



Preservation is predictable: quantifying the effect of taphonomic biases on ecological disparity in birds

Jonathan S. Mitchell

Abstract.—Evolutionary inferences from fossil data often require accurately reconstructing differences in richness and morphological disparity between fossil sites across space and time. Biases such as sampling and rock availability are commonly accounted for in large-scale studies; however, preservation bias is usually dealt with only in smaller, more focused studies. Birds represent a diverse, but taphonomically fragile, group commonly used to infer environmental conditions in recent (Pleistocene and later) fossil assemblages, and their relative scarcity in the fossil record has led to controversy over the timing of their radiation. Here, I use simulations to show how even weak taphonomic biases can distort estimates of richness, and render variance sensitive to sample size. I then apply an ecology-based filtering model to recent bird assemblages to quantify the distortion induced by taphonomy. Certain deposit types, such as caves, show less evidence of taphonomic distortion than others, such as fluvial and lacustrine deposits. Archaeological middens unsurprisingly show some of the strongest evidence for taphonomic bias, and they should be avoided when reconstructing Pleistocene and early Holocene environments. Further, these results support previously suggested methods for detecting fossil assemblages that are relatively faithfully preserved (e.g., presence of difficult-to-preserve taxa), and I use these results to recommend that future large-scale studies include facies diversity along with metrics such as rock volume, or compare only sites with similar taphonomic histories.

Jonathan S. Mitchell. *Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois U.S.A., 60637. E-mail: mitchelljs@uchicago.edu*

Accepted: 13 November 2014
Published online: 24 February 2015

Introduction

Fossil assemblages vary dramatically in composition and disparity across time and space, yet each fossil assemblage represents a biased picture of the living community from which it is derived (Johnson 1960; Lawrence 1968; Raup 1975; Valentine 1989; Behrensmeyer et al. 2000). Without accounting for the biases, direct comparisons between assemblages may reflect differences in preservation or sampling more than they reflect biological differences, obscuring the true history of change in clades and ecosystems. For clades with a high variance in preservation potential (e.g., those that occupy many different habitats, differ dramatically in body size or shell thickness) preservation bias will have a strong effect, as the highly preservable members are the most likely drivers of the patterns observed in the fossil record, even if they had lower abundances or were species-poor.

Birds are a species-rich, ecologically diverse clade and birds occur on every modern continent and in every habitat save the deep

sea. However, because of their typically small size, hollow bones, and lack of teeth, their fossil record is poorly resolved compared to other groups (Turvey and Blackburn 2011), although it is better than many realize and is improving rapidly (Ksepka and Boyd 2012). As a result of the low preservation potential for most birds, fossil sites that preserve a rich avifauna are usually either taphonomically exceptional deposits (Lagerstätten) or composed mostly of low-pneumaticity, large-bodied “water birds.” The patchiness of the bird record results in low stratigraphic sampling and poor resolution, and has rendered the timing of the origins of major bird clades highly contentious (Cooper and Penny 1997; Brown et al. 2008; Hackett et al. 2008; Pacheco et al. 2011; Ksepka and Boyd 2012; Ksepka et al. 2014).

Ecological preservation bias is known to have a strong effect on which members of clades and ecosystems successfully fossilize even in highly preservable groups such as mammals (Damuth 1982), bivalves (Kidwell et al. 2005), and brachiopods (Holland 2003).

Accounting for differences in worker effort (Raup 1975; Alroy 2010) and rock outcrop exposure (Sepkoski et al. 1981; Peters and Heim 2010) between fossil assemblages is widely acknowledged as necessary for making biologically meaningful comparisons through time and across space. Nonetheless, because it is difficult to quantify preservation potential (we cannot know the “true” assemblage’s composition for prehistoric sites), preservation bias has been examined primarily in small-scale comparisons (although see Behrensmeier et al. 2005; Turvey and Blackburn 2011; Kosnik et al. 2011). Large-scale analyses have generally excluded aberrant preservation modes (Alroy 2010), focused on sites that retain taphonomic control (“bellwether”) taxa that are difficult to preserve (Bottjer and Jablonski 1988), or ignored the distinction of preservation bias and focused only on rock volume and sampling, implicitly assuming that preservation will only add noise rather than obscure biological patterns (e.g., Brusatte et al. 2014; Newham et al. 2014).

Through simulations, I show that even weak preservational biases will strongly mislead estimates of relative disparity and richness when the type or strength of bias varies among sites (i.e., when sites with different taphonomic histories are pooled together). Accounting for differing biases among sites is crucial for any study that seeks to accurately detect biological differences across space and time, rather than spurious differences due to preservation. By comparing extremely young fossil, subfossil, and recent death assemblages with modern avian communities from the same regions of North America, I demonstrate that ecological biases in preservation are ubiquitous within Aves. These biases render cross-taphonomic regime comparisons of both richness and disparity suspect. A key result is that variance in preservation potential (a function of variance in ecological traits and variance in bias strength) is the critical factor, suggesting that these qualitative results are extendable to all groups, even as the strength of the biases varies.

Methods

Simulations.—Quantifying the effect of biases in preservation on estimates of taxonomic

richness and morphological disparity requires knowing the original community composition, the strength and type of bias, and the resulting fossilized community composition. Before attempting to handle empirical data, I used simulations to assess whether or not bias had a measurable effect, even at low strengths.

To understand how taphonomy affects estimates of means and variances of ecological data in fossil assemblages, I simulated a “living assemblage” with 1000 species, and “fossilized” it by randomly sampling pools of species ranging from 10 to 1000 species. For each species in the simulation, I generated a random morphological value from a normal distribution arbitrarily centered on 10 with a standard deviation of 2. Preservation probability was modeled as a single variable that exponentially decreased with the ecological trait at a rate (λ) of 1, to simulate a single axis of preservation (e.g., body size, or propensity to live in forests). For each iteration (from 10 to 1000 species) the species were selected stochastically, with the probability a given species was preserved determined by its trait value and the exponential decay distribution.

The choice of simulating an exponential decay relationship between ecology and preservation was arbitrary, and any decay distribution will give a qualitatively similar result, with the degree of bias being controlled by the variance in preservation potential (a function of the morphological variance and decay parameters). Non-decay models could also be appropriate; for instance, Miller et al. (2013) explored the use of a logistic regression of traits to predict preservation probability. These models will produce different patterns of bias, but the qualitative result of the fossilized subsample being distorted relative to the true living community will remain, albeit at varying strengths.

To understand how taphonomy affects comparisons of richness among assemblages, I simulated 500 “living” communities with 50 species each, where abundance was log-normally distributed, as in most extant assemblages (although see Williamson and Gaston 2005). Again, each species received a random, normally distributed trait (mean = 10, SD = 2), and again taphonomic bias was modeled as an

exponential function of that trait (rate=1). Rather than sample species as above, however, these simulations sampled individuals, such that a highly abundant, and highly preservable species could constitute the entire “fossil assemblage.” These simulated communities were subjected to rarefaction and Shareholder's Quorum Sampling (SQS; Alroy 2010). The simulated results were then compared with empirical data on avian dead assemblages (= extant mass mortality, archaeological middens, and true death assemblages) from the Aleutian Islands (see below).

Bird Assemblage Data.—In order to assess how biases in preservation affect real data sets, I gathered ecological data on living birds, and compiled data on bird occurrences in both the modern world and recent stratigraphic record. Differences between the modern and subfossil/fossil bird assemblages are the result of both biological responses (e.g., to climate change, to humans) and taphonomic filters; however, by using extremely young subfossil/fossil deposits, and focusing on the taphonomically fragile clade Aves, I argue that the differences observed predominantly originated from taphonomic filters (e.g., 500 years ago Alaskan bird assemblages were not composed solely of sturdy-boned, large-bodied alcids). By comparing the ecological characteristics of birds that were successfully preserved with those that were not, I quantified the strength and mode of taphonomic filters. Although the precise parameter estimates do incorporate “noise” as a result of non-taphonomic community change, the premise of the model is that, owing to the fragile nature of avian remains, the biological noise to taphonomic signal ratio is as low as reasonably possible.

