1 **Embracing scale-dependence to achieve a deeper understanding of**

2 **biodiversity and its change across communities**

3

4 *Authors and affiliations:*

5 Jonathan M. Chase*; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-

6 Leipzig and Department of Computer Science, Martin Luther University, Halle-Wittenberg;

7 jonathan.chase@idiv.de

8

9 Brian J. McGill*; School of Biology and Ecology & Mitchell Center for Sustainability Solutions,

10 University of Maine; mail@brianmcgill.org

11

12 Daniel J. McGlinn; Biology Department, College of Charleston, Charleston, SC;

13 mcglinndj@cofc.edu

14

15 Felix May; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig and

16 Leuphana University Lüneburg, Universitätsallee 1, D-21335 Lüneburg, Germany;

17 felix.may@leuphana.de

18

19 Shane A. Blowes; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-

20 Leipzig; shane.blowes@idiv.de

21

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

23

1   Tiffany M. Knight; Dept. Community Ecology, Helmholtz Centre for Environmental Research –

2   UFZ; Institute of Biology, Martin Luther University Halle-Wittenberg; German Centre for

3   Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig. tiffany.knight@idiv.de

4

5   Oliver Purschke; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-

6   Leipzig; oliver.purschke@idiv.de

7

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

9

10 *Joint first authors

11

12 Statement of Authorship: JC and BM conceived the study and the overall approach, and all

13 authors participated in multiple working group meetings to develop and refine the approach. BM

14 collected the data for the meta-analysis that led to Fig. 2,3; JC collected the data for the meta-

15 analysis that led to Figure 4 and S1; SB and FM did the analyses for Figures 2-4; DM, FM and

16 XX wrote the code for the analysis used for the recipe and case study in Figure 6. JC, BM and

17 NG wrote first drafts of most sections, and all authors contributed substantially to revisions.

18

19 Data accessibility statement: All data for meta-analyses and case study will be deposited in a

20 publically available repository with DOI upon acceptance (available in link for submission).

21

22

23 *Running Title:* Scale-dependence in biodiversity comparisons

1

2   *Key Words:* Species-area relationship; species richness; evenness; Hill number; Simpson's index;

3   rarefaction; scale-dependence

4

5   *Type of Article:* Reviews and Syntheses

6

7   *Number of references*: 78

8   *Number of words in abstract*: 184

9   *Number of words in main text*: 7432

10   *Number of Tables:* 1

11   *Number of Figures:* 6

12

13

14

15

16   *Mailing address of corresponding author:*

17   Jonathan Chase

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

19   Deutscher Platz 5e

20   04103 Leipzig

21   Germany

1   **Abstract:**

2   Because biodiversity is multidimensional and scale-dependent, it is challenging to estimate its

3   change. However, it is unclear (1) how much scale-dependence matters for empirical studies,

4   and (2) if it does matter, how exactly we should quantify biodiversity change. To address the

5   first question, we analyzed studies with comparisons among multiple assemblages, and found

6   that rarefaction curves frequently crossed, implying reversals in the ranking of species richness

7   across spatial scales. Moreover, the most frequently measured aspect of diversity—species

8   richness—was poorly correlated with other measures of diversity. Second, we collated studies

9   that included spatial scale in their estimates of biodiversity change in response to ecological

10   drivers and found frequent and strong scale-dependence, including nearly 10% of studies which

11   showed that biodiversity changes switched directions across scales. Having established the

12   complexity of empirical biodiversity comparisons, we describe a synthesis of methods based on

13   rarefaction curves that allow more explicit analyses of spatial and sampling effects on

14   biodiversity comparisons. We use a case study of nutrient additions in experimental ponds to

15   illustrate how this multi-dimensional and multi-scale perspective informs the responses of

16   biodiversity to ecological drivers.

17

1   **Introduction**

2   Biodiversity is multidimensional and scale-dependent. It is multi-dimensional, because it is

3   determined by a combination of the total numbers of species in a given area and the total and

4   relative abundances of those species in that area. And it is scale-dependent, because all measures

5   of diversity necessarily increase non-linearly with increasing sampling effort and spatial scale.

6   However, the nature of this scaling depends on the total numbers of species, the total and relative

7   abundances of individuals, as well as their spatial distribution. This makes comparisons of

8   biodiversity from place to place (and time to time) more challenging than comparisons of most

9   other variables in ecology (e.g., McGill 2011a).

10

11 Understanding how biodiversity changes in the face of natural and anthropogenic drivers can be

12 greatly enhanced by taking a scale-explicit and multi-dimensional perspective on biodiversity

13 and its change. Of course, ecologists have long recognized the complexity in making

14 comparisons of biodiversity in response to natural and anthropogenic drivers and continuously

15 sought better solutions to this problem (e.g., Preston 1960, MacArthur and MacArthur 1961,

16 Hurlbert 1971, Hill 1973, Lande 1996, Gotelli and Colwell 2001, Jost 2006, Magurran and

17 McGill 2011, Chao and Jost 2012, Chase and Knight 2013). Nevertheless, there is little

18 consensus or consistency as to which (if any) solutions are best. And as a likely result, the vast

19 majority of empirical studies ignore these complexities and typically measure and compare only

20 a single variable (usually species richness), measured at a single spatial scale (usually determined

21 by the grain size and extent of the sampling).

22

5

1   To illuminate the neglect of spatial scale and multi-dimensionality in biodiversity studies, we

2   evaluated papers published from 2012-2017 in Ecology Letters that compared (taxonomic)

3   diversity (but not compositional). Of these, 80% (44/55) only reported one measure of diversity

4   (90% of those were species richness), and 75% (41/55) compared diversity at only one spatial

5   scale; just three exemplary studies (Solar et al. 2015, Martinson et al. 2017, Livne-Luzon et al.

6   2017), measured multiple dimensions of diversity at multiple scales, which we advocate to more

7   fully understand biodiversity change.

8

9   Simply estimating a difference (i.e., effect size) of any measure of diversity at a single scale and

10   then subjecting that difference to standard statistical analyses does not capture the fact that the

11   magnitude, and even direction, of biodiversity change critically depends on which measure, and

12   at which spatial scale, the data were collected (e.g., Whittaker et al. 2001, Chase and Leibold

13   2002, Rahbek 2005, Dumbrell et al. 2008, Sandel and Smith 2009, Keil et al. 2011). This is not

14   just a theoretical issue, but is a concrete problem that can affect biodiversity conservation and

15   management. For example, Hill and Hamer (2004) reviewed studies of the effects of logging on

16   tropical forest bird diversity and found that bird diversity tended to decrease with logging when

17   the scale in which diversity was measured was relatively small (i.e., less than 25 ha), but increase

18   with logging in studies in which diversity was measured at larger spatial scales. In an analysis

19   and review of the effects of logging on butterfly diversity, Hamer and Hill (2000) likewise

20   illustrated scale-dependent effects of logging, but in a different direction (i.e., minor effects of

21   logging on diversity at small scales, but larger negative effects at larger scales).

22

1   Does it matter that biodiversity is multidimensional, where change in an estimate of species

2   richness at a given scale might differentially reflect changes in the total and relative abundances

3   of individuals in a community? Many have argued that different measures of biodiversity (e.g.,

4   evenness, richness) are highly correlated and thus perhaps redundant. If so, one measure of

5   diversity and its change would likely suffice. Likewise, does it matter that these biodiversity

6   changes can be scale-dependent? While there are certainly many case studies where scale-

7   dependence has been observed, is this common? How often do measured effect sizes vary in

8   magnitude and/or direction such that conclusions about biodiversity change would be altered?

9

10   Here, we first review the multidimensional concept of biodiversity and its scaling to establish

11   general hypotheses about the scale-dependent nature of biodiversity measures and their

12   comparisons. We collated datasets where species assemblages from multiple sites were

13   compared within a study and asked whether measures of biodiversity were correlated. Next, we

14   collated studies in which biodiversity changes were measured at more than one spatial scale, so

15   that we could estimate how frequent scale-dependence occurred in the magnitude and direction

16   of effect sizes. Finally, after showing that biodiversity changes are indeed typically multivariate

17   and scale-dependent, we describe a recipe that synthesizes a number of measures of biodiversity

18   to more fully evaluate how biodiversity changes, and we illustrate our approach with a case study

19   from a previously published experiment on aquatic mesocosms.

20

21   **The multidimensional nature of biodiversity and its scaling**

22   Biodiversity is a summary variable that is influenced by the abundances and distributions of

23   populations of multiple species. Three factors in particular combine to influence diversity

1   measures at any given sample or scale (e.g., He and Legendre 2002, McGlinn and Palmer 2009,

2   McGill 2011a, Chase and Knight 2013): (1) the total numbers of individuals (N) in a community;

3   (2) the Species Abundance Distribution (SAD), which includes both the total numbers of species

4   in the community (S) and their relative abundances; (3) the spatial distributions of individuals,

5   which influences patterns of intraspecific aggregation or clumping. Each of these components

6   can be independently and interactively influenced by ecological drivers, leading to

7   multidimensional and scale-dependent biodiversity change. We recognize that a critical

8   component of biodiversity change is in compositional shifts, which uses a different set of metrics

9   to detect changes (e.g., Baselga 2010, Dornelas et al 2014, Legendre 2014, Hillebrand et al.

10   2018).   But here we only focus on the quantifications of biodiversity within and across

11   communities, which presents more than enough challenge in itself.

12

13   When comparing communities, any differences in the N, the SAD and/or aggregation will

14   influence how measurements and comparisons of biodiversity change with sampling scale or

15   effort. Figure 1 illustrates one way to depict this scaling—via individual-based rarefaction

16   curves— and shows two qualitatively distinct ways in which it can differ. Differences in

17   individual-based rarefaction curves are determined only by differences in the SAD (both the total

18   S and their relative abundance) (e.g., Gotelli and Colwell 2001, Cayuela et al. 2015)

19

20   In Figure 1A, three communities (labelled A, B, and C) differ in a straightforward way; both

21   species richness and the relative abundances of the three communities are ranked A>B>C. Thus,

22   if we were to compare species richness or other commonly measured estimates of diversity (e.g.,

23   Shannon, Simpson), they would rank the same. In such a case, it would not be necessary to

8

1   worry much about the multidimensionality of diversity estimates (i.e., because different

2   dimensions are correlated). However, even when the rankings of S and other diversity measures

3   are consistent, there can still be scale-dependence when the SAD changes among communities

4   (e.g., Gotelli and Colwell 2001, Cao et al. 2007, Chase and Knight 2013, Cayuela et al. 2015).

5   Nevertheless, this scale-dependence is moderate and would be predictable, consistently leading

6   to larger differences between communities as scale increases (i.e., rarefaction curves diverge).

7

8   In Figure 1B, three communities (labelled A, B, and C) differ in a more complex way.

9   Community A is dominated by a few very common species and has several rarer species (i.e., is

10   less even), and thus has a shallower slope near its base than Community B, but higher total S.

11   That is, the difference between the communities flip-flop; at small scales, B is richer than A, but

12   at large scales, A is richer than B. The comparison between Communities B and C is somewhat

13   similar; they have the same total S, but different levels of evenness (and diversity), again

14   decoupling the different measures of diversity. Finally, Communities A and C differ in both S

15   and diversity measures in the same direction.

