The feasible set approach is part of a more general framework to understanding macroecological patterns that has been emerging over the last decade. These approaches, which also include applications of the Maximum Entropy Principle, propose that some empirical patterns are emergent statistical properties constrained by numerical inputs (state variables), while the ecological processes operate only indirectly through their effects on the constraints themselves (Harte 2011; Supp et al. 2012; Locey and White 2013; Frank 2014). The constraints, which are usually descriptive statistics of the system (i.e., state variables or moments of the distributions), either strictly limit the possible configurations that a system can take ("hard" constraints; e.g., there are a limited number of ways to allocate 100 individuals into 5 groups) or limit the expected characteristics of the system ("soft" constraints"; e.g., 20 individuals are expected on average for each group but the observed values may vary; see Haegeman and Etienne 2010). In our study the system is hard-constrained by the total number of individuals observed (Q_i) and the number of groups they belong to (N_i), which forces the mean of configurations in the feasible set to match the mean of empirical data. The power-law relationship between the mean and the variance then arises for most possible configurations.

While the application of the feasible set approach is independent of assumptions about processes, the distribution of possible states of the system (in our case, variance) may shift with different definitions of what constitutes a unique configuration. This can be seen by the differences between the feasible set of partitions and the feasible set of compositions (Fig. 2 – 2, Fig. D – 1). Locey and White (2013) used partitions for the SAD with the goal of focusing only on the possible forms of the macrostates for a pattern that is characterized by an unordered list of species abundances with no information on species identity. In the case of TL, we have expanded our consideration by also considering integer compositions, which is the same as simultaneously choosing quantities of individuals at random to be assigned to specific sites or surveys. These two choices makes intuitive sense because we are interested in patterns that are related to how entire populations of individuals are distributed across space and through time, not the specific location of individual organisms. There are two other configurations that we did not consider, both of which focus on the distribution of individual organisms. One is the equivalent of randomly selecting the location of each of a set of known individuals among sites or years (known as surjections in combinatorics); e.g., the composition (1, 1) can be expanded into two configurations (A, B) and (B, A), where A and B represent labeled individuals. The other is the case where elements of the configuration are labeled but the order/arrangement of their bins is not accounted for; e.g., knowing that individual A and B are in the same site but not which site they are located in.

While there is a reasonable justification for choosing to focus on how populations are distributed rather than how individuals are distributed, both partition- and composition-based approaches satisfy this consideration. In equivalent discussions on how to set up maximum entropy based models, it has been argued that these choices must be made on the basis of comparisons to empirical data (Haegeman and Etienne 2010). Our study shows that the feasible set of partitions provides a more adequate characterization for our compilation of ecological TL relationships both in the fit of the power-law form (Table 2 – 1) and in the distribution of the exponent (Fig. 2 – 1D), however it remains to be seen if this statement holds more broadly, especially in non-ecological systems.

The constraint-based explanation for the general form of Taylor's Law is robust, with both feasible sets yielding mean-variance relationships that are well-characterized by power-laws, with high R^2 values and exponents falling within the range expected from empirical relationships. In combination with the observation of TL patterns in non-ecological systems, this suggests that the power-law form of TL as well as the general bound between 1 and 2 for the exponent may be explained with a simple set of constraints on Q_i and N_i. It remains to be determined which combinatorial approach is the most appropriate or whether this depends on the system. Moreover, a broad-stroke explanation, such as expecting empirical patterns to be generally similar to the majority of possible outcomes does not preclude the influence of other processes and constraints, be they ecological or statistical. Specifically, we find that neither the variance for a given (Q_i, N_i)

pair nor the exponent for a particular relationship can be accurately estimated by merely examining the central tendency of the feasible set defined by (Q_i, N_i) alone (Fig. 2 – 2). This suggests that more information is required to understand the detailed form of individual mean-variance relationships, either through direct operation of ecological processes (e.g., density-dependent growth, dispersal limitation, etc.) or through their indirect effects on additional constraints.

References

Anderson, R. M., D. M. Gordon, M. J. Crawley, and M. P. Hassell. 1982. Variability in the abundance of animal and plant species. Nature 296:245–248.

Azevedo, R. B., and A. M. Leroi. 2001. A power law for cells. Proceedings of the National Academy of Sciences of the United States of America 98:5699–5704.

Ballantyne, F. 2005. The upper limit for the exponent of Taylor’ s power law is a consequence of deterministic population growth. Evolutionary Ecology Research 7:1213–1220.

Ballantyne, F., and A. J. Kerkhoff. 2007. The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. Oikos 116:174–180.

Bona, M. 2006. A walk through combinatorics: an introduction to enumeration and graph theory (2nd ed.). World Scientific Publishing Company, New Jersey.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Towards a metabolic theory of ecology. Ecology 85:1771–1789.

Cohen, J. E., M. Xu, and W. S. F. Schuster. 2013. Stochastic multiplicative population growth predicts and interprets Taylor’s power law of fluctuation scaling. Proceedings of the Royal Society B: Biological Sciences 280:20122955.

de Menezes, M. A., and A.-L. Barabási. 2004. Fluctuations in network dynamics. Physical Review Letters 92:028701–028704.

Dewar, R. C., and A. Porté. 2008. Statistical mechanics unifies different ecological patterns. Journal of Theoretical Biology 251:389–403.

Eisler, Z., I. Bartos, and J. Kertész. 2008. Fluctuation scaling in complex systems: Taylor’s law and beyond. Advances in Physics 57:89–142.

Eisler, Z., and J. Kertész. 2006. Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. Physical Review E 73:046109.

Frank, S. A. 2014. Generative models versus underlying symmetries to explain biological pattern. Journal of Evolutionary Biology 27:1172–1178.

Fronczak, A., and P. Fronczak. 2010. Origins of Taylor's power law for fluctuation scaling in complex systems. Physical Review E 81:066112.

Haegeman, B., and R. S. Etienne. 2010. Entropy maximization and the spatial distribution of species. The American Naturalist 175:E74–E90.

Haegeman, B., and M. Loreau. 2008. Limitations of entropy maximization in ecology. Oikos 117:1700–1710.

Hanski, I. 1980. Spatial patterns and movements in Coprophagous beetles. Oikos 34:293–310.

Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. Oxford University Press, Oxford.

He, F., and K. J. Gaston. 2003. Occupancy, spatial variance, and the abundance of species. The American Naturalist 162:366–375.

Jaynes, E. T. 2003. Probability theory: the logic of science. Cambridge University Press, Cambridge.

Jørgensen, B., J. R. Martínez, and M. Tsao. 1994. Asymptotic behaviour of the variance function. Scandinavian Journal of Statistics 21:223–243.

Kaltz, O., P. Escobar-Páramo, M. E. Hochberg, and J. E. Cohen. 2012. Bacterial microcosms obey Taylor's law: effects of abiotic and biotic stress and genetics on mean and variance of population density. Ecological Processes 1:5.

Kalyuzhny, M., Y. Schreiber, R. Chocron, C. H. Flather, R. Kadmon, D. A. Kessler, and N. M. Shnerb. 2014. Temporal fluctuation scaling in populations and communities. Ecology 95:1701–1709.

Kendal, W. S. 2003. An exponential dispersion model for the distribution of human single nucleotide polymorphisms. Molecular Biology and Evolution 20:579–590.

Kendal, W. S., and B. Jørgensen. 2011. Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. Physical Review E 83:066115.

Kilpatrick, A. M., and A. R. Ives. 2003. Species interactions can explain Taylor's power law for ecological time series. Nature 422:65–68.

Kleiber, M. 1932. Body size and metabolism. Hilgardia 6:315–351.

Locey, K. J., and D. J. McGlinn. 2013. Efficient algorithms for sampling feasible sets of macroecological patterns. PeerJ PrePrints 1:e78v1.

Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. Ecology letters 16:1177–1185.

Perry, J. N. 1994. Chaotic dynamics can generate Taylor's power law. Proceedings of the Royal Society B: Biological Sciences 257:221–226.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge.

Severs, C., and J. A. White. 2010. On the homology of the real complement of the k-parabolic subspace arrangement. arXiv: 1012.3387 [math.CO].

Shipley, B., D. Vile, and E. Garnier. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. Science 314:812–814.

Supp, S. R., X. Xiao, S. K. M. Ernest, and E. P. White. 2012. An experimental test of the response of macroecological patterns to altered species interactions. Ecology 93:2505–2511.

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature 189:732–735.

Taylor, L. R., and R. A. J. Taylor. 1977. Aggregation, migration and population mechanics. Nature 265:415–421.

Taylor, L. R., R. A. J. Taylor, I. P. Woiwod, and J. N. Perry. 1983. Behavioural dynamics. Nature 303:801–804.

Taylor, L. R., and I. P. Woiwod. 1980. Temporal stability as a density-dependent species characteristic. Journal of Animal Ecology 49:209–224.

Taylor, L. R., and I. P. Woiwod. 1982. Comparative synoptic dynamics. I. Relationships between inter- and intra-specific spatial and temporal variance/mean population parameters. Journal of Animal Ecology 51:879–906.

Taylor, L. R., I. P. Woiwod, and J. N. Perry. 1978. The density-dependence of spatial behaviour and the rarity of randomness. Journal of Animal Ecology 47:383–406.

Taylor, R. A. J. 1981a. The behavioural basis of redistribution. II. Simulations of the Δ-model. Journal of Animal Ecology 50:587–604.

Taylor, R. A. J. 1981b. The behavioural basis of redistribution I. The Δ-model concept. Journal of Animal Ecology 50:573–586.

Tweedie, M. C. K. 1984. An index which distinguishes between some important exponential families. Pages 579–604 *in* J. K. Ghosh & J. Roy, eds., Statistics: Applications and New directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta.

White, E. P., K. M. Thibault, and X. Xiao. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93:1772–1778.

Xiao, X., E. P. White, M. B. Hooten, and S. L. Durham. 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. Ecology 92:1887–1894.

Yenni, G. M. 2013. Self-limitation as an explanation for species' relative abundances and the long-term persistence of rare species. All Graduate Theses and Dissertations. Paper 1958.

Table 2 – 1. Summary of the mean-variance relationships constructed from sampling the feasible sets of partitions or compositions, as well as empirical TL relationships.

	Significance at α = 0.05	Curvature at α = 0.05	Average R^2	Proportion of *b* between 1 and 2
Relationships from partitions	95.2%	8.49%	0.88	85.1%
Relationships from compositions	98.0%	27.8%	0.93	75.0%
Empirical TLs (spatial)	86.7%	11.1%	0.80	77.8%
Empirical TLs (temporal)	97.1%	4.29%	0.86	88.6%

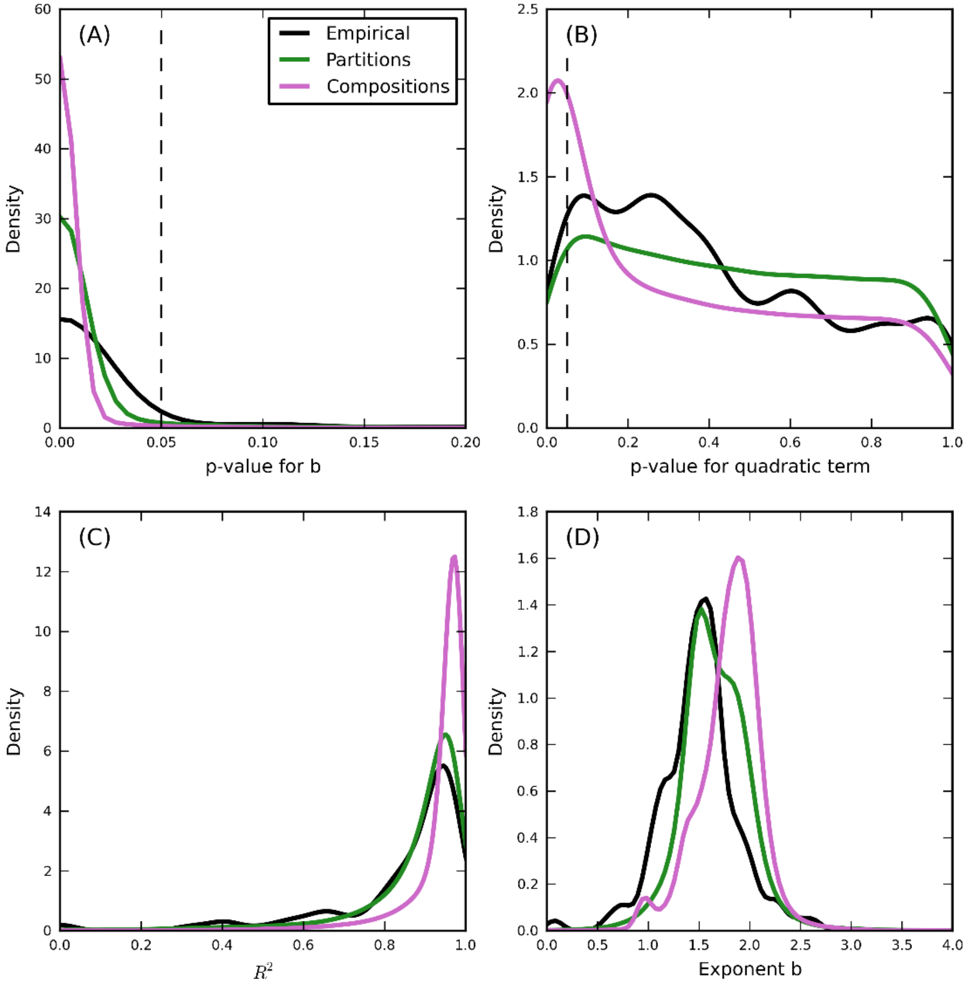


Figure 2 – 1. Density distribution of statistics. (A) p-values for the exponent *b* (B) p-values for the quadratic term (C) R^2 of the power-law relationship on logarithmic scale (D) values of the exponent *b*. Empirical results from the spatial TLs and the temporal TLs are not qualitatively different and are thus pooled. The dashed vertical lines in (A) and (B) represent the significance level of 0.05.

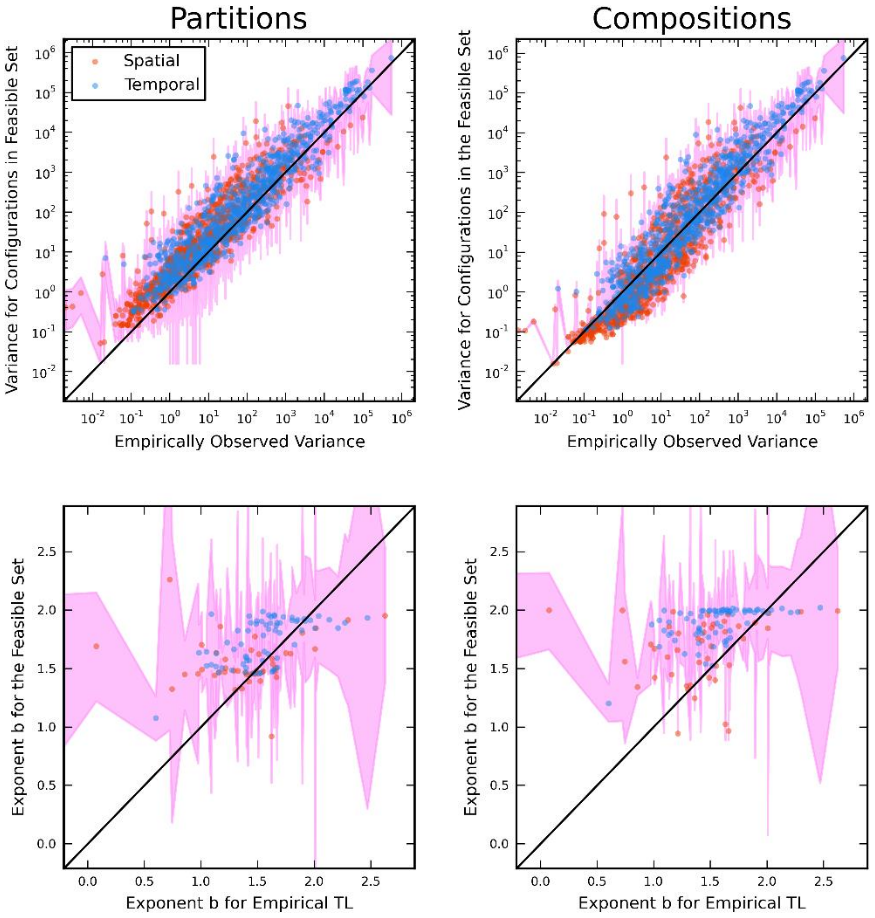


Figure 2 – 2. Observed values plotted against average values from the feasible sets. Top panels: The observed variance in each (m_i, s_i^2) pair from each empirical TL relationship plotted against the average value of the 1000 s_{ij}^2's. Bottom panels: The exponent *b* estimated for each empirical TL relationship, plotted against the average among the estimates of *b* for the 1000 sets of (m_i, s_{ij}^2) for each empirical relationship. Results are presented separately for partitions and for compositions. Pink shade in each subplot represents the 95% quantile. The diagonal lines are the 1-to-1 line. Note that in the top panels the lower (2.5%) quantile of some (m_i, s_i^2) pairs equals zero, which cannot be displayed on logarithmic scale and thus is replaced with the smallest value above zero among the s_{ij}^2's (0.01565) for visualition purpose.

CHAPTER 3

A STRONG TEST OF THE MAXIMUM ENTROPY THEORY OF ECOLOGY *

Abstract

The Maximum Entropy Theory of Ecology (METE) is a unified theory of biodiversity that predicts a large number of macroecological patterns using only information on the species richness, total abundance, and total metabolic rate of the community. We evaluated four major predictions of METE simultaneously at an unprecedented scale using data from 60 globally distributed forest communities including over 300,000 individuals and nearly 2000 species. METE successfully captured 96% and 89% of the variation in the rank distribution of species abundance and individual size, but performed poorly when characterizing the size-density relationship and intraspecific distribution of individual size. Specifically, METE predicted a negative correlation between size and species abundance, which is weak in natural communities. By evaluating multiple predictions with large quantities of data, our study not only identifies a mismatch between abundance and body size in METE, but also demonstrates the importance of conducting strong tests of ecological theories.

---

*Coauthored by: Xiao, X., D. J. McGlinn, and E. P. White. (2014) A strong test of the Maximum Entropy Theory of Ecology. The American Naturalist *in press*.

Introduction

The structure of ecological communities can be quantified using a variety of relationships, including many of the most well-studied patterns in ecology such as the distribution of individuals among species (the species abundance distribution or SAD), the increase of species richness with area (the species area relationship or SAR), and the distributions of energy consumption and body size (Brown 1995; Rosenzweig 1995; McGill et al. 2007; White et al. 2007). With the increasing consensus that these patterns are not fully independent, a growing number of unified theories have been proposed to identify links between the patterns and unite them under a single framework (e.g., Hanski and Gyllenberg 1997; Hubbell 2001; Harte 2011; see McGill 2010 for a review). Among these unified theories there are generally two different approaches, one based on processes and the other based on constraints. With the process-based approach, characteristics of the community are captured by explicitly modeling a few key ecological processes (e.g., Hanski and Gyllenberg 1997; Hubbell 2001). While this approach has the potential to directly establish connection between patterns and processes, it has been found that the same empirical patterns can result from different processes (Cohen 1968; Pielou 1975), and process-specific parameters are often hard to obtain (Hubbell 2001; Jones and Muller-Landau 2008). Alternatively, the constraint-based approach suggests that many macroecological patterns are emergent statistical properties arising from general constraints on the system, while processes are only indirectly incorporated through their effect on the constraints (e.g., Harte 2011; Locey and White 2013). This approach attempts to provide a general explanation of the observed patterns that does not rely on specific processes, which allows predictions to be made with little detailed information about the system.

One of the newest and most parsimonious constraint-based approaches is the Maximum Entropy Theory of Ecology (METE; Harte et al. 2008; Harte et al. 2009; Harte 2011). METE adopts the Maximum Entropy Principle from information theory, which identifies the most likely (least biased) state of a system given a set of constraints (Jaynes 2003). Assuming that the allocation of individuals and energy consumption within a community is constrained by three state variables (total species richness, total number of individuals, and total energy consumption), METE makes predictions for the SAD as well as multiple patterns related to energy use. Spatial patterns such as the SAR and the endemics area relationship can also be predicted with an additional constraint on the area sampled (Harte et al. 2008; Harte et al. 2009; Harte 2011). METE is one of the growing number of theoretical approaches that attempt to synthesize traditionally distinct areas of macroecology dealing with the distributions of individuals and the distributions of energy and biomass (Dewar and Porté 2008; Morlon et al. 2009; O'Dwyer et al. 2009), and thus provides a very general characterization of the structure of ecological systems. With no specific assumptions about biological processes, it can potentially be applied to any community where the values of the state variables can be obtained.

