Paleobiology

cambridge.org/pab

Spatial standardization of taxon occurrence data—a call to action

Gawain T. Antell[1], Roger B. J. Benson[2] and Erin E. Saupe[3]

[1]Department of Earth and Planetary Sciences, University of California, Riverside, California 92521, U.S.A.
[2]Richard Gilder Graduate School and Division of Paleontology, American Museum of Natural History, New York 10024, U.S.A.
[3]Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, U.K.

On The Record

Cite this article: Antell GT, Benson RBJ, Saupe EE (2024). Spatial standardization of taxon occurrence data—a call to action. *Paleobiology* **50**, 177–193. https://doi.org/10.1017/pab.2023.36

Received: 18 August 2023
Accepted: 20 November 2023

Corresponding author:
Gawain T. Antell;
Email: gawain.antell@ucr.edu

Non-technical Summary

The fossil record is notoriously incomplete. The spatial distribution of fossils reflects in part the geography of biodiversity gradients, areas of sediment deposition and present-day rock exposure, and locations of wealthy nations with long-standing investments in Western science. Importantly for paleobiologists, the geographic location and size of fossil sampling gaps varies through time, between environments, and from one group of organisms to another. This spatial structure in recorded fossil occurrences has many consequences for ecological and evolutionary investigations. If the fossil record is taken at face value, results and conclusions will be inaccurate, sometimes to the point of being misleading. Therefore, it is essential to standardize the spatial distribution of fossil occurrences (the total area covered by sites and the spread across sites) before addressing research questions about diversity dynamics, geographic range size, or other ecological variables. We review sources of spatial structure in the fossil record, means to account for them, and possible consequences of leaving them unaddressed. Several of the tools we discuss are compiled into a new software package named divvy, in the R language of data analysis. We call for the paleobiology community to take up spatial standardization as a routine consideration in studying the informative but patchy fossil record.

Abstract

The fossil record is spatiotemporally heterogeneous: taxon occurrence data have patchy spatial distributions, and this patchiness varies through time. Large-scale quantitative paleobiology studies that fail to account for heterogeneous sampling coverage will generate uninformative inferences at best and confidently draw wrong conclusions at worst. Explicitly spatial methods of standardization are necessary for analyses of large-scale fossil datasets, because nonspatial sample standardization, such as diversity rarefaction, is insufficient to reduce the signal of varying spatial coverage through time or between environments and clades. Spatial standardization should control both geographic area and dispersion (spread) of fossil localities. In addition to standardizing the spatial distribution of data, other factors may be standardized, including environmental heterogeneity or the number of publications or field collecting units that report taxon occurrences. Using a case study of published global Paleobiology Database occurrences, we demonstrate strong signals of sampling; without spatial standardization, these sampling signatures could be misattributed to biological processes. We discuss practical issues of implementing spatial standardization via subsampling and present the new R package divvy to improve the accessibility of spatial analysis. The software provides three spatial subsampling approaches, as well as related tools to quantify spatial coverage. After reviewing the theory, practice, and history of equalizing spatial coverage between data comparison groups, we outline priority areas to improve related data collection, analysis, and reporting practices in paleobiology.

Introduction

Quantitative paleobiologists seek to measure attributes of recorded fossil occurrences that truthfully reflect the biological attributes of past ecosystems, such as biodiversity dynamics, community structure, or functional ecology. This goal is complicated by the strong confounding signatures of stochasticity and sampling on fossil-derived estimates (Signor et al. 1982; Behrensmeyer and Kidwell 1985; Smith 2001; Bush and Bambach 2004; Kiessling 2005; Patzkowsky and Holland 2012; Vilhena and Smith 2013; Raja et al. 2022; Nanglu and Cullen 2023). Stochasticity stems from many, unidentifiable sources, including misidentification, measurement error, and site-to-site variability. Sampling, meanwhile, encompasses an expansive set of processes that systematically bias fossil observation probability, such as taphonomy, sedimentation, erosion, and collector and taxonomist effort. (See Table 1 for additional definitions.)

© The Author(s), 2024. Published by Cambridge University Press on behalf of Paleontological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

PALEOBIOLOGY A PUBLICATION OF THE Paleontological SOCIETY

CAMBRIDGE UNIVERSITY PRESS

Table 1. Glossary of disciplinary terms relevant to taxon occurrences and spatial standardization.

Term	Definition
	A. Biology
α-diversity	Richness at the finest spatial grain of study, e.g., a single site
β-diversity	Variation and turnover in taxon composition between sites, often parameterized as (dis)similarity, replacement, and/or nestedness
γ-diversity	Total richness across the largest spatial scale of study
Collection	A taxon occurrence list tied to a single point coordinate on the Earth, e.g., records from a stratigraphic horizon at a site
Species–area effect	The positive relationship between spatial coverage and observed richness
	B. Geography
Area	Cumulative square kilometers of observation sites
Dispersion	Degree of aggregation between observation sites
Extent	Dimensions of the bounds around all observation sites
Minimum spanning tree	Metric from network analysis that connects all observation sites with line segments between successively nearest neighboring sites
Occupancy	Count of sites (e.g., raster grid cells) where a taxon occurs
Raster	A grid specifying the division of an area into cells with given coordinate reference system
Spatial heterogeneity	Uneven distribution of entities (e.g., habitats or taxa) over a landscape
Wallacean shortfall	Undersampling of geographic range distributions for living or fossil taxa, especially as the result of systematic, spatial differences in research investment
	C. Research practices
CARE	Indigenous data governance principles of Collective benefit, Authority of Indigenous peoples, Responsibility, and Ethics
FAIR	Open access principles that data be Findable, Accessible, Interoperable, and Reusable
Low-power problem	The finding that many bodies of literature have insufficient sample size or replicates to be able to detect null relationships correctly
	D. Statistics
Coverage	Relative to the full distribution of a focal variable (e.g., total land area in a time bin or true frequency distribution of taxa), the scope represented by observations
Power	Probability of correctly rejecting the null hypothesis if it were false, i.e., avoiding false negatives
Power analysis	Calculation of required sample size for a given effect size, significance threshold, and power, or power associated with a given empirical sample size, effect size, and significance threshold
Rarefaction	Type of subsampling that equalizes the quantity of a given unit (e.g., number of taxon occurrences) in each subsample
Sample	Set of observations taken to represent an entire group, usually because it is impractical or impossible to observe every group member
Sampling	Processes that influence the probability of recording an observation, e.g., fossilization
Stochasticity	Random error, e.g., from misidentification or measurement error (contrast with systematic bias)
Subsampling	The process of taking a subset of a sample, e.g., bootstrapping or rarefaction
Systematic bias	Differences between observed and true values that are consistent in direction or magnitude (contrast with stochasticity)

For 70 years, paleobiologists have detailed the effects of fossil preservation, rock exposure area, and research effort on paleobiodiversity estimates (e.g., Gregory 1955; Raup 1976, 1977). Correspondingly, empiricists have developed many frameworks to account for such sampling disparities—for instance, taxon-free metrics for live–dead analysis across intervals of environmental change (Smith et al. 2020), stratigraphic paleobiology to tease apart the geological structure of the fossil record (Patzkowsky and Holland 2012), and rarefaction techniques to standardize the completeness of reported taxon richness (Alroy et al. 2001; Alroy 2010; Close et al. 2018). These individual methods are well suited to studies of restricted geographic scope, such as faunal investigations of a single formation. Increasingly, however, such local or regional studies are eclipsed in the quantitative paleobiology literature by analyses at larger spatial scales, from continents or ocean basins to planet-wide scope, fueled by the rapid growth of digital fossil occurrence databases in the last two decades (Deng et al. 2020; Dillon et al. 2023).

When datasets span many strata, taxa, or time periods, manually correcting for the individual influences of fossilization, excavation, and publication practices becomes prohibitively time-consuming. Consequently, many analyses bypass these steps with the assumption that stochasticity and sampling structure contribute less signal to results than does original biological signal (e.g., Sepkoski et al. 1981; Benton and Emerson 2007). However, fossil data violate this assumption frequently, leading to imprecise or inaccurate results. For instance, differential sampling of equatorial and temperate zones can mask true shifts in latitudinal

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

biodiversity gradients through time (Allison and Briggs 1993; Alroy et al. 2001; Bush and Bambach 2004; Menegotto and Rangel 2018; Jones et al. 2021) and distort mean global temperature reconstructions (Jones and Eichenseer 2022). Differential spatial coverage between time steps can induce changes in estimated diversity dynamics, such as introducing false signals of unconstrained diversification (Close et al. 2017; Dunne et al. 2023). In the following sections, we review further evidence and present a case study demonstrating the potential reversal of study conclusions between face-value and standardized fossil data.

An efficient, holistic approach to account for biases in paleontological data at large scales is to standardize the spatial coverage of taxon occurrences, because geographic data distribution is an emergent property arising in large part from sampling and stochastic processes. Correcting for multiple sampling processes individually (e.g., applying richness rarefaction and facies analysis) may fail to account fully for spatial heterogeneity, but standardizing for spatial heterogeneity within and between datasets can mitigate the influence of multiple sampling biases at once. Traditional standardization methods such as richness rarefaction cannot replace spatial standardization because spatial heterogeneity interacts with the original structure of biological communities —regardless of subsequent sampling processes. That is, bias resulting from the primary species–area relationship (detailed in the following section, "Processes That Structure Fossil Occurrence Distributions," and Fig. 1 and Table 1) will remain even after stripping away biases of taphonomy, stratigraphy, and research effort. In contrast, spatial standardization will reduce the influence of the species–area effect from biological and sampling processes combined. Richness rarefaction or other nonspatial standardization may still be necessary as additions to spatial standardization but should not be treated as substitutes.

Here, we (1) detail biological, geological, and historical sources of spatial heterogeneity in published fossil occurrence data; (2) explain why standardization of both the spatial area and spatial dispersion of taxon occurrences is indispensable; (3) discuss practical considerations for analyzing spatially subsampled data; (4) illustrate a case where spatial standardization substantively changed paleoecological conclusions; and (5) introduce the R package divvy to implement spatial subsampling methods. Concluding notes review the state of spatial standardization in quantitative paleobiology and suggest data access and methodological development priorities to pursue.

Processes That Structure Fossil Occurrence Distributions

Biodiversity is uneven in its spatial distribution in the present day and throughout recorded evolutionary history, at all spatial scales. Biogeographers describe this unevenness by partitioning species richness into three hierarchical spatial levels (Fig. 1). At the local scale (e.g., a single quadrat, raster grid cell, or quarry), taxonomic richness is termed alpha diversity. At the largest scale of the whole study area (e.g., the extent of a research station, ocean basin, or planet), total richness is termed gamma diversity. Turnover, the factor by which gamma exceeds mean alpha, is termed beta diversity, and increases as the total area of sampling increases or as the dispersion of sampling locations spreads (Whittaker 1960, 1972; Connor and McCoy 1979; Tuomisto 2010). This biological link between diversity and spatial coverage, the species–area effect (Preston 1962; MacArthur and Wilson 1967), carries across all spatial grains throughout the Phanerozoic (Sepkoski 1976; Barnosky et al. 2005).

