ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/235371173

Taxon Categories and the Universal Species-Area Relationship

Article *in* The American Naturalist · February 2013

DOI: 10.1086/668821 · Source: PubMed

CITATIONS
9

READS
100

4 authors:



John Harte
University of California, Berkeley
238 PUBLICATIONS **11,960** CITATIONS
SEE PROFILE



Justin Kitzes
University of Pittsburgh
40 PUBLICATIONS **2,415** CITATIONS
SEE PROFILE



Erica A. Newman
The University of Arizona
19 PUBLICATIONS **79** CITATIONS
SEE PROFILE



Andrew J Rominger
Santa Fe Institute
18 PUBLICATIONS **177** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Wildfire on Pacific islands View project



DynaMETE View project

All content following this page was uploaded by John Harte on 22 May 2014.

The user has requested enhancement of the downloaded file.



CHICAGO JOURNALS



The University of Chicago

Taxon Categories and the Universal Species-Area Relationship: A Comment on Šizling et al., "Between Geometry and Biology: The Problem of Universality of the Species-Area Relationship."
Author(s): John Harte, Justin Kitzes, Erica A. Newman, and Andrew J. Rominger
Reviewed work(s):
Source: *The American Naturalist*, Vol. 181, No. 2 (February 2013), pp. 282-287
Published by: The University of Chicago Press for The American Society of Naturalists
Stable URL: http://www.jstor.org/stable/10.1086/668821
Accessed: 25/01/2013 12:04

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press, The American Society of Naturalists, The University of Chicago are collaborating with JSTOR to digitize, preserve and extend access to *The American Naturalist.*

http://www.jstor.org

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

Comment

Taxon Categories and the Universal Species-Area Relationship

(A Comment on Šizling et al., "Between Geometry and Biology: The Problem of Universality of the Species-Area Relationship")

John Harte,[1,2,*] Justin Kitzes,[2] Erica A. Newman,[1,2] and Andrew J. Rominger[2]

1. Energy and Resources Group, University of California, Berkeley, California 94720; 2. Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720

Submitted January 2, 2012; Accepted September 27, 2012; Electronically published January 4, 2013

abstract: A theory of macroecology based on the maximum information entropy (MaxEnt) inference procedure predicts that the log-log slope of the species-area relationship (SAR) at any spatial scale is a specified function of the ratio of abundance, $N(A)$, to species richness, $S(A)$, at that scale. The theory thus predicts, in generally good agreement with observation, that all SARs collapse onto a specified universal curve when local slope, $z(A)$, is plotted against $N(A)/S(A)$. A recent publication, however, argues that if it is assumed that patterns in macroecology are independent of the taxonomic choices that define assemblages of species, then this principle of "taxon invariance" precludes the MaxEnt-predicted universality of the SAR. By distinguishing two dimensions of the notion of taxon invariance, we show that while the MaxEnt-based theory predicts universality regardless of the taxonomic choices that define an assemblage of species, the biological characteristics of assemblages should under MaxEnt, and do in reality, influence the realism of the predictions.

Keywords: MaxEnt, information entropy, species-area relationship, macroecology, scaling, universality, maximum entropy theory of ecology, taxon invariance.

Introduction

A recently proposed MaxEnt-based theory of ecology (METE) predicts numerous metrics of macroecology, including abundance and metabolic rate distributions, spatial aggregation patterns of individuals within species, and both endemics-area and species-area relationships (SARs). Numerous empirical tests indicate general support for the predictions (Harte et al. 2008, 2009). A full derivation of the theoretical predictions, detailed comparisons with data, and characterization of systems for which the theory may fail to make accurate predictions are described by Harte (2011). Predicted mathematical forms of the metrics are derived from constraints arising from the values of four state variables: an arbitrary but specified spatial scale (A_0), the total number of species (S_0), total number of individuals (N_0), and total metabolic rate of all individuals (E_0) at that spatial scale. One prediction of METE that we will expand on below is that all SARs, if properly rescaled, collapse onto a single curve, a phenomenon we call SAR universality.