I collated ecological data on extant birds from the Cornell Birds of North America database (<http://bna.birds.cornell.edu/bna/>), with body masses from the CRC Handbook of Avian Body Masses (Dunning 1992). These data include binary characters describing the foraging habits of birds in various ecosystems (marine, lake, open, and forested), and binary characters describing the inclusion of various food categories in the birds' diet (terrestrial arthropods, arboreal insects, volant insects/vertebrates,

seeds, foliage, fruits, nectar, aquatic arthropods, aquatic plants, aquatic vertebrates and terrestrial vertebrates). For each bird species, I used the sources listed above to determine its primary mode of wing use (flightless, burst flight, travel between foraging patches, sallying, pursuit flight, wing-propelled diving, or soaring) and leg use (standing, killing, perching, walking, running, wading, paddling, or diving), and scored these as two further multivariate variables (ecological data available at Dryad and in the online supplement of Mitchell and Makovicky 2014). The Gower distances between species were found based on their ecological traits, and the species were ordinated using principal coordinates analysis (PCo) to generate continuous variables. The average of PCo scores for all species within a genus were used in subsequent analyses, as many bird species are defined by non-osteological characters (e.g., song, color), and there is low intrageneric variance in ecology at the coarse scale of these variables.

I downloaded extant bird assemblages for North America from the site records on ebird.org, and binned the assemblages by the 38 “Bird Conservation Regions” delimited there (data on Dryad and in Mitchell and Makovicky 2014). The probability of observing a given genus in a particular region recorded at ebird.org (correlated with abundance) was averaged across the more than 50 years of data, resulting in decadal time-averaged live data (which evens out yearly stochasticity). Using the literature, I pulled data on 139 dead assemblages (recent dead + subfossil deposits; see below) of birds (sources and data files on Dryad), ranging in age from ~100 thousand years ago, to surveys of bird mass deaths by U.S. National Park rangers. These dead assemblages were classified by the type of data present on the bird genera (presence/absence, total bone counts, minimum number of individuals [MNI], or whole bodies), and sites without an estimate of MNI or a count of individuals were removed. Each dead assemblage also contained geographic data, allowing them to be matched to the Bird Conservation Regions from the modern data. For the results below, only those sites with either an MNI or whole-body count and at least five genera were

included, leaving a total of 53 dead bird assemblages. Thirty-six of these assemblages have associated carbon dates, and they range from 100-Kyr-old Pleistocene deposits to a modern mass death (median age: 2490.5 years). These sites were classified by their deposit type (fluvial/sand, lake, archaeological midden, cave, or peat); the modern mass death was classified as “gas” because it resulted from a volcanic outgassing event (Bond et al. 2012; data on Dryad). These are collectively referred to by the catchall term “dead assemblages,” which includes “death assemblages” sensu Johnson (1960) along with both the Pleistocene fossil and Holocene subfossil assemblages.

Of note is that the Aleutian Islands off the coast of Alaska are their own Bird Conservation Region, and the Aleutians have a rich history of archaeological middens, especially Baldur and Kiska Islands. Further, the modern mass death from volcanic outgassing occurred on Kiska Island. These features make the Aleutian islands an ideal test case for the sensitivity of richness and ecological disparity (sum of variances along PCo axes) estimates to taphonomic bias through time in a single geographic area.

Habitat Filtering Model.—The Community Assembly via Trait Selection model (CATS model; Shipley et al. 2006; Shipley 2010) is a model designed to find the probability that a species from a regional pool will occur in a local community when trait-based habitat filtering is occurring. The model was designed with the intention of modeling the ecological process of habitat filtering; however, the model itself merely quantifies how biased traits are in a subset (local community or dead assemblage) relative to the generating set (regional species pool or living assemblage). Because the model simply measures how ecologically biased a subset of taxa is relative to a generating set, this ecological approach is readily repurposed for use in taphonomy by considering the probability that a member of the living assemblage (i.e., “regional pool,” or generating set) will occur in the fossil assemblage (i.e., “local community,” or subset) as a function of their ecological traits. All that is needed is a matrix of ecological traits for the living and dead taxa, as well as the proportional

abundances of the species in the dead assemblage.

This model also takes the proportional abundances of species in the living community as a “prior” for the probability a given species will end up preserved. This allows living abundances to be included, modeling the probability that a living species will be preserved as a function of its abundance-independent preservation potential (i.e., ecological traits) and its abundance, allowing for hard-to-preserve but abundant taxa to (potentially) have the same preservation potential as readily preserved but rare taxa. CATS modeling is equivalent to a generalized linear model from the Poisson family where log-local species abundance is a function of the trait vectors, with log-regional abundance as an offset term (the prior in MAX-ENT models), and can be interpreted as such (Warton et al. 2014). The following equation from Warton et al. (2014) gives the structure of the model:

$$\ln \mu_i = \ln q_i + \beta_0 + x_i' \beta$$

In the taphonomic formulation μ_i represents the proportional abundance of taxon i in the fossil deposit, and q_i is the proportional abundance in the modern (modeled as an offset). The vector x_i' is composed of taxon i 's ecological or morphological traits. Again, following Warton et al. (2014) the regression slope coefficients (β) and intercept are estimated by maximizing the Poisson likelihood of this function.

I fit the CATS filtering model to each of the 53 dead assemblages, using the aggregate trait values of the preserved taxa, as well as the proportional abundances and matrix of ecological traits for the living genera. This yielded the trait- and abundance-mediated probability that a living genus from a given living assemblage would occur in the dead assemblages from that same region. I then computed the likelihood of the occurrences in the dead assemblage, using the probabilities from this model for each site, and compared it with the likelihood in a scenario where all living taxa were equally likely to be preserved (regardless of abundance or ecology), and then with the likelihood from the living abundances alone (regardless of ecology). Instead of using the extant abundance as the probability, I used

uniform probability of occurrence to compare with the filtering model, for three reasons: first, the extant abundance is included as an offset ("prior") in the model, so including it as a separate model would be redundant; second, use of the extant abundances without ecological traits produces extremely low likelihoods across all sites; and finally, abundance changes so rapidly across geological time that using the extant abundances as hard constraints (as opposed to an offset) would likely be inappropriate for even young subfossil sites.

The preservation probability for each genus at each locality can be predicted by the filtering model; therefore, the likelihood of the filtering model at a site is simply the probability of observing the exact distribution of fossil counts at that site. For instance, if a living assemblage has a large number of bird species, but only loons are found in the dead assemblage, the likelihood is the probability of finding a loon (given either the uniform or filtering model) raised to the power of the number of loons. For the filtering model, the number of parameters is equal to the number of fit axes, whereas the uniform model has no free parameters (the probability that a genus from the living assemblage is preserved in the dead assemblage is one over the total number of genera in the living assemblage, and so is not free to vary to maximize fit).

Measuring Taphonomic Distortion.—I computed three statistics to summarize the difference between the ecological disparity present in the living community and the disparity observed in each dead assemblage. First, I assessed the effect of taphonomic distortion on estimates of ecological variance. If taxa were preserved at random with respect to ecology, variance observed in a fossil assemblage would be an unbiased estimate of the true variance and independent of sampled richness. However, if preservation is related to ecology, then the number of sampled species will be associated with variance. I regressed generic richness against functional dispersion (abundance-weighted ecological disparity [Shiple 2010]) for the modern assemblages, and found the residual between the predicted relationship for the modern and the observed relationship in each dead assemblage.

