Ecology, 87(6), 2006, pp. 1411–1423

EMPIRICAL EVALUATION OF NEUTRAL THEORY

Brian J. McGill,[1,3] Brian A. Maurer,[1] and Michael D. Weiser[2]

[1]*Department of Fisheries and Wildlife, Natural Resources Building, Room 13, Michigan State University, East Lansing, Michigan 48823 USA*
[2]*Department of Ecology and Evolutionary Biology, Biosciences West, University of Arizona, Tucson, Arizona 85721 USA*

Abstract. We describe a general framework for testing neutral theory. We summarize similarities and differences between ten different versions of neutral theory. Two central predictions of neutral theory are that species abundance distributions will follow a zero-sum multinomial distribution and that community composition will change over space due to dispersal limitation. We review all published empirical tests of neutral theory. With the exception of one type of test, all tests fail to support neutral theory. We identify and perform several new tests. Specifically, we develop a set of best practices for testing the fit of the zero-sum multinomial (ZSM) vs. a lognormal null hypothesis and apply this to a data set, concluding that the lognormal outperforms neutral theory on robust tests. We explore whether a priori parameterization of neutral theory is possible, and we conclude that it is not. We show that non-curve-fitting predictions readily derived from neutral theory are easily falsifiable. In toto, there is a current overwhelming weight of evidence against neutral theory. We suggest some next steps for neutral theory.

Key words: community ecology; neutral theory; testing theories.

Introduction

Neutral theories of biodiversity assert that all individuals of all species are competitively identical. Any variation in traits between species has no impact on their overall abundance or on their speciation rates. This contradicts 100 years of community ecology (Elton 1927, Gause 1934, Odum 1959, MacArthur 1972, Roughgarden 1979, Pianka 1988, Maurer 1999). Thus, neutral theory has provided a much-needed challenge and invigoration to community ecology. Scientists are now going back and paying much more careful attention and applying new levels of rigor to the study of niches (Wisheu 1998, Peterson et al. 1999, Silvertown et al. 1999, Gregory and Gaston 2000, McKane et al. 2002, Chase and Liebold 2003). At the same time, a large number of empirical tests of neutral theory have been conducted (e.g., Condit et al. 2002, Clark and McLachlan 2003, McGill 2003*a*, Ricklefs 2003).

The main goal of this paper is to provide a review of empirical tests of neutral theory. We start by defining a framework that allows for careful evaluation of the tests. We then summarize neutral model theory and predictions, discussing to which organisms neutral theory should apply, describing types and strengths of tests, and reviewing empirical tests. We add some new tests to fill in gaps identified in published tests. We conclude with some suggestions of useful directions in which to proceed with neutral theory.

Which Neutral Model Are We Testing?

There are at least 10 different neutral community models proposed (see Table 1, Chave 2004, Holyoak and Loreau 2006). A more detailed discussion of the differences between the models can be found in Appendix A, but some critical attributes are listed below and the correspondence of these attributes to specific models is given in Table 1.

Is it a metacommunity model?—The metacommunity (Hubbell 2001, Leibold et al. 2004) or regional pool (Ricklefs 1987) is the large spatial setting within which local dynamics in abundance occur. A purely local model of neutral dynamics results in fixation to a single species (Hubbell and Foster 1986), which is clearly unrealistic. In all neutral models developed in 2000 and after, emigration from the metacommunity to the local community slows fixation to a single species.

Is it a spatially explicit metacommunity?—Metacommunities can be modeled in one of two ways. Spatially explicit metacommunity models create a number of local communities (with an explicit spatial relationship between the local communities influencing migration) and model the metacommunity as the sum of the local communities. Spatially implicit metacommunities model only two groups, one local community, and one metacommunity. In these models the metacommunity is assumed to be so large with changes occurring so slowly that the metacommunity is completely independent of and unchanging relative to the local community. This is sometimes called the fast-slow assumption (fast

Manuscript received 18 May 2004; revised 12 April 2005; accepted 13 April 2005. Corresponding Editor: M. Holyoak. For reprints of this Special Feature, see footnote 1, p. 1368.

[3] Present address: Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec H3A 1B1 Canada. E-mail: mail@brianmcgill.org

TABLE 1. A summary of 10 versions of the neutral model.

Reference	Has metacommunity model?	Spatially explicit metacommunity?	Constant local population (zero-sum)?	Finite metapopulation?	Species abundances affect other abundances?
Caswell (1976)	no†	NA	yes/no‡	NA	no
Hubbell (1979), Hubbell and Foster (1986)	no†	NA	yes	NA	yes
Hubbell (2001)	yes	no	yes	yes	yes
Bell (2000)	yes	no	yes§	yes	yes
Bell (2001, 2003)	yes	yes	yes§	yes	yes
Volkov et al. (2003)	yes	no	yes	no	no
McKane et al. (2000), Vallade and Houchmandzadeh (2003), McKane et al. (2004)	yes	no	yes	yes¶	no
Etienne and Olff (2004)	yes	no	yes	yes	yes
He (2005)	yes	no	yes	no	no
Etienne (2005)	yes	no	yes	yes	yes

Notes: The column definitions are described in more detail in *Which neutral model are we testing?* Abbreviations are: NA, not applicable; SAD, species abundance distribution; SAR, species–area relationship. Several different implementations of spatially explicit models have been omitted from this table but are compared in Appendix A.

† Caswell explicitly modeled and Hubbell discussed qualitatively the role of migration from outside the local community, but there is no quantitative model of a metacommunity or regional pool.

‡ Caswell analyzed models both with fixed and with varying population sizes, but he cited theorems that prove the results he studied are independent of this fact.

§ Bell's models have only an upper limit on population size (carrying capacity), but populations can be below this level.

¶ An infinite metapopulation was also presented, which was faster to solve analytically.

local dynamics, slow metacommunity dynamics). The fast-slow assumption is convenient for modeling, but it is biologically unrealistic to assume that the metacommunity is so big as to be unchanging but small enough that all individuals in the metacommunity have equal probability of migrating into the local community.

Is there a constant local population (zero-sum)?—Some models have the same number of individuals in the local community at every time step. This is known as the zero-sum assumption. Hubbell (2001) makes the zero-sum assumption a central assumption in the neutral model. However, Caswell (1976) showed that zero-sum and variable size local community models produce very similar results.

Is the metapopulation finite?—All spatially explicit and some spatially implicit metacommunity models assign a particular population size to the metacommunity, while some implicit models allow the metacommunity size to vary, possibly approaching infinity. A varying metacommunity effectively breaks the zero-sum assumption for the metacommunity.

Do species abundances affect each other?—It is easier to model the dynamics of just one species at a time using a mean field model (McKane et al. 2000), but in fact the abundances of different species are not independent of each other (Etienne and Olff 2004), especially if a zero-sum assumption is made. It is unclear how big an effect the non-independence of species has in the real world, although it clearly becomes bigger as the number of species and individuals becomes smaller.

Is the model analytical or Monte Carlo?—Two general modeling techniques exist, analytical solutions that produce formulas, and Monte Carlo simulations on a computer that use random numbers. Both of these techniques have been used to analyze neutral models, although to date only Monte Carlo models have been used for spatially explicit metacommunities.

WHAT PREDICTIONS OF NEUTRAL MODELS CAN WE TEST?

Considering different neutral models leads to various predictions, some common to all models and others that are specific to particular models. These predictions can be grouped into three categories.

Species abundance distribution (SAD).—All of the neutral models predict the shape of the species abundance distribution, which is a histogram of the number of individuals of various species. Empirical data have suggested for decades that in almost all communities the SAD is hyperbolic (hollow-curved) and modal on a log-abundance scale (Motomura 1932, Raunkiaer 1934, Preston 1948). Neutral theory predicts a probability distribution known as the Ewens distribution in molecular neutral theory (Johnson et al. 1997), and with some changes due to the addition of migration and a metacommunity, the zero-sum multinomial distribution (ZSM) in neutral ecology (Hubbell 2001). Although the various neutral models frequently derive slightly different forms of the ZSM depending on their assumptions, it should be emphasized that biologically these distinctions are small. We use the abbreviations anZSM and mcZSM to describe analytically calculated vs. Monte Carlo versions of the ZSM.

