Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities

Authors and affiliations:

Jonathan M. Chase*; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig and Department of Computer Science, Martin Luther University, Halle-Wittenberg; jonathan.chase@idiv.de

Brian J. McGill*; School of Biology and Ecology & Mitchell Center for Sustainability Solutions, University of Maine; mail@brianmcgill.org

Daniel J. McGlinn; Biology Department, College of Charleston, Charleston, SC; mcglinndj@cofc.edu

Felix May; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig; felix.may@idiv.de

Shane A. Blowes; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig; shane.blowes@idiv.de

Xiao Xiao; School of Biology and Ecology, University of Maine; xiao@weecology.org

Tiffany M. Knight; Dept. Community Ecology, Helmholtz Centre for Environmental Research – UFZ; Institute of Biology, Martin Luther University Halle-Wittenberg; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig. tiffany.knight@idiv.de

Oliver Purschke; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig; oliver.purschke@idiv.de

Nicholas J. Gotelli; University of Vermont, Department of Biology; Nicholas.Gotelli@uvm.edu

*Joint first authors

Statement of Authorship: JC and BM conceived the study and the overall approach, and all authors participated in multiple working group meetings to develop and refine the approach. JC collected the data for the meta-analysis that led to Figure 2 and S1; BM collected the data for the meta-analysis that led to Fig. 6,7; SB and FM did the meta-analyses; FM did the simulations for Fig. 5; DM, FM and XX wrote the code for the analysis used for the recipe and case study in Figure 8. JC, BM and NG wrote first drafts of most sections, and all authors contributed substantially to revisions.

Data accessibility statement: All data for meta-analyses and case study will be deposited in a publically available repository with DOI upon acceptance (available in link for submission).

Running Title: Scale-dependence in biodiversity comparisons

Key Words: Species-area relationship; species richness; evenness; Hill number; Simpson's index; rarefaction; scale-dependence

Type of Article: Reviews and Syntheses

Number of references: 74

Number of words in abstract: 195

Number of words in main text: 7497

Number of Tables: 1

Number of Figures: 8

Mailing address of corresponding author:

Jonathan Chase

German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig

Deutscher Platz 5e

04103 Leipzig

Germany

Abstract:

Although spatial scale plays a critical role in estimates of biodiversity, most empirical studies ignore scale or use only simple controls for sampling intensity. In a meta-analysis of studies that measured biodiversity responses to numerous ecological drivers across scales, we found nearly universal scale-dependence: effect sizes either greatly increased or decreased across scales, and nearly 10% switched directions across scales. Next, we show how accumulation and rarefaction curves can be used to dissect the effects of biodiversity scaling based on three components: total number of individuals (N), the shape of the species abundance distribution (SADs), and the pattern of spatial aggregation. In a second meta-analysis of studies with multiple rarefaction curves, ~30% of them crossed, and the three components of biodiversity were uncorrelated, contradicting conventional wisdom that most biodiversity measures are strongly intercorrelated. These results imply that there is no single 'magic metric' or scale for measuring biodiversity, and that multiple measures and scales are necessary to quantify biodiversity patterns. We use a case study of nutrient additions in experimental ponds to illustrate how this multi-scale perspective reveals the responses of biodiversity to ecological drivers, and allows for a more informed search for possible mechanisms.

Introduction

Biodiversity is fundamentally a scale-dependent metric, making comparisons complex and challenging. This challenge has been recognized for decades, and ecologists have searched for solutions to this problem (e.g., Preston 1960, MacArthur and MacArthur 1961, Hurlbert 1971, Hill 1973, Lande 1996, Gotelli and Colwell 2001, Jost 2006, Chao and Jost 2012, Chase and Knight 2013). Nevertheless, most empirical studies measure a single variable, species richness, at a single spatial scale. And theoretical ecologists have continued the search for a single 'magic' metric of biodiversity and its comparison.

Rather than aiming towards a single metric or appropriate scale for comparisons of biodiversity, we argue that it is better to embrace scale as a modifier when comparing biodiversity measures (e.g., across treatments). Many studies demonstrate that the magnitude, and even direction, of biodiversity change in response to natural and anthropogenic factors depends on spatial scale (e.g., Chase and Leibold 2002, Dumbrell et al. 2008, Keil et al. 2011, Powell et al. 2013a). The focus on species richness at a single spatial scale limits our ability to understand and synthesize how biodiversity changes in response to natural and anthropogenic factors. However, it is unclear how general scale-dependence is, and how exactly to account for it in empirical analyses.

The scale-dependence of biodiversity change is not just a theoretical issue, but is a concrete problem affecting biodiversity conservation and management. For example, Hill and Hamer (2004) showed that tropical forest bird diversity tended to decrease following logging when the scale of studies was relatively small (i.e., less than 25 ha), whereas studies at scales greater than 25 ha tended to show an increase in bird diversity in logged forests. Here, understanding the

scale-dependence of the result of the disturbance-diversity relationship illuminates the likely mechanisms. Diversity declined at small scales because forest specialists were negatively influenced by logging, whereas diversity increased at larger scales because species with habitat preferences for forest edge or open areas benefitted from logging.

In the first part of this paper, we present a meta-analysis of scale-dependent changes in species richness between treatments of ecological drivers. Next, we discuss methods and tools that capture key aspects of scale dependence to achieve an intermediate "Goldilocks approach" – not more complicated than necessary, but complex enough to adequately describe biodiversity change. We focus on three components underlying species richness: the numbers of individuals (*N*), the shape of the species abundance distribution (SAD) and the aggregation of species in space, and we highlight metrics that can capture aspects of these components. In another meta-analysis comparing empirical rarefaction curves (i.e., studies with information about SADs), we show that these components are not strongly correlated, indicating that biodiversity change can emerge from changes on any of these components, and that rarefaction curves can often cross, qualitatively switching the direction of biodiversity change across scales. Finally, we provide a recipe and case study showing how our approach can illuminate much about the patterns, and potential underlying processes, of biodiversity change.

Empirical evidence that scale-dependence is the rule, not the exception

Sample grain (i.e., the size of the sampling unit) and sample extent (i.e., the collective area encompassed by all sampling units) each influence observed values of biodiversity and the measure of biodiversity change between different communities (e.g., Palmer and White 1994,

Scheiner et al. 2000, Sandel and Smith 2009). Sample grain size is often chosen on the basis of sampling efficiency, time constraints, environmental impact concerns, and/or tradition. For example, sample grain for herbaceous plants is often $1m^2$, whereas sample grain in plankton surveys is determined by the net size and tow length. Although most biodiversity studies are not explicitly designed to address scale dependence, it can still be detected by quantifying the effect of sampling effort (number of individuals or samples collected) on estimates of biodiversity.

We use two examples to illustrate scale-dependence of biodiversity responses to ecological drivers (Figure 1). First, Powell et al. (2013a) examined the effect of a dominant plant invader on native plant biodiversity in replicate invaded and uninvaded communities (Fig 1a) and found that plant species richness decreased with the presence of the invader in $1m^2$ plots (t-test: t=3.9; df=4; P=0.018) but was not significantly different in a 500 m^2 area (the total extent of the community considered: t=2.5; df=4; P=0.08). Importantly, the effect size resulting from these two grains is quite different (log-response ratio of 1.39 for $1m^2$ vs. 0.1 for $500m^2$), calling into question the popular practice of combining studies on species richness across grains into a simple meta-analysis (e.g., Whittaker 2010, Chase and Knight 2013). In the second example, Lan et al. (2015a) examined the influence of sample grain on responses of plant species richness to nitrogen deposition in Mongolia (Fig. 1b). Although there was a significant negative relationship between the level of N addition and species richness at both smaller ($1m^2$) and larger ($25m^2$) grains, the slope of the relationship was much steeper at the larger grain size (y=19.54-0.30x) relative to the smaller grain size (y=7.89-013x), and one would conclude a much stronger effect of N deposition when considering larger relative to smaller grains. Conserving biodiversity is an important target for many land managers, and these examples illustrate that the applied

implications of these results (i.e., how concerned to be about threats of invaders and N deposition) depend on the scale of the study.

Are these scale-dependent results the rule or the exception? Although multiple-scale analyses are uncommon in the biodiversity literature, there are enough to allow for a quantitative meta-analysis. We performed a meta-analysis to discern the frequency, direction, and magnitude of changes in effect size of ecological drivers as a function of the spatial scale (sampling grain and/or extent) of a study. To do so, we identified possible studies using an ISI Web of Science search with the key words to identify studies that measured an index of diversity ("species richness or diversity or biodiversity") AND an indication that scale was explicitly considered ("scale or grain or extent") AND "ecology" (to eliminate hits that were in other fields). This yielded ~8,500 papers, from which we then read titles and if promising, abstracts, to determine whether studies measured diversity at multiple scales in response to an ecological driver. For relevant studies, we examined the results, figures, tables, and supplemental information to determine whether species richness at more than one scale in response to an ecological factor could be extracted. We found 103 comparisons within 52 studies (several studies reported responses from more than one driver, taxonomic group, or study site). A list of the studies and their data are given in the Supplemental material (Appendix 1).

Figure 2a shows the conceptual expectation for comparisons of the (log ratio) effect sizes of an ecological driver on species richness at small (x-axis) and the large (y-axis) scales. If scale-dependence were not important, most of the studies would fall near the 1:1 line, whereas the points would fall in the gray triangles if scale-dependence were important. If scale dependence influenced the qualitative direction of the effect (e.g., reversal from positive to negative), the

points would fall in the 'flip flops with scale' area of the figure. Figure 2b shows that scale dependence influences the magnitude of the effect of the factor on biodiversity in most comparisons in our meta-analysis, and the direction of the effect in eight comparisons.

In Figure 2b, there was only a weak relationship between effect sizes measured at large vs. small spatial scales (R^2=*0.14*); thus observations at one scale poorly predict effect sizes at another scale. Moreover, there was a trend for studies to show larger effects at smaller, relative to larger spatial scales (χ^2 = 3.1765, df = 1, p = 0.07). We categorized the studies based on the ecological factor considered: disturbance, environmental change (e.g., climate factors, N-deposition), fragmentation/dispersal (e.g., fragmentation experiments, propagule addition experiments), grazing/predation, invasive species, and land use (e.g., intensity of urbanization, agriculture). Studies of invasive species, land use change, and grazers/predators found significantly higher effect sizes at small compared to larger spatial scales (see supplemental Figure S1). This implies that effect size estimates from studies on these factors at a single small spatial scale may overestimate their effects relative to that which would be observed at larger scales. Indeed, in a meta-analysis, Powell et al. (2011) found that the effect sizes of invasive species on plant richness systemically declined with increasing spatial scale. Because such small-scale estimates of biodiversity change are often used as inputs, for example, into 'biodiversity scenario' models that project future diversity loss (e.g., Alroy 2017, Newbold et al. 2017), we suspect that these will overestimate the actual change observed at more realistic larger scales.

