Comparing process-based and constraint-based approaches for modeling macroecological patterns

Xiao Xiao,[1,2,3,4,8] James P. O'Dwyer,[5] and Ethan P. White[1,2,6,7]

[1]*Department of Biology, Utah State University, Logan, UT 84322-5305, USA*
[2]*Ecology Center, Utah State University, Logan, UT 84322-5205, USA*
[3]*School of Biology and Ecology, University of Maine, Orono, ME 04469, USA*
[4]*Senator George J. Mitchell Center for Sustainability Solutions, University of Maine, Orono, ME 04469, USA*
[5]*Department of Plant Biology, University of Illinois, Urbana, IL 61801, USA*
[6]*Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, 32611, USA*
[7]*Informatics Institute, University of Florida, Gainesville, FL 32611, USA*

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 numerical constraints. Differentiating the extent to which patterns in a given system are determined statistically, and where it requires explicit ecological processes, has been difficult. We tackled this challenge by directly comparing models from a constraint-based theory, the Maximum Entropy Theory of Ecology (METE) and models from a process-based theory, the size-structured neutral theory (SSNT). Models from both theories were capable of characterizing the distribution of individuals among species and the distribution of body size among individuals across 76 forest communities. However, the SSNT models consistently yielded higher overall likelihood, as well as more realistic characterizations of the relationship between species abundance and average body size of conspecific individuals. This suggests that the details of the biological processes contain additional information for understanding community structure that are not fully captured by the METE constraints in these systems. Our approach provides a first step towards differentiating between process- and constraint-based models of ecological systems and a general methodology for comparing ecological models that make predictions for multiple patterns.

Key words: constraints; maximum entropy theory of ecology; mechanisms; model comparison; neutral theory; processes.

Introduction

Patterns of biodiversity that are aggregated across large numbers of individuals often take similar shapes across ecosystems and taxonomic groups (Brown 1995). Understanding why such patterns seem to be universal, for example the skewed distribution of individuals among species (the species abundance distribution) (Fisher et al. 1943, McGill et al. 2007) and the uneven allocation of body size among individuals (the individual size distribution) (Enquist and Niklas 2001, Muller-Landau et al. 2006b), is one of the central pursuits of macroecology (Brown 1999, McGill and Nekola 2010).

This task is not trivial because common patterns are often associated with multiple models that have different assumptions about mechanisms yet make similar or even identical predictions (Frank 2014). For example, more than 20 models exist for the species-abundance distribution (SAD) all making realistic predictions with many rare species and a few abundant ones, but with mechanisms ranging from purely statistical to population dynamics to resource partitioning (Marquet et al. 2003, McGill et al. 2007). Moreover, many macroecological patterns are not independent. For example, the species-area relationship at small spatial scales can be derived from the shape of the SAD and the level of intraspecific aggregation (Harte 2011, McGill 2011), while the SAD itself can be obtained as a spatially autocorrelated sample from the regional species pool (McGill 2011). This combination of equivalent models with different processes and interrelated patterns makes determining process using a single pattern challenging and instead calls for unified theoretical frameworks that are capable of capturing multiple patterns as well as their intercorrelations with a minimal set of assumptions (Marquet et al. 2014).

Theories that have been proposed for macroecological patterns tend to fall into two conceptually distinct categories (Brown 1999, Frank 2014). Similar patterns may arise directly from fundamental ecological processes if the same processes dominate across multiple systems. Theories in this category include the theory of island biogeography (MacArthur and Wilson 1967), which

explains the species richness on islands as the equilibrium between immigration and extinction, and the neutral theory of biodiversity (Hubbell 2001), which shows that demographic stochasticity can lead to community-level diversity patterns. Alternatively, patterns may arise as emergent statistical phenomena with forms determined primarily by some set of numerical constraints on the system (Frank 2014), where processes operate only indirectly through their effects on the constraints. Theories built on constraints include the feasible set (Locey and White 2013), and recent applications of the maximum entropy principle to ecology (Shipley et al. 2006, Dewar and Porté 2008, Harte 2011). Neither of these approaches relies on the operation of specific processes but instead on the fact that many possible combinations of processes and states of the system produce similar empirical outcomes (Frank 2014).

In this study we examined two theoretical frameworks, the Maximum Entropy Theory of Ecology (METE) (Harte 2011) and the size-structured neutral theory (SSNT) (O'Dwyer et al. 2009), which are two of the most comprehensive theories in macroecology. Both theories are able to predict two distinct sets of patterns, those of biodiversity as well as body size and energy use. METE is a constraint-based theory, where patterns arise as the most likely (least biased) state of a community constrained by a set of state variables, such as species richness, the total number of individuals, and the total energy consumption across all individuals. SSNT is an extension of the neutral theory of ecology (Hubbell 2001) and is a process-based theory, where the patterns arise as the steady state of a dynamic system governed by individual birth, death, and growth in size. Both theories make predictions for multiple patterns of biodiversity as well as biomass and energy use, providing a multifaceted characterization of community structure.