Fossil occurrences undergo additional spatial structuring from geological processes after organisms die. Taphonomy, sedimentation, and erosion all affect the geographic distribution, abundance, and present-day exposure of fossils (Signor et al. 1982; Behrensmeyer and Kidwell 1985; Smith 2001; Patzkowsky and Holland 2012; Vilhena and Smith 2013). So long as these factors are randomly and identically distributed across regions or time intervals of comparison, large-scale biodiversity studies may proceed without systematic bias in conclusions (Bush et al. 2004). There are also cases where multiple sources of systematic bias differ in direction, with the net effect that a study can ignore all these confounding factors and still manage to arrive at results with the same direction and approximate magnitude as when samples are standardized (Bush and Bambach 2004; Marcot 2014). However, such studies probably represent a rare minority of cases, and therefore it is prudent to account for the spatial structure of fossil distribution as a matter of course (Benson et al. 2021).

Human research effort applies a final spatial filter on the distribution of specimens that enter the paleontological (and neontological) record. For fossil data to enter an occurrence database, people must unearth, identify (in binomial Latin taxonomy), georeference (in Cartesian coordinates), and report the findings (usually in English). Given these obstacles to publishing fossil occurrence data, it should be unsurprising that studies dating back decades have noted the co-distribution of paleobiodiversity with modern-day research effort and in-country wealth (e.g., Raup 1977; Kiessling 2005). The underrepresentation of fossil occurrences in the Global South stems from long-standing and ongoing imperial extraction of material and intellectual resources that deprives people in these areas from studying and communicating their paleontological heritage (Monarrez et al. 2022; Raja et al. 2022). Unequal research investment globally shapes the density of published knowledge about extant species, too; ecologists use the terms Wallacean, Hutchinsonian, and Linnean shortfalls for the sampling gaps that truncate recorded geographic ranges, environmental occupancy, and counts of described species, respectively (Hortal et al. 2015; Oliveira et al. 2016).

Biological, geological, and historical spatial processes add bias to all ecological metrics derived from fossil distributions, not only metrics related to spatial traits. For instance, exploratory analyses indicate that community connectedness metrics vary with spatial sampling coverage in a similar way that geographic range size measurements do (C. Malanoski personal communication 2022). Similarly, spatial unevenness affects estimates from all large-scale datasets, including neontological occurrence datasets. The Paleobiology Database (PBDB) has become one of the most popular sources for fossil occurrences reported globally, resulting in 450 publications to date (https://paleobiodb.org, accessed 16 March 2023), and thus has received particular scrutiny for its spatial biases. However, spatial standardization is equally relevant for records from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org), Ocean Biodiversity Information System (OBIS; https://obis.org), Neotoma Paleoecology Database (https://www.neotomadb.org), and other datasets spanning upward of a continent or ocean basin (Boakes et al. 2010; Beck et al. 2014; Menegotto and Rangel 2018; Moudrý and Devillers 2020).

Methods to Standardize Spatial Coverage

All the biological, geological, and historical processes described in the preceding section contribute to spatial heterogeneity of fossil

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

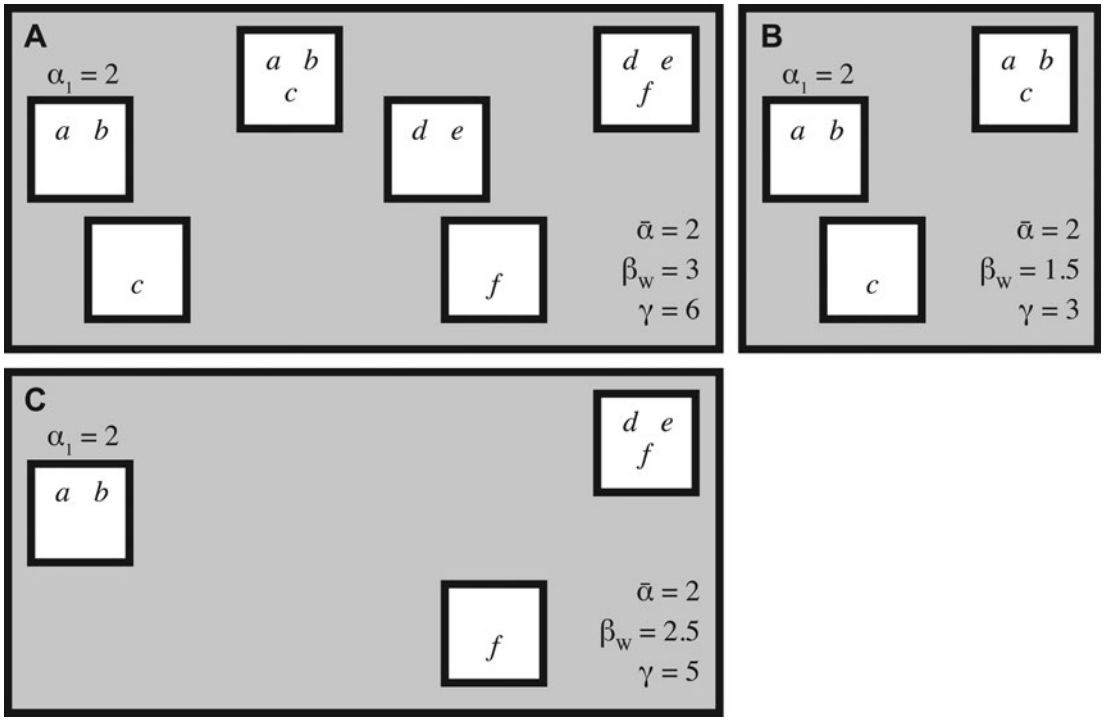


Figure 1. A schematic of the species–area effect, in map view. The total sampling area (gray boxes) in A and C is twice as large as in B; these bounding regions could represent the total preserved outcrop area from three time steps or continents of comparison. Individual sampling sites within a study region are indicated with clear boxes, and species occurrences are represented with lowercase letters. Species count at an individual site is alpha diversity (annotated at only one site in each panel, for simplicity). Total species count within a study area is gamma diversity. There are many metrics for beta diversity related to species turnover between sites, but a simple and original measure is the ratio of gamma to mean alpha (Whittaker 1960, 1972). Note that both beta and gamma diversity increase as sampling area doubles from B to A, even though the distributions of alpha diversity, species' geographic range size, and site density are identical. Without accounting for the difference in sampling area, (paleo)ecologists might falsely infer time bin A more diverse than B and with smaller proportional range sizes. C also has larger beta and gamma diversity than B, despite the same number and cumulative area of sampled sites, because the dispersion between sites is larger.

occurrence data. Unfortunately, accounting for these processes individually—for instance, by restricting analysis to sites of similar preservation potential and performing richness rarefaction—fails to remove signatures of spatial heterogeneity from results (Bush et al. 2004; Close et al. 2020). Even site–occupancy models borrowed from ecology, which simultaneously estimate taxon detection rates and occurrence rates (Reitan et al. 2022), overlook the differential influences of spatial turnover that arise whenever data from one comparison group are distributed differently in geographic space from those of another group. As discussed elsewhere, sampling correction using residuals-based biodiversity estimates is also unsuitable: not only are these methods sensitive to the choice of sampling proxy, but the final model in a succession lacks the information to estimate errors appropriately (Brocklehurst 2015; Sakamoto et al. 2017; Dunhill et al. 2018).

The only adequate way to control for biases deriving from spatial structure is with explicit spatial standardization (e.g., Antell et al. 2020; Close et al. 2020; Womack et al. 2021). In particular, analyses must control for both the area and dispersion components of spatial coverage. Additional factors that may be relevant to standardize include data list structures (e.g., PBDB collections) and habitat distribution.

Area

With more sites in a dataset, more taxa tend to be recovered, as discussed earlier with regard to the species–area effect (Fig. 1). An immediate if partial remedy to the influence of area on biodiversity metrics is rarefaction on the number of sites. Rarefaction is a class of subsampling that equalizes the quantities of a given unit. To limit confusion in the previous sections, only standardization of taxonomic richness is referred to as "rarefaction," while standardization of spatial coverage is referred to as "subsampling." We continue to use rarefaction over subsampling as the term for richness standardization throughout this piece, although either terminology is appropriate—for instance, "coverage-based rarefaction" (Chao and Jost 2012) and "shareholder-quorum subsampling" (Alroy 2010) are equivalent methods to equalize the coverage of frequency distribution curves for diversity estimation.

Rarefaction of sites is intended to equalize sampling area on a map, but note the definition of a site may vary between studies. For example, a paleoecologist interested in comparing terrestrial vertebrate diversity between habitats might subsample (rarefy) an equal number of quarries from each paleoenvironment, while a paleobiogeographer interested in North American plant diversity through time might subsample an equivalent number of equal-area raster grid cells from each time step. The area or number of sites to set as a quota for standardization should form a sample size large enough to adequately characterize the variable of interest. As a common rule of thumb, a sample size of six data points is often taken as a bare minimum in statistics to estimate a mean value with acceptable precision. However, given the large site-to-site variability of many metrics in paleoecology, a minimum of twice this may be more appropriate. As precedents in paleobiological literature, Womack et al. (2021) set a quota of 13 raster cells, and Antell et al. (2020) selected 10, 12, or 15 raster cells depending on the dataset and mode of analysis.

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

Dispersion

Even after variable numbers and square kilometers of sites between comparison datasets are accounted for, the spatial distribution of occurrences still imprints a signature on ecological parameter estimates. For the same amount of sampled area, a larger spacing between sites tends to correspond to more community turnover and larger beta diversity (Fig. 1), again due to the species–area effect. Therefore, spatial standardization must account for the dispersion of occurrences as well as areal coverage.

Subsetting occurrences into regions defined by a standard bounding extent can limit the variance of site dispersion across time steps or environments, at least to a moderate degree. Objective ways to define regional subsample boundaries include circles centered on random occurrences (Antell et al. 2020), minimum spanning trees split at their longest branches into subtrees (Close et al. 2017), and latitude–longitude boxes (Marcot et al. 2016). Precursors to these automated regionalization methods date back to early studies on public fossil database records; for example, the first step in a collections-based subsampling procedure proposed by Alroy (2000) was the omission of the small subset of faunal lists from eastern North America, "to minimize the biogeographic spread of sampling through time" (p. 716).