16

17   These represent two extreme scenarios for diversity comparisons between communities. In

18   Figure 1A, diversity measures are largely correlated and scale-dependence is unidirectional,

19   suggesting that we may not need to worry much about which measures of diversity we use for

20   comparisons, and at which scales. Alternatively, in Figure 1B, diversity measures are decoupled

21   and scale dependence is strong, suggesting that more complex measurements and comparisons of

22   biodiversity will be necessary. Furthermore, the curves in Figure 1 only include differences in

23   the SAD between communities, whereas variation also in N and aggregation will lead to even

1   more complex SACs, which could further decouple diversity measures and create more, and

2   more complex, forms of scale-dependence (see also Chase and Knight 2013, Blowes et al. 2017,

3   McGlinn et al. 2018).

4

5   Next, we empirically evaluate these possibilities, with the extremes being: (i) biodiversity change

6   between communities is relatively "well behaved", where different measures of biodiversity are

7   largely correlated and scale-dependence is moderate and unidirectional (Figure 1A), or (ii)

8   biodiversity change is more complicated where different measures are poorly correlated and

9   scale-dependence is strong and multi-directional (Figure 1B). As discussed above, few

10   exemplary data exist where multiple measures and multiple scales of biodiversity change are

11   collected and compared. And so we used two separate meta-analyses described in the following

12   sections, one to compare correlations (or lack thereof) among diversity measures, and another

13   containing data measured at multiple scales to evaluate the degree and direction of scale-

14   dependence in biodiversity responses to ecological drivers.

15

16   **Empirical evidence that biodiversity change is often multidimensional**

17   To determine which of the caricatures in Figure 1 is more indicative of biodiversity change

18   among communities, we directly compared SADs and the resulting rarefaction curves from 37

19   datasets in which two or more local communities were compared (data and references listed in

20   Appendix 2, taken from a larger assemblage of datasets amassed by McGill 2011b). We used

21   these studies to evaluate (i) how often rarefaction curves cross, suggesting that sampling scale

22   can critically influence conclusions about the magnitude and direction of biodiversity change,

23   and (ii) how measures of diversity (S, S_{PIE}) and numbers of individuals (N) co-vary among

10

1 communities. Because there was no spatial information available in these datasets, we were only

2 able to compare rarefaction curves, eliminating potential influences of differences in spatial

3 aggregation, which could have further complicated their relationships.

4

5 Example datasets include bird surveys along standardized routes in North America (i.e., the

6 Breeding Bird Survey; Pardieck et al. 2017), benthic marine nematodes collected from cores in

7 relation to pollution (Lambshead 1986), arthropods associated with different macrophyte species

8 in streams (Harrod 1964), and trees from different sections of the Barro Colorado Island

9 permanent forest plot (Condit et al. 2012). Most datasets compared 2-20 communities; for

10 datasets with >20 communities (e.g., the North American breeding birds), we randomly selected

11 20 sites so that larger datasets did not numerically dominate the results.

12

13 First, we evaluated whether the rankings of rarefaction curves were relatively well-behaved, as in

14 Figure 1A, or whether curves were more likely to cross, as in Figure 1B (see also Lande et al.

15 2000, Thompson and Withers 2003). Such crossings indicate not just a change in magnitude, but

16 also a change in direction of biodiversity change due to sampling and spatial scale. We

17 compared curves for each pair of communities within a single dataset (see Figure 2A for two

18 examples). Across all datasets, a total of 2203 pairwise comparisons were made, with 732 pairs

19 of rarefaction curves crossing (33% of all pairs, with some curves crossing multiple times)

20 (Figure 2B). Importantly, the crossing points were not ecologically trivial (e.g., crossings only at

21 very small N). The average crossing occurred at 24% of the total abundance in a community and

22 the average vertical separation on curves that crossed was 6.3 species (versus an average of ~40

23 species in a community).

1

2   Next, we examined the relationships between three measures often collected and compared

3   among communities. Specifically, we compared the total N measured in a community, the

4   number of species (S) from that community, and a measure of diversity that takes relative

5   abundances into account. For the latter, we used a bias-corrected version of Hurlbert's (1971)

6   Probability of Interspecific Encounter; $PIE = \left(\frac{N}{N-1}\right) * \left(1 - \sum_{i=1}^{S} p_i^{\,2}\right)$; where N is the total

7   number of individuals in the entire community, S is the total number of species in the

8   community, and p_i is the proportion of each species i. We used PIE because it is a relatively

9   unbiased estimator of evenness, it is intuitive and familiar to ecologists, and is equivalent to the

10   slope of the rarefaction curve at its base (between N=1 and N=2) (Olszewski 2004).

11   Specifically, a more even community (higher PIE) accumulates species more quickly with

12   increasing N than a less even community (lower PIE). PIE is equivalent to 1-Simpson's index

13   (also called the Gini-Simpson index) (Jost 2006), but we prefer the PIE terminology because it

14   has a strong intuitive interpretation (i.e., the probability of interspecific encounter), as well as

15   connection to the individual-based rarefaction curve. We converted PIE into an effective number

16   of species, S_{PIE}, so that comparisons with S are more intuitive (e.g., Hill 1973, Jost 2006, 2007);

17   S_{PIE} is equivalent to 1/Simpson's index. S_{PIE} is the number of equally abundant species it would

18   take to yield a given value of PIE (see McGlinn et al, 2018 for justification for why we take the

19   effective number of species from the bias corrected PIE). In a community in which all species

20   had identical abundances, S_{PIE} would be equal to S, but as the community becomes increasingly

21   dominated, S stays the same, but PIE decreases towards 0 and S_{PIE} decreases towards 1. Other

22   measures of diversity when converted to an effective number of species are also potentially

23   informative, such as Shannon's index (e.g., Jost 2006, 2007, Chao et al. 2014a), but here we

12

1   focus on S and S_{PIE} because S is most sensitive to changes in the abundance of rare species,

2   whereas PIE is more sensitive to changes in the abundance of common species.

3

4   We examined bivariate relationships between S, N and S_{PIE} using hierarchical linear models with

5   a fixed effect, and modeled variation among studies as a random effect on slopes and intercepts.

6   N and S were modelled assuming Poisson error and a log-link function, and S_{PIE} was log-

7   transformed and modelled assuming Gaussian error; N, S_{PIE} and S entered all models as log-

8   transformed covariates. All models were fit in R (R Development Core Team 2017) using the

9   brms package (Bürkner 2017). To compare variation across versus between datasets, we

10   quantified the marginal (fixed effect only) and conditional R^2 (fixed + random effects; Gelman et

11   al. 2017).

12

13   First, we regressed S as a function of N (Figure 3A). As expected from sampling, there was an

14   overall increasing relationship between N and S across all studies. However, the relationship had

15   low explanatory power ($R^2 = 0.12$). The variation explained when the relationships were allowed

16   to vary between studies increased dramatically (R^2 increases to 0.91). While most studies

17   showed a positive relationship between N and S, slopes and intercepts among studies varied

18   widely, and more than 30% (12/37) of studies showed that the 95% credible interval of the

19   study-level slope overlapped zero. Clearly, much of the relationship between N and S, as well as

20   its variation among studies, is due to sampling and methodological differences. However, the

21   fact that some types of communities within a study have large differences in N, with little

22   variation in S, whereas other communities within a study have smaller N differences, but large S

23   differences indicates that N and S changes are not completely coupled. For example, there are

1   many instances where N can dramatically increase or decrease from one place to another, but if

2   there is a concomitant change in the relative abundances of species, we might expect little

3   variation in S.

4

5   Second, we regressed S as a function of S_{PIE} to determine whether diversity measures were

6   consistent in their differences among communities (Figure 3B). Again, we found a significant,

7   but weak ($R^2 = 0.11$), relationship between these two measures. As in our analysis of S and N,

8   the fit of the model was much higher when we incorporated a random effect to capture variation

9   in the slope and intercept among studies ($R^2 = 0.85$). Here, most studies (20/37) showed that the

10   95% credible interval of the study-level slope overlapped zero, though some were positively

11   associated (16/37) and one showed negative associations.

12

13   In sum, our analyses above show that rarefaction curves often cross and that different measures

14   among communities are often decoupled (N, S, S_{PIE}). This supports the hypothesis that multiple

15   measures are often necessary to understand treatment effects on biodiversity (see also e.g.,

16   Stirling and Wilsey 2001, Wilsey et al. 2005, Soininen et al. 2012).

17

18   **Empirical evidence that biodiversity change is often scale-dependent**

19
20   Explicit consideration of spatial scale has provided some resolution to contradictory results in the

21   literature. For example, Chase and Leibold (2002) showed that the influence of productivity on

22   species richness was much stronger at larger spatial scales than at smaller ones (see also e.g.,

23   Whittaker et al. 2001, Gardezi and Gonzalez 2008, Andrew et al. 2012). Likewise, Powell et al.

14

1   (2011, 2013) showed that invasive species can simultaneously have strong impacts on diversity

2   at small scales, but much weaker effects at larger scales.

3

4   Although such multiple-scale analyses remain uncommon in the biodiversity literature, there are

5   enough to allow for a synthetic meta-analysis. We identified possible studies using an ISI Web

6   of Science search with the key words "species richness or diversity or biodiversity" AND "scale

7   or grain or extent" AND "ecology" (to eliminate hits that were in other fields). This yielded

8   ~8,500 papers, from which we could extract species richness at more than one scale in response

9   to ecological treatment from 103 comparisons within 52 studies (several studies reported

10 responses from more than one driver, taxonomic group, or study site). A list of the studies and

11 their data are given in the Supplemental material (Appendix 2).

12

13 With this collection of studies, we asked how frequent, and in which direction, the change in

14 biodiversity was influenced by spatial scale. We coded the log ratio effect size for any case

15 where species richness was lower in the treatment than the control as negative, and any case

16 where species richness was higher in the treatment than in the control as positive. Unfortunately,

17 we do not have estimates of uncertainty for most studies (i.e., most studies did not provide an

18 estimate of variance at the largest scale, and it was not possible to extract variance for many

19 smaller scale studies), and so cannot explore potential publication bias using traditional

20 approaches (e.g., funnel plots). We further note that there is likely some bias because our search

21 criteria explicitly included 'scale' (or 'grain', 'extent') as a search term. Thus, our meta-analysis

22 should be taken as a first indication of the frequency and direction of scale-dependent

23 biodiversity change.

1

2 The analysis shown in Figure 4 shows that scale-dependence in frequent, and a non-trivial

3 modifier of effect sizes in empirical studies. Most studies showed consistent positive (upper

4 right quadrant) or negative (lower left quadrant) effects of the treatment at both small and large

5 spatial scales. However, this obscures the reality that there are many instances where strong

6 effects at one scale were weak at the other. If effect sizes were scale-free then all points would

7 fall on the 1:1 line, but this line explains only 30% of the variance. Furthermore, nearly 12% of

8 comparisons (12/103) showed a reversal in direction of species richness change from negative to

9 positive (upper left quadrant), or from positive to negative (lower right quadrant) from smaller to

10 larger scales, several of which are quite substantial. While this is lower than the 33% of crossing

11 rarefaction curves (implying reversals) that we found in the analysis above (Figure 3), it again

12 implies that biodiversity change can not only vary in magnitude, but can often switch direction.