Dispersal limitation.—Most neutral models predict that species composition changes across space because species have limited dispersal distances, which becomes more acute when the species is rare. This is in contrast to the traditional assumption that changes in community structure are due to adaptations of different species to

TABLE 1. Extended.

Analytical or Monte Carlo?	Predicts SAD?	Predicts SAR?	Other predictions
analytical	yes	collector	Shannon-Weaver diversity
Monte Carlo	yes	no	time to extinction
Monte Carlo	yes	collector and SAR	phylogeny
Monte Carlo	yes	collector	many
Monte Carlo	yes	collector and SAR	many
analytical	yes	no	left skew increases as migration decreases
analytical	yes	no	dynamics of SAD
Markov Chain Monte Carlo on analytical formula	yes	no	none
analytical	yes	no	none
analytical	yes	no	none

local conditions (Whittaker 1975). The exact form of this prediction varies, including predictions about species–area relationships (SARs), the decay of similarity with distance, beta diversity, and so on. But they all trace to the same underlying process and we collectively label these predictions "dispersal limitation." There are also clearly non-neutral invocations of dispersal limitation (e.g., metapopulation dynamics). It is worth noting that predictions of SARs may truly involve spatial structure or may just predict diversity as a function of number of individuals sampled independently of spatial structure. This second prediction is more properly called a collectors curve and is really a prediction about SADs and their sampling (Ugland et al. 2003), not dispersal limitation.

Miscellaneous.—Neutral theory is such a powerful, self-consistent, and elegant theory that a great many predictions can be derived. Thus, there are a large number of predictions made by only one or two of the models (although these predictions presumably could be made by other models). These are summarized in Table 1 and discussed further in Appendix A.

Thus, despite the diversity of models and assumptions, neutral theory appears to have converged on two main predictions and a number of auxiliary predictions.

HOW GENERAL ARE NEUTRAL MODELS?

Hubbell (2001) claimed that neutral theory provides a "unified theory of biodiversity and biogeography." However, it is unlikely that a single theory in ecology applies at all spatial, temporal, and taxonomic scales. Tests at inappropriate scales or scopes are unreasonable tests. Moreover, the degree of universality is central to judging the importance of neutral theory.

There are differing viewpoints on the scales across which neutral theory applies. Hubbell (2001) and Bell (2003) presented empirical data taken from spatial scales ranging from 0.25 m^2 to the entire British Isles or a substantial chunk of the Pacific Ocean and taxonomic groups ranging from copepods to herbs to trees to bats to bees, and they suggest that the neutral model is explanatory for these data. Others suggested limits to the applicability of neutral theory. Chave (2004) suggested, "neutral theory is explicitly concerned with communities of resident organisms (like plants, corals, or non-migratory animals)." Hubbell (2001:6) defined neutrality as applying to a "trophically defined community" (2001:54, 312), but later suggested that neutral theory can apply across multiple trophic guilds (2001:312). The main limit Hubbell (2001:54, 312) placed is that neutral theory applies only when the assumption of zero-sum dynamics (constant population size) is met. He gave three examples (2001:54) of when violation of the zero-sum dynamic assumption can be expected: variation in resource supply across space, severe disturbance regimes, or "attempting to aggregate taxa that are trophically too dissimilar." Hubbell (2001:55) explicitly stated that zero-sum dynamics does not require a constant carrying capacity, allowing temporal heterogeneity.

There are several problems with this seemingly precise statement. First, Bell (2000, 2001, 2003) allows populations to fluctuate below carrying capacity and yet produces similar results. Caswell (1976) demonstrates that the neutral results for constant population vs. infinitely growing populations are very similar. In a paper with Volkov and colleagues, Hubbell (Volkov et al. 2003) described a neutral model with metacommunities that grow without limit but converged to the same SAD as his earlier finite metapopulation (Hubbell 2001). Finally, Alonso and McKane (2004) showed that all of the anZSM models place no requirement of spatial proximity of a local community—a local community of size J is mathematically equivalent to any random, independent sample of size J from the metacommunity (albeit the migration parameter may change). Thus zero-sum dynamics may not be as important as initially thought. This leaves no model assumptions that point to specific organisms. Either the neutral theory should apply generally to all organisms, or we will have to determine empirically which organisms behave neutrally.

Chave (2004) has attempted to empirically measure which organisms exhibit equivalent per-capita fitness to answer this question. But caution is needed. Niche-based scenarios can generate per-capita fitness equivalence as well. Imagine one species per niche in a set of non-overlapping niches, with each species at the carrying capacity for that niche. In this scenario, each species has equivalent fitness but drift has played no role in determining abundances.

Beyond the question of what taxa might exhibit neutrality, there is the practical question of what is a reasonable size for a local community (Hubbell 2001:313)? A more useful phrasing of this question is "over what scales is the variation among species in adaptation to heterogeneous environmental conditions irrelevant to population dynamics?" If adaptation to heterogeneous environmental conditions matters, then neutrality is broken; fitness at a given site is not equal among species, and abundance across sites is driven by the frequency of underlying environmental conditions, not drift. We suggest that the importance of adaptation to environmental variation has been proven true at the very smallest scales (sites for single organisms) or at the global scales (tundra vs. tropics). It is possible, but unproven, that neutrality to heterogeneity occurs in intermediate scales. Another approach based on a well-known result out of population genetics (Roughgarden 1979), is that selection dominates when $4Ns \gg 1$ and drift dominates when $4Ns \ll 1$ (where N is population size or J in our notation, and s is strength of selection or variation in adaptation to environmental heterogeneity). Thus the spatial scales at which neutrality dominates may depend on population size, N. This can be surprisingly restrictive—if one species is just 1% better adapted to a particular site than another species, then neutrality applies only if $N \ll 25$ individuals (see also Fuentes 2004).

In summary, we do not believe, contrary to previous claims, that there are good modeling grounds for predicting which taxonomic groups will be neutral. Further, neutrality should apply mostly to spatial scales where the effects of environmental heterogeneity are not important (if such scales exist) and where N is small enough for $4Ns \ll 1$.

How Do We Test Large-Scale Theories Such as Neutral Models?

The method of testing a theory in ecology is well developed for small spatial and temporal scales: use manipulative experiments with the obvious null hypothesis that the manipulation has no effect. Appropriate methods of testing large-scale theories where such experiments are impossible are less clear (Diamond 1986, Brown and Maurer 1989). For example, Marquet et al. (2003) identified more than 30 different mechanistic explanations for SADs. All provide an adequate fit to empirical data, and most have not been decisively rejected. In particular, a number of theories based on niche partitioning (Lewontin et al. 1978, Sugihara 1980, Ginzburg et al. 1988, Tokeshi 1993, Chave et al. 2002, Wilson et al. 2003) and neutral theory (Hubbell 2001) both produce SADs that fit empirical data fairly well.

McGill (2003*b*) suggested that there are various levels of rigor of testing macroecological theories. These levels dealt exclusively with the scenarios of no null hypothesis or a statistical null hypothesis. Here we extend this framework to include modes of scientific inference based on contrasting multiple plausible hypotheses (Platt 1964, Hilborn and Mangel 1997). It has been argued that statistical null hypotheses are "trivial" and necessarily false and therefore provide a weak mode of inference (Quinn and Dunham 1983, Yoccoz 1991, Anderson et al. 2000). Stronger modes of inference involve multiple potentially realistic alternate hypotheses (Platt 1964, Hilborn and Mangel 1997, Burnham and Anderson 1998). Thus we propose four levels of increasing strength of inference:

Level A) *Single theory test*.—This level tests whether empirical data are consistent with or contradictory to a theory. It contains two sublevels.

A1) *Correct shape approach*.— The model displays the same general shape or relation as empirical data (e.g., the hollow curve of species abundance distributions).

A2) *Curve-fitting approach*.—The model fits the empirical data well when the parameters are chosen via curve fitting.

Level B) *Null hypothesis test*.—This level tests whether the theory fits empirical data better than a null hypothesis (H_0) representing a simple scenario.

B1) *Hypotheticodeductive approach*.—The model fits the empirical data significantly better than H_0 after penalization for number of parameters.