The diameter or length to set as a maximum limit of subsample dispersion ideally would be informed by empirical data about the extent of biogeographic regions relevant to the study. For instance, a subsample region significantly larger than an average (paleo)continent or ocean basin would be too wide to limit beta diversity in many cases. On the other hand, a subsample region smaller than the average geographic range size of focal taxa would unnecessarily truncate observations in a study of range size. Practically, the subsampling parameter controlling dispersion is often set at the smallest size that still allows subsamples in every comparison group to attain the quota for included sites or area. Antell et al. (2020) reported results from circular subsamples with diameters of 3000 km (1500 km radius; Fig. 2A) and 1500 km (750 km radius) in conjunction with sensitivity tests for site rarefaction quotas of 6, 10, and 12 sites. Close et al. (2017) defined subsample regions based on minimum spanning trees connecting the centroids of occupied raster grid cells and set the maximum summed tree length at 3200 km ± 10% (Fig. 2B). Close et al. (2020) modified this subsampling procedure and reported results at summed tree lengths of ~2400, 4000, and 5600 km (~1500, 2500, and 3500 miles). Womack et al. (2021) also used minimum spanning tree length as a metric for standardization, with a cutoff of 1000–1100 km.

Collections and Other Data Lists

One common, nonspatial method of correcting for differential sampling effort is rarefaction of occurrence lists (also called faunal lists in the paleobiological literature, although the method is equally applicable to flora). For PBDB data, the primary list structure is the "collection," a term with only a loose definition that refers to any sampling unit tied to a single point coordinate on the Earth. Collections may contain any number of specimen records from any number of clades or publications; these taxon lists may be equivalent to expedition localities, stratigraphic horizons, or other episodes of field collection, but such definitions are inconsistent between studies in the database.

Paleobiologists have published many variations of list rarefaction. The basic logic follows earlier theory in ecology (Shinozaki 1963); a later addition of weighting by number of occurrences included in lists was meant to mitigate sampling biases reflected in list length (reviewed in Alroy 2000). Rarefaction of fossil occurrence lists has been discussed as an indirect approach to spatial subsampling. Previous work has taken the number of collections as a proxy for beta diversity and the length of collections for alpha diversity (Bush et al. 2004). However, these proxies are imperfect in both theory and practice: accounting solely for the number of lists or the number of occurrences they contain cannot directly control the geographic dispersion of sites. Many authors agree the performance of list rarefaction methods varies with the spatial structure of occurrences, and so any such method represents an incomplete correction for fossil sampling biases (Alroy et al. 2001; Bush et al. 2004). Researchers at the time of development of list-based methods admitted dissatisfaction with their proxies for this reason and remarked, "routines that directly control the geographic and environmental composition of a subsample need to be developed" (Bush et al. 2004: p. 668).

Modern computing power and geographic information systems allow us to answer the call for explicitly spatial standardization methods. The site-based rarefaction described earlier is broadly analogous to unweighted list rarefaction, given the one-to-one correlation between reference counts and raster grid cell counts in global PBDB datasets (Alroy et al. 2008). The additional control on dispersion described in the preceding subsection further standardizes beta diversity. Ideally, standardization procedures would include an additional step to subsample the frequency distribution of taxon abundance data at each site and thereby control alpha diversity at the local scale (Bush et al. 2004). An analogous step in species distribution modeling for extant taxa is subsampling the number of occurrences of a focal taxon within each equal-area grid cell of a study area (Beck et al. 2014). Regrettably, harmonized abundance data are not yet available for many large composite fossil datasets.

Habitat Heterogeneity

An additional confounding factor when attempting to account for geographic distribution of data is that habitat heterogeneity influences the species–area effect independently of area itself (Furness et al. 2023). Although the ecological community has yet to agree on a common measure of habitat heterogeneity (Loke and Chisholm 2022), it is readily apparent that occurrences spanning multiple environments likely differ in their estimated ecological attributes compared with occurrences from a homogeneous environment. Standardizing for dispersion can mitigate differences in habitat heterogeneity to a limited degree but fails to address the problem directly. The most rigorous way to account for differential coverage of paleo-habitats between comparison groups is facies analysis, a cornerstone of stratigraphic paleobiology (Patzkowsky and Holland 2012). This approach is most feasible when the study area lies within a single basin where it is possible to construct a comprehensive model of sequence stratigraphy from field observations.

When environmental occurrence data cannot be resolved in a sequence–stratigraphic framework, as for most public database and museum records or other inherited datasets, it is worth attempting habitat standardization through grosser means. Extracting the "environment" and "lithology" fields associated with specimen occurrences in the PBDB, for instance, can categorize records into coarse divisions such as shallow- versus deep-water, siliciclastic versus carbonate, and fine- versus coarse-

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

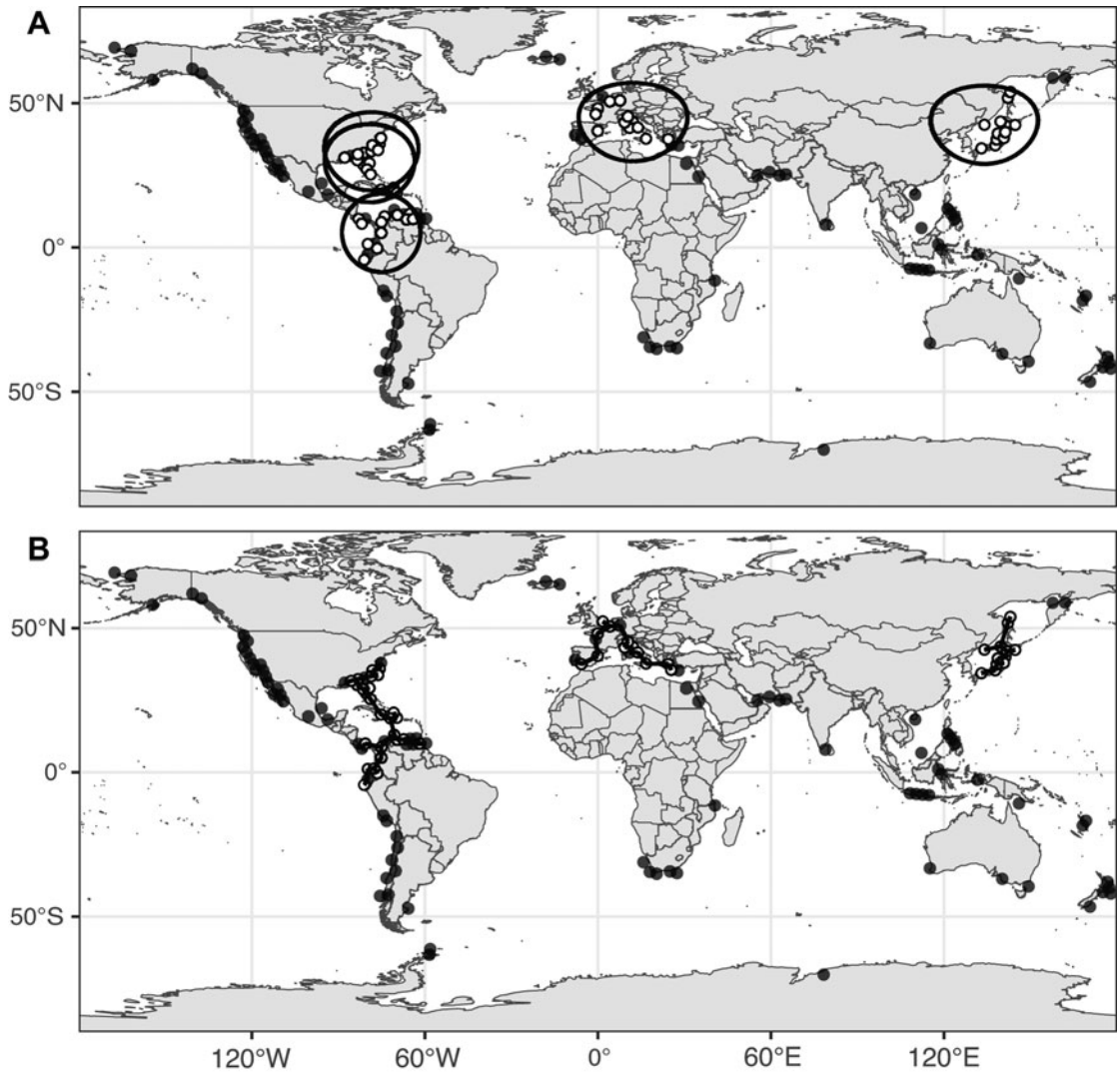


Figure 2. Five spatial subsamples of Pliocene bivalve occurrences from the Paleobiology Database (available as data object *bivalves* in the R package divvy). For each subsample, site dispersion is constrained by a circle of 3000 km diameter (A) or a minimum spanning tree with maximum great circle distance of 3000 km (B). Within each subsampling region, the number of occurrence sites is rarefied to 12 (open circles). Sites are raster grid cells of approximately equal area and shape. The random points to initiate subsamples are identical in A and B. Note that subsamples here are impervious to potential biogeographic barriers, for example, the Isthmus of Panama, which was not emergent for the full duration of the Pliocene. Subsamples can also overlap with each other, as shown in southeastern North America for two circular subsamples and three minimum spanning trees. Subsamples with overlapping regional boundaries may differ in the random subsets of sites they contain.

grained substrate settings (Nürnberg and Aberhan 2013; Antell et al. 2020). Environmental data in large public databases should be treated with appropriate suspicion. As with all variables in big data, inconsistencies in data-entry lexicons and outright errors abound. Data vetting is critical to any analytic workflow, not only as a preliminary step but recurring throughout investigation, as additional data inconsistencies may appear. An alternative approach to estimate water depth is to pair occurrence coordinates with paleo–digital elevation models (e.g., Close et al. 2020), although such models also contain large imprecision.

Considerations for Analysis of Spatial Subsamples

Statistical Properties of Subsamples

There are several ways analysis differs when based on subsamples instead of a single dataset of full geographic extent. Subsampling is already common in paleobiology in the forms of bootstrapping tests and richness estimators (including shareholder-quorum subsampling), and many ecology texts are available to describe these applications and their properties (e.g., Bolker 2008; Chao and Jost 2012; Close et al. 2018). Here, we review statistical features to consider in the specific case of spatial subsampling.

One obvious feature of a subsample is that it contains only a subset of the information in the full dataset. Nevertheless, after iterating a subsampling procedure, many or all observations may be represented in at least one subsample and so still contribute to the overall analysis. Each subsample generates ecological estimates that are meaningfully comparable to those of any another subsample. This equivalency facilitates fair ecological comparisons between time steps or environments, or between organismal groups that differ in fossil record coverage.

Variance among subsamples broadly reflects variance among geographic regions. To the degree that the area and location of

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

subsamples correspond to bioregions (Fig. 2), error bars from subsampled estimates within a category (e.g., time bin) thus indicate a first-order signal of biogeographic heterogeneity. Empirically, total variance reflects the sum signal of both biogeographic heterogeneity and stochasticity, in the sense of true random differences. Estimating the relative contributions of these factors may be cumbersome or impossible in most cases. Therefore, variance of subsampled estimates should be interpreted as arising both from error in estimating true values within regions and from spread in the true values across regions. It may be misleading to label ranges of subsampled estimates as confidence intervals, which implies variance surrounding a single, meaningful population mean. A single mean across the world rarely exists for biodiversity data. A neutral phrasing to discuss variance among subsample estimates within a category could be just that: a given quantile range of values across subsamples, corresponding to regional variation in population means, with error.