We evaluated two existing models of METE and two models that we derived for SSNT, to explore whether community structure in biodiversity and body size can be adequately captured by constraints or processes. One of the METE models, ASNE (Harte 2011, see Methods for details), has been shown in previous studies to have mixed performance among its predictions (Newman et al. 2014, Xiao et al. 2015), while the other models have not been thoroughly tested with empirical data. Using data from 76 forest communities we examined the models' ability to characterize three major macroecological patterns and compared their performance using a single joint distribution that encapsulates these and other predictions as marginal or conditional distributions. Direct comparison of multiple models from the two theoretical frameworks, using a large number of datasets and multiple empirical patterns, allows strong inference to be made about the relative performance of the models and, by extension, the ability of current constraint-based and process-based approaches to characterize community-level macroecological patterns of diversity and body size.

METHODS

Theoretical frameworks

This section briefly outlines the underlying assumptions of METE and SSNT, and the specification of the two models under each theory. For mathematical configurations and derivation of the predicted patterns, see Appendix S1.

Maximum Entropy Theory of Ecology (METE).—The Maximum Entropy Theory of Ecology (METE) is a theory built on the maximum entropy principle (MaxEnt; Jaynes 2003). MaxEnt states that the least biased state of a system is the one with the highest information entropy (Shannon entropy). Given a set of constraints that the system has to satisfy, this state can be obtained by optimization using the method of Lagrange multipliers, with no tunable parameters aside from the constraints.

Among existing applications of MaxEnt to ecology (e.g., Shipley et al. 2006, Pueyo et al. 2007, Dewar and Porté 2008), METE is arguably the most comprehensive, encompassing three distinct branches of ecological patterns – the spatial distributions of individuals and species, the distributions of individuals among species and higher taxonomic ranks, and the allocation of body size and energy use at different taxonomic levels. We examined two existing models of METE, ASNE (Harte et al. 2008, Harte 2011), where the acronym stands for Area, Species, Number of individuals, and Energy, and the newly developed AGSNE (Harte et al. 2015), where the additional "G" stands for Genera or other higher taxonomic ranks (family, order, etc.). In this study, we focused on the non-spatial patterns in ASNE and AGSNE, which are predicted independently from the spatial patterns. In non-spatial ASNE, the allocations of individuals and of body size within a community are regulated by three state variables: species richness S, total abundance N, and total metabolic rate within the community E_{METE}. Non-spatial AGSNE requires an additional input G for a higher taxonomic group, which we took to be the number of genera within the community.

Size-structured neutral theory (SSNT).—Size-structured neutral theory (SSNT) is an extension of Hubbell's neutral theory of ecology (NTE; Hubbell 2001). In NTE, macroecological patterns emerge as the steady state of the community where individuals go through the processes of birth, death, and speciation. SSNT introduces a size component into NTE, where the size of each individual increases through time. Ontogenetic growth thus introduces variation in individual sizes, and also variation in the average size and total biomass across different species. The structure of the community in SSNT is governed by the forms and values of the three demographic parameters b (birth rate), m (mortality rate), and g (rate of growth).

We examined two realized models of SSNT. In the simplest model (SSNT_N, with "N" for neutral), all three

demographic parameters are assumed to be constant for all individuals regardless of their species identities or other individual characteristics. This is called the "completely neutral case" (O'Dwyer et al. 2009). Note that while the assumption of b and m being constant holds regardless of the unit used for size, g can only be constant in one particular set of size units, e.g., constant growth in diameter as a function of current diameter does not translate into constant growth in cross-sectional area or volume as a function of current area or volume. In SSNT_N, we made the intentionally naïve assumption that g was constant across individuals when measured as the increase in diameter D (i.e., $g(D) = \frac{dD}{dt} = $ constant).

In the second model, termed SSNT_M where M stands for metabolism or metabolic theory, we incorporated insights from the metabolic theory of ecology (MTE; Brown et al. 2004), and made the more realistic assumption that g was a function of size, while b and m were still held constant. Specifically, MTE predicts that a plant's growth rate measured as increase in biomass is proportional to the plant's metabolic rate (Enquist et al. 1999, West et al. 1999, Muller-Landau et al. 2006a), which translates into constant growth rate when size is measured in units of diameter raised to the power of 2/3, $D^{2/3}$.

Macroecological patterns.—All four models can predict the same set of three major macroecological patterns: the species-abundance distribution (SAD; distribution of individuals among species), the individual size distribution (ISD; distribution of body size among individuals regardless of their species identity), and the size-density relationship (SDR; relationship between average body size within each species and the abundance of the species) (Cotgreave 1993). AGSNE is also able to predict higher-order patterns, such as the distribution of individuals and body size among genera, which we did not examine in this study (but see Harte et al. 2015).

Table 1 summarizes the predicted forms of the patterns in the four models. λ's in ASNE and AGSNE are Lagrange multipliers (Jaynes 2003), determined by the state variables S, N and E_{METE} in ASNE (see Harte 2011 and Xiao et al. 2015 for detailed derivation), and by G, S, N and E_{METE} in AGSNE (Harte et al. 2015; Appendix S1). τ's in the SSNT models are ratios of the demographic parameters, and are also fully determined by the variables S, N, and a measure of total body size E_{SSNT}, with $E_{SSNT\_N} = \sum D_i$ and $E_{SSNT\_M} = \sum D_i^{2/3}$ (see Appendix S1). Note that patterns of body size (ISD and SDR) predicted by the METE models are of the same unit as metabolic rates (B), which scales with size in trees with good approximation as the square of diameter (D): $B \propto D^2$ (West et al. 1999), wheras the basic unit of size in SSNT is D. For the purpose of comparison, we converted patterns of size from the models into the same units. The ISDs in ASNE and AGSNE were converted to unit of D. The SDRs predicted by these two models do not have simple analytical forms in unit of D, so we converted the predictions of SSNT_N and SSNT_M to unit of B (i.e., D^2) instead (Table 1; Appendix S1).

