Oikos

OIK-05968

Chase, J. M., McGill, B. J., Thompson, P. L., Antão, L. H., Bates, A. E., Blowes, S. A., Dornelas, M., Gonzalez, A., Magurran, A. E., Supp, S. R., Winter, M., Bjorkman, A. D., Bruelheide, H., Byrnes, J. E. K., Sarmento Cabral, J., Elahi, R., Gomez, C., Guzman, H. M., Isbell, F., Myers-Smith, I. H., Jones, H. P., Hines, J., Vellend, M., Waldock, C. and O'Connor, M. 2019. Species richness change across spatial scales. – Oikos doi: 10.1111/oik.05968

Appendix 1

Methods and results for representative simulation models

Methods for non-equilibrium model

First, we used a non-equilibrium model similar to that used by Cassey et al. (2006) to address similar questions. Such a non-equilibrium scenario may be a reasonable assumption for many systems over the past several hundred years as anthropogenic pressures continue to increase through time. Specifically, we modelled the effects of immigration and population losses in a simulated, spatially structured metacommunity to explore the scale-dependence of species richness change through time. In our simulation, we assumed that changes in species richness are only due to probabilistic local population loss and immigration events from the regional species pool. Therefore, the simulation serves as a null expectation for how biodiversity change should depend on spatial scale in the absence of speciation, trait differences and/or habitat heterogeneity.

We simulated a global metacommunity consisting of 1000 local communities, nested equally within three larger scales of communities at four spatial scales. Therefore, we have 1 global metacommunity consisting of 10 regions, each consisting of 10 sub regions, that each consist of 10 local communities. We initially seeded this global metacommunity from a pool of 1000 species (Figure A1) such that:

1. The species pools for each of the 10 regions were drawn at random from the global pool with a 0.3 probability of selecting each species.
2. The species pools for each of the 100 sub regions were drawn from drawn at random from the species pools in region in which they were nested, with a 0.3 probability of selecting each species.

3. The species pools for each of the 1000 local communities were drawn from drawn at random from the species pools in sub region in which they were nested, with a 0.3 probability of selecting each species.

After seeding the metacommunity, we removed all species from each nested community that were not present in at least one of the component local communities. This resulted in only a subset of the 1000 species being present in the metacommunity. This initial seeding of the metacommunity served as our pre-change condition, to which changes in biodiversity that resulted from colonization and extirpation events are compared.

Following the initialization of the metacommunity we simulated biodiversity change over 100 time steps. During each time step, in each local community:

1. Species *i* colonizes the focal community with a probability:

$p_{ic} = 0.01\left(\frac{A_{is}+A_{ir}k+A_{ig}m}{\sum_{j=1}^{S} A_{js}+A_{jr}k+A_{jg}m}\right)$, where A_{is} is the number of communities in the same sub region occupied by species *i*, A_{ir} is the number of communities in other subregions within the same region occupied by species *i*, and A_{ig} is is the number of communities in other regions occupied by species *i*. A_{js}, A_{jr} *and* A_{jg} are represent these same values but for each other species *j*. *k* and *m* are scaling parameters that determine the degree to which dispersal is constrained within nested subsets of the metacommunity.
2. Species *i* is lost from the focal community with a probability $p_{il} = p_l \times ext_i$, where p_l is the local extirpation frequency, and ext_i is the extirpation probability for species *i*. Values of *p* were drawn at random from a beta distribution ($\alpha = 2$, $\beta = 5$), such that some species were more prone to extirpation than others. This parameter ext_i was the only non-neutral component of the model.

Species were only lost from large spatial scales when they were lost from all nested communities comprising that region. We considered three patterns of dispersal (global: k = 1, m = 1; intermediate: k = 0.1, m = 0.01; localized: k = 0.01, m = 0.001) crossed with three scenarios of local extirpation frequency (rare - $p_l = 0.01$; intermediate - $p_l = 0.03$; frequent - $p_l = 0.1$). In all scenarios, the colonization probability p_c was set to 0.001. In each simulation, we calculated species richness change from the initial conditions for all communities at all four spatial scales using the log response ratio.