Power Analysis

No standardization procedure fully eliminates the biases it aims to correct. Rather, the goal of standardization is to reduce the effect size of biases to less than that of the hypothesized signal under investigation. Unfortunately, in many cases, the strength of expected signal is unknown, which may have been the impetus for conducting the study in the first place. In studies where analysis returns null results, uncertainty about a signal's effect size can lead to uncertainty of whether to infer the true absence of that signal or merely insufficient power to detect it (Altman and Bland 1995). Power analyses can foresee and sometimes remediate such scenarios.

A power analysis generates an equation to relate effect size, sample size, significance threshold (probability of rejecting the null hypothesis when true; ceiling for allowed chance of type I error), and statistical power (probability of rejecting the null hypothesis when false; limit on type II error). Experimental, resource-intensive studies such as clinical trials solve this equation for the minimum sample size sufficient to detect a given effect size at a given significance threshold (usually 5%) and power (usually 80%). For analysts of natural experiments such as the fossil record, sample size is usually a given, and an alternative goal may be to solve for the power to detect the signal at a given strength or range of strengths. Many bespoke software tools are already developed for power analysis of specific statistical tests or models, such as the R packages SIMR for generalized linear mixed-effects models (Green and MacLeod 2016) and pwr for *t*-tests, analysis of variance, Pearson correlation, and other common parametric tests (Champely 2020). Simulations are a flexible strategy to approach power analysis within any statistical framework (Bolker 2008).

An example of simulation-based power analysis in paleobiology is one conducted by Antell et al. (2020), reanalyzed in the case study presented in this article. After PBDB data were spatially standardized, results indicated a lack of relationship between the predictor (regional species count) and dependent variable (geographic range size). To estimate the study's power to detect the hypothesized nonzero relationship between these variables, the authors simulated a biotic signal at a range of magnitudes and calculated the probability of correctly recovering the signal given empirical subsample sizes. In this case, the analysis retained strong power to detect true signal even after subsampling (Antell et al. 2020: SI fig. 3, SI Methods section S3.3). However, there is no guarantee of such an outcome for every PBDB study. Spatially explicit neutral models are a related class of simulation with recent application in paleobiology; Dunne et al. (2023) used this approach to demonstrate that the putative tetrapod radiation after the Carboniferous rainforest collapse can be explained entirely through increased spatial sampling coverage instead of increased endemism and speciation.

Power analysis remains an underutilized tool with valuable potential in paleobiology and macroecology. If more studies in these fields were to estimate their power or significance, it is possible many would find the present sample size and distribution of fossil occurrences insufficient to reliably detect the signals they aim to identify. The low-power problem is well documented in related fields such as ecology and animal behavior (Smith et al. 2011; Kimmel et al. 2023), and the large number of study designs, analytic modes, and tested relationships practiced in paleoscience make it susceptible to the problem (Ioannidis 2005). A finding of insufficient power is useful: although it would be wise to exercise restraint in pursuing an investigation until sufficient data became available to address the research question robustly (Nanglu and Cullen 2023), the power analysis alone would be a valuable contribution to identify the exact scope of further data required and redirect collection efforts to address that targeted need.

Case Study: Consequences of Analyzing Non-standardized Data

A recent study by Antell and others (2020) set out to quantify the degree to which species' geographic distributions through time reflected the number of species that might be in competition. The classic and intuitive theory of ecological release posits that competition restricts species' resource use and thereby abundance and geographic distribution: when many species compete over limited resources, each species on average will have a smaller share, reducing reproduction and population expansion (Roughgarden 1972). This expectation of an inverse diversity–distribution relationship has large-scale consequences for both macroecology and macroevolution, as a possible explanation for diversity-dependent mathematical patterns of diversification—the self-regulation of extinction and speciation rates (Alroy et al. 2008; Rabosky 2013; Aguilée et al. 2018; Foote 2023; Pie et al. 2023). Testing the relationship between diversity and geographic range size at the large scales of these hypotheses is far from straightforward, however, because both variables share tight correlations with the geographic coverage of sampling.

One method of standardizing estimates of geographic range size for heterogeneous sampling coverage through time is calculating proportional occupancy: the number of occupied sites or raster grid cells as a fraction of all sampled sites or cells. Proportional occupancy has many precedents in the paleobiological literature (e.g., Foote et al. 2007; Harnik et al. 2012; Finnegan et al. 2015). Here, we reanalyze the data from Antell et al. (2020) to calculate mean proportional occupancy of species in each time bin. Data consist of PBDB occurrences from ~17,000 brachiopod and bivalve species from all marine sites throughout the post-Cambrian Phanerozoic, binned to equal-area grid cells (average width 100 km) in 63 time bins (Appendix Table A1). The correlation between mean proportional occupancy and species count, as a proxy for number of potential competitors, is plotted in Figure 3A. Corrected for time-series correlation by pre-whitening the predictor time series as residuals of a first-order autoregressive model (as in all correlations reported throughout this section), the

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

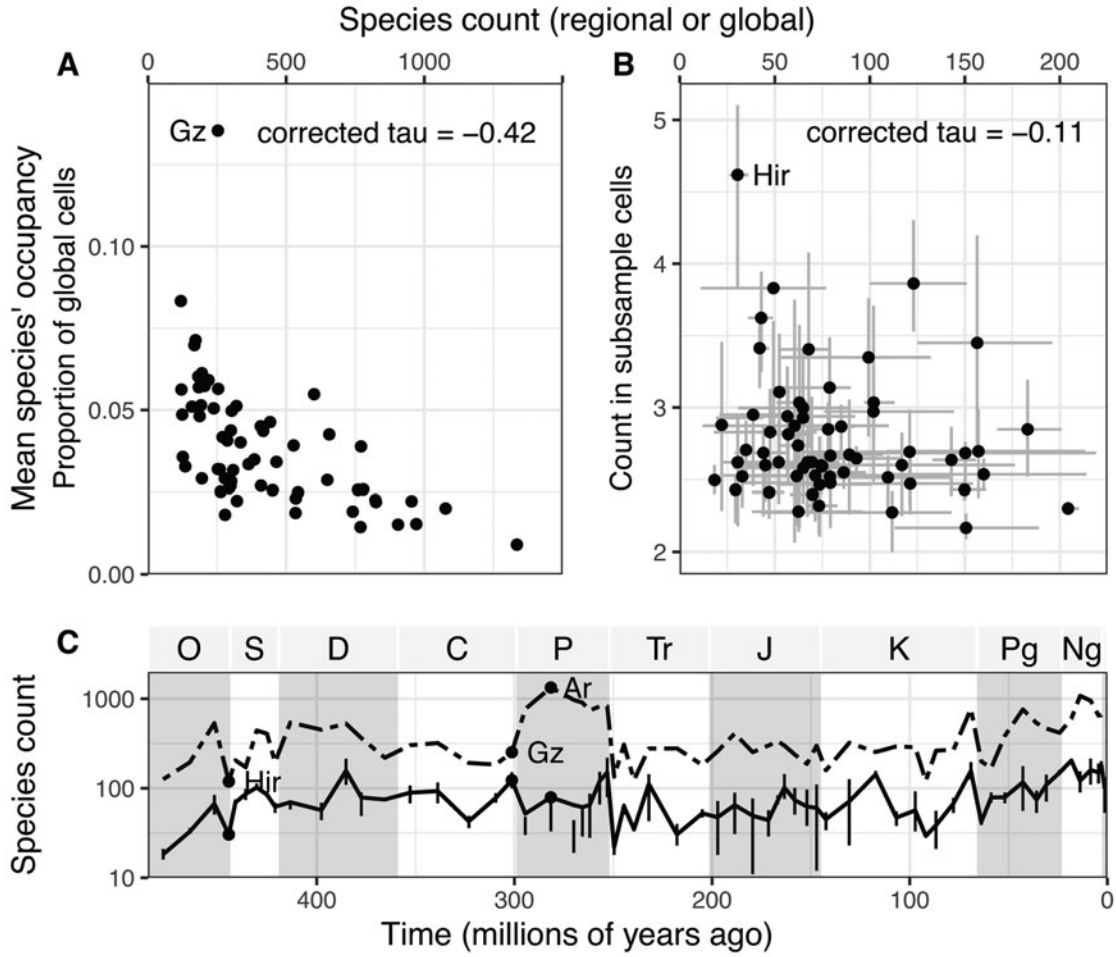


Figure 3. Scatter plots indicate the relationship between species count and mean per-species occupied grid cells in 63 time bins, either as a proportion of all occupied grid cells (A) or as a count within subsample regions of 12 cells (B). Outlier points are labeled by geological stage and overplotted on C: Ar, Artinskian; Gz, Gzhelian; Hir, Hirnantian. C, Species count in each stage, either tallied globally (dashed line) or within subsampled regions (solid line). Note logarithmic y-axis scale in C. Error bars in B and C denote interquartile range across 500 replicate subsampled regions. Geological periods: O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; N, Neogene.

Kendall's tau correlation coefficient is −0.42 (95% confidence interval [−0.56, −0.27]). This result is a stunningly strong correlation, in agreement with the negative relationship posited by ecological theory—but is it trustworthy?

The species–area effect induces strong relationships between observed richness and geographic sampling coverage. Figure 4A,B plots species count against spatial coverage of sampling; positive relationships appear in both plots, with magnitudes large enough to explain the entirety of the focal correlation in the preceding paragraph. Species count increases approximately linearly as a function of the number of equal-area grid cells in a time bin (Fig. 4A), with a nonparametric correlation (Kendall's tau) of 0.41 (95% CI = [0.26, 0.56]; Appendix Fig. A1A). Species count also increases monotonically as a function of the dispersion of sampled sites in a time bin (summed distance of minimum spanning tree connecting all occupied cell centroids), as plotted in Figure 4B. The correlation magnitude was similar, with tau of 0.44 (95% CI = [0.31, 0.56]; Appendix Fig. A1B). However, the shape of the latter relationship appears nonlinear, consistent with an exponential or power law form, as is common for species–area relationships (Matthews et al. 2016).

The relationship between geographic sampling coverage and range size as measured by proportional occupancy is equally strong but negative in direction. Proportional occupancy decreases sharply as a function of either equal-area grid cells in a time bin (Fig. 4C and Appendix Fig. A1C, tau = −0.67, 95% CI = [−0.77, −0.55]) or dispersion of those grid cells (Fig. 4D and Appendix Fig. A1D, tau = −0.65, 95% CI = [−0.74, −0.54]). This result can be explained by fluctuations in coverage of the enormous study area through time, which disproportionately impact the numerator and denominator of fractional occupancies. With more extensive sampling, linear increases in the denominator (total sampling area, with a maximum size of the planet's surface area) outpace modest increases in the numerator (observed range size, which has an asymptotic limit at the true range size of a species). Furthermore, the scaling of mismeasurement in proportional range size is unequal between taxa: as study area increases beyond the extent of a species' geographic range, the difference in proportional occupancy of widespread species will be greater than that of restricted species (mathematical proof in Antell et al. (2020): SI Methods S1).