In a recent article, Šizling et al. (2011) invoke a concept they denote "taxon invariance" (Storch and Šizling, 2008) and argue that it is incompatible with the METE prediction of a universal SAR. Taxon invariance asserts that the functions describing the macroecological metrics for an assemblage of species should not depend on the taxonomic choices used to define the assemblage. Šizling et al. also adjoin to the above meaning of taxon invariance a stronger statement: a postulated additivity rule. Applied to the METE SAR prediction, the additivity rule asserts that if the predicted universal SAR is evaluated separately for two subassemblages of species, then the sum of the two SARs should be identical to the predicted universal SAR evaluated for the entire assemblage. They then show that the METE universal SAR violates that rule and that the only universal SAR that would be compatible with the additivity rule is in conflict with empirical data. Here we examine the implications of taxon invariance and the additivity rule in the context of MaxEnt-based theory and demonstrate that (1) METE satisfies taxon invariance in the sense that it predicts universality of the SAR for any chosen taxonomic assemblage, (2) predictions that violate the additivity rule are expected within the MaxEnt inference framework and those violations do not invalidate METE, and (3) biological criteria allow choices of assemblages to be made for which the METE prediction of a universal SAR is most accurate.

* Corresponding author; e-mail: jharte@berkeley.edu.

Am. Nat. 2013. Vol. 181, pp. 282–287. © 2013 by The University of Chicago. All rights reserved.
0003-0147/2013/18102-53551$15.00.
DOI: 10.1086/668821

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

Basics of the MaxEnt Inference Procedure

The maximum information entropy inference procedure (MaxEnt) is a basis for deriving theory and making predictions. The procedure has been proven to yield least-biased probability distributions that are consistent with prior knowledge in the form of constraints on those distributions (Jaynes 1979, 2003). "Least biased" here means that the resulting distributions do not implicitly or explicitly embody any information other than the known constraints, which represent prior knowledge about the system. In graphical terms, they are the smoothest possible distributions that satisfy the constraints.

The detailed forms of the probability distributions predicted by MaxEnt depend on the known constraints on the system. They include specified ranges of values the independent variables can take and knowledge of the means of specified functions of those variables. When different analysts start with different prior knowledge constraining a probability distribution, they will generally infer different predictions for the form of that distribution. The more constraints that are known and that are empirically valid, the more reliable (or at least no less reliable) a MaxEnt inference will be. MaxEnt predictions do not necessarily accurately describe nature, but they are the best possible predictions compatible with prior knowledge of constraints (Jaynes 1982). Constraints are generally formulated in terms of the state variables characterizing the system being described.

Census data in ecology are rarely complete because of false absences and presences. Thus empirical values of the state variables are only estimates of their true values and an empirical SAR is only an estimate of the true SAR. METE uses the observed (imperfect) values of the state variables to constrain predictions for observed (imperfect) metrics such as the SAR. Given the constraints, least-biased predictions are obtained by this procedure, but these predictions are not necessarily predictions of the true metrics. Were more complete data available, improved values for the state variables would predict improved values for the more accurately measured metrics.

In METE, the ratios of the values of the state variables provide the constraints that determine the shape of the macroecological metrics. The values of the state variables, in turn, depend on the choice of the ecological system and the assemblage of species being studied. If there are S species and N individuals at scale A, then the species-abundance distribution (SAD) at that scale is predicted by METE to be the upper-truncated log series distribution:

$$\Phi(n|S, N) \approx \frac{1}{\log\,[1/(1 - e^{-\beta})]}\frac{e^{-\beta n}}{n}, \tag{1}$$

where the approximation is valid for $e^{-S_0} \ll 1$, and $\beta(A)$, which depends only on N/S, is a combination of Lagrange multipliers resulting from maximizing information entropy conditional on the constraints that arise from ratios of the state variables. Term $\beta(A)$ is determined from the equation

$$\frac{N}{S}\sum_{n=1}^{N}\frac{e^{-\beta n}}{n} = \sum_{n=1}^{N} e^{-\beta n}. \tag{2}$$

In most realistic cases $\beta \ll 1$, so the term $(1 - e^{-\beta})$ in equation (1) can be replaced with $1/\beta$ and equation (2) can be replaced with $\beta \ln\,(1/\beta) \approx S/N$.