Previous studies have evaluated the performance of METE with separate datasets for different patterns and have shown that METE generally provides good characterizations of these patterns across geographical locations and taxonomic groups (Harte et al. 2008; Harte et al. 2009; Harte 2011; White et al. 2012a; McGlinn et al. 2013). However, these tests are relatively weak as they focus on one pattern at a time (McGill 2003). As a unified theory with multiple predictions, METE allows stronger tests to be made by testing the ability of the theory to characterize multiple patterns simultaneously for the same data (McGill 2003; McGill et al. 2006). In this study, we conduct a strong test of the non-spatial predictions of METE using data from 60 globally distributed forest communities to simultaneously evaluate four predictions of the theory (Fig. 3 – 1) including the SAD (the distribution of individuals among species) and energetic analogs of the individual size distribution (ISD; the distribution of body size among individuals regardless of their species identity) (Enquist and Niklas 2001; Muller-Landau et al. 2006), the size-density relationship (SDR; the correlation between species abundance and average individual size within species) (Cotgreave 1993), and the intraspecific individual size distribution (iISD; the distribution of body size among individuals within a species) (Gouws et al. 2011). Our analysis shows mixed support for METE across its four predictions, with METE successfully capturing the variation in some patterns while failing to do so for other. We discuss the ecological implications of our findings, as well as the importance of conducting strong multi-pattern tests in the evaluation of ecological theories.

Materials and Methods

1. Predicted patterns of METE

METE assumes that allocation of individuals and energy consumption within a community is constrained by three state variables: species richness (*S₀*), total number of individuals (*N₀*), and total metabolic rate summed over all individuals in the community (*E₀*) (Harte et al. 2008; Harte et al. 2009; Harte 2011). Define $R(n, \varepsilon)$ as the joint probability that a species randomly picked from the community has abundance n and an individual randomly picked from such a species has metabolic rate between (ε, $\varepsilon + \Delta\varepsilon$), two constraints are then established on the ratio between the state variables:

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot nR(n,\varepsilon) = \frac{N_0}{S_0} \qquad (1)$$

which represents the average abundance per species, and

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot n\varepsilon R(n,\varepsilon) = \frac{E_0}{S_0} \qquad (2)$$

which represents the average total metabolic rate per species. Note that the lower limit of individual metabolic rate is set to be 1, and all measures of metabolic rate are rescaled accordingly.

The forms of the four macroecological patterns that METE predicts can then be derived from $R(n, \varepsilon)$ (see Harte 2011 and Appendix E for detailed derivation) and are given by the following four equations. SAD takes the form

$$\Phi(n) \approx \frac{1}{Cn} e^{-(\lambda_1+\lambda_2)n} \qquad (3)$$

which is an upper-truncated Fisher's log-series distribution. Here λ_1 and λ_2 are Lagrange multipliers obtained by applying the Maximum Entropy Principle with respect to the constraints, and C is the proper normalization constant. The Individual-level Energy Distribution (which is the energetic equivalent of the ISD) takes the form

$$\Psi(\varepsilon) = \frac{S_0}{N_0 Z} \cdot \frac{e^{-\gamma}}{(1-e^{-\gamma})^2} \cdot \left(1 - (N_0+1)e^{-\gamma N_0} + N_0 e^{-\gamma(N_0+1)}\right) \qquad (4)$$

where $\gamma = \lambda_1 + \lambda_2 \cdot \varepsilon$. Conditioned on abundance n, the Species-level Energy Distribution (which is the energetic equivalent of the iISD) is given by

$$\Theta(\varepsilon|n) = \frac{n\lambda_2 e^{-\lambda_2 n\varepsilon}}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_0}} \qquad (5)$$

which is an exponential distribution with parameter $\lambda_2 n$. The expected value of the iISD $\Theta(\varepsilon|n)$ then gives the Average Species Energy Distribution (which is the energetic equivalent of the SDR), i.e., the expected average metabolic rate (size) for individuals within a species with abundance n:

$$\bar{\varepsilon}(n) = \frac{1}{n\lambda_2(e^{-\lambda_2 n} - e^{-\lambda_2 n E_0})} \cdot \left[e^{-\lambda_2 n}(\lambda_2 n + 1) - e^{-\lambda_2 n E_0}(\lambda_2 n E_0 + 1)\right] \qquad (6)$$

It should be noted that this derivation shows that the iISD and the SDR are closely related to one another since the SDR is the expectation of the iISD. As a result, the two patterns are expected to yield similar fits to the theory and provide similar insights into its performance.

2. Data

METE predicts the iISD to be an exponential distribution (Eqn 5; also see Fig. 3 – 1D) where the smallest size class is the most abundant, regardless of species identity or abundance. However, most animal species exhibit interior modes of adult body size (e.g., Koons et al. 2009; Gouws et al. 2011; but see Harte 2011) and large variation in minimum (and maximum) body size among species associated with these modal values (Gouws et al. 2011). In other words, the body sizes of conspecifics are clustered around some intermediate value, while individuals that are much larger or smaller are rare. Consequently, assembling all individuals across species in such communities often yields multimodal ISD (Thibault et al. 2011), as opposed to the monotonically decreasing relationship predicted by METE (Eqn 4; also see Fig. 3 – 1B). Thus animal communities are expected a priori to violate two of the predictions of METE. Therefore, to ensure that the performance of METE was not trivially rejected because of the life history trait of determinate growth, in our analysis we focused exclusively on trees, which are known to have iISDs (Condit et al. 1998b) and ISDs (Enquist and Niklas 2001; Muller-Landau et al. 2006) that are well characterized by monotonically declining distributions and which arguably have the greatest prevalence of high quality individual level size data among indeterminately growing taxonomic groups.

We compiled forest plot data from previous publications, publicly available databases, and data obtained through personal communication (Table 3 – 1). All plots have been fully surveyed with size measurement for all individuals above plot-specific minimum thresholds. For those plots where surveys have been conducted multiple times, we adopted data from the most recent one unless otherwise specified (see Table 3 – 1). We excluded records of ferns, palms, and herbs, if they existed. Individuals that were dead, not identified to species/morphospecies, and/or missing size measurements were excluded. Individuals with size measurements below or equal to the designated minimum thresholds were excluded as well, because it is unclear whether these size classes were thoroughly surveyed. Overall our analysis encompassed 60 plots that were at least 1 ha in size and had a richness of at least 14 (Table 3 – 1), with 1943 species/morphospecies and 379022 individuals in total.

3. Analyses

The scaling relationship between diameter and metabolic rate can be described with good approximation by metabolic theory as $B \propto D^2 \cdot e^{-E/kT}$, where B is metabolic rate, D is diameter, T is temperature, E is the activation energy, and k is the Boltzmann's constant (West et al. 1999; Gillooly et al. 2001). Assuming that E is constant across species and T is constant within a community, the temperature-dependent term $e^{-E/kT}$ is constant within a community, and can be dropped when the metabolic rate of individuals are rescaled. We thus used $(D/D_{min})^2$ as the surrogate for individual metabolic rate, where D_{min} is the diameter of the smallest individual in the community, which sets the minimal individual metabolic rate to be 1 following METE's assumption (see Eqn 2). For individuals with multiple stems, we adopted the pipe model to combine the records, i.e., $D = \sqrt{\sum d_i^2}$, where d_i's were diameter of individual stems (Ernest et al. 2009). Since metabolic rate scales as D^2, the pipe model preserves the total area as well as the total metabolic rate for all stems combined.

We obtained the Lagrange multipliers λ_1 and λ_2 in each community with inputs S_0, N_0, and E_0 (i.e., the sum over the rescaled individual metabolic rates) (see Appendix E). Predictions for the four ecological patterns were obtained from Eqns 3-6 and further transformed to facilitate comparison with observations. For the SAD and the ISD, we converted the predicted probability distributions (Eqns 3 & 4) to rank distributions of abundance (i.e., abundance at each rank from the most abundant species to the least abundant species) and size (i.e., scaled metabolic rate at each rank from the largest individual to the smallest individual across all species) (Harte et al. 2008; Harte 2011; White et al. 2012a), which were compared with the empirical rank distributions of abundance and size. For the SDR, predicted average metabolic rate was obtained from Eqn 6 for species with abundance n, which was compared to the observed average metabolic rate for that species. For the iISD, we converted the predicted exponential distribution (Eqn 5) into a rank distribution of individual size for each species, and compared the scaled metabolic rate predicted at each rank to the observed value.

The explanatory power of METE for each pattern was quantified using the coefficient of determination R^2, which was calculated as

$$R^2 = 1 - \sum_i [\log_{10}(obs_i) - \log_{10}(pred_i)]^2 \,/\, \sum_i [\log_{10}(obs_i) - \overline{\log_{10}(obs_i)}]^2 \tag{7}$$

where obs_i and $pred_i$ were the ith observed value and METE's prediction, respectively. Both observed and predicted values were log-transformed for homoscedasticity. Note that R^2 measures the proportion of variation in the observation explained by the prediction; it is based on the 1:1 line when the observed values are plotted against the predicted values, not the regression line. Thus it is possible for R^2 to be negative, which is an indication that the prediction is worse than taking the average of the observation.

While R^2 between predicted and observed values provides an intuitive measure of the predictive power of the theory, it ignores the variation that can arise from random sampling even when the predicted distribution is valid. To address this issue, we conducted a bootstrap analysis, where we drew 500 random samples from the predicted distribution for each pattern (Eqn 3 for SAD, Eqn 4 for ISD, Eqn 5 for SDR and iISD), and examined the fit of the theory to the bootstrap samples using both R^2 and the Kolmogorov-Smirnov (K-S) statistic (see Appendix F for details). If METE fits the empirical data as well as it fits the bootstrap samples, then the theory matches the data and the residual variation is consistent with random sampling. If instead METE fits the bootstrap samples better than the empirical data, it indicates that there are meaningful deviations of empirical data from the theory's predictions. By comparing the fits to empirical data to those from data simulated from the theory, this analysis provides additional insights into patterns like the SAD that are expected to be well fit by many theories (Connolly et al. 2009; Locey and White 2013) and into patterns like the iISD where large amounts of variation about the predicted values may be expected due to sampling.

Python code to replicate our analyses together with a processed subset of datasets is deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.5fn46 (Xiao et al. 2014). Data included in the deposit are specifically designed for the replication of our analyses, and may lack spatial/temporal components or other useful information in the original data. Readers interested in using the data for purposes other than replicating our analyses are advised to obtain the raw data from the original sources.

Results

The results for all forest plots combined are summarized in Fig. 4 – 2, with observations plotted against predictions for each macroecological pattern. METE provides excellent predictions for the SAD ($R^2 = 0.96$) and the ISD ($R^2 = 0.89$), though the largest size classes deviate slightly but consistently in the ISD. However, the SDR ($R^2 = -2.24$) and the iISD ($R^2 = 0.15$) are not well characterized by the theory.

Further examination of the four macroecological patterns within each community confirms METE's ability to consistently characterize the SAD (all R^2 values > 0.60, 59/60 R^2 values > 0.8) and the ISD (all R^2 values > 0.48, 49/60 R^2 values > 0.8), as well as its inadequacy in characterizing the SDR (all R^2 values below zero) and the iISD (maximal $R^2 = 0.30$, 49/60 R^2 values < 0). Results from bootstrap analysis (Appendix F) are also largely consistent with the direct interpretation of the goodness of fit statistics. METE provides comparable characterization for the empirical and the bootstrap SADs in most communities, while its fit is consistently worse to the empirical SDRs and the iISDs than to the bootstrap samples (Fig. F – 2). For the ISD, however, the analysis reveals that METE characterizes bootstrap samples consistently better than its fit to empirical data (Fig. F – 2), which implies that the empirical ISD significantly deviates from METE's prediction despite the theory's ability to capture the general shape of the pattern (Fig. 4 – 2B). This is consistent with model comparison in Appendix G where we show that alternative models provide a better fit to the distribution (Table G – 1).

Discussion

Macroecological theories increasingly attempt to make predictions across numerous ecological patterns (McGill 2010), by either directly modeling ecological processes or imposing constraints on the system. Among the constraint-based theories, METE is unique in that it makes simultaneous predictions for two distinct sets of ecological patterns, synthesizing traditionally separate areas of macroecology dealing with distributions of individuals and distributions related to body size and energy use (see also Dewar and Porté 2008; Morlon et al. 2009; O'Dwyer et al. 2009). Using only information on the species richness, total abundance, and total energy use as inputs, METE attempts to characterize various aspects of community structure without additional tunable parameters or assumptions, making it one of the most parsimonious of the current unified theories.

Our analysis shows that METE accurately captures the general shape of the SAD (allocation of individuals among species) and ISD (allocation of energy/biomass among individuals) within and among 60 forest communities (Fig. 4 – 2A, B). The SAD and the ISD are among the most well-studied patterns in ecology, and numerous models exist for both patterns. For instance, with metabolic theory and demographic equilibrium models, Muller-Landau et al. (2006) identified four possible predictions for the ISD under different assumptions of growth and mortality rates. For the SAD more than twenty models have been proposed (Marquet et al. 2003; McGill et al. 2007), ranging from purely statistical to mechanistic.

Our study demonstrates METE's high predictive power for these two patterns, but it does not imply that it is the best model when each pattern is considered independently. Indeed, our results reveal a consistent departure of individuals in the largest size class from the ISD predicted by METE (Fig. 4 – 2B), which may result from mortality unrelated to energy use (Muller-Landau et al. 2006). Bootstrap analysis (Appendix F) further confirms that such deviation is more severe than expected from the effect of random sampling alone. The discrepancy between the high R^2 of the ISD both within and across communities, and the seemingly poor fit of the pattern revealed by bootstrapping, results from the two different ways that goodness of fit is evaluated by the two analyses. While METE is able to predict the rank size of individuals (Fig. 4 – 2B) with high accuracy (illustrated by the high R^2 between predicted and observed values), the empirical ISDs are still significantly different from the predicted distribution (illustrated by higher deviation of empirical data from the predicted form when compared to bootstrap samples). Indeed, while METE has been shown to frequently outperform the most common model of the SAD (the lognormal) for a variety of taxonomic groups including plants (White et al. 2012a), model comparisons for the ISD using AIC suggest that the maximum likelihood Weibull distribution (one of the distributions for tree diameter in Muller-Landau et al. 2006) almost always outperforms METE (though METE's performance is comparable to that of the other two distributions, the exponential and the Pareto; see Appendix G). Quantitatively comparing theories that make multiple predictions is challenging and there is no general approach for properly comparing models that make different numbers of predictions. When comparing general theories to single prediction models with multiple tunable parameters it is not surprising that theories such as METE fail to provide the best quantitative fit (White et al. 2012b). However, as a constraint-based unified theory, METE's strength lies in its ability to link together ecological phenomena that were previously considered distinct, and to make predictions based on first principles with minimal inputs. The general agreement between METE's predictions and the observed SAD and ISD (as measured by the R^2 for the rank distributions) supports the notion that the majority of variation in these macroecological patterns can be characterized by variation in the state variables S_0, N_0, and E_0 alone (Harte 2011; Supp et al. 2012; White et al. 2012a).

While METE performs well in characterizing the SAD and ISD, it performs poorly when predicting the distribution of energy at the species level (Fig. 4 – 2C, D). This is not that surprising given that the iISD and the SDR (which is the expectation of the iISD) provide a more detailed perspective on the community structure by examining the intercorrelation of abundance and size. The deviations of the empirical patterns from the predictions reveal a mismatch between the predicted metabolic rate of individuals and their species' abundances. METE predicts a monotonically decreasing relationship between species abundance and average intraspecific metabolic rate, i.e., species with higher abundance are also smaller in size on average and are more likely to contain smaller individuals (Eqns 5, 6, Fig. 4 – 1C). Evaluating the total (instead of average) intraspecific metabolic rate, this relationship translates roughly into Damuth's energetic equivalence rule (Damuth 1981), where the total energy consumption within a species does not depend on species identity or abundance (Harte et al. 2008; Harte 2011). While Damuth's rule has been argued to apply at global scales (Damuth 1981; White et al. 2007), our results indicate that it does not hold locally, in concordance with a number of previous studies (Brown and Maurer 1987; Blackburn and Gaston 1997; White et al. 2007).

The consistency of our results across 60 forest communities (as well as confirmative evidence from a concurrent study of a single herbaceous plant community; Newman et al. 2014) provides strong evidence for METE's mixed performance among the four macroecological patterns. However, several limitations of the study are worth noting. First, we only analyzed a single taxonomic group (trees). This was in part because individual level size data collected in standardized ways is available for a large number of tree communities, and in part based on a prior knowledge that the form of the ISD and the iISD (Condit et al. 1998; Enquist and Niklas 2001; Muller-Landau et al. 2006) had a reasonable chance of being well characterized by the theory (see Methods). While we know that the SAD predictions of the theory perform well in general (White et al. 2012a), further tests are necessary to determine if the simultaneous good fit of the ISD predictions is supported in other taxonomic groups. There is some evidence that this result holds in invertebrate communities (Harte 2011). Second, we estimated the metabolic rate of individuals based on predictions of metabolic theory rather than direct measurement. It is possible that directly measured metabolic rates could result in different fits to the theory (but see Newman et al. 2014, which adopts a different method to obtain metabolic rate yet reaches similar conclusions).

Models and theories can be evaluated at multiple levels which yield different strengths of inference (McGill 2003; McGill et al. 2006), progressing from matching theory to empirical observations on a single pattern, to testing against a null hypothesis, to evaluating multiple *a priori* predictions, to eventually comparing between multiple competing models. With quantitative predictions on various ecological patterns, METE and other unified theories allow for simultaneous examination of multiple predictions, which provides a much stronger test compared to curve-fitting for a single pattern and can often reveal important insight into theories that are otherwise overlooked by single pattern tests (e.g., Adler 2004). As a comprehensive analysis on the performance of METE in predicting abundance and energy distributions in the same datasets, our study demonstrates the importance of moving towards stronger tests in ecology, especially when multiple intercorrelated predictions are available; while previous studies have shown that METE does an impressive job characterizing a single pattern (White et al. 2012a; McGlinn et al. 2013), concurrently evaluating all predictions of the theory identifies a slight yet consistent discrepancy between the observed and the predicted size distribution, as well as a mismatch between species' abundance and individual size.

The fact that METE fails to provide good characterization of all four patterns of community structure and performs more poorly than alternative models in some cases can be interpreted in two ways. First, the aspects of community structure that are poorly characterized by the theory may be more adequately characterized by explicitly modeling ecological processes. For example, O'Dwyer et al. (2009) has developed a model that incorporates individual demographic rates of birth, death, and growth, which likewise yields predictions of abundance and body size distributions. It is worth noting, however, that the process-based approach and the constraint-based approach do not have to be mutually exclusive. While O’Dwyer et al. (2009) suggested that size-related patterns may reflect ecological processes, the agreement between their model and METE in the predicted SAD (both log-series), as well as METE’s performance for the ISD, support the idea that information in the underlying processes can be summarized in constraints alone for some macroecological patterns. Alternatively, the constraint-based approach may be sufficient in characterizing patterns of abundance and of body size, but the current form of METE may be incorrect. Specifically, the limitations revealed in our analyses may be remedied by either relaxing the current constraints to remove the implicit negative correlation between species-level average body size and abundance (Fig. 4 – 1C) from the theory, or by adding additional constraints to the system so that energetic equivalence among species no longer holds (Harte and Newman 2014). While the success of METE in characterizing the general shape of the SAD and the ISD adds to the growing support for the constraint-based approach for studying macroecological patterns, further work is clearly needed to develop unified theories for community structure whether they are based on specific biological processes or emergent statistical properties.

References

Adler, P. B. 2004. Neutral models fail to reproduce observed species-area and species-time relationships in Kansas grasslands. Ecology 85:1265–1272.