Proportional occupancy is inadequate to standardize differential spatial coverage of fossil occurrences through time. Detection

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

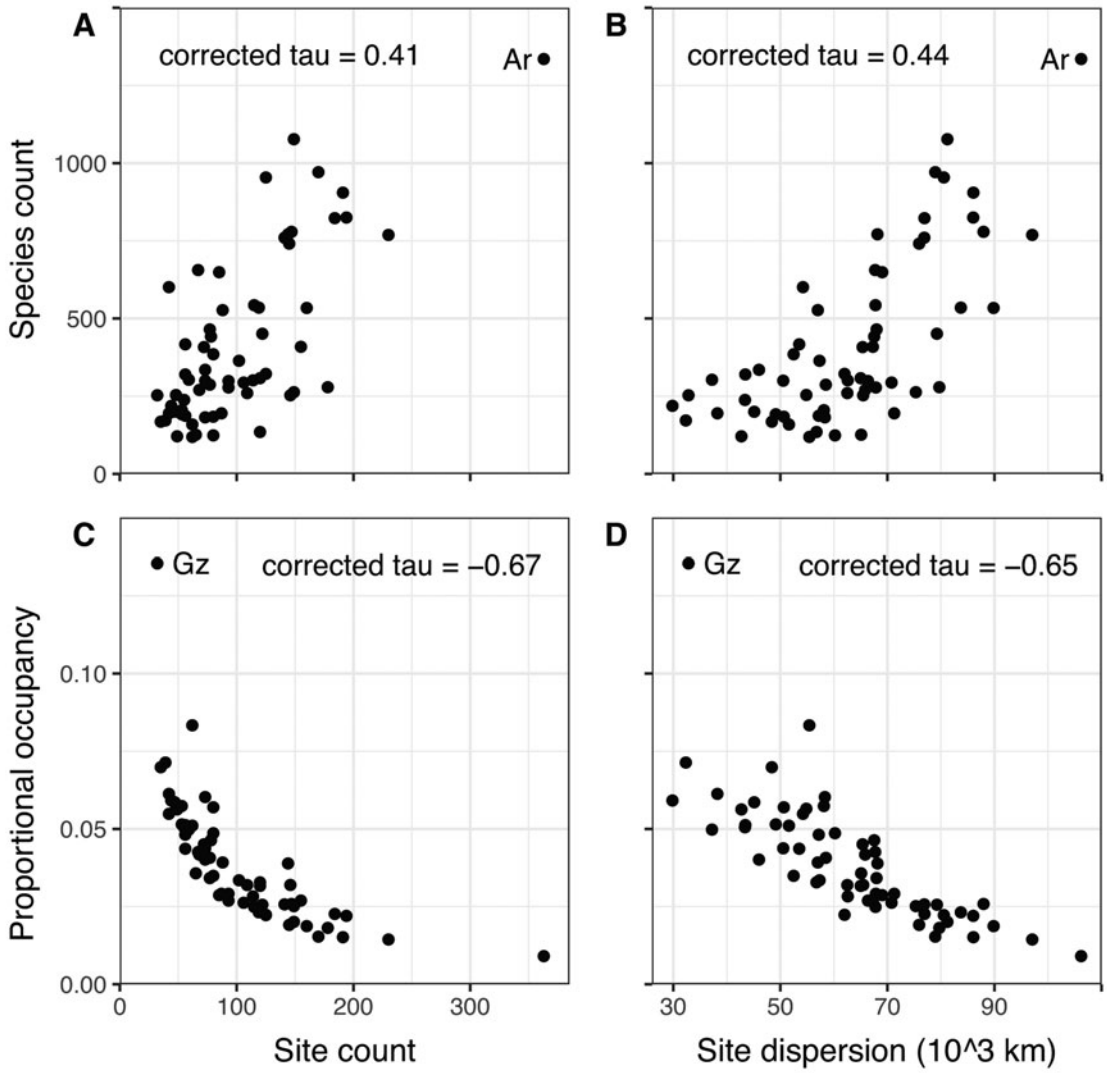


Figure 4. Scatter plots indicate the pairwise relationship between either species count (A and B) or mean proportional occupancy of equal-area grid cells (C and D) and spatial sampling coverage, measured as either a count of grid cells (A and C) or summed length of minimum spanning tree connecting occupied cell centroids (B and D). Outlier points are labeled by the earliest geological stage of a time bin, here and on the timescale in Fig. 3C: Ar, Artinskian; Gz, Gzhelian.

ratios, a similar metric used in ecology, have received analogous criticism (Reitan et al. 2022). However, spatial subsampling represents a viable alternative to measure geographic range size as well as species richness, because the method can control sampling area and dispersion directly. The original 2020 study tested 12 variations of subsampling methods for the global occurrence dataset. Here, we reanalyze data from the main text results: 500 replicates of subsampling with a quota of 12 equal-area grid cells in regions defined by a 1500 km radius. The cells within each circular region were drawn with weighted probabilities inversely proportional to the square of the distance from the subsampling center point, a procedure designed to further limit dispersion of samples where sufficient data were available closer to the center.

Subsampling successfully diminished the dominating signature of spatial coverage in the study variables: the corrected tau correlation between species count and site dispersion centered on zero (Appendix Table A2). Large fluctuations in species count present in the global curve were strongly moderated in subsampled estimates, particularly in the Permian and Cenozoic (Fig. 3C), matching the substantive changes in global richness curves for all marine genera following regional subsampling (Close et al. 2020). Geographic range size was measured as mean count of occupied cells (out of 12) among species in each subsample, excluding singly occurring species; dividing by the total number of cells to derive proportional occupancy was unnecessary, because every subsample contained the same number of cells, by design. Species count and mean occupancy were substantially more independent in subsampled data (Fig. 3B, Appendix Table A2; tau = −0.11, 95% CI = [−0.25, 0.01]), compared with face-value data (Fig. 3A). When range size was measured as median summed minimum spanning tree length, this independence was even clearer (tau = −0.06, 95% CI = [−0.19, 0.08]). Thus, the overall conclusions drawn from unstandardized data—that a negative relationship exists between range size and species count, congruent with the hypothesis of ecological release—dissipates entirely after accounting for the joint influence of heterogeneous spatial coverage on range size and species count.

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

divvy: Spatial Subsampling and Analytic Tools in R

The preceding sections reviewed both theoretical and empirical justifications for spatial subsampling in paleobiology and ecology (and also note Barnosky et al. 2005; Marcot et al. 2016; Close et al. 2017, 2020; Antell et al. 2020: SI Methods S3; Benson et al. 2021; and references therein). However, despite vigorous discussion of this topic, accessible tools for spatial subsampling of taxon occurrence data have remained limited. This lack of shared protocols and standards for data processing hinders the wider adoption of spatial standardization in quantitative paleobiology (Dillon et al. 2023). Although the code for many spatial subsampling methods has been made public alongside the papers it supports, there is seldom comprehensive documentation, human-readable syntax, and ongoing maintenance (e.g., bug patches) for publication-associated code to the same standards of standalone published software. Additionally, many spatial analysis scripts in paleobiology and ecology have used functions from the R packages sp and raster, which were recently deprecated, and the dependent packages maptools, rgdal, and rgeos, which were retired in 2023. There is a need for formal, actively maintained tool kits to perform common spatial analysis steps in paleobiology, similar to the divDyn and palaeoverse packages of tools to streamline common data manipulation tasks (e.g., time-binning) and perform common calculations (e.g., extinction and origination rates) in paleobiology (Kocsis et al. 2019; Jones et al. 2023).

The new R package divvy (Antell 2023) helps address current research community needs by implementing three versions of spatial subsampling methods, as well as related tools to quantify spatial coverage of taxon occurrences. Each function is fully documented, with references and examples, and the undergirding spatial calculations are built on the sf and terra packages—the replacements for now-unsupported R spatial packages. Spatial subsampling in divvy is operationalized in the following functions:

1. *cookies*: imposes a radial constraint on the spatial bounds of a subsample and standardizes area by rarefying the number of localities (Fig. 2A);
2. *clustr*: aggregates sites that are nearest neighbors (connecting them with a minimum spanning tree) to impose a maximum diameter on the spatial bounds of a subsample, and optionally rarefies localities (Fig. 2B); and
3. *bandit*: rarefies the number of localities within bands of equal latitudinal range.

These functions are adapted from previously published paleobiological methods. Circular subsampling follows the assemblage-based subsampling framework of Antell et al. (2020), wherein the user can specify the number of subsampling iterations, radius of a subsampling region, number of sites, and whether to select sites at random or with weighted probability to tighten their spatial aggregation. Nearest-neighbor subsampling modifies the procedure of Close et al. (2020), wherein the user can specify the number of subsampling iterations, maximum distance across a subsampling region (spanning tree), and, optionally, number of sites. Site rarefaction was not conducted in the original study but is added as a feature in divvy, for comparability with the other methods and in keeping with the theory described earlier, to standardize both area/sites and dispersion of occurrences. The third subsampling method, rarefaction of sites within latitudinal bands, has precedent in Marcot et al. (2016). The overall extent of latitudinal bins is unequal; there is more available surface in equally spaced latitudinal bands near the equator due to the spherical shape of the Earth, and longitudinal distributions of sampling also differ between bands in most empirical cases. Rarefaction within latitudinal bands accounts for only the area/site count and not longitudinal dispersion of subsampled localities. However, given the prevalence of paleobiological studies investigating gradients across latitudinal bins, site rarefaction alone represents an important improvement in standardization.

Additional functions available in divvy (v. 1.0.0) include *uniqify* to subset an occurrence dataset to unique taxon-coordinate combinations, *sdSumry* to calculate basic spatial coverage and diversity metadata for a dataset or its subsamples, *rangeSize* to calculate five measures of geographic range size, and *classRast* to generate a raster containing the most common environment or trait for point occurrences falling in each grid cell. These helper functions are designed to assist in basic data exploration, formatting, and analysis, regardless of any spatial subsampling. For instance, the analytic script to generate Figures 1 and 2 uses *uniqify* and *sdSumry* to quickly compute species count, total sampled sites, and dispersion of sampled sites in each time bin of the non-standardized full dataset.

There are several vignettes published as part of divvy and available through the package index at the Comprehensive R Archive Network (CRAN) https://cran.r-project.org/web/packages/divvy/index.html or the package website https://gawainantell.github.io/divvy. The subsampling tutorial gives rationale, practical considerations, and code demonstrations for three types of subsampling on one of the PBDB datasets included in the package. The case study compares geographic range size between different marine environments and ecological groups of Silurian brachiopods. The conceptual walkthrough illustrates subsampling steps with diagrams. Together, these articles give example code in empirical contexts for all divvy functions as complements to the short, abstract examples included in the help documentation.