13 Overall, effect size at one scale is a poor predictor of effect size at a different scale.

14

15 Given the poor fit of the 1:1 line to predict effect sizes across scales, we can be reasonably

16 certain that scale-dependence is important. However, we can also use the information to

17 evaluate the nature of this scale-dependence in a deeper way. For example, the simple

18 alternative hypotheses we presented in Figure 1 illustrate different ways that scale-dependence

19 can occur. If biodiversity change is 'well-behaved' as in Figure 1A, we might expect that there

20 would be a consistent increase in the difference in species richness with increasing scale. This

21 would mean that points should be more likely to fall below the 1:1 line in the negative-negative

22 (lower left) quadrat (larger reductions in richness at larger scales); and above the 1:1 line in the

23 positive-positive (upper right) quadrat (larger increases in richness at larger scales). While there

1   are many cases that fall in these zones, this pattern is in fact less frequent than the opposite,

2   where effect sizes tend to decrease with increasing scale. Indeed, an unconstrained regression

3   shows that the slope is significantly positive, but significantly less than 1 (slope=0.65 ± 0.074;

4     $p<0.01$). This relationship is still relatively weak, but explains more of the variation than the

5   regression constrained to 1:1 in effect sizes across scales ($R^2 = 0.43$), and suggests that overall

6   effect sizes at smaller scales are on average larger than effect sizes at larger scales.

7

8   We categorized the different ecological drivers from these studies, and found that studies from

9   three drivers in particular— invasive species, land use change, and grazers/predators—showed

10   significantly larger effect sizes at small compared to larger spatial scales (see supplemental

11   Figure S1). These findings reinforce an earlier meta-analysis by Powell et al. (2011), who found

12   that the effect sizes of invasive species on plant richness systematically declined with increasing

13   spatial scale. Because such small-scale estimates of biodiversity change are often used as inputs,

14   for example, into 'biodiversity scenarios' models that project future diversity loss (e.g., Alroy

15   2017, Newbold et al. 2017), we suspect that these analyses may overestimate the actual change

16   observed at larger management-relevant scales.

17

18   **Towards a framework to estimate scale-dependent multidimensional biodiversity change**

19

20   Our above empirical review emphasizes what has been suspected for some time, but not

21   heretofore systematically quantified; comparisons of biodiversity and its change from site to site

22   are not well captured by a single number (Figure 2, 3) nor a single scale of observation (Figure

23   4). However, from a methodological perspective, the recognition that biodiversity is

1   multidimensional and scale-dependent is nothing new. There have been hundreds of

2   publications aimed at developing, analyzing and comparing various methodologies that can

3   capture some of this complexity.

4

5   Three of the most commonly used and recommended approaches for dealing with the

6   multidimensional scaling problem in biodiversity studies include: (1) converting species richness

7   into other measures that incorporate the relative abundances of species, such as Shannon's or

8   Simpson's diversity indices (e.g., Hill 1973, Lande 1996, Jost 2006, 2007); (2) comparing

9   species richness values after controlling for sampling effort in the numbers of samples or

10   individuals through rarefaction (e.g., Hurlbert 1971, Simberloff 1972, Gotelli and Colwell 2001,

11   Cayuela et al. 2015); and (3) extrapolating species richness values to a hypothetical asymptote

12   based on the estimation of the number of undetected species (e.g., Chao 1984, Colwell and

13   Coddington 1994, Chao et al. 2009). Importantly, these have been combined into a single

14   statistical framework that uses both interpolation and extrapolation (e.g., Chao and Jost 2012,

15   Colwell et al. 2012), and can be applied to Hill numbers, a family of diversity metrics that give

16   different weighting to the importance of rare and common species (e.g., Chao et al. 2014a).

17

18   Despite these advances, this framework does not fully solve the problem of spatial scale. For

19   example, rarefactions that control for the numbers of individuals will still identify scale-

20   dependent species richness rankings that depend on the numbers of individuals to which species

21   richness is rarefied (e.g., Gotelli and Colwell 2001, Cao et al. 2007, Chase and Knight 2013,

22   Cayuela et al. 2015). On the other hand, diversity extrapolations compare only the maximum

23   diversity in a community at a hypothetical sampling asymptote (e.g., the largest scale in Figure

1   1), which may require different levels of sampling intensity to achieve for different treatment

2   groups (Chao et al. 2009) and do not capture differences in the rise to that asymptote. Metrics

3   that quantify differences in evenness (e.g., Shannon's, Simpson's), can capture some of the

4   differences observed (e.g., Hill 1973, Chao et al. 2014a), but as we described above, few studies

5   use more than a single biodiversity metric which is necessary in order to fully evaluate the

6   complexity of biodiversity change, nor do they consider what any similarities or differences in

7   the metrics mean.

8

9   **Constructing and comparing elements of rarefaction curves**

10   Gotelli and Colwell (2001) contrasted several types of rarefaction and accumulation curves

11   which contain complementary information on how species diversity varies with sampling effort

12   and scale. Here, we focus on two that are useful for making comparisons across scales and

13   dissecting the relationships between different factors associated with biodiversity measurements:

14   sample-based rarefaction and individual-based rarefaction.

15

16   Sample-based rarefaction (SBR) considers increases in the numbers of species as the numbers of

17   sampling units increases (see Chiarucci et al. 2008 for early uses of this concept). The

18   rarefaction curve is an average of different accumulations, in which the ordering of the

19   accumulated samples is random. In our discussion, we modify the original SBR which is not

20   spatially explicit to create the spatial SBR (sSBR), which keeps spatial structure explicit (Figure

21   5A). That is, one starts with a single plot, then adds the closest plot, etcetera until all plots are

22   included, and the processes are repeated across starting plots (Chiarucci et al. 2009) (we prefer

23   the sSBR terminology rather than 'spatially constrained rarefaction' coined by Chiarucci et al.

19

1 [2009], because their term does not differentiate whether the rarefaction is sample- or individual-

2 based). It is important to note that sSBR is one form of what investigators have generically

3 called species-area relationships (Scheiner 2003). The sSBR, however, differs in some important

4 ways from the nested species area relationships that are most commonly used in macroecological

5 studies (e.g., Storch 2016), even though it is possible to estimate one from the other with certain

6 assumptions (e.g., Azaele et al. 2015, Kunin et al. 2018). We focus here on sSBR curves and

7 their derivatives, rather than nested species area relationships, because the sSBR is more closely

8 aligned with the way that the vast majority of field biologists collect and analyze their data.

9

10 From the sSBR, we can derive the most commonly used metrics associated with species richness

11 (S) using Whittaker's (1960) multiplicative diversity partition $\gamma=\alpha*\beta$ (Table 1) (see also Jost

12 2007, Tuomisto 2010) (see Crist and Veech 2006 for an additive partition within a similar

13 approach). Specifically, we can estimate species richness at the small scale, which we call α-

14 diversity and abbreviate as $^{\alpha}S$. Typically, this will be derived as an average value from a number

15 of replicates or sampling plots. Next, we can estimate species richness at the large scale, which

16 we call γ-diversity and abbreviate as $^{\gamma}S$. This will typically be derived from the total numbers of

17 species observed among all replicates or samples (nevertheless, it is important to note that simply

18 adding up species encountered in from multiple sampling plots within a larger regional extent

19 will typically be far from a complete survey of the true γ-diversity in that entire extent). Finally,

20 we can estimate the turnover in species from smaller to larger scales as β-diversity ($\beta=\gamma/\alpha$),

21 which we abbreviate as β_S.

22

1   Our definitions of α-, β- and γ-diversity are not intended to define any particular features of a

2   given community, such as zones where species interact and coexist, or where a given species

3   pool is defined. It is our view that it is unlikely that any discrete scale can be reasonably defined

4   to be explicitly 'local' or 'regional', but rather these demarcations are a somewhat arbitrary way

5   to define smaller and larger scale estimates of diversity and their comparison. Nevertheless, for

6   specific comparisons, it is critical to make sure that α- and γ-scale comparisons are done with

7   similar sampling effort (e.g., Gotelli and Colwell 2001). We also note that our use of β-diversity

8   in the context of the diversity partition is intended to capture variation within a given treatment

9   of an ecological driver, and is quite different than approaches that compare compositional

10 differences (Tuomisto 2010, Anderson et al. 2011).

11

12 Critically, the spatially explicit and sample-based structure of the sSBR and values derived from

13 it (e.g., α–, β–, γ–diversity) are jointly determined by the total numbers of individuals in a

14 community (N), the shape of the SAD (including total species numbers and their relative

15 abundances), and their intraspecific aggregations or clumping, and these components can have

16 bi-directional effects. Furthermore, changes in any one of the components can lead to similar

17 looking sSBRs, and changes in two or more components can lead to quite complex outcomes

18 (e.g., Chase and Knight 2013, Powell et al. 2013). Thus, one of our main goals in evaluating the

19 multidimensional and scale-dependent way by which biodiversity changes across space or time is

20 to dissect the influence of each of the components on the observed sSBR. While we discuss N,

21 the SAD and aggregation as the components influencing changes in diversity, it is important to

22 note that our approach is agnostic to mechanism. Thus, while changes in observed S that are

23 correlated with changes in total N in the community is consistent with some forms of the 'more

1   individuals hypothesis' (sensu Srivastava and Lawton 1999), for example, we cannot necessarily

2   infer that change in N causes change in S in a unidirectional way, but simply that they are

3   correlated (e.g., Storch et al. 2018).

4

5   Whereas the sSBR is the most complex curve that preserves information on differences among

6   two or more communities N, SAD and aggregation, the individual-based rarefaction (IBR) is the

7   most simple (Figure 5B). Specifically, in the IBR, individuals are pooled across all samples in a

8   treatment and selected at random, so that any spatial structure in intraspecific aggregation and

9   differences in the density of individuals in sampling plots is removed (e.g., Hurlbert 1971,

10   Simberloff 1972, Gotelli and Colwell 2001). While the distance along the x-axis (number of

11   individuals) is determined by N, the shape of the IBR is determined only by the SAD. The

12   height of the IBR is the total numbers of species (S) observed among all of the samples pooled,

13   and the steepness of the slope near the origin is determined by the relative abundances

14   (evenness) of species represented in the pooled samples (Hurlbert 1971, Gotelli and Colwell

15   2001, Olszwewski 2004, Chase and Knight 2013, Cayuela et al. 2015).

16

17   From the IBR, we can derive a number of parameters of interest (Figure 5B-D; Table 1). To

18   derive these parameters across scales, we calculate two IBRs, one using only the N and S from

19   samples collected at the α-scale (i.e., from a single replicate or plot, or several plots within a

20   locality embedded within a metacommunity) and one from the γ-scale across all replicates or

21   localities within a given treatment or type (a multi-scale version of this approach, including

22   benchmark tests and evaluation of statistical error, is presented in McGlinn et al. 2018). From

1   the α- and γ-scale IBRs, and the measures we can derive from these, we can begin to understand

2   the relationship between N, the SAD and aggregation and the observed diversity patterns.