TABLE 1. Analytical forms of the patterns predicted by the four models with interpretations. λ's and λ's are Lagrange multipliers for ASNE and AGSNE, respectively. τ's are parameters for SSNT_N and SSNT_M. C's are normalization constants. γ in $\Psi_{ASNE}(D)$ is defined as $\gamma = \lambda_1 + \lambda_2 D^2$, and γ' in $\Psi_{AGSNE}(D)$ is defined as $\gamma' = \lambda_2' + \lambda_3' \cdot D^2$.

Patterns	Species abundance distribution (SAD)	Individual size distribution (ISD)	Size-density relationship (SDR)
ASNE	$\Phi_{ASNE}(n) \approx \frac{1}{C_{ASNE1}} \frac{e^{-(\lambda_1 + \lambda_2)n}}{n}$	$\Psi_{ASNE}(D) = \frac{D}{C_{ASNE2}} \cdot \frac{e^{-\gamma}}{(1-e^{-\gamma})^2} \cdot (1-(N+1)e^{-\gamma N} + Ne^{-\gamma(N+1)})$	$\bar{\varepsilon}_{ASNE}(n) = \frac{1}{n\lambda_2\left(e^{-\lambda_2 n} - e^{-\lambda_2 nE}\right)} \cdot \left[e^{-\lambda_2}(\lambda_2 n+1) - e^{-\lambda_2 nE}(\lambda_2 nE+1)\right]$
AGSNE	$\Phi_{AGSNE}(n) \approx \frac{1}{C_{AGSNE1}} \cdot \frac{e^{-(\lambda_1' + \lambda_2' + \lambda_3')n}}{1-e^{-(\lambda_1' + \lambda_2' + \lambda_3')n}} \frac{1}{n}$	$\Psi_{AGSNE}(D) \approx \frac{D}{C_{AGSNE2}} \sum_{m=1}^{S} \frac{me^{-(\lambda_1' + \lambda_2')m}}{(1-e^{-\gamma'm})^2}$	$\bar{\varepsilon}_{AGSNE}(m,n) \approx [e^{-\lambda_3' mn}(\lambda_3' mn+1) - e^{-\lambda_3' mnE}(\lambda_3' mnE+1)]/$ $[\lambda_3' mn(e^{-\lambda_3' mn} - e^{-\lambda_3' mnE})]$
SSNT_N	$\Phi_{SSNT\_N}(n) = -\frac{1}{\ln(1-\tau_1)} \frac{\tau_1^n}{n}$	$\Psi_{SSNT\_N}(D) = \tau_2 e^{-\tau_2(D-1)}$	$\bar{\varepsilon}_{SSNT\_N} = \frac{2}{\tau_2^2} + \frac{2}{\tau_2} + 1$
SSNT_M	$\Phi_{SSNT\_M}(n) = -\frac{1}{\ln(1-\tau_1)} \frac{\tau_1^n}{n}$	$\Psi_{SSNT\_M}(D) = \frac{2}{3}\tau_3 e^{-\tau_3(D^{2/3}-1)} \cdot D^{-1/3}$	$\bar{\varepsilon}_{SSNT\_M} = \frac{6}{\tau_3^3} + \frac{6}{\tau_3^2} + \frac{3}{\tau_3} + 1$
Interpretation	The probability that a randomly selected species has abundance n.	The probability that a randomly selected individual from the community has diameter between $(D, D + \Delta D)$ regardless of species identity.	The average individual metabolic rate within a species with abundance n (and that the species belongs to a genus with m species in AGSNE). Note that metabolic rate scales as D^2 instead of D.

Data

We used forest census data to empirically evaluate the models. This type of data consistently includes individual level size measurements, allowing the compilation of large numbers of communities with the necessary information for fitting and evaluating the models. Forest data sample all individuals of every species down to a certain minimum size, thus avoiding 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) often exhibit multimodal ISDs (Ernest 2005, Thibault et al. 2011), whereas the ISDs for trees are in general monotonically decreasing (Enquist and Niklas 2001, Muller-Landau et al. 2006b), and therefore consistent with the qualitative form predicted by the four models (Table 1).

We combined the data compiled by (Xiao et al. 2015), which encompassed 60 forest communities worldwide, with data on 20 additional communities from (Bradford et al. 2014). All communities have been fully surveyed with species identity and measurement of size (diameter or equivalent) for each individual above community-specific size thresholds (ranging from 10 to 100 mm). In cases where multiple surveys are available for a community, we used those from the most recent survey unless otherwise specified. We excluded individuals that were not identified to genus or species, and removed four communities in (Xiao et al. 2015) from our analysis where >10% of the individuals were not identified to genus. We also excluded individuals that were dead 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 were included in the surveys. Overall the compilation encompasses 76 communities with 2030 species and 378806 individuals from four continents (Asia, Australia, North America, and South America) (Table 2).