Discussion

State of the Field

Because of the spatial structure in biodiversity itself, differential geographic coverage of observations—whether of living species or fossil taxa—will always bias ecological comparisons between time intervals or world regions if left unmitigated. Geological and human sampling processes further distort inferences of biodiversity distributions. These influences cannot be addressed adequately by the inclusion of rarefaction or other nonspatial standardization procedures during analysis (Bush et al. 2004; Close et al. 2020). Explicit spatial standardization is necessary to discern truthful information about past ecosystems.

The magnitude and direction of bias from non-uniform spatial coverage varies with context. There could exist cases where estimated spatial sampling signal is weaker than the theorized signal of the primary phenomenon of interest, in which instances, analysts could justifiably disregard heterogeneous geographic sampling. Nevertheless, it is prudent in every study to consider the possible ways and extent to which variable spatial coverage could affect results and inferences. When analyses are repeated with spatial standardization, the outcome may not only be adjustment of point estimates or refinement of confidence intervals, but reversals or nullifications of the biggest conclusions, as in the case study presented earlier.

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

A proliferation of negative results from spatially standardized data is perhaps unsurprising. The primacy of sampling signal in raw occurrence data, coupled with the pressure to publish positive results, potentially means that many findings in the paleobiological literature could reflect biases that are misinterpreted as biological signal. Conversely, it is possible some past conclusions will be strengthened after controlling for spatial structure of sampling. Considering the ample potential for negative results, routine use of power analyses may aid interpretation by estimating the probability that analysis would be able to detect a true signal if present, given empirical data size and distribution.

Not only are results based on unstandardized data potentially misleading or outright wrong, but unqualified visualizations of global curves are uninterpretable. Due to geological and human sampling distributions and the species–area effect, any global diversity curve reflects not only diversity but also area and dispersion of observation. When a global diversity curve is presented without additional information about its compartmentalization across spatial regions, it is impossible to deduce how much of the global richness at any given time step was contributed by original biological diversity and how much by variation in the completeness of knowledge about the fossil record (Raup 1976; Vilhena and Smith 2013; Benson et al. 2021). Analogous problems occur with global curves of other parameters besides diversity, for example, diversification and extinction rates (Allen et al. 2023) and proxy temperature averages (Jones and Eichenseer 2022).

Fortunately, improvements in data access and computing capacity in recent decades have enabled the development and distribution of spatial-standardization analysis tools. Multiple research teams have designed methods to account for unequal area and dispersion of taxon occurrences between time steps or other comparison categories, several of which are formalized in the new R package divvy (Antell 2023). As access to software for spatial standardization and other standardization methods improves, more research teams might reevaluate foundational research questions in quantitative paleobiology with tighter statistical control. Given that global curves have been a staple of the discipline and feature prominently in discussion of the diversification of life on Earth (e.g., Sepkoski et al. 1981; Alroy 2010), broad adoption of spatial standardization in future studies might impel far-ranging revisions for long-standing assumptions about patterns and processes in paleoecology and evolutionary biology.

Development of Data Collection, Analysis, and Reporting

We must confront the uneven spatial distributions of recorded paleontological knowledge not only scientifically through statistical means but also societally through material means: subsampling can work around data gaps but cannot fill them. Sustained investment in Western science in many former colonizing countries has generated overproportionate quantities of fossil data for a minority of Earth's surface, primarily in the Global North and at the expense of the Global South (Rodney 2018; Raja et al. 2022). Generating a truly global map of fossil biodiversity will require specimen repatriation to former colonized countries, reparations to support scientific capacity-building, decolonization of access to literature, de-Anglicization of publishing, and accreditation of traditional knowledge and classification systems (Liboiron 2021; Nuñez et al. 2021). Resultant data should be stewarded under the FAIR guiding principles of open access (i.e., Findable, Accessible, Interoperable, and Reusable) and in accordance with CARE Principles of Indigenous Data Governance (i.e., managed for Collective benefit on intergenerational timescales, respecting sovereign Authority to control access, practicing Responsibility towards Indigenous worldviews and relationships, and following Ethics for minimizing harm and ensuring justice in future use; https://www.gida-global.org/care) (Wilkinson et al. 2016; Carroll et al. 2020; Jennings et al. 2023).

With respect to data analysis, underdeveloped methods include (1) the treatment of abundance data within sites, (2) rarefaction of data lists such as references and collections, and (3) refining spatial subsampling methods for sensitivity to changing geographic configurations of bioregions and continents through time. First, harmonizing the disparate practices for reporting abundance data (e.g., as counts, proportions, or ordinal values) will enable big data analyses to equalize the frequency distribution coverage of taxa within each site. This step might prove a necessary addition to controlling the number and dispersal of sites, to standardize alpha diversity more completely. Whether or not sampling bias would be further reduced with rarefaction of the number of data citations or sources (e.g., references and collections in PBDB data) is unclear. Studies that have rarefied these data lists within regionally constrained subsamples (e.g., Marcot 2014; Close et al. 2017; Womack et al. 2021) omitted rarefaction of sites, while work that rarefied sites omitted rarefaction of data lists (e.g., Antell et al. 2020). As discussed earlier in "Collections and Other Data Lists," sites and lists may be partially redundant data structures. However, it is possible that spatial standardization could be refined by thoughtful application of rarefaction for both sites and lists. When abundance data are lacking but large disparities in collector effort exist across sites, standardizing publication counts might serve as a particularly salient addition to spatial standardization.

A third issue unaddressed by major subsampling approaches is (paleo)continental configuration. It remains an open question whether sensitivity to the geography of coastlines and major biogeographic barriers is desirable for spatial subsampling. At present, the three subsampling routines implemented in divvy are agnostic to global geography. Until there are widely adopted, objective, automated methods to partition bioregions, the choice of where to draw uncrossable limits for regional subsamples will involve manual analytic choices (e.g., Close et al. 2017). This challenge is heightened by the extremeness of continental reconfiguration that has occurred over Phanerozoic-scale study intervals: entire oceans and seaways have opened and closed over that time span, preventing any through-ranging analysis of marine biodiversity tracked in the same regions.

It is important for the fidelity of future studies that the paleobiology community continue to discuss and converge upon data-processing and analytic standards, including but not limited to spatial subsampling methods. The larger the number of possible methods to analyze a dataset, the less likely an individual reviewer will have enough familiarity with the specific method used to provide technical critique, and the more opportunities (whether unconscious or accidental) for authors to select a method that generates results in line with expected answers (Ioannidis 2005; Simmons et al. 2011). Analysts should always select methodological approaches to address research questions in the most direct and robust ways, without prejudice against negative or inconclusive results. Another safeguard against misplaced conclusions is to run many variations of an analysis and note how sensitive the results may be to methodological choices. Finally, we should welcome null findings, circumspect conclusions, and corrections to

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

prior publications as valuable and generative contributions to paleobiological knowledge.

At the publication stage of projects, there are evidence-backed strategies that research communities can employ to strengthen the impartiality and transparency of reported findings. Trials in other scientific disciplines also offer lessons about interventions unlikely to change publishing culture—for example, education campaigns about scientific integrity have yet to demonstrate clear empirical benefits (Marusic et al. 2016). Similarly, raising expectations of reviewers and editorial teams is unrealistic as a solution (Nosek et al. 2012). Concrete guidance in the form of reviewer checklists might prove more effective, for instance, with items such as reporting of sample size, geographic data coverage, effect size, standardization procedures, and results repeated when excluded data points are included (Simmons et al. 2011; Nosek et al. 2012). Additionally, journals could shift publication standards from perceived importance to explicit criteria of soundness, especially given the poor track record of peer review at identifying importance (for a review, see Nosek et al. 2012).

Call to Action

Some of the earliest writings in quantitative paleobiology demonstrated the need to correct synoptic diversity curves for spatial heterogeneity of sampling through time (Gregory 1955; Raup 1976). Now, more than half a century later, the unprecedented availability of fossil occurrence data and computational infrastructure has actualized the possibility of doing so. Adopting spatial standardization as a routine component of analysis is a grand challenge for quantitative paleobiology (Dillon et al. 2023) but also a grand opportunity to make the field more truthful, more reproducible, and more credible to ecologists, conservation biologists, and practitioners who draw on life sciences findings to inform policy.

Acknowledgments. The Natural Environment Research Council (grant NE/V011405/1; E.E.S., G.T.A, R.B.J.B), a Leverhulme Prize (E.E.S.), and University of California President's Postdoctoral Fellowship Program (G.T.A.) supported this work. A. Dunhill and an anonymous reviewer contributed considerate feedback on the article. The authors thank the University of California library system for assisting with publication fees.

Competing Interest. The authors declare no competing interests.

Data Availability Statement. The divvy software package (version 1.0.0) can be installed from the Comprehensive R Archive Network (CRAN; https://cran.r-project.org/web/packages/divvy) as binaries or with the R command `install.packages("divvy")`. The development version is available to install from GitHub at https://github.com/GawainAntell/divvy. All data analyzed in the case study are archived at Zenodo (Antell et al. 2019) and maintained at https://github.com/GawainAntell/EcoRelease, along with the script to generate results herein. Analyses were programmed in the R statistical computing environment, v. 4.3.1 (R Core Team 2023).

Literature Cited

Aguilée, R., F. Gascuel, A. Lambert, and R. Ferriere. 2018. Clade diversification dynamics and the biotic and abiotic controls of speciation and extinction rates. *Nature Communications* **9**:3013.

Allen, B. J., M. E. Clapham, E. E. Saupe, P. B. Wignall, D. J. Hill, and A. M. Dunhill. 2023. Estimating spatial variation in origination and extinction in deep time: a case study using the Permian–Triassic marine invertebrate fossil record. *Paleobiology* **49**:509–526.

Allison, P. A., and D. E. Briggs. 1993. Paleolatitudinal sampling bias, Phanerozoic species diversity, and the end-Permian extinction. *Geology* **21**:65–68.

Alroy, J. 2000. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* **26**:707–733.

Alroy, J. 2010. The shifting balance of diversity among major marine animal groups. *Science* **329**:1191–1194.

Alroy, J., C. Marshall, R. Bambach, K. Bezusko, M. Foote, F. Fürsich, T. Hansen, S. Holland, L. Ivany, and D. Jablonski. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences USA* **98**:6261–6266.

Alroy, J., M. Aberhan, D. J. Bottjer, M. Foote, F. T. Fürsich, P. J. Harries, A. J. Hendy, S. M. Holland, L. C. Ivany, and W. Kiessling. 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321**:97–100.

Altman, D. G., and J. M. Bland. 1995. Statistics notes: absence of evidence is not evidence of absence. *BMJ* **311**:485.

Antell, G. T. 2023. divvy: spatial subsampling of biodiversity occurrence data, R package version 1.0.0. https://cran.r-project.org/web/packages/divvy.

Antell, G. T., W. Kiessling, M. Aberhan, and E. E. Saupe. 2019. Data and code for "Marine biodiversity and geographic distributions are independent on large scales" (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.3491853.

Antell, G. T., W. Kiessling, M. Aberhan, and E. E. Saupe. 2020. Marine biodiversity and geographic distributions are independent on large scales. *Current Biology* **30**:115–121.e5.