B2) *A priori parameters approach*.—The model fits the empirical data significantly better than H_0 after penalization for number of parameters and when the model parameters are chosen independently of the empirical data.

Level C) *Multiple, complex predictions approach*.—A single model is tested using multiple a priori predictions. These predictions are more complex than the data-fitting predictions tested in levels A and B. Two examples of complex predictions are predictions of correlations and predictions about dynamic processes. Each of the complex predictions is then tested at least at level B (i.e., against an appropriate H_0).

Level D) *Model comparison test*.—Realistic alternate models (as opposed to null hypotheses) are contrasted against each other. Platt (1964) suggested this leads to strong inference and rapid progress in science. Rejecting a theory relative to a null hypothesis leads to little advancement in scientific knowledge (H_0 is not then accepted). By selecting among alternative theories, we are necessarily progressing, even if further refinement is needed (Platt 1964). In the context of neutral theory, alternative models come from niche or adaptive theory.

D1) *Best theory approach*.—The alternative models are ranked according to their match to empirical reality according to some score (such as r^2), and the best model is selected.

D2) *Last standing approach*.—Rigorous attempts are made to falsify all models in a Popperian fashion until only one model remains unfalsified, which is then accepted as the best model. This is the approach advocated by Platt (1964). It closely matches the preferred modality for NSF proposals as well.

D3) *Model weighting approach*.—This level not only involves multiple realistic theories, but assigns weights to them according to their explanatory or predictive power. The classical analysis of variance and partitioning of sums of squares is a linear example of such a technique. Akaike weights achieve a similar but not identical result (Burnham and Anderson 1998).

D4) *Conditional weights approach*.—This approach identifies how the model weights of D3 depend on the scale or context. A theory may have greater explanatory power at large spatial scales or in marine environments while another may be better for small scales or terrestrial environments. This level of model comparison answers the question, "under what conditions is one model better than another in explaining the data?"

Note that within levels A–C, the levels of tests are sequentially hierarchical; a model that does not show the correct shape (i.e., is rejected at A1) would not be supported in more rigorous tests (i.e., A2–C). The strongest support for a theory is a test at level C, but the strongest rejection of a theory is at level A1. Support for a theory at level D is not directly comparable with support for a theory at levels A–C. In our opinion, the ultimate goal of science is to develop and test theories at level D4. However, it is appropriate for a research program to start at the lower levels and work upwards.

Nearly all of the evidence presented in early tests of neutral theory (Caswell 1976, Hubbell 1979, 2001, Bell 2000, 2001, 2003) was presented at level A1 or A2, and by our current proposal must be considered weak. We give an illustrative example of exactly why we consider level A tests to be weak evidence using the ZSM SAD prediction. Specifically, level A tests (Bell 2001, Hubbell 2001) which have shown that the ZSM fits empirical data well are a simple consequence of the power of curve fitting. We demonstrate this in Fig. 1. The ZSM is unusually flexible in its ability to fit SADs (and produce successful level A tests) because each parameter is independent. J sets the scale, θ controls the shape to the right of the mode and m controls the shape to the left of the mode. Thus for any given empirical SAD, the odds that θ and m can be chosen to fit the data well are very high. Clearly stronger tests are needed.

EMPIRICAL TESTS OF NEUTRAL THEORY

Given the large number of published but as yet unsynthesized tests of neutral theory and the testing framework above, it is opportune to assess the balance of evidence. Because the tests cover many different predictions, a true meta-analysis is not possible. We limit ourselves to simply counting tests that support or contradict neutral theory, but we do this in the context of our testing hierarchy. This has two implications. First, some tests are stronger and more important than others (recall that support at high levels and rejections at low levels are strongest). Second, level D of the hierarchy makes it clear that we need to move towards tests that simultaneously evaluate multiple theories (such as neutral and niche). To date, almost all tests of neutral theory have been below level D and involve only a single hypothesis from neutral theory. We assess the consequences of this limitation in the discussion.

We list in Table 2 all tests known to us that (1) apply to the versions of neutral theory that include a metacommunity (after year 2000); (2) are at least in press or published as a dissertation and submitted for review to a journal; and (3) are level B tests or above (or rejections at level A). For sake of historical completeness, we also include a few tests that violate conditions 1 and 3, but do not count them in the summary statistics. A one-paragraph summary of each test listed in Table 2 is given in Appendix B. The outcomes are summarized in Table 3. A perusal of Table 2 and Table 3 suggests several patterns in the tests. Only the ZSM SAD prediction had any empirical support—the other predictions were always rejected. The ZSM SAD only had support in level B1 tests and there the support is mixed. This highlights three holes in the tests: (1) level B1 tests of the ZSM SAD prediction were very inconsistent, even though they often used the same data; (2) higher (level B2–C) tests of the ZSM SAD prediction were absent; and (3) level D tests were underrepresented. We address the first two holes by presenting new results in the next three sections.

LEVEL B1 TESTS OF THE ZSM SAD: NEUTRAL VS. NULL LOGNORMAL

Given the power of curve fitting to provide successful level A tests, one important and seemingly simple test is whether the ZSM fits empirical SADs at level B1 (i.e., better than a null hypothesis). Unfortunately, this seemingly simple test has in fact proved complicated. Seven different papers (McGill 2003*a*, Volkov et al. 2003, Alonso and McKane 2004, Etienne and Olff 2004, 2005, Olszewski and Erwin 2004, He 2005) have used seven different methods and obtained conflicting results. Even within one paper with consistent methods (Volkov et al. 2003, Etienne and Olff 2005, He 2005), the results

TABLE 2. Tests of the neutral model known to the authors.

Reference	Organism	Sedentary trophic guild?	Prediction tested
Zhang and Lin (1997), Yu et al. (1998), Fuentes (2004)	NA (model)	NA	genericness of neutrality
Terborgh et al. (1996), Pitman et al. (2001)	tropical trees	yes	dispersal limitation (nonexistence of everywhere abundant species)
Pandolfi (1996)	coral reef invertebrates	yes	reassembly events should be random sample of species pool
Fuller et al. (2004)	invertebrate pool detritivores	no	ZSM fits empirical SAD
Harpole and Tilman (2005)	temperate grassland plants	yes	ZSM fits empirical SAD
Condit et al. (2002)	tropical trees	yes	dispersal limitation (shape of decay of similarity with distance curve)
McGill (2003*a*)	tropical trees	yes	ZSM fits empirical SAD
	birds	no	ZSM fits empirical SAD
Volkov et al. (2003)	tropical trees	yes	ZSM fits empirical SAD
He (2005)	tropical trees, insects, birds, fish	yes/no	ZSM fits empirical SAD
Etienne and Olff (2005)	tropical trees, fish	yes/no	ZSM fits empirical SAD
Ricklefs (2003)	tropical trees	yes	speciation rates
Adler (2004)	temperate grassland plants	yes	SAR and STR
Wootton (2005)	marine intertidal invertebrates	yes	neutral model parameters from full community should predict abundances in community with dominant species removed
Clark and McLachlan (2003)	temperate trees	yes?	dispersal limitation (increasing variation with space and time)
Fuller et al. (2004)	temperate trees	no	dynamics of SAD
McGill et al. (2005)	small mammals	no	constancy of community structure in space and time
Fargione et al. (2003)	temperate grassland plants	yes	invasion by seeding should be equally successful (neutral) in all guilds despite preexisting guild structure
Harpole and Tilman (2005)	temperate grassland plants	yes	neutral abundances uncorrelated with plant traits and with site traits
Gilbert and Lechowicz (2004)	temperate trees and understory	yes	dispersal limitation (distance explains community structure better than environment)
This paper	tropical trees	yes	ZSM fits empirical SAD
This paper	tropical trees	yes	parameters meaningful
This paper	various	no	correlations in community structure

Notes: The first three rows are tests of pre-2000 neutral theories without metacommunities. The column "Sedentary trophic guild" is explained in *How general are neutral models?* The column "Test level" is addressed in *How do we test large-scale theories such as neutral models?* Abbreviations are: NA, not applicable; ZSM, zero-sum multinomial; SAD, species abundance distribution; STR, species–time relationship.