Biodiversity scaling theory indicates that scale-dependence is inevitable

Species accumulation curves (SACs) are a general way to depict how species richness increases with increasing sampling effort (e.g., area, plots, numbers of individuals). SACs have an intercept at the origin (i.e., zero area or zero individuals have zero species), and rise non-linearly with increasing sample effort, often in a decelerating way (though they can also accelerate over some spatial grains, especially at larger scales). As a result, SACs between two or more communities cannot be parallel, unless they are identical (i.e. lie on top of each other). This inevitably leads to scale-dependence of most biodiversity measures. SACs of two communities will either converge, diverge, or crisscross with increased sampling effort if they differ in any of three underlying components that influence the shape of the SAC: the numbers of individuals (N), the size of the species pool and relative abundance of species (which we refer to collectively as the species abundance distribution (SAD)), and intraspecific aggregation (clumping) (e.g., He and Legendre 2002 McGlinn and Palmer 2009, McGill 2011a, Chase and Knight 2013).

To illustrate how non-parallel SACs create scale-dependent biodiversity change, Figure 3 gives an example with three hypothetical communities (labelled A, B, C) each with a different SAC shape. The three vertical lines represent three scales (grains or extents), from which we can see that the grain size influences on the rankings of species density among the communities. Curves between two communities can diverge (e.g., A and C), consistent with the case study on nitrogen addition by Lan et al. (2015) described in Figure 1b and studies from the meta-analysis that showed larger effect sizes at larger spatial grains (Figure 2b). SACs between two communities can converge (e.g., B and C), consistent with the case study on invasive species by Powell et al. (2013a) described in Figure 1a and studies from the meta-analysis that showed smaller effect

sizes at larger spatial grains (Figure 2b). Finally, SACs can intersect and cross, so that biodiversity responses change sign with scale (e.g., A and B), consistent with the studies from the meta-analysis that showed shifts from positive to negative effect sizes with increasing spatial scale (Figure 2b). See James and Wamer (1982) for empirical examples of several SACs of bird diversity in different woodland habitat types that qualitatively resemble these hypothetical cases.

Three of the most commonly used and recommended approaches for dealing with the scaling problem in biodiversity studies include: (1) converting species richness into less 'biased' measures that incorporate the relative abundances of species, such as Shannon's or Simpson's diversity indices (e.g., Hill 1973, Lande 1996, Jost 2006, 2007); (2) Comparing species richness values after controlling for sampling effort in the numbers of samples or individuals through rarefaction (e.g., Hurlbert 1971, Simberloff 1972, Gotelli and Colwell 2001, Cayuela et al. 2015); or (3) Extrapolating species richness values to a hypothetical asymptote based on the estimation of the number of undetected species (e.g., Chao 1984, Colwell and Coddington 1994, Chao et al. 2009). Importantly, these have been combined into a single approach that combines interpolation and extrapolation (e.g., Chao and Jost 2012, Colwell et al. 2012), and explicitly incorporates measures of species relative abundances in the community (e.g., Chao et al. 2014a).

Despite these advances, the above methods typically aim towards a single, incomplete solution. For example, rarefactions that control for the numbers of individuals will still identify scale-dependent species richness rankings depending on the numbers of individuals to which species richness is rarefied (e.g., Gotelli and Colwell 2001, Cao et al. 2007, Chase and Knight 2013, Cayuela et al. 2015). On the other hand, if we were to use species richness extrapolations, we

are assuming that the only value of interest is the maximum number of species that is achieved at some hypothetical asymptote (e.g., scale 3 in Figure 3); focus only on the asymptote ignores important structural differences in communities at smaller spatial grains and extents. Finally, approaches that explicitly deal with differences in evenness (e.g., Shannon's diversity), can capture some of the differences observed among the SACs (e.g., Hill 1973, Chao et al. 2014a), but most studies only consider one measure at a time. Nevertheless, the approach we advocate below is similar to comparing comparison of multiple metrics that differentially weight common and rare species (e.g., Hill numbers; Hill 1973, Jost 2006, Chao et al. 2014a).

Next, we describe a framework that can more explicitly tease apart how the underlying components of biodiversity, namely N, the SAD and aggregation, influence comparisons. This approach synthesizes many different types of biodiversity metrics that are frequently used in empirical studies. In a companion paper, we provide a detailed exploration of the statistical methodology necessary to make these comparisons in a spatially continuous framework, including benchmark tests and evaluation of statistical error, and provide an R package (*mobr*) that can be used to analyze both spatially implicit and explicit data on changes in biodiversity among treatments (McGlinn et al. *submitted*).

Contrasting accumulation and rarefaction curves

Here, we overview two types of SACs—sample-based accumulation and individual-based rarefaction—that can reveal the influence of different components that underlie changes in species richness across scales (Gotelli and Colwell 2001). We illustrate these curves and how they connect to more traditional metrics of biodiversity in Figure 4 and in Table 1.

Sample-based accumulation curve—This curve illustrates how the number of species (y-axis) accumulates as the number of samples (i.e., number of plots, quadrats, net sweeps, etc) (x-axis) is increased in a spatially-explicit way (Figure 4a). To accumulate samples spatially-explicitly, one starts with a single plot, then adds the closest plot, then the next closest plot, etcetera until all plots are included. Because the choice of the initial starting plot influences the shape of this curve, we recommend averaging the curve across multiple starting plots (also called spatially constrained rarefaction; Chiarucci et al. 2009). This average curve, and measurements of species richness at a given sampling scale, is sensitive to changes in all three components that underlie species richness (SAD, N, and aggregation) (e.g., Powell et al. 2013a, Chase and Knight 2013).

Individual-based rarefaction curve—In contrast to the above curve, which is the most similar to what the ecologist observes in the field, and is the most complex (retaining information on all three components), the individual-based rarefaction curve is the least complex and most derived from what is observed in the field. Specifically, the individual-based rarefaction curve plots species richness (y-axis) against number of individuals (x-axis) (Figure 4b). Individuals are pooled across all samples and selected at random, so that any spatial structure in intraspecific aggregation and differences in the numbers of individuals between plots is lost (e.g., Hurlbert 1971, Simberloff 1972, Gotelli and Colwell 2001). The shape of the rarefaction curve (i.e., slope and asymptote) depends only on the SAD. Specifically, an individual-based rarefaction curve is influenced both by the relative abundances of the community (a more even community has a steeper sloped curve than a less even community), as well as the total number of species in the community (a community with higher richness has a higher asymptote than one with lower richness) (Hurlbert 1971, Olszewski 2004, Cayuela et al. 2015). The slope of the rarefaction

curve at its base is equivalent to the probability of interspecific encounter (Hurlbert 1971, Olszewski 2004) (see below for more details).

In individual-based rarefaction curves, we can pool and randomly sample individuals at multiple spatial scales. For example, if we have multiple sampling plots (e.g., replicates) within a given treatment, we can create an individual-based rarefaction curve for each sample and then average these curves to get a mean curve seen at the α–scale (scale that represents the number of individuals found in a single sample; dashed line in Figure 4B). Or we can pool individuals across all plots into one large pool (γ-scale, solid line in Figure 4B). The difference between the two curves indicates some site-to-site variability (i.e., β-diversity). If the γ-rarefaction curve rises much more steeply than the α-rarefaction curve then this indicative of high β-diversity (spatial turnover of species composition). This results from the spatial aggregation of individuals within a species (causing a single plot to contain more individuals of one or a few species and thus fewer total species than if the individuals were spatially random) (Olszewski 2004, Dauby and Hardy 2012) (Figure 4b)

Single-scale biodiversity measures——Here we consider the case for which we only know one value for the observed numbers of species (S) and for number of individuals (N) per treatment. In this case, we focus on comparing S and N across treatments. While not a biodiversity metric per se, N is a critical feature of natural communities and differences in N can be quite useful for making inference about any observed changes in biodiversity metrics such as S. For example, if S and N change between treatments, then at least some of the difference in S is likely due to

changes in N, whereas if N does not differ strongly, then the effects on S are more likely due to changes in the SAD and/or aggregation.

Second, we can explicitly evaluate the influence of changes in N on any changes in S by estimating rarefied species richness. Rarefaction is usually accomplished by rarefying the expected species richness given the lowest N observed among all samples. However, differences in rarefied species richness are also scale-dependent (e.g., Cao et al. 2007, Chase and Knight 2013, Cayuela et al. 2015), and so a better approach is to measure rarefied richness at multiple values of N, which we refer to as S_n. Values of S_n can be compared to each other quickly across all scales by visually comparing whether one rarefaction curve is above or below another.

Third, many empiricists are interested in the total numbers of species in a given community, knowing that many might be undetected using typical survey methods. Chao and others (e.g., Chao 1984, Chao et al. 2009, Colwell et al. 2012, Chiu et al. 2014a) have created a family of estimators that estimate the asymptote of the individual-based rarefaction curve (see e.g., Colwell and Coddington 1994, Reese et al. 2014 for other types of estimators with similar goals). We refer to this as $S_{asymptote}$, although we note that its accuracy requires adequate sampling and other assumptions to be met (Chao et al. 2009, Reese et al. 2014). $S_{asymptote}$ is highly correlated with observed S (McGill 2011b), but we can also use it to estimate the numbers of undetected species at either scale ($S_{asymptote} - S$)(Chao et al. 2009) as an indicator the number of undetected rare species; this parameter is referred to as f_0.

Fourth, S is an imperfect descriptor of diversity in a community. Even after controlling for any differences in N, we could see differences in the shape of the individual-based rarefaction curve in its rising slope that are not captured simply by looking at S_n or $S_{asymptote}$. Specifically, while the asymptote of the curve is described by the total S in the community, the shape of the curve and how it rises is determined by the SAD, including the commonness and rarity of species (i.e., evenness of the community). Olszewski (2004) showed how the slope of the rarefaction curve at its base (between N=1 and N=2) is equivalent to a relatively unbiased estimator of evenness in a community, the Probability of Interspecific Encounter (PIE) (Hurlbert 1971); with the PIE simply being 1-Simpson's original index (Jost 2006). A more even community (higher PIE) accumulates species more quickly with increasing N than a less even community (lower PIE).

Given that individual-based rarefaction curves capture elements of the total species in a community as well as their relative abundances, they can also be used to visualize the diversity continuum known as Hill numbers. Hill (1973) showed that some of the most familiar metrics used to measure species diversity, including S, Shannon's index, and Simpson's index, differed in how they weight rare species (denoted with q). Reyni's (1961) and Tsallis' (1988) entropies represent similar weighting continua (see also Keylock 2005, Jost 2006, Mendes et al. 2008, Chao et al. 2014a). Specifically, the upper-right part of the rarefaction curve where all species are weighted equally regardless of their abundance (q=0) gives us S, and the slope at the lower-left (base) of the rarefaction curve pinpoints the other end of the Hill continuum (q=2) at PIE (=1-Simpson's index), where common species are emphasized and rare species are down-weighted; Shannon's index (q=1) is somewhere between these extremes.

For statistical comparisons, it is useful to convert diversity entropies (such as Shannon's or PIE) into an effective number of species (e.g., Hill 1973, Jost 2006, 2007). For PIE, the effective number of species (S_{PIE}) is the number of equally abundant species it would take to yield a given value of PIE. For example, in a community where all species had the exact same abundances, S_{PIE} would be equal to S. As the community becomes almost completely dominated by one species, but S stays the same so that all others are extremely rare, PIE decreases towards 0 and S_{PIE} decreases towards 1. While Jost (2006, 2007) and others (e.g., Chao et al. 2014a) have suggested that Shannon's index (which is intermediate between S and Simpson's index (=1-PIE) in its weighting of rare species) has the best statistical properties, we argue that it is more informative to focus on more extreme measures at S and PIE, which capture more of the potential variation among communities. Nevertheless, a more complete view of points along the rarefaction curve can only benefit from comparisons of multiple orders of the Hill continuum.