Second, if preservation is biased with respect to ecology, then the mean value of ecological traits in the fossil assemblage will be shifted in the direction of the bias relative to the mean value of the living community from which it was generated (e.g., a bias toward larger body sizes will produce a larger mean size in the dead assemblage). I computed the magnitude and direction of the centroid shift between the living and dead disparity, using the first five ecological axes from the PCo (63% of the variance from axes with non-negative eigenvalues). Only the first five axes were used, as these were the axes fit in the filtering model. Finally, if ecological bias is consistent within or among deposit types, then the centroid shift will tend to be in a particular direction. As an example, if large-bodied taxa are preferentially preserved in lacustrine deposits, then not only will there be a shift in the mean value of the fossil assemblage relative to the living community, but that shift will consistently be toward larger body sizes across multiple lake sites (if no bias is present, then the angles will be distributed uniformly and reflect only minor sampling error).

These statistics summarize the loss of disparity as well as the magnitude and direction of the location shift imposed by taphonomic biases. Under a model without ecological bias in preservation, these metrics will be small, random and only weakly correlated with one another, as the metrics will reflect standard sampling error. I tested how well the observed distortions match the model expectations by subsampling living communities with both uniform and filtering-modeled probabilities for genera, and I generated a distribution of expected correlations to compare with the observed correlation among distortion metrics.

To measure distortion in fossil preservation, it is necessary to assume that the ancient assemblages strongly resembled what is observed in the modern communities. This assumption is problematic even in these young deposits, because climate change, anthropogenic alterations, community changes, and evolution have all occurred even over the small time scales examined here. I fit the filtering model to each dead assemblage separately to estimate the strength and ecological bias of each site

individually. The correlations between distortion metrics and the deposit-type breakdown of ecology-taphonomy correlations described above test this assumption. If the ancient living community were extremely different from the modern community, then the different deposit types would show no consistent patterns, and the distortion metrics would be weakly to uncorrelated with one another as they would be measuring both community change and taphonomic bias. Likewise, if the living community changed dramatically between the time of deposition and the modern, the information content of the modern relative abundances would be low (even used weakly as an offset in a linear model, as done here), and poorly supported by the Akaike weight tests.

Results

When taxa from the simulated assemblages are sampled without respect to simulated traits (uniformly), the “fossilized” mean and variance cluster around the true values and are sample-size-independent estimators, as expected (Fig. 1, black). However, introducing preservation bias to the simulations results in substantially erroneous estimation (Fig. 1, gray), with both

mean and variance being sensitive to sample size until all species in the living assemblage have been collected (a situation that never occurs in the fossil record).

Any model of taphonomic bias (not simply exponential) will produce an offset between the true and preserved means and variances. The shape and magnitude of the offset is determined by the variance in preservation potential, which is itself a function of the strength of bias, the type of bias (exponential, logistic, linear, etc.), and the variance in organism traits present in the living community (Fig. 2). An easy way to conceptualize these simulations is with the trait of interest being body size. If larger body sizes are preferentially preserved, then the observed average fossil body size will be higher than the true average body size, and the variance in body size will be spuriously low (as only the large end of the size spectrum is preserved). This effect holds true even if abundance is inversely correlated with body size, although under some non-exponential models of preservation bias there is hypothetically an equilibrium point, where a species’ body size and its abundance exactly even out to produce no signal of bias.

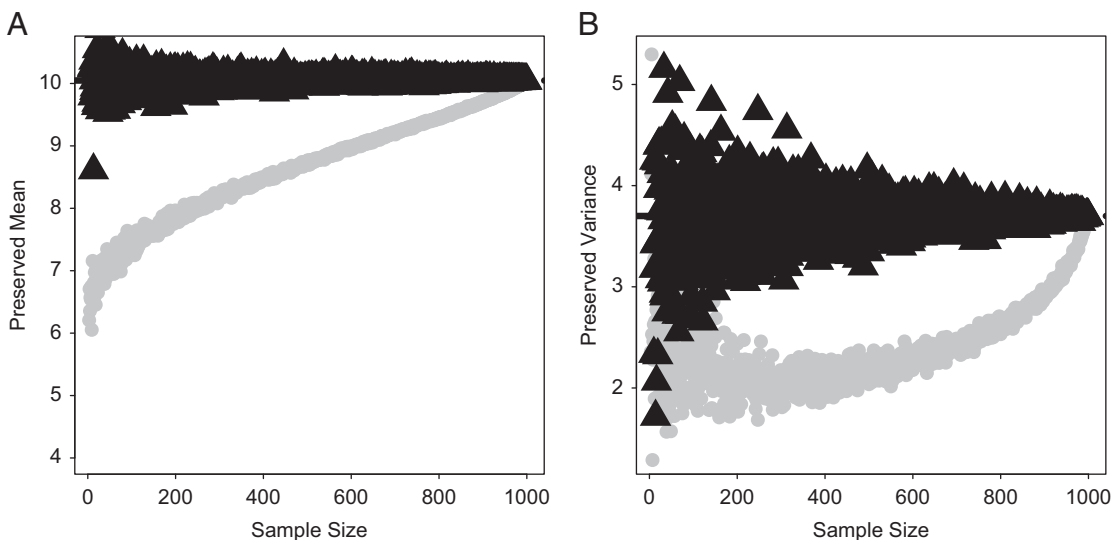


FIGURE 1. Simulations showing how estimates of mean and variance vary with the number of species sampled under biased and uniform preservation. Graphs show the relationship between the subsampled mean (A) and variance (B) at different sample sizes, with uniform preservation (black) and biased preservation (gray). Until nearly all species are sampled, biased preservation results in an offset estimate of means (direction of offset is related to both the magnitude and direction of bias) and a lower estimate of variance (only related to the magnitude of the bias).

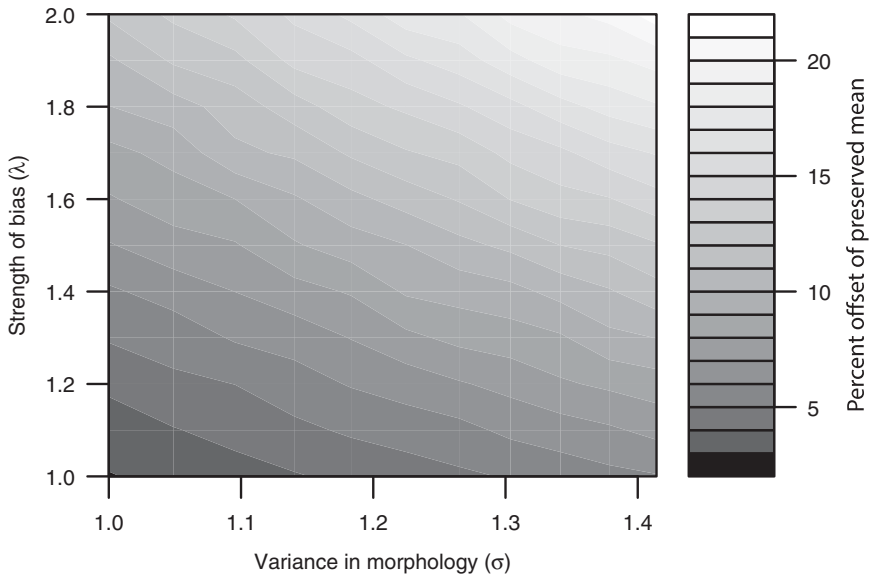


FIGURE 2. Simulation results showing how the distortion introduced by preservation bias (percent offset between the true and “fossilized” means) varies with increasing bias strength (λ) and increasing ecological/morphological variance in the “living community” (σ -squared).