Barnosky, A. D., M. A. Carrasco, and E. B. Davis. 2005. The impact of the species–area relationship on estimates of paleodiversity. *PLoS Biology* **3**:e266.

Beck, J., M. Böller, A. Erhardt, and W. Schwanghart. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* **19**:10–15.

Behrensmeyer, A. K., and S. M. Kidwell. 1985. Taphonomy's contributions to paleobiology. *Paleobiology* **11**:105–119.

Benson, R. B., R. Butler, R. A. Close, E. Saupe, and D. L. Rabosky. 2021. Biodiversity across space and time in the fossil record. *Current Biology* **31**:R1225–R1236.

Benton, M. J., and B. C. Emerson. 2007. How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaeontology* **50**:23–40.

Boakes, E. H., P. J. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O'Connor, and G. M. Mace. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology* **8**:e1000385.

Bolker, B. M. 2008. *Ecological models and data in R*. Princeton University Press, Princeton, N.J.

Brocklehurst, N. 2015. A simulation-based examination of residual diversity estimates as a method of correcting for sampling bias. *Palaeontologia Electronica* **18**(3):1–15.

Bush, A. M., and R. K. Bambach. 2004. Did alpha diversity increase during the Phanerozoic? Lifting the veils of taphonomic, latitudinal, and environmental biases. *Journal of Geology* **112**:625–642.

Bush, A. M., M. J. Markey, and C. R. Marshall. 2004. Removing bias from diversity curves: the effects of spatially organized biodiversity on sampling-standardization. *Paleobiology* **30**:666–686.

Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, and R. Rowe. 2020. The CARE principles for Indigenous Data Governance. *Data Science Journal* **19**:1–12.

Champely, S. 2020. pwr: basic functions for power analysis, R package version 1.3-0. https://cran.r-project.org/web/packages/pwr.

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

Close, R., R. B. Benson, E. Saupe, M. Clapham, and R. Butler. 2020. The spatial structure of Phanerozoic marine animal diversity. *Science* **368**:420–424.

Close, R. A., R. B. Benson, P. Upchurch, and R. J. Butler. 2017. Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. *Nature Communications* **8**:15381.

Close, R. A., S. W. Evers, J. Alroy, and R. J. Butler. 2018. How should we estimate diversity in the fossil record? Testing richness estimators using

sampling-standardised discovery curves. *Methods in Ecology and Evolution* **9**:1386–1400.

Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species-area relationship. *American Naturalist* **113**:791–833.

Deng, Y., J. Fan, Y. Wang, Y. Shi, J. Yang, and Z. Lu. 2020. Current status of paleontological databases and data-driven research in paleontology. *Acta Metallurgica Sinica* **26**(4):361-383.

Dillon, E. M., E. M. Dunne, T. M. Womack, M. Kouvari, E. Larina, J. R. Claytor, A. Ivkić, M. Juhn, P. S. M. Carmona, and S. V. Robson. 2023. Challenges and directions in analytical paleobiology. *Paleobiology* **49**:377–393.

Dunhill, A. M., B. Hannisdal, N. Brocklehurst, and M. J. Benton. 2018. On formation-based sampling proxies and why they should not be used to correct the fossil record. *Palaeontology* **61**:119–132.

Dunne, E. M., S. E. Thompson, R. J. Butler, J. Rosindell, and R. A. Close. 2023. Mechanistic neutral models show that sampling biases drive the apparent explosion of early tetrapod diversity. *Nature Ecology and Evolution* **7**:1480–1489.

Finnegan, S., S. C. Anderson, P. G. Harnik, C. Simpson, D. P. Tittensor, J. E. Byrnes, Z. V. Finkel, D. R. Lindberg, L. H. Liow, and R. Lockwood. 2015. Paleontological baselines for evaluating extinction risk in the modern oceans. *Science* **348**:567–570.

Foote, M. 2023. Diversity-dependent diversification in the history of marine animals. *American Naturalist* **201**:680–693.

Foote, M., J. S. Crampton, A. G. Beu, B. A. Marshall, R. A. Cooper, P. A. Maxwell, and I. Matcham. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science* **318**:1131–1134.

Furness, E. N., E. E. Saupe, R. J. Garwood, P. D. Mannion, and M. D. Sutton. 2023. The jigsaw model: a biogeographic model that partitions habitat heterogeneity from area. *Frontiers of Biogeography* **15**(1). https://doi.org/10.21425/F5FBG58477.

Green, P., and C. J. MacLeod. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* **7**:493–498.

Gregory, J. T. 1955. Vertebrates in the geologic time scale. *Geological Society of America Special Paper* **62**:593–608.

Harnik, P. G., C. Simpson, and J. L. Payne. 2012. Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society of London B* **279**:4969–4976.

Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **46**:523–549.

Ioannidis, J. P. 2005. Why most published research findings are false. *PLoS Medicine* **2**:e124.

Jennings, L., T. Anderson, A. Martinez, R. Sterling, D. D. Chavez, I. Garba, M. Hudson, N. A. Garrison, and S. R. Carroll. 2023. Applying the "CARE Principles for Indigenous Data Governance" to ecology and biodiversity research. *Nature Ecology and Evolution* **7**:1547–1551.

Jones, L. A., and K. Eichenseer. 2022. Uneven spatial sampling distorts reconstructions of Phanerozoic seawater temperature. *Geology* **50**:238–242.

Jones, L. A., C. D. Dean, P. D. Mannion, A. Farnsworth, and P. A. Allison. 2021. Spatial sampling heterogeneity limits the detectability of deep time latitudinal biodiversity gradients. *Proceedings of the Royal Society of London B* **288**:20202762.

Jones, L. A., W. Gearty, B. J. Allen, K. Eichenseer, C. D. Dean, S. Galván, M. Kouvari, P. L. Godoy, C. S. Nicholl, and L. Buffan. 2023. palaeoverse: a community-driven R package to support palaeobiological analysis. *Methods in Ecology and Evolution* **14**:2205–2215.

Kiessling, W. 2005. Habitat effects and sampling bias on Phanerozoic reef distribution. *Facies* **51**(1–4):24–32.

Kimmel, K., M. L. Avolio, and P. J. Ferraro. 2023. Empirical evidence of widespread exaggeration bias and selective reporting in ecology. *Nature Ecology and Evolution* **7**:1525–1536.

Kocsis, Á. T., C. J. Reddin, J. Alroy, and W. Kiessling. 2019. The R package divDyn for quantifying diversity dynamics using fossil sampling data. *Methods in Ecology and Evolution* **10**:735–743.

Liboiron, M. 2021. Decolonizing geoscience requires more than equity and inclusion. *Nature Geoscience* **14**:876–877.

Loke, L. H., and R. A. Chisholm. 2022. Measuring habitat complexity and spatial heterogeneity in ecology. *Ecology Letters* **25**:2269–2288.

MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, N.J.

Marcot, J. D. 2014. The fossil record and macroevolutionary history of North American ungulate mammals: standardizing variation in intensity and geography of sampling. *Paleobiology* **40**:238–255.

Marcot, J. D., D. L. Fox, and S. R. Niebuhr. 2016. Late Cenozoic onset of the latitudinal diversity gradient of North American mammals. *Proceedings of the National Academy of Sciences USA* **113**:7189–7194.

Marusic, A., E. Wager, A. Utrobicic, H. R. Rothstein, and D. Sambunjak. 2016. Interventions to prevent misconduct and promote integrity in research and publication. *Cochrane Database of Systematic Reviews*, **4**, MR000038. https://doi.org/10.1002/14651858.MR000038.pub2.

Matthews, T. J., F. Guilhaumon, K. A. Triantis, M. K. Borregaard, and R. J. Whittaker. 2016. On the form of species–area relationships in habitat islands and true islands. *Global Ecology and Biogeography* **25**:847–858.

Menegotto, A., and T. F. Rangel. 2018. Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. *Nature Communications* **9**:4713.

Monarrez, P. M., J. B. Zimmt, A. M. Clement, W. Gearty, J. J. Jacisin, K. M. Jenkins, K. M. Kusnerik, A. W. Poust, S. V. Robson, and J. A. Sclafani. 2022. Our past creates our present: a brief overview of racism and colonialism in Western paleontology. *Paleobiology* **48**:173–185.

Moudrý, V., and R. Devillers. 2020. Quality and usability challenges of global marine biodiversity databases: an example for marine mammal data. *Ecological Informatics* **56**:101051.

Nanglu, K., and T. M. Cullen. 2023. Across space and time: a review of sampling, preservational, analytical, and anthropogenic biases in fossil data across macroecological scales. *Earth-Science Reviews* **244**:104537.

Nosek, B. A., J. R. Spies, and M. Motyl. 2012. Scientific utopia. II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science* **7**:615–631.

Nuñez, M. A., M. C. Chiuffo, A. Pauchard, and R. D. Zenni. 2021. Making ecology really global. *Trends in Ecology and Evolution* **36**:766–769.

Nürnberg, S., and M. Aberhan. 2013. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology* **39**:360–372.

Oliveira, U., A. P. Paglia, A. D. Brescovit, C. J. de Carvalho, D. P. Silva, D. T. Rezende, F. S. F. Leite, J. A. N. Batista, J. P. P. P. Barbosa, and J. R. Stehmann. 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. *Diversity and Distributions* **22**:1232–1244.

Patzkowsky, M. E., and S. M. Holland. 2012. *Stratigraphic paleobiology*. University of Chicago Press, Chicago.

Pie, M. R., R. Divieso, and F. S. Caron. 2023. Clade density and the evolution of diversity-dependent diversification. *Nature Communications* **14**:4576.

Preston, F. W. 1962. The canonical distribution of commonness and rarity. Part I. *Ecology* **43**:185–215.

Rabosky, D. L. 2013. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Annual Review of Ecology, Evolution, and Systematics* **44**:481–502.

Raja, N. B., E. M. Dunne, A. Matiwane, T. M. Khan, P. S. Nätscher, A. M. Ghilardi, and D. Chattopadhyay. 2022. Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature Ecology and Evolution* **6**:145–154.

Raup, D. M. 1976. Species diversity in the Phanerozoic: an interpretation. *Paleobiology* **2**:289–297.

Raup, D. M. 1977. Systematists follow the fossils. *Paleobiology* **3**:328–329.

R Core Team. 2023. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austra. https://www.R-project.org.

Reitan, T., T. H. Ergon, and L. H. Liow. 2022. Relative species abundance and population densities of the past: developing multispecies occupancy models for fossil data. *Paleobiology* **49**:23–38.

Rodney, W. 2018. *How Europe underdeveloped Africa*, new ed. Verso Books, Brooklyn, N.Y.

Roughgarden, J. 1972. Evolution of niche width. *American Naturalist* **106**:683–718.

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

Sakamoto, M., C. Venditti, and M. J. Benton. 2017. "Residual diversity estimates" do not correct for sampling bias in palaeodiversity data. *Methods in Ecology and Evolution* **8**:453–459.

Sepkoski, J. J., Jr. 1976. Species diversity in the Phanerozoic: species-area effects. *Paleobiology* **2**:298–303.