Methods for equilibrium model

To contrast with the non-equilibrium scenario above, we used an individual-based neutral model, following Hubbell (2001), to simulate a metacommunity in quasi-equilibrium. This quasi-

equilibrium is achieved because probabilistic extinctions are slow and are balanced out by rare speciation events.

We simulated a global metacommunity consisting of three regions, each containing 10 subregions, each comprised of 10 local communities. The local communities each contain *J* individuals. In each time step, *D* individuals die in each local community, chosen at random. Then:

1. With a probability (*1-m-v*) replace the individual with a copy of another drawn from the remaining local community
2. With a probability (1-*v*) replace the individual with an immigrant drawn from the metacommunity, weighted by nested structure of the metacommunity such that:
    a. *1-m* is the probability that the parent of the immigrant is from another community in the same subregion.
    b. $m-m^2$ is the probability that the parent of the immigrant is from the same region, but not the same subregion.
    c. m^2 is the probability that the parent of the immigrant is a different region of the metacommunity.
3. With a probability (*v*) a new species is born

We started the simulation by assuming that all individuals in a local community are the same species and the identity of these species differs for each local community. We started the simulations with the following parameters: $J = 50$, $D = 5$, $m = 0.1$ and $v = 0.005$. We then ran the model for 1000 time steps, which is sufficient to allow biodiversity at all scales in the metacommunity to reach equilibrium (no long term directional changes in biodiversity). We then simulated some change to influence biodiversity change by changing *J*, *m* or both and then ran the model for an additional 1000 time steps to allow biodiversity to reach a new equilibrium.
We explored the following scenarios in a factorial contrast:

1. a reduction or increase in carrying capacity – reduced or increased *J* by 10
2. an increase or decrease in long distance dispersal – increased *m* by 0.2, or decreased *m* by 0.05

Again, one could exhaustively explore other parameter values, but we felt that this was sufficient to produce a range of biodiversity change patterns that differ in how they depend on scale. We then estimated biodiversity change for each local community, subregion, region, and the entire global metacommunity as the mean richness of the final 100 time steps before and after the parameters were changed using the log response ratio of pre-change and post-change equilibrium species richness as above.

Appendix 2

Analyses for case studies

For each of the case studies presented above, we nested smaller scale surveys (in the case of corals and breeding birds) or regional checklists (in the case of Hawaiian birds and European plants) into larger aggregate surveys or checklists. This allowed us to compare species richness change at multiple nested spatial scales; change was measured yearly for the surveys, and as a log-ratio comparing species richness at time 1 and time 2 for the checklists. Because there are more observations at smaller scales than at larger scales (which have several smaller-scale estimates nested within), simple OLS regression of species richness change regressed against scale may be considered an overly liberal test of the significance of the slope, and thus we chose not to present p-values in the text, as the quantile regression fits illustrate the patterns in the data we wish to highlight.

Nevertheless, analyses in the macroecological literature often use regression-based approaches on data with similar nested structures (e.g. hundreds to thousands of studies on nested species-area relationships), and thus we used simple regressions simply to illustrate the scale-dependence of the relationships. Corals showed a significant slope negative slope and quadratic term ($p < 0.01$). The quadratic model was not significant for the North American birds ($p = 0.25$) but a linear model showed a significant positive slope with scale ($p < 0.01$). The Hawaiian birds showed a significant negative slope ($p = 0.02$) and a marginally significant quadratic term ($p = 0.09$). The European plants did not show a significant relationship with scale. When we used a more conservative approach, taking only the means at each scale (and circumventing the nesetedness problem) and fitting a linear model, the relationship was still significant for the North American birds ($p < 0.01$); however, because there were only 5 scales analyzed in the coral case study, we had less confidence that the slope was negative ($p = 0.18$). We did not fit this conservative approach to the Hawaiian Birds because this would leave us with two few scales for appropriate analysis. When we averaged the LRS and area at the country, region, and subregion scales, the European plants showed a marginally significant negative slope (–0.03) with scale ($p < 0.08$).

Table A1. Regional classification of European countries used in the European plant case study.