Beyond aspects of ecological disparity, preservation bias can mislead subsampled richness estimates, as well. When no bias is present (the preservation potential of a species is associated solely with abundance) the rarefaction curves for simulated communities generated under identical conditions are tight, and not significantly different from one another, as one would expect (Fig. 3A, black). Likewise, estimates of richness under uniform and biased subsampling vary significantly even when using a fair subsampling metric like SQS (Fig. 3C). In the case of the Aleutian Islands (Fig. 3B), a significant difference is perceived between the subfossil avian assemblages (3325–200 years old) and the recent death assemblage (8 years old; Bond et al. 2012). All of the assemblages 200 years and older are from archaeological middens, whereas the recent death assemblage is from a volcanic outgassing. Again, for the Aleutian Islands, a sudden drop in richness is associated with the switch to a different taphonomic mode in the modern. In the simulations, when bias is identical across iterations, both rarefaction and SQS correctly infer no significant difference among sites (note clustering of black rarefaction curves in Fig. 3A and low variance

in black points for both Shannon's H and subsampled richness for SQS simulations in Fig. 3C). However, for simulations where the fossilization process distorts the evenness of preserved communities in different ways (e.g., a living community fossilized under two or more taphonomic regimes) both rarefaction and SQS detect differences between assemblages that are taphonomic (not ecological). These models were designed to detect sampling bias, and are performing exactly as they were designed to in these simulations (to describe relative differences among preserved samples), but not in the way they are commonly used (to describe relative differences in ancient living assemblages).

Preservation bias distorts the evenness preserved in the fossil record, with greater bias resulting in lower evenness (i.e., higher bias means a higher proportion of the sample is made of the few readily preserved taxa; Fig. 3C). When the fossilization process distorts the evenness of preserved communities in different ways (e.g., a living community fossilized under two or more taphonomic regimes) both rarefaction and SQS will be positively misled and will detect spurious differences between assemblages. This raises the question

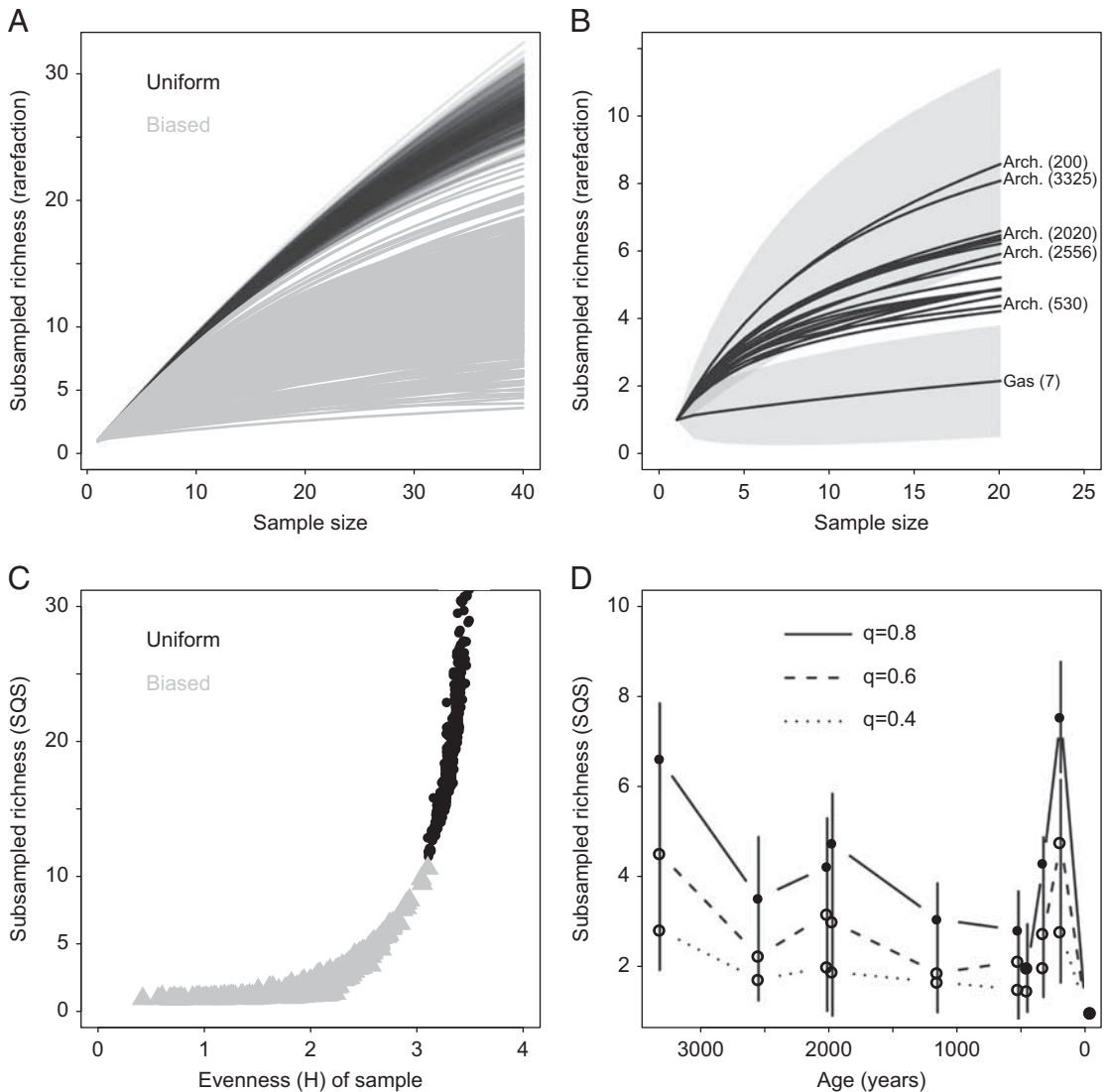


FIGURE 3. Simulated and empirical data showing how deposits formed under different levels of preservation bias can have dramatically different richness estimates, even when drawn from the same living community. A, Rarefaction curves from randomly generated “living communities” of 50 species that were “fossilized” either with (gray) or without (black) bias. B, Empirical example of rarefaction curves from the Aleutian Islands, with ages (years B.P.) for select sites on the right; the youngest assemblage is from a volcanic outgassing whereas all others are from archaeological middens. Confidence intervals are shown for the two most recent deposits (200-year-old midden and the gas-induced mass death from 2007 C.E.). The variance in richness between these two recent sites is significant, but these observed differences could have been produced by the taphonomic biases illustrated in panel A. C, SQS subsampled richness values from randomly generated “living communities” of 100 species that were “fossilized” by sampling 10% of the individuals either with (gray) or without (black) bias (quorum = 0.4). SQS takes subsamples of different sizes as a function of the evenness of the total sample, and so evenness is plotted here. Bias preservation directly affects evenness, and so also affects richness comparisons between subsamples from SQS. D, Empirical example of how SQS subsampled richness through time for the Aleutian Islands at quorum levels of 0.8, 0.6, and 0.4. Error bars show one standard deviation above and below the mean subsampled richness for each site after 1000 runs, and the differences between the two more recent sites are significant ($p < 0.01$) for all three quorum values, but given the expected variance in richness from taphonomic processes shown in panel C, the biological meaning of these significant differences is questionable.

