**Appendix S1**

*Supporting information for section 1:* Existing time-series of biodiversity monitoring programs are a spatially biased representation of Earth’s diversity, as well as the anthropogenic drivers that cause diversity change.

The datasets collated by Vellend et al. (2013) and Dornelas et al. (2014) included a large number of observations of biodiversity taken from many locations around the globe. However, the studies represent an opportunistic collection of sites that lack comparability. So our question is whether the studies collated for their datasets were sufficiently spatially representative of Earth’s biodiversity, as well as human impacts that influence biodiversity, to justify their extrapolation of findings about local diversity change to the global scale.

To assess how well the datasets assembled by Vellend et al. and Dornelas et al. capture a globally representative sample of species diversity, we performed an analysis of spatial representativeness, following protocols established by Schmill et al. (2014) for the GLOBE Global Collaboration Engine (see work flow diagram in Fig. S1, below). We compared the sampling locations represented in each data synthesis to existing, well-resolved global maps of species richness, as well as existing global maps that serve as proxies of human influence on terrestrial and marine ecosystems: Global Forest Cover Change, Crop and pasture density, Human Influence Index and Marine Impacts maps (citations and comparisons are summarized in Table S1). All steps in the spatial data analysis are outlined in Fig. S1.

The first step was to transform all of the maps to categorical data. The species richness and human impacts maps (as well as croplands and urbanization as in Fig. S2) were reclassified by binning the data into quartiles. For the map of Forest Cover Change, we kept the original categories of ‘Loss’, ‘Gain’, and ‘Loss + Gain’.

The second step was to generate random points over each map, and then extract data from the underlying pixels to those points. For each of the six maps used in our analysis (Table S1), we randomly sampled *N* locations, where *N* was equal to the 346 studies collated for terrestrial data layers used in Vellend et al. comparisons, or 10,000 sampling locations for the marine data layers used in the Dornelas et al. comparisons. We repeated this sampling 1,000 times for each dataset, in order to generate a distribution of values. Although Dornelas et al. had >400,000 sampling locations represented in their dataset, simulations showed that 10,000 randomly chosen points of the marine data layers was sufficient to capture the true mean and standard deviation of global maps, while allowing for similar replication (*n*=1,000) for generating a distribution of values (Table S2).

We utilized the Google Earth Engine to extract initial data from the Hansen et al (2014) Forest Cover Change dataset. Because the FCC dataset is at a much finer resolution than the other maps (30m), we compensated for potential inaccuracies in the reported latitude and longitude in Vellend et al. points by creating a 200-m buffer around each study site and summing the total number of pixels of each class covered by the buffers. We compared those totals to the average number of pixels per category from the random points. These results were exported from Google Earth Engine to R for further analysis. We present results from the three change categories (gain, loss, loss + gain; Fig. 2). Pixels categorized as No Change represented both any forest pixels that remained forest as well as any non-forest pixels that did not convert to forest during the study period. Buffered points (pixels) covering non-forest areas were not masked from the analysis, but this should not affect our conclusions because comparisons of observed/expected for the change categories are based on the proportional global area covered by those categories alone.

For each sample of *N* locations, we quantified the amount of spatial bias in those locations using Hellinger’s distance *d,* through the distrEx package in R. Hellinger’s distance quantifies similarities between the distribution of values captured by the set of points and a uniform distribution, with equal probability across all four quartile bins.  The Hellinger distance is defined as:

*d= 0.5 \int |√{dP}-√{dQ}|^2*

where *P* is the uniform distribution and *Q* is the distribution of values captured by the sample or random points. To be clear, the density of values for either randomly cast points or the sample locations are each compared to the uniform distribution scenario, not to each other, to calculate *d.* Then these respective values of *d* can be compared to assess relative biases, with a *d=*0 as the baseline of zero bias.

As *N* gets increasingly large and approaches the sampling of every location *L* on a map, there is no spatial bias in a sample and *d* becomes zero, because the map has been exhaustively sampled. However, because *N* is always less than *L*, even a random sampling of *N* points will have some spatial bias associated with that sample. By performing this random sampling effort 1000 times for each map, we generated expected distributions of *d* (as shown in Fig. 1) that represent the amount of spatial bias one would expect to occur for a sampled composed of *N* studies, but where those studies were chosen in an unbiased manner, and thus, were spatially representative of species richness or human impact across space. We generate one value of *d* each for the Vellend and Dornelas sets of sample locations. A spatially representative sample of Earth’s diversity, or of human impacts on Earth’s ecosystems, should fall within the expected distributions of *d* generated from the randomly cast points. But in fact, the collection of studies collated for the Vellend et al. and Dornelas et al. syntheses lied a minimum of 8, and a maximum of 41 standard deviations outside of the mean of randomly generated distributions of Hellinger’s *d* (see Table S1, Fig. 1, Fig. S1). This indicates that both data syntheses were based on sampling locations that exhibited rather extreme spatial bias with respect to spatial trends in species richness and human impacts.