Two-scale biodiversity measures—Above, we discussed each measure at a single sampling scale. However, we often have several plots (or replicates) for a given treatment. Using Whittaker's (1960) concept of multiplicative diversity partitioning ($\gamma=\alpha*\beta$), we define ${}^{\alpha}S$ as the *average* S per plot (=observational grain) and ${}^{\gamma}S$ as the *count* of species (S) across all samples, plots or replicates of a given comparison (=observational extent). Finally, we define the ratio of ${}^{\gamma}S$ to ${}^{\alpha}S$ as β_s (= ${}^{\gamma}S/{}^{\alpha}S$) to quantify how species are gained from small to large scales (i.e., β-diversity)(we do not discuss other ways β-diversity is sometimes calculated, which address different questions (Tuomisto 2010a,b; Anderson et al. 2011). Although β-diversity is a useful concept, we note the caveat that its calculation assumes a linear difference from two scales of an inherently non-linear SAC, and thus is itself scale-dependent (e.g., MacNally et al. 2004, Nekola

and McGill 2014, Zhang et al. 2015, Tuomisto et al. 2017). However, other approaches to measuring scaling relationships (e.g., via the slope [z] of the species area relationship) are influenced by the same parameters we describe here.

Just as with $^{\gamma}S$ and $^{\alpha}S$, β_S is influenced by N, SAD, and aggregation. However, we can use the individual-based rarefaction curve collected from small and large scales to isolate the influence of aggregation on β-diversity from the influence of N and SAD (see McGlinn et al. *submitted* for a method based on a continuous scale of spatial grain to estimate the aggregation effect). Because the individual-based rarefaction curve randomly samples individuals regardless of their spatial position, it eliminates the influence of aggregation. But if we sample randomly from only a subset of the spatial distribution (i.e., the α-scale), and there is strong intraspecific aggregation, the rarefaction curves from the smaller and larger scales will capture different amounts of this heterogeneity (Figure 4b). We can quantify this aggregation by comparing the slope of the rarefaction curve (i.e, PIE) of the two scales; the bigger the difference in PIE between the α- and γ-scale, the more aggregation there is. For comparisons, we convert to the effective numbers of species and take the ratio of small-scale ($^{\alpha}S_{PIE}$) to large-scale ($^{\gamma}S_{PIE}$) to yield an estimate of β-diversity ($\beta_{S_{PIE}}$) that is driven primarily by aggregation of regionally common species.

As above, the β-diversities described here are equivalent to those derived from Hill numbers calculated when all species have equal weights (=β_S) and when common species are given much higher weights than rare species (=$\beta_{S_{PIE}}$) (Jost 2007, Tuomisto 2010a,b). It may also be useful to compare other values of β-diversity using this framework, including β_{S_n} or $\beta_{S_{asymptote}}$.

Our use of rarefaction curves makes a very important assumption that numbers of individuals are counted and that this is meaningful. This assumption is often not upheld in empirical case studies on organisms where N is difficult or impossible to distinguish (e.g., herbaceous plant communities), or in organisms where the sizes of individuals are so different that biomass might be a more reasonable way to indicate dominance in a community rather than N. Nevertheless, it is still possible to calculate 1-Simpson's index (=PIE) when other estimates of relative abundances are taken, though the interpretation is slightly different. In addition, there are imperfect ways to approximate the rarefaction methods we advocate here, for example using biomass or percent cover as a proxy for numbers of individuals.

Comparing Species Accumulation and Rarefaction Curves

To illustrate how N, SAD, and aggregation alter the shape of accumulation and rarefaction curves, we varied them in a simple spatial simulation using the R package mobsim (see https://github.com/MoBiodiv/mobsim) (May et al. 2017). We do not present all combinations of changes, but rather show how changing key components influence the sample-based accumulation (left panels) and individual-based rarefaction curves (right panels) (Figure 5).

The effect of changing N

When only the number of individuals changes, the SACs differ, but the individual-based rarefaction curves are on top of each other (but one extends further to the right; Figure 5A). This is sometimes referred to as the 'more individuals hypothesis' (Srivistava and Lawton 1999, Currie et al. 2004, Hurlbert and Jetz 2010), and can arise when some process (e.g., energy, predation) alters the numbers of individuals that can persist in a site.

The effect of changing the SAD

Although it is common to summarize the shape of the SAD with a single measure of evenness, we argue that it is better to distinguish two types of changes to the SAD: 1) Increasing dominance, which decreases PIE; or 2) adding more individuals of the rare species which will increase S, but have little effect on PIE (see also Thompson and Withers 2003). Any combination of these two scenarios is possible. Thus, dominance can change without affecting the rare species (Figure 5B), rare species can increase while dominance is unchanged (Figure 5C), both dominance and rare species can increase or decrease in parallel (Figure 5D), or dominance can increase while rare species decline, leading to crossing curves (Figure 5E). Changes in the shape of the SAD are most often associated with changes from one community to another in the processes that shape coexistence, dominance and the persistence of rare species.

The effect of changing spatial aggregation

If individuals in a community are spatially random, the individual-based rarefaction would be identical at smaller and larger scales, as would the univariate parameters that describe this (${}^{\alpha}S_{\mathrm{PIE}} = {}^{\gamma}S_{\mathrm{PIE}}$). However, if the individuals in a community are strongly aggregated, the smaller-scale curve will be shallower than the larger-scale curve (${}^{\alpha}S_{\mathrm{PIE}} < {}^{\gamma}S_{\mathrm{PIE}}$), leading to $\beta_{S_{\mathrm{PIE}}}$. Aggregation can emerge from habitat heterogeneity, dispersal limitation, and a number of other processes.

Components of species richness vary independently

Is this complex decomposition of rarefaction curves really necessary? Empirical ecologists often suggest that because different measurements of biodiversity are often strongly correlated (e.g.,

species richness and Shannon's index), only one measure is needed to understand the response of biodiversity to ecological factors. Here we explicitly examine this claim to determine whether the extra complexity we advocate is necessary.

In order to evaluate how strongly correlated the different measures of biodiversity are, we analyzed datasets for which we could calculate rarefaction curves from multiple sites within a given study. We specifically analyzed 37 datasets from a larger assemblage of datasets amassed by McGill 2011b that had species abundance data from at least two comparable local communities (data and references listed in Appendix 2). We compared whether differences among sites in N and the SAD (indicated by S_{PIE} and S) were correlated. There was not enough spatial information within these datasets for us to examine aggregation. Example datasets include data from bird surveys along standardized routes in North America, marine benthic invertebrates collected from trawls along an environmental gradient, and arthropods on different host plants. Some of these datasets occurred along natural (e.g. latitudinal) or human-caused (e.g. oil leakage) environmental gradients. Other data sets were variable, but not arranged along any known gradient. Twenty-two of the datasets came from terrestrial ecosystems and 15 were from marine ecosystems, but we found no consistent differences between habitats. Most datasets compared 2-20 communities for comparisons; datasets with >20 communities had 20 randomly selected so that they did not numerically dominate the results.

First, we tested the idea that N, S and S_{PIE} are redundant (correlated) by looking at the three bivariate relationships between S, N and S_{PIE} using hierarchical linear models. Specifically, we modeled S_{PIE} as a function of N, and S as a function of either N or S_{PIE} using models with a fixed

effect to estimate the strength of the relationship across datasets, and random slopes and intercepts to examine variation between datasets. N and S were modelled assuming Poisson error and a log-link function, and S_{PIE} was log-transformed and modelled assuming Gaussian error; N, S_{PIE} and S entered all models as log-transformed covariates. All models were fit in R (R Development Core Team 2017) using the lme4 package (Bates et al. 2015). To compare variation across versus between datasets, we quantified the marginal (fixed effect only) and conditional R^2 (fixed + random effects) using the piecewiseSEM package (Lefcheck 2016).

Figure 6 shows that N and S_{PIE} were unrelated across all studies (R^2 = 0.02), and that the relationships between S and N, and S and S_{PIE} were weak across studies (R^2 = 0.14-0.16). Importantly, Figure 6 also shows that there is considerable variation in the bivariate relationships within individual datasets with trends in all directions. In particular, the variation explained when the relationships were allowed to vary between studies increased dramatically (R^2 increases to 0.81 – 0.95). Thus, specific ecosystems have correlations between these variables, but we do not know *a priori* what type or strength of correlation. Because these variables vary largely independently and in inconsistent directions, they likely contribute unique information that provides a more nuanced understanding of how biodiversity differs between sites.

Second, we evaluated a commonly held assumption that real-world rarefaction curves of comparable data do not usually cross, so that the rank order of diversity differences among communities is typically consistent (even if the effect sizes change) (but see, e.g., Lande et al. 2000, Thompson and Withers 2003). For this analysis, we compared individual-based rarefaction curves for each pair of communities within a single dataset (Figure 7A). Across all

datasets, a total of 2203 pairwise comparisons were made and 732 pairs of rarefaction curves crossed each other (33% of all pairs) (some crossing multiple times) (Figure 7B). Importantly, the crossing points were not ecologically trivial (e.g., crossings only at very small N). The average crossing occurred at 24% of the total abundance in a community and the average vertical separation on curves that crossed was 6.3 species (vs an average of ~40 species in a community).

In sum, these analyses show that the different components (N, S, S_{PIE}) of rarefaction curves among two or more communities are relatively independent, and that rarefaction curves often cross. Importantly, however, this analysis was able to compare only spatially implicit curves, and we were not able to address the role of spatial aggregation, which would further complicate comparisons that take a scale-agnostic perspective.

Recommended protocol for analyses dissecting biodiversity data with a case study

Here, we provide an empirical recipe for using multiple metrics at multiple scales to achieve deeper insights into the patterns and potential processes by which ecological drivers influence biodiversity. We first describe how one might address these questions generically, and then we illustrate this recipe with a re-analysis of an experiment on the effect of nutrient additions on macroinvertebrates and amphibians in experimental ponds (for full details, see Chase 2010).

In a companion paper (McGlinn et al. *submitted*), and the *mobr* statistical package (https://github.com/MoBiodiv/mobr), we provide the methodology and R code for these analyses and how to dissect the influence of N, SAD and aggregation in a spatially continuous way. For these analyses, we used one-way PERMANOVA to assess the treatment effect on diversity at the

α-scale (=1 mesocosm) (permuted treatment group labels 999 times for significance). At the γ-scale (=15 mesocosms), we permuted treatment group labels across samples, pooled the groups, and calculated the difference in diversity between treatments for each permutation.

Results are illustrated in Figure 8; some results are not shown for brevity:

Step 1: Do treatments affect the numbers of individuals of all species (N)?

If treatments differ in N, then any difference in S between treatments may, in part, be due to treatment effects on N. We expect that N often scales linearly with increasing sampling (unlike S), and so it should not matter whether the average N per plot (α-scale), or the total N across plots (γ-scale), are compared. In our case study, we found no influence of the nutrient addition treatment on N measured at either scale (P>0.3) (not shown in Figure).

Step 2: Do treatments affect diversity responses at the α-scale?

a) Do treatments affect $^{\alpha}S$?