Country	Subregion	Region
Austria	Central	Western
Baltic States	Baltic	Northern

Belgium	Central	Western
Bulgaria	East South	Eastern
Czech Republic	East North	Eastern
Denmark	Scandinavia	Northern
Finland	Scandinavia	Northern
France	Central	Western
Germany	Central	Western
Greece	Mediterranean	Southern
Hungary	East South	Eastern
Iceland	Scandinavia	Northern
Ireland	British Isles	Western
Italy	Mediterranean	Southern
Netherlands	Central	Western
Norway	Scandinavia	Northern
Poland	East North	Eastern
Portugal	Iberia	Southern
Romania	East South	Eastern
Spain	Iberia	Southern
Sweden	Scandinavia	Northern
Switzerland	Central	Western

United Kingdom	British Isles	Western

Appendix 3

Methods for building cross-scale dataset

In all, we present the log ratio richness change from 1429 data sources collated from a number of different sources (see below). All data and sources are available at https Figshare link.

Local scale richness change estimates

We developed a synthetic dataset to examine species richness change through time across taxa and spatial scales from local to the nearly global scales. The dataset consisted of two qualitatively distinct types of data. First, we used data from relatively local-scale ecological surveys that were taken over 2 or more time periods, typically on the order of decades. This includes data included in Vellend et al. (2017b) on plants, Dornelas et al. (2014) and the expanded BioTIME database (Dornelas et al. 2018) on a number of different taxa, and Elahi et al. (2015) on marine organisms. Although these datasets were heterogeneous in their focus, habitat, taxa, temporal scale, and measurements (i.e., species richness versus composition), they all share in common quantification of species richness for at least two time periods with consistent methodology.

From these datasets, we extracted time-series that met the following criteria: 1) sampling methods were consistent over time, and this was assessed based on information available in the original studies, 2) the sampling area (grain size of sampled units, m^2) was clearly reported as well as number of samples, and 3) the sampling location had to be consistent over the duration of the time-series. For the purposes of this study, we only used the first and last estimates of species change to calculate the log ratio difference in species richness (LRS). Thus, for each study, we had a single estimate of species richness change.

We considered the total sampled area reported across all replicates within that time-series to assign a spatial scale (area) to each diversity estimate. When multiple estimates were reported within a single time and region, we aggregated the sampled areas by adding the area and the species lists to estimate a (pooled) species richness per (aggregated) sampled area. When estimates were based on reported sample based rarefaction estimates, the total sampled area was calculated as the size of each sampled area multiplied by number of samples used for the rarefaction. In no case did we estimate species richness for unsampled areas. Combining sampled area and diversity data from different sampling locations that were not adjacent to each other in this way introduced uncertainty associated with how species are aggregated in the landscape (Azaele et al. 2016). This method of pooling areas might overestimate diversity by ignoring spatial beta diversity for cases where species are distributed non-randomly across a site. However, this uncertainty was preferable to uncertainty associated with the alternative approach of extrapolating observed species diversity to unobserved areas.

Larger-scale checklist data

We augmented the above data with additional estimates of change from a qualitatively distinct kind of data to expand the range of taxa and spatial scales in our analyses. This came from a literature review of species richness estimates (or lists of species presences/absences) and its change over (typically) long time periods (decades to centuries). We identified studies by searching the literature (Google scholar, ISI) using search terms including, but not limited to, "resurvey", "historic", "extinction", "alien", "invasion", "richness", "change", and "turnover". In addition, we located data from existing species inventory databases (e.g., state, province or national check- and red lists, protected area checklists, IUCN redlists, NatureServe, USDA plants) where information on historical and current richness is available, as well as extinctions and colonizations. Finally, to fill biogeographic, taxonomic, and spatiotemporal gaps, we added datasets opportunistically and conducted more targeted searches (i.e. located via personal networks, gray literature, or using more specific search terms to target specific groups).

For each dataset, we recorded the taxon, location, total spatial grain and extent sampled based on what was reported in the original study (km^2) (or taken from other sources, such as for the size of geopolitical units (e.g. area of country) or biogeographic units (e.g. area of islands or continents), and the years in which each sample was taken. For many datasets, year was estimated between long time periods (e.g. before human colonization or expansion). We recorded species richness at each point in time (or in several cases, estimated this from numbers of extant, extinct and nonnative species), and calculated LRS as above.