† Fit for 100 m to 50 km but did not fit for longer and shorter distances. The curve was fit to the data and therefore was likely to show a good fit for some subset of the data, which the authors conclude is best interpreted as contradicting neutral theory.

‡ ZSM fit worse than lognormal for Pasoh data; ZSM fit better than lognormal for BCI.

§ He (2005) showed that his version of the ZSM fit best for birds and fish, the Volkov version fit best on BCI, and the lognormal was best for aphids, but that both his neutral theory and the lognormal fit well and should not be differentiated on these results.

¶ There was no clear winner for ZSM vs. lognormal for BCI tropical trees, but tropical fish clearly fit ZSM better than the lognormal.

differed, for example finding the lognormal superior for Pasoh but the ZSM superior for BCI (Volkov et al. 2003).

These contradictory results might be because some data sets are non-neutral and some are neutral, but we suggest they are due to methodological issues. In this section we summarize similarities and differences in methods used, propose a set of best practices, and analyze the BCI tropical tree data set with these best practices. We summarize the issues here in the main text and provide additional technical details in Appendix C.

All authors used the lognormal distribution as a null hypothesis, but there are important differences in methods on five other points.

1) Hubbell (2001) and McGill (2003*a*) used the Monte Carlo ZSM (mcZSM), while the remaining authors used the anZSM.

2) Different authors used different measures of goodness of fit to empirical data. McGill (2003*b*) pointed out that there is no a priori best measure of fit. All commonly used measures of goodness of fit contain biases by emphasizing goodness of fit in some regions of the curve over others—choosing a measure biased in the region where a given theory fits especially well will change the results. In particular, the use of binned data on a $\log_2$ scale (Volkov et al. 2003) pays undue attention to fits of one to eight individuals at the cost of the rest of the distribution. At the same time, the

TABLE 2. Extended.

Test level	Spatial scale	Temporal scale	Favors neutral?
NA	NA	hundreds of generations	no
A1	thousands of km	point in time	no
C	tens of km	100 000 yr	no
A2	tens of cm	point in time	no
A2	several m^2	point in time	no
B1	thousands of km	point in time	yes/no†
B1	50 ha	point in time	no
B1	40 km	point in time	no
B1	50 ha	point in time	yes/no‡
B1	hectares	point in time	yes/no§
B1	varies	point in time	yes/no¶
B2	250 000 km^2	evolutionary time	no
B2	1 m^2 to 0.1 ha	dozens of years	no
B2	0.36 m^2	9 yr	no
C	hundreds of km	10 000 yr	no
C	tens of cm	9 yr	no
C	thousands of km	hundreds of thousands of years	no
C	9 m^2	4 yr	no
C	typically several m^2	point in time	no
D3	10 km^2	point in time	no
B1	50 ha	point in time	no
B2	50 ha	point in time	no
C	various	various	no

ZSM does exceptionally well in the area of one to eight individuals due to the flexibility of the m parameter focused precisely on this region (Fig. 1).

3) Parameter estimation is usually based on varying parameters to maximize some goodness of fit measure to the data. There is a certain circularity if this same measure of goodness of fit is then used to evaluate how well the curve using these parameters fits the data. One may be testing the flexibility of the distribution more than the quality of the underlying model (Fig. 1).

4) Several different versions of the lognormal have been used.

5) The authors used different burdens of proof ranging from slightly better to statistically significantly better.

Among all of these differences, we pick out the most desirable approaches, statistically speaking, and suggest current best practices:

1) use the analytical ZSM over the Monte Carlo ZSM due to the large number of simulations required to attain accuracy in the tails of the mcZSM;

2) explore goodness of fit on a variety of measures on a variety of different scales of abundance to lead to a more robust test of overall fit;

3) use parameter estimation methods that do not rely on maximizing fit according to the measures of fit used for testing to avoid valuing the flexibility of the function being fit;

4) use data collected at a single point in time without time averaging (or at least a time span small relative to the generation time);

5) assign confidence intervals to parameter estimates and/or P values to null hypothesis tests, and;

6) use the continuous, untruncated lognormal rather than Preston's approximation or the Poisson lognormal.

We note that although the method of Etienne and Olff (2004, 2005) does not conform to all of these requirements, it also seems like a very robust method that should be explored further.

To date, no test has been performed following all of these best practices. The BCI tropical tree data set (Pyke et al. 2001, Condit et al. 2002) was the subject of six tests, of which two (McGill 2003*a*, Etienne and Olff 2004) favored the lognormal, two (Volkov et al. 2003, He 2005) favored the ZSM (but the lognormal for data sets other than BCI), one found a tie (Etienne and Olff 2005), and one did not make an explicit comparison (Alonso and McKane 2004), but all of these violated at least one best practice. Here we performed a test on the BCI data using all five best practices. We implemented MATLAB computer code to replicate the Volkov et al. (2003) version of the analytical ZSM and their estimation methods (the algorithm and issues are described more fully in Appendix D and the source code is available in the Supplement). However, to meet best practice 3 we estimated parameters by maximizing generalized likelihood (Hubbell 2001:292) rather than the χ^2 score on Preston bins (Volkov et al. 2003); this had only minor effects on the outcomes. We derived very similar parameters to Volkov et al. ($m = 0.09335$ vs. Volkov's $m = 0.1$ and $\theta = 47.879$ vs. 47.226). This method actually provided a slightly better fit by the Volkov χ^2–criteria than Volkov's original estimates (3.11 vs. 3.20) and is thus more favorable to the ZSM. We used eight different measures of goodness of fit originally reported by McGill (2003*a*), and now that an analytical formula for the ZSM is available, we added log-likelihood (Hilborn and Mangel 1997, Alonso and McKane 2004). We found that using a variety of scales (arithmetic, logarithmic, data-driven or likelihood, and mixed or 10 + 1 bin) and hence arguably robust, the analytical ZSM performed worse than the lognormal on eight out of nine measures (Table 4). The measure of goodness of fit that Volkov et al. reported (χ^2 on Preston bins) was the only measure where the anZSM does better than the lognormal. Following best practice 4, we tested whether the anZSM performed statistically significantly better than the lognormal for each measure (Table 4). Necessarily $P > 0.5$ for eight measures, and $P = 0.18$ (i.e., not significantly better) for the one Preston χ^2 measure. The overall conclusion is that neither distribution is significantly better than the other