To address whether the treatment influences the number of species in a single plot, we compare $^{\alpha}S$ (or $^{\alpha}S_{\text{asymptote}}$) between the reference and treatment samples. In our case study, $^{\alpha}S$ was equivalent to $^{\alpha}S_{\text{asymptote}}$ for each treatment, and thus we just show results for $^{\alpha}S$. However, we note that if we had poorly sampled the diversity of each sample, $^{\alpha}S_{\text{asymptote}}$ would have been a more appropriate response variable. If we find that the treatment does affect the species per sample, this could be due to effects of treatments on N and/or the SAD. Importantly, however, it is still useful to explore the steps below, even if there is no difference in $^{\alpha}S$, because it is possible that (1) changes to N and the

SAD can have equal but opposite effects on their influence on $^{\alpha}S$ (e.g., if the treatment changes both N and the SAD), or (2) that effects only emerge at larger scales. In our case study, $^{\alpha}S$ did not differ between the treatments (Fig. 8a).

b) Do treatments affect $^{\alpha}S_n$? Dissecting the influence of N

If there are large differences in N among treatments (as in Step 1), it is possible that the results for $^{\alpha}S$ and $^{\alpha}S_n$ will differ owing to the treatment effect on N. For example, if a significant treatment effect on $^{\alpha}S$ is no longer significant for $^{\alpha}S_n$, we can conclude that the effect on $^{\alpha}S$ was primarily due to differences in N between treatments, rather than due to differences the SAD. However, if there remains a difference in $^{\alpha}S_n$, we can conclude that some of the treatment difference in $^{\alpha}S$ was due to changes in the SAD. It is even possible that the effects can shift between $^{\alpha}S$ and $^{\alpha}S_n$. For example, McCabe and Gotelli (2000) showed that $^{\alpha}S$ was higher for stream invertebrates in undisturbed treatments, but that this effect was reversed when the effect of disturbance on N was discounted; whereas $^{\alpha}S_n$ was higher in disturbed treatments. Nevertheless, in our case study, because there was no difference in N between treatments, the effects on $^{\alpha}S_n$ were the same as for $^{\alpha}S$ and did not differ among treatments (result not shown in figure).

c) Do treatments differ in $^{\alpha}S_{PIE}$? Dissecting the influence of SAD

If there is a difference in $^{\alpha}S_n$ from above, we can infer that the SAD has changed between the treatments. However, the SAD can change by influencing either the common species or the rare species in a community. A difference in $^{\alpha}S_{PIE}$ between treatments emphasizes a

change in the dominance patterns of common species in a community. Thus, if a treatment decreases the value of ${}^{\alpha}S_{PIE}$, this indicates that the treatment has caused species to become increasingly dominant, and if it increases ${}^{\alpha}S_{PIE}$, this means that dominance has decreased and the community has become more even. Alternatively, it is possible that ${}^{\alpha}S_{PIE}$ could not vary (or vary little), with a large influence of the treatment on ${}^{\alpha}S_{n}$. This would imply that rarer species that do not influence PIE are responding to the treatment. It is also possible to directly estimate how a treatment might influence rare species (e.g., f_0 described above), but this requires an estimate of undetected (rare) species, which was not true in our case study. Instead, our case study showed that that there was no influence of treatment on ${}^{\alpha}S_{PIE}$ (Figure 8b) or ${}^{\alpha}S$ (above), and thus treatments did not differ in the shape of the SAD at the local scale.

In sum, none of the response variables of interest were significantly different at the α-scale. Had we stopped here, at a majority of biodiversity experiments do, we might conclude that there was no interesting influence of nutrients on biodiversity. However, by looking at the γ-scale (i.e., species richness across all mesocosms), we see that this conclusion is wrong.

Step 3: Do treatments affect diversity responses at the γ-scale?

a) Do treatments affect ${}^{\gamma}S$?

As in Step 2a, but at the regional scale. Again, any observed treatment level differences could result from changes in N or the regional-scale SAD. And importantly, the local and regional scale SAD need not be the same. For example, if there is strong intraspecific aggregation among plots (which is typical; McGill 2010), as might result from habitat

heterogeneity or dispersal limitation, we can expect the shapes of regional and local SADs to differ. In our case study, and in contrast to the local sale results ($^{\alpha}S$), we found a significant increase in $^{\gamma}S$ in the nutrient added treatments relative to the controls (Fig. 8c).

b) Do treatments affect $^{\gamma}S_n$? Dissecting the influence of regional N

As in Step 2b, but at the regional scale. As above, rarefying species richness to a common N removes any influence of N on S, such that any observed differences in $^{\gamma}S_n$ are due to changes in the regional SAD, which itself is a product of the local SAD and aggregation. In our case study, because we found no differences in N among treatments, the results for $^{\gamma}S_{\mathrm{n}}$ was the same as for $^{\gamma}S$ (not shown).

c) Do treatments differ in $^{\gamma}S_{PIE}$? Dissecting the influence of the regional SAD

As in Step 2c, but at the regional scale. As above, this allows us to determine whether any differences observed in species richness are due to changes in the dominance of common species, which would influence $^{\gamma}S_{PIE}$, or instead due to changes in rare species, which would only influence $^{\gamma}S_n$ (and direct measures of rarity, such as f_0). In our case study, there were no differences in $^{\gamma}S_{\mathrm{PIE}}$ among treatments (Fig. 8d), despite the fact that $^{\gamma}S$ (and $^{\gamma}S_{\mathrm{n}}$) did change.

Overall, our case study showed that γ-scale results differed from α-scale results. Specifically, $^{\gamma}S$ increased with nutrient additions, but not $^{\gamma}S_{PIE}$. This suggests that it is the rare species end of the SAD that changed among the treatments; added nutrients allowed more rare species to persist regionally, but not locally. Further, the qualitatively different results at the γ-and α-scale implies that β-diversity was influenced by the nutrient addition treatment.

Step 4: Do treatments affect β-diversity?

β-diversity results reflect the contrast between the α- and γ-scale. Just as with the other scales, we can evaluate β-diversity as turnover of all species (β_S) and of the primarily dominant species ($\beta_{S_{PIE}}$). If the differences between the α- and γ-scale patterns were driven by one treatment causing strong shifts in dominance and interspecific aggregations from site to site, as might happen if a treatment creates strong compositional heterogeneity, we would expect this to be reflected in treatment-level effects on $\beta_{S_{PIE}}$. If instead, differences between the α- and γ-scale were driven by one treatment causing more by site to site replacements in rare species, either as a result of ecological drift, dispersal limitation, or some frequency-dependent process, this would be reflected only in β_S. In our case study, we found strong effects of nutrient addition on the β_S (Fig. 8e), but not $\beta_{S_{PIE}}$ (Fig. 8h). This emphasizes that a core group of common species were present across replicates in both treatments, and that a group of rarer species were largely responsible for the treatment-level responses at the γ-scale. Specifically, there was more turnover among those rare species in the high nutrient treatment; this likely resulted from ecological drift and/or priority effects that were initiated at the beginning of the experiment.

Conclusions

There are multiple pathways by which ecological factors can influence biodiversity across multiple scales, but most of studies continue to rely on comparisons of a single summary variable—usually species richness (S)—at a single spatial scale. As a consequence, despite thousands of published studies quantifying how species richness changes in response to natural and anthropogenic drivers, we know much less than we think about how and why biodiversity

changes from place to place and time to time. This is particularly problematic when trying to achieve synthesis across multiple studies of multiple ecological drivers, through meta-analyses and other means, because effect sizes are highly confounded by spatial scale (see also Chase and Knight 2013). We are currently limited in our ability to create realistic 'biodiversity scenario' models that project future biodiversity loss in response to changing ecological conditions.

To move forward, it is critical to consider the factors that underlie S and how it scales from local to regional spatial scales. Specifically, S at any given scale is determined the number of individuals and the SAD, as well as the degree of spatial clumping or interspecific aggregation that alters the SAD from smaller to larger scales. Fortunately, there is a rich literature on other measures of biodiversity that can explicitly complement comparisons of S, and there are easy ways that empiricists can explicitly deal with issues of spatial scaling (i.e., replicates nested within treatments). We have advocated for measures of biodiversity that can explicitly account for the factors underlying S and its scaling via consideration of different aspects of the individual-level rarefaction curve that emphasize different underlying components (e.g., S_n, S_{PIE}, at the α and γ scales). For a majority of studies, these can be estimated in a straightforward way with data that are already, or could be, collected.

The approach we advocate will often require more complex collecting and reporting of data (i.e., total relative abundances at multiple scales) and more complex analyses with multiple response variables. This will create a more nuanced view of what biodiversity is and how it varies—biodiversity is not a single number, and it cannot be compared at a single scale to estimate how it responds to ecological factors. However, this more nuanced view is necessary to resolve long-

standing debates. For example, Blowes et al. (2017) recently used an approach similar to the one we advocate here to show that the debate about whether environmental versus historical biogeographic controls influence global biodiversity patterns of reef associated fishes can be resolved by dissecting patterns of species richness in a scale-explicit way.

Finally, there are many extensions of the approach that we advocate which are necessary to be able to fully understand, and synthesize, how biodiversity changes in time and space. First, as we mentioned above, the approach we have taken here views scale in a categorical way (e.g., α, β, γ–diversity), while scale is continuous. In a companion paper, we develop a scale-continuous methodology (McGlinn et al. *submitted*). Second, we have only focused on taxonomic diversity, although interest in other measures of diversity has increased greatly in recent years. These measures, such as functional and phylogenetic diversity, show patterns of scaling similar to taxonomic diversity (e.g., Morlon et al. 2011, Smith et al. 2013) and will thus show scale-dependence when making comparisons. Approaches similar to those advocated here are emerging for these other types of diversity (Chao et al. 2014b, 2015, Chiu et al. 2104b), and so we anticipate that a family of approaches for comparing scale-dependent diversity responses to ecological drivers at multiple levels of organization will soon emerge.

Acknowledgements

This paper emerged from several working group meetings and extended visits funded by the German Centre of Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (funded by the German Research Foundation; FZT 118) and the Alexander von Humboldt Foundation as part of the Alexander von Humboldt Professorship of TMK. DJM was also supported by the College of

Charleston startup funding. NJG was supported by U.S. NSF DEB 1257625. BJM was supported by USDA Hatch grant to MAFES #1011538 and NSF ABI grant #1660000. In addition, we thank F. Walter for help extracting data for the meta-analysis presented in Figure 2, and J. Belmaker, N. Sanders, D. Storch, and a number of other colleagues that helped us to develop and reform these ideas and tools.

Literature Cited

Alroy, J., (2017). Effects of habitat disturbance on tropical forest biodiversity. Proc. Natl. Acad. Sci. 114, 6056–6061.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., et al, (2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecol. Lett., 14, 19-28.

Bates, D., Maechler, M., Bolker, B. and Walker, S., (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7), pp.1-23.

Blowes, S. A., Belmaker, J., and Chase, J. M. (2017). Global reef fish richness gradients emerge from divergent and scale-dependent component changes. Proc. R. Soc. B 284, 20170947

Cao, Y., Hawkins, C.P., Larsen, D.P. and Van Sickle, J., (2007). Effects of sample standardization on mean species detectabilities and estimates of relative differences in species richness among assemblages. Am. Nat., 170, 381-395.

Cayuela, L., Gotelli, N.J. and Colwell, R.K., (2015). Ecological and biogeographic null hypotheses for comparing rarefaction curves. Ecol. Monog., 85, 437-455.