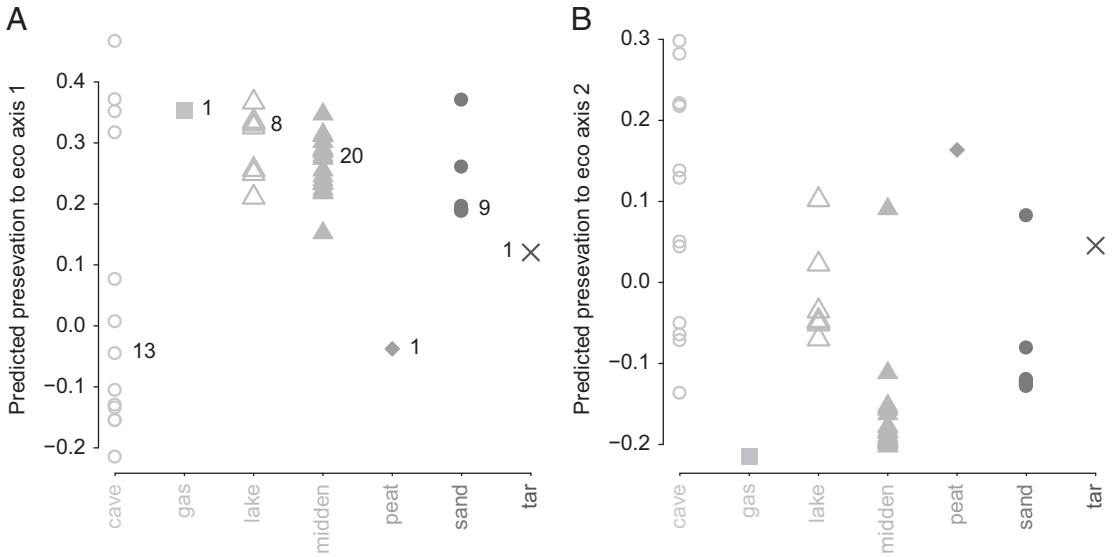


FIGURE 4. For each assemblage of dead birds, the correlation between the preserved species' predicted preservation probabilities and their ecological PCo scores along the first (A; 39.5% of variance) and second (B; 22.1% of variance) axes. Sites are grouped by taphonomic mode, with the number of sites for each taphonomic mode given in left plot. Correlations here show the direction of preservation bias at each site.

of whether the differences seen between Aleutian island sites from different taphonomic regimes (Fig. 3B, D), or any arbitrary collection of fossil assemblages, is real or an artifact of changing preservational modes (see *Discussion*).

An important consideration is how different ecologies are related to preservation potential. To understand how different ecological traits were related to preservation, I used the correlation at each site between the preservation potential estimated by the trait-based filtering model from Shipley (2010) for each bird, and that bird's ecological score. This is not a metric of fit, but rather these correlations show how the magnitude and direction of the model-estimated biases vary across sites (Fig. 4). A strong correlation value for a site indicates that a given ecological axis is a good predictor of preservation potential for that site, whereas a weak correlation suggests that the ecological axis under examination has little bearing on preservation at a particular site. The first ecological axis (PCo1) is strongly associated with the difference between aquatic and terrestrial taxa (correlations: 0.61 Marine, 0.73 Wetlands, 0.63 Lakes, -0.65 Forests and -0.33 Open), and the second axis

(PCo2) divides terrestrial birds (-0.49 Forests, and 0.51 Open). Both axes are also positively correlated with body mass (0.43 and 0.33). The positive correlation between predicted preservation and ecological axis 1 in lakes, middens, and sandy channels suggests that large-bodied aquatic birds are preserved at those sites more often than would be expected either at random or due to their proportional abundances. Likewise, the negative correlation with the second axis in middens and sandy channels implies preferential preservation of closed-habitat birds. Cave deposits, on the other hand, do not have a single general bias, suggesting that comparisons of richness and proportional abundance between cave deposits should be done with caution, as even within that category there is high variance in the strength, direction, and magnitude of preservation bias. Predicted preservation potential is positively, albeit weakly, correlated with the first morphological axes at all sites (range of Pearson's $r = 0.03$ to 0.32 , mean of 0.16 , $p < 0.01$ for 20/53 sites), likely reflecting the fact that larger birds are generally more likely to be preserved. The Akaike weight of the filtering model (compared to the uniform) for each of the 53 sites is plotted in Figure 5.

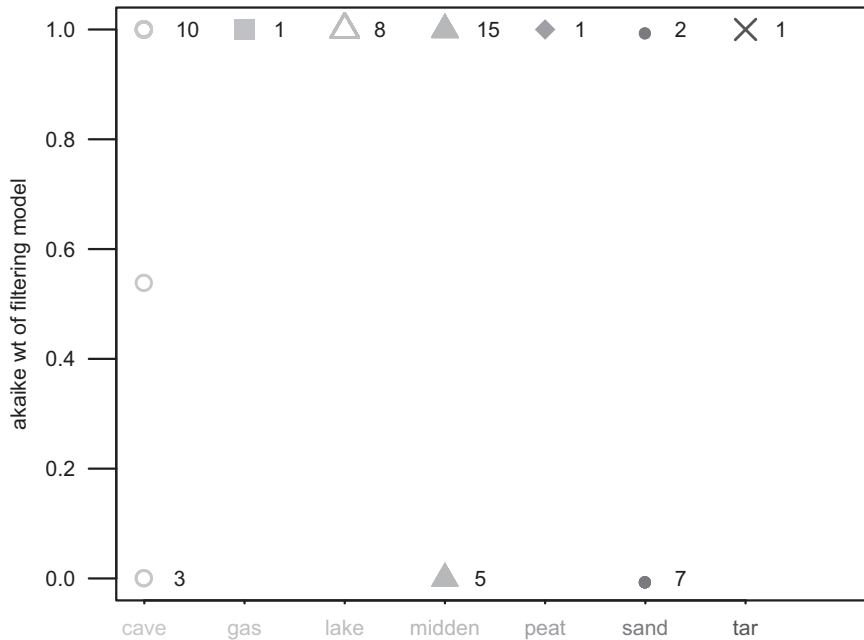


FIGURE 5. Support (Akaike weight) for the ecological filtering model for each site, separated by taphonomic mode, with the numbers shown at top representing how many sites show strong support (Akaike weight ≥ 0.9) for the filtering model, and numbers at bottom representing how many do not.

All three distortion metrics are uncorrelated with both age of the assemblages (Fig. 6, center column) and number of individuals sampled (Fig. 6, right column). This suggests that for taphonomic modes other than (potentially) cave deposits, the differences observed are not primarily due to changes in community composition through time or to sampling effort, but rather to changes in preservation potential. The Aleutian Island middens provide strong evidence for this, with no consistent patterns through time (Fig. 3, right) or with the various distortion metrics (Fig. 6); these results suggest either real changes to the biological community, or taphonomic fluctuations (changes in human predation patterns) below the resolution of this model, although I lean toward the latter because of the short time period involved.

The correlations between distortion metrics are all statistically significant ($p < 0.01$), and the correlation between the centroid shift and the variance residual is the highest of all combinations (Fig. 7A). The strength of this correlation is higher than predicted by either uniform subsampling or even by the filtering model, although the filtering model is significantly closer (Fig. 7B). The biological significance of

the strong correlations is that all three metrics are capturing aspects of taphonomic distortion such that sites can be broadly broken into “low” or “high” distortion. However, given that the correlations are higher than expected by the filtering model, and that for most sites the filtering model is preferred over uniform subsampling, the actual preservation biases present in birds either are stronger than modeled or include factors not present in this ecological data set (e.g., bone pneumaticity, life span). Channel and cave deposits show the largest departure from the filtering model, although overall the model is supported across 37 of the sites, suggesting that it is useful in comparing many fossil assemblages with one another.