Analyses

We applied the four models to each empirical community, and examined their abilities to characterize community structure in abundance and body size. Diameter values in each community were rescaled as $D = D_{original}/D_{min}$, where D_{min} is the diameter of the smallest individual in the community after the exceptional individuals were excluded (see 2. Data), so that D has a minimal value of 1 in each community following METE's assumption (see Harte 2011). Although SSNT does not make such an assumption on minimal size, it predicts the ISD to be an exponential distribution (Table 1), the fit of which is unaffected by this transformation of D. Multiple branches from the same individual were combined to determine the basal stem diameter with the pipe model, which preserves the total area as well as the metabolic rate of the branches

Table 2. Summary of datasets.

Dataset	Description	Area of Individual Plots (ha)	Number of Plots	Survey Year	References
CSIRO	Tropical rainforest	0.5	20	1985–2012*	1
Serimbu	Tropical rainforest	1	1	1995†	2–5
La Selva	Tropical wet forest	2.24	5	2009	6, 7
Eno-2	Tropical moist forest	1	1	2000–2001	8
BCI	Tropical moist forest	50	1	2010	9–11
DeWalt Bolivia forest plots	Tropical moist forest	1	2	N/A	12
Luquillo	Tropical moist forest	16	1	1994–1996‡	13, 14
Sherman	Tropical moist forest	5.96	1	1999	15–17
Cocoli	Tropical moist forest	4	1	1998	15–17
Western Ghats	Wet evergreen/moist/dry deciduous forests	1	34	1996–1997	18
UCSC FERP	Mediterranean mixed evergreen forest	6	1	2007	19
Shirakami	Beech forest	1	2	2006	4, 5, 20
Oosting	Hardwood forest	6.55	1	1989	21, 22
North Carolina forest plots	Mixed hardwoods/pine forest	1.3–5.65	5	1990–1993§	23–25

Notes: [1]Bradford et al. 2014, [2]Kohyama et al. 2001, [3]Kohyama et al. 2003, [4]Lopez-Gonzalez et al. 2009, [5]Lopez-Gonzalez et al. 2011, [6]Baribault et al. 2011*a*, [7]Baribault et al. 2011*b*, [8]Pitman et al. 2005, [9]Condit 1998*a*, [10]Hubbell et al. 1999, [11]Hubbell et al. 2005, [12]DeWalt et al. 1999, [13]Zimmerman et al. 1994, [14]Thompson et al. 2002, [15]Condit 1998*b*, [16]Pyke et al. 2001, [17]Condit et al. 2004 [18]Ramesh et al. 2010, [19]Gilbert et al. 2010, [20]Nakashizuka et al. 2003, [21]Reed et al. 1993, [22]Palmer et al. 2007, [23]Peet and Christensen 1987, [24]McDonald et al. 2002, [25]Xi et al. 2008.

*We chose the most recent survey in each plot before documented disturbances.

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

‡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.

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

(Ernest et al. 2009). Predictions of the models in each community were obtained with the variables S, N, and $E_{METE} = \sum_i D_i^2$ for ASNE, G, S, N, and E_{METE} for AGSNE, S, N, and $E_{SSNT\_N} = \sum_i D_i$ for SSNT_N, and S, N, and $E_{SSNT\_M} = \sum_i D_i^{2/3}$ for SSNT_M.

As an overall measure of model performance, we define the joint distribution $P(n, D_1, D_2, ..., D_n)$ 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 three macroecological patterns, where the SAD is the marginal distribution of n with D_i's integrated out from $P(n, D_1, D_2, ..., D_n)$, the ISD is the marginal distribution of D_i, and the SDR is the expectation of the conditional distribution of D_i given n. The form of this joint distribution predicted by each of the four models is listed in Table 3 (see Appendix S1 for derivations).

We first compared the performance of the four models using the likelihood of $P(n, D_1, D_2, ..., D_n)$ in each community, then examined each of the three macroecological patterns individually. To quantify the predictive power of the models, we converted the SAD and the ISD 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 models' predictions. For example, for the SAD we compared the predicted versus observed abundances of the most abundant species in the community, the second most abundant species, all the way down to the least abundant species (Harte 2011, White et al. 2012, Xiao et al. 2015). For the SDR, we compared the observed average metabolic rate (D'^2) within each species to those expected from the models. The explanatory power of the models for each pattern was quantified using the coefficient of determination R^2:

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

where obs_i and $pred_i$ were the ith value of abundance or size (diameter for the ISD, metabolic rate for the SDR) in the observed and predicted ranked distributions,

respectively. Note that it is possible for the coefficient of determination to be negative, which indicates that the prediction is worse than taking the geometric mean of the observed values. Finally, we examined if the empirical patterns were significantly different from the models' predictions by bootstrap analysis (Clauset et al. 2009, Connolly et al. 2009, Xiao et al. 2015), where we generated random samples from the predicted patterns and quantified their deviation from the predictions (pred$_i$'s) using both R^2 and the Kolmogorov–Smirnov statistic, which were compared with empirical deviations (Appendix S2).

Python Code to fully replicate our analyses is deposited in the Dryad Digital Repository (doi:10.5061/dryad.93ct6).

Results

The log-likelihoods of the joint distribution $P(n, d_1, d_2, ..., d_n)$ of the SSNT models are higher than those of the METE models in all 76 communities (Fig. 1), which implies that SSNT models consistently do a better job characterizing the overall community structure in the allocations of individuals and of body size. Comparing models under the same theoretical framework, the log-likelihood of AGSNE is higher than that of ASNE in all 76 communities, while the log-likelihood of SSNT_M is higher than SSNT_N in 59.