The equation for the number of species in an area A′ in terms of the number in area $A > A'$ can be written as

$$S(A') = S(A)\sum_{n=1}^{N}[1 - \Pi(0|n, A, A')]\Phi(n|S, N). \tag{3}$$

Here, the function Π is the species-level spatial abundance distribution (Harte et al. 2008) and $1 - \Pi(0|n, A, A')$ is the probability that if there are n individuals in A, then the species is present in A'. Function $\Phi(n|S, N)$ is given by equations (1) and (2). METE predicts that Π takes the form of an upper-truncated geometric distribution, and Φ is an upper-truncated log series distribution (Harte 2008, 2011).

METE then predicts (Harte 2011) a SAR behavior characterized by

$$z(A) \approx \frac{\ln\left[\frac{\ln\,(1/\beta(A))}{\ln\,(1/\beta(A)) - 1}\right]}{\ln\,(2)}, \tag{4}$$

where $z(A)$ is the log-log slope of the SAR at area A. Different choices of taxonomic groups result in different values of S and N at every scale, and this can lead to different predictions for the numerical value of $\beta(A)$ and thus $z(A)$.

Universality

The form of $z(A)$ in equation (4) results in a universal SAR in the following sense: the log-log slope of the SAR at any spatial scale is a specified function only of the ratio of the total number of individuals, $N(A)$, and total number of species, $S(A)$, at that scale. Starting at any specified scale at which S and N are known, equations (2) and (4) can be iterated to yield S and N at larger or smaller scales; the resulting universal SAR is shown as the predicted curve in figure 1. If, for any choice of a taxonomic group, data on species richness from many spatial scales are collected and the log-log SAR slopes plotted at each scale against