Discussion

Differences observed between fossil assemblages, or between fossil and modern assemblages, are the result of both biological differences between the true communities, and the differential preservation and persistence of taxa from those communities (Johnson 1960; Lawrence 1968; Damuth 1982; Behrensmeyer et al. 2000; Cooper et al. 2006). Differential preservation

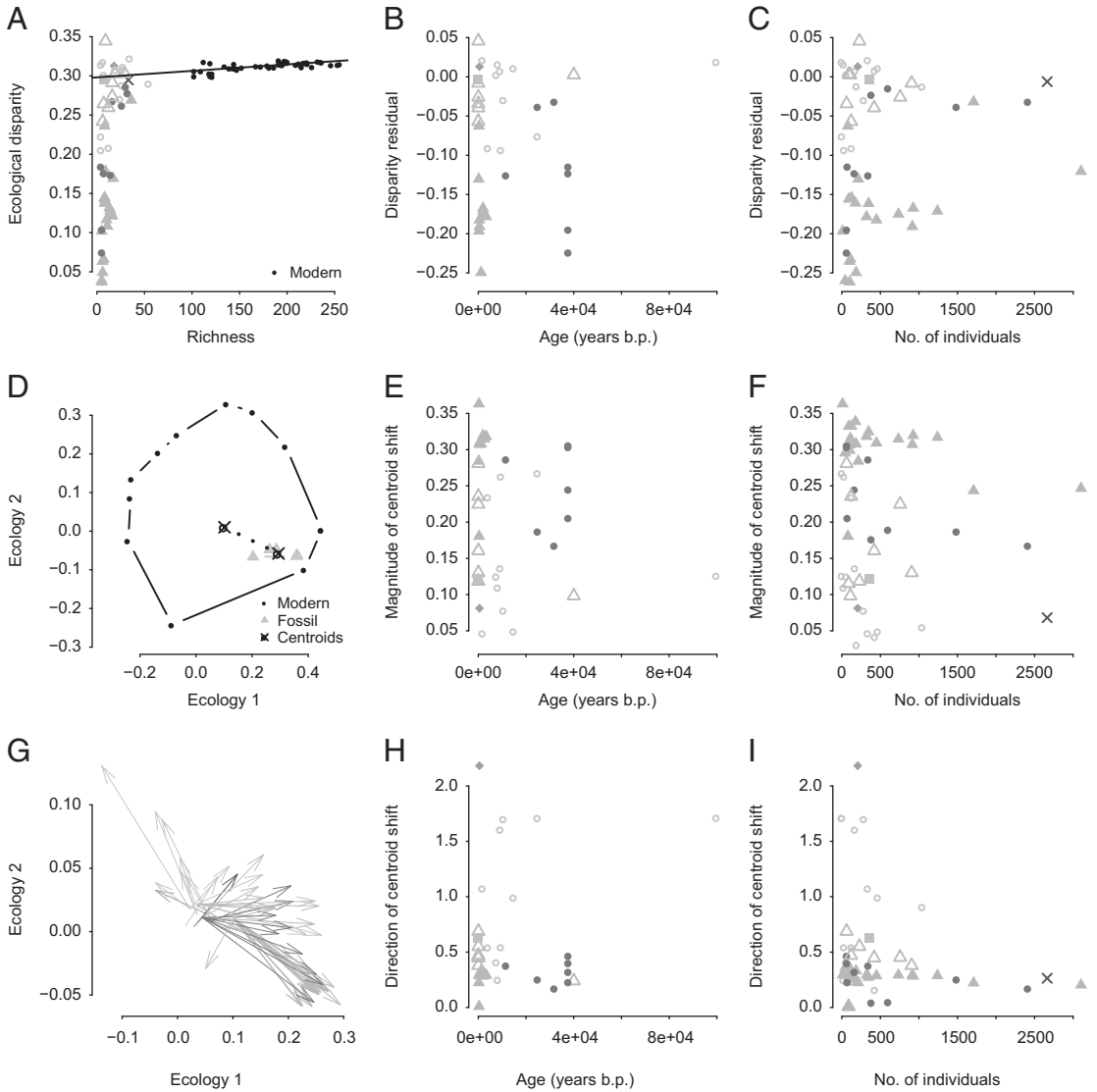


FIGURE 6. Graphs illustrating how the various metrics of ecological distortion were measured (left column), and their correlation with deposit age (center) and sample size (right) for the different sites, with symbols as in Fig. 4. Top row: Ecological disparity shows the difference (residual) between the disparity observed in the preserved birds from young dead assemblages, and the expected disparity for that many genera given the linear relationship in the modern. Center row: The magnitude of the centroid's shift from the extant North American bird assemblages to the preserved assemblages from the same geographic region (Bird Conservation Region). Bottom row: The direction of the centroid's shift in radians from the extant to the dead North American bird assemblages. Note that only the cave deposits show any sign of a correlation with sample size or age, and that all but one site show a positive shift along the first two ecological axes (which are positively correlated with body size).

results from interactions between organismal traits, such as ecology and body size, and the physical processes of deposition. Preservation biases change the relationship between the abundance of an organism in the living community (*A in* Damuth 1982) and its abundance in the death assemblage (*D in* Damuth 1982). Some

traits, such as the presence of hard parts, generally have a strong positive influence on preservation potential; however, in certain deposits (e.g., the Burgess Shale), the relationship can be inverted. This means that both deposit type and organismal traits must be taken into account when comparing richness or disparity

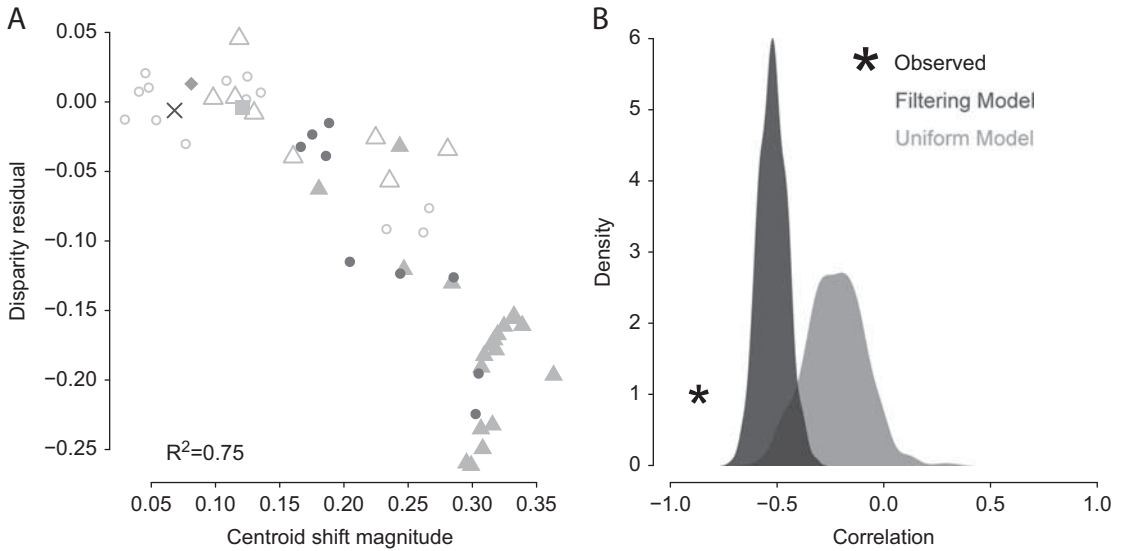


FIGURE 7. Correlation between the richness-disparity residual and the magnitude of the centroid shift in subfossil assemblages (A) and the expected correlation under uniform (B, light gray) and biased (B, dark gray) preservation. Observed correlation depicted in (A) is plotted in (B) as a black asterisk (*), to compare the observed correlation to model expectations. Symbols as in Figure 4. When sites are composed of birds that are more ecologically similar to one another than expected from the richness (the residual), the sites also have a center that is strongly shifted, suggesting that the taxa not only are all ecologically similar, but they also do not accurately reflect the “average” birds that lived there before fossilization (magnitude of center shift). This relationship among metrics of ecological distortion is not expected under uniform subsampling (B), and is strong evidence for ecological bias in preservation.

between two or more fossil assemblages. Further, in attritional deposits (sensu Johnson 1960), time-averaging and the range of generation times in focal taxa add additional post-death biases (Vermeij and Herbert 2004), although time-averaging also smoothes out year-to-year noise and allows for the preservation of otherwise rare taxa (Olszewski 1999). In sum, comparing sites with different degrees of time-averaging is likely problematic, but comparing sites with similar degrees is likely to give a more realistic signal of long-scale ecological patterns.