Chao, A., (1984). Nonparametric estimation of the number of classes in a population. Scand. J. Stat., 11, 265-270

Chao, A. and Jost, L., (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology, 93, 2533-2547.

Chao, A., Colwell, R.K., Lin, C.W. and Gotelli, N.J., (2009). Sufficient sampling for asymptotic minimum species richness estimators. Ecology, 90, 1125-1133.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. and Ellison, A.M., (2014a). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol. Monog., 84, 45-67.

Chao, A., Chiu, C.H. and Jost, L., (2014b). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. Annu. Rev. Ecol. Syst, 45, 297-324.

Chao, A., Chiu, C.H., Hsieh, T.C., Davis, T., Nipperess, D.A. and Faith, D.P., (2015). Rarefaction and extrapolation of phylogenetic diversity. Methods Ecol. Evol., 6, 380-388.

Chase, J.M. and Leibold, M.A., (2002). Spatial scale dictates the productivity–biodiversity relationship. Nature, 416, 427-430.

Chase, J.M., (2010). Stochastic community assembly causes higher biodiversity in more productive environments. Science, 328, 1388-1391.

Chase, J. M., and Knight, T. M. (2013). Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. Ecol. Lett., 16, 17-26.

Chiarucci, A., Bacaro, G., Rocchini, D., Ricotta, C., Palmer, M. and Scheiner, S., (2009). Spatially constrained rarefaction: incorporating the autocorrelated structure of biological communities into sample-based rarefaction. Comm. Ecol., 10, 209-214.

Chiu, C.H., Wang, Y.T., Walther, B.A. and Chao, A., (2014a). An improved nonparametric lower bound of species richness via a modified Good–Turing frequency formula. Biometrics, 70, 671-682.

Chiu, C.H., Jost, L. and Chao, A., (2014b). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol. Monog., 84, 21-44.

Colwell, R.K. and Coddington, J.A., (1994). Estimating terrestrial biodiversity through extrapolation. Phil. Trans. Roy. Soc B, 345, 101-118.

Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.Y., Mao, C.X., Chazdon, R.L. et al. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. J. Plant Ecol., 5, 3-21.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.F., Hawkins, B.A., et al. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecol. Lett., 7, 1121-1134.

Dauby, G. and Hardy, O.J., (2012). Sample-based estimation of diversity sensu stricto by transforming Hurlbert diversities into effective number of species. Ecography, 35, 661-672.

Dumbrell, A.J., Clark, E.J., Frost, G.A., Randell, T.E., Pitchford, J.W. and Hill, J.K., (2008). Changes in species diversity following habitat disturbance are dependent on spatial scale: theoretical and empirical evidence. J. Appl. Ecol., 45, 1531-1539.

Gotelli, N.J. and Colwell, R.K., (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol. Lett., 4, 379-391.

He, F. and Legendre, P., (2002). Species diversity patterns derived from species–area models. Ecology, 83, 1185-1198.

Hill, M.O., (1973). Diversity and evenness: a unifying notation and its consequences. Ecology, 54, 427-432.

Hill, J.K. and Hamer, K.C., (2004). Determining impacts of habitat modification on diversity of tropical forest fauna: the importance of spatial scale. J. Appl. Ecol., 41, 744-754.

Hurlbert, A.H. and Jetz, W., (2010). More than "more individuals": the nonequivalence of area and energy in the scaling of species richness. Am. Nat., 176, E50-E65.

Hurlbert, S.H., (1971). The nonconcept of species diversity: a critique and alternative parameters. Ecology, 52, 577-586.

James, F.C. and Wamer, W. O. (1982). Relationships between temperate forest bird communities and vegetation structure. Ecology, 63, 159-171.

Jost, L., (2006). Entropy and diversity. Oikos, 113, 363-375.

Jost, L., (2007). Partitioning diversity into independent alpha and beta components. Ecology, 88, 2427-2439.

Keil, P., Biesmeijer, J.C., Barendregt, A., Reemer, M. and Kunin, W.E., (2011). Biodiversity change is scale-dependent: an example from Dutch and UK hoverflies (Diptera, Syrphidae). Ecography, 34, 392-401.

Keylock, C. J. (2005), Simpson diversity and the Shannon–Wiener index as special cases of a generalized entropy. Oikos, 109, 203–207.

Lan Z, Jenerette GD, Zhan S, Li W, Zheng S, Bai Y (2015a). Testing the scaling effects and mechanisms of N-induced biodiversity loss: evidence from a decade-long grassland experiment. J. Ecol. 103, 750-760.

Lan Z, Jenerette GD, Zhan S, Li W, Zheng S, Bai Y (2015b). Data from: Testing the scaling effects and mechanisms of N-induced biodiversity loss: evidence from a decade-long grassland experiment. Dryad Digital Repository. https://doi.org/10.5061/dryad.h7893

Lande, R., (1996). Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos, 5-13.

Lande, R., DeVries, P.J. and Walla, T.R., (2000). When species accumulation curves intersect: implications for ranking diversity using small samples. Oikos, 89, 601-605.

Lefcheck, J.S., (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Meth. Ecol. Evol., 7, 573-579.

MacNally, R., Fleishman, E., Bulluck, L.P. and Betrus, C.J., (2004). Comparative influence of spatial scale on beta diversity within regional assemblages of birds and butterflies. J. Biogeog., 31, 917-929.

MacArthur, R.H. and MacArthur, J.W., (1961). On bird species diversity. Ecology, 42, 594-598.

May, F., Gerstner, K. . McGlinn, D. J., Xiao, X and Chase, J. M. (2017) mobsim: An R package for the simulation and measurement of biodiversity across spatial scales. bioRxiv (2017): 209502.

McCabe, D.J. and Gotelli, N.J., (2000). Effects of disturbance frequency, intensity, and area on assemblages of stream macroinvertebrates. Oecologia, 124, 270-279.

McGill, B. J. (2010). Towards a unification of unified theories of biodiversity. Ecol. Lett. 13, 627-642.

McGill, B. J. (2011a). Linking biodiversity patterns by autocorrelated random sampling. Am. J. Bot., 98, 481-502.

McGill, B. J. (2011b). Species abundance distributions. In. A. Magurran and B. McGill (eds) Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford (201:1) 105-122.

McGlinn, D.J. and Palmer, M.W., (2009). Modeling the sampling effect in the species–time–area relationship. Ecology, 90, 836-846.

McGlinn, D. Xiao, X., May, F. Gotteli, N. J., Blowes, S. A., Knight, T. M. et al. Submitted. MoB (Measurement of Biodiversity): a method to separate the scale-dependent effects of species abundance distribution, density, and aggregation on diversity change. Submitted to Meth. Ecol. Evol.

Mendes, R.S., Evangelista, L.R., Thomaz, S.M., Agostinho, A.A. and Gomes, L.C., (2008). A unified index to measure ecological diversity and species rarity. Ecography, 31, 450-456.

Morlon, H., Schwik, D. W. Bryant, L. A., Marquet, P. A., Rebelo, A. G. et al. (2011). Spatial patterns of phylogenetic diversity. Ecol. Lett. 14, 141-149.

Nekola, J.C. and McGill, B.J., (2014). Scale dependency in the functional form of the distance decay relationship. Ecography, 37, 309-320.

Newbold, T., Boakes, E.H., Hill, S.L., Harfoot, M.B. and Collen, B., (2017). The present and future effects of land use on ecological assemblages in tropical grasslands and savannas in Africa. Oikos, 126, 1760–1769.

Olszewski, T.D., (2004). A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. Oikos, 104, 377-387.

Palmer, M.W. and White, P.S., (1994). Scale dependence and the species-area relationship. Am. Nat., 144, 717-740.

Powell, K.I., Chase, J.M. and Knight, T.M., (2011). A synthesis of plant invasion effects on biodiversity across spatial scales. Am. J. Bot., 98, 539-548.

Powell, K.I., Chase, J.M. and Knight, T.M., (2013a). Invasive plants have scale-dependent effects on diversity by altering species-area relationships. Science, 339(6117), pp.316-318.

Powell, K.I., Chase, J.M. and Knight, T.M., (2013b) Data from: Invasive plants have scale-dependent effects on diversity by altering species-area relationships. Dryad Digital Repository. https://doi.org/10.5061/dryad.qq08m

Preston, F.W., (1960). Time and space and the variation of species. Ecology, 41, 611-627.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reese, G. C., Wilson, K. R. and Flather, C. H. (2014), Performance of species richness estimators across assemblage types and survey parameters. Global Ecology and Biogeography, 23: 585–594.

Rényi, A., (1961). On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California.

Sandel, B. and Smith, A.B., (2009). Scale as a lurking factor: incorporating scale-dependence in experimental ecology. Oikos, 118, 1284-1291.

Scheiner, S. M., Cox, S.B., Mittelbach, G.G., Osenberg, C. and Kaspari, M., (2000). Species richness, species–area curves and Simpson's paradox. Evol. Ecol. Res. 2, 791-802.

Simberloff, D., (1972). Properties of the rarefaction diversity measurement. Am. Nat. 106, 414-418.

Smith, A.B., Sandel, B., Kraft, N.J. and Carey, S., (2013). Characterizing scale-dependent community assembly using the functional-diversity-area relationship. Ecology, 94, 2392-2402.

Srivastava, D.S. and Lawton, J.H., (1998). Why more productive sites have more species: an experimental test of theory using tree-hole communities. Am. Nat., 152, 510-529.

Thompson, G.G. and Withers, P.C., (2003). Effect of species richness and relative abundance on the shape of the species accumulation curve. Aust. Ecol., 28, 355-360.

Tsallis, C.. (1988). Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys., 52, 479–487.

Tuomisto, H., (2010a). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography, 33, 2-22.

Tuomisto, H., (2010b). A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. Ecography, 33, 23-45.

Tuomisto, H., Ruokolainen, K., Vormisto, J., Duque, A., Sánchez, M., Paredes, V.V. et al., (2017). Effect of sampling grain on patterns of species richness and turnover in Amazonian forests. Ecography, 40, 840-852.

Whittaker, R.H., (1960). Vegetation of the Siskiyou mountains, Oregon and California. Ecol. Monog., 30, 279-338.

Whittaker, R.J., (2010). Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness-productivity relationship. Ecology, 91, 2522-2533.

Zhang, Y., Ma, K., Anand, M., Ye, W. and Fu, B., (2015). Scale dependence of the beta diversity-scale relationship. Comm. Ecol., 16, 39-47.

1 **Table 1.** Definitions and interpretations of the biodiversity metrics for scale-explicit analyses.