Baribault, T. W., R. K. Kobe, and A. O. Finley. 2011. Data from: tropical tree growth is correlated with soil phosphorus, potassium, and calcium, though not for legumes. Ecological Monographs. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.r9p70.

Baribault, T. W., R. K. Kobe, and A. O. Finley. 2012. Tropical tree growth is correlated with soil phosphorus, potassium, and calcium, though not for legumes. Ecological Monographs 82:189–203

Blackburn, T. M., and K. J. Gaston. 1997. A critical assessment of the form of the interspecific relationship between abundance and body size in animals. The Journal of Animal Ecology 66:233–249.

Brown, J. H. 1995. Macroecology. University of Chicago Press, Chicago.

Brown, J. H., and B. A. Maurer. 1987. Evolution of species assemblages: Effects of energetic constraints and species dynamics on the diversification of the North American avifauna. The American Naturalist 130:1–17.

Cohen, J. E. 1968. Alternate derivations of a species-abundance relation. The American Naturalist 102:165–172.

Condit, R. 1998a. Tropical forest census plots. Springer-Verlag and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas.

Condit, R. 1998b. Ecological implications of changes in drought patterns: shifts in forest composition in Panama. Climatic Change 39:413–427.

Condit, R., S. Aguilar, A. Hernández, R. Pérez, S. Lao, G. Angehr, S. P. Hubbell, et al. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. Journal of Tropical Ecology 20:51–72.

Condit, R., R. Sukumar, S. P. Hubbell, and R. B. Foster. 1998. Predicting population trends from size distributions: a direct test in a tropical tree community. The American Naturalist 152:495–509.

Connolly, S. R., M. Dornelas, D. R. Bellwood, and T. P. Hughes. 2009. Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. Ecology 90:3138–3149.

Cotgreave, P. 1993. The relationship between body size and population abundance in animals. Trends in Ecology & Evolution 8:244–248.

Damuth, J. 1981. Population density and body size in mammals. Nature 290:699–700.

DeWalt, S. J., G. Bourdy, L. R. ChÁvez de Michel, and C. Quenevo. 1999. Ethnobotany of the Tacana: Quantitative inventories of two permanent plots of Northwestern Bolivia. Economic Botany 53:237–260.

Dewar, R. C., and A. Porté. 2008. Statistical mechanics unifies different ecological patterns. Journal of Theoretical Biology 251:389–403.

Enquist, B. J., and K. J. Niklas. 2001. Invariant scaling relations across tree-dominated communities. Nature 410:655–660.

Ernest, S. K. M., E. P. White, and J. H. Brown. 2009. Changes in a tropical forest support metabolic zero-sum dynamics. Ecology Letters 12:507–515.

Gilbert, G. S., E. Howard, B. Ayala-Orozco, M. Bonilla-Moheno, J. Cummings, S. Langridge, I.

M. Parker, et al. 2010. Beyond the tropics: forest structure in a temperate forest mapped plot. Journal of Vegetation Science 21:388–405.

Gillooly, J. F., J. H. Brown, G. B. West, V. M. Savage, and E. L. Charnov. 2001. Effects of size and temperature on metabolic rate. Science 293:2248–2251.

Gouws, E. J., K. J. Gaston, and S. L. Chown. 2011. Intraspecific body size frequency distributions of insects. PLoS ONE 6:e16606.

Hanski, I., and M. Gyllenberg. 1997. Uniting two general patterns in the distribution of species. Science 275:397–400.

Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. Oxford University Press, Oxford.

Harte, J., and E. A. Newman. 2014. Maximum information entropy: a foundation for ecological theory. Trends in Ecology & Evolution 29:384–389.

Harte, J., A. B. Smith, and D. Storch. 2009. Biodiversity scales from plots to biomes with a universal species-area curve. Ecology Letters 12:789–97.

Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 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. Princeton University Press, Princeton.

Hubbell, S. P., R. Condit, and R. B. Foster. 2005. Barro Colorado forest census plot data.

Hubbell, S. P., R. B. Foster, S. T. O’Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, et al. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science 283:554–557.

Jaynes, E. T. 2003. Probability theory: the logic of science. (G. L. Bretthorst, ed.). Cambridge University Press, Cambridge.

Jones, F. A., and H. C. Muller-Landau. 2008. Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. Journal of Ecology 96:642–652.

Kohyama, T., E. Suzuki, T. Partomihardjo, and T. Yamada. 2001. Dynamic steady state of patch-mosaic tree size structure of a mixed dipterocarp forest regulated by local crowding. Ecological Research 16:85–98.

Kohyama, T., E. Suzuki, T. Partomihardjo, T. Yamada, and T. Kubo. 2003. Tree species differentiation in growth, recruitment and allometry in relation to maximum height in a Bornean mixed dipterocarp forest. Journal of Ecology 91:797–806.

Koons, D. N., R. D. Birkhead, S. M. Boback, M. I. Williams, and M. P. Greene. 2009. The effect of body size on cottonmouth (Agkistrodon piscivorus) survival, recapture probability, and behavior in an Alabama swamp. Herpetological Conservation and Biology 4:221–235.

Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. Ecology Letters 16:1177-1185.

Lopez-Gonzalez, G., S. L. Lewis, M. Burkitt, T. R. Baker, and O. L. Phillips. 2009. ForestPlots.net Database. www.forestplots.net. Date of extraction [06, 07, 2012].

Lopez-Gonzalez, G., S. L. Lewis, M. Burkitt, and O. L. Phillips. 2011. ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. Journal of Vegetation Science 22:610–613.

Marquet, P. A., J. A. Keymer, and H. Cofre. 2003. Breaking the stick in space: of niche models, metacommunities and patterns in the relative abundance of species. Pages 64–86 *in* T. M. Blackburn & K. J. Gaston, eds., Macroecology: Concepts and Consequences. Blackwell Science, Oxford, UK.

McDonald, R. I., R. K. Peet, and D. L. Urban. 2002. Environmental correlates of oak decline and red maple increase in the North Carolina piedmont. Castanea 67:84–95.

McGill, B. J. 2003. Strong and weak tests of macroecological theory. Oikos 102:679–685.

McGill, B. J. 2010. Towards a unification of unified theories of biodiversity. Ecology Letters 13:627–642.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, 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.

McGlinn, D. J., X. Xiao, and E. P. White. 2013. An empirical evaluation of four variants of a universal species–area relationship. PeerJ 1:e212.

Morlon, H., E. P. White, R. S. Etienne, J. L. Green, A. Ostling, D. Alonso, B. J. Enquist, et al. 2009. Taking species abundance distributions beyond individuals. Ecology letters 12:488–501.

Muller-Landau, H. C., R. S. Condit, K. E. Harms, C. O. Marks, S. C. Thomas, S. Bunyavejchewin, G. Chuyong, et al. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. Ecology Letters 9:589–602.

Nakashizuka, T., M. Saito, K. Matsui, A. Makita, T. Kambayashi, T. Masaki, T. Nagaike, et al. 2003. Monitoring beech (Fagus crenata) forests of different structure in Shirakami Mountains. Tohoku Journal of Forest Science 8:67–74.

Newman, E. A., M. E. Harte, N. Lowell, M. Wilber, and J. Harte. 2014. Empirical tests of within- and across-species energetics in a diverse plant community. Ecology 95:2815-2825.

Nishimura, T. B., and E. Suzuki. 2001. Allometric differentiation among tropical tree seedlings in heath and peat-swamp forests. Journal of Tropical Ecology 17:667–681.

Nishimura, T. B., E. Suzuki, T. Kohyama, and S. Tsuyuzaki. 2006. Mortality and growth of trees in peat-swamp and heath forests in central Kalimantan after severe drought. Plant Ecology 188:165–177.

O'Dwyer, J. P., J. K. Lake, A. Ostling, V. M. Savage, and J. L. Green. 2009. An integrative framework for stochastic, size-structured community assembly. Proceedings of the National Academy of Sciences of the United States of America 106:6170–6175.

Palmer, M. W., R. K. Peet, R. A. Reed, W. Xi, and P. S. White. 2007. A multiscale study of vascular plants in a North Carolia Piedmont forest. Ecology 88:2674–2674.

Peet, R. K., and N. L. Christensen. 1987. Competition and tree death. BioScience 37:586–595.

Pielou, E. C. 1975. Ecological diversity. Wiley, New York.

Pitman, N. C. A., C. E. Cerón, C. I. Reyes, M. Thurber, and J. Arellano. 2005. Catastrophic natural origin of a species-poor tree community in the world's richest forest. Journal of Tropical Ecology 21:559–568.

Pyke, C. R., R. Condit, S. Aguilar, and S. Lao. 2001. Floristic composition across a climatic gradient in a neotropical lowland forest. Journal of Vegetation Science 12:553–566.

Ramesh, B. R., M. H. Swaminath, S. V. Patil, R. Pélissier, P. D. Venugopal, S. Aravajy, C. Elouard, et al. 2010. Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India. Ecology 91:3118–3118.

Reed, R. A., R. K. Peet, M. W. Palmer, and P. S. White. 1993. Scale dependence of vegetation-environment correlations: a case study of a North Carolina piedmont woodland. Journal of Vegetation Science 4:329–340.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge.

Supp, S. R., X. Xiao, S. K. M. Ernest, and E. P. White. 2012. An experimental test of the response of macroecological patterns to altered species interactions. Ecology 93:2505–2511.

Thibault, K. M., E. P. White, A. H. Hurlbert, and S. K. M. Ernest. 2011. Multimodality in the individual size distributions of bird communities. Global Ecology and Biogeography 20:145–153.

Thompson, J., N. Brokaw, J. K. Zimmerman, R. B. Waide, E. M. Everham, D. J. Lodge, C. M. Taylor, et al. 2002. Land use history, environment, and tree composition in a tropical forest. Ecological Applications 12:1344–1363.

West, G. B., J. H. Brown, and B. J. Enquist. 1999. A general model for the structure and allometry of plant vascular systems. Nature 400:664–667.

White, E. P., S. K. M. Ernest, A. J. Kerkhoff, and B. J. Enquist. 2007. Relationships between body size and abundance in ecology. Trends in Ecology & Evolution 22:323–330.

White, E. P., K. M. Thibault, and X. Xiao. 2012a. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93:1772–1778.

White, E. P., X. Xiao, N. J. B. Issac, and R. M. Sibly. 2012b. Methodological tools. Pages 1–20 *in* R. M. Sibly, J. H. Brown, & A. Kodric-Brown, eds., Metabolic Ecology: A Scaling Approach. Wiley, Chichester.

Xi, W., R. K. Peet, J. K. Decoster, and D. L. Urban. 2008. Tree damage risk factors associated with large, infrequent wind disturbances of Carolina forests. Forestry 81:317–334.

Xiao, X., S. Aravajy, T. W. Baribault, N. Brokaw, N. L. Christensen, Dasappa, S. J. DeWalt, et al. 2014. Data from: a strong test of the Maximum Entropy Theory of Ecology. The American Naturalist, Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.5fn46.

Zimmerman, J. K., E. M. Everham III, R. B. Waide, D. J. Lodge, C. M. Taylor, and N. V. L. Brokaw. 1994. Responses of tree species to hurricane winds in subtropical wet forest in Puerto Rico: Implications for tropical tree life histories. The Journal of Ecology 82:911–922.

Table 3 – 1. Summary of datasets.

Dataset	Description	Area of Individual Plots (ha)	Number of Plots	Survey Year	References
Serimbu	Tropical rainforest	1	2	1995[1]	1, 2, 3, 4
La Selva	Tropical wet forest	2.24	5	2009	5, 6
ACA Amazon Forest Inventories	Tropical moist forest	1	1	2000-2001	7
BCI	Tropical moist forest	50	1	2010	8, 9, 10
DeWalt Bolivia forest plots	Tropical moist forest	1	2	N/A	11
Lahei	Tropical moist forest	1	3	1998	3, 4, 12, 13
Luquillo	Tropical moist forest	16	1	1994-1996[2]	14, 15
Sherman	Tropical moist forest	5.96	1	1999	16, 17, 18
Cocoli	Tropical moist forest	4	1	1998	16, 17, 18
Western Ghats	Wet evergreen / moist / dry deciduous forests	1	34	1996-1997	19
UCSC FERP	Mediterranean mixed evergreen forest	6	1	2007	20
Shirakami	Beech forest	1	2	2006	3, 4, 21
Oosting	Hardwood forest	6.55	1	1989	22, 23
North Carolina forest plots	Mixed hardwoods / pine forest	1.3 – 5.65	5	1990-1993[3]	24, 25, 26

[1]Kohyama et al. (2001) [2]Kohyama et al. (2003) [3]Lopez-Gonzalez et al. (2009) [5]Lopez-Gonzalez et al. (2011) [5]Baribault et al. (2011) [6]Baribault et al. (2012) [7]Pitman et al. (2005) [8]Condit (1998a) [9]Hubbell et al. (2005) [10]Hubbell et al. (1999) [11]DeWalt et al. (1999) [12]Nishimura et al. (2006) [13]Nishimura and Suzuki (2001) [14]Zimmerman et al. (1994) [15]Thompson et al. (2002) [16]Condit (1998b) [17]Condit et al. (2004) [18]Pyke et al. (2001) [19]Ramesh et al. (2010) [20]Gilbert et al. (2010) [21]Nakashizuka et al. (2003) [22]Reed et al. (1993) [23]Palmer et al. (2007) [24]McDonald et al. (2002) [25]Peet and Christensen (1987) [26]Xi et al. (2008)

---

[1] One plot has a more recent survey in 1998, however it lacks species ID.

[2] We chose Census 2 because information for multiple stems is not available in Census 3, and the unit of diameter is unclear in Census 4. Data from both parts a and b are used.

[3] We chose survey individually for each plot based on expert opinion to minimize the effect of hurricane disturbance.

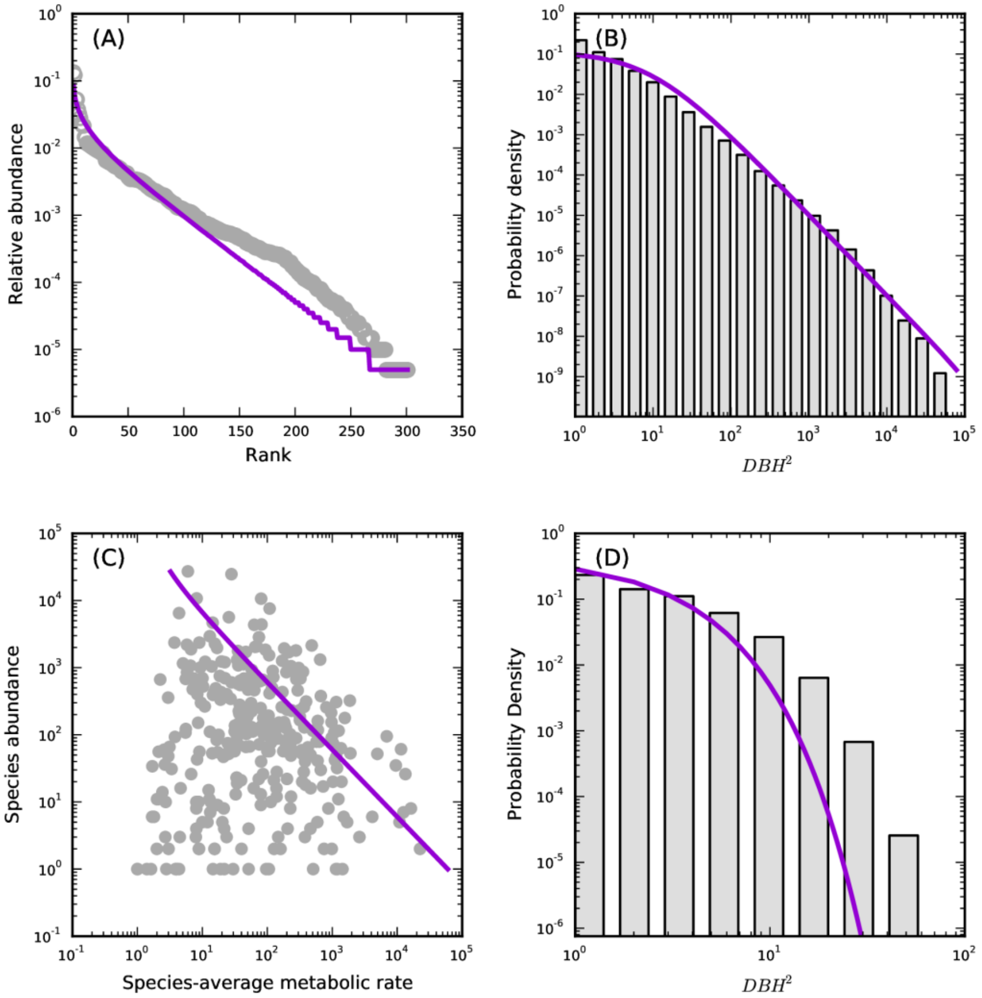


Figure 3 – 1. An illustration of the four patterns with data from Barro Colorado Island. (A) Rank-abundance distribution; (B) Individual size distribution (ISD); (C) Size-density relationship (SDR); (D) Intraspecific individual size distribution (iISD) of the most abundant species, *Hybanthus prunifolius*. Grey dots or bars in each panel represent empirical observations and magenta curve represents METE's prediction.

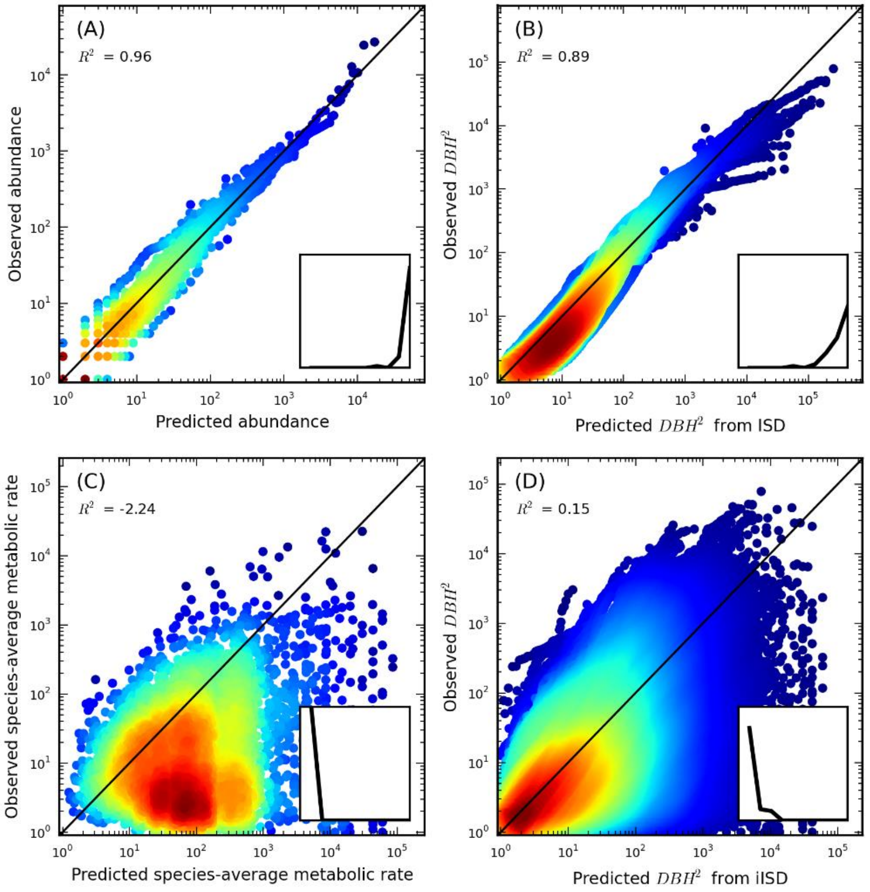


Figure 3 – 2. METE's predictions plotted against empirical observations across 60 communities. (A) SAD (each data point is the abundance of a species at a single rank in one community); (B) ISD (each data point is the metabolic rate of an individual at a single rank in one community); (C) SDR (each data point is the average metabolic rate within one species in one community); (D) iISD (each data point is the metabolic rate of an individual at a single rank belonging to a specific species in one community). The diagonal black line in each panel is the 1:1 line. The points are color-coded to reflect the density of neighbouring points, with warm (red) colors representing higher densities and cold (blue) colors representing lower densities. The inset reflects the distribution of R^2 among 60 communities from negative (left) to 1 (right).