Further examination of individual patterns show that the models predict nearly identical forms for the SAD (i.e., upper-truncated log-series in ASNE, near log-series in AGSNE (Harte et al. 2015), untruncated log-series in SSNT_N and SSNT_M; see Table 1), which not surprisingly translates into equally good performance when evaluated with empirical data (Fig. 2, first column). All four models are also able to characterize the ISD reasonably well with high predictive power ($R^2_{ASNE} = 0.89$, $R^2_{AGSNE} = 0.90$, $R^2_{SSNT\_N} = 0.86$, $R^2_{SSNT\_M} = 0.96$). Three of the models show systematic deviations for the largest individuals, however, with the two METE models tending to over predict the size of the largest individuals and SSNT_N tending to under predict (Fig. 2, second column).

The discrepancy of the two sets of models lies mainly in their predictions of the correlation between individual body

TABLE 3. Joint distribution $P(n, D_1, D_2, ..., D_n)$ for the four models. Z in AGSNE is a constant. See Table 1 for the interpretation of the other symbols and parameters, and Appendix S1 for derivations.

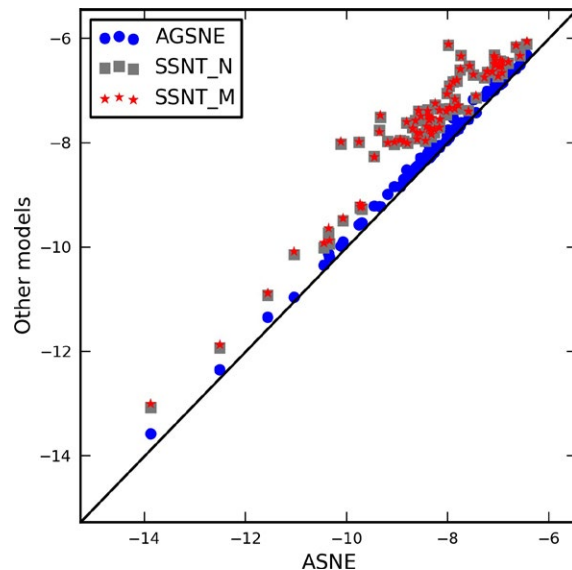| Model	Predicted joint distribution
ASNE	$P_{ASNE}(n, D_1, D_2, ..., D_n) = \dfrac{1}{C_{ASNE1}n} e^{-(\lambda_1 + \lambda_2)n} \prod_{i=1}^{n} \dfrac{2n\lambda_2 D_i e^{-\lambda_2 n D_i^2}}{e^{-\lambda_2 n} - e^{-\lambda_2 n E_{METE}}}$
AGSNE	$P_{AGSNE}(n, D_1, D_2, ..., D_n) = \left(\dfrac{2G}{ZS}\right)^n [\Phi_{AGSNE}(n)]^{1-n} \prod_{i=1}^{n} D_i \dfrac{t}{(1-t)^2} [1 - (S+1)t^S + St^{S+1}]$ where $t(D_i) = e^{-(\lambda_1' + \lambda_2' n + \lambda_3' n D_i^2)}$
SSNT_N	$P_{SSNT\_N}(n, D_1, D_2, ..., D_n) = -\dfrac{1}{\log(1 - \tau_1)} \dfrac{\tau_1^n}{n} \prod_{i=1}^{n} \tau_2 e^{-\tau_2 \cdot (D_i - 1)}$
SSNT_M	$P_{SSNT\_N}(n, D_1, D_2, ..., D_n) = -\dfrac{1}{\log(1 - \tau_1)} \dfrac{\tau_1^n}{n} \cdot \prod_{i=1}^{n} \dfrac{2}{3}\tau_3 e^{-\tau_3(D_i^{2/3} - 1)} \cdot D_i^{-1/3}$

Fig. 1. Comparison of the log-likelihood (l) of the joint distribution $P(n, d_1, d_2, …, d_n)$ for the four models in each of the 76 forest communities. l of AGSNE, METE_N, and METE_M are compared with that of ASNE, which has the lowest likelihood in all 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. Note that values of l depend on the number of individuals within the community, and thus are not comparable across communities.

size and species abundance. The two METE models both predict that individual body size depends on the species that the individual belongs to – in ASNE average individual body size is negatively correlated with species abundance (Harte 2011), whereas in AGSNE it is negatively correlated with both species abundance and number of species within genus (Harte et al. 2015). The SDR predicted by ASNE has been shown to be unrealistic in plant communities (Newman et al. 2014, Xiao et al. 2015), and our results show that AGSNE improves ASNE's prediction only marginally ($R^2_{ASNE} = -2.11$, $R^2_{AGSNE} = -2.00$). SSNT, on the other hand, predicts that there is no correlation between individual size and species characteristics, leading to better, but still far from good, agreement with empirical data for the SDR ($R^2_{SSNT\_N} = 0.09$, $R^2_{SSNT\_M} = 0.02$). These results are robust when the two models are examined in each of the 76 communities individually (Fig. 3) – the four models yield nearly identical R^2 values for the SAD and comparable R^2 values for the ISD across communities (with SSNT_M having the highest predictive power on average), while the two SSNT models consistently outperforms the two METE models for the SDR.