Metric	Definition	Interpretation
N	Total numbers of individuals	Measure of density of individuals. Because N scales roughly linearly with area (i.e. density is scale-independent) we do not need to measure at multiple scales.
${}^{\alpha}S$, ${}^{\gamma}S$	Observed richness of species from α-scale (average of observations) and from γ-scale (sum of observations) local scale observations.	Number of species at local scale (=α-diversity) and large scale (=γ-diversity)
${}^{\alpha}S_n$, ${}^{\gamma}S_n$	The expected richness for n randomly sampled individuals (Hurlbert 1971). Can be calculated from α- or γ-scale.	Estimate of richness at α- or γ-scale after controlling for differences due to aggregation and number of individuals (i.e., only reflects SAD)
${}^{\alpha}S_{\text{asymptote}}$, ${}^{\gamma}S_{\text{asymptote}}$	Extrapolated richness at α- or γ-scale (most typically estimated via Chao1 estimator; Chao 1984).	Richness at α- or γ-scale that includes unknown species but is highly correlated with S (McGill, 2011b)

${}^{\alpha}f_0$, ${}^{\gamma}f_0$	Richness of undetected species ($S_{\text{asymptote}} - S$, Chao et al. 2009). Can be measured at α- or γ-scale.	Measure of rarity, slope at top of rarefaction curve, more sensitive to rare species than S
${}^{\alpha}$PIE , ${}^{\gamma}$PIE	Probability of intraspecific encounter ($S_{n=2} - S_{n=1}$, Hurlbert 1971, Olszweski 2004)	Measure of evenness, slope at base of the rarefaction curve, and sensitive to common species
${}^{\alpha}S_{\text{PIE}}$, ${}^{\gamma}S_{\text{PIE}}$	Equally abundant species needed to yield PIE (Jost 2006) (*=1/(1-PIE))*	Effective number of species of PIE (=1-Simpson's) that is easier to compare with S
β_S	Ratio of total treatment ${}^{\gamma}S$ and average plot ${}^{\alpha}S$ (Whittaker, 1960)	More species turnover results in larger β_S because of increases in spatial aggregation, N, and/or unevenness of the SAD.
$\beta_{S_{\text{PIE}}}$	Ratio of total treatment ${}^{\gamma}S_{\text{PIE}}$ and ${}^{\alpha}S_{\text{PIE}}$ (Olszewski 2004, Jost 2007)	Like β_S but emphasizes common species. Higher $\beta_{S_{\text{PIE}}}$ due to aggregation.

1

Figure Legends:

Figure 1. (A) Effects of a dominant invasive plant (Dianella ensifolia) on the species richness of plants in the understory of Hardwood Hammock forests in central Florida, USA (data from Powell et al. 2013b). Data show species richness for three pairs of invaded and uninvaded sites at two nested spatial grains ($1m^2$ and $500m^2$). (B) Effects of N addition (6 treatments) to a Mongolian grassland on species richness measured at two spatial grains ($1m^2$ and $25m^2$) (data from Lan et al. 2015b).

Figure 2. (A) Hypothetical comparison of effect sizes (i.e., log response ratio) measured at larger and smaller spatial scales. (B) Results of a meta-analysis of scale-dependent responses to a number of different ecological factors. Points represent the log response ratio of an ecological factor in a given comparison measured at the smallest (x-value) and largest (y-value) scale. We defined the reference treatment as the one with the highest level of richness at the small scale and took the absolute value of the log-ratio effect size, so that the distribution of effect sizes was limited mostly to the positive direction. The solid line indicates the 1:1 line expected if effect sizes were not scale-dependent. Points above and below this line indicate effect sizes that are larger or smaller, respectively, as scale increases; points that are negative on the y-axis indicate those where the direction of the effect shifted from smaller to larger scales. Because of the way the reference treatment was defined, it is not possible to get negative values at the small scale. Overall, effect sizes at the smallest scale was correlated with the effect size observed at the larger scale ($P<0.05$), however, this relationship had poor predictive power ($R^2=0.14$).

Figure 3. A hypothetical depiction of three different communities (labelled A, B, C) each with a different Species Accumulation Curve (SAC). Each community can be sampled at different

scales (area or number of individuals), labelled Scale 1-3, showing how sampling scale can lead to different rankings of community in terms of species richness. Curves can diverge, leading to increasing in differences with scale (compare communities A vs C), they can converge, leading to decreases in differences with scale (communities B vs. C), or they can intersect and cross, where species richness is greater for one community at one scale (B>A at scale 1), but greater for another community at another scale (A>B at scales 2 & 3).

Figure 4. (A) Sample-based accumulation curve where average number of individuals in a sample is defined as the α-scale (first vertical line), and the total number of individuals across samples is defined as the γ-scale (second vertical line). From this, we can derive local richness ($^{\alpha}S$), regional richness ($^{\gamma}S$), and β-richness ($\beta_s=^{\gamma}S/^{\alpha}S$). (B) individual-based rarefaction curve from α-scale samples (dashed line) and γ-scale samples (solid line). From this, we can visualize S, PIE for each scale ($^{\alpha}PIE$, $^{\gamma}PIE$), and the difference between them (β_{PIE}). $^{\alpha}S_{PIE}$, $^{\gamma}S_{PIE}$ and $\beta_{S_{PIE}}$ are not readily shown on the figure, but can be calculated from PIE by the effective numbers of species (see Jost 2006). Other metrics presented in Table 1 are not shown for clarity.

Figure 5. Results from a simple spatial simulation model in which diversity parameters were varied. The description on the left for each scenario in A-F indicates how one community, the 'reference' (open circles for accumulation; dashed lines for rarefaction)) was varied relative to the one in which biodiversity change was imposed (closed circles; solid lines for rarefaction). The left column show species accumulation curves (mean $\pm$ 1 standard deviation) and the right column shows rarefaction curves and 95% confidence intervals. The final row—in which

aggregation was manipulated—is illustrated with two scales for the individual-based rarefaction from the 'change' scenario. See text for detail.

Figure 6. Bivariate relationships between N, S_{PIE} and S for 346 communities across the 37 datasets. (A) S_{PIE} as a function of N; (B) S as a function of N; (C) S as a function of S_{PIE}. Black lines depict the relationships across studies (and correspond to R^2 fixed); colored points and lines show the relationships within studies. All axes are log-scale.

Figure 7. Representative rarefaction curves, the proportion of curves that crossed, and counts of how often curves crossed. (A) Rarefaction curves for different local communities within two datasets: marine invertebrates (nematodes) along a gradient from a waste plant outlet (Lambshead 1986), and trees in a Ugandan rainforest (Eggeling 1947); axes are log-transformed. (B) Counts of how many times pairs of rarefaction curves (from the same community) crossed; y-axis is on a log-scale.

Figure 8. Effect of nutrient additions on several measurements of biodiversity from Table 1. Each biodiversity measure was calculated at the α-scale (1 mesocosm) (Panels A,B), γ-scale (15 mesocosms) (Panels C,D), as well as the β-scale (i.e. turnover across scales, γ/α)(Panels E,F). See Chase (2010) for details on the experimental design and the mobr package (McGlinn et al. submitted, https://github.com/MoBiodiv/mobr) for details on the statistical methods

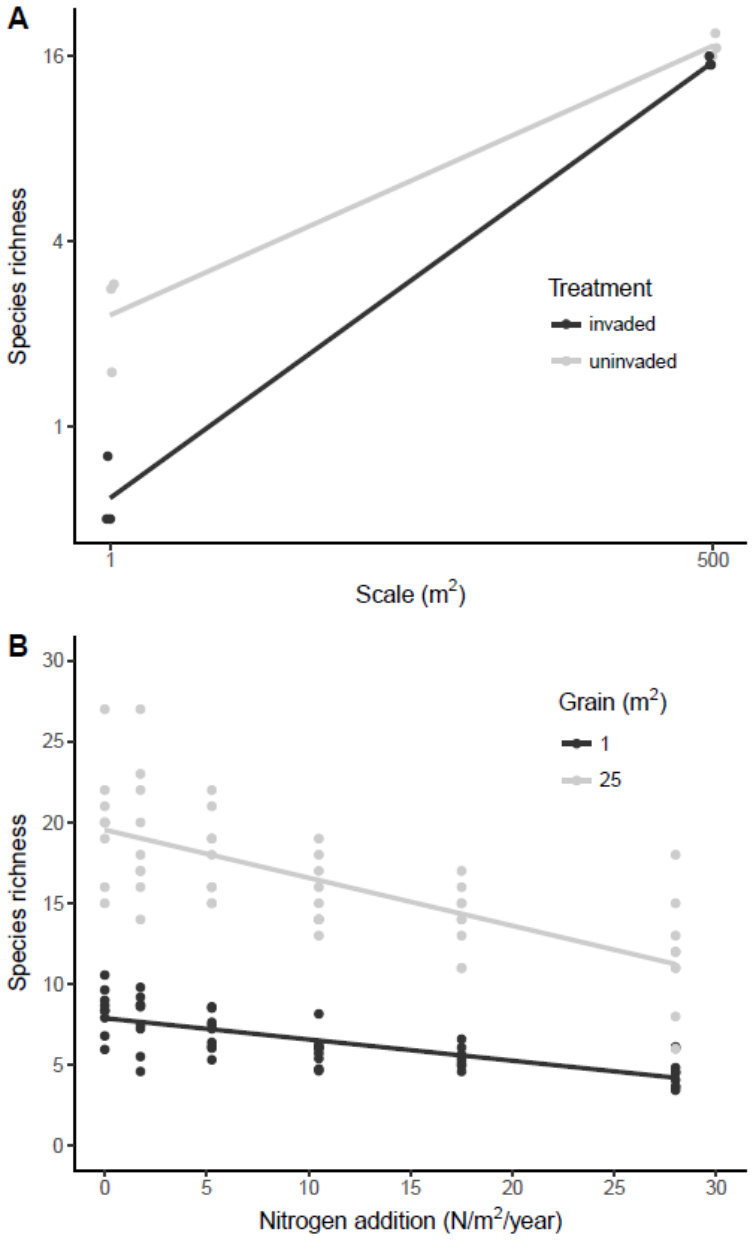
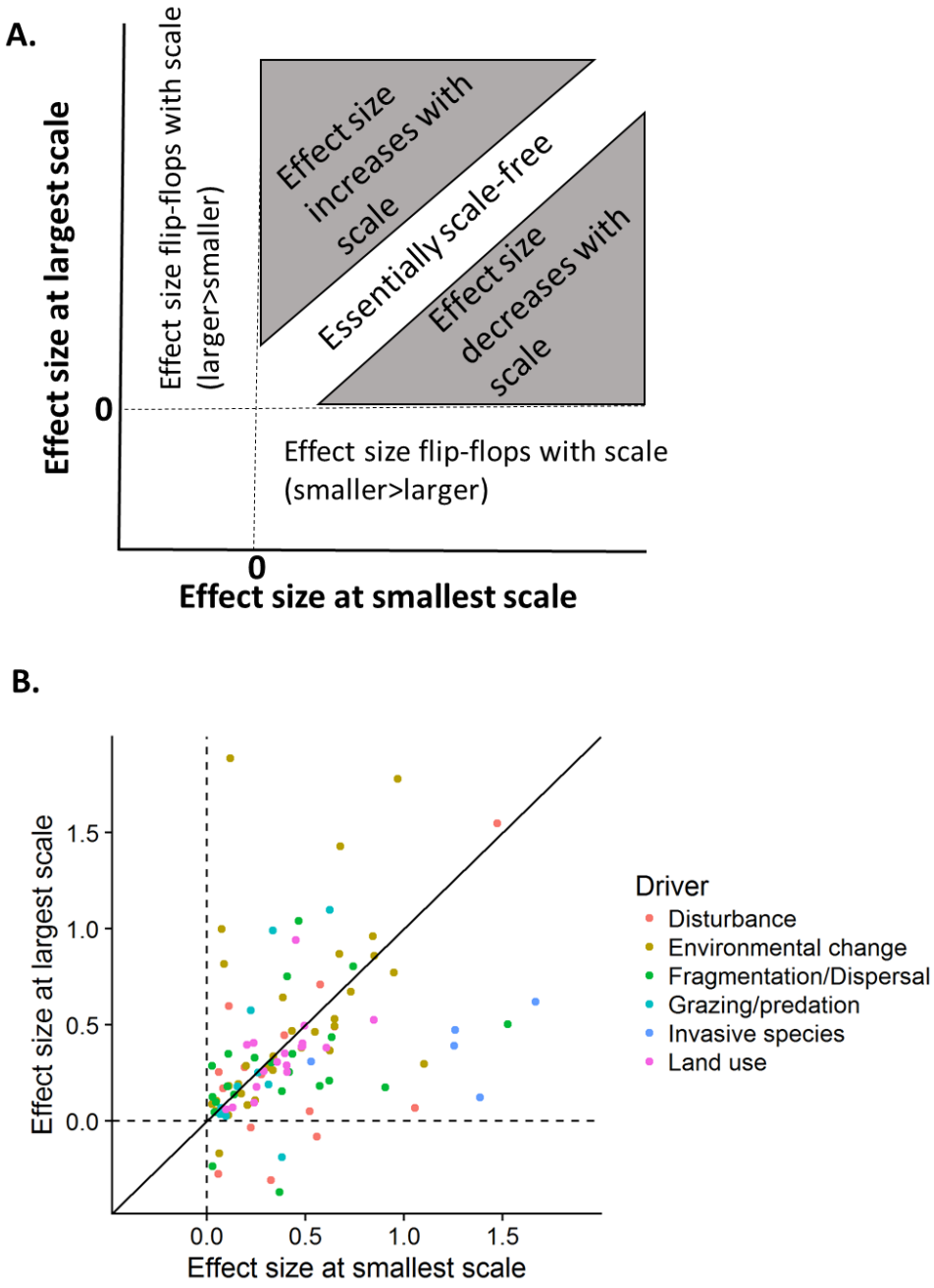