CHAPTER 4

DIRECT COMPARISON OF CONSTRAINT- AND PROCESS-BASED THEORIES USING MULTIPLE PATTERNS *

Abstract

Ecological patterns arise from the interplay of many different processes, and yet the emergence of consistent phenomena across a diverse range of ecological systems suggests that many patterns may in part be determined by statistical or combinatorial constraints. Differentiating the extent to which patterns in a given system are determined statistically, and where this description breaks down and we require explicit ecological processes, has been difficult because methods for comparing predictions across multiple patterns simultaneously are poorly developed. We addressed this challenge by directly comparing a constraint-based theory (Maximum Entropy Theory of Ecology; METE) and a process-based theory (size-structured neutral theory; SSNT) across multiple ecological communities, formulating both theories to predict a single joint distribution. SSNT consistently outperformed METE in characterizing this joint distribution among 60 forest communities. This suggests that, as currently formulated, the demographic processes in SSNT provide better predictions than the constraints in METE when evaluating a suite of patterns simultaneously. This approach provides a first step towards differentiating between process- and constraint-based models of ecological systems and a general approach for comparing ecological theories that make predictions for multiple patterns.

---

* Coauthored by: Xiao, X., J. P. O’Dwyer, and E. P. White.

Introduction

Ecological systems can be characterized by a variety of macroecological patterns occurring across a broad array of ecosystems and taxonomic groups (Brown 1995). These include some of the most well-studied patterns in ecology, such as the uneven distribution of individuals among species (the species abundance distribution or SAD; Fisher et al. 1943; McGill et al. 2007) and the allocation of body size among individuals (the individual size distribution or ISD; Enquist and Niklas 2001; Muller-Landau et al. 2006). In many cases, the same overall shape for a given pattern occurs across multiple ecological systems, and there are two distinct hypotheses to explain this universality. Ecological systems exhibiting the same patterns could all be governed by similar fundamental processes, operating in similar ways. For example, the theory of island biogeography (MacArthur and Wilson 1967) predicts that the species richness on islands is determined by the equilibrium between immigration and extinction, and neutral theory (Hubbell 2001) predicts that community-level diversity patterns arise from individual-level demographic processes. The alternative hypothesis is that general patterns may arise as emergent statistical phenomena from a set of numerical constraints on the systems, while processes operate only indirectly through their effects on the constraints. For example, recent applications of the Maximum Entropy Principle to ecology (e.g., Shipley et al. 2006; Dewar and Porté 2008; Harte 2011) and the feasible set (Locey and White 2013) rely not on the operation of specific processes but instead on the idea that many possible combinations of processes and states of the system produce similar empirical patterns.

Understanding even the broad categories of mechanism underlying ecological patterns is difficult. There are often multiple models based on a variety of different mechanisms that make similar or even identical predictions. For example, more than 20 models exist for the SAD all predicting a realistic hollow-curve shape, but with mechanisms ranging from population dynamics to resource partitioning to purely statistical (McGill et al. 2007). While it is relatively easy for multiple models to make good predictions for a single pattern such as the SAD (Frank

2014), stronger tests of model predictions can be obtained by evaluating predictions for multiple patterns (McGill 2003; McGill et al. 2006). Unified theories that unite multiple patterns under a single theoretical framework (McGill 2010) thus allow for stronger tests to be conducted on the underlying mechanisms.

We examine the performance of two unified theories, the Maximum Entropy Theory of Ecology (METE; Harte 2011) and the size-structured neutral theory (SSNT; O’Dwyer et al. 2009). METE adopts the constraint-based view of community structure, where the ecological patterns arise as the most likely (least biased) state with the assumption that the patterns are constrained by species richness, the total number of individuals, and the total energy consumption of the community. SSNT is a process-based model where the patterns arise as the steady state of a dynamic system governed by birth, death, and growth in size of individuals. Both theories make predictions for an array of ecological patterns including patterns of biodiversity and patterns of biomass and energy use, making these two of the most comprehensive unified theories (see also Dewar and Porté 2008; Morlon et al. 2008). To provide a general comparison between theories with multiple predictions, we show that the predictions of each theory can be summarized in a comprehensive joint distribution that encapsulates all of the predicted patterns. This allows for a direct comparison of the two theories using a likelihood-based method.

We compared the predictions of the two theories using data from 60 forest communities. While both models make identical predictions for the SAD, indicating that either common demographic processes or statistical constraints provide equally accurate explanations for this pattern, we found that when evaluated using the likelihood of the joint distribution, SSNT consistently outperformed METE. This implies the given constraints are not sufficient to describe the overall distributions of individuals and energy consumption in these communities, and that ecological processes provide additional meaningful information. A detailed examination of individual patterns showed that the two theories differ primarily in their predictions for the relationship between individual body size within a species and that species' abundance. These results provide insight into the current state of the process-based versus constraint-based approaches to understanding ecological systems and represent one of the first attempts at strong tests of macroecological theories where the performance of a theory is compared not to a null model but to a meaningful alternative, using multiple predictions, and multiple datasets simultaneously.

Methods

1. Theoretical frameworks

METE proposes that the allocations of individuals and of body size within a community are regulated by three state variables (Harte 2011): species richness S, total abundance N, and total metabolic rate within the community E_{METE} (the subscript is added to distinguish E_{METE} from its counterpart in SSNT with a potentially different unit; see below). In contrast, SSNT introduces a size component to the original neutral theory (Hubbell 2001), with individuals in the community going through the processes of birth, death, and growth in size (O'Dwyer et al. 2009). The structure of the community in SSNT is governed by the forms and values of the demographic parameters b (birth rate), m (mortality rate), and g (rate of growth). Our study adopts the simplest assumption that all three demographic parameters are constant for all individuals regardless of their species identities or body size (i.e., the completely neutral case in O'Dwyer et al. 2009). Note that while the assumption of b and m being constant holds regardless of the unit used for body size, g can only be constant at one particular scale (e.g., constant growth rate measured in diameter does not translate into constant growth rate measured in area or volume).

Both theories predict the same set of four major patterns: the SAD (distribution of individuals among species), the ISD (distribution of body size among individuals regardless of their species identity), the size-density relationship (SDR; relationship between species abundance and average body size within species) (Cotgreave 1993), and the intraspecific individual size distribution (iISD; distribution of body size among individuals within a given species) (Gouws et al. 2011). However, their predictions for the three patterns of body size are not necessarily in the same unit. In METE, body size has the same unit as E_{METE}, or metabolic rate (B), which scales with good approximation as the square of diameter (D): $B \propto D^2$ (West et al. 1999). In contrast, body size in predictions of SSNT has the unit arbitrarily defined by the assumption on growth rate g. Here we adopt the most intuitive assumption that g is constant with respect to diameter D: $g(D) = {}^{dD}/_{\mathrm{d}t} = constant$. In other words, the growth rate measured as the increase in diameter is constant for all individuals regardless of their size, leading to predictions in unit of D for SSNT. To adequately compare the performance of the theories, we converted the ISD and the iISD in METE also to predictions of d with the transformation

$$f(D) = g(B)\left|\frac{d}{dD}B\right| = g(dD^2) \cdot 2D$$

where $f(D)$ is the distribution in unit of D and $g(B)$ is the original distribution predicted by METE in unit of B (Casella and Berger 2001; Stegen and White 2008). Since the SDR predicted by METE does not have a simple analytical form in unit of d, we converted SSNT's prediction to unit of B (D^2) instead.

Table 4 – 1 summarizes the predicted forms of the four patterns in METE and SSNT. Parameters λ_1 and λ_2 in METE are Lagrange multipliers (Jaynes 2003) determined by the state variables S, N and E_{METE} (see Harte 2011 and Appendix E for detailed derivation). Parameters in SSNT in the completely neutral case are ratios of the demographic parameters, b/m and m/g, which can be calculated with equations $\frac{N}{S} = -\frac{b/m}{1-b/m}\log\left(1-\frac{b}{m}\right)$ and $\frac{m}{g} = \frac{N}{E_{\mathrm{SSNT}}-N}$, where E_{SSNT} is the summed diameter across individuals in the community (see Appendix H for derivation). Thus the predictions of SSNT under the completely neutral assumption are also fully determined

by the three variables S, N, and E_{SSNT}, though they are not assumed to constrain the system like state variables in METE. Note that while METE predicts a strong negative correlation between species abundance and average body size within species (see $\Theta_{\mathrm{METE}}(D|n)$ and $\bar{\varepsilon}_{\mathrm{METE}}(n)$ in Table 4 – 1), SSNT predicts that there is no relationship between the two, leading to an iISD that takes the same form of the ISD (i.e., individuals in each species are a random sample from the community) and SDR that is constant independent of abundance ($\Theta_{\mathrm{SSNT}}(D)$ and $\bar{\varepsilon}_{\mathrm{SSNT}}$ in Table 4 – 1).

We define the joint distribution $P(n, D_1, D_2, \ldots, D_n)$ as the probability that a species randomly selected from the community has abundance n, while individuals within the species have diameter D_i's with i ranging from 1 to n. This distribution combines all major non-spatial predictions of the theories, where the SAD is the marginal distribution of n with D_i's integrated out from $P(n, D_1, D_2, \ldots, D_n)$, the ISD is the marginal distribution of D_i, and the iISD is the conditional distribution of D_i given n.

For METE, where the values of D_i's depend on species abundance n (see Table 4 – 1),

$$P_{\mathrm{METE}}(n, D_1, D_2, \ldots, D_n) = \Phi_{\mathrm{METE}}(n) \cdot \prod_{i=1}^{n} \Theta_{\mathrm{METE}}(D_i|n)$$

$$= \frac{1}{Cn} e^{-(\lambda_1+\lambda_2)n} \prod_{i=1}^{n} \frac{2n\lambda_2 D e^{-\lambda_2 n D_i^2}}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_{\mathrm{METE}}}}$$

where C is the normalization constant for the SAD (see Table 4 – 1).

For SSNT, where the values of D_i's are independent of n,

$$P_{\mathrm{SSNT}}(n, D_1, D_2, \ldots, D_n) = \Phi_{\mathrm{SSNT}}(n) \cdot \prod_{i=1}^{n} \Theta_{\mathrm{SSNT}}(D_i) = -\frac{1}{\log(1-\frac{b}{m})} \frac{(\frac{b}{m})^n}{n} \cdot \prod_{i=1}^{n} \frac{m}{g} \cdot e^{-\frac{m}{g}(D_i-1)}$$

2. Data

We focused exclusively on trees in the empirical examination of the two theories. Data on tree communities consistently includes individual level size measurements allowing the compilation of large numbers of communities with the necessary data for evaluating the theories. Tree data samples all individuals of every species down to a certain minimum size. This avoids issues with not detecting juvenile organisms (other than those below the minimum size), which may bias the empirical size distributions. In addition, determinately growing organisms (e.g., birds and mammals) which often exhibit multimodal ISDs (Ernest 2005; Thibault et al. 2011) and unimodal iISDs (Koons et al. 2009; Gouws et al. 2011), whereas the ISDs (Enquist and Niklas 2001; Muller-Landau et al. 2006) and the iISDs (Condit et al. 1998) for trees are in general monotonically decreasing, which is consistent with METE and SSNT's general predictions.

We used data compiled for Chapter 3, which encompassed 60 forest communities worldwide. All communities have been fully surveyed with species identity and measurement of size (diameter or equivalent) for each individual no smaller than a community-specific threshold. If data from multiple surveys are available for one community, we adopted those from the most recent survey unless otherwise specified (see Table 3 – 1). We excluded individuals that were dead, not identified to species/morphospecies, or missing size measurements, as well as those with sizes below or equal to the specified threshold, since not all individuals in these size classes had been included in the surveys. Overall the compilation encompasses 1943 species/morphospecies with 379022 individuals.

3. Analyses

We applied METE and SSNT to each empirical community, and examined their abilities to characterize community structure. Diameter values in each community were rescaled as $D = D_{\mathrm{original}} / D_{\mathrm{min}}$, where D_{min} is the diameter of the smallest individual in the community, so that D has a minimal value of 1 following METE's assumption (see Harte 2011). Multiple branches from the same individual were combined to determine the basal stem diameter with the pipe model, which preserves the total area (metabolic rate) of the branches (Ernest et al. 2009). Predictions of METE and SSNT in each community were obtained with the variables S, N, and E, where $E_{\mathrm{METE}} = \sum_i D_i^2$ and $E_{\mathrm{SSNT}} = \sum_i D_i$.

We first compared the general performance of the two theories with the likelihood of the comprehensive joint distribution $P(n, D_1, D_2, \ldots, D_n)$ in each community, then examined each of the four patterns individually. To quantify the predictive power of the theories, we converted the SAD, the ISD, and the iISD into rank values, where the abundance of species or the diameter of individuals were ranked from the highest to the lowest, and the value at each rank was compared to the theories' predictions (White et al. 2012). For the SDR, we compared the observed average metabolic rate (diameter squared) within each species to those expected from the theories. The explanatory power of METE or SSNT for each pattern was quantified using the coefficient of determination R^2:

$$R^2 = 1 - \frac{\sum_i [\log_{10}(obs_i) - \log_{10}(pred_i)]^2}{\sum_i \left[\log_{10}(obs_i) - \overline{\log_{10}(obs_i)}\right]^2} \qquad (14)$$

where obs_i and $pred_i$ were the ith value of abundance or size (diameter for the ISD and the iISD, metabolic rate for the SDR) in the observed and predicted ranked distributions, respectively. Finally, we examined the empirical patterns to determine if they were significantly different from the theories' predictions by bootstrap analysis (Connolly et al. 2009; Appendix F), where the deviation between the observed and the predicted patterns, quantified with both R^2 and the Kolmogorov-Smirnov statistic, was compared to that for a random sample from the predictions (Appendix I).

Results

The log-likelihood of the joint distribution $P(n, D_1, D_2, \ldots, D_n)$ of SSNT is higher than that of METE in all 60 communities (Fig. 4 – 1), implying that SSNT does a better job characterizing the overall community structure in the allocations of individuals and of body size. Individual examination for each pattern (Fig. 4 – 2) shows that the two theories have almost identical performance in terms of the SAD, which is to be expected from the form of their predictions (i.e., upper-truncated maximum likelihood log-series versus untruncated maximum likelihood logseries; see Table 4 – 1). For the ISD, the two theories have similar predictive power ($R^2_{METE} = 0.89$, $R^2_{SSNT} = 0.86$) despite the difference in their predicted analytical forms (Table 4 – 1), though METE tends to over predict the size of the largest individuals, while SSNT tends to under predict. The discrepancy of the two theories lies mainly in their predictions of the interaction between individual body size and species abundance. METE predicts a negative relationship between the average individual body size within a species and its abundance, which has been shown to be unrealistic in plant communities (Chapter 3; Newman et al. 2014). SSNT, on the other hand, predicts that there is no relationship, leading to better agreement with empirical data for the SDR ($R^2_{METE} = -2.24$, $R^2_{SSNT} = 0.06$) and the iISD ($R^2_{METE} = 0.15$, $R^2_{SSNT} = 0.50$), though substantial scatter still exists (Fig. 4 – 2). On the other hand, the bootstrap analysis (Appendix I) shows that the discrepancy between the theories' predictions and the observations for the ISD and the iISD is almost ubiquitously higher than expected from random sampling. This suggests that neither METE (Fig. F – 2) nor SSNT (Fig. I – 1) with its current formulation is able to fully capture the observed variation in the size distributions, despite their high R^2 values for the ISD. The discrepancy for the SDR in SSNT is less severe with the majority of the communities indistinguishable from random samples of the predicted pattern (Fig. I – 1), which implies that SSNT's prediction of no relationship between species abundance and individual body size is more or less accurate.

Discussion

The Maximum Entropy Theory of Ecology (Harte 2011) and the size-structured neutral theory (O’Dwyer et al. 2009) are two of the most comprehensive unified theories in ecology, with predictions encompassing both patterns of biodiversity and patterns of energy consumption and biomass. Our study evaluated the two theories by directly comparing their performance on multiple patterns simultaneously. This provides a strong test of these two macroecological theories by confronting each with a meaningful alternative, instead of a null model, which is the highest level of model evaluation suggested by McGill et al. (2006). These two theories also represent two distinct perspectives in ecology, with METE assuming that ecological patterns can be determined using statistical constraints, while SSNT builds in the explicit ecological processes of birth, death and growth. We introduced a joint distribution that encapsulates all of the previous predictions of these model as marginal or conditional distributions, and showed that SSNT consistently outperforms METE in characterizing this overall joint distribution of biodiversity and body size distributions in forest communities (Fig. 4 – 1). This results mainly from the distinct predictions of the two theories on the relationship between species abundance and body size distribution within species. METE predicts that common species with high abundances are more likely to contain small individuals, which has been shown not to hold among plants (Chapter 3; Newman et al. 2014). SSNT, on the other hand, predicts no correlation between the two, which is in better agreement with empirical data but still far from fully capturing the observed variation (Fig. 4 – 2).

While our study has formulated METE and SSNT such that they take equivalent sets of variables as inputs to make predictions for the same set of patterns, they represent two different views on the underlying mechanisms. As a constraint-based approach, METE applies the Maximum Entropy Principle (Jaynes 2003) with the constraints defined by ratios of the state variables S, N, and E_{METE}, leading to predictions arising as the most likely state of the system. Since METE makes no explicit assumptions on ecological processes, the processes only operate indirectly through their potential effects on the values of the state variables. In contrast, SSNT stems from a process-based view of community structure, where patterns emerge from the interactions of birth, death, and growth of individuals. While predictions of SSNT in the completely neutral scenario can be quantified with an equivalent set of input variables (S, N, and E_{SSNT}; see Methods: 1. Theoretical frameworks), they represent summary statistics for the demographic parameters (b, m, and g) which then directly give rise to the patterns. The fact that SSNT outcompetes METE suggests that the demographic processes contain meaningful information that helps to shape the macroecological patterns, the effect of which is not simply summarized in the values of the state variables alone.

Despite the equivalent or superior performance of SSNT compared to METE in forest communities among all patterns that we examined, it would be premature to reject METE as a general theory or its underlying constraint-based view as a potential explanation for macroecological patterns. Our conclusions are limited to the current formulations of the theories, and it is possible that improved models from either theoretical perspective could lead to changes in the relative strengths of the two approaches. For example, modifications to METE that decouple the relationship between abundance and body size could improve the relative performance of the constraint-based modeling approach (Harte and Newman 2014), and relaxing the assumption of size-independent growth and mortality may allow SSNT to better account for the variation of the size-related patterns (O'Dwyer et al. in prep).

These improvements are needed because neither METE nor SSNT is yet capable of fully capturing the empirical patterns evaluated here (Fig. 4 – 2, Appendix I) and elsewhere (e.g., Clark and McLachlan 2003; McGill et al. 2006; Newman et al. 2014). These theories are both still being developed, and fully characterizing a broad suite of patterns is difficult for any ecological theory. Our evaluation and comparison of these two unified ecological theories should help point the way forward for future development.

While the constraint-based and the process-based approaches have generally been adopted by distinct theories (such as METE and SSNT), they do not necessarily have to be mutually exclusive. The fact that the three variables S, N and E_{SSNT} are sufficient to characterize the shapes for all predictions in SSNT in the completely neutral case, and that SSNT predicts the same form for the SAD (log-series) as METE with the same inputs (i.e., S and N), strongly suggest that part of the effects of the demographic processes propagate through the state variables. On the other hand, multiple configurations that exist for the same set of constraints can often be tied to (and may eventually be informed from) process-based mechanistic models (Haegeman and Etienne 2010). The attempts to model ecological systems completely with constraints or processes may thus represent two extremes of a continuous spectrum, among which multiple models exist that lean towards one approach or the other, yet all provide adequate characterization of the system if properly formulated. We look forward to future studies that combine new theoretical development with strong empirical tests to further elucidate the entangled effects of constraints versus processes in structuring ecological systems.

References

Brown, J. H. 1995. Macroecology. University Of Chicago Press, Chicago.

Casella, G., and R. L. Berger. 2001. Statistical inference (2nd ed.). Cengage Learning, Stamford.

Clark, J. S., and J. S. McLachlan. 2003. Stability of forest biodiversity. Nature 423:635–638.

Condit, R., R. Sukumar, S. P. Hubbell, and R. B. Foster. 1998. Predicting population trends from size distributions: a direct test in a tropical tree community. The American Naturalist 152:495–509.