3

4   ***Disentangling the influence of N by comparing rarefied richness values***—We can first

5   examine whether any differences in $^{\alpha}S$ or $^{\gamma}S$ observed from the sSBR result from differences in

6   total N among treatments. Specifically, with the IBR, we can calculate the numbers of species

7   expected when rarefied to a common N, known as rarefied richness, abbreviated as S_n. Because

8   differences in rarefied richness are also scale-dependent (e.g., Cao et al. 2007, Chase and Knight

9   2013, Cayuela et al. 2015), we advise comparing S_n at multiple values of N. Further, we can

10   compare values of S_n at different scales of sampling, which we denote $^{\alpha}S_n$ and $^{\gamma}S_n$, respectively.

11

12   ***Disentangling the influence of SAD by comparing elements of individual-based rarefaction***

13   ***curves***—If the shapes of the IBRs differ between treatments, we can infer that SADs differ via

14   changes in the numbers of species (S), the relative abundances (evenness) of species, or both (as

15   shown above in Figure 1). To evaluate this, we can perform statistical comparisons at two

16   extremes of the IBR, the upper limit (S) at the right extreme of the curve and the slope at the

17   base of the IBR. The upper limit of the IBR can be calculated in a number of ways, depending

18   on which sampling controls are most appropriate, including the rarefied richness S_n at

19   standardized values of N (e.g., Gotelli and Colwell 2001), S calculated to a fixed coverage of

20   sample completeness (e.g., Chao and Jost 2012) or an extrapolation to a hypothetical asymptote

21   where coverage is complete (e.g., Chao 1984, Chao et al. 2009, 2014a). The slope at the base of

22   the IBR, as described above, is equivalent to the bias-corrected probability of interspecific

23   encounter (PIE) (Hurlbert 1971, Olszwewski 2004), and represents a measure of evenness in the

1   community. The PIE can be measured from samples at the α-scale ($^{\alpha}$PIE), or at the γ-scale

2   ($^{\gamma}$PIE). Because PIE is the same as 1-Simpson's diversity index, we can convert it to an effective

3   number of species (S_{PIE}) as above, and calculate this quantity at each spatial scale: $^{\alpha}S_{PIE}$ and

4   $^{\gamma}S_{PIE}$. Both PIE and S_{PIE} can provide useful information for comparisons between two or more

5   communities (e.g., Dauby and Hardy 2012). Because S and PIE (or S_{PIE}) are indicative of

6   different parts of the IBR, and also reflect different locations along the continuum of the Hill

7   numbers (e.g., Hill 1973, Jost 2006), they conceptually link the IBR with Hill numbers, which

8   differentially weight common versus rare species.

9

10 ***Disentangling β-diversity due to aggregation using individual-based rarefaction curves***—Raw

11 measures of β-diversity from the sSBR are influenced by N and the shape of the SAD, as well as

12 by spatial aggregation. To tease apart the specific influence of aggregation on β-diversity, we

13 can compare the γ-scale IBR (which randomly samples individuals regardless of their spatial

14 position), to the α-scale IBR (which preserves effects of aggregation within plots or replicates).

15

16 First, we can compute $β_{PIE}$ as difference between the slopes defined by $^{\alpha}$PIE and $^{\gamma}$PIE at the base

17 of the respective IBRs (Figure 5B-D) (Olszwewski 2004, Dauby and Hardy 2012). And we can

18 convert this to an effective number of species by taking the ratio of small-scale ($^{\alpha}S_{PIE}$) to large-

19 scale ($^{\gamma}S_{PIE}$) to yield $β_{S_{PIE}}$ (Jost 2007, Tuomisto 2010). Second, we can compute $β_{S_n}$ as the ratio

20 of the number of species expected from a given *n* from the γ-scale IBR ($^{\gamma}S_n$) to the number of

21 species observed from a given *n* from the α-scale IBR ($^{\alpha}S_n$). Three qualitative outcomes of these

22 different components of β-diversity are possible. If there is no aggregation, the γ-scale and α-

1   scale IBRs will not differ in their shapes and there will be no differences due to spatial clumping

2   in the landscape. Thus β_{PIE}, $\beta_{S_{PIE}}$, and β_{S_n} will equal 1 (Figure 5B). Second, if there is strong

3   aggregation in the community, there can be large differences in γ-scale and α-scale IBRs, leading

4   to high levels of β_{PIE} (and $\beta_{S_{PIE}}$), as well as β_{S_n} (Figure 5C). Finally, it is possible that the γ-scale

5   and α-scale IBRs can differ in their shapes, but do not differ in PIE, which measures turnover in

6   the most common species, but misses changes that might be due to aggregation of rarer species

7   in the community. Here, we would observe β_{PIE} (or $\beta_{S_{PIE}}$) equal to 0, but β_{S_n} greater than 0

8   (Figure 5D).

9

10   **Recommended protocol for analyses dissecting biodiversity data with a case study**

11   Next, we provide a recipe for using the multiple metrics at multiple scales, as presented above

12   and overviewed in Table 1. We describe how one might address these questions generically, and

13   illustrate this recipe with a re-analysis of an experiment on the effect of nutrient additions on

14   macroinvertebrates and amphibians in experimental ponds, showing how deeper insights can be

15   gained than are possible with a more traditional perspective (for full details, see Chase 2010).

16

17   R code for these analyses, as well as a series of vignettes and another case study, are available in

18   *mobr* statistical package (https://github.com/MoBiodiv/mobr) (McGlinn et al. 2018). We

19   specifically used one-way PERMANOVA to assess the treatment effect on estimates of diversity

20   measured at the α-scale (=1 mesocosm) (treatment group labels were permuted 999 times to

21   generate a null distribution, and then the tail probability for a treatment effect was evaluated by

22   comparing the observed difference to the null distribution). At the γ-scale (=15 mesocosms), the

23   null distribution was generated by permuting treatment group labels across samples, pooling the

1   groups, and then calculating the difference in diversity between treatments for each permutation.

2   Results are illustrated in Figure 6; some results are not shown for brevity.

3

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

5   If treatments differ in N, then any difference in S between treatments may, in part, be due to

6   treatment effects on N. We expect that N scales linearly with increasing sampling (unlike S),

7   and so it should not matter whether the average N per plot (α-scale), or the total N across plots

8   (γ-scale), are compared. In our case study, we found no influence of the nutrient addition

9   treatment on N measured at either scale ($P > 0.3$; not shown in Figure).

10

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

12   Here, we can compare several types of α-scale diversity. Comparing effects of the treatment on

13    $^{\alpha}S$ with treatment effects on $^{\alpha}S_n$ allow us to discern any influence of changes in N on the results.

14   For example, if any treatment-level effects on $^{\alpha}S$ disappear for $^{\alpha}S_n$, we can conclude that the

15   effect was due to treatment effects on N. However, if there remains a difference in $^{\alpha}S_n$, we can

16   conclude that the treatment effect on $^{\alpha}S$ was due, at least in part, to changes in the SAD. It is

17   even possible that the effects can shift between $^{\alpha}S$ and $^{\alpha}S_n$. For example, McCabe and Gotelli

18   (2000) showed that $^{\alpha}S$ was higher for stream invertebrates in undisturbed treatments, but that this

19   effect was reversed when the effect of disturbance on N was discounted; $^{\alpha}S_n$ was higher in

20   disturbed treatments. A difference in $^{\alpha}S_{PIE}$ implies a change in the dominance patterns of

21   common species in a community. Alternatively, it is possible that $^{\alpha}S_{PIE}$ could not vary (or vary

22   little), while there is still an influence of the treatment on $^{\alpha}S_n$. This would imply that rarer

23   species are responding to the treatment.

1

2   From our case study, we observed that $^{\alpha}S$ did not differ between the treatments (Fig. 6A), and

3   because there was no difference in N, the effects of N on $^{\alpha}S$ and $^{\alpha}S_n$ did not differ among

4   treatments (result not shown in figure). Furthermore, there was no influence of treatment on

5    $^{\alpha}S_{PIE}$ (Figure 6B), indicating that nutrient addition treatment had no identifiable effects on N or

6   any measures of diversity that capture differences in the shape of the SAD at the local scale. Had

7   we stopped here, we might conclude that there was no interesting influence of nutrients on

8   biodiversity. However, by looking at the γ-scale, we see that this conclusion is wrong.

9

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

11   Comparisons of γ-scale diversity measures are a straightforward extension of the α-scale

12   measures, except we use the totals observed in all samples, rather than the average number from

13   a single sample. And the relationships among the measures are the same, where $^{\gamma}S$ is the total

14   richness of species, $^{\gamma}S_n$ is the richness of species rarefied to a common N, and $^{\gamma}PIE$ and $^{\gamma}S_{PIE}$

15   consider the relative abundances of species and conversion to an effective number of species,

16   respectively.

17

18   From our case study, and in contrast to the local-scale result, we found a significant increase in

19    $^{\gamma}S$ (Fig. 6C) and $^{\gamma}S_n$ (not shown) with nutrient additions. However, there was no influence of

20   nutrient additions on $^{\gamma}PIE$ (not shown) or $^{\gamma}S_{PIE}$ (Figure 6D). This suggests that it is the rare

21   species end of the SAD that changed among the treatments; nutrient enrichment allowed more

22   rare species to persist regionally, but not locally. Further, the qualitatively different results at the

23   γ-and α-scale imply that β-diversity was influenced by the nutrient addition treatment.

1

2   *Step 4: Do treatments affect β-diversity?*

3   First, we can test for differences in β_S among treatments. However, as discussed above, any

4   differences could occur due to a combination of changes in N, the SAD or aggregation. If the

5   treatment influences aggregation, we would expect a difference in β_{S_n} between treatments. If the

6   treatment affects the aggregation of common species, we would also expect to see treatment-

7   level effects on $\beta_{S_{PIE}}$.

8

9   As anticipated from the scale-dependent results we observed at the α– and γ-scale (Figure 6A,

10   6C), we found strong effects of nutrient addition on β_S (Fig. 6E), and on β_{S_n} (Fig. 6F). And

11   because we found no difference in the α– and γ-scale results for S_{PIE}, we unsurprisingly found no

12   treatment-level effect on $\beta_{S_{PIE}}$ (Fig. 6G). Overall, these patterns suggest that a core group of

13   common species were present across replicates in both treatments, and that a group of rarer

14   species were largely responsible for the treatment-level responses at the γ-scale. Specifically,

15   there was more turnover among those rare species in the high nutrient treatment, which may have

16   resulted from ecological drift and/or priority effects.

17

18   **Conclusions**

19   There are multiple pathways by which ecological factors can influence biodiversity across

20   multiple scales, but most studies continue to rely on comparisons of a single summary variable—

21   usually species richness (S)—at a single spatial scale. As a consequence, despite thousands of

22   published studies quantifying how species richness changes in response to natural and

23   anthropogenic drivers, we know much less than we think about how and why biodiversity

28

1   changes from place to place and time to time. This is particularly problematic when trying to

2   achieve synthesis across multiple studies of multiple ecological drivers, through meta-analyses

3   and other means, because effect sizes are highly confounded by spatial scale (see also Chase and

4   Knight 2013). We are currently limited in our ability to create realistic 'biodiversity scenarios'

5   models that project future biodiversity loss in response to changing ecological conditions.