Figure 1.



1

2 **Figure 2.**

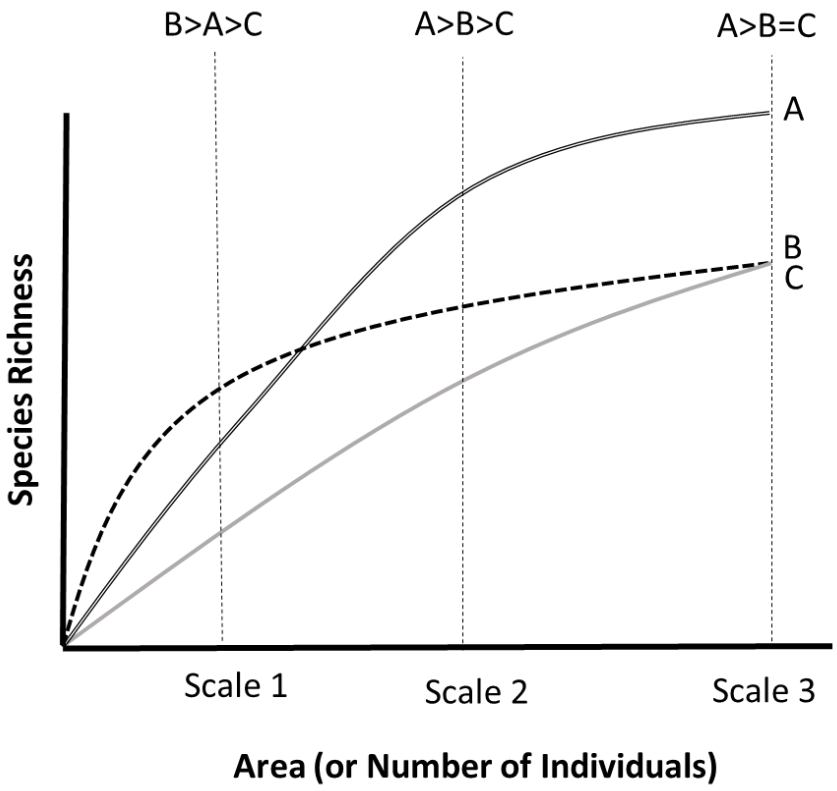


Figure 3.

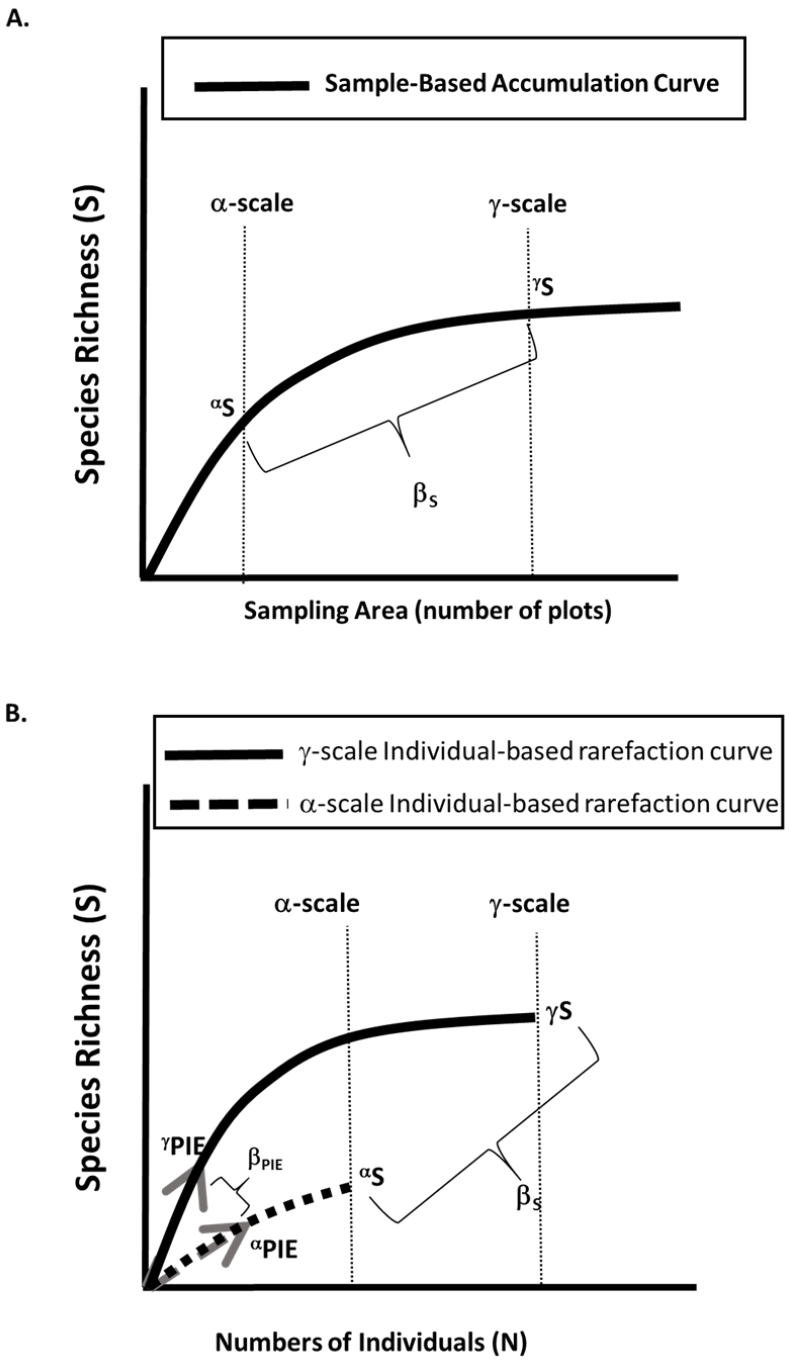


Figure 4.

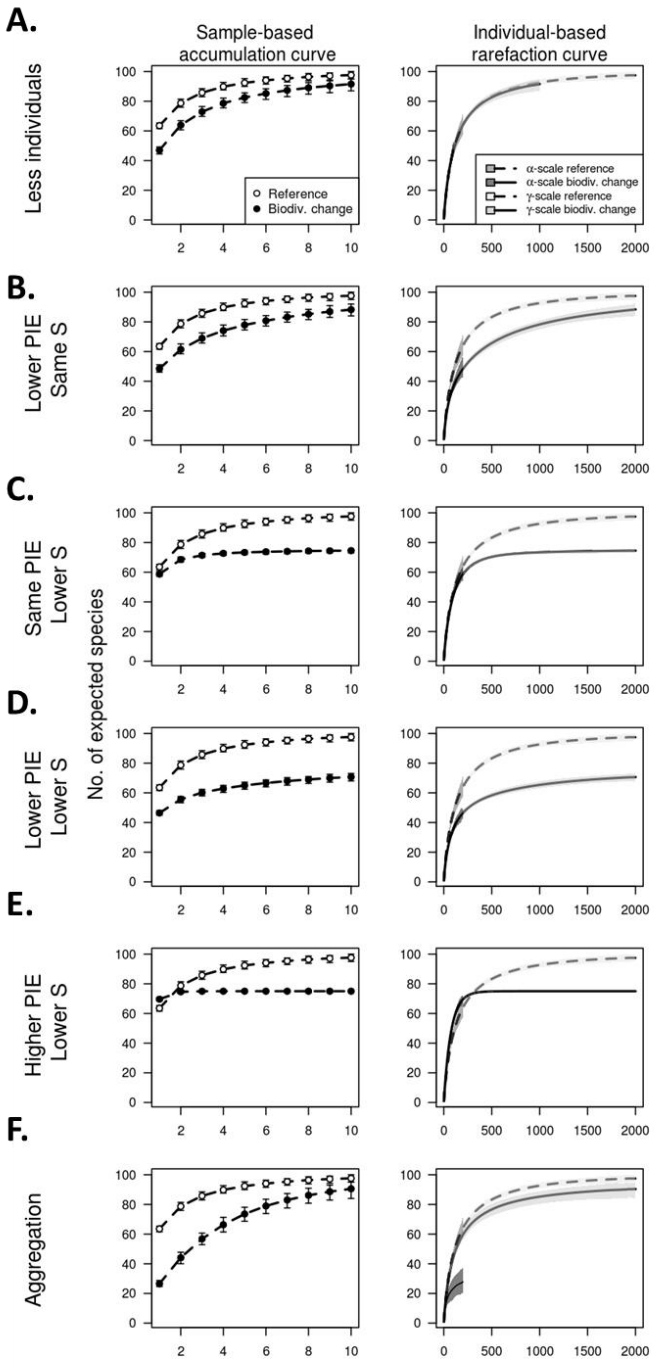


Figure 5.