Connolly, S. R., M. Dornelas, D. R. Bellwood, and T. P. Hughes. 2009. Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. Ecology 90:3138–3149.

Cotgreave, P. 1993. The relationship between body size and population abundance in animals. Trends in Ecology & Evolution 8:244–248.

Dewar, R. C., and A. Porté. 2008. Statistical mechanics unifies different ecological patterns.

Journal of Theoretical Biology 251:389–403.

Enquist, B. J., and K. J. Niklas. 2001. Invariant scaling relations across tree-dominated communities. Nature 410:655–660.

Ernest, S. K. M. 2005. Body size, energy use, and community structure of small mammals. Ecology 86:1407–1413.

Ernest, S. K. M., E. P. White, and J. H. Brown. 2009. Changes in a tropical forest support metabolic zero-sum dynamics. Ecology Letters 12:507–515.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 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 12:42–58.

Frank, S. A. 2014. Generative models versus underlying symmetries to explain biological pattern. Journal of Evolutionary Biology 27:1172–1178.

Gouws, E. J., K. J. Gaston, and S. L. Chown. 2011. Intraspecific body size frequency distributions of insects. PLoS ONE 6:e16606.

Haegeman, B., and R. S. Etienne. 2010. Entropy maximization and the spatial distribution of species. The American Naturalist 175:E74–E90.

Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. Oxford University Press, Oxford.

Harte, J., and E. A. Newman. 2014. Maximum information entropy: a foundation for ecological theory. Trends in Ecology & Evolution 29:384–389.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton.

Jaynes, E. T. 2003. Probability theory: the logic of science. (G. L. Bretthorst, ed.). Cambridge University Press, Cambridge.

Koons, D. N., R. D. Birkhead, S. M. Boback, M. I. Williams, and M. P. Greene. 2009. The effect of body size on cottonmouth (Agkistrodon piscivorus) survival, recapture probability, and behavior in an Alabama swamp. Herpetological Conservation and Biology 4:221–235.

Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. Ecology Letters 16:1177–1185.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton.

McGill, B. J. 2003. Strong and weak tests of macroecological theory. Oikos 102:679–685.

McGill, B. J. 2010. Towards a unification of unified theories of biodiversity. Ecology Letters

13:627–642.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, 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., G. Chuyong, R. Condit, S. Hubbell, D. Kenfack, D. Thomas, R. Valencia, et al. 2008. A general framework for the distance-decay of similarity in ecological communities. Ecology Letters 11:904–917.

Muller-Landau, H. C., R. S. Condit, K. E. Harms, C. O. Marks, S. C. Thomas, S. Bunyavejchewin, G. Chuyong, et al. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. Ecology Letters 9:589–602.

Newman, E. A., M. E. Harte, N. Lowell, M. Wilber, and J. Harte. 2014. Empirical tests of within- and across-species energetics in a diverse plant community. Ecology.

O'Dwyer, J. P., J. K. Lake, A. Ostling, V. M. Savage, and J. L. Green. 2009. An integrative framework for stochastic, size-structured community assembly. Proceedings of the National Academy of Sciences of the United States of America 106:6170–6175.

Shipley, B., D. Vile, and É. Garnier. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. Science 314:812–814.

Stegen, J. C., and E. P. White. 2008. On the relationship between mass and diameter distributions in tree communities. Ecology Letters 11:1287–1293.

Thibault, K. M., E. P. White, A. H. Hurlbert, and S. K. M. Ernest. 2011. Multimodality in the individual size distributions of bird communities. Global Ecology and Biogeography 20:145–153.

West, G. B., J. H. Brown, and B. J. Enquist. 1999. A general model for the structure and allometry of plant vascular systems. Nature 400:664–667.

White, E. P., K. M. Thibault, and X. Xiao. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93:1772–1778.

Table 4 – 1. Analytical forms of the patterns predicted by METE and SSNT with interpretations.

<table>
<tr><th>Patterns</th><th>METE</th><th>SSNT</th></tr>
<tr><td>SAD</td><td>$\Phi_{\mathrm{METE}}(n) \approx \frac{1}{Cn} e^{-(\lambda_1+\lambda_2)n}$</td><td>$\Phi_{\mathrm{SSNT}}(n) = -\frac{1}{\ln(1-\frac{b}{m})} \frac{(b/m)^n}{n}$</td></tr>
<tr><td colspan="3">Interpretation: the probability that a randomly selected species has abundance n.
Additional parameter C in $\Phi_{\mathrm{METE}}(n)$ is the normalization constant.</td></tr>
<tr><td>ISD</td><td>$\Psi_{\mathrm{METE}}(D) = \frac{2S}{NZ} \cdot D \cdot \frac{e^{-\gamma}}{(1-e^{-\gamma})^2} \cdot \left(1-(N+1)e^{-\gamma N} + Ne^{-\gamma(N+1)}\right)$</td><td>$\Psi_{\mathrm{SSNT}}(D) = \frac{m}{g} e^{-\frac{m}{g}(D-1)}$</td></tr>
<tr><td colspan="3">Interpretation: the probability that a randomly selected individual from the community has diameter between (D, $D + \Delta D$) regardless of species identity.
γ in $\Psi_{\mathrm{METE}}(D)$ is defined as $\gamma = \lambda_1 + \lambda_2 \cdot D^2$, and Z is the normalization constant.</td></tr>
<tr><td>SDR</td><td>$\bar{\varepsilon}_{\mathrm{METE}}(n) = \frac{1}{n\lambda_2(e^{-\lambda_2 n} - e^{-\lambda_2 n E_{\mathrm{METE}}})} \cdot \left[e^{-\lambda_2}(\lambda_2 n + 1) - e^{-\lambda_2 n E_{\mathrm{METE}}}(\lambda_2 n E_{\mathrm{METE}} + 1)\right]$</td><td>$\bar{\varepsilon}_{\mathrm{SSNT}} = \frac{2g^2}{m^2} + \frac{2g}{m} + 1$</td></tr>
<tr><td colspan="3">Interpretation: the average individual metabolic rate within a species with abundance n.
Note that metabolic rate scales as D^2 instead of D.</td></tr>
<tr><td>iISD</td><td>$\Theta_{\mathrm{METE}}(D|n) = \frac{2n\lambda_2 D e^{-\lambda_2 n D^2}}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_{\mathrm{METE}}}}$</td><td>$\Theta_{\mathrm{SSNT}}(D) = \frac{m}{g} \cdot e^{-\frac{m}{g}(D-1)}$</td></tr>
<tr><td colspan="3">Interpretation: the probability that a randomly selected individual within a given species with abundance n has diameter between (D, $D + \Delta D$).</td></tr>
</table>

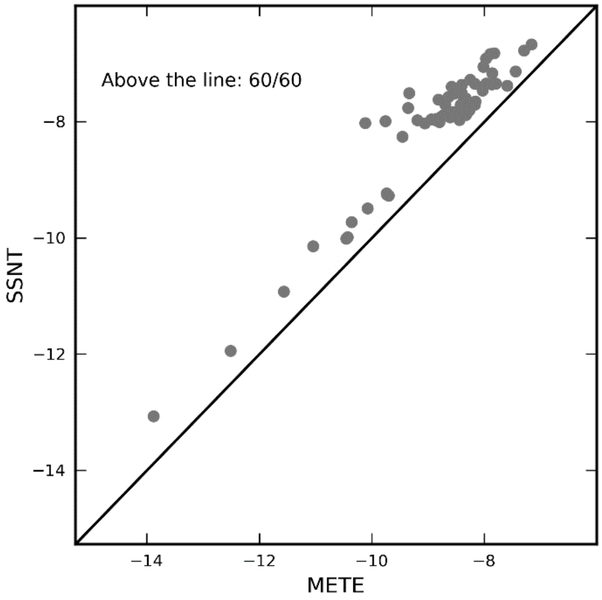
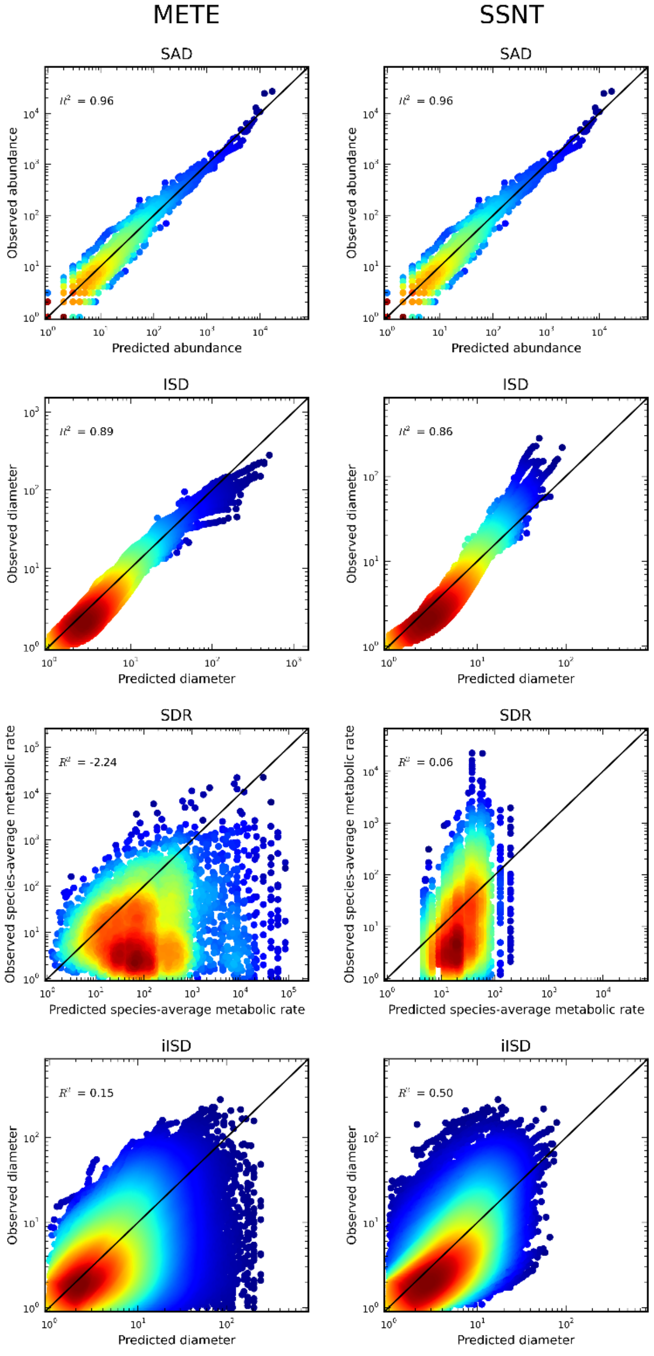


Figure 4 – 1. Comparison of the log-likelihood (*l*) of the joint distribution for METE and SSNT in each of the 60 forest communities. The diagonal line is the one-to-one line. For better visualization, *l* is transformed to –log(-*l*), which is a monotonic transformation that does not change the position of the points with respect to the diagonal line.

Figure 4 – 2. Comparison of the performance of METE and SSNT for each of the four macroecological patterns. Each point in the subplot represents the abundance of one species in a community for the SAD, the diameter of one individual in a community for the ISD, the average metabolic rate (squared diameter) within one species in a community for the SDR, and the diameter of one individual from a given species in a community for the iISD. The colors represent density of the points, where warmer (redder) colors correspond to denser regions. The diagonal line represents the one-to-one line between the predicted values and the observed values.


METE
SSNT
SAD
R^2 = 0.96
Observed abundance
Predicted abundance
SAD
R^2 = 0.96
Observed abundance
Predicted abundance
ISD
R^2 = 0.89
Observed diameter
Predicted diameter
ISD
R^2 = 0.86
Observed diameter
Predicted diameter
SDR
R^2 = -2.24
Observed species-average metabolic rate
Predicted species-average metabolic rate
SDR
R^2 = 0.06
Observed species-average metabolic rate
Predicted species-average metabolic rate
iISD
R^2 = 0.15
Observed diameter
Predicted diameter
iISD
R^2 = 0.50
Observed diameter
Predicted diameter

CHAPTER 5

CONCLUSION

I explored two conceptually different types of approaches as explanations for macroecological patterns, one based on processes and the other on constraints. The process-based approaches attempt to identify the few key ecological processes driving each pattern, directly establishing the link between patterns and processes. Alternatively, the constraint-based approaches attempt to encapsulate the overall effect of all processes with a set of constraints, which then give rise to patterns as emergent statistical properties.

By not explicitly modeling the processes, the constraint-based approaches do not rely on specific knowledge of the system and thus can be broadly applied, making them particularly suitable for patterns that arise across systems regulated by very different processes. My research shows how a constraint-based approach (the feasible set) provides a general explanation for a major ecological pattern, Taylor’s Law. I show that the power-law relationship between the mean and the variance of one or multiple populations is expected simply because the vast majority of possible configurations of a system under two constraints take similar forms. It adds Taylor’s Law to the growing list of macroecological patterns with statistical explanations, and suggests that similar insights may be gained for other general patterns that are observed across systems with different underlying processes.

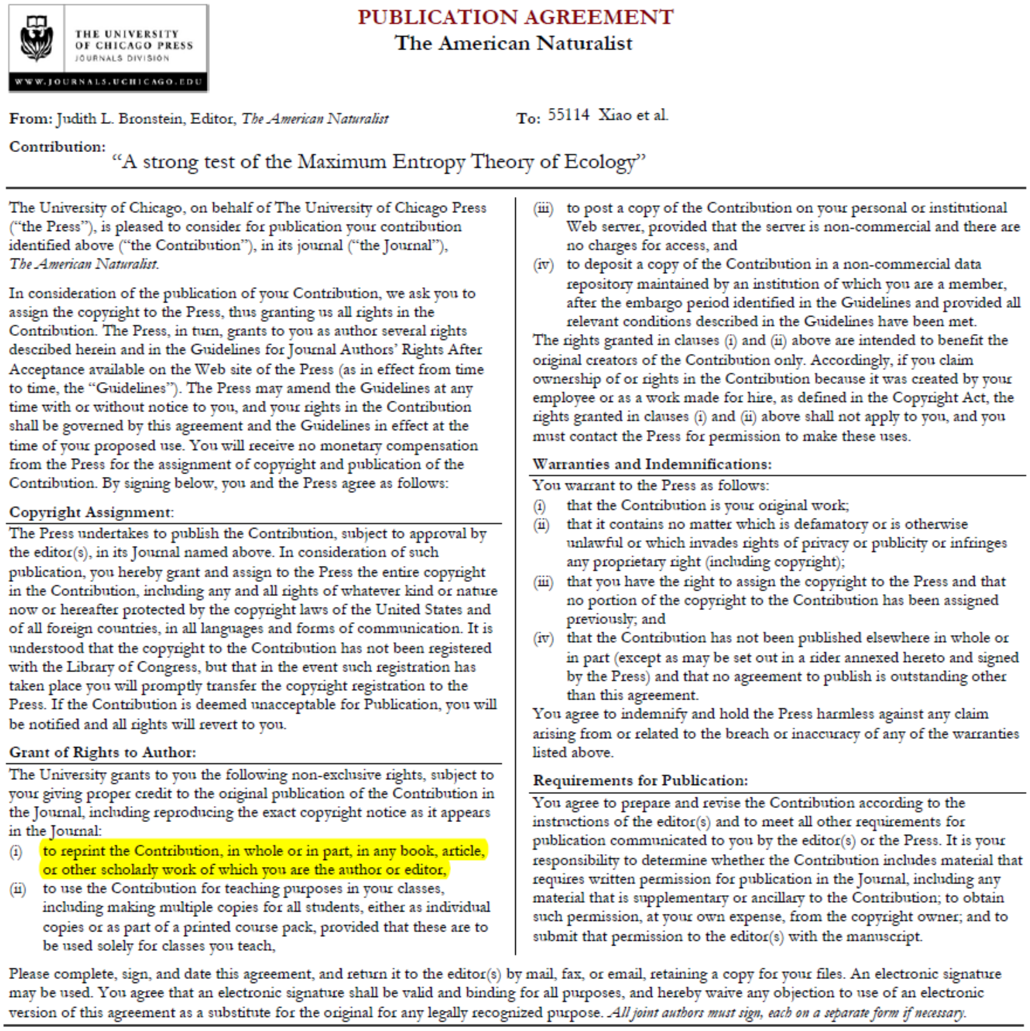
On the other hand, not all macroecological patterns can be fully explained by constraints alone. As the evaluation of the Maximum Entropy Theory of Ecology shows, the constraints currently adopted by even the most comprehensive constraint-based approach are not sufficient to characterize some of the patterns related to body size. Moreover, the comparison between this theory and size-structured neutral theory suggests that demographic processes such as birth, death and growth contain ecologically meaningful information not fully encapsulated in the constraints.

The results of my research suggest that the process- and constraint-based approaches are both valid explanations for macroecological patterns, with some patterns directly tied to processes (e.g., competitive exclusion), some patterns mainly regulated by constraints (e.g., Taylor's Law), and some patterns being somewhere in between with both processes and constraints playing a role. The two kinds of mechanisms may also operate at different levels as shown in my study of Taylor's Law, where the power-law form of the pattern with an exponent between 1 and 2 is well-explained by the two constraints, while the exact shape of the relationship in each system may be tied to system-specific characteristics beyond the constraints. Future studies that combine theoretical development with strong empirical evaluations are needed to further elucidate the roles of the processes and constraints in shaping macroecological patterns.

APPENDICES

APPENDIX A:

PERMISSION TO REPRINT CHAPTER 2



THE UNIVERSITY OF CHICAGO PRESS
JOURNALS DIVISION
WWW.JOURNALS.UCHICAGO.EDU

PUBLICATION AGREEMENT
The American Naturalist

From: Judith L. Bronstein, Editor, *The American Naturalist*

To: 55114 Xiao et al.

Contribution: "A strong test of the Maximum Entropy Theory of Ecology"

The University of Chicago, on behalf of The University of Chicago Press ("the Press"), is pleased to consider for publication your contribution identified above ("the Contribution"), in its journal ("the Journal"), *The American Naturalist*.

In consideration of the publication of your Contribution, we ask you to assign the copyright to the Press, thus granting us all rights in the Contribution. The Press, in turn, grants to you as author several rights described herein and in the Guidelines for Journal Authors' Rights After Acceptance available on the Web site of the Press (as in effect from time to time, the "Guidelines"). The Press may amend the Guidelines at any time with or without notice to you, and your rights in the Contribution shall be governed by this agreement and the Guidelines in effect at the time of your proposed use. You will receive no monetary compensation from the Press for the assignment of copyright and publication of the Contribution. By signing below, you and the Press agree as follows:

Copyright Assignment:

The Press undertakes to publish the Contribution, subject to approval by the editor(s), in its Journal named above. In consideration of such publication, you hereby grant and assign to the Press the entire copyright in the Contribution, including any and all rights of whatever kind or nature now or hereafter protected by the copyright laws of the United States and of all foreign countries, in all languages and forms of communication. It is understood that the copyright to the Contribution has not been registered with the Library of Congress, but that in the event such registration has taken place you will promptly transfer the copyright registration to the Press. If the Contribution is deemed unacceptable for Publication, you will be notified and all rights will revert to you.

Grant of Rights to Author:

The University grants to you the following non-exclusive rights, subject to your giving proper credit to the original publication of the Contribution in the Journal, including reproducing the exact copyright notice as it appears in the Journal:

(i) to reprint the Contribution, in whole or in part, in any book, article, or other scholarly work of which you are the author or editor,
(ii) to use the Contribution for teaching purposes in your classes, including making multiple copies for all students, either as individual copies or as part of a printed course pack, provided that these are to be used solely for classes you teach,
(iii) to post a copy of the Contribution on your personal or institutional Web server, provided that the server is non-commercial and there are no charges for access, and
(iv) to deposit a copy of the Contribution in a non-commercial data repository maintained by an institution of which you are a member, after the embargo period identified in the Guidelines and provided all relevant conditions described in the Guidelines have been met.

The rights granted in clauses (i) and (ii) above are intended to benefit the original creators of the Contribution only. Accordingly, if you claim ownership of or rights in the Contribution because it was created by your employee or as a work made for hire, as defined in the Copyright Act, the rights granted in clauses (i) and (ii) above shall not apply to you, and you must contact the Press for permission to make these uses.