We then extended these analyses to assess how this measured bias was affected by potential over- or under-sampling across the categories using the ln(observed/expected). We calculated expected number of cases (points) per category, adjusted for the proportional total land (marine) area covered by that category.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table S1. Summary of spatial comparisons. On the left is the dataset (Vellend et al. or Dornelas et al.) and global map that are being compared to one another. Random sampling of points from the global map was used to generate expected values of Hellinger's distance *d (*mean  SD given at right). The observed value of d for the dataset is then shown, along with the deviation from expected value, reported in units of standard deviations. | | | | |
|  |  | **Hellinger's distance, *d*** | | |
| **Dataset being compared** | **Global map being compared to** | Expected ( SD) | Observed | Deviation (in SD's) |
| Vellend et al. (2013) | 1. Terrestrial species richness (Fig. 3d, Kreft & Jetz 2007) | 0.077 (0.018) | 0.221 | 8.000 |
|  | 2. Global forest cover change (Hansen et al. 2011) | 0.022 (0.009) | 0.127 | 11.667 |
|  | 3. Cropland & pasture density (HYDE 3.1 human-induced land-use change, Goldewijk et al. 2011) | 0.030 (0.012) | 0.252 | 18.500 |
|  | 4. Global human influence index of land use & infra-structure (WCS 2005) | 0.029 (0.012) | 0.363 | 27.833 |
|  |  |  |  |  |
| Dornelas et al. (2014) | 5. Global marine biodiversity (UNEP World Conservation Center’s map, Tittensor et al. 2010) | 0.078 (0.004) | 0.242 | 41.000 |
|  | 6. Human impacts on marine ecosystems  (Halpern et al. 2008) | 0.056 (0.005) | 0.214 | 31.600 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table S2. Comparison of how well data extracted to different sample sizes of randomly generated points, cast over rasters of species richness and marine impacts, are able to capture the true global distribution of values, as measured by Hellinger's Distance. One million random points were generated over the global area, and divided into different sample sizes to compare variation in representation across sample sizes. | | | | | | | |
| No. points, *N* | 10,000 |  | 100,000 |  | 400,000 |  |  |
| No. replicate draws | 1,000 |  | 100 |  | 25 |  |  |
| **Hellinger's distance** | **mean** | **sd** | **mean** | **sd** | **mean** | **sd** |  |
| Species Richness | 0.077 | 0.004 | 0.077 | 0.001 | 0.077 | 0.001 |  |
| Marine Impacts | 0.056 |  | 0.056 |  | 0.056 |  |  |

Figure S1. Workflow for analyses on spatial representation presented in section 1 of main text

Load raster map (e.g. Terr Plant Sp Richness)

Generate random points over raster area

Load points from collection (e.g. Vellend etal)

In R

In Google Earth Engine

Compare HD (d) values

Computer Hellinger’s Distance

Computer Hellinger’s Distance

Send to R

Distribution of values from random points

Distribution of values from collection points

Uniform distribution

Histograms of pixels per class for collection and random sampled

Extract data to collection points

Extract data to random sample points

Reclassify to Quartiles

Mask NAs

Load points from Vellend etal collection

Load Hansen FCC data

Create random points

Extract data from buffer areas

Create 200 m buffers

Create 200 m buffers

*This was done iteratively in sets of 1 million points for marine points, then combined, due to memory/computing issues*

Figure S2. Comparison of the Vellend et al. (2013) dataset to global maps of land converted to cropland/pasture habitat, or converted to human infrastructure similarly revealed biases ranging from 18 to 27 standard deviations from a representative sample

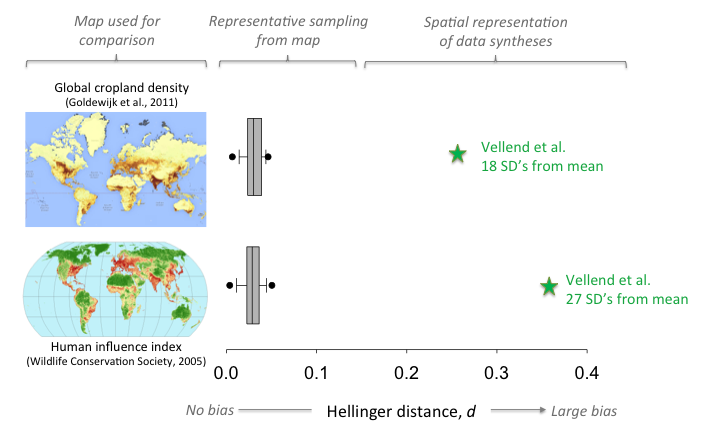


Figure S2 Citations/dois to access these calculations on GLOBE:

Allington, Ginger,"Global Cropland Density - Vellend Assessment". *GLOBE Representativeness Analysis*. Published September 11, 2015. http://globe.umbc.edu/app/#/analysis/rep/published/view/1844

Allington, Ginger,"Rep Assessment of Vellend v. Human Influence Index". *GLOBE Representativeness Analysis*. Published September 11, 2015. http://globe.umbc.edu/app/#/analysis/rep/published/view/1846

*Supporting information for section 2*: showing that estimates of biodiversity change are systematically biased when syntheses are based on datasets composed primarily of short time series. All code for the analyses described in the text and this Supplemental Material can be found at the following repository on github: <https://github.com/jebyrnes/Biodiversity>

To demonstrate how the duration of a time-series influences the estimate of biodiversity change, we simulated long time series of species richness and estimated effect sizes on short duration samples selected from the full series. We generated time series of species richness by iterating a discrete-time form (*St*+1=*S*t + *I*t – *E*t) of the equilibrium model of island biogeography (MacArthur and Wilson 1964, Fig. S3a and 3d) and altering the balance of extinction and colonization.