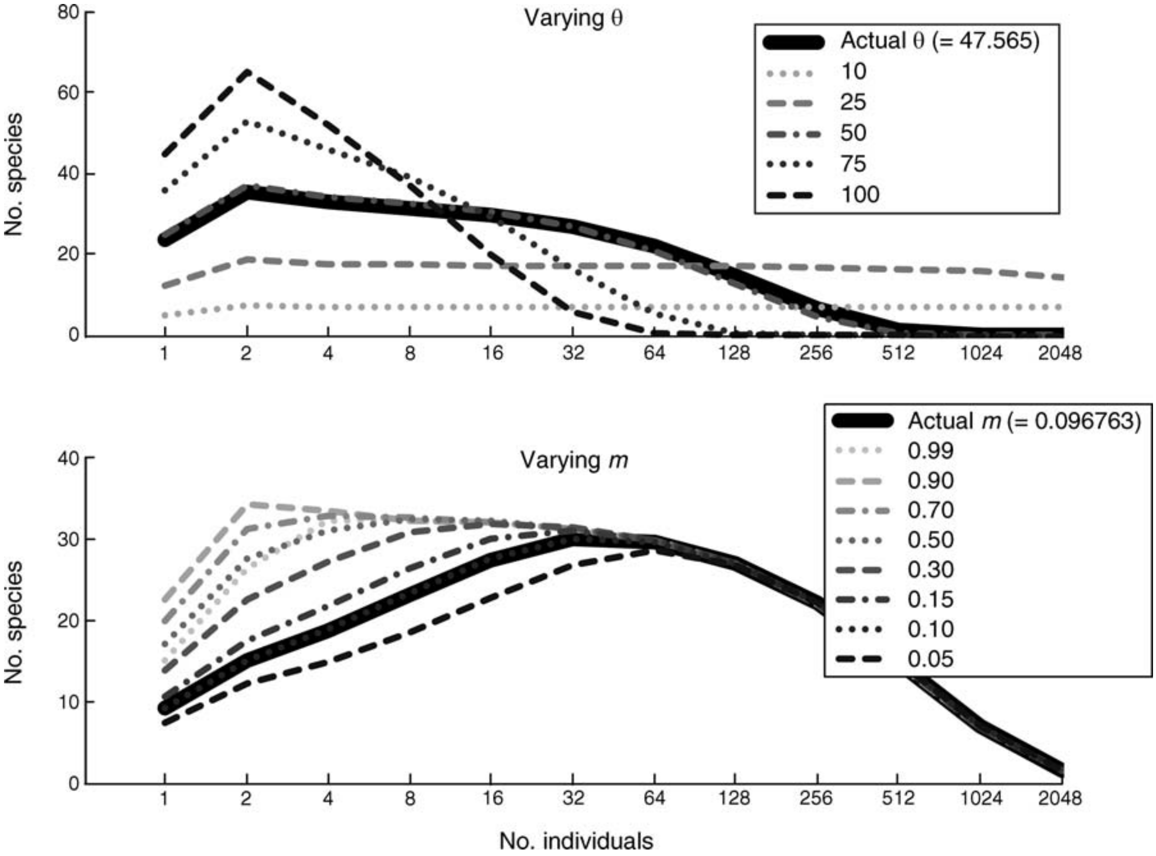

FIG. 1. Effect of parameters. This figure shows why the curve-fitting approach for the zero-sum mulitnomial (ZSM) is so successful. Both graphs use the BCI (Barro Colorado Island, Panama) tropical tree data set. The first parameter, J (local community size), sets the scale of the x-axis (and is held at $J = 20\ 000$ throughout this figure). (A) The second parameter, θ, is varied to obtain a good fit of the data on the right-hand side. Note that even if S is used as a parameter and θ is implicitly calculated (Volkov et al. 2003), the effect is the same. (B) The third parameter, m (migration), is varied to obtain a good fit on the left-hand side (with θ held constant at 47.88, the value for the BCI tropical tree data). The fit on the left-hand side (abundances of one individual to about four to eight individuals are especially well fit due to the use of m as a curve-fitting parameter). Hubbell (2001) was the first to suggest this basic approach of fitting θ to the right-hand side and then m to the left-hand side. Even if θ (or S) and m are fit simultaneously using minimization of some criterion, the effect is the same. Note the $\log_2$ x-axis scale.

although the null lognormal tended to be a better fit. This means that the ZSM SAD prediction failed a robust, best-practice level B1 test. In general level B1 tests have done little to improve our understanding of mechanisms of the SAD (Marquet et al. 2003, McGill 2003*a*, Etienne and Olff 2004), and so we turn to higher level tests in the hope that they provide clearer results.

LEVEL B2 TEST OF THE ZSM SAD: A PRIORI PARAMETERS?

A stronger test of a hypothesis is to compare empirical data to model predictions using realistic, a priori parameter estimates (i.e., without using model parameters to maximize fit). The most desirable way to develop a priori estimates is to have a mechanistic model that gives specific meaning to the parameters and allows for independent, direct measures of the appropriate parameters. The ability to do this is claimed as a strength of neutral theory (Hubbell 2001, Bell 2003, Alonso and McKane 2004). In fact, it is surprisingly difficult to measure a priori parameters for neutral theory. The parameter J, local community size, is simple to calculate. The parameter m, the percentage of born individuals whose parents reside outside the local community, is clearly defined. But it is very hard to measure m directly and there is considerable debate about the accuracy of attempts to measure m (or the equivalent for gene flow) indirectly by molecular sequencing (Whitlock and McCauley 1999). It is also well understood that m is scale dependent. Similarly, the third parameter, θ, is dimensionless, but has a direct interpretation $\theta = 2J_M\nu$ where J_M is the population size of the metacommunity and ν is the speciation rate (but see Volkov et al. 2003 where $J_M = \infty$ and $\theta = S_M P_0 \nu / b$). Thus in principle it should be possible to estimate θ. In practice, ν is rarely known and extraordinarily difficult to measure. Moreover, J_M is ill defined. While the idea of a metacommunity or regional pool is useful, it is impossible to give a precise

TABLE 3. Summary of results of empirical tests of neutral theory listed in Table 2 and their level of support.

Parameter	A2	B1	B2	C	D	All levels
ZSM SAD	0/2	5/11				5/13
Dispersal limitation			0/1	0/1	0/6	0/8
Miscellaneous			0/1	0/9		0/10
All predictions	0/2	5/11	0/2	0/10	0/6	5/31

Notes: The columns indicate the level of the test. The rows indicate the prediction tested. The cells contain the notation x/y to indicate that x out of y tests supported the neutral theory. All pre-2000 tests were omitted from this table. The tests given in this paper were included, but leaving them out did not change the results qualitatively. Similarly, breaking out tests in sedentary, single-trophic-level guilds from tests in other organisms was uninformative. The level B1 tests of the ZSM SAD (zero-sum multinomial, species abundance distribution) were counted individually (if a paper tested trees and birds it was counted as two tests). Several other papers also contain multiple tests on different groups of organisms (six in Gilbert and Lechowicz [2004], four in Fargione et al. [2003]) or on different types of gradients (three in Harpole and Tilman [2005]) and were also counted as separate tests.

physical interpretation to it—it is in reality a continuum with much contribution from nearby and little contribution from far away. Attempts to estimate θ (Ricklefs 2003, 2006) generate estimates of ν that are impossibly high.

A second, less desirable (but still at level B2), method for a priori parameter measurement is to estimate the parameters through curve fitting on one set of data and then apply them to a second, consistent set of data. Adler (2004) and Wootton (2005) have done this for non-SAD predictions of neutral theory. We attempt this for the ZSM SAD prediction using the BCI tropical tree data (Pyke et al. 2001, Condit et al. 2002). Using the source code described above, we have estimated θ and m for each of 50 1-ha plots, 10 (nonoverlapping) 5-ha plots, five 10-ha plots, and two 25-ha plots contained within the 50-ha BCI plot. The results are shown in Fig. 2.

Could the 50 1-ha plots be fit by the same parameters with some random variation? Assume a null hypothesis, H_0, that all 50 plots are from the ZSM with θ and m are chosen by taking the mean across the plots, giving $\theta = 36.33$ and $m = 0.929$. Allow J to vary to the actual value for each 1-ha plot. We then estimated the fit of the ZSM with mean parameters to the empirical data using a χ^2 statistic on Preston-binned data (the measure of fit found most favorable to neutral theory above). We found that seven of the 50 plots could reject the null hypothesis at $\alpha = 0.05$ level (or six at the $\alpha = 0.01$ level). The odds of this number or more rejections of the null hypothesis in 50 trials by chance according to the binomial distribution is $P < 0.005$ (or $P < 0.0000001$ for $\alpha = 0.01$). We conclude that the individual 50 ha plots are not random samples from the ZSM with a constant set of parameters. Interpretation of this as a deterministic response to spatial heterogeneity would also contradict neutral theory.

An examination of Fig. 2 also shows that θ systematically increases with scale while m systematically decreases with scale. The changes in θ and m with scale are each significant in an unbalanced, one-way ANOVA analysis on effect of plot size at $P < 0.001$. Does this change in parameters with scale contradict neutral theory? It certainly does for θ. Recall that $\theta = 2J_M\nu$ or equivalently $\nu = \theta/(2J_M)$. The fact that θ changes significantly between 1 ha and 50 ha (36.33 vs. 47.57) is equivalent to saying that the 50-ha plot experiences a 31% higher rate of speciation in the metacommunity than do the 1-ha plots. Yet both the 1-ha and 50-ha plots are small enough that they should have nearly identical metacommunities, since Ricklefs (2003) estimated the metacommunity for tropical trees to be roughly 25 000 000 ha or 1×10^{10} individual trees. Unlike θ, m should change with scale, and it changes in the right direction in our analysis (m gets smaller, meaning fewer individuals come from outside the local community as

TABLE 4. Comparison of lognormal and ZSM distributions for nine different measures of goodness of fit on the BCI tropical tree data.