Appendix 4

Analyses of cross-scale data

As noted in the main text, this analysis was done in the spirit of an exploratory analysis because we *a priori* would not expect any single model structure to fit all of the data, and instead, we expected that different systems would exhibit different kinds of scale-dependence. We report p-values and AIC values only as a courtesy. Several irreconcilable statistical issues as well as a lack of clear a priori hypotheses prevented a full hypothesis testing analysis. For each taxon group analyzed such as terrestrial birds or marine fish, we analyzed LRS ($\ln(S_{last}/S_{first})$) versus scale (area in m^2). This gives a common measure of change across datasets with a roughly normal distribution. Within each taxonomic group (where more than 40 data points were available), we ran several regression models:

A. Intercept only – calculated the mean and standard deviation and a test to see if the mean LRS was significantly different from zero using a t-test;

B. Linear OLS – a simple linear regression of the form LRS~$\log_{10}$(scale). We report only the coefficient for scale (determine if LRS changed with scale) and the p-value from the t-test for this coefficient;

C. Quadratic OLS – of the form LRS~$\log_{10}$(scale)+($\log_{10}$(scale)-mean($\log_{10}$scale))^2;

D. Quadratic OLS with Study type - of the form LRS~$\log_{10}$(scale)+($\log_{10}$(scale)-mean($\log_{10}$scale))^2+StudyType (where study type was a binary dummy variable coded as 0 for lists and 1 for surveys);

E. Quadratic GLS – because variance was non-constant, a weighted GLS model (with MLE) was used. The variances appeared to vary in a non-monotonic fashion across scale (Figure 2) so there was no obvious appropriate error model, but we used var~exp(k*$\log_{10}$(scale)) with the fixed terms as in model (C).

We report both the estimated coefficients and the p-values of individual coefficients based on their t-tests (Table A2). We compared the AIC of model (C) with (E) to see which variance model was best. We have not done a full model selection, nor a full p-value analysis (e.g. comparing quadratic to linear or adjusting for multiple tests) because they would be misleading and overreach the nature of these highly heterogeneous data. Additionally p-values are of limited meaning when there are large numbers of data points, but low explanation of the variance by models as in our data. We therefore report p-values (and some AIC values) but we limit the strengths of our interpretations accordingly.

Table A2 presents the results from our analyses of all of the data collected. However, we caution that analyses of some of the taxonomic groups are not likely very meaningful. For example,

several taxa had few observations (e.g. marine plants, terrestrial invertebrates, freshwater fish), were difficult to compare/calculate area (e.g. freshwater fish from a single lake versus entire regions with only a small percentage of land area covered by freshwater), or had few observations at large scales that typically involved checklists (e.g. marine invertebrates). Thus, even though we present all of our analyses in Table A2, Figure 2 only illustrates the results for the four taxonomic groupings that we were most confident in the data being representative to allow a reasonably comparative approach (note, marine fishes also had few observations from checklists at larger scales, but we included them as a marine example).

Note the following three points argue against overly strong interpretations of these data: 1) There are large numbers of data points; even relationships with low explanatory power (R^2 values often around 0.01) were still statistically significant. 2) The variance is clearly not homoscedastic with spatial scale, but has a complex structure. In Figure 2, the variation often appears hourglass shaped (two bulges), probably driven by the data sampling intensities at different scales. This makes choice of a model for the variance structure difficult and does not conform to any of the standard models (e.g. power law or exponential function of x-variable). 3) The factors of potential interest are highly confounded (e.g. spatial and temporal scales) (Figure A1): a) the Pearson correlation between log-transformed spatial scale and temporal duration of the time-series ($r = 0.756$; $p < 2e\text{-}16$); b) Welch's two-sample t-tests for log-transformed scale separated into data collection methods (surveys versus checklists) was also highly significant (means of $10^{-4.36}$ versus $10^{2.37}$ respectively); and c) a similar t-test for log-transformation of duration of the time-series versus study type was also significant ($10^{1.44}$ versus $10^{2.37}$ years). Models (C) and (D) were qualitatively identical (quadratic terms were significant or not significant the same in both models) so only model (D) is reported in Table A2. Similarly model (A) is a better estimate of the mean height of the data so intercepts are not reported in the other models.