Warranties and Indemnifications:

You warrant to the Press as follows:

(i) that the Contribution is your original work;
(ii) that it contains no matter which is defamatory or is otherwise unlawful or which invades rights of privacy or publicity or infringes any proprietary right (including copyright);
(iii) that you have the right to assign the copyright to the Press and that no portion of the copyright to the Contribution has been assigned previously; and
(iv) that the Contribution has not been published elsewhere in whole or in part (except as may be set out in a rider annexed hereto and signed by the Press) and that no agreement to publish is outstanding other than this agreement.

You agree to indemnify and hold the Press harmless against any claim arising from or related to the breach or inaccuracy of any of the warranties listed above.

Requirements for Publication:

You agree to prepare and revise the Contribution according to the instructions of the editor(s) and to meet all other requirements for publication communicated to you by the editor(s) or the Press. It is your responsibility to determine whether the Contribution includes material that requires written permission for publication in the Journal, including any material that is supplementary or ancillary to the Contribution; to obtain such permission, at your own expense, from the copyright owner; and to submit that permission to the editor(s) with the manuscript.

Please complete, sign, and date this agreement, and return it to the editor(s) by mail, fax, or email, retaining a copy for your files. An electronic signature may be used. You agree that an electronic signature shall be valid and binding for all purposes, and hereby waive any objection to use of an electronic version of this agreement as a substitute for the original for any legally recognized purpose. *All joint authors must sign, each on a separate form if necessary.*

Figure A – 1. Part of the Publication Agreement from The American Naturalist. Permission is given to reprint the work "in whole or in part, in any book, article, or other scholarly work" of which I am the author or editor (see highlight).

APPENDIX B:

LETTERS OF RELEASE FROM DRS. KENNETH J. LOCEY, DANIEL J. MCGLINN, AND JAMES P. O'DWYER

TO WHOM IT MAY CONCERN:

I hereby release the rights to reproduce the manuscript, "A Process-Independent Explanation for the General Form of Taylor's Law" to my coauthor Xiao Xiao for use in her dissertation.

Kenneth J. Locey

27 September 2014

25 September 2014

TO WHOM IT MAY CONCERN:

I hereby release the rights to reproduce the manuscript, “A Strong Test of the Maximum Entropy Theory of Ecology” to my coauthor Xiao Xiao for use in her dissertation.

Daniel J. McGlinn

25 September 2014

TO WHOM IT MAY CONCERN:

I hereby release the rights to reproduce the manuscript, “Direct Comparison of Constraint- and Process-Based Ecological Theories Using Multiple Patterns” to my coauthor Xiao Xiao for use in her dissertation.

APPENDIX C:

CHAPTER 2: INFORMATION ON COMPILED DATASETS

Table C – 1. Summary of datasets.

Study ID	Taxon	Type	Number of Data Sets	Number of Data Sets Included	Reference
1	fish	spatial	1	1	Stanfield et al. 2013
2	bacteria	spatial	16	0	Kaltz et al. 2012
3	arthropod	temporal	2	2	Karban et al. 2012
4	arthropod	spatial	2	2	Hui et al. 2012
5	arthropod	spatial	3	3	Thein and Singh 2011
7	arthropod	spatial	4	4	Costa et al. 2010
8	nematode	spatial	2	0	Aminayanaba 2010
10	arthropod	spatial	1	1	Lessio and Alma 2006
11	fungi	spatial	4	0	Sallam et al. 2007
12	arthropod	spatial	2	0	Nachman 2006
13	mammal	spatial	1	0	McMahon et al. 2005
14	arthropod	spatial	1	1	Sileshi and Magongoya 2004
15	invertebrate	spatial	1	1	Clarke et al. 2002
16	arthropod	spatial	1	1	Silva et al. 2003
17	arthropod	spatial	8	0	Parker et al. 2002
18	annelid	spatial	11	6	Jiménez et al. 2001
19	arthropod	spatial	2	2	Floater 2001

20	arthropod	spatial	1	1	Schexnayder et al. 2001
21	mollusc	spatial	2	2	Eleutheriadis and Lazaridou-Dimitriadou 2001
23	mollusc	spatial	1	1	Babineau 2000
24	mollusc	spatial	1	1	Staikou 1998
25	arthropod	spatial	1	1	Pahl 1969
26	mollusc	spatial	1	1	Todd 1978
27	protist	spatial	4	4	Buzas 1970
28	plant	spatial	3	3	Crawley and Weiner 1991
29	fish	spatial	1	0	Van Damme and Hamerlynck 1992
30	nematode	spatial	2	0	Warren and Linit 1992
31	arthropod	spatial	5	3	Rosewell et al. 1990
32	arthropod	temporal	2	0	Samways 1990
33	Echinorhynchidae	spatial	1	1	Brattery 1986
34	nematode	spatial	4	2	Wheeler et al. 1987
35	arthropod	spatial	1	1	Purrington et al. 1989
36	invertebrate	spatial	2	2	He and Gaston 2003
37	bird	temporal	1	1	Dickson et al. 1993
38	bird	temporal	1	1	Gaston and Blackburn 2000
39	bird	temporal	4	4	Holmes et al. 2012
40	bird	temporal	1	0	Sandercock 2009
41	bird	temporal	1	1	Waide 2012
43	bird	temporal	1	1	Vickery and Nudds 1984

44	bird	temporal	1	1	Williamson 1983
45	herp	temporal	3	3	How 1998; Thompson and Thompson 2005; Wilgers et al. 2006
46	arthropod	temporal	3	3	Anderson 2003
47	arthropod	temporal	4	4	Holmes 1997
48	mollusc	temporal	4	0	Willig and Bloch 2004
49	arthropod	temporal	37	23	Pollard et al. 1986
50	mammal	temporal	1	0	Grant 1976
51	mammal	temporal	1	0	Bestelmeyer 2007
52	mammal	temporal	16	15	Kaufman 2010
53	mammal	temporal	1	0	Merritt 1999
54	mammal	temporal	9	7	Friggens 2008
55	mammal	temporal	7	2	Stapp 2006
56	mammal	temporal	8	2	SANParks 1989; SANParks 1997; SANParks 2004; SANParks 2009
57	plant	temporal	1	0	Venable 2008
58	plant	temporal	1	0	Adler et al. 2007
59	plant	temporal	1	0	Clark and Clark 2006
60	plant	temporal	1	0	Zachmann et al. 2010
62	mammal	temporal	1	0	Ernest et al. 2009

Data sources include the Desert Laboratory, supported by NSF grants DEB 9419905 (LTREB), DEB 0212782 (LTREB), and DEB 0717466 (LTREB); HBES LTER (with data on bird abundance at Hubbard Brook and Lepidoptera larvae abundance in northern hardwood forests, provided by Richard T. Holmes), a collaborative effort at the Hubbard Brook Experimental Forest operated and maintained by the USAD Forest Service, Northern Research Station, Newtown Square, PA; Jornada Basin LTER, supported by the NSF grant DEB-1235828; KNZ LTER, supported by the NSF Long Term Ecological Research Program; Luquillo LTER, supported by grants BSR-8811902, DEB 9411973, DEB 0080538, DEB 0218039, DEB 0620910 and DEB 0963447 from NSF to the Institute for Tropical Ecosystem Studies, University of Puerto Rico, and to the International Institute of Tropical Forestry USA Forest Service, with additional support from the U.S. Forest Service (Dept. of Agriculture) and the University of Puerto Rico; SANParks; Sevilleta LTER, supported by the National Science Foundation Long Term Ecological Research program with NSF grant numbers BSR 88-11906, DEB 9411976, DEB 0080529 and DEB 0217774; and Virginia Coast Reserve LTER, supported by NSF Grants BSR-8702333-06, DEB 9211772, DEB 9411974, DEB 0080381 and DEB 0621014. Results in this study are derived from and thus a modified version of the original data.

References

Adler, P. B., W. R. Tyburczy, and W. K. Lauenroth. 2007. Long-term mapped quadrats from Kansas prairie: Demograhic information for herbaceous plants. Ecology 88:2673.

Aminayanaba, O. A. 2010. Determination of cluster size of Pratylenchus Penetrans phytonematode in a rose field. Journal of Agriculture and Social Research 10:178–185.

Anderson, J. 2003. Arthropod Pitfall Traps at LTER II NPP sites. Jornada LTER: ARTHPIT3 [Database].

Babineau, D. 2000. *Estimating the population size of razor clams using a model assisted sampling design and analysis*.

Bestelmeyer, B. 2007. Small Mammal Exclosure Study. Jornada LTER: smesrdnt.

Brattery, J. 1986. Life history and population biology of larval Acanthocephalus lucii (Acanthocephala: Echinorhynchidae) in the isopod Asellus aquaticus. Journal of Parasitology 72:633–645.

Buzas, M. A. 1970. Spatial homogeneity: Statistical analyses of unispecies and multispecies populations of Foraminifera. Ecology 51:874–879.

Clark, D. B., and D. A. Clark. 2006. Tree growth, mortality, physical condition, and microsite in an old-growth lowland tropical rain forest. Ecology 87:2132.

Clarke, R. T., M. T. Furse, R. J. M. Gunn, J. M. Winder, and J. F. Wright. 2002. Sampling variation in macroinvertebrate data and implications for river quality indices. Freshwater Biology 47:1735–1751.

Costa, M. G., J. C. Barbosa, P. T. Yamamoto, and R. M. Leal. 2010. Spatial distribution of Diaphorina citri Kuwayama (Hemiptera: Psyllidae) in citrus orchards. Scientia Agricola 67:546–554.

Crawley, M. J., and J. Weiner. 1991. Plant size variation and vertebrate herbivory: Winter wheat grazed by rabbits. Journal of Applied Ecology 28:154–172.

Dickson, J. G., R. N. Conner, and H. J. Williamson. 1993. Neotropical migratory bird communities in a developing pine plantation. Proceedings of the annual conference of the Southeastern Association of Fish and Wildlife Agencies (pp. 439–446).

Eleutheriadis, N., and M. Lazaridou-Dimitriadou. 2001. The life cycle, population dynamics, growth and secondary production of Bithynia graeca (Westerlund, 1879) (Gastropoda) in Lake Kerkini, North Greece. Journal of Molluscan Studies 67:319–328.

Ernest, S. K. M., T. J. Valone, and J. H. Brown. 2009. Long-term monitoring and experimental manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona, USA. Ecology 90:1708–1708.

Floater, G. J. 2001. Habitat complexity, spatial interference, and “minimum risk distribution”: A framework for population stability. Ecological Monographs 71:447–468.

Friggens, M. 2008. Small Mammal Mark-Recapture Population Dynamics at Core Research Sites at the Sevilleta National Wildlife Refuge, New Mexico. Albuquerque, NM: Sevilleta Long Term Ecological Research Site Database: SEV008.

Gaston, K. J., and T. M. Blackburn, eds. 2007. Appendix II: Eastern Wood Breeding Bird Data. Pattern and Process in Macroecology (pp. 355–357). Blackwell Science Ltd, Malden, MA, USA.

Grant, P. R. 1976. An 11-year study of small mammal populations at Mont St. Hilaire, Quebec. Canadian Journal of Zoology 54:2156–2173.

He, F., and K. J. Gaston. 2003. Occupancy, spatial variance, and the abundance of species. The American Naturalist 162:366–375.

Holmes, R. T. 1997. Long-term trends in abundance of Lepidoptera larvae in northern hardwood forests. knb-lter-hbr.82.2 (http://tropical.lternet.edu/knb/metacat/knb-lter-hbr.82.2/lter).

Holmes, R. T., S. Sillett, N. Rodenhouse, F. W. Sturges, and T. W. Sherry. 2012. Bird Abundances at Hubbard Brook (1969-2010) and on three replicate plots (1986-2000) in the White Mountain National Forest. Durham, NH. Hubbard Brook Data Archive [Database]. http://hubbardbrook.org/data/dataset.php?id=81.

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.

Hui, C., C. Boonzaaier, and L. Boyero. 2012. Estimating changes in species abundance from occupancy and aggregation. Basic and Applied Ecology 13:169–177.

Jiménez, J. J., J.-P. Rossi, and P. Lavelle. 2001. Spatial distribution of earthworms in acid-soil savannas of the eastern plains of Colombia. Applied Soil Ecology 17:267–278.

Kaltz, O., P. Escobar-Páramo, M. E. Hochberg, and J. E. Cohen. 2012. Bacterial microcosms obey Taylor's law: effects of abiotic and biotic stress and genetics on mean and variance of population density. Ecological Processes 1:5.

Karban, R., P. Grof-Tisza, J. L. Maron, and M. Holyoak. 2012. The importance of host plant limitation for caterpillars of an arctiid moth (Platyprepia virginalis) varies spatially. Ecology 93:2216–2226.

Kaufman, D. W. 2010. Seasonal summary of numbers of small mammals on the LTER traplines in prairie. Konza Prarie LTER: CSM04.

Lessio, F., and A. Alma. 2006. Spatial distribution of nymphs of Scaphoideus titanus (Homoptera: Cicadellidae) in grapes, and evaluation of sequential sampling plans. Journal of Economic Entomology 99:578–582.

McMahon, C. R., M. N. Bester, H. R. Burton, M. A. Hindell, and C. J. A. Bradshaw. 2005. Population status, trends and a re-examination of the hypotheses explaining the recent declines of the southern elephant seal Mirounga leonina. Mammal Review 35:82–100.

Merritt, J. 1999. Long Term Mammal Data from Powdermill Biological Station. Virginia Coast Reserve Long-Term Ecological Research Project Data Publication knb-lter-vcr.67.11.

Nachman, G. 2006. The effects of prey patchiness, predator aggregation, and mutual interference on the functional response of Phytoseiulus persimilis feeding on Tetranychus urticae (Acari: Phytoseiidae, Tetranychidae). Experimental and Applied Acarology 38:87–111.

Pahl, P. J. 1969. On testing for goodness-of-fit of the negative binomial distribution when expectations are small. Biometrics 25:143–151.

Parker, B. L., S. D. Costa, M. Skinner, and M. El Bouhssini. 2002. Sampling Sunn Pest (Eurygaster integriceps Puton) in overwintering sites in Northern Syria. Turkish Journal of Agriculture and Forestry 26:109–117.

Pollard, E., M. L. Hall, and T. J. Bibby. 1986. Monitoring the abundance of butterflies, 1976-1985. Joint Nature Conservation Committee.

Purrington, F. F., J. E. Bater, M. G. Paoletti, and B. R. Stinner. 1989. Ground beetles from a remnant oak-maple-beech forest and its surroundings in northeastern Ohio (Coleoptera: Carabidae). The Great Lakes Entomologist 22:105–110.

Rosewell, J., B. Shorrocks, and K. Edwards. 1990. Competition on a divided and ephemeral resource: Testing the assumptions. I. aggregation. Journal of Animal Ecology 59:977–1001.

Sallam, M. N., C. A. McAvoy, P. R. Samson, and J. J. Bull. 2007. Soil sampling for Metarhizium anisopliae spores in Queensland sugarcane fields. BioControl 52:491–505.

Samways, M. J. 1990. Species temporal variability: Epigaeic ant assemblages and management for abundance and scarcity. Oecologia 84:482–490.

Sandercock, B. K. 2009. Bird Populations. Konza Prarie LTER CBP01. http://www.konza.ksu.edu/knz/pages/data/Knzdsdetail.aspx?datasetCode=CBP01.

SANParks. 1989. Species data for Mountain Zebra National Park. peggym.3.4. (http://tropical.lternet.edu/knb/metacat/peggym.3.4/lter).

SANParks. 1997. Census totals for large herbivores in the Kruger National Park summarized by year and region 1965-1997. judithk.814.4. (http://tropical.lternet.edu/knb/metacat/judithk.814.4/lter).

SANParks. 2004. Golden Gate Highland National Parks Census Data. peggym.113.6 (http://tropical.lternet.edu/knb/metacat/peggym.113.6/lter).

SANParks. 2009. Karoo Natioanal Park Census Data. 1994 - 2009. peggym.117.10. (http://tropical.lternet.edu/knb/metacat/peggym.117.10/lter).

Schexnayder, H. P., T. E. Reagan, and D. R. Ring. 2001. Sampling for the sugarcane borer (Lepidoptera: Crambidae) on sugarcane in Louisiana. Journal of Economic Entomology 94:766–771.

Sileshi, G., and P. L. Magongoya. 2004. Incidence of snout beetles Diaecoderus spp. on planted fallow species and mize in agroforestry practices. Proceedings of the Regional Agroforestry Conference on Agroforestry Impacts on Livelihoods in Southern Africa: Putting Research into Practice 257–262.

Silva, L., V. Vieira, and J. Tavares. 2003. Sampling plans for Pseudaletia unipuncta (Lepidoptera: Noctuidae) larvae in Azorean pastures. Environmental Entomology 32:1211–1218.

Staikou, A. E. 1998. Aspects of life cycle, population dynamics, growth and secondary production of the pulmonate snail Cepaea vindobonensis (Férussac, 1821) in northern Greece. Journal of Molluscan Studies 64:297–308.

Stanfield, L. W., N. P. Lester, and I. C. Petreman. 2013. Optimal effort intensity in backpack electrofishing surveys. North American Journal of Fisheries Management 33:277–286.

Stapp, P. 2006. SGS-LTER Long-Term SGS-LTER Long-Term Monitoring Project: Small Mammals on Trapping Webs. Shortgrass Steppe LTER: LTMntrSmlMamWebs.

Thein, M. M., and V. S. Singh. 2011. A method for transforming rice stinkbug counts in rice. International Rice Research Notes 29.2:28–30.

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.

Todd, C. D. 1978. Changes in spatial pattern of an intertidal population of the nudibranch mollusc Onchidoris muricata in relation to life-cycle, mortality and environmental heterogeneity. Journal of Animal Ecology 47:189–203.

Van Damme, P. A., and O. Hamerlynck. 1992. The infection dynamics and dispersion pattern of Lernaeocera branchialis L. on 0+ whiting (Merlangius merlangus L.) in the Oosterschelde (SW Netherlands). Journal of Fish Biology 41:265–275.

Venable, D. L. 2008. Long-term population dynamics of individually mapped Sonoran Desert winter annuals from the Desert Laboratory, Tucson AZ. University of Arizona.

Vickery, W. L., and T. D. Nudds. 1984. Detection of density-dependent effects in annual duck censuses. Ecology 65:96–104.

Waide, R. B. 2012. Avian populations Long-Term Monitoring. Luquillo LTER project of the Institute for Tropical Ecosystem Studies: Bird abundance – point counts [Database].

Warren, J. E., and M. J. Linit. 1992. Within-wood spatial dispersion of the Pinewood nematode, Bursaphelenchus xylophilus. Journal of Nematology 24:489–494.

Wheeler, T. A., C. M. Kenerley, M. J. Jeger, and J. L. Starr. 1987. Effect of quadrat and core sizes on determining the spatial pattern of Criconemella sphaerocephalus. Journal of Nematology 19:413–419.

Wilgers, D. J., E. A. Horne, B. K. Sandercock, and A. W. Volkmann. 2006. Effects of rangeland management on community dynamics of the herpetofauna of the tallgrass prairie. Herpetologica 62:378–388.

Williamson, M. 1983. The land-bird community of Skokholm: Ordination and turnover. Oikos 41:378–384.

Willig, M., and C. P. Bloch. 2004. Community Ecology of Land Snails Survey (Long-term population dynamics of snails in the tabonuco forest). Luquillo LTER project of the Institute for Tropical Ecosystem Studies: Lterdb107 [Database].

Zachmann, L., C. Moffet, and P. Adler. 2010. Mapped quadrats in sagebrush steppe: long-term data for analyzing demographic rates and plant–plant interactions. Ecology 91:3427–3427.