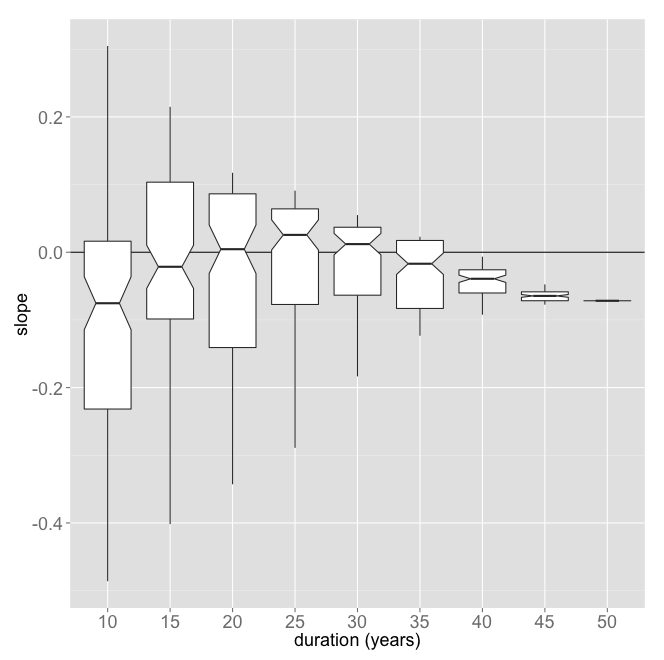
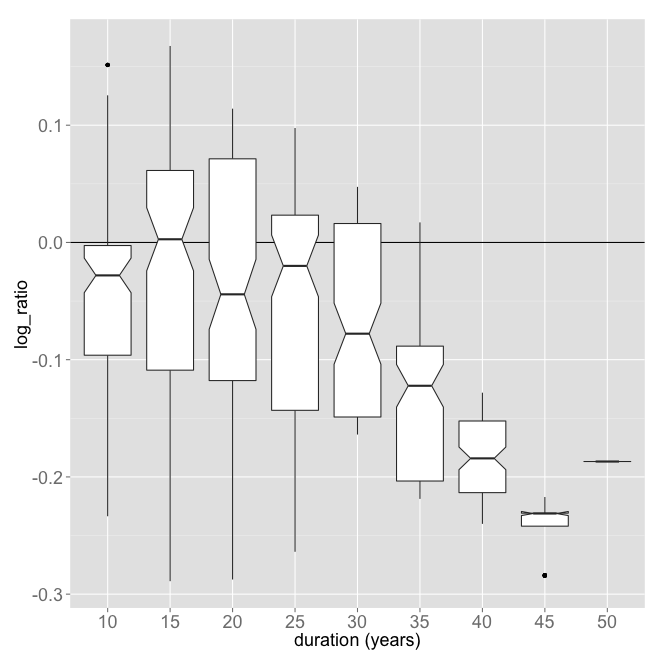
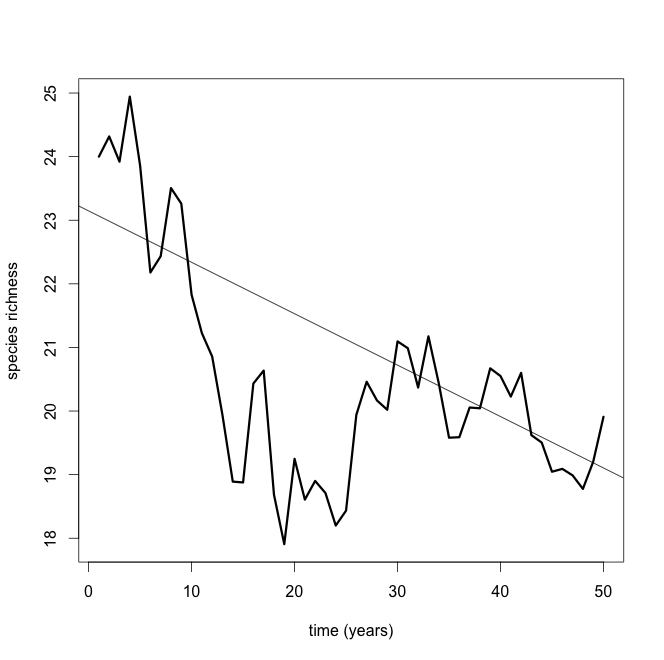
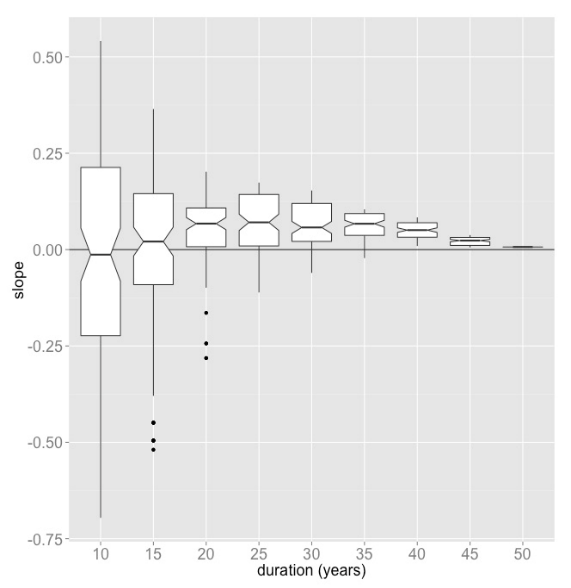
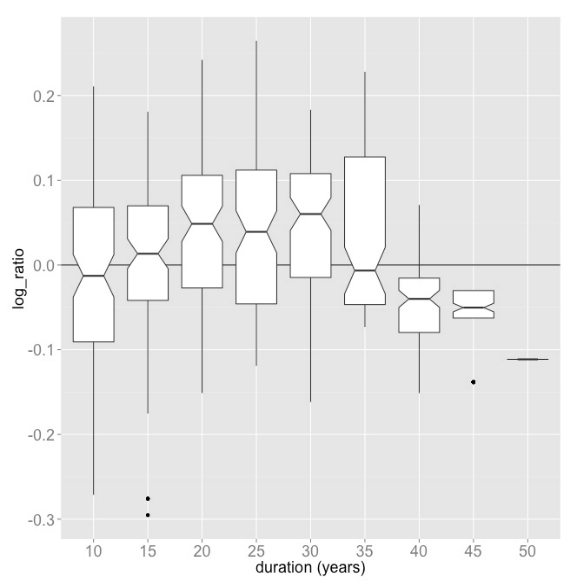
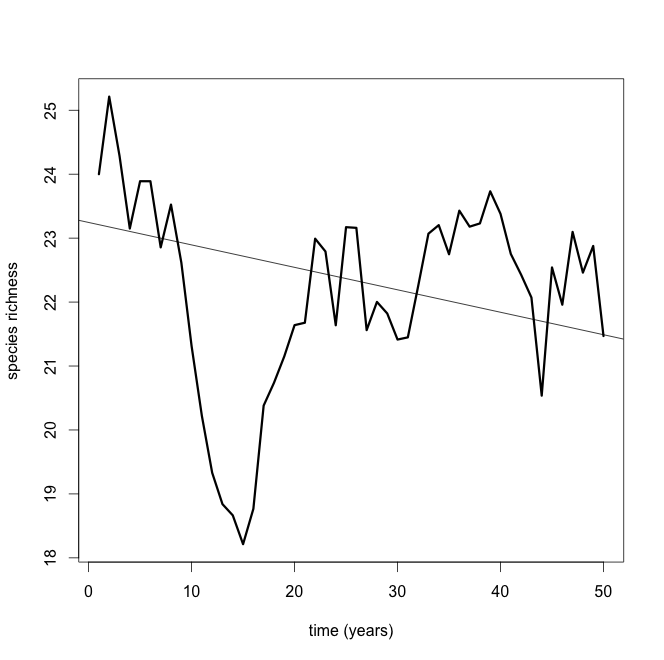
*A* is area, *q* is extinction probability (q ~ *N*(0.65, 0.07), *c* is colonization probability (c ~ *N*(0.55, 0.07), *P* is the number of species in the mainland pool, *D* is the distance of island from the mainland, and *m* is scaling factor for distance. Values for extinction and colonization probabilities were chosen to induce a 10% decline in species richness on average. We chose this as the most direct demonstration that short duration series will lead to biased estimates of net change in species richness. The results are similar for systematic increases in local species richness (not shown). The estimates of effect size in this figure were derived by calculating the log response ratio and the slope of the regression between species richness and time over 10 to 45 year intervals in increments of 5 years. Ten random replicate series were sampled for each interval from the series shown in Fig. S3a and S3b. Fig. S3 shows how the estimates of the negative trends for these two representative time series depend on the duration over which the effect size was calculated. In Fig. S3b and S3c we see that only the longest duration sample time series (>30 years) approach the true estimate of the log response ratio (-0.18) and the slope (-0.088). Analyses of sample-series shorter than 30 years produce substantial underestimates of the known trend in species richness and the median values are not significantly different from zero.