Recognizing these qualifications, our analyses summarized in Table A2 indicate several scale-dependent responses, often at least roughly consistent with the non-linear expectations presented in the main text. However, there is also a great amount of variation in LRS within and among taxonomic groups and scales. Overall, we found that the linear model (B) usually trended for LRS to increase with scale but only terrestrial plants were significant and terrestrial birds were borderline significant. In the quadratic model (D) when both spatial scale and study type (survey versus list) were included in a multiple regression model, the study type was never close to significant, arguing that study type was not causing artefacts and that the differences could be attributed to scale (although spatial and temporal scale remain correlated). The datasets with fewer data points did not show a significant quadratic term, as might be expected given the power to detect such patterns decreases with data availability. The GLS model (E) was similar but the quadratic term in terrestrial

plants was no longer significant and freshwater fish became significant. The best variance model (homoscedasticity of OLS or exponential heteroscedasticity of GLS) varied widely between data sets, and in most cases, the difference in fit between the two was small. We interpret this and an examination of the visual plots as indicating neither OLS nor exponential GLS is a strong model of variance structure.

Table A2. Summary of statistical fits of different models for the cross-scale data.

	Data statistics		Model A — Intercept only			Model B — Linear OLS			
						Intercept		Slope vs scale	
	No. obs.	% Survey	Mean	SD	p	Coef	p	Coef	p
Terrestrial plants	818	58.1%	0.073	0.309	<0.0001	0.0908	<0.0001	0.0049	0.0894
Terrestrial birds	288	3.1%	0.073	0.222	<0.0001	0.0491	0.0394	0.0246	0.0025
Terrestrial mammals	173	6.4%	0.129	0.469	0.0004	0.0859	0.4386	-0.0233	0.6419
Marine invertebrates	134	99.3%	0.098	0.497	0.0237	-0.2942	0.1624	-0.0637	0.0597
Marine fish	101	98.0%	0.028	0.435	0.5243	0.0924	0.5073	0.0125	0.7434
Terrestrial invertebrates	69	8.7%	-0.016	0.327	0.6795	-0.0959	0.1964	0.0075	0.6719
Freshwater fish	49	4.1%	0.167	0.120	<0.0001	-0.0404	0.5954	0.0428	0.0095
Marine plants	19	-	-	-	-	-	-	-	

Table A2. Continued

	Model D — Quadratic OLS					
	Slope vs Scale		$Scale^2$		Study type	
	Coef.	p	Coef	p	Coef	p
Terrestrial plants	0.0143	0.0571	-0.0028	0.0072	0.0500	0.4411
Terrestrial birds	0.0236	0.0064	-0.0125	0.0019	-0.0256	0.7634
Terrestrial mammals	-0.0202	0.7101	0.0049	0.7745	-0.0228	0.8865
Marine invertebrates	-0.0392	0.2539	-0.0094	0.5122	-0.2951	0.6101
Marine fish	0.0015	0.9664	-0.0151	0.1144	-0.4215	0.2708
Terrestrial invertebrates	-0.0006	0.9724	0.0047	0.5693	0.1434	0.4883
Freshwater fish	0.0428	0.0418	0.0053	0.0816	-0.0927	0.4101
Marine plants	-	-	-	-	-	

Table A2. Continued

	Model E — Quadratic variance GLS			
	Slope vs Scale		$Scale^2$	
	Coef.	p	Coef	p
Terrestrial plants	0.0049	0.0894	-0.0006	0.4161
Terrestrial birds	0.0246	0.0025	-0.0133	0.0003
Terrestrial mammals	-0.0233	0.6419	0.0058	0.7140
Marine invertebrates	-0.0637	0.0597	0.0056	0.5370
Marine fish	0.0125	0.7434	-0.0116	0.1346
Terrestrial invertebrates	0.0075	0.6719	0.0039	0.4518
Freshwater fish	0.0438	0.0095	0.0045	0.0240
Marine plants	-	-	-	-