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

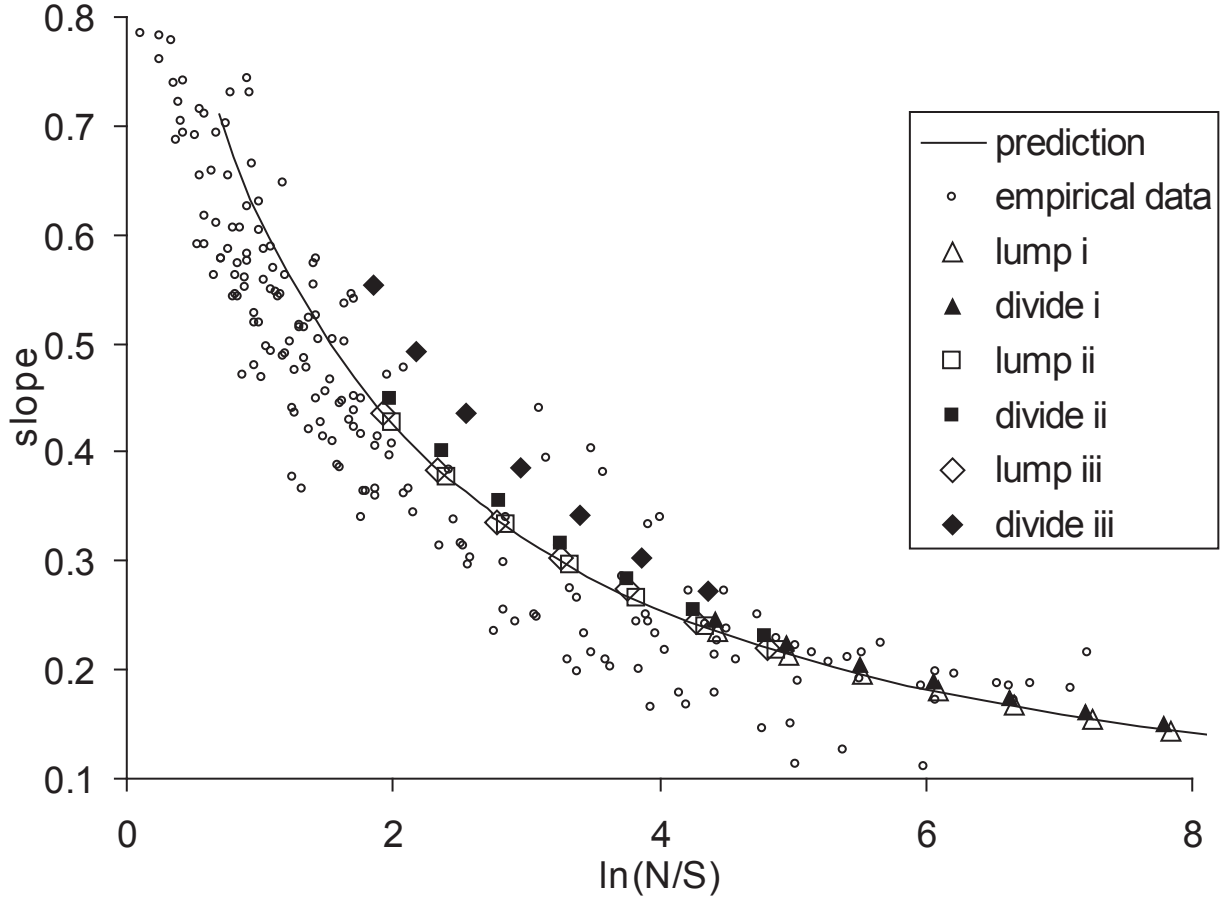


Figure 1: Universality and taxon invariance of the SAR. The solid line is the predicted relationship between log-log slope z of the SAR and the ratio N/S. The small open circles are empirical data from numerous SARs (sources of data are cited in Harte et al. 2009). Larger data symbols describe numerical results of dividing or lumping taxa, using the state variable assignments shown in table 1. The open symbols are the lumped predictions, calculated from $S = S_1 + S_2$. Filled symbols describe the slopes of the SAR curves derived from upscaling with the universal curve from S_1 and S_2 separately and then adding the two results at each spatial scale. Numerical species richness predictions for both the divider and the lumper are shown in table 1.

the corresponding N/S values, then METE predicts that the graphs will collapse onto that predicted universal curve in which z is a unique, decreasing, function of the local-scale value of N/S (Harte 2011). The numerical value of the remaining state variable E, metabolic rate, only very weakly influences this SAR prediction for reasons explained in Harte (2011).

As seen in figure 1, data from a large number of habitats, taxonomic groups, and spatial scales do indeed appear to be consistent with this nontrivial prediction. Both within and across habitats and taxa, the scatter around the universality prediction is very small compared to the deviations from a power-law SAR model (constant slope prediction), which would correspond to a horizontal line in figure 1. Many models and theories have been proposed to explain the shapes of SARs, but to our knowledge METE is the first theory to predict without adjustable parameters the empirically realistic behavior shown in the figure.

Taxon Invariance, the Additivity Rule, and SAR Universality

The universal SAR curve in figure 1 is predicted for any choice of an assemblage of species, and in that sense, the METE universal SAR is taxon-invariant. Where any given assemblage lies along the universal curve depends on the values of the state variables at some fixed scale. To shed light on the Šizling et al. observation about additivity (which postulates a stronger requirement on theory), consider two analysts, with the same census data gathered at

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

some spatial scale A, seeking to estimate species richness at some larger spatial scale, A'. Analyst I looks at the entire assemblage and uses the value of $S(A)$ and $N(A)$ as state variables to predict with the METE universal SAR the value of $S' = S(A')$. Analyst II divides the species list at scale A into two sublists, $S_1(A)$ and $S_2(A)$, where the subdivision is nonoverlapping. Associated with these two categories of species, S_1 and S_2, the total abundance, N, at scale A is divided into $N = N_1 + N_2$. Analyst II also uses the METE universal SAR to upscale each list separately, infer $S_1(A')$ and $S_2(A')$, and then add the upscaled results together to predict species richness at A'.