APPENDIX D:

CHAPTER 2: EXAMPLES SHOWING DISTRIBUTION WITHIN THE FEASIBLE SETS FOR INDIVIDUAL (Q, N) PAIRS AND DATASETS

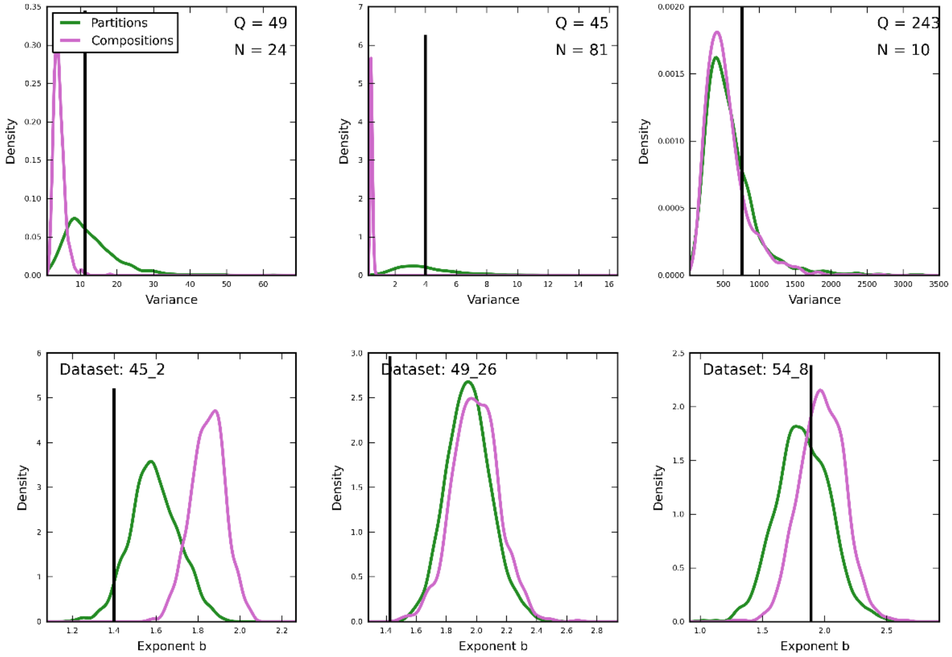


Figure D – 1. Examples showing the distribution within the feasible sets. Top panels: variance calculated for configurations sampled from the feasible sets. Bottom panels: exponent *b* estimated for (m_i, s_{ij}^2) pairs generated from the feasible sets. The black vertical line in each subplot represents the empirical variance or *b* of the data from which the feasible sets are constructed.

APPENDIX E:

CHAPTER 3: DERIVATIONS FOR THE EQUATIONS

The equations we adopted in our analysis (see Methods**:** 1. Predicted patterns of METE) are largely identical to those in Harte (2011), except for a few minor modifications. Below we briefly summarize the derivations, and derive those that are slightly different. See Harte (2011) for the step-by-step procedure.

The distribution of central significance on which all other predictions are based is *R*(*n*, *ε*), the joint probability that a species randomly picked from the community has abundance *n* and an individual randomly picked from such a species has metabolic rate between (*ε*, *ε* +Δ*ε*). By maximizing information entropy $I = -\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot R(n,\varepsilon)\log(R(n,\varepsilon))$ with respect to the constraint on average abundance per species

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot nR(n,\varepsilon) = \frac{N_0}{S_0} \quad \text{(Eqn 1 in the main text; Eqn 7.2 in Harte 2011)}$$

and the constraint on total metabolic rate per species

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot n\varepsilon R(n,\varepsilon) = \frac{E_0}{S_0} \quad \text{(Eqn 2 in the main text; Eqn 7.3 in Harte 2011)}$$

as well as the normalization condition $\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot R(n,\varepsilon) = 1$ (Eqn 7.1 in Harte 2011), *R*(*n*, *ε*) can be obtained as

$$R(n,\varepsilon) = \frac{1}{Z} e^{-\lambda_1 n} e^{-\lambda_2 n\varepsilon} \quad \text{(Eqn 7.13 in Harte 2011)}$$

where the normalization constant *Z* is given by

$$Z = \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot e^{-\lambda_1 n} e^{-\lambda_2 n\varepsilon} \quad \text{(Eqn 7.14 in Harte 2011)}$$

With reasonable approximations, the Lagrange multipliers λ_1 and λ_2 are given by

$$\sum_{n=1}^{N_0} e^{-(\lambda_1+\lambda_2)\cdot n} \Big/ \sum_{n=1}^{N_0} \frac{e^{-(\lambda_1+\lambda_2)n}}{n} \approx \frac{N_0}{S_0} \quad \text{(Eqn 7.26 in Harte 2011)}$$

$$\lambda_2 \approx \frac{S_0}{E_0 - N_0} \quad \text{(Eqn 7.27 in Harte 2011)}$$

Derivation for equations not found in Harte (2011):

1. Species-abundance distribution (SAD; Eqn 3 in main text)

From Eqn 7.23 in Harte (2011):

$$\Phi(n) = \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot R(n,\varepsilon) = \frac{e^{-(\lambda_1+\lambda_2)n} - e^{-(\lambda_1+E_0\lambda_2)n}}{\lambda_2 Z n} \qquad \text{(Eqn E1)}$$

Note that this distribution is properly normalized, i.e., $\sum_{n=1}^{N_0} \Phi(n) = 1$.

Given that E_0 is large, the second term in the numerator, $e^{-(\lambda_1+E_0\lambda_2)n}$, is much smaller than the first term $e^{-(\lambda_1+\lambda_2)n}$. Dropping the second term,

$$\Phi(n) \approx \frac{e^{-(\lambda_1+\lambda_2)n}}{\lambda_2 Z n} \qquad \text{(Eqn E2)}$$

This approximation leads to the familiar Fisher's log-series distribution, upper-truncated at N_0. However, the form in Eqn E2 is not properly normalized, which can cause problems when the SAD is converted to the RAD (rank-abundance distribution). To ensure the proper normalization of Φ(*n*), we replace the constant term in the Eqn E2, $\lambda_2 Z$, with constant *C*, where

$$C = \sum_{n=1}^{N_0} \frac{e^{-(\lambda_1+\lambda_2)n}}{n} \qquad \text{(Eqn E3)}$$

2. The energetic analog of the individual size distribution (ISD; Eqn 4 in main text)

From Eqn 7.6 in Harte (2011):

$$\begin{aligned}\Psi(\varepsilon) &= \frac{S_0}{N_0}\sum_{n=1}^{N_0} n \cdot R(n,\varepsilon) \\ &= \frac{S_0}{N_0 Z}\sum_{n=1}^{N_0} n \cdot e^{-\lambda_1 n}\, e^{-\lambda_2 n\varepsilon} \\ &= \frac{S_0}{N_0 Z}\sum_{n=1}^{N_0} n \cdot e^{-(\lambda_1+\lambda_2\varepsilon)n} \\ &= \frac{S_0}{N_0 Z} \cdot e^{-(\lambda_1+\lambda_2\varepsilon)} \cdot \frac{1-(N_0+1)e^{-N_0(\lambda_1+\lambda_2\varepsilon)} + N_0 e^{-(N_0+1)(\lambda_1+\lambda_2\varepsilon)}}{(1-e^{-(\lambda_1+\lambda_2\varepsilon)})^2}\end{aligned}$$

$$= \frac{S_0}{N_0 Z} \cdot \frac{e^{-\gamma}}{(1-e^{-\gamma})^2} \cdot (1 - (N_0 + 1)e^{-\gamma N_0} + N_0 e^{-\gamma(N_0+1)}) \qquad \text{(Eqn E4)}$$

where $\gamma = \lambda_1 + \lambda_2 \cdot \varepsilon$. Note that Eqn E4 is not identical to Eqn 7.24 in Harte (2011), which contains a minor error (J. Harte, pers. comm.). However, the trivial difference is unlikely to invalidate or significantly change any published results.

3. The energetic analog of the size-density relationship (Eqn 6 in main text)

From Eqn 7.25 in Harte (2011):

$$\Theta(\varepsilon|n) = \frac{n\lambda_2 e^{-\lambda_2 n\varepsilon}}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_0}} \qquad \text{(Eqn E5)}$$

Then

$$\bar{\varepsilon}(n) = \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot \varepsilon \cdot \Theta(\varepsilon|n)$$

$$= \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot \varepsilon \cdot \frac{n\lambda_2 e^{-\lambda_2 n\varepsilon}}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_0}}$$

$$= \frac{n\lambda_2}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_0}} \int_{\varepsilon=1}^{E_0} d\varepsilon \cdot \varepsilon \cdot e^{-\lambda_2 n\varepsilon}$$

$$= \frac{1}{n\lambda_2 (e^{-\lambda_2 n} - e^{-\lambda_2 n E_0})} \cdot \left[e^{-\lambda_2 n}(\lambda_2 n + 1) - e^{-\lambda_2 n E_0}(\lambda_2 n E_0 + 1)\right] \qquad \text{(Eqn E6)}$$

References

Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. Oxford University Press.

APPENDIX F:

CHAPTER 3: BOOTSTRAP ANALYSIS

We conducted a bootstrap analysis to examine if the deviation of the empirical data from the distributions predicted by METE was comparable to that of random samples drawn from the distributions themselves. In each community, we obtained the Lagrange multipliers λ_1 and λ_2 with empirically observed state variables S_0, N_0, and E_0, which determined the form of the predicted patterns (Eqns 3-5 in main text). We drew 500 bootstrap samples from the predicted distribution for each pattern. For the SAD, samples of size S_0 were drawn from the upper-truncated log-series distribution defined by Eqn 3. For the ISD, samples of size N_0 were drawn from the distribution defined by Eqn 4. For the SDR and the iISD, samples for each species given its abundance n were drawn from the exponential distribution defined by Eqn 5. The SDR of each sample community was then obtained by taking the average body size within a bootstrap sample for each species. Note that this sampling scheme assumes independence among values within each bootstrap sample. As a result, the values of the original state variables are unlikely to be preserved in the bootstrap samples. However, given that the discrepancy is not systematic, and that the results of the bootstrap analysis are highly consistent both across samples and across communities (see Fig. F – 2 below), we conclude that the assumption of independence should not qualitatively affect our results.

The deviation between empirical data or bootstrap samples and METE's predictions were quantified using R^2 and the Kolmogorv-Smirnov (K-S) statistic. The K-S statistic is defined as

$$D_n = \sqrt{n}\sup|F_n(x) - F(x)| \qquad \text{(Eqn F1)}$$

where n is sample size, $F_n(x)$ is the empirical cumulative distribution function, and $F(x)$ is the reference (predicted) cumulative distribution function. Therefore the K-S statistic directly measures the largest discrepancy in shape between two distributions across multiple points. Note that the statistic is defined for distributions only and thus cannot be applied to the SDR. However, since the SDR and the iISD are closely related (see Methods in the main text), the iISD results can provide insights for the SDR. Fig. F – 1 illustrates the comparison between the empirical data and the bootstrap samples for the SAD and the iISD when evaluated with the two statistics, using data from one community (USCS FERP) as an example.

We converted the test statistics within each community into quantiles so that results can be pooled across communities. The quantile for each pattern in a community was calculated for each of the two statistics as the proportion of bootstrap samples that had larger deviations from METE's prediction (i.e., lower value of R^2, or higher value of K-S statistic) than the empirical data. For example, Fig. F – 1A shows that 77% of the bootstrap SADs have a lower R^2 than the empirical SAD in the community USCS FERP. For the iISD, where bootstrap samples were independently generated for each species, the quantile of the K-S statistic for a given community was calculated as the average quantile across all species having 10 or more individuals, weighted by their abundances.

Comparisons between the empirical data and the bootstrap samples for the four ecological patterns across all 60 communities are summarized in Fig. F – 2. Results from the two statistics are qualitatively consistent (though note again that the K-S statistic cannot be applied to the SDR, which is not a probability distribution). While the bootstrap analysis confirms that METE provides a satisfactory characterization for the empirical SAD but not for the empirical SDR or the iISD, it shows that the empirical ISD cannot be fully accounted for by METE's prediction, despite the relatively high R^2 within and across communities for this pattern.

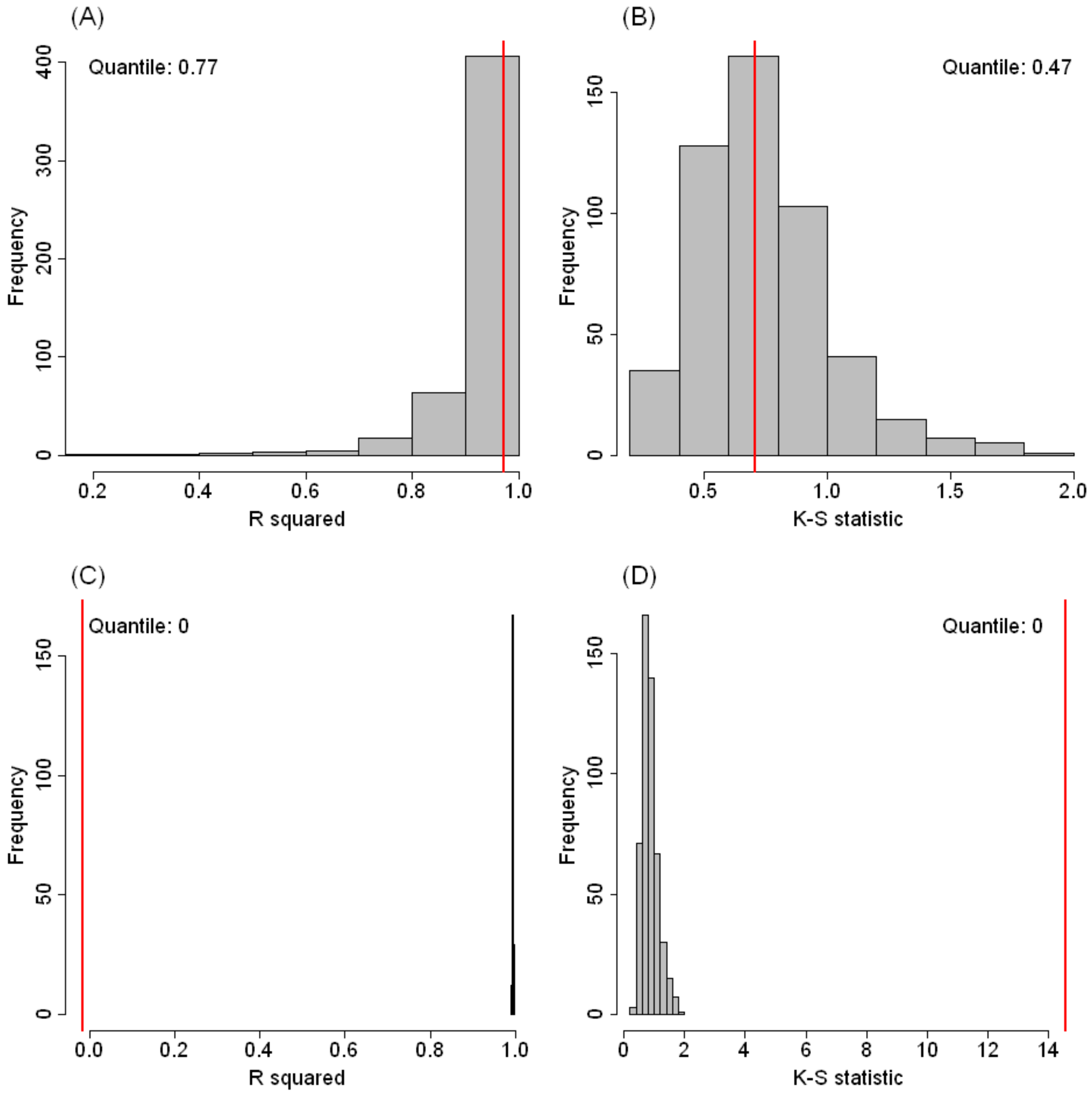
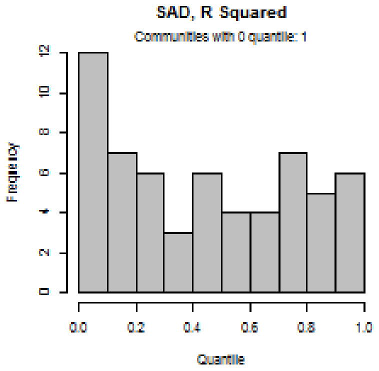
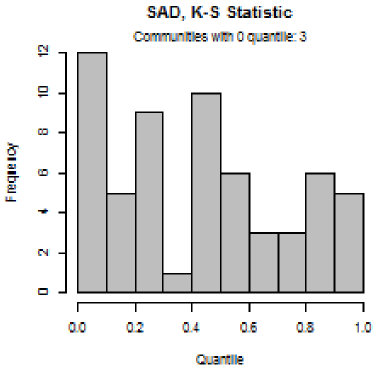


Figure F – 1. Illustration of the bootstrap analysis using data from UCSC FERP as an example. (A) and (B) show the results for the SAD when evaluated with R^2 (A) and the K-S statistic (B), while (C) and (D) show the results for the iISD. In each panel the histogram represents the frequency distribution of the test statistic among the 500 bootstrap samples, while the red vertical line represents the test statistic of the empirical data. Note that for the iISD the K-S statistic was individually obtained for each species, and the illustration in (D) only includes the results for one species, *Pseudotsuga menziesii*.

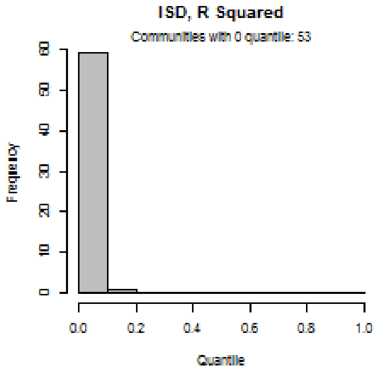
Figure F – 2. Results of the bootstrap analysis across all 60 communities. The histogram in each panel is the frequency distribution of the quantile values across the 60 communities for one pattern using one statistic (R^2 or K-S statistic), where each quantile value represents the quantile of the empirical statistic among that of the 500 bootstrap communities. The number of communities where the quantile equals zero (i.e., where the empirical data have a larger deviation from the predicted pattern than any of the bootstrap samples) is also given. Note that for the iISD, the quantile of the K-S statistic is a pooled value across all species with abundance > 10 in a community, and thus can only equal zero when the quantiles for all species are zero.


SAD, R Squared

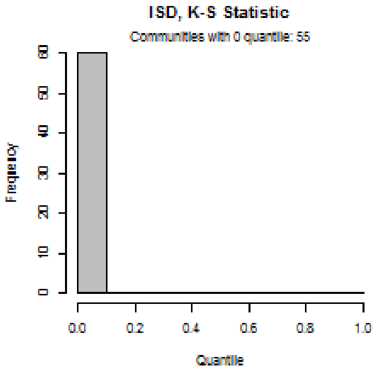
Communities with 0 quantile: 1


SAD, K-S Statistic

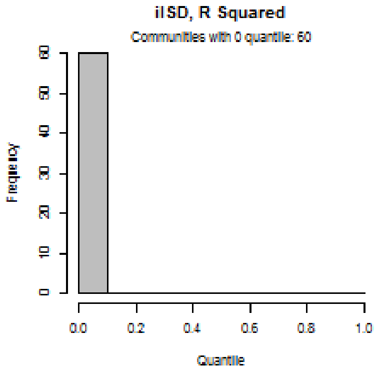
Communities with 0 quantile: 3


ISD, R Squared

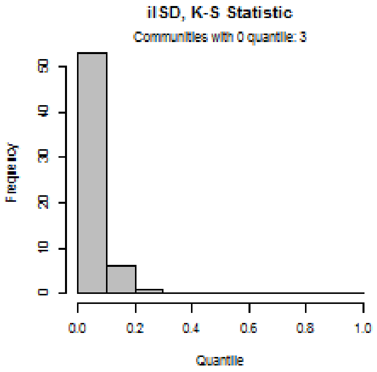
Communities with 0 quantile: 53


ISD, K-S Statistic

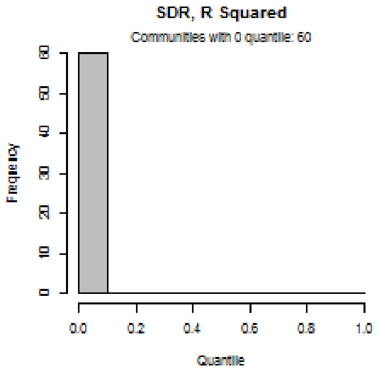
Communities with 0 quantile: 55


iISD, R Squared

Communities with 0 quantile: 60


iISD, K-S Statistic

Communities with 0 quantile: 3


SDR, R Squared

Communities with 0 quantile: 60

APPENDIX G:

CHAPTER 3: MODEL COMPARISON FOR THE ISD

Muller-Landau et al. (2006) proposed four possible distributions (exponential, Pareto, Weibull, and quasi-Weibull) for diameter in old-growth forests, under different assumptions of growth and mortality. Here we compare the fit of three of the four distributions (exponential, Pareto, and Weibull) to the fit of the ISD predicted by METE (Eqn 8) using data from the 60 forest communities. The quasi-Weibull distribution, which has been shown to provide the best fit for the majority of communities (Muller-Landau et al. 2006), is not evaluated due to the difficulty in obtaining its maximum likelihood parameters when it is left-truncated.