Table A2. Continued

	Model AIC scores (Model C vs E)	
Terrestrial plants	391.4532	137.7995
Terrestrial birds	-63.7176	-62.2645
Terrestrial mammals	235.3891	237.3873
Marine invertebrates	196.6703	192.5339
Marine fish	123.6297	122.4970
Terrestrial invertebrates	45.5948	-71.9930
Freshwater fish	-66.6704	-66.5679
Marine plants	-	-

Table A2. – Statistical results of different models on different taxa groups. Taxa are sorted by the number of time-series available (column TS). %Survey indicates what fraction of the time-series were from surveys (e.g. BioTIME data) with the remainder being checklist type of data. Model A shows the mean LSR and a p-value for whether this mean is statistically significant from zero. Model B is linear and the slope coefficient and p-value for a test of a slope different from zero are shown. AIC scores compares model C (OLS quadratic) with model E (GLS quadratic) to see what error model fits best. Model D shows the coefficient (slope) and p-value for the linear, quadratic and study-type variables. A significant linear model indicates scale-dependence. A significant quadratic model indicates non-linear scale-dependence. Model E shows the slope and quadratic terms.

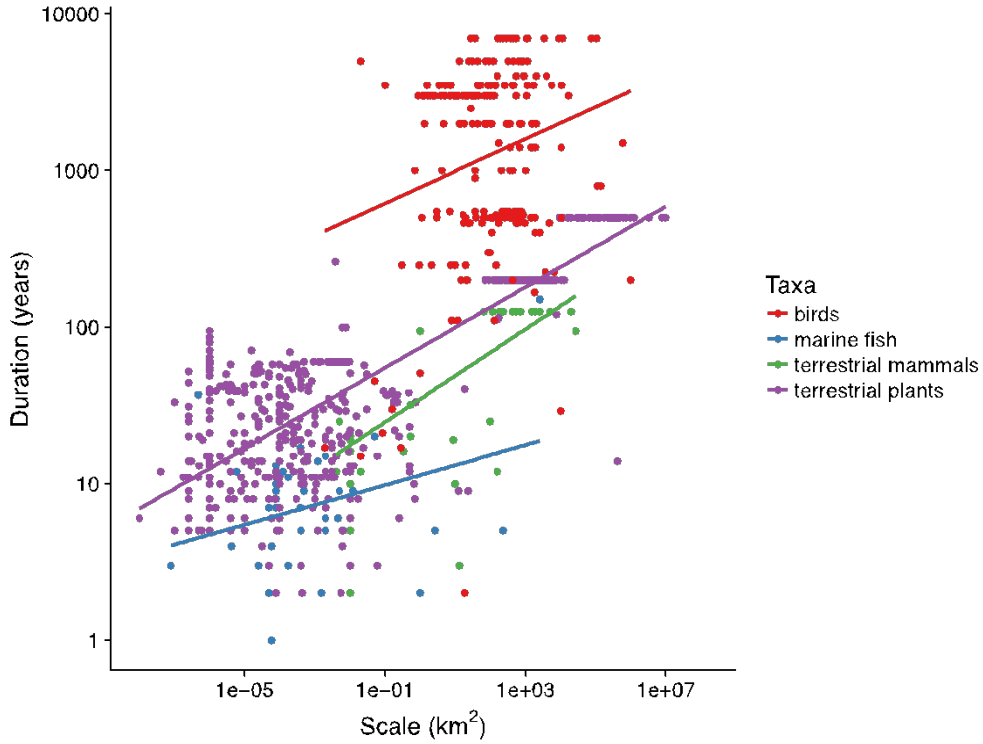

Figure A1. Illustration of the potential for covariance between the spatial scale of the study and the temporal duration of the study. Our main goal with this figure is simply to show that while the patterns from this analysis help us to understand that different perspectives on species richness change, the complexity of error structures and co-variance in the data suggest that any inferential statistics are cautionary.