Although both analysts use the same theory to upscale species richness, they may reach different conclusions because they impose different constraints that, as we have explained, will generally lead to different outcomes in a MaxEnt inference procedure. Both analysts will use the same universal SAR predicted by METE to upscale, but they reach different conclusions because their differing initial constraints lead them to upscale species richness along different portions of that curve. The discrepancy between the predictions of the two analysts will disappear if $N_1/S_1 = N_2/S_2$. This is because the shape of the SAR depends only on the quantity β (eq. [4]), and β in turn depends only on N/S.

Figure 1 illustrates the discrepancy between the two analysts' projections with three numerical examples for the case $A' = 256A$, using as initial conditions several different sets of values of S, N, S_1, S_2, N_1, and N_2 at area A as shown in table 1. Figure 2 illustrates the same phenomenon for the case of a serpentine meadow ecosystem. In this example, the data at the 1-m^2 scale provide the state variables (number of plant species and their total abundance) and species richness is upscaled using METE to 64 m^2. Calculating the SAR separately for each of two subassemblages of meadow plant species (in our example, the lower- and the higher-abundance species) and then adding those two SARs is seen in figure 2 to lead to higher species richness than obtained if the SAR is calculated from the entire species list.

Table 1: State variables at scale $A = 1$ and results of upscaling to $A' = 256$ for three choices (cases i–iii) of dividing species into two categories

	Case i	Case ii	Case iii
S_1	10	8	10
S_2	10	12	200
N_1	800	60	500
N_2	200	40	500
$(S_1 + S_2)'$	55.52	112.49	1,214.3
$S_1' + S_2'$	58.70	124.68	1,935.4

Note: The last two rows give the predicted upscaled species richness values for the case in which the taxa are lumped $(S_1 + S_2)'$, and divided $S_1' + S_2'$. Note that only the most lopsided division of the species list (case iii) yields values of upscaled species richness that differ appreciably between the two analysts. At each scale from $A' = 2$ to 128, the z values for the three cases are shown in figure 1.

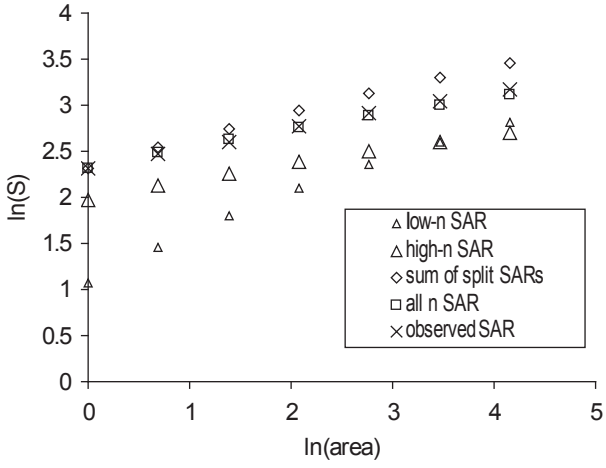


Figure 2: The species-area relationship for a serpentine meadow, illustrating the effect on the predicted SAR of dividing the species into subassemblages. The empirical data are accurately described by using METE to scale up the entire species list from 1 to 64 m^2 (crosses vs. open squares). Dividing the list into low- and high-abundance species (open small and large triangles) and using METE to scale up each list results in the SAR denoted by open trapezoids.

Šizling et al. (2011) show that there is a universal SAR that is compatible with the additivity rule but that it differs dramatically from empirical data. It would in fact be a convex, or downward accelerating, curve instead of the concave decelerating curve in figure 1. Hence, if the additivity rule is postulated to be a rigid constraint on theory, then a realistic universal SAR is ruled out.

Making Taxonomic Choices

Because the additivity rule generally fails, as expected, in a MaxEnt calculation, an interesting question arises: is a more realistic answer obtained by analyst I (the lumper), who upscales an entire assemblage, or analyst II (the divider), who upscales the two subassemblages separately and adds the results? To answer this it is useful to first examine some extreme situations, which will serve to highlight the relevant factors that should influence the decision to apply METE to the entire assemblage or to the subassemblages.