The example time series in Fig. S3d provides an informative case where the trend over the entire 50-yr period is negative, but the series is punctuated by a sharp loss and then a recovery (*t* = 10-20). This case is useful because it represents systems recovering from some historical disturbance such as clear-cutting (see Fig. 4), which we suggest are over-represented in Vellend et al. (2013). For this example, all estimates shorter than 40 years are biased (significantly positive) or are not distinguishable from zero (Fig. S3e). Even the estimate after 45 years, although significantly negative, underestimates the true value of biodiversity change. Fig. S3f shows a similar picture for the estimates of the slope. This systematic bias occurs because shorter duration samples do not capture the variance in the full time series. This arises because temporal autocorrelation in the time series will result in biased estimates and confidence intervals when obtained from short series (Bence 1995). Long periods of recovery to a historic baseline, capturing large variance in species richness, as shown in Fig. S3d, will dominate estimates from short duration samples of that series, leading to estimates of no net change or net gains in richness, even if the overall trend is negative. The take-home message of these simulations is that short duration samples of a longer time trend do not provide enough information to reliably estimate the overall trend if that longer-term trend is unevenly distributed across periods of loss and gain.

Given that short time series can provide unreliable estimates of a known trend, how well can a meta-analytic dataset composed of many short time series and just few long time series do at estimating a regional trend? We generated 100 independent times series of fluctuating species richness that were 50 years in length, to represent a large set of communities. We considered two cases 1) where colonization balanced extinction (Fig. S4a) and 2) where extinction rate was greater than colonization resulting in a net decline in species richness from 22 to 17 species on average (Fig. S5a). For each case we generated a 1000 shorter time series, 100 each of 10 to 50 years in duration in increments of 5 years from each of the 100 full-length series. Fig. S4b (case 1) and S5b (case 2) show the distribution of log response ratio effect sizes calculated over all 1000 time series. In Fig. S4b we see that short and long time series provide an unbiased estimate of the effect size when there is no trend in species richness. In Fig. S5b we see that only effect sizes calculated over the longest duration time series (>40 years) provide a reasonable estimate of the overall mean trend across the 100 time series. In Fig. S4b we again see that short and long time series provide an unbiased estimate of the mean effect size, but short time series (<20 years) show large uncertainty around the estimate. In Fig. S4d and e, we see that as expected the distribution of effect sizes is centered on zero for both metrics of biodiversity change. In Fig. S5d we see that the distribution of response ratios from all 1000 estimates does not have an expected effect size of -0.25, but rather a mean -0.12. Fig. S5c and S5e show the same output for the least squares estimates, revealing the need for long duration time series to provide a significant estimate of the average slope across many time series. This example shows that meta-analytic datasets composed predominantly of many short time series may not provide reliable estimates of average trends in biodiversity across many communities even if the true effect size is large and consistently negative across all communities. One way to detect an effect of the length of a time-series *a posteriori* is to use study duration as a predictor in a model testing for biodiversity change in the meta-analytic dataset (see main text).

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

Figure S3. (A) and (D) two 50-year stochastic time series of declining species richness from a simple island biogeography model of community dynamics; (B) box-plots showing variation in the log response ratios and (C) regression slopes, calculated over shorter time samples of the 50 year times series in panel (A) from 10 year to 50 intervals in increments of 5 years; (E) box-plots summarizing the variation in the log response ratio estimated from the time series in panel (D); (F) shows the variation in slope estimates as a function of duration using OLS linear regression estimates of the trend.