Measure	Lognormal	ZSM	anZSM better than lognormal?	P	anZSM statistically significantly better?
r^2	0.9996	0.9992	no	0.5004	no
r^2 mean corrected	0.9960	0.9923	no	0.5039	no
r^2 correlation	0.9966	0.9955	no	0.8410	no
K (Kolmogorov-Smirnov)	0.0565	0.0805	no	N/A	no
ln(likelihood/AIC)	−1157.01	−1161.08	no	0.9956	no
χ^2 arithmetic	12.61	27.21	no	0.8342	no
χ^2 Preston $\log_2$	5.83	3.11	yes	0.1812	no
χ^2 10 + 1 bins	10.26	17.31	no	0.7465	no
χ^2 arithmetic 5 bins	4.57	10.93	no	0.7054	no

Notes: Measures are the same as McGill (2003*a*) with ln(likelihood) added. The first two columns give the measures of fit for the lognormal and analytical ZSM (zero-sum multinomial). The third column indicates whether the ZSM had a score indicating a better fit. The fourth column gives statistical significance (probability that this difference could be observed if lognormal is correct). These are stated so that $P < 0.05$ indicates the anZSM was statistically significantly better than the lognormal, while a $P > 0.95$ indicates the lognormal was statistically significantly better than the anZSM. A $P = 0.50$ indicates that the theories are equally likely. Note that this P statistic begins to take on the flavor of a level C3 test although it is not a true partitioning of variance. Values of P were calculated using the F statistic for ratios of χ^2 and r^2, using the χ^2 asymptotic distribution of the AIC for likelihood, and using the Fisher z transformation for correlation for the r^2 correlation statistic.

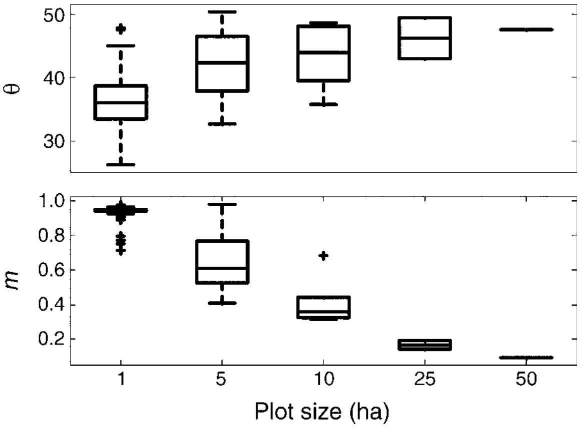

FIG. 2. Variation in parameters with sample plot size. These plots are all from within the 50-ha BCI tropical tree plot (Condit et al. 1996, Pyke et al. 2001). Parameters were calculated for 50 1-ha subplots, 10 5-ha subplots, and so on. Both plots are box plots with the top and bottom of the box describing the upper and lower quartile markers of the data, the horizontal line representing the median, and the bars describing the range of the data. Plus marks represent outliers. (A) Estimates of θ. (B) Estimates of the migration parameter, m.

the local community gets bigger). An exact calculation of m requires knowledge of the average dispersal kernel for trees at BCI and a fairly complicated model. However, a simple and crude geometric model (Fig. 3) suggests that m does not scale correctly. Specifically, stepping from 25 ha to 50 ha, then $(1 - m)$ should go from 83% to less than 87% ($m = 0.17, f = 0.75$) when in fact it goes to 91% (more than twice as high). In short, $(1 - m)$ seems to increase in this case faster than a literal physical interpretation of m would allow. Thus, neutral theory fails to have parameters that remain constant across consistent data. This makes it impossible for neutral theory to have successful, confirmatory test at level B2 (at least for the BCI tropical tree data used herein).

LEVEL C TESTS OF THE ZSM SAD

Continuing with our objective of adding higher level tests for the ZSM SAD prediction, we suggest that there are obvious level C tests of neutral theory that use previously published empirical data and have been overlooked. Specifically neutral models predict that any given species trait (e.g., body size) should be uncorrelated with abundance in the local community since abundance in the local community is determined entirely through drift. Instead, the distribution of any species trait (e.g., body size) should be merely a random sampling from the metacommunity distribution (Dolman and Blackburn 2004, Harpole and Tilman 2005), which in turn, under the metacommunity neutral model, should show no correlation with metacommunity abundances. There are a number of well known studies that falsify these predictions. Body mass in local communities is not a random sample of the metacommunity (Brown and Nicoletto 1991, Marquet and Cofre 1999). Body mass is correlated significantly with abundance at both local (Damuth 1981, 1991, Enquist et al. 2001) and metacommunity scales (Nee et al. 1991). Likewise, assembly rules cause local assemblages to have a taxonomic structure that is not representative of the taxonomic structure of the metacommunity (Diamond 1975, Fox 1987, Gotelli and Graves 1996, Kelt and Brown 1999, Weiher and Keddy 1999). These patterns are all in clear contradiction of the neutral theory prediction that species traits will be random within a local community (within the constraints of the metacommunity).

CONCLUSIONS AND PROSPECTUS

We set out a context in which to test neutral theory and then analyzed published empirical tests within this context. We suggested that although there are a diversity of neutral theories, they make similar predictions and that two nearly universal predictions are the ZSM SAD and dispersal limitation. We also argued that based on model assumptions, the neutral theory should apply to all organisms but the relevant spatial scales are hard to

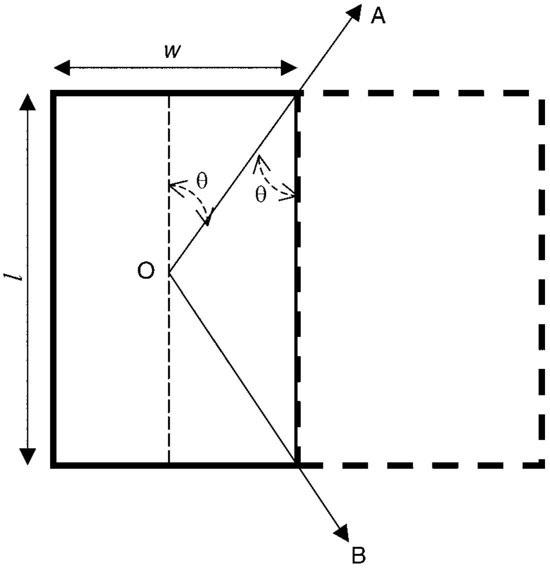

FIG. 3. Crude geometric model for scaling of m. Imagine we have a rectangular plot of width w and length l. Assume that this is adjacent to a second plot of the same shape. Note that $\tan(\theta) = (w/2)/(l/2) = w/l$. Because alternate exterior angles are congruent, the interior angle for the two rays (AOB) will be $180 + 2\theta = 180 + 2\arctan(w/l)$. The fraction of the external propagule rain that comes from outside the second, adjacent, rectangular plot will be greater than the fraction $f = (180 + 2\arctan(w/l))/360$. Now take m_1 as the fraction of individuals in the target plot that arrive from outside the plot. Then $(1 - m_1)$ individuals are born from inside the target plot. The fraction of individuals arriving in the target plot from the adjacent plot $= m_{\text{adj}} < m_1(1 - f)$. Then the estimate for m_2, the fraction of individuals born outside the combined plot of two rectangles, can be derived from looking at the fraction of individuals that come from inside the combined plots $(1 - m_2)$. But $(1 - m_2) = 2([1 - m_1] + m_{\text{adj}})/2 = (1 - m_1) + m_{\text{adj}} < (1 - m_1) + (1 - f)m_1$.

determine. Empirical testing should be helpful in identifying what subsets of organisms and spatial scales show neutrality. We proposed that there are varying levels of strength of prediction and test, and we have argued that neutral theory started with weak tests and is now progressing to stronger tests. Finally we showed that published empirical tests overwhelmingly reject neutral theory predictions other than the ZSM SAD. Weaker tests were more likely to support the predictions of neutral theory.