Suppose the entire assemblage contains both birds and insects. Analyst I only knows the total value of S and N at some scale, while analyst II, who subdivides the assemblage into birds and insects, will have twice as much initial

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

information: the values of S and N for birds and insects separately. As discussed above, within the MaxEnt framework having more information necessarily suggests that analyst II should be able, using METE, to make a more accurate prediction of total species richness at some larger scale. Under METE, analyst II will infer an SAR for the birds and an SAR for the insects. By adding these two SARs, the least-biased SAR for birds plus insects will be obtained. Using METE, analyst I will not be able to infer anything about the SAR of birds or insects separately and will make a less accurate prediction of the combined SAR of birds and insects.

Essential to the above argument, however, is the absence of ambiguity in the scaling-up procedure. To see how ambiguity can arise, consider a different total assemblage, consisting just of plant species as in figure 2. Analyst II, the divider, knows the state variables for two subassemblages of the plants, defined on the basis of abundance at the initial small scale. Upscaling the subassemblages separately, analyst II faces a problem. There is no way to know, for example, if rare species at small scale upscale only to rare species at large scale or if abundant species at small scale upscale only to rare species at large scale. In other words, the two upscaled species lists may contain some species in common. As a result, when analyst II adds the results of the upscaling calculation for the two subassemblages, the result may overestimate the correct answer, just as was shown in figure 2. In this example, there is no ambiguity in analyst I's calculation, and so upscaling the entire assemblage is expected to yield a more reliable result.

The critical point is that under METE, both analysts will use the exact same universal SAR curve to upscale their data. The fact that different results are obtained depending on how the analysts choose to subdivide the species list should not be surprising, for different choices lead to different constraints and thus to different outcomes under MaxEnt. The failure of theory to respect the additivity rule does not invalidate the MaxEnt inference procedure. Each analyst is making the correct (i.e., least-biased) inference from his or her chosen set of state variables and the constraints imposed by the numerical values of those state variables. In this second case, where subassemblages are defined by an abundance criterion, the prediction of analyst I is expected to be more accurate because there is inherent ambiguity in the conclusions of analyst II.

The empirical success of the universal SAR curve predicted by METE (fig. 1) supports the assumption that good taxonomic and ecological decisions were made in defining species and delimiting communities in the data sets shown in the figure. In all cases, the assemblages that satisfy the universal SAR prediction were defined broadly enough to avoid the problem of ambiguity but narrowly enough to not mix taxonomic groups as distinct as birds and insects. Based on the above analysis, a good taxonomic decision should maximize the number of state variables so that the amount of constraining information is maximized while at the same time avoiding ambiguity.

Pursuing this, two criteria can be explored for determining when a subdivision of species might lead to more accurate predictions. The first arises from consideration of the species abundance distribution predicted by METE, the log series. If that distribution characterizes the abundance distribution of a group of species that compete for resources so that substitution of individuals within that group under a zero-sum resource constraint is possible, then the SAR for a group of species with common resource niches might not equal the sum of the SARs of any arbitrary subset of subgroups of that larger group. The most accurate SAR would be that obtained for the entire group as a whole.

A second approach involves evolutionary history and raises the possibility of a marriage of insight between macroecological and macroevolutionary theory. Consider two clades separated by a given amount of evolutionary time. We hypothesize that with greater evolutionary separation, the SAR predicted by scaling up the two groups of species separately and then adding the curves will be more accurate relative to scaling up the two clades as one group. The depth of evolutionary divergence invokes not only biology (distant relatives are likely to also be biological divergent) but also information, as temporally distant divergences are less likely to result in phylogenetic uncertainty. It would be of considerable interest to test what role that plays in characterizing assemblages that obey the universal SAR prediction.

The failure of MaxEnt-based theory to respect the additivity rule thus suggests a way to obtain insight into the ecological and evolutionary structure of communities. Such insight can be obtained from empirical determination as to whether subdividing into two or more subassemblages and adding SARs for subassemblages generates a more accurate SAR prediction than is obtained by not subdividing. Whether some criteria such as those suggested above for defining groups of species yields the most accurate SARs must be resolved empirically. The answer might of course differ across habitats, types of taxa, spatial scales, nutrient and water availability, or other broad ecological and evolutionary features of ecosystems.