The bootstrap analysis (Appendix S2) shows that the discrepancy between the models' predictions and the observations for the ISD is almost ubiquitously higher than expected from random sampling in all four models. This suggests that none of the models is able to fully capture the observed variation in the size distributions of individuals, despite their high predictive power. A similar

pattern is observed for the SDR predictions for the METE models. In contrast, despite the low predictive power, SDR predictions for the two SSNT models suggest that the majority of the communities are indistinguishable from random samples of the predicted pattern (Figs. B3, B4). This implies that SSNT's prediction of no correlation between species abundance and individual body size is more or less accurate.

DISCUSSION

In this study, we compared the performance of METE (Harte 2011) and SSNT (O'Dwyer et al. 2009), two of the most comprehensive theories to date in macroecology, using two realized models for each. Both theories attempt to unify multiple aspects of community structure under a single theoretical framework, predicting patterns of biodiversity as well as patterns of energy consumption and body size. Using data from 76 forest communities worldwide, we showed that the two models of SSNT consistently provide a better characterization of overall community structure than the two models of METE (Fig. 1). This disparity results primarily from the ability of SSNT_N and SSNT_M to more accurately characterize the relationship between species abundance and body size distributions within species, whereas the predictions of ASNE and AGSNE on this relationship deviate from empirical patterns (Newman et al. 2014, Xiao et al. 2015).

By comparing multiple competing models on multiple predictions simultaneously using an extensive set of data, our study achieves the strongest level of model evaluation suggested by McGill et al. (2006), and provides insights into the role of the underlying mechanisms of the theories. In METE, the macroecological patterns arise as the most likely state of the system assuming that the system is constrained by a small number of state variables. METE makes no explicit assumptions about ecological processes, leaving their influence to operate indirectly through their potential effects on the values of the state variables. In SSNT, patterns emerge directly from the interactions of the demographic processes including birth, death, and growth. The fact that SSNT performs better than METE suggests that the demographic processes contain meaningful information that helps to characterize the patterns, the effect of which is not currently captured by the state-variable based models.

Although the differences between the models are important, the fact that all four models are capable of adequately characterizing the shapes of the SAD and the ISD across a large number of communities with simple assumptions and limited inputs is impressive. Moreover, the maximum likelihood parameters for SSNT_N and SSNT_M are also fully determined by S, N, and E (see Appendix S1), so that these variables serve as summary statistics for the demographic parameters. These results imply that METE and SSNT contain overlapping information. Although these demographic processes explain a higher proportion of the variation in the empirical
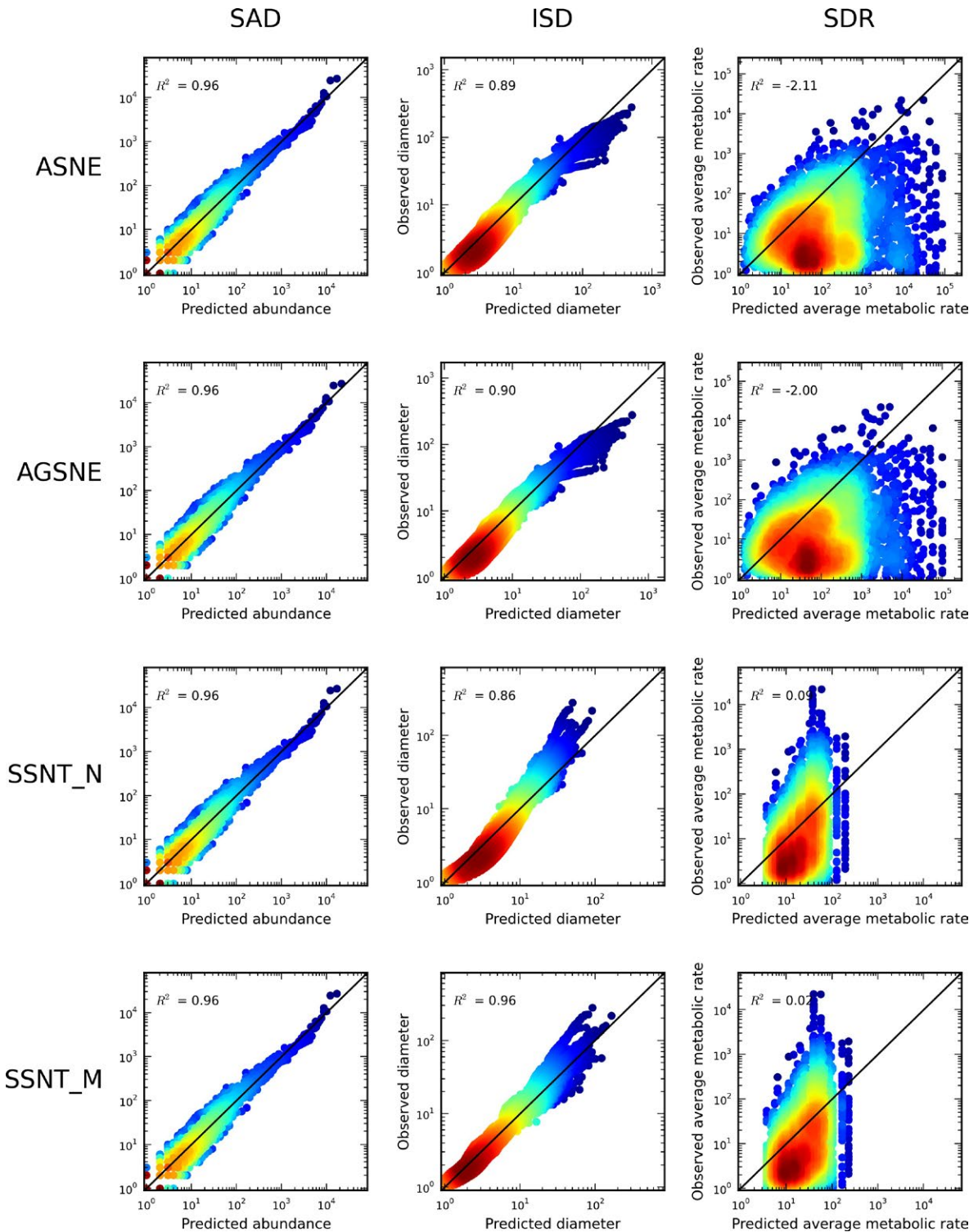
Fig. 2. Comparison of the performance of the four models for each of the three macroecological patterns. Each point in the subplot represents the abundance of one species in a community for the SAD, the diameter of one individual in a community for the ISD, and the average metabolic rate (squared diameter) within one species in a community for the SDR. 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.
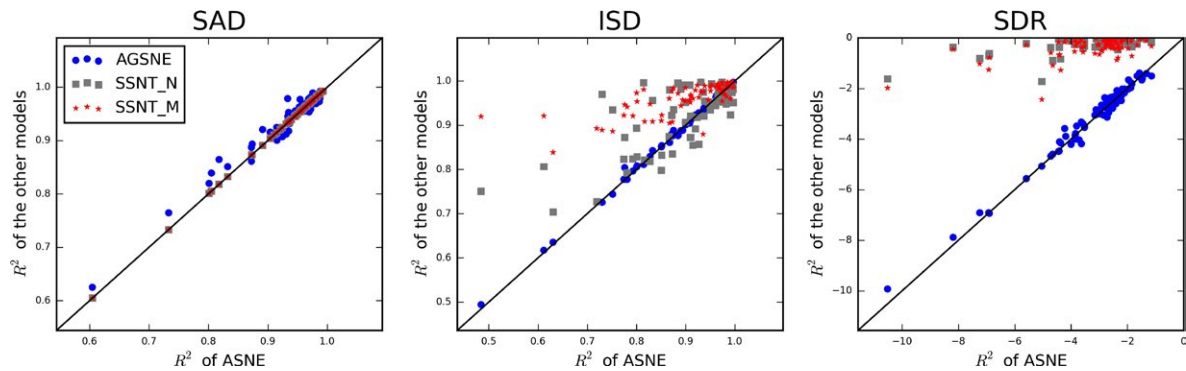
Fig. 3. Comparison of R^2 values for the three macroecological patterns predicted by the four models in each of the 76 forest communities. R^2 of AGSNE, METE_N, and METE_M are compared with that of ASNE. The diagonal line is the one-to-one line.