All distributions are left-truncated to account for the fact that individuals below the minimal threshold in each community where excluded from the datasets. With the minimal size rescaled as 1 across communities (see Methods**)**, the left-truncated exponential distribution takes the form

$$f(D) = \lambda e^{-\lambda(D-1)} \qquad \text{(Eqn G1)}$$

the left-truncated Pareto distribution takes the form

$$f(D) = \frac{\alpha}{D^{\alpha+1}} \qquad \text{(Eqn G2)}$$

the left-truncated Weibull distribution takes the form

$$f(D) = \frac{k}{\lambda}(\frac{D}{\lambda})^{k-1} e^{-({}^{D}/_{\lambda})^{k}} / e^{-({}^{1}/_{\lambda})^{k}} \qquad \text{(Eqn G3)}$$

where the diameter D >= 1 for all three distributions.

Parameters in Eqns G1, G2 and G3 were obtained with maximum likelihood method (MLE) for each community. While analytical solutions exist for parameters in Eqn G1 and Eqn G2, MLE solutions for parameters in Eqn G3 can only be obtained numerically. The three distributions of D were then transformed into distributions of D^2 (surrogate for metabolic rate; see Methods) to be consistent with METE’s prediction (Eqn 8) as:

$$g(D^2) = \frac{1}{2D} f(D) \qquad \text{(Eqn G4)}$$

where f(D) is the left-truncated exponential, Pareto, or Weibull distribution in Eqns G1, G2 or G3.

The fit of the ISD predicted by METE and the other three distributions was evaluated with Akaike's Information Criterion (AIC; Burnham and Anderson 2002). AIC$_c$, a second-order variant of AIC which corrects for finite sample size, was computed for each distribution as

$$AIC_c = 2k - 2\ln(L) + \frac{2k(k+1)}{n-k-1} \qquad \text{(Eqn G5)}$$

where k is the number of parameters in the corresponding distribution, n is the number of individuals in the community, and L is the likelihood of the distribution across all individuals (Burnham and Anderson 2002). Within a community, the distribution with a lower AIC$_c$ value provides a better fit.

Our results show that overall the Weibull distribution provides the best fit for the ISD, which outperforms the other three distributions (i.e., has the smallest AIC$_c$ value) in 50 out of 60 communities. While METE is exceeded by the Weibull distribution in all except 3 communities, its performance is comparable to that of the other two distributions, with METE outperforming the exponential distribution in 24 communities and the Pareto distribution in 33 (Table G – 1).

Table G – 1. The AIC_c value of the four distributions of ISD across communities. The distribution with the best fit (lowest AIC_c value) for each community is in bold.

Dataset	Site	AIC_c -exponential	AIC_c-Pareto	AIC_c -Weibull	AIC_c -METE
FERP	FERP	85971.15	82823.11	**81893.76**	88390.74
ACA	eno-2	3047.892	3123.951	**3037.737**	3048.544
WesternGhats	BSP104	8447.378	8232.82	**8147.375**	8597.933
WesternGhats	BSP11	9670.786	9737.739	**9565.319**	9756.008
WesternGhats	BSP12	8072.348	7580.985	**7580.105**	8005.097
WesternGhats	BSP16	6505.854	6465.984	**6371.536**	6473.227
WesternGhats	BSP27	4158.854	4352.934	**4154.657**	4168.587
WesternGhats	BSP29	5200.085	5601.832	**5186.167**	5246.872
WesternGhats	BSP30	**5228.032**	5550.478	5229.22	5272.148
WesternGhats	BSP36	5363.257	4997.568	**4994.507**	5613.485
WesternGhats	BSP37	6648.723	**5882.951**	5940.894	6702.201
WesternGhats	BSP42	4862.353	4579.541	**4572.774**	4912.597
WesternGhats	BSP5	6316.684	**5868.932**	5879.056	6344.512
WesternGhats	BSP6	8362.132	8224.467	**8144.515**	8368.706
WesternGhats	BSP65	10730.14	10597.32	10418.12	**10323.55**
WesternGhats	BSP66	6127.039	6078.716	**5969.159**	6118.758
WesternGhats	BSP67	5733.979	6116.641	**5713.447**	5970.901
WesternGhats	BSP69	9639.039	9839.743	**9566.506**	9677.272
WesternGhats	BSP70	7568.366	7643.62	7475.877	**7471.337**
WesternGhats	BSP73	**13866.8**	14638.34	13867.97	14056.6

WesternGhats	BSP74	10384.88	10164.99	**10043.66**	10178.07
WesternGhats	BSP75	**3828.718**	4032.776	3830.225	3844.366
WesternGhats	BSP79	10012.15	10192.38	**9943.069**	10014.63
WesternGhats	BSP80	10351.04	10721.97	**10333.53**	10392.1
WesternGhats	BSP82	7775.241	8109.038	**7766.727**	7779.842
WesternGhats	BSP83	**10080.84**	10603.67	10082.84	10184.62
WesternGhats	BSP84	9941.77	10676.22	**9906.56**	10087.81
WesternGhats	BSP85	4090.759	4051.023	**3986.417**	4092.965

References

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

Muller-Landau, H. C., R. S. Condit, K. E. Harms, C. O. Marks, S. C. Thomas, S. Bunyavejchewin, G. Chuyong, et al. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. Ecology Letters 9:589–602.

APPENDIX H:

CHAPTER 4: DERIVATION FOR PREDICTIONS OF SSNT

Predictions of SSNT have been presented in detail in O’Dwyer et al. (2009) for both the general case with arbitrary demographic parameters *b* (birth rate), *m* (mortality rate), *g* (growth rate) and *v* (speciation rate), and the special case where all parameters are constant across individuals in the community (i.e., the completely neutral case). Here we adopt the completely neutral case, and further derive the form of the predictions following the notation in the main text.

SSNT predicts that the size component of the size does not affect the SAD, which still takes the same form as in the original neutral theory:

$$\Phi_{\mathrm{SSNT}}(n) = \frac{v}{bn}(\frac{b}{m})^n \qquad \text{(Eqn H1; modified from Eqn 3 in O'Dwyer et al. 2009)}$$

The mean size spectrum, or the average number of individuals per species in a given size class, is given by

$$< n(D) >= \frac{v}{g(1-\frac{b}{m})} e^{-\frac{m}{g}D} \qquad \text{(Eqn H2; modified from Eqn 17 in O'Dwyer et al. 2009)}$$

Transforming the above equation into probability distribution (ISD) yields

$$\Psi_{\mathrm{SSNT}}(D) = \frac{S}{N} < n(D) >= \frac{Sv}{Ng(1-\frac{b}{m})} e^{-\frac{m}{g}D} \qquad \text{(Eqn H3)}$$

where *S* is species richness and *N* is the total abundance in the community.

Since Eqns H1 and H3 are probability distributions, they have to be properly normalized, requiring that

$$\frac{v}{b} = -\frac{1}{\ln\left(1-{}^{b}/_{m}\right)} \qquad \text{(Eqn H4)}$$

$$\frac{Sv}{Ng(1-\frac{b}{m})} = \frac{m}{g} \qquad \text{(Eqn H5)}$$

Solving Eqns H4 and H5 simultaneously yields a solution for the parameter *b*/*m* for the SAD:

$$\frac{N}{S} = -\frac{{}^{b}\!/_{m}}{1 - {}^{b}\!/_{m}} \log\left(1 - \frac{b}{m}\right)$$

which insures that both normalization conditions (Eqns H4 and H5) are satisfied. Therefore the SAD predicted by SSNT is a log-series distribution with *b*/*m* being the maximum likelihood parameter (White et al. 2012). The parameter *m*/*g* characterizing the ISD (Eqn H3), however, can take arbitrary values. We adopt the additional assumption that *m*/*g* is also the maximum likelihood parameter. Given the observed ISD which is lower-truncated at 1 with rescaling, it follows that

$$\frac{m}{g} = \frac{1}{\bar{D} - 1} = \frac{1}{\frac{E_{\mathrm{SSNT}}}{N} - 1} = \frac{N}{E_{\mathrm{SSNT}} - N}$$

The SDR measured as average metabolic rate, or D^2, can then be calculated as the expected value of the iISD (which in the completely neutral case takes the same form as the ISD) converted to distribution of $\varepsilon = D^2$:

$$\bar{\varepsilon}_{\mathrm{SSNT}} = \mathrm{E}(\Theta_{\mathrm{SSNT}}(\varepsilon)) = \int_1^\infty \varepsilon \cdot \frac{m}{2g} \cdot \frac{1}{\varepsilon^{0.5}} e^{-\frac{m}{g}(\varepsilon^{0.5}-1)} d\varepsilon$$

$$(let\ t = \varepsilon^{0.5} - 1) = \frac{m}{2g} \int_0^\infty (t+1) e^{-\frac{m}{g}t} d(t^2 + 2t + 1)$$

$$= \frac{m}{g} \int_0^\infty (t+1)^2 e^{-\frac{m}{g}t} dt = \frac{2g^2}{m^2} + \frac{2g}{m} + 1$$

References

O’Dwyer, J. P., J. K. Lake, A. Ostling, V. M. Savage, and J. L. Green. 2009. An integrative framework for stochastic, size-structured community assembly. Proceedings of the National Academy of Sciences of the United States of America 106:6170–6175.

White, E. P., K. M. Thibault, and X. Xiao. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93:1772–1778.

APPENDIX I:

CHAPTER 4: BOOTSTRAP ANALYSIS

In the main text we examined the performance of METE and SSNT with two metrics – the log-likelihood of the joint distribution $P(n, D_1, D_2, \ldots, D_n)$, which quantifies the general performance of a theory compared to another in characterizing the overall pattern of abundance and body size; the R^2 value between observed values and predicted values, which quantifies the explanatory power of a theory for a single pattern. However, neither metric takes into account the intrinsic variation in a probability distribution, which may translate into discrepancy between the observations and the predictions even when the predicted form is accurate.

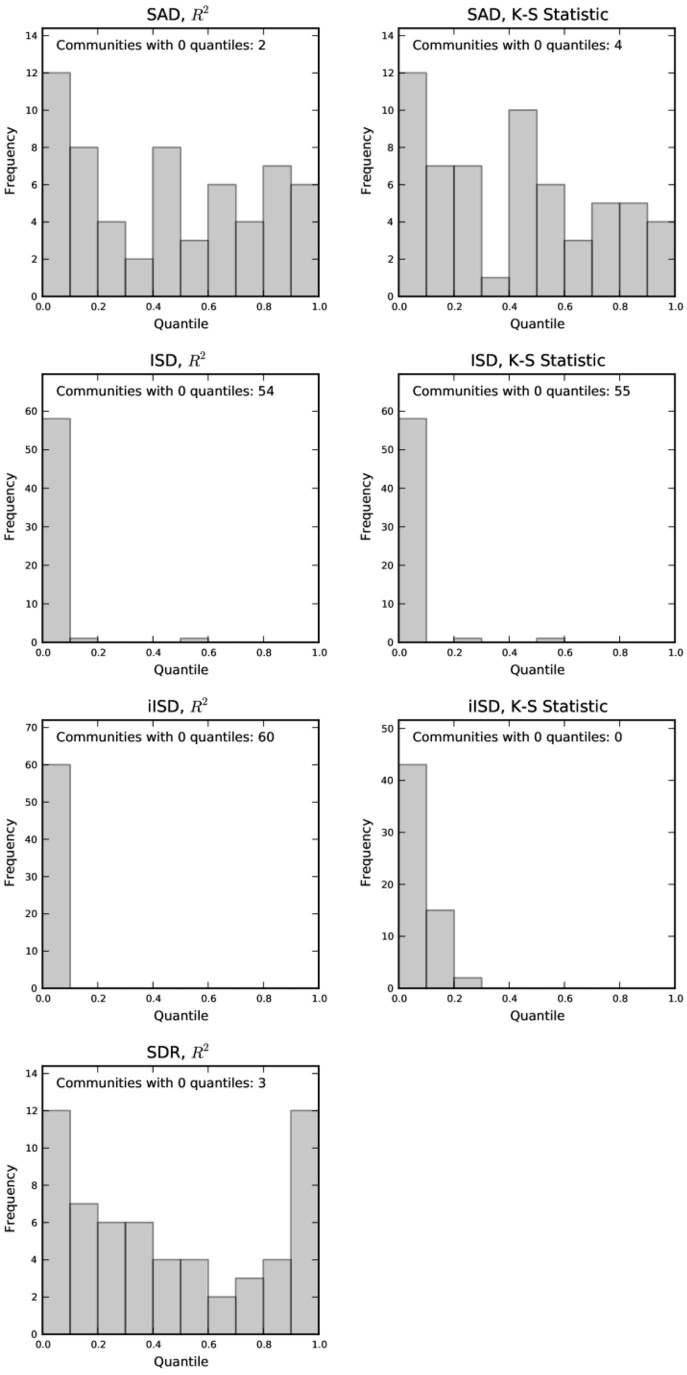
Here we examine the discrepancy between random samples from a distribution and the predicted (rank) values as a measure of the intrinsic variation, which is then compared to the discrepancy between the predicted values and the observations. If the discrepancy calculated for the observations is no larger than that for the random samples, it would imply that the observations are indistinguishable from a random sample from the predicted distribution. Alternatively, if the discrepancy for the observations is significantly higher, it would imply that the observations do not fully conform to the predicted distribution.

We followed the same procedure as in Appendix F. We drew 500 random samples from the distributions predicted for the SAD, the ISD, and the iISD by METE and SSNT (see Table 4 – 1 in the main text), with the parameterization empirically obtained from S, N and E for each community. The SDR was then obtained as the average values of the iISD converted to D^2 for a given species. The discrepancy between a random sample and the values predicted by the theories was measured with two metrics, R^2 and the Kolmogorov-Smirnov (K-S) statistic.

We computed the R^2 for all four patterns, and the K-S statistic for the three patterns except for the SDR, which is not a probability distribution and thus the K-S statistic does not apply. We compared the statistics obtained for empirical observations to those obtained for random samples of the predicted distributions by calculating the proportion (quantile) of random samples that have equal or higher discrepancy (lower values of R^2 or larger K-S statistic) than the observations. For the iISD, where there is one distribution (and thus one K-S statistic) for each species, we computed the quantile as the average across all species with no less than 10 individuals in the community.

As Figs F – 2 and I – 1 show, the log-series SAD predicted by both METE and SSNT provides a satisfying characterization of the empirical distribution of abundance among species in the majority of communities (i.e., a non-negligible proportion of random samples show equal or higher discrepancy compared to the observed values). The empirical patterns of the ISD and the iISD differ from the predictions of both theories. However, SSNT significantly improves the fit of the SDR, where the pattern in most communities is indistinguishable from random samples from SSNT's prediction. This reflects that SDR is a higher-level pattern with lesser degree of details compared to individual-level patterns such as the ISD and the iISD. It also implies that SSNT's prediction of no relationship between body size and species abundance may be more or less on target, despite the fact that the empirical ISD (and the iISD) does not conform to the predicted exponential distribution.

Figure I – 2. Results of the bootstrap analysis for SSNT. The histogram in each panel shows the frequency distribution of the quantile values among the 60 communities for a given pattern, where each quantile value represents the proportion of random samples (among 500) that have equal or higher discrepancy (lower R^2 or larger K-S statistic) from the predicted values compared to the observations.


SAD, R^2
Communities with 0 quantiles: 2
SAD, K-S Statistic
Communities with 0 quantiles: 4
ISD, R^2
Communities with 0 quantiles: 54
ISD, K-S Statistic
Communities with 0 quantiles: 55
iISD, R^2
Communities with 0 quantiles: 60
iISD, K-S Statistic
Communities with 0 quantiles: 0
SDR, R^2
Communities with 0 quantiles: 3
Frequency
Quantile
0.0
0.2
0.4
0.6
0.8
1.0

CURRICULUM VITAE

XIAO XIAO
Department of Biology
Utah State University, Logan UT 84322
Email: xiao@weecology.org
Website: http://xiao-xiao.org/

Education

- 12/2014 Ph.D. in Ecology
  Utah State University, Logan, UT
- 05/2011 M.S. in Statistics
  Utah State University, Logan, UT
- 05/2008 B.Sc. in Biology (Honours) with Minor in Mathematics
  McGill University, Montreal, Canada

Honors and Awards

- Diversity Fellowship, Utah State University *2008 – 2011*
- First Class Honours in Biology, McGill University *2008*
- Science Undergraduate Research Award, McGill University *2007*
- Dean's Honour List, McGill University *2006 - 2008*

Skills

- Programming: Python, R, MATLAB, Version Control
- Quantitative: Bayesian statistics, spatial statistics, time series analysis, hierarchical modeling, GLM, stochastic processes, ODE, PDE

Publications

Journal Articles

Xiao, X., D. J. McGlinn, and E. P. White. *In press.* A strong test of the Maximum Entropy Theory of Ecology. The American Naturalist.

McGlinn, D. J., **X. Xiao**, and E. P. White. 2013. An empirical evaluation of four variants of a universal species-area relationship. PeerJ 1: e212.

Xiao, X., and G. F. Fussmann. 2013. Armstrong-McGehee mechanism revisited: competitive exclusion and coexistence of nonlinear consumers. Journal of Theoretical Biology 339: 26-35.

Supp, S. R., **X. Xiao**, S. K. M. Ernest, and E. P. White. 2012. An experimental test of the response of macroecological patterns to altered species interactions. Ecology 93: 2505-2511.

White, E. P., K. M. Thibault, and **X. Xiao**. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93: 1772-1778.

Xiao, X., E. P. White, M. B. Hooten, and S .L. Durham. 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power-laws. Ecology 92: 1887-1894.

Book Chapters

White, E. P., **X. Xiao**, N. J. B. Issac, and R.M. Sibly. 2012. Methodological tools. 7-20 *in* J. H. Brown, R. M. Sibly, and A. Kodric-Brown, editors. Metabolic Ecology: A Scaling Approach. Wiley-Blackwell.

Manuscripts

McGlinn, D. J., **X. Xiao**, J. Kitzes, and E. P. White. Exploring the spatially explicit predictions of the Maximum Entropy Theory of Ecology. *Submitted to Global Ecology and Biogeography*.

Software and Data Products

Data from: A strong test of the Maximum Entropy Theory of Ecology. By **Xiao, X.**, S. Aravajy, T. W. Baribault, N. Brokaw, N. L. Christensen, Dasappa, S. J. DeWalt, et al. 2014. The American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.5fn46.

METE: Python-based tool for fitting and modeling the Maximum Entropy Theory of Ecology. By E.P. White, K.M. Thibault, **X. Xiao**, D.J. McGlinn, and S.R. Supp. https://github.com/weecology/METE

Presentations

Talks

"Feasible Set: A mechanism-free explanation for Taylor's Law".

- Gordon Research Conference on Unifying Ecology across Scales, Biddeford, ME, USA, July 22nd, 2014.

"Confront big theory with big data: a strong test of the Maximum Entropy Theory of Ecology".

- INTECOL, London, UK, August 21st, 2013.
- iEvoBio, Snowbird, UT, USA, June 25th, 2013. (***Winner of iEvoBio travel award***)

Posters

Xiao, X., E. P. White, and D. J. McGlinn.

- Ecological Society of America, 2012.
- Gordon Research Conference on Metabolic Basis of Ecology, 2012.

Xiao, X., E. P. White, M. B. Hooten, and S. L. Durham.

- Gordon Research Conference on Metabolic Basis of Ecology, 2010.

Teaching Experience

TA for Biology I/II (Biology 1610/1620) lab, Utah State University (2012-2013).
TA for Software Carpentry Bootcamp, Utah State University (03/23/2013-03/24/2013)

Professional and Community Service

Manuscript Reviewing

Ecography, *Ecological Monographs*, *Ecology*, *Journal of Theoretical Biology*, *PLOS Computational Biology*, *Population Ecology*