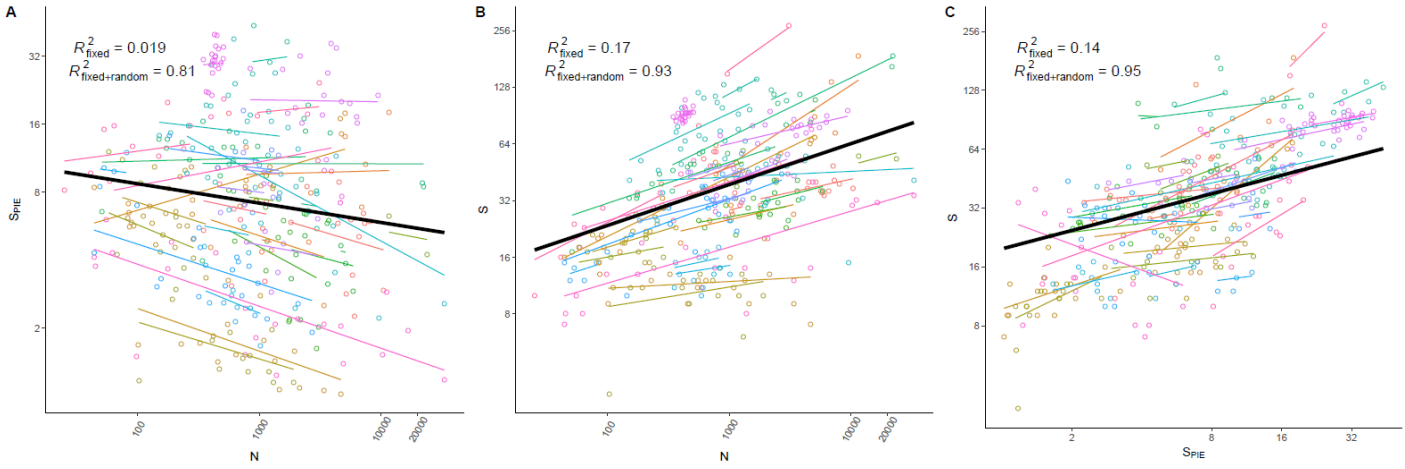


Figure 6.

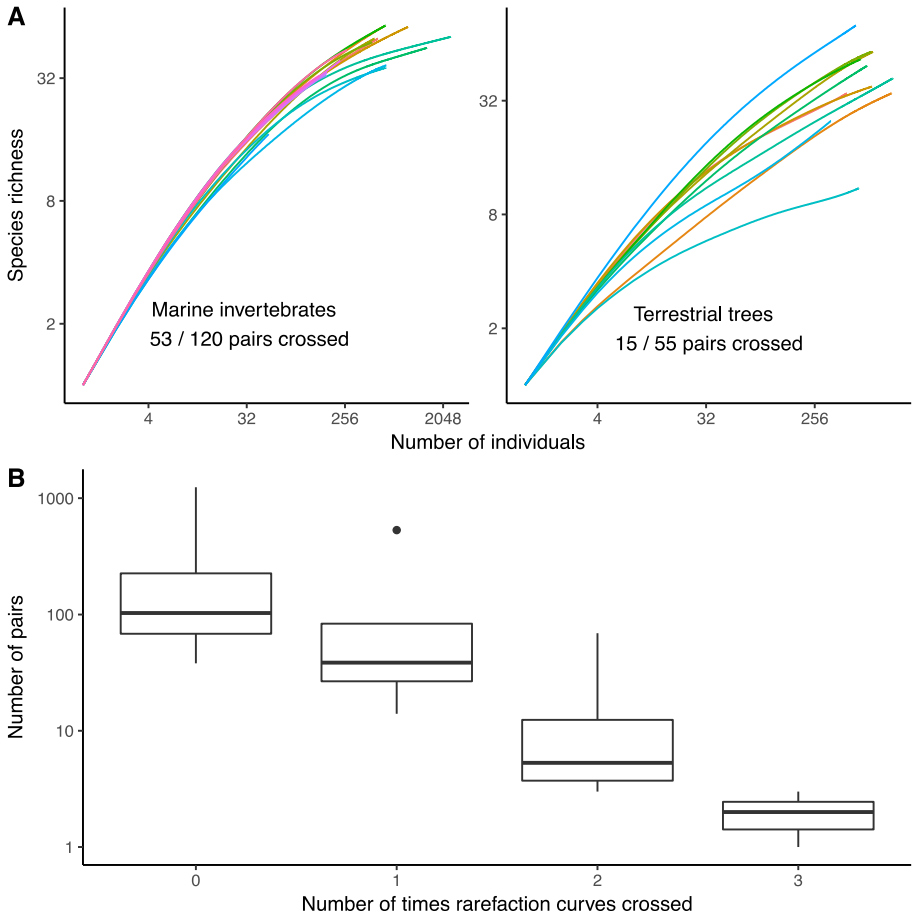


Figure 7.

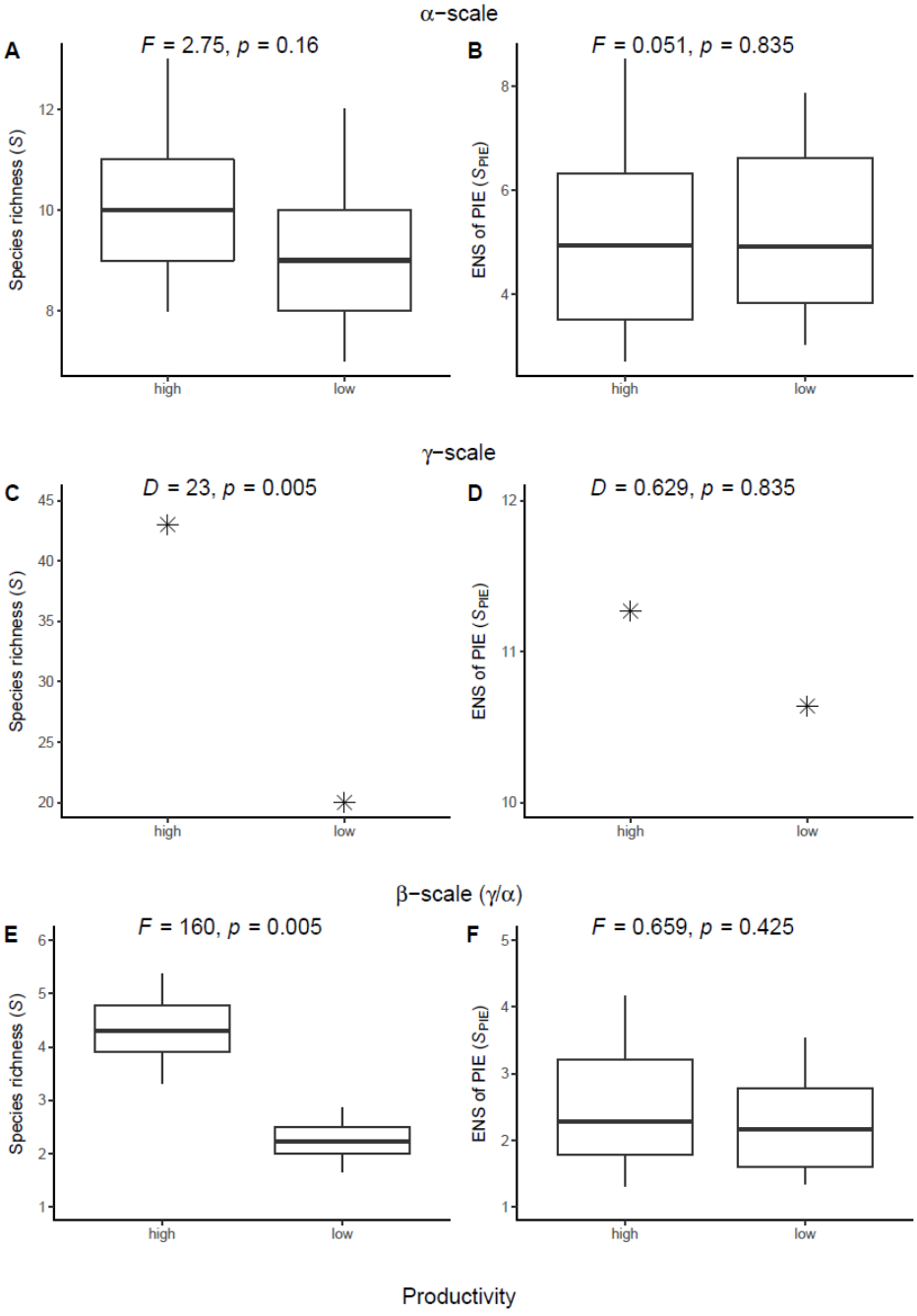


Figure 8.

Supplementary Information

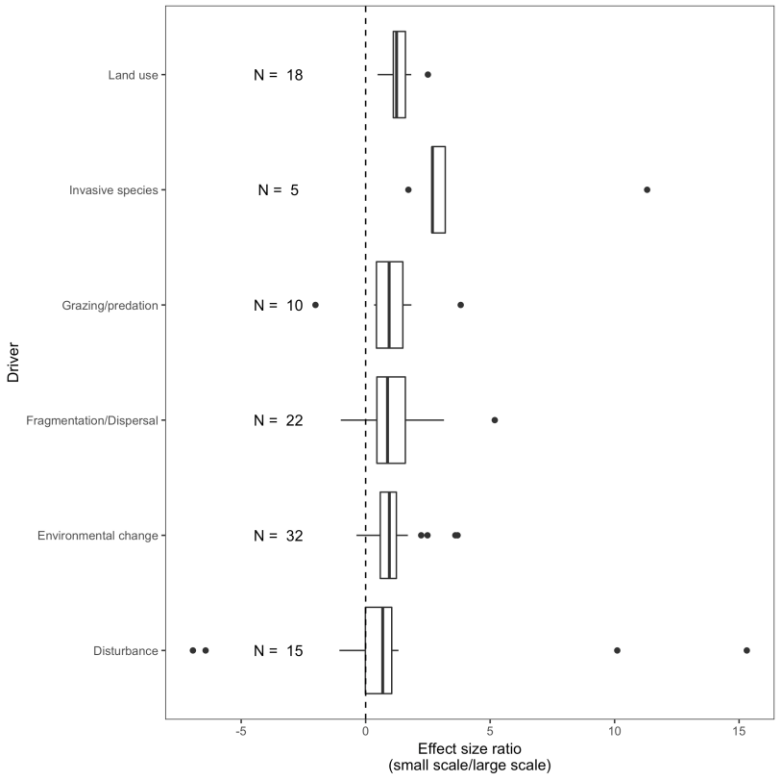


Figure S1. Results showing the ratio of log-response ratio effect sizes from experiments where species richness responses were measured at two spatial scales (small scale/large scale). The dashed line at 0 would indicate studies where the effect sizes were the same at the smaller and larger scale. For each category of ecological driver, the means are above 1, indicating that the measured effect size is larger at the smaller relative to larger size, and this difference is statistically significant for land use, invasive species, and grazing/predation.

Appendix 1. Effect sizes, metadata, and references for studies used in the scale-dependent meta-analysis presented in Figure 2.

File available here:

https://www.dropbox.com/sh/s3vmrwya1b7khz2/AAAnWDucBJYdxPdT_tOk2olOa?dl=0

(will be deposited in appropriate repository on acceptance)

Appendix 2. Raw data and references for rarefaction curves and analyses presented in Figures 6 and 7. Subset of studies presented in McGill 2011b.

File available here:

https://www.dropbox.com/sh/s3vmrwya1b7khz2/AAAnWDucBJYdxPdT_tOk2olOa?dl=0

(will be deposited in appropriate repository on acceptance)

Appendix 3. Raw data for the analyses of invertebrate and amphibian communities in nutrient addition and reference treatments presented in Figure 8. From Chase (2010).

File available here:

https://www.dropbox.com/sh/s3vmrwya1b7khz2/AAAnWDucBJYdxPdT_tOk2olOa?dl=0

(will be deposited in appropriate repository on acceptance)