A

B

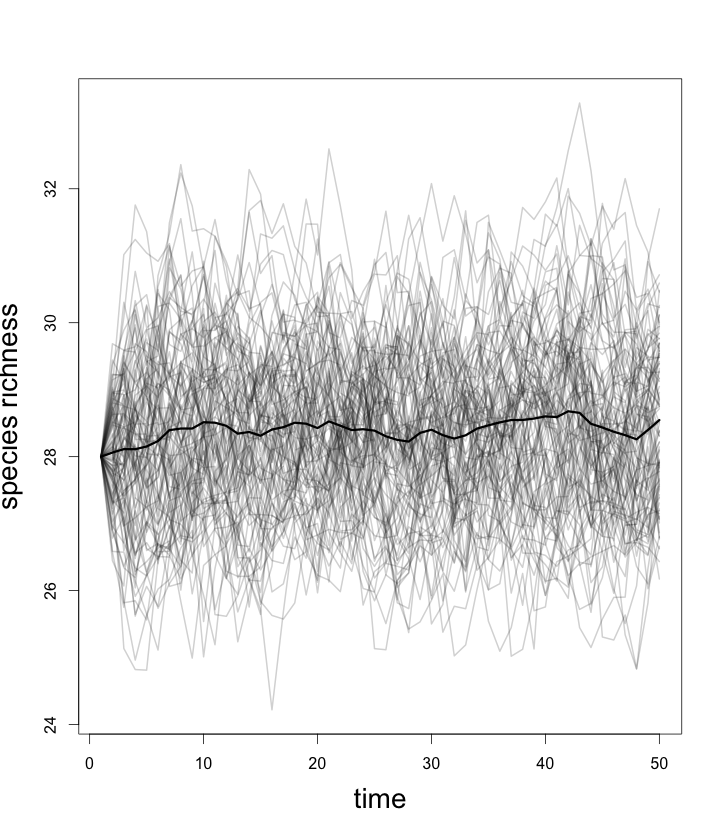
C

D

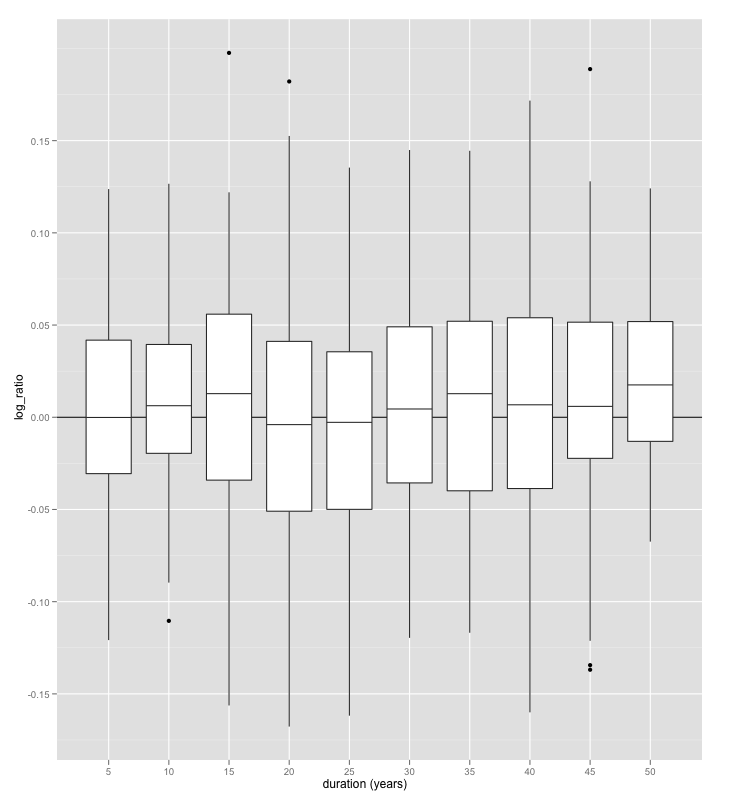
E

F

Figure S4. Panel (A) simulated dataset composed of 100 50-year time series of species number derived from an island biogeography model of community dynamics where colonization rate equals extinction rate; the black line shows the mean number of species over the 100 replicates. We created a meta-analytic data set by sampling 1000 shorter time series, 100 each of 10 to 50 years in length in increments of 5 years. Panel (B) shows the range of log response ratios and (C) OLS regression slope estimates derived from the dataset showing the effect of time series duration on the estimate. Panel (D) and (E) show the distributions of the log response ratios and slopes, respectively, estimated from all 1000 time series.