What should ecologists do with neutral theory at this point given the current negative balance of evidence? We are in agreement with Lakatos (1978) "so-called 'refutations' are not the hallmark of empirical failure, as Popper has preached, since all programmes grow in a permanent ocean of anomalies." But in the next sentence, Lakatos suggests, "What really counts are dramatic, unexpected, stunning predictions: a few of them are enough to tilt the balance." Originally, it appeared that the good fit of the empirical data (level A tests) was such a stunning prediction. But with careful analysis we see that neutral theory falls far short of dramatic on this front. Indeed there is very little weight of empirical support for neutral theory to date. The primary research program for neutral theory is to develop these "dramatic, unexpected, stunning predictions" (Nee and Stone 2003).

Another important research direction (Condit et al. 1996, Fuller et al. 2005) for neutral theory is to move towards level D tests. It is one thing to falsify neutral theory, but science advances more if we show that adaptive niche theory (or some alternative) can explain where the neutral fails (a level D test). Rejecting neutral theory at levels A–C does not in itself support niche theory; a level D test is needed. Clark and McLachlan (2003) and McGill et al. (2005) have come close to a level D test, but neither study produced quantitative estimates of the niche model predictions. Gilbert and Lechowicz (2004) are unique to date in producing a Level D3 test. Their results suggest that neutral theory explains a very small percentage of variance in comparison to niche theory. Should this result prove general, then there may be little point in advancing to Level D. More work at these more sophisticated model-testing levels is needed.

Despite all of this, neutral theory remains an extraordinarily and perhaps uniquely elegant ecological theory. Such elegance is indicative that neutral theory will have some important role in ecology. Debate over the "correct" null hypothesis (Gotelli and Graves 1996) has made several important research programs such as assembly rules (Weiher and Keddy 1999) and nestedness (Wright et al. 1998) so complex that few dare to use these tools. We suggest that even should neutral theory turn out to explain a relatively small fraction of variance, it is still an extraordinarily good null hypothesis for a very wide variety of questions. We emphatically note that we use the term "null" theory here in its strict, statistical sense of the H_0, a hypothesis to be rejected by an alternative theory under study; we explicitly do not mean the wider sense of "null" hypothesis which has come to mean a mechanistic theory based on stochastic principles. Using the neutral theory as an alternative to niche theory in a level D test accomplishes much the same effect. For neutral theory to succeed as either a null (Enquist et al. 2002) or alternative theory, research needs to focus on standardizing tools for implementing it, understanding the implications of assumptions behind the analytical ZSM solutions, and the publication of computer code so that a public debate can be held about the merits of approaches. To this end we have provided computer code that implements the spatially explicit metacommunity (McGill et al. 2005), the mcZSM (McGill 2003*a*), and the anZSM in the Supplement.

ACKNOWLEDGMENTS

Special thanks to Peter Adler, Mike Fuller, Ben Gilbert, Marcelino Fuentes, and Stan Harpole for sharing unpublished manuscripts with us. Many thanks also to Peter, Marcelino, and Marty Lechowicz for helpful comments on this manuscript. Marcel Holyoak gave many helpful suggestions and spent time above and beyond the call of an editor to improve this manuscript. We thank Maria Dornelas, Gyorgy Abrusan, Joan Tracey, and Eugene Gallagher for testing preliminary versions of the source code for fitting the analytical ZSM. B. J. McGill was funded by an NSF Interdisciplinary Informatics Postdoctoral Fellowship.

LITERATURE CITED

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

Alonso, D., and A. J. McKane. 2004. Sampling Hubbell's neutral theory of biodiversity. Ecology Letters **7**:901–910.

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management **64**:912–923.

Bell, G. 2000. The distribution of abundance in neutral communities. American Naturalist **155**:606–617.

Bell, G. 2001. Neutral macroecology. Science **293**:2413–2418.

Bell, G. 2003. The interpretation of biological surveys. Proceedings of the Royal Society of London B **270**:2531–2542.

Brown, J. H., and B. A. Maurer. 1989. Macroecology: the division of food and space among species on continents. Science **243**:1145–1150.

Brown, J. H., and P. F. Nicoletto. 1991. Spatial scaling of species composition: body masses of North American land mammals. American Naturalist **138**:1478–1512.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer, New York, New York, USA.

Caswell, H. 1976. Community structure: a neutral model analysis. Ecological Monographs **46**:327–354.

Chase, J. M., and M. A. Liebold. 2003. Ecological niches: linking classical and contemporary approaches. University of Chicago Press, Chicago, Illinois, USA.

Chave, J. 2004. Neutral theory and community ecology. Ecology Letters **7**:241–253.

Chave, J., H. C. Muller-Landau, and S. A. Levin. 2002. Comparing classical community models: theoretical consequences for patterns of diversity. American Naturalist **159**:1–23.

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

Clark, J. S., and J. S. McLachlan. 2004. The stability of forest biodiversity: reply. Nature **427**:696–697.

Condit, R., S. P. Hubbell, J. V. LaFrankie, R. Sukumar, N. Manokaran, R. B. Foster, and P. S. Ashton. 1996. Species–area and species–individual relationships for tropical trees: a comparison of three 50-ha plots. Journal of Ecology **84**:549–562.

Condit, R., N. Pitman, E. G. Leigh, J. Chave, J. Terborgh, R. B. Foster, P. Nunez, S. Aguilar, R. Valencia, G. Villa, H. C. Muller-Landau, E. Losos, and S. P. Hubbell. 2002. Beta-diversity in tropical forest trees. Science **295**:666–669.

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

Damuth, J. 1991. Of size and abundance. Nature **351**:268–269.

Diamond, J. 1986. Overview: laboratory experiments, field experiments, and natural experiments. Pages 3–22 *in* J. Diamond and T. J. Case, editors. Community ecology. Harper and Row, Cambridge, UK.

Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.

Dolman, A. M., and T. M. Blackburn. 2004. A comparison of random draw and locally neutral models for the avifauna of an English woodland. BMC Ecology ⟨http://www.biomedcentral.com/1472–6785/4/8⟩

Elton, C. 1927. Animal ecology. University of Chicago Press, Chicago, Illinois, USA.

Enquist, B. J., J. H. Brown, and G. B. West. 2001. Allometric scaling of plant energetics and population density. Nature **395**:163–165.

Enquist, B. J., J. Sanderson, and M. D. Weiser. 2002. Modeling macroscopic patterns in ecology. Science **295**:1835–1837.

Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. Ecology Letters **8**:253–260.

Etienne, R. S., and H. Olff. 2004. A novel genealogical approach to neutral biodiversity theory. Ecology Letters **7**: 170–175.

Etienne, R. S., and H. Olff. 2005. Confronting different models of community structure to species–abundance data: a Bayesian model comparison. Ecology Letters **8**:493–504.

Fargione, J., C. S. Brown, and D. Tilman. 2003. Community assembly and invasion: an experimental test of neutral vs. niche processes. Proceedings of the National Academy of Sciences (USA) **100**:8916–8920.

Fox, B. J. 1987. Species assembly and the evolution of community structure. Evolutionary Ecology **1**:201–213.

Fuentes, M. 2004. Slight differences among individuals and the unified neutral theory of biodiversity. Theoretical Population Biology **66**:199–203.

Fuller, M. M., T. N. Romanuk, and J. Kolasa. 2005. Effects of predation and variation in species relative abundance on the parameters for the neutral models. Community Ecology **6**: 229–240.

Gause, G. F. 1934. The struggle for existence. Williams and Wilkins New York, New York, USA.

Gilbert, B., and M. J. Lechowicz. 2004. Neutrality, niches, and dispersal in a temperate forest understory. Proceedings of the National Academy of Sciences (USA) **101**:7651–7656.

Ginzburg, L. R., H. R. Akcakaya, and J. Kim. 1988. Evolution of community structure: competition. Journal of Theoretical Biology **133**:513–523.

Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington, D.C., USA.

Gregory, R. D., and K. J. Gaston. 2000. Explanations of commonness and rarity in British breeding birds: separating resource use and resource availability. Oikos **88**:515–526.

Harpole, W. S., and D. Tilman. 2005. Non-neutral patterns of species abundances and species traits. Ecology Letters **9**:15–23.

He, F. 2005. Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics. Functional Ecology **19**:187–193.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Holyoak, M., and M. Loreau. 2006. Reconciling empirical ecology with neutral community models. Ecology **87**:1370–1377.