Discussion

Two distinct concepts have been postulated by Šizling et al. (2011), and confusion is created by conflating them. Taxon invariance in one sense states that the same theory has to be used to describe any assemblage, however it is

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

defined; the theory cannot be adjusted to match the taxa. In the case of METE, once the numerical values of the state variables are specified, then the metrics of macroecology can be predicted. METE is taxon invariant in just that sense; that kind of taxon invariance is a reasonable requirement to be imposed on theories, but more on the grounds of Occam's razor than for any biological reason. If theories have to be recalibrated for every taxonomic assemblage being studied, falsification is difficult and theory becomes unwieldy.

While taxon invariance is a statement about theories, not about nature, the additivity rule goes beyond this and asserts an additive relationship between the results of two different ways of defining state variables: dividing and lumping. That additive relationship is generally violated by METE, which we have seen is a natural consequence of the dependence of MaxEnt predictions on prior knowledge expressed as constraints on relevant probability distributions. Imposing the additivity rule and SAR universality together results in a contradiction with observation.

This suggests that the additivity rule is an unrealistic requirement to be imposed on theory. In contrast to the above taxon invariance requirement, the additivity rule is an observation about data manipulations and is not a required constraint on theory. It is respected by nature in the sense that the sum of the empirical SARs for two subassemblages equals the empirical SAR for the assemblage. A theory designed to predict the assemblage and not predefined subassemblages will not necessarily respect the additivity rule. For MaxEnt theory in general and METE in particular, there are better and worse ways to set up the state variables and, hence. the constraints that drive the theoretical predictions: if the subassemblages are very distinct (birds and insects) then the least-biased estimate of the SAR for the whole assemblage is obtained by adding the predicted SARs for the subassemblages (because there is more known constraint information). If the subassemblages lead to ambiguity, a more accurate estimate will be obtained by upscaling the whole, not by adding the predicted SARs of the subassemblages. Biological considerations are required to resolve the trade-off between ambiguity and additional constraint information.

Conclusion

Taxon invariance does not challenge the universality of METE's predicted SAR. Universality in METE comes from the fact that standardizing by the richness and abundance of species at any spatial scale leads to the prediction that all log-log SAR slopes collapse onto one curve (fig. 1). Taxon invariance is respected by the theory because any taxonomic choice of an assemblage will produce the same macroecological pattern. METE's success, however, depends on the assumption that taxa are organized in an ecologically and evolutionarily meaningful and relevant way. We suggest that ecological and/or evolutionary insights may be obtained by determining empirically which criteria for grouping species yield more accurate answers for scaling species richness. The merit of pursuing that agenda is reinforced by the fact that evidence supporting the idea that theory should obey the additivity rule is scant (Šizling et al. 2011), while METE has been tested successfully on many data sets (Harte 2011).

Acknowledgments

We thank M. Wilber for useful conversations about the universality of the species-area relationship. This work was funded in part by the National Science Foundation and by the Gordon and Betty Moore Foundation.

Literature Cited

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

Harte, J., A. Smith, and D. Storch. 2009. Biodiversity scales from plots to biomes with a universal species-area curve. Ecology Letters 12:789–797.

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

Jaynes, E. 1979. Where do we stand on maximum entropy? Pages 15–118 *in* R. Levine and M. Tribus, eds. The maximum entropy principle. MIT Press, Cambridge, MA.

———. 1982. On the rationale of maximum entropy methods. Proceedings of the Institute of Electrical and Electronics Engineers 70: 939–952.

———. 2003. Probability: the logic of science. Cambridge University Press, Cambridge.

Šizling, A., W. Kunin, E. Šizlingova, J. Reif, and D. Storch. 2011. Between geometry and biology: the problem of universality of the species-area relationship. American Naturalist 178:602–611.

Storch D., and A. L. Šizling. 2008. The concept of taxon invariance in ecology: do diversity patterns vary with changes in taxonomic resolution? Folia Geobotanica 43:329–344.

Associate Editor: Kevin J. Gaston
Editor: Mark A. McPeek

This content downloaded on Fri, 25 Jan 2013 12:04:05 PM
All use subject to JSTOR Terms and Conditions

View publication stats