A



B

C

D

E

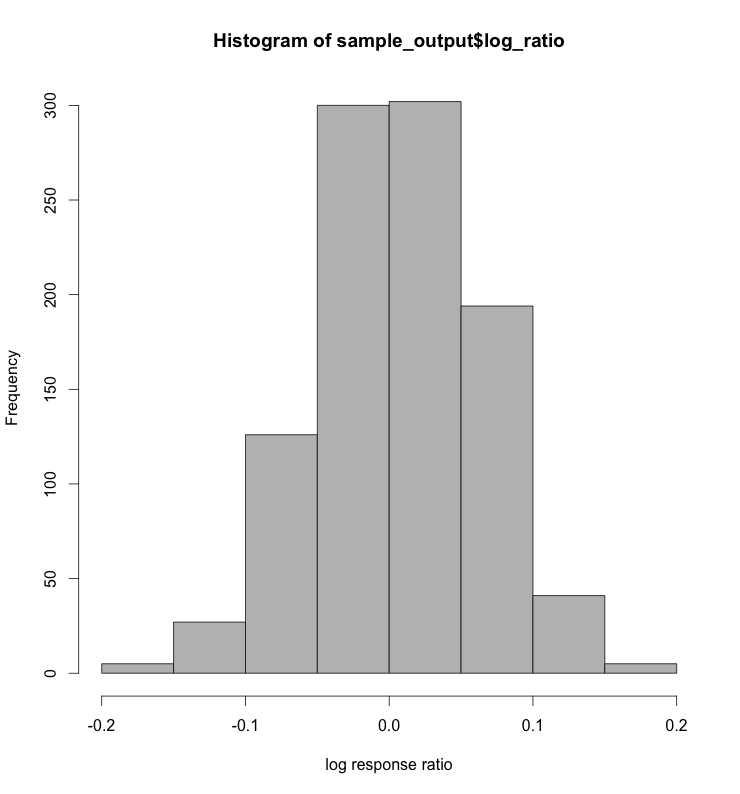
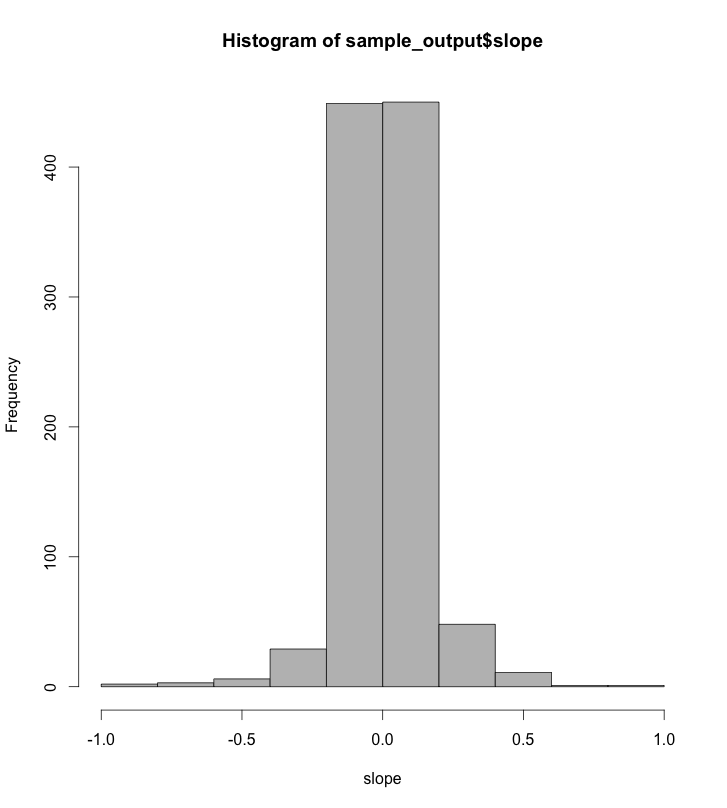
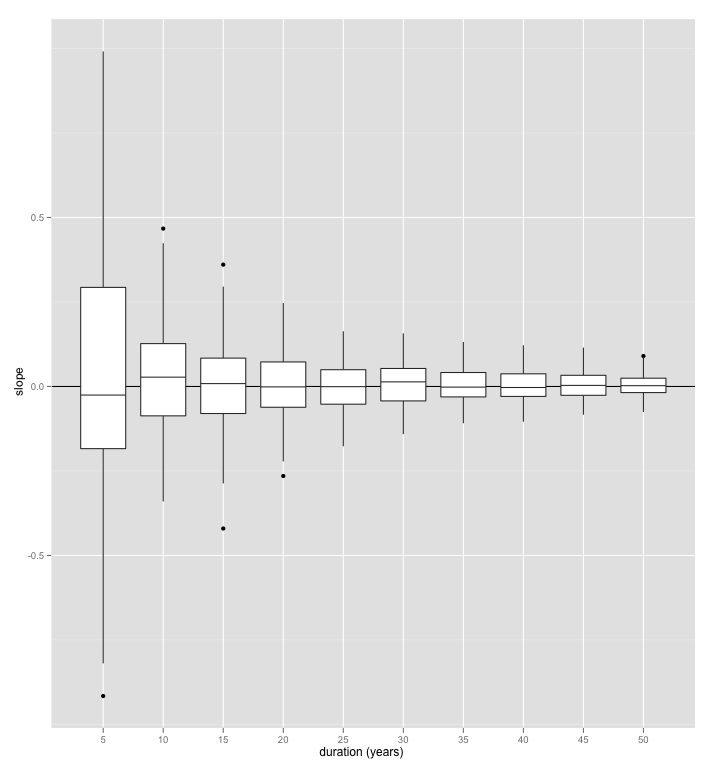