Hubbell, S. P. 1979. Tree dispersion, abundance and diversity in a tropical dry forest. Science **203**:1299–1309.

Hubbell, S. P. 2001. A unified theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.

Hubbell, S. P., and R. B. Foster. 1986. Biology, chance, and history and the structure of tropical rain forest tree communities. Pages 314–328 *in* J. Diamond and T. Case, editors. Community ecology. Harper and Row, Cambridge, Massachusetts, USA.

Johnson, N. L., S. Kotz, and N. Balkrishnan. 1997. Discrete multivariate distributions. John Wiley and Sons, New York, New York, USA.

Kelt, D. A., and J. H. Brown. 1999. Community structure and assembly rules: confronting conceptual and statistical issues with data on desert rodents. Pages 58–74 *in* E. Weiher and P. A. Keddy, editors. Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press, Cambridge, UK.

Lakatos, I. 1978. Introduction: science and pseudoscience. Pages 1–8 *in* J. Worrall and G. Currie, editors. The methodology of scientific research programs. Cambridge University Press, Cambridge, UK.

Leibold, M. A., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and A. Gonzalez. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters **7**:601–613.

Lewontin, R. C., L. R. Ginzburg, and S. D. Tuljapurkar. 1978. Heterosis as an explanation for large amounts of genic polymorphism. Genetics **88**:149–170.

MacArthur, R. H. 1972. Geographical ecology: patterns in the distribution of species. Princeton University Press, Princeton, New Jersey, USA.

Marquet, P. A., and H. Cofre. 1999. Large temporal and spatial scales in the structure of mammalian assemblages in South America: a macroecological approach. Oikos **85**:299–309.

Marquet, P. A., J. A. Keymer, and C. Hernan. 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 and K. J. Gaston, editors. Macroecology: concepts and consequences. Blackwell Science, Oxford, UK.

Maurer, B. A. 1999. Untangling ecological complexity. University of Chicago Press, Chicago, Illinois, USA.

McGill, B. J. 2003*a*. A test of the unified neutral theory of biodiversity. Nature **422**:881–885.

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

McGill, B. J., E. A. Hadly, and B. A. Maurer. 2005. Community inertia of Quaternary small mammal assemblages in North America. Proceedings of the National Academy of Sciences (USA) **102**:16701–16706.

McKane, A., D. Alonso, and R. V. Sole. 2000. Mean-field stochastic theory for species-rich assembled communities. Physical Review E **62**:8466–8484.

McKane, A., D. Alonso, and R. V. Sole. 2004. Analytic solution of Hubbell's model of local community dynamics. Theoretical Population Biology **65**:67–73.

McKane, R. B., L. C. Johnson, G. R. Shaver, K. J. Nadelhoffer, E. B. Rastetter, B. Fry, A. E. Giblin, K. Kietland, B. L. Kwiatkowski, J. A. Laundre, and G. Murray.

2002. Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. Nature **413**:68–71.

Motomura, L. 1932. On the statistical treatment of communities. Zoological Magazine **44**:379–383.

Nee, S., A. F. Read, J. J. D. Greenwood, and P. H. Harvey. 1991. The relationship between abundance and body size in British birds. Nature **351**:312–313.

Nee, S., and G. Stone. 2003. The end of the beginning for the neutral theory. Trends in Ecology and Evolution **18**:433–434.

Odum, E. P. 1959. Fundamentals of ecology. Second edition. W B Saunders, Philadelphia, Pennsylvania, USA.

Olszewski, T. D., and D. H. Erwin. 2004. Dynamic response of Permian brachiopod communities to long-term environmental change. Nature **428**:738–741.

Pandolfi, J. 1996. Limited membership in Pleistocene reef coral assemblages from the Huon Peninsula, Papua New Guinea: constancy during global change. Paleobiology **22**:152–176.

Peterson, A. T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservativism of ecological niches in evolutionary time. Science **285**:1265–1267.

Pianka, E. R. 1988. Evolutionary ecology. Fourth edition. Harper and Row, Cambridge, Massachusetts, USA.

Pitman, N. C. A., J. W. Terborgh, M. R. Silman, P. Nunez, D. A. Neill, C. E. Ceron, W. A. Palacios, and M. Aulestia. 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. Ecology **82**:2101–2117.

Platt, J. R. 1964. Strong inference. Science **146**:347–353.

Preston, F. W. 1948. The commonness and rarity of species. Ecology **29**:254–283.

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.

Quinn, J. F., and A. E. Dunham. 1983. On hypothesis testing in ecology and evolution. American Naturalist **122**:602–617.

Raunkiaer, C. 1934. The life forms of plants and statistical plant geography; begin the collected papers of C. Raunkiaer. Clarendon, Oxford, UK.

Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes. Science **235**:167–171.

Ricklefs, R. E. 2003. A comment on Hubbell's zero-sum ecological drift model. Oikos **100**:185–192.

Ricklefs, R. E. 2006. The unified theory of biodiversity: do the numbers add up? Ecology **87**:1424–1431.

Roughgarden, J. 1979. Theory of population genetics and evolutionary ecology: an introduction. First edition. MacMillan Publishing, New York, New York, USA.

Silvertown, J., M. E. Dodd, D. J. G. Gowing, and J. O. Mountford. 1999. Hydrologically defined niches reveal a basis for species richness in plant communities. Nature **400**: 61–63.

Sugihara, G. 1980. Minimal community structure: an explanation of species–abundance patterns. American Naturalist **116**:770–787.

Terborgh, J., R. B. Foster, and P. V. Nunez. 1996. Tropical tree communities: a test of the nonequilibrium hypothesis. Ecology **77**:561–567.

Tokeshi, M. 1993. Species abundance patterns and community structure. Advances in Ecological Research **24**:111–186.

Ugland, K. I., J. S. Gray, and K. E. Ellingsen. 2003. The species–accumulation curve and estimation of species richness. Journal of Animal Ecology **72**:888–897.

Vallade, M., and B. Houchmandzadeh. 2003. Analytical solution of a neutral model of biodiversity. Physical Review E 68:061902_061901–061902_061905.

Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan. 2003. Neutral theory and relative species abundance in ecology. Nature **424**:1035–1037.

Weiher, E., and P. A. Keddy, editors. 1999. Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press, Cambridge, UK.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: Fst <> 1/(4Nm+1). Heredity **82**: 117–125.

Whittaker, R. H. 1975. Communities and ecosystems. Second edition. MacMillan Publishers, New York, New York, USA.

Wilson, W. G., P. Lundberg, D. P. Vazquez, J. B. Shurin, M. D. Smith, W. Langford, K. L. Gross, and G. G. Mittelbach. 2003. Biodiversity and species interactions: extending Lotka-Volterra community theory. Ecology Letters **6**:944–952.

Wisheu, I. C. 1998. How organisms partition habitats: different types of community organization can produce identical patterns. Oikos **83**:246–258.

Wootton, J. T. 2005. Field parameterization and experimental test of the neutral theory of biodiversity. Nature **433**:309–312.

Wright, D. H., B. D. Patterson, G. M. Mikkelson, A. Cutler, and W. Atmar. 1998. A comparative analysis of nested subset patterns of species composition. Oecologia **113**:1–20.

Yoccoz, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America **72**:106–111.

Yu, D. W., J. Terborgh, and M. D. Potts. 1998. Can high tree species richness be explained by Hubbell's null model? Ecology Letters **1**:193–199.

Zhang, D. Y., and K. Lin. 1997. The effects of competitive asymmetry on the rate of competitive displacement: how robust is Hubbell's neutral theory. Journal of Theoretical Biology **188**:361–367.

APPENDIX A

Detailed comparison of neutral models (*Ecological Archives* E087-080-A1).

APPENDIX B

Summaries of empirical tests (*Ecological Archives* E087-080-A2).

APPENDIX C

Details of statistical methods for comparing SAD fits (*Ecological Archives* E087-080-A3).

APPENDIX D

A description of the algorithms and usage of code for fitting analytical ZSM (*Ecological Archives* E087-080-A4).

SUPPLEMENT 1

The Matlab source code containing the algorithm described in Appendix D to fit the analytical ZSM (*Ecological Archives* E087-080-S1).