6

7   To move forward, it is critical to consider the multidimensional and scale-dependent nature of

8   biodiversity and its change. Fortunately, there is a rich literature on other measures of

9   biodiversity that can explicitly complement comparisons of S, and there are easy ways that

10   empiricists can explicitly deal with issues of spatial scaling (i.e., replicates nested within

11   treatments). We have advocated for measures of biodiversity and its scaling that can explicitly

12   be used to disentangle the influence of N, the SAD and aggregation via consideration of different

13   aspects of the individual-level rarefaction curve that emphasize different underlying components

14   (e.g., S_n, S_{PIE}, at the α and γ scales). For a majority of studies, these can be estimated in a

15   straightforward way with data that are already, or could be, collected.

16

17   The approach we advocate will often require more complex collecting and reporting of data (i.e.,

18   absolute species abundances at multiple scales) and more complex analyses of multiple response

19   variables. This will create a more nuanced view of what biodiversity is and how it varies—

20   biodiversity is not a single number, and it cannot be compared at a single scale to estimate how it

21   responds to ecological factors. However, this more nuanced view is necessary to resolve long-

22   standing debates. For example, Blowes et al. (2017) recently used this approach to show that the

23   debate about whether environmental versus historical biogeographic controls influence global

1   biodiversity patterns of reef-associated fishes can be progressed by dissecting patterns of species

2   richness in a scale-explicit way.

3

4   Finally, there are many extensions of the approach that we advocate which are necessary to be

5   able to fully understand, and synthesize, how biodiversity changes in time and space. First, as

6   we mentioned above, the approach we have taken here views scale in a discrete two-scale way

7   (e.g., α, β, γ–diversity), while scale is continuous. In a companion paper, we develop a multi-

8   scale methodology (McGlinn et al. 2018). Second, we have only focused on taxonomic

9   diversity, although interest in other measures of diversity has increased greatly in recent years.

10   These measures, such as functional and phylogenetic diversity, show patterns of scaling similar

11   to taxonomic diversity (e.g., Morlon et al. 2011, Smith et al. 2013) and will thus show scale-

12   dependence when making comparisons. Approaches similar to those advocated here are

13   emerging for these other types of diversity (Chao et al. 2014b, 2015, Chiu et al. 2104), and so we

14   anticipate that a family of approaches for comparing scale-dependent diversity responses to

15   ecological drivers at multiple levels of organization will soon emerge.

16

1　Storch, two anonymous reviewers and the handling editor, and a number of other colleagues that

2　helped us to develop and reform these ideas and tools.

3

4　**Literature Cited**

5　Alroy, J., (2017). Effects of habitat disturbance on tropical forest biodiversity. Proc. Natl. Acad.

6　Sci. 114, 6056–6061.

7

8　Azaele, S., Maritan, A., Cornell, S.J., Suweis, S., Banavar, J.R., Gabriel, D. and Kunin, W.E.,

9　(2015). Towards a unified descriptive theory for spatial ecology: Predicting biodiversity patterns

10　across spatial scales. Methods Ecol. Evol. 6, 324-332.

11

12　Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., et al,

13　(2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist.

14　Ecol. Lett., 14, 19-28.

15

16　Andrew, M.E., Wulder, M.A., Coops, N.C. and Baillargeon, G., (2012). Beta-diversity gradients

17　of butterflies along productivity axes. Global Ecol. Biogeog. 21: 352-364.

18

19　Baselga, A., (2010). Partitioning the turnover and nestedness components of beta diversity.

20　Global Ecol. Biogeog., 19, 134-143.

21

22　Blowes, S. A., Belmaker, J., and Chase, J. M. (2017). Global reef fish richness gradients emerge

23　from divergent and scale-dependent component changes. Proc. R. Soc. B 284, 20170947

1

Bürkner, P-C. (2017). brms: An R package for Bayesian multilevel models using Stan. J. Stat.

Soft., 80, 1-28.

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.

1

2   Chao, A., Chiu, C.H. and Jost, L., (2014b). Unifying species diversity, phylogenetic diversity,

3   functional diversity, and related similarity and differentiation measures through Hill numbers.

4   Annu. Rev. Ecol. Syst, 45, 297-324.

5

6   Chao, A., Chiu, C.H., Hsieh, T.C., Davis, T., Nipperess, D.A. and Faith, D.P., (2015).

7   Rarefaction and extrapolation of phylogenetic diversity. Methods Ecol. Evol., 6, 380-388.

8

9   Chase, J.M. and Leibold, M.A., (2002). Spatial scale dictates the productivity–biodiversity

10 relationship. Nature, 416, 427-430.

11

12 Chase, J.M., (2010). Stochastic community assembly causes higher biodiversity in more

13 productive environments. Science, 328, 1388-1391.

14

15 Chase, J. M., and Knight, T. M. (2013). Scale-dependent effect sizes of ecological drivers on

16 biodiversity: why standardised sampling is not enough. Ecol. Lett., 16, 17-26.

17

18 Chiarucci, A., Bacaro, G., Rocchini, D. and Fattorini, L., (2008). Discovering and rediscovering

19 the sample-based rarefaction formula in the ecological literature. Comm. Ecol., 9, 121-123.

20

21 Chiarucci, A., Bacaro, G., Rocchini, D., Ricotta, C., Palmer, M. and Scheiner, S., (2009).

22 Spatially constrained rarefaction: incorporating the autocorrelated structure of biological

23 communities into sample-based rarefaction. Comm. Ecol., 10, 209-214.

1

2   Chiu, C.H., Jost, L. and Chao, A., (2014). Phylogenetic beta diversity, similarity, and

3   differentiation measures based on Hill numbers. Ecol. Monog., 84, 21-44.

4

5   Colwell, R.K. and Coddington, J.A., (1994). Estimating terrestrial biodiversity through

6   extrapolation. Phil. Trans. Roy. Soc B, 345, 101-118.

7

8   Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.Y., Mao, C.X., Chazdon, R.L. et al. (2012). Models

9   and estimators linking individual-based and sample-based rarefaction, extrapolation and

10   comparison of assemblages. J. Plant Ecol., 5, 3-21.

11

12   Condit R, Lao, S, Pérez, R. Dolins, S. B., Foster, R. B. and Hubbell, S. P. (2012). Barro

13   Colorado Forest Census Plot Data, 2012 Version. http://dx.doi.org/10.5479/data.bci.20130603

14

15   Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. and Magurran,

16   A.E., (2014). Assemblage time series reveal biodiversity change but not systematic loss. Science,

17   344, 296-299.

18

19   Dauby, G. and Hardy, O.J., (2012). Sample-based estimation of diversity sensu stricto by

20   transforming Hurlbert diversities into effective number of species. Ecography, 35, 661-672.

21

1 Dumbrell, A.J., Clark, E.J., Frost, G.A., Randell, T.E., Pitchford, J.W. and Hill, J.K., (2008).

2 Changes in species diversity following habitat disturbance are dependent on spatial scale:

3 theoretical and empirical evidence. J. Appl. Ecol., 45, 1531-1539.

4

5 Gardezi, T. and Gonzalez, A. (2008). Scale dependence of species-energy relationships:

6 Evidence from fishes in thousands of lakes. Am. Nat. 171, 800-815.

7

8 Gelman, A., Goodrich, B., Gabry, J., Ali, I. (2017) R-squared for Bayesian regression models.

9 *http://www.stat.columbia.edu/~gelman/research/unpublished/*

10

11 Gotelli, N.J. and Colwell, R.K., (2001). Quantifying biodiversity: procedures and pitfalls in the

12 measurement and comparison of species richness. Ecol. Lett., 4, 379-391.

13

14 Harrod, J. (1964). The distribution of invertebrates on submerged aquatic plants in a chalk

15 stream. J. Anim. Ecol., 33, 335-348.

16

17 He, F. and Legendre, P., (2002). Species diversity patterns derived from species–area models.

18 Ecology, 83, 1185-1198.

19

20 Hill, M.O., (1973). Diversity and evenness: a unifying notation and its consequences. Ecology,

21 54, 427-432.

22

1   Hill, J.K. and Hamer, K.C., (2004). Determining impacts of habitat modification on diversity of

2   tropical forest fauna: the importance of spatial scale. J. Appl. Ecol., 41, 744-754.

3

4   Hillebrand, H., Blasius, B., Borer, E.T., Chase, J.M., Downing, J.A., Eriksson, et al., 2018.

5   Biodiversity change is uncoupled from species richness trends: Consequences for conservation

6   and monitoring. Journal Appl. Ecol., 55, 169-184.

7

8   Hurlbert, S.H., (1971). The nonconcept of species diversity: a critique and alternative

9   parameters. Ecology, 52, 577-586.

10

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

12

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

14 2427-2439.

15

16 Keil, P., Biesmeijer, J.C., Barendregt, A., Reemer, M. and Kunin, W.E., (2011). Biodiversity

17 change is scale-dependent: an example from Dutch and UK hoverflies (Diptera, Syrphidae).

18 Ecography, 34, 392-401.

19

20 Kunin, W.E., Harte, J., He, F., Hui, C., Jobe, R.T., Ostling, A., Polce, C., Šizling, A., Smith,

21 A.B., Smith, K. and Smart, S.M., (2018). Upscaling biodiversity: estimating the species–area

22 relationship from small samples. Ecological Monographs 88, 170-187.

23

1   Lande, R., (1996). Statistics and partitioning of species diversity, and similarity among multiple

2   communities. Oikos, 5-13.

3

4   Lande, R., DeVries, P.J. and Walla, T.R., (2000). When species accumulation curves intersect:

5   implications for ranking diversity using small samples. Oikos, 89, 601-605.

6

7   Legendre, P., 2014. Interpreting the replacement and richness difference components of beta

8   diversity. Global Ecol. Biogeog., 23, 1324-1334.

9

10   Livne-Luzon, S., Ovadia, O., Weber, G., Avidan, Y., Migael, H., et al. (2017). Small-scale

11   spatial variability in the distribution of ectomycorrhizal fungi affects plant performance and

12   fungal diversity. Ecol. Lett., 20, 1192-1202.

13

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

15

16   Magurran, A. E., and McGill, B. J. (eds) (2011) Biological diversity: Frontiers in measurement

17   and assessment. Oxford University Press.

18

19   May, F., Gerstner, K. . McGlinn, D. J., Xiao, X and Chase, J. M. (2017) mobsim: An R package

20   for the simulation and measurement of biodiversity across spatial scales. bioRxiv (2017):

21   209502.

22

1   Martinson, V.G., Douglas, A.E. and Jaenike, J., (2017). Community structure of the gut

2   microbiota in sympatric species of wild Drosophila. Ecol. Lett., 20, 629-639.