My results show that, under conditions where there is a strong bias in preservation based on organismal traits, comparisons between assemblages are readily misled. This means that estimates of richness and ecological/morphological disparity are compared among sites with different taphonomic histories at the researcher’s peril. Each of those sites shows a significant signal of taphonomic bias, and the direction of the bias is in the intuitive direction (abundance, body size, and aquatic habits are all positively correlated with preservation). It is vital to document the effect of

taphonomic biases on fossil assemblages before we can move from the preservation-biased fossil record to an accurate understanding of the biological history it represents.

As an example empirical case, the Aleutian Island deposits are primarily middens, formed over dozens to hundreds of years by the people living there, and the birds people choose to catch and eat are biased by all manner of factors. All of these sites show a strong signal of ecological distortion, despite being some of the youngest dead assemblages in the data set. By taking a nonrandom subsample of the avifauna, and doing so repeatedly over a long time, the evenness of the preserved avifauna is vastly altered compared to the evenness of the original avifaunas that lived alongside the hunters. Through selective hunting and time-averaging, the deposits have become distorted, and these distortions strongly affect subsampling techniques designed to compare richness. Although human-derived middens do not exist in the deep time fossil record, other highly biased taphonomic regimes do. This means that although the particulars of middens are a

recent phenomenon, the larger message is that paleobiologists should be broadly concerned about direct comparisons between deposits with differing taphonomic histories.

This message is emphasized when the highly biased midden deposits are compared with the gas-induced mass death on Kiska island (one of the Aleutian islands), where a localized and instantaneous kill agent caused a mass mortality in the local avifauna. The richness of the birds in that death assemblage is inferred to be significantly lower than even the 200-year-old midden deposit when subsampling techniques are applied. However, the evenness of the gas-induced and midden samples are distorted from the original community in different ways, as are their degrees of time-averaging. This supports the contention that comparisons between sites with different taphonomic modes are unreliable. If deposit type is correlated with both the types of preservational biases and the amount of time-averaging, as some studies suggest (e.g., Kidwell et al. 2005), then future studies should attempt to incorporate data on deposit types as a covariate with other aspects of the fossil record, such as overall sampling and rock volume.

Despite advances in understanding taphonomic processes, both large- and small-scale comparisons between sites and time slices still rely almost exclusively on subsampling procedures. This is problematic, as subsampling can only account for variances in preserved taxonomic richness between the sites/times as a result of differential worker effort, and cannot account for variances in what portion of the true community was preserved for workers to find. Therefore, observed differences do not necessarily reflect true biological differences. A single living assemblage, subjected to two different taphonomic filters, can preserve entirely different fossil assemblages. If one site preserves a large fraction of the true community, while another simply preserves very few taxa (i.e., has a stronger bias—an extreme case being monotoxic beds), then subsampling estimates are expected to detect a significant difference between the assemblages. If our goal as paleontologists is to describe the fossil record itself, then this significant difference is meaningful, because

the respective compositions of the fossil deposits truly differ. However, if our goal is to accurately describe the history of life, then this difference is spurious and positively misleading, because the compositions of the ancient assemblages were the same (see *Results*).

As an example of the importance of taphonomic considerations, Terry (2010) showed that raptor deposits have high fidelity between living and dead assemblages for rodents, and Hadly (1999) has shown that cave deposits (a combination of woodrat middens and carnivore accumulations) in an arid-landscape cave can show high fidelity between living and dead mammalian sagebrush communities, with caves preserving even large-bodied taxa like ungulates and carnivorans. Hadley (1999) noted that even if deposits are drawn from the same living community, comparing deposits from different taphonomic regimes can positively mislead researchers. Subsampling techniques performed across taphonomic regimes will not mitigate these inherent biases.

Beyond simply the richness measures, disparity measures such as variance in body size are also biased by taphonomic processes. Raptor accumulations, for example, preserve only the subset of body sizes preferred as prey, and do not reflect a true estimate of body size variance for the whole ecosystem. Additionally, owls preferentially predate nocturnal taxa, like murids, whereas hawks preferentially predate diurnal taxa, like sciurids and reptiles (Marti and Kochert 1995), so differences in taphonomic composition can occur even within the “raptor accumulation” deposit type. Vertebrate death assemblages show high spatial fidelity (Miller 2012; Miller et al. 2013), which means that variations in richness within a geological formation may be a function of the number of distinct localized deposits it contains, and not necessarily a function of the regional richness that actually existed (e.g., Lyson and Longrich 2011). This is very useful for intraformational ecological comparisons, but again the number of deposit types represented within a formation should be considered along with total volume when comparing among formations.

Different taphonomic filters applied to a single living assemblage has been readily

documented in taxa as different as marine invertebrates (e.g., Kidwell et al. 2005) and terrestrial vertebrates (e.g., Behrensmeyer et al. 2000, 2003). Although such taphonomic filters are widely recognized, the ability to quantify their effect in large-scale analyses has been hindered by concurrent biological change. For death assemblages and subfossil or fossil deposits, comparisons with the modern reflect changes in the assemblage composition that are imposed by ecological reorganization as well as by preservation biases (e.g., Valentine 1989; Vermeij and Herbert 2004; Kidwell et al. 2005; Miller 2012). In the case of birds, which have a very high variance in preservation potential (data above; Behrensmeyer et al. 2003; Turvey and Blackburn 2011), the effect of preservation biases is up-weighted, allowing them to be readily detected and quantified. Beyond the results presented above for birds and the hypothetical examples, strong empirical evidence exists for preservation biases affecting even clades with “good” preservation potential, like mammals. For instance, fluvial deposits that destroy small bones are known to preserve a different suite of taxa than stagnant pools (Badgley 1986; Rogers and Brady 2010). Additionally, Damuth (1982) showed that several Pleistocene mammal assemblages were strongly biased, with anomalously low abundance of small-bodied taxa.

These results suggest that future work on reconstructing the true biological patterns of richness and disparity through time should also take taphonomic differences into account directly. If a given region or time period has a higher richness, but also a greater variance in depositional types, any apparent patterns in richness and/or disparity could be a spurious signal of taphonomy. Even if one individual site is better sampled (more specimens) than another, the better-sampled site may still preserve a more distorted and less accurate view of the living assemblage if it is from a deposit type with stronger biases.

My simulations and empirical data support past conclusions that repeated sampling of taphonomically “fragile” taxa is a good indication that a large fraction of the true richness (and disparity) has been discovered (Bottjer and Jablonski 1988). Restricting large-scale analyses

only to sites that preserve taphonomically fragile taxa (e.g., small, thin-shelled, aragonite bivalves), or restricting smaller-scale analyses to comparisons between sites with similar taphonomic histories, is recommended. Significant differences between sites in richness or disparity, detected with subsampling methods like rarefaction, cannot be interpreted if those sites differ in taphonomic mode.

In this study, I use simulations to demonstrate the effect of varying strengths of preservational biases, and to establish the role of such biases in generating assemblages that mislead subsampling estimates of richness and morphological disparity. I also use empirical examples from death assemblages and from subfossil and fossil deposits of birds to highlight the effect of such biases on real data sets. Given their small average size and the frailty of their bones, birds are widely recognized as experiencing strong preservational biases (Turvey and Blackburn 2011; Ksepka and Boyd 2012). Although community compositional changes have certainly occurred between the dead and living avian communities, I argue here that such changes are small relative to the changes imposed by preservation; this makes birds both an ideal case for quantifying the effect of preservational biases on fossil community reconstruction and a group for which the consideration of preservational biases in paleoenvironmental, paleoecological, and evolutionary reconstructions is of paramount importance.