Sepkoski, J. J., Jr., R. K. Bambach, D. M. Raup, and J. W. Valentine. 1981. Phanerozoic marine diversity and the fossil record. *Nature* **293**:435–437.

Shinozaki, K. 1963. Notes on the species-area curve. *Proceedings of the Tenth Annual Meeting of the Ecological Society of Japan*. Tokyo, Japan.

Signor, P. W., J. H. Lipps, L. Silver, and P. Schultz. 1982. Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. *In* L. T. Silver and P. H. Schultz, eds. *Geological implications of impacts of large asteroids and comets on the Earth. Geological Society of America Special Paper* 190:291–296.

Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* **22**:1359–1366.

Smith, A. B. 2001. Large–scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society of London B* **356**:351–367.

Smith, D. R., I. C. Hardy, and M. P. Gammell. 2011. Power rangers: no improvement in the statistical power of analyses published in *Animal Behaviour*. *Animal Behaviour* **1**:347–352.

Smith, J. A., G. P. Dietl, and S. R. Durham. 2020. Increasing the salience of marine live–dead data in the Anthropocene. *Paleobiology* **46**:279–287.

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

Vilhena, D. A., and A. B. Smith. 2013. Spatial bias in the marine fossil record. *PLoS ONE* **8**:e74470.

Whittaker, R. H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* **30**:279–338.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. *Taxon* **21**:213–251.

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, and P. E. Bourne. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**(1):1–9.

Womack, T. M., J. S. Crampton, and M. J. Hannah. 2021. Spatial scaling of beta diversity in the shallow-marine fossil record. *Paleobiology* **47**:39–53.

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

Appendix

Appendix Table A1. Time bins to divide the global Phanerozoic dataset ($n = 63$). A species' occurrence record was included in a time bin if the name or age of both its maximum and minimum occurrence estimates fell within the onset and terminus ages (in Ma) for the bin. During binning, ages were rounded to the nearest 0.01 Ma for boundaries younger than 10 Ma, 0.1 Ma for boundaries 50–150 Ma, or 1 Ma for boundaries older than 150 Ma. The number of unique occurrences for each species is tallied for each time bin. Replicated from table S1 in Antell et al. (2020).

Name	Onset	Terminus	Duration	Occurrences
Tremadocian, Floian	485.4	470.0	15.4	196
Dapingian, Darriwilian	470.0	458.4	11.6	347
Sandbian, Katian	458.4	445.2	13.2	1226
Hirnantian	445.2	443.8	1.4	394
Rhuddanian, Aeronian	443.8	438.5	5.3	411
Telychian	438.5	433.4	5.1	337
Sheinwoodian, Homerian	433.4	427.4	6.0	1072
Gorstian, Ludfordian	427.4	423.0	4.4	924
Pridoli	423.0	419.2	3.8	313
Lochkovian, Pragian	419.2	407.6	11.6	1141
Emsian, Eifelian	407.6	387.7	19.9	914
Givetian	387.7	382.7	5.0	1293
Frasnian	382.7	372.2	10.5	937
Famennian	372.2	358.9	13.3	371
Tournaisian	358.9	346.7	12.2	641
Visean	346.7	330.9	15.8	607
Serpukhovian, Bashkirian	330.9	315.2	15.7	344
Moscovian, Kasimovian	315.2	303.7	11.5	561
Gzhelian	303.7	298.9	4.8	613
Asselian, Sakmarian	298.9	290.1	8.8	1716
Artinskian, Kungurian	290.1	272.9	17.2	2937
Roadian	272.9	268.8	4.1	1596
Wordian	268.8	265.1	3.7	1621
Capitanian	265.1	259.1	6.0	1342
Wuchiapingian	259.1	254.1	5.0	2451
Changhsingian	254.1	251.9	2.2	2411
Induan, Olenekian	251.9	247.2	4.7	430
Anisian	247.2	242.0	5.2	782
Ladinian	242.0	237.0	5.0	284
Carnian	237.0	227.0	10.0	503
Norian	227.0	208.5	18.5	689
Rhaetian	208.5	201.3	7.2	481
Hettangian, Sinemurian	201.3	190.8	10.5	581
Pliensbachian	190.8	182.7	8.1	1296
Toarcian	182.7	174.1	8.6	866
Aalenian, Bajocian	174.1	168.3	5.8	588
Bathonian	168.3	166.1	2.2	659
Callovian	166.1	163.5	2.6	719
Oxfordian	163.5	157.3	6.2	437

(*Continued*)

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

Appendix Table A1. (*Continued.*)

Name	Onset	Terminus	Duration	Occurrences
Kimmeridgian	157.3	152.1	5.2	304
Tithonian	152.1	145.0	7.1	486
Berriasian	145.0	139.8	5.2	317
Valanginian, Hauterivian,	139.8	125.0	14.8	656
Barremian				
Aptian	125.0	113.0	12.0	509
Albian	113.0	100.5	12.5	522
Cenomanian	100.5	93.9	6.6	590
Turonian	93.9	89.8	4.1	187
Coniacian, Santonian	89.8	83.6	6.2	709
Campanian	83.6	72.1	11.5	422
Maastrichtian	72.1	66.0	6.1	3197
Danian	66.0	61.6	4.4	363
Selandian, Thanetian	61.6	56.0	5.6	272
Ypresian	56.0	47.8	8.2	661
Lutetian, Bartonian	47.8	37.8	10.0	1893
Priabonian	37.8	33.9	3.9	993
Rupelian	33.9	27.8	6.1	773
Chattian, Aquitanian	27.8	20.4	7.4	632
Burdigalian	20.4	16.0	4.5	817
Langhian, Serravallian	16.0	11.6	4.3	2056
Tortonian, Messinian	11.6	5.3	6.3	1760
Zanclean	5.3	3.6	1.7	934
Piacenzian	3.6	2.6	1.0	1333
Pleistocene	2.6	0.0	2.6	2204

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press

Appendix Table A2. Kendall's τ (tau) coefficient estimates and 95% quantiles (from 500 subsamples) for pairwise nonparametric correlations between subsampled species count, mean occupied grid cells (excluding singly occurring species, and out of 12 cells in a subsample), and aggregation of sampling sites (summed length of minimum spanning tree connecting subsampled cell centroids). In each correlation, the predictor time series was pre-whitened with a first-order autoregressive model; the residuals of this model were correlated with the response series to account for temporal autocorrelation.

	Correlation with predictor time series			
	Species count		Sampling aggregation	
Response series	Mean	95%	Mean	95%
Mean occupancy	−0.11	(−0.25, 0.01)	−0.17	(−0.32, −0.03)
Species count	1.00	(1.00, 1.00)	−0.01	(−0.14, 0.12)

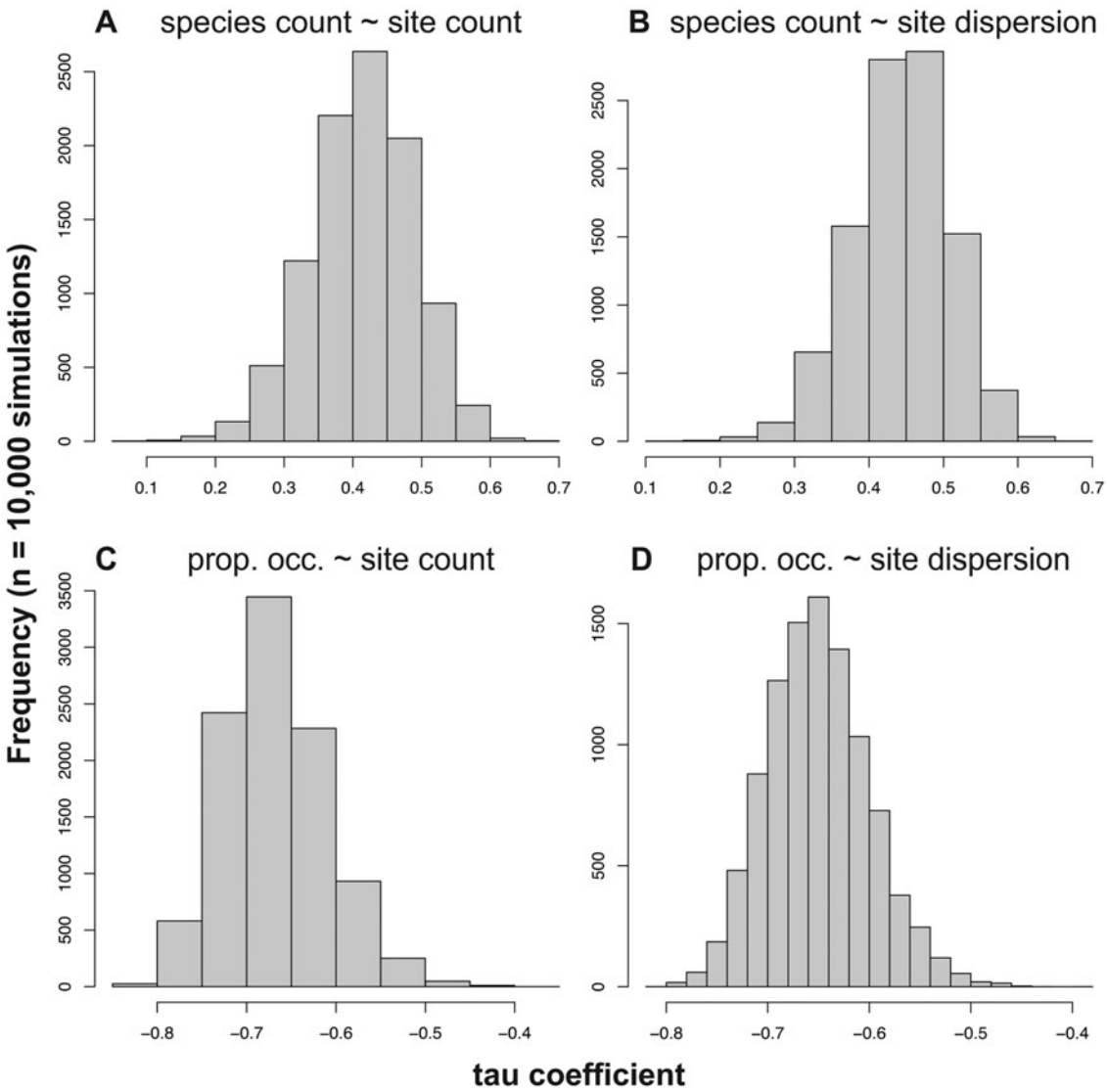


Figure A1. Kendall's τ (tau) coefficient distributions for correlations between (A) global species count and total occupied cells (sampling area), (B) global species count and summed minimum spanning tree length between occupied cells (sampling dispersion), (C) mean proportional species occupancy and sampling area, and (D) mean proportional occupancy and sampling dispersion. Figure 4 panels plot the corresponding scatter plot for each correlation.

https://doi.org/10.1017/pab.2023.36 Published online by Cambridge University Press