3

4   McCabe, D.J. and Gotelli, N.J., (2000). Effects of disturbance frequency, intensity, and area on

5   assemblages of stream macroinvertebrates. Oecologia, 124, 270-279.

6

7   McGill, B. J. (2011a). Linking biodiversity patterns by autocorrelated random sampling. Am. J.

8   Bot., 98, 481-502.

9

10 McGill, B. J. (2011b). Species abundance distributions. In. A. Magurran and B. McGill (eds)

11 Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford

12 (201:1) 105-122.

13

14 McGlinn, D.J. and Palmer, M.W., (2009). Modeling the sampling effect in the species–time–area

15 relationship. Ecology, 90, 836-846.

16

17 McGlinn, D. Xiao, X., May, F. Gotelli, N. J., Blowes, S. A., Knight, T. M. et al. Submitted.

18 MoB (Measurement of Biodiversity): a method to separate the scale-dependent effects of species

19 abundance distribution, density, and aggregation on diversity change. Submitted to Meth. Ecol.

20 Evol. (available also on https://www.biorxiv.org/content/early/2018/01/07/244103)

21

22 Mendes, R.S., Evangelista, L.R., Thomaz, S.M., Agostinho, A.A. and Gomes, L.C., (2008). A

23 unified index to measure ecological diversity and species rarity. Ecography, 31, 450-456.

1

2   Morlon, H., Schwik, D. W. Bryant, L. A., Marquet, P. A., Rebelo, A. G. et al. (2011). Spatial

3   patterns of phylogenetic diversity. Ecol. Lett. 14, 141-149.

4

5   Newbold, T., Boakes, E.H., Hill, S.L., Harfoot, M.B. and Collen, B., (2017). The present and

6   future effects of land use on ecological assemblages in tropical grasslands and savannas in

7   Africa. Oikos, 126, 1760–1769.

8

9   Olszewski, T.D., (2004). A unified mathematical framework for the measurement of richness

10   and evenness within and among multiple communities. Oikos, 104, 377-387.

11

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

13   Nat., 144, 717-740.

14

15   Pardieck, K. L., Ziolkowski Jr, D. J., Lutmerding, M, Campbell, K., and Hudson, M-, A. R..

16   2017. North American Breeding Bird Survey Dataset 1966 - 2016, version 2016.0. U.S.

17   Geological Survey, Patuxent Wildlife Research Center. <www.pwrc.usgs.gov/BBS/RawData/>;

18   doi:10.5066/F7W0944J.

19

20   Powell, K.I., Chase, J.M. and Knight, T.M., (2011). A synthesis of plant invasion effects on

21   biodiversity across spatial scales. Am. J. Bot., 98, 539-548.

22

1   Powell, K.I., Chase, J.M. and Knight, T.M., (2013). Invasive plants have scale-dependent effects

2   on diversity by altering species-area relationships. Science, 339(6117), pp.316-318.

3

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

5

6   Rahbek, C., (2005). The role of spatial scale and the perception of large-scale species-richness

7   patterns. Ecology Letters, 8, 224-239.

8

9   R Core Team (2017). R: A language and environment for statistical computing. R Foundation for

10 Statistical Computing, Vienna, Austria.   URL http://www.R-project.org/.

11

12 Sandel, B. and Smith, A.B., (2009). Scale as a lurking factor: incorporating scale-dependence in

13 experimental ecology. Oikos, 118, 1284-1291.

14

15 Scheiner, S.M., (2003). Six types of species-area curves. Global Ecol. Biogeog., 12, 441-447.

16

17 Scheiner, S. M., Cox, S.B., Mittelbach, G.G., Osenberg, C. and Kaspari, M., (2000). Species

18 richness, species–area curves and Simpson's paradox. Evol. Ecol. Res. 2, 791-802.

19

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

21 418.

22

1   Smith, A.B., Sandel, B., Kraft, N.J. and Carey, S., (2013). Characterizing scale-dependent

2   community assembly using the functional-diversity-area relationship. Ecology, 94, 2392-2402.

3

4   Solar, R.R.D.C., Barlow, J., Ferreira, J., Berenguer, E., Lees, A.C., et al., (2015). How pervasive

5   is biotic homogenization in human-modified tropical forest landscapes? Ecol. Lett., 18, 1108-

6   1118.

7

8   Srivastava, D.S. and Lawton, J.H., (1998). Why more productive sites have more species: an

9   experimental test of theory using tree-hole communities. Am. Nat., 152, 510-529.

10

11 Storch, D., (2016). The theory of the nested species–area relationship: geometric foundations of

12 biodiversity scaling. J. Veg. Sci., 27, 880-891.

13

14 Storch, D., Bohdalková, E. and Okie, J., (2018). The more-individuals hypothesis revisited: the

15 role of community abundance in species richness regulation and the productivity-diversity

16 relationship. Ecology Letters 21, 920-937

17

18 Thompson, G.G. and Withers, P.C., (2003). Effect of species richness and relative abundance on

19 the shape of the species accumulation curve. Aust. Ecol., 28, 355-360.

20

21 Tuomisto, H., (2010). A diversity of beta diversities: straightening up a concept gone awry. Part

22 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography, 33, 2-22.

23

1   Whittaker, R.H., (1960). Vegetation of the Siskiyou mountains, Oregon and California. Ecol.

2   Monog., 30, 279-338.

3

4   Whittaker, R. J. Willis, K. J., and Field, R. (2001). Scale and species richness: towards a

5   general, hierarchical theory of species diversity. J. of Biogeog. 28: 453-470

6

1

2 **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)
${}^{\alpha}PIE$, ${}^{\gamma}PIE$	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
${}^{\alpha}S_{PIE}$, ${}^{\gamma}S_{PIE}$	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

43

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.
β_{PIE}	Ratio of total treatment $^\gamma PIE$ and $^\alpha PIE$ (Olszewski 2004)	Higher difference between $^\gamma PIE$ and $^\alpha PIE$ indicates higher levels of aggregation.
β_{S_n}	Ratio of total treatment $^\gamma S_n$ and $^\alpha S_n$	Like β_S but emphasizes aggregation due to common and rare species.
$\beta_{S_{PIE}}$	Ratio of total treatment $^\gamma S_{PIE}$ and $^\alpha S_{PIE}$ (Jost 2007)	Like β_S but emphasizes aggregation due to common species.

1

1   **Figure Legends:**

2   **Figure 1.** A. Individual-based rarefaction curves of three hypothetical communities (labelled

3   A,B, C) where ranked differences between communities are consistent across scales. B.

4   Individual-based rarefaction curves of three hypothetical communities (labelled A,B, C) where

5   rankings between communities switch because of differences in the total numbers of species, and

6   their relative abundances. Dotted vertical lines illustrate sampling scales where rankings switch.

7   These curves were generated using the sim_sad function from the mobsim R package (May et al.

8   2018).

9

10   **Figure 2.** Bivariate relationships between N, S_{PIE} and S for 346 communities across the 37

11   datasets taken from McGill (2011b)(see Appendix 1). (A) S as a function of N; (B) S as a

12   function of S_{PIE}. (N vs S_{PIE} not shown). Black lines depict the relationships across studies (and

13   correspond to R^2 fixed); colored points and lines show the relationships within studies. All axes

14   are log-scale. Insets are histograms of the study-level slopes, with the solid line representing the

15   slope across all studies. Gray bars indicate the study-level slope did not differ from zero, blue

16   indicates a significant positive slope, and red indicates a significant negative slope.

17

18   **Figure 3.** Representative rarefaction curves, the proportion of curves that crossed, and counts of

19   how often curves crossed. (A) Rarefaction curves for different local communities within two

20   datasets: marine invertebrates (nematodes) along a gradient from a waste plant outlet

21   (Lambshead 1986), and trees in a Ugandan rainforest (Eggeling 1947); axes are log-transformed.

22   (B) Counts of how many times pairs of rarefaction curves (from the same community) crossed;

23   y-axis is on a log-scale.

24

1   **Figure 4.** Results of a meta-analysis of scale-dependent responses to a number of different

2   ecological drivers (see Appendix 2). Points represent the log response ratio comparing species

3   richness in control compared to treatments in a given comparison measured at the smallest (x-

4   value) and largest (y-value) scale. The solid line indicates the 1:1 line expected if effect sizes

5   were not scale-dependent. Points above and below this line indicate effect sizes that are larger or

6   smaller, respectively, as scale increases; points in the upper left and lower right quadrats

7   represent cases where the direction of change shifted from positive to negative, or vice versa,

8   with increasing scale. The dashed line indicates the best fit correlation, which is significantly

9   different than the 1:1 line (P<0.01), indicating that overall, effect sizes tend to be larger at

10   smaller scales than at larger scales. Colors for points indicate categorizations into different

11   ecological drivers.

12
13
14   **Figure 5.** (A) Spatial Sample-Based Rarefaction (sSBR) as defined in text. The average

15   number of species in a sample is defined as the α-scale, and the total number of species across

16   samples is defined as the γ-scale. From this, we can derive local richness ($^{\alpha}S$), regional richness

17   ($^{\gamma}S$), and β-richness ($\beta_s = {^{\gamma}S}/{^{\alpha}S}$). (B), (C), (D) individual-based rarefaction curves (IBR) from α-

18   scale samples (dashed lines) and γ-scale samples (solid lines). From this, we can visualize S_n

19   and PIE for each scale, and the differences between them (β_{Sn}, β_{PIE}). Other metrics presented in

20   Table 1 are not shown for clarity. (B) illustrates a case where there is no aggregation in the

21   community, and the IBRs from the α- and γ-scale completely overlap. (C) illustrates a case

22   where there is considerable aggregation in the community, and the IBRs from the α- and γ-scale

23   are completely different; here, β_{Sn} and β_{PIE} are high. (D) illustrates a case where there is some

46

1   aggregation in the community, it is mostly manifest among the rarer species, such that there is no

2   β_{PIE}, but there is β_{Sn}.