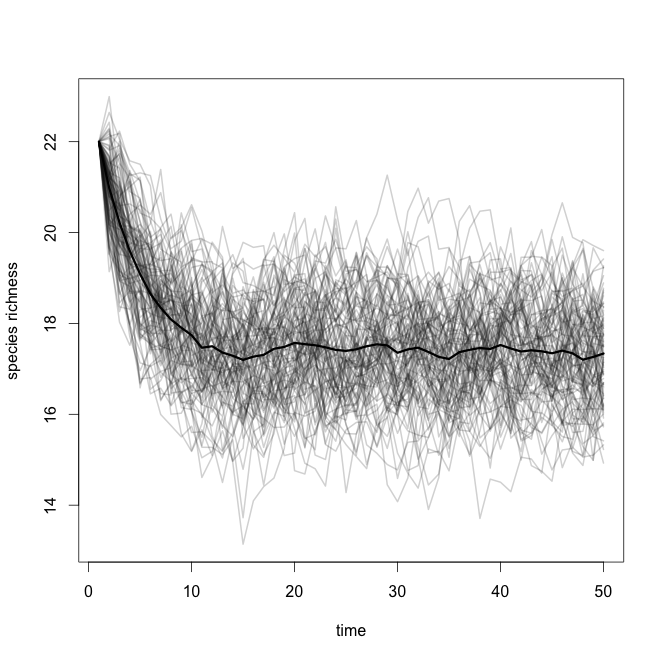
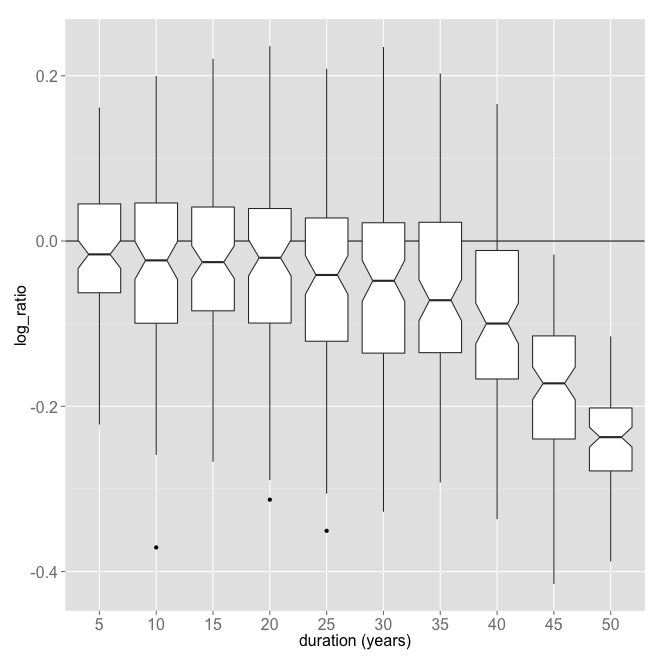
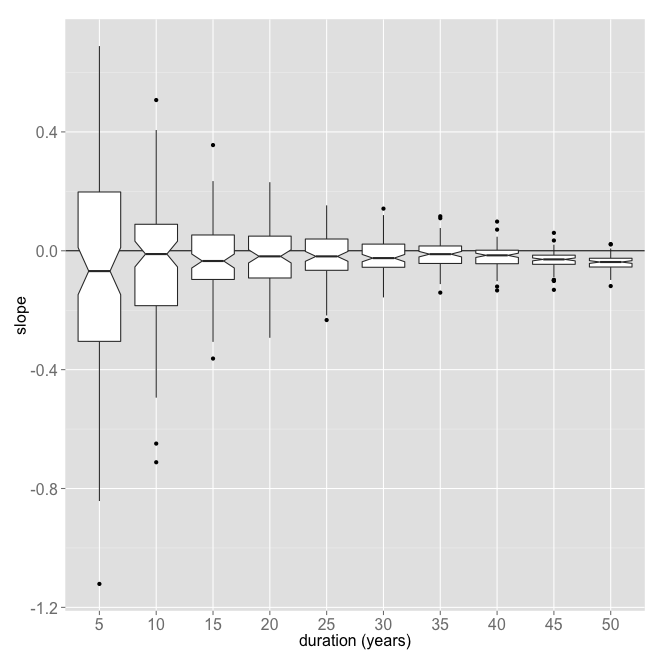