patterns, their effects are likely to at least be partially channeled through the constraints.

Our study is one step towards the goal of disentangling the effects of different mechanisms on macroecological patterns. Though we have adopted model comparison for stronger inference, we do not advocate rejecting the theoretical framework (METE) with the poorer fitting models or its underlying constraint-based view as a potential explanation for patterns. There are three reasons for being cautious about over interpreting these results. First, METE and SSNT are general theories built on first principles, whereas our conclusions are limited to their current realized models based on specific assumptions. Models under the same theoretical framework yield different predictions with different assumptions and inputs, which can be evaluated with empirical data and improved with additional information. This is demonstrated in our study by comparing the two models from the same theory – with an additional constraint G, AGSNE has consistently higher likelihood than ASNE (Fig. 1), while incorporating information from the metabolic theory in SSNT_M eliminates the systematic deviation in the predicted ISD (Fig. 2). Future models will likely be developed with alternative implementations leading to new and/or improved predictions.

Second, our inference is limited by the scope of the data. Though the models have the potential to be applied to a wide variety of systems, we focused exclusively on trees, where data of full surveys are readily available with species identity and body size for all individuals. Although our results are consistent across forest communities of different types and sizes (Fig. 3), it remains to be seen if they can be generalized to other taxa. Previous studies suggest that the size distributions predicted by ASNE and AGSNE may be more accurate when applied to invertebrates (Harte 2011, Harte et al. 2015). Third, patterns that can be unified under the same theoretical framework do not necessarily have to arise from the same underlying mechanism. Indeed, there is increasing evidence that the SAD is driven by statistical properties of the system (White et al. 2012, Locey and White 2013,

Blonder et al. 2014), while patterns that show spatial or taxonomical variation, such as the patterns of body size, are more likely to be tied to ecological processes (Blonder et al. 2014).