3

4

5   **Figure 6.** Effect of nutrient additions on several measurements of biodiversity from Table 1 (see

6   data in Appendix 3). Each biodiversity measure was calculated at the α-scale (1 mesocosm)

7   (Panels A,B), γ-scale (15 mesocosms) (Panels C,D), as well as the β-scale (i.e. turnover across

8   scales, γ/α)(Panels D,E,F). See Chase (2010) for details on the experimental design and the mobr

9   package (McGlinn et al. submitted, https://github.com/MoBiodiv/mobr) for details on the

10   statistical methods

11

12

13

14

1
2

3
4
5
6
7
8   **Figure 1.**

1

2   **Figure 2.**

1
2
3 **Figure 3.**
4
5

1

2

3
4   **Figure 4.**

5

6

1

3   **Figure 5.**
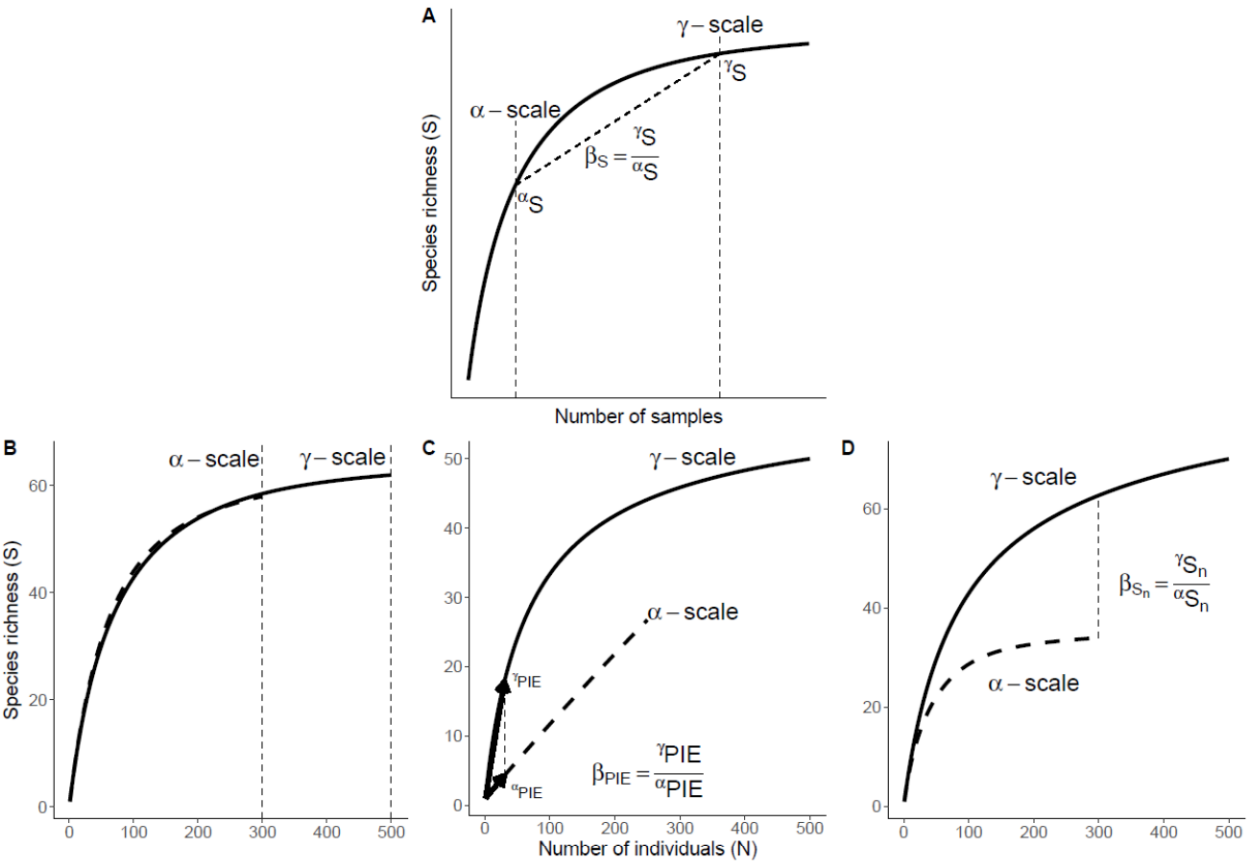
4

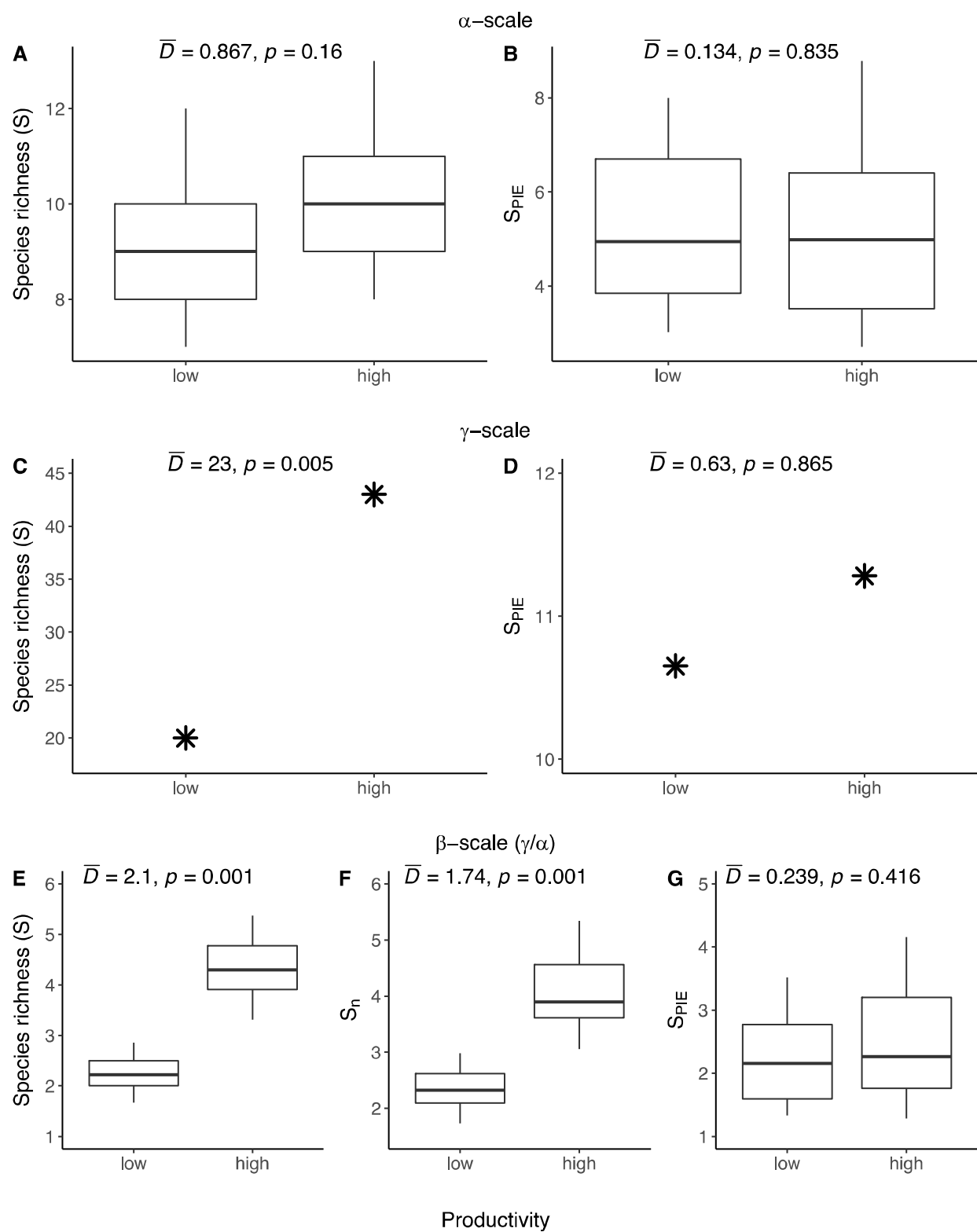52

1

Figure 6.