Acknowledgments

This manuscript was greatly improved by comments from two anonymous reviewers, J. Miller, P. Makovicky, M. Foote, K. Angielczyk, M. Webster, J. Leonard-Pingel, K. Voorhies, and T. Sosa. Discussion with members of the Field Museum’s Bird Division and use of the collections also improved this paper. This work was financially supported by the Hinds Fund, and National Science Foundation grants EAPSI 1107676 and DDIG 1311389.

Literature Cited

- Alroy, J. 2010. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology* 53:1211–1235.

- Badgley, C. 1986. Counting individuals in mammalian fossil assemblages from fluvial environments. *Palaios* 1:328–338.
- Behrensmeyer, A. K., S. M. Kidwell, and R. A. Gastaldo. 2000. Taphonomy and paleobiology. *Paleobiology* 26:103–147.
- Behrensmeyer, A. K., C. T. Stayton, and R. E. Chapman. 2003. Taphonomy and ecology of modern avifaunal remains from Amboseli Park, Kenya. *Paleobiology* 29:52–70.
- Behrensmeyer, A. K., F. T. Fürsich, R. A. Gastaldo, S. M. Kidwell, M. A. Kosnik, M. Kowalewski, R. E. Plotnick, R. R. Rogers, and J. Alroy. 2005. Are the most durable shelly taxa also the most common in the marine fossil record? *Paleobiology* 31:607–623.
- Bond, A. L., W. C. Evans, and I. L. Jones. 2012. Avian mortality associated with a volcanic gas seep as Kiska Island, Aleutian Islands, Alaska. *Wilson Journal of Ornithology* 124:146–151.
- Bottjer, D. J., and D. Jablonski. 1988. Paleoenvironmental patterns in the evolution of Post-Paleozoic benthic marine invertebrates. *Palaios* 3:540–560.
- Brown, J., J. Rest, J. Garcia-Moreno, M. Sorenson, and D. Mindell. 2008. Strong mitochondrial DNA support for a Cetacean origin of modern avian lineages. *BMC Biology* 6:6. doi:10.1186/1741-7007-6-6.
- Brusatte, S. R. Butler, P. Barrett, M. Carrano, D. Evans, G. Lloyd, P. Mannion, M. Norell, D. Peppé, P. Upchurch, and T. Williamson. 2014. The extinction of the dinosaurs. *Biological Reviews*, doi:10.1111/brv.12128.
- Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* 275:1109–1113.
- Cooper, R. A., P. A. Maxwell, J. S. Crampton, A. G. Beu, C. M. Jones, and B. A. Marshall. 2006. Completeness of the fossil record: estimating losses due to small body size. *Geology* 34:241–244.
- Damuth, J. 1982. Analysis of the preservation of community structure in assemblages of fossil mammals. *Paleobiology* 8: 434–446.
- Dunning, J. B. 1992. *CRC Handbook of avian body masses*, 2nd ed. CRC Press, Boca Raton, Fla.
- Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, F. H. Sheldon, D. W. Steadman, C. C. Witt, and T. Yuri. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- Hadly, E. A. 1999. Fidelity of terrestrial vertebrate fossils to a modern ecosystem. *Palaeogeography, Palaeoclimatology, Palaeoecology* 149:389–409.
- Holland, S. M. 2003. Confidence limits on fossil ranges that account for facies changes. *Paleobiology* 29:468–479.
- Johnson, R. G. 1960. Models and methods for analysis of the mode of formation of fossil assemblages. *Geological Society of America Bulletin* 71:1075–1086.
- Kidwell, S. M., M. M. R. Best, and D. S. Kaufman. 2005. Taphonomic trade-offs in tropical marine death assemblages: differential time-averaging, shell loss, and probable bias in siliclastic vs. carbonate facies. *Geology* 33:729–732.
- Kosnik, M. A., J. Alroy, A. K. Behrensmeyer, F. T. Fürsich, R. A. Gastaldo, S. M. Kidwell, M. Kowalewski, R. E. Plotnick, R. R. Rogers, and P. J. Wagner. 2011. Changes in shell durability of common marine taxa through the Phanerozoic: evidence for biological rather than taphonomic drivers. *Paleobiology* 37:303–331.
- Ksepka, D. T., and C. A. Boyd. 2012. Quantifying historical trends in the completeness of the fossil record and the contributing factors: an example using Aves. *Paleobiology* 38:112–125.
- Ksepka, D. T., J. L. Ware, and K. S. Lamm. 2014. Flying rocks and flying clocks: disparity in the fossil and molecular dates for birds. *Proceedings of the Royal Society of London B* 281(2): 0140677.
- Lawrence, D. R. 1968. Taphonomy and information losses in fossil communities. *Geological Society of America Bulletin* 79: 1315–1330.
- Lyns, T. R., and N. R. Longrich. 2011. Spatial niche partitioning in dinosaurs from the Late Cretaceous (Maastrichtian) of North America. *Proceedings of the Royal Society of London B* 278: 1158–1164; doi: 10.1098/rspb.2010.1444.
- Marti, C. D., and M. N. Kochert. 1995. Are red-tailed hawks and great horned owls diurnal-nocturnal dietary counterparts? *Wilson Bulletin* 107:615–628.
- Miller, J. H. 2012. The spatial fidelity of skeletal remains: elk wintering and calving grounds revealed by bones on the Yellowstone landscape. *Ecology* 93:2474–2482.
- Miller, J. H., P. Druckenmiller, and V. Bahn. 2013. Antlers of the Arctic Refuge: capturing multi-generational patterns of calving ground use from bones on the landscape. *Proceedings of the Royal Society B* 280:20130275; doi:10.1098/rspb.2013.0275.
- Mitchell, J. M., and P. J. Makovicky. 2014. Low ecological disparity in Mesozoic birds. *Proceedings of the Royal Society of London B* 281; doi: 10.1098/rspb.2014.0608.
- Newham, E., R. Benson, P. Upchurch, and A. Goswami. 2014. Mesozoic mammaliaform diversity: the effect of sampling corrections on reconstructions of evolutionary dynamics. *Palaeogeography, Palaeoclimatology, Palaeoecology* 412:32–44.
- Olszewski, T. 1999. Taking advantage of time-averaging. *Paleobiology* 25:226–238.
- Pacheco, M. A., F. U. Battistuzzi, M. Lentino, R. Aguilar, S. Kumar, and A. A. Escalante. 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Molecular Biology and Evolution* 28:1927–1942. doi: 10.1093/molbev/msr014.
- Peters, S. E., and N. A. Heim. 2010. The geological completeness of paleontological sampling in North America. *Paleobiology* 36:61–79.
- Raup, D. M. 1975. Taxonomic diversity estimation using rarefaction. *Paleobiology* 1:333–342.
- Rogers, R. R., and M. E. Brady. 2010. Origins of microfossil bonebeds: insights from the Upper Cretaceous Judith River Formation of north-central Montana. *Paleobiology* 36:80–112.
- 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.
- Shibley, B. 2010. Community assembly, natural selection and maximum entropy models. *Oikos* 119:604–609.
- Shibley, B., D. Vile, and E. Garnier. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science* 314:812–814.
- Terry, R. 2010. On raptors and rodents: testing the ecological fidelity and spatiotemporal resolution of cave death assemblages. *Paleobiology* 36:137–160.
- Turvey, S. T., and T. M. Blackburn. 2011. Determinants of species abundance in the Quaternary vertebrate fossil record. *Paleobiology* 37:537–546.
- Valentine, J. W. 1989. How good was the fossil record? Clues from the California Pleistocene. *Paleobiology* 15:83–94.
- Vermeij, G. J., and G. S. Herbert. 2004. Measuring proportional abundance in fossil and living assemblages. *Paleobiology* 30:1–4.
- Warton, D. I., B. Shibley, and R. Hastie. 2014. CATS regression a model-based approach to studying community assembly. *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.12280.