One weakness prevalent across all four models is their inability to characterize the SDR, the relationship between species abundance and the body size of individuals within species, despite their success in independently predicting the distribution of individuals among species and the allocation of body size among individuals. Our results agree with previous studies showing that the SDR exhibits significant variation at the local scale (Lawton 1990, Cotgreave 1993), not strongly abundance-dependent as the METE models predict. The prediction of the SSNT models that the SDR results from random draws is more in line with empirical observations (Appendix S2), but they too lack predictive power (Fig. 2). Although part of the variation may result from the limitation of data we used, e.g., species having different growth rates were surveyed at different life stages, it could also indicate species-specific size biases in resource use (White et al. 2007). One potential remedy that may improve the predicted SDR as well as lead to additional predictions is to take an approach alternative to the two that we have addressed, and model macroecological patterns by directly stacking models of individual species. This approach has shown promise in predicting other patterns, such as biodiversity across space (Guisan and Rahbek 2011, D'Amen et al. 2015). Similar models could potentially be developed to model the abundance and body size of species based on their traits, and to obtain the macroecological patterns from the species-level predictions. Such models will likely sacrifice parsimony for accuracy, however, and require a lot more parameters than the models that we examined.

Another potentially fruitful route to push the two theories forward is to unify the constraint- and the process-based approaches, which have generally been adopted by distinct theories but do not necessarily have to be mutually exclusive. Results of our study imply that part of the effects of the demographic processes propagate through

the constraints, while other studies (e.g., Haegeman and Etienne 2010) state that different configurations for the same set of constraints can often be tied to (and may eventually be informed from) process-based mechanistic models. Indeed, an exciting new model is being developed where the maximum entropy principle is combined with demographic processes to characterize not only the steady state but also temporal dynamics of a system (Umemura and Harte 2015). The attempts to model ecological systems completely with constraints or processes may thus represent two extremes of a continuous spectrum, along 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 biological processes in structuring ecological systems.

Literature Cited

Baribault, T. W., R. K. Kobe and A. O. Finley. 2011a. Data from: tropical tree growth is correlated with soil phosphorus, potassium, and calcium, though not for legumes. Dryad Digital Repository http://dx.doi.org/10.5061/dryad.r9p7.

Baribault, T. W., R. K. Kobe, and A. O. Finley. 2011b. Tropical tree growth is correlated with soil phosphorus, potassium, and calcium, though not for legumes. Ecological Monographs 82:189–203.

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

Bradford, M. G., H. T. Murphy, A. J. Ford, D. Hogan, and D. J. Metcalfe. 2014. Long-term stem inventory data from tropical rain forest plots in Australia. Ecology 95:2362.

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

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

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.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. SIAM Review 51:661–703.

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, and R. B. Foster. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. Journal of Tropical Ecology 20:51–72.

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.

D'Amen, M., A. Dubuis, R. F. Fernandes, J. Pottier, L. Pellissier, and A. Guisan. 2015. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. Journal of Biogeography 42:1255–1266.

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.

Enquist, B. J., G. B. West, E. L. Charnov, and J. H. Brown. 1999. Allometric scaling of production and life-history variation in vascular plants. Nature 401:907–911.

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.

Gilbert, G. S., E. Howard, B. Ayala-Orozco, M. Bonilla-Moheno, J. Cummings, S. Langridge, I. M. Parker, J. Pasari, D. Schweizer, and S. Swope. 2010. Beyond the tropics: forest structure in a temperate forest mapped plot. Journal of Vegetation Science 21:388–405.

Guisan, A., and C. Rahbek. 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. Journal of Biogeography 38:1433–1444.

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, UK.

Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. Maximum entropy and the state-variable approach to macroecology. Ecology 89:2700–2711.

Harte, J., A. Rominger and W. Zhang. 2015. Integrating macroecological metrics and community taxonomic structure. Ecology Letters 18:1068–1077.

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

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

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

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

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.

Lawton, J. H. 1990. Species richness and population dynamics of animal assemblages. Patterns in body size: abundance space. Philosophical Transactions of the Royal Society B: Biological Sciences 330:283–291.

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.

MacArthur, R. H. and E. O. Wilson. 1967. The Theory of Island Biogeography. Princeton University Press, Princeton, New Jersey, USA.

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, and K. J. Gaston, editors. Macroecology: concepts and Consequences. Blackwell Science, Oxford.

Marquet, P. A., et al. 2014. On Theory in Ecology. BioScience 64:701–710.

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

McGill, B. J. 2011. Linking biodiversity patterns by autocorrelated random sampling. American Journal of Botany 98:481–502.

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

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

McGill, B. J., B. A. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. Ecology 87:1411–1423.

Muller-Landau, H. C., et al. 2006a. Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. Ecology letters 9:575–588.

Muller-Landau, H. C., et al. 2006b. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. Ecology Letters 9:589–602.

Nakashizuka, T., et al. 2003. Monitoring dynamics of beech forests with 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.

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.

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

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.

Pueyo, S., F. He, and T. Zillio. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. Ecology Letters 10:1017–1028.

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, and S. Ramalingam. 2010. Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India. Ecology 91:3118.

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.

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

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, D. García-Montiel, and M. Fluet. 2002. Land use history, environment,

and tree composition in a tropical forest. Ecological Applications 12:1344–1363.

Umemura, K. and J. Harte. 2015. Dynamic extension of the maximum entropy theory of ecology (METE) using a master equation approach. Poster presented at the 100th Ecological Society of America Annual Meeting.

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. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93:1772–1778.

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., D. J. McGlinn, and E. P. White. 2015. A strong test of the Maximum Entropy Theory of Ecology. The American Naturalist 185:E70–E80.

Zimmerman, J. K., E. M. III Everham, 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.

Supporting Information