Figure S5. Panel (A) simulated dataset composed of 100 50-year time series of species number derived from an island biogeography model of community dynamics (see github repository); the black line shows the expected decline in mean trend over the 100 replicates. We created a meta-analytic data set by sampling 1000 shorter time series, 100 each of 10 to 50 years in length in increments of 5 years. Panel (B) shows the range of log response ratios and (C) OLS regression slope estimates derived from the dataset showing the effect of time series duration on the estimate. Panel (D) and (E) show the distributions of the log response ratios and slopes, respectively, estimated from all 1000 time series.

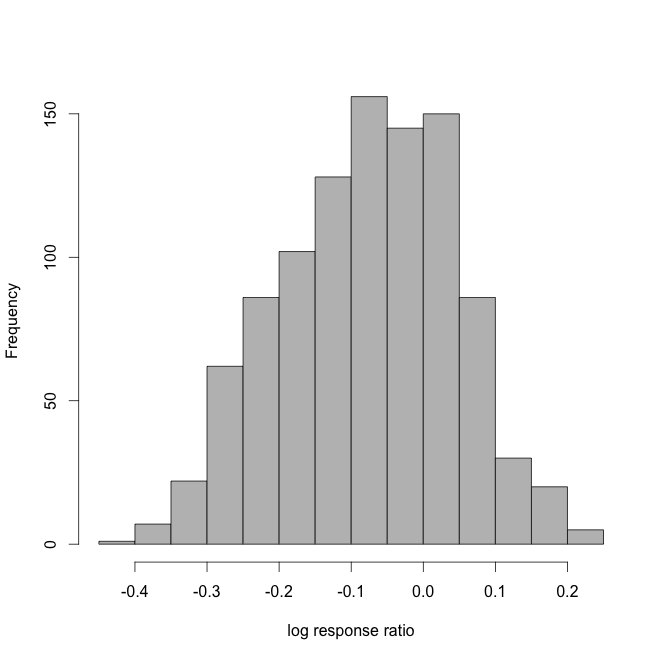
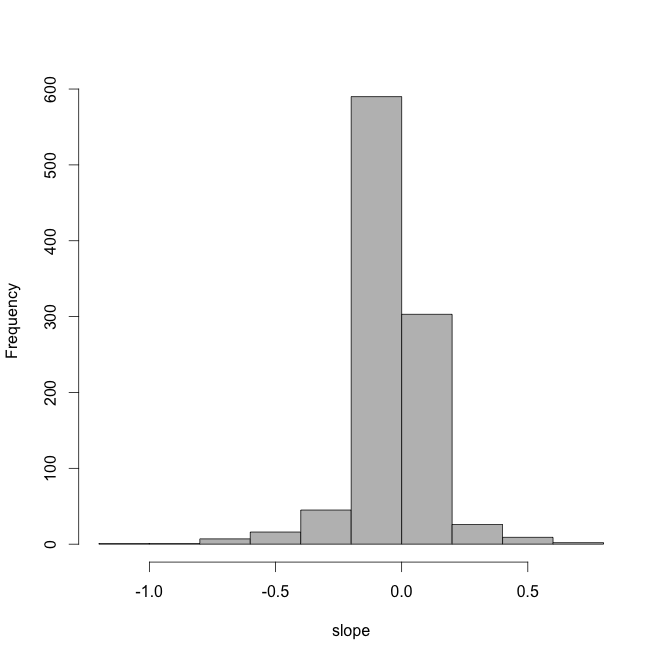


A

B

C

D



E