1
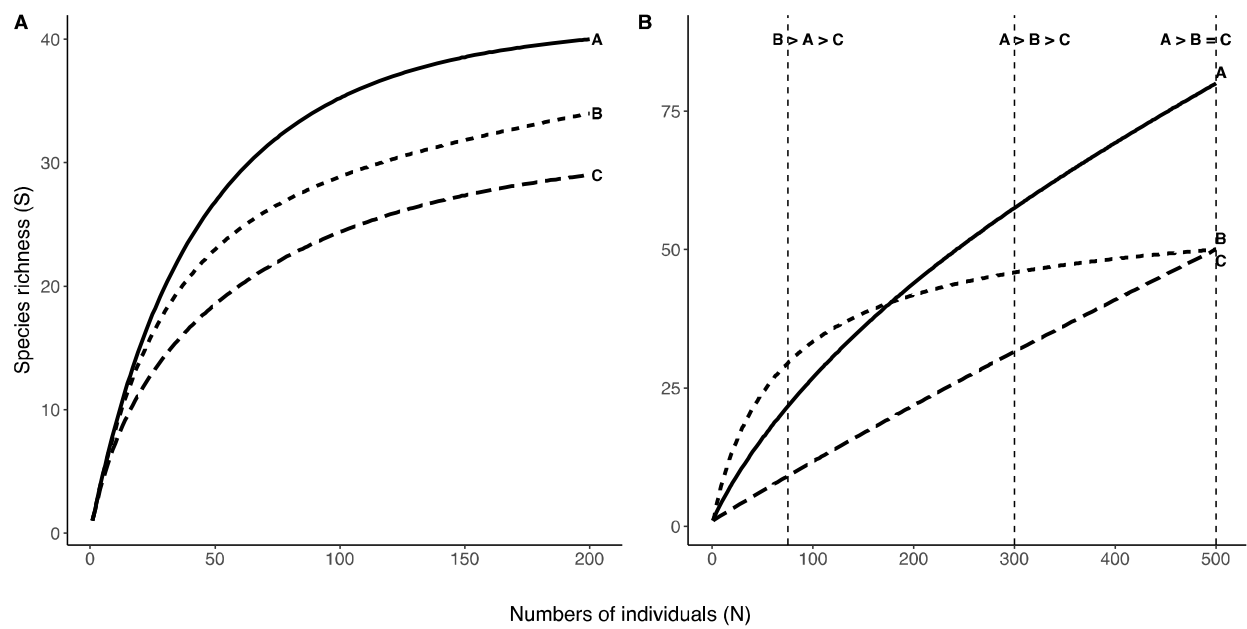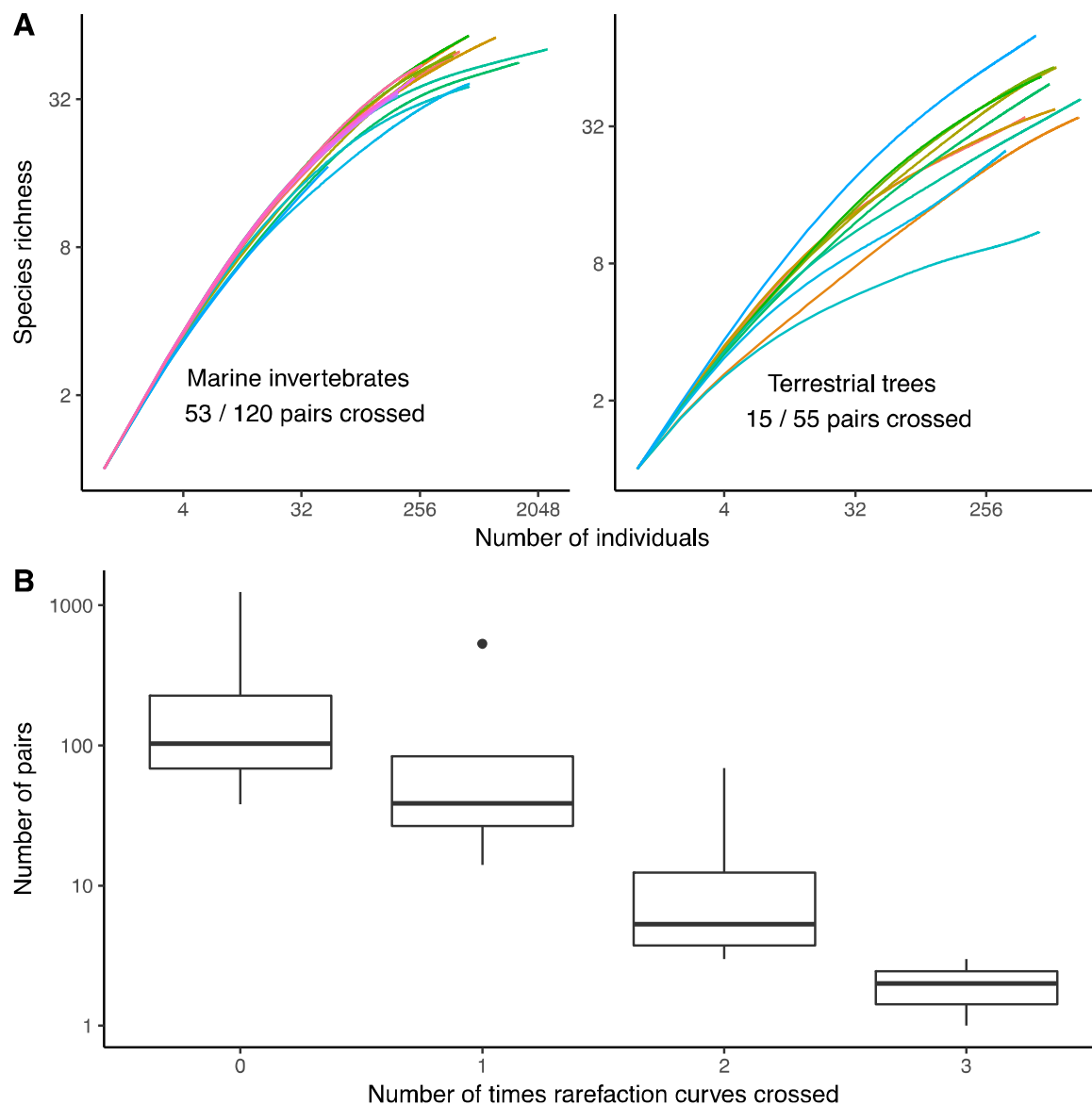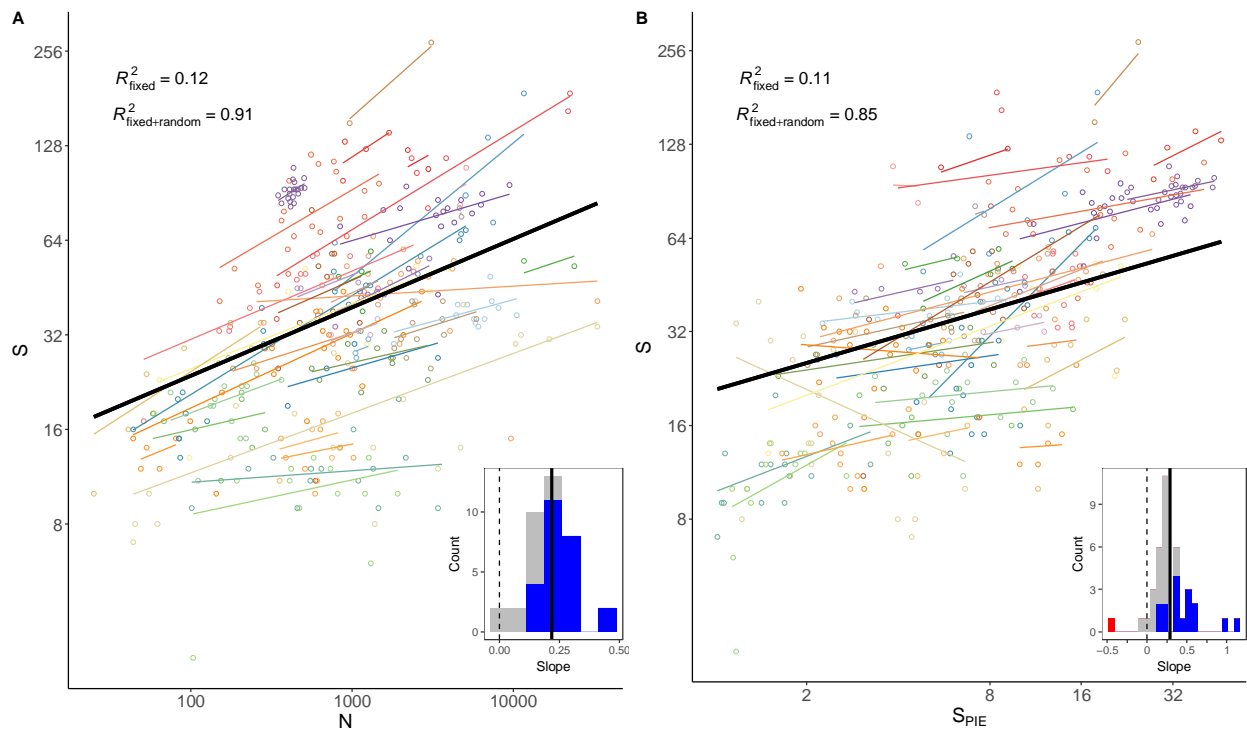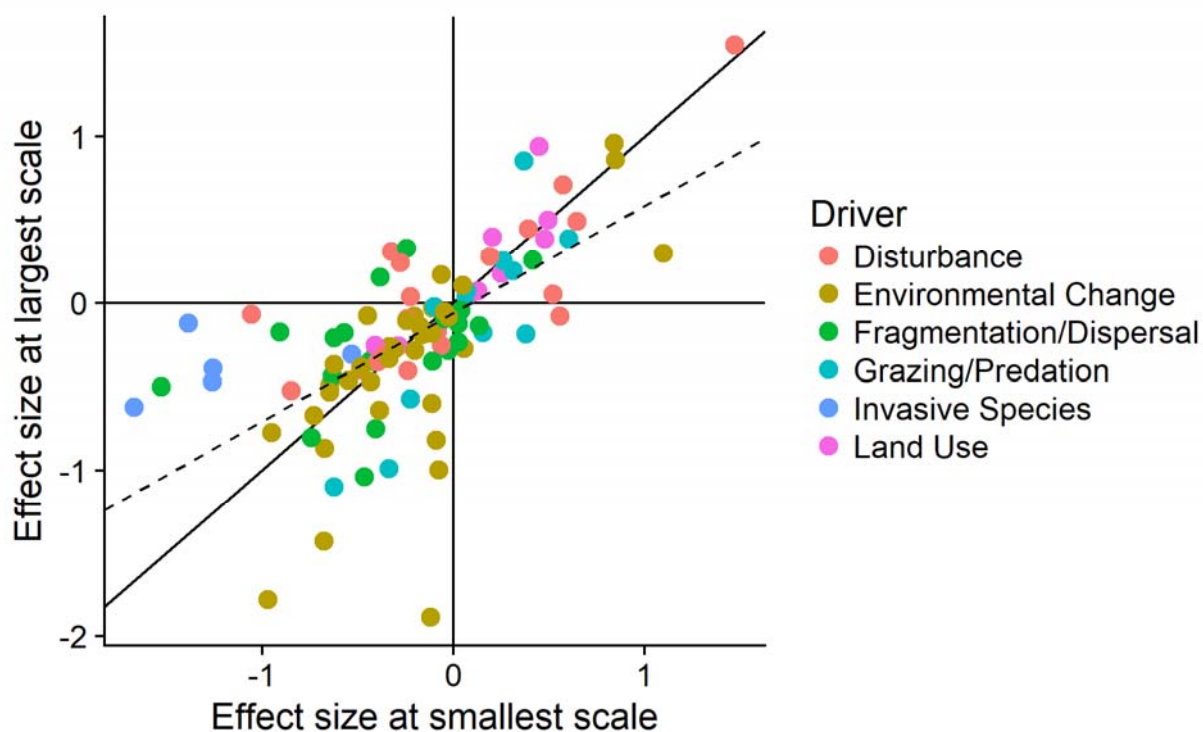
2   **Supplementary Information**
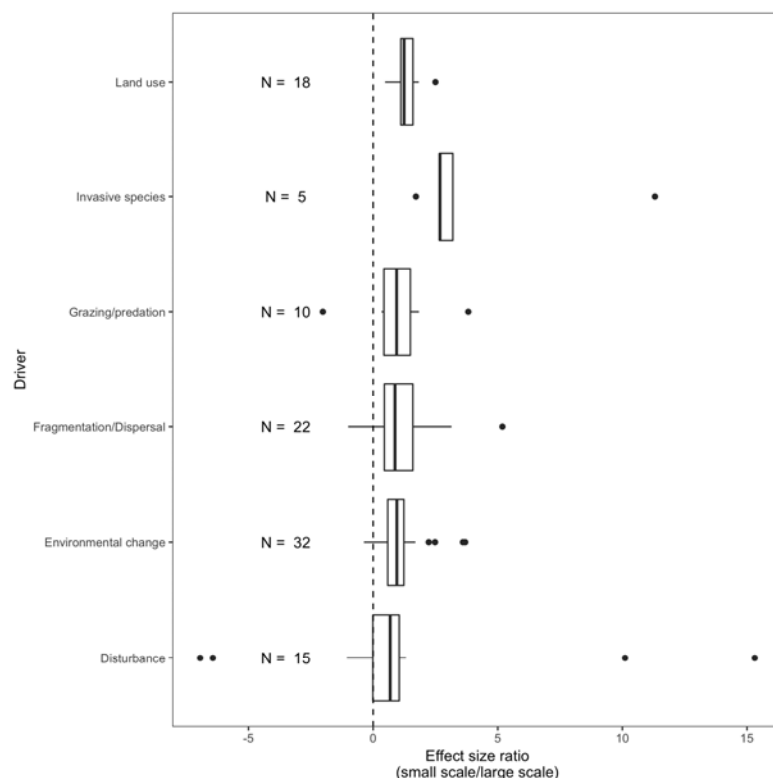
3

4

5

6   **Figure S1**. Results showing the ratio of log-response ratio effect sizes from experiments where

7   species richness responses were measured at two spatial scales (small scale/large scale). The

8   dashed line at 0 would indicate studies where the effect sizes were the same at the smaller and

9   larger scale. For each category of ecological driver, the means are above 1, indicating that the

10   measured effect size is larger at the smaller relative to larger size, and this difference is

11   statistically significant for land use, invasive species, and grazing/predation.

12

1   Appendix 1. Raw data and references for rarefaction curves and analyses presented in Figures 2

2   and 3. Subset of studies presented in McGill 2011b.

3   File available here:

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

5   (will be deposited in appropriate repository on acceptance)

6

7   Appendix 2. Effect sizes, metadata, and references for studies used in the scale-dependent meta-

8   analysis presented in Figure 4.

9

10   File available here:

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

12   (will be deposited in appropriate repository on acceptance)

13

14   Appendix 3. Raw data for the analyses of invertebrate and amphibian communities in nutrient

15   addition and reference treatments presented in Figure 6. From Chase (2010).

16   File available here:

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

18   (will be deposited in appropriate repository on acceptance)

19

20

21