Paleobiology, 41(4), 2015, pp. 633–639
DOI: 10.1017/pab.2015.26

A more precise speciation and extinction rate estimator

John Alroy

Abstract.—A new turnover rate metric is introduced that combines simplicity and precision. Like the related three-timer and gap-filler equations, it involves first identifying a cohort of taxa sampled in the time interval preceding the one of interest (call the intervals i_0 and i_1). Taxa sampled in i_0 and i_1 are two-timers (t_2); those sampled in i_0 and i_2 but not i_1 are part-timers (p); and taxa sampled only in either i_1, i_2, or i_3 are newly notated here as either s_1, s_2, or s_3. The gap-filler extinction proportion can be reformulated as $(s_1 - s_3)/(t_2 + p)$. The method proposed here is to substitute s_3 with the second-highest of the three counts when the expected ordering $s_1 \geq s_2 \geq s_3$ is violated. In simulation, this new estimator yields values that are highly correlated with those produced by the gap-filler equation but more precise. In particular, it rarely produces highly negative values even when sample sizes are quite small. It is mildly upwards biased when sampling is extremely poor and turnover rates are extremely low, but it is otherwise highly accurate. Examples of Phanerozoic extinction rates for four major marine invertebrate groups are given to illustrate the method's improved precision. Based on the results, the procedure is recommended for general use.

John Alroy. Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia.
E-mail: john.alroy@mq.edu.au

Accepted: 2 June 2015
Published online: 7 October 2015

Introduction

Simplicity, accuracy, and precision are the desired qualities of statistical estimators, and finding a balance of these properties has proven difficult with respect to measuring speciation and extinction rates. The recently proposed gap-filler (GF) equation (Alroy 2014) was intended to have these properties, but its precision is not very high. The purpose of this note is to introduce a heuristic improvement to the GF estimator that increases its precision and even its accuracy but is extremely simple. The GF equation's general properties are briefly restated below because readers may not be familiar with it.

To save space, no direct comparisons are made with older range-based equations such as those of Alroy (1996) and Foote (1999, 2000) and even older ones that were reviewed by Foote (1994). Such comparisons have already been made in detail (Alroy 2014), and the upshot is that these equations are often precise but have been argued to be biased by the Signor-Lipps, edge, Pull of the Recent effects, and by a taxonomic artifact related to the undersplitting of rare taxa. In other words, older equations suggest that extinction rates are elevated in intervals that precede mass extinctions but are not immediately before them; create gradual declines as the Recent is approached; and produce lower turnover rates for rare higher taxa, which is counterintuitive. The reason that older methods are biased in these ways is that they employ full age-ranges instead of examining patterns of sampling across limited temporal windows, which is the approach taken here and in related earlier studies (Alroy 2008, 2014).

Method

Moving-window methods (Alroy 2008, 2014) assume that there is a focal time interval (call it i_1) and that a cohort has been defined by only considering taxa sampled in the preceding one (i_0). Four categories of taxa have been discussed previously: those sampled in i_0 and i_1 (two-timers or t_2), those sampled in i_0, i_1, and i_2 (three-timers or t_3), those sampled in i_0 and i_2 but not in i_1 (part-timers or p), and those sampled in i_0 and i_3 but not in i_2 (gap-fillers or g). The three-timer (3T) and gap-filler (GF) equations use these counts and were respectively presented by Alroy (2008) and Alroy (2014). In this paper I put aside 3T because it has been shown to be much less precise than GF.

© 2015 The Paleontological Society. All rights reserved. 0094-8373/15

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press

The GF equation for the instantaneous per interval extinction rate μ is the following log ratio:

$$\mu = \ln[(t_2+p)/(t_3+p+g)]. \quad (1)$$

GF can be rearranged to yield an extinction proportion (i.e., the number of extinctions per taxon per interval) (Alroy 2014):

$$E = 1-(t_3+p+g)/(t_2+p). \quad (2)$$

Another version is:

$$E = (t_2 - t_3 - g)/(t_2+p). \quad (3)$$

The new equation is best understood by reformulating equation (2) in yet another way. Let us first define two new categories: s_1 = those taxa sampled in i_0 and i_1 but not later (first-interval taxa) and s_3 = those sampled in i_0 and i_3 but not in i_1 or i_2 (third-interval taxa). The s_1 and s_3 patterns have both been illustrated previously (Alroy 2014: Fig. 2) but have not been the subject of any specific discussion.

Now assume that sampling is incomplete but uniform. When turnover is zero, one should expect s_1 to equal s_3. When turnover is very fast, s_3 should limit on zero. The difference $s_1 - s_3$ should therefore be some function of the turnover rate. By careful inspection of Fig. 2 in Alroy (2014), it will become apparent that $s_1 - s_3 = t_2 - t_3 - g$ and therefore that:

$$E = (s_1 - s_3)/(t_2+p). \quad (4)$$

The problem with this equation is that the four counts are often small, creating high amounts of random sampling error. I note in passing that there are two simple ways to decrease the variance by increasing the denominator. The first is to add in the very smallest count:

$$E = (s_1 - s_3)/(t_2+p+s_3). \quad (5)$$

This metric performs well in simulations resembling those presented below and it could hardly be simpler. It does introduce an obvious downwards bias, but the amount of bias is small.

A second equation that can be shown by simulation to be similarly precise and more accurate is

$$E = s_1/(t_2+p+s_3) - s_3/(t_2+p+s_1+s_2), \quad (6)$$

where s_2 = the number of taxa sampled in i_0 and i_2 but not in i_1 or i_3 (one of two patterns producing the part-timer count p, just as s_1 is a subset of t_2 and s_3 is a subset of g).

Regardless of equation (6)'s merits, it is trumped in simulation by a very simple algorithmic approach. Consider the relations among the first-, second-, and third-interval taxa. Given that sampling is reasonably uniform albeit partial and that extinction is a one way street, the counts should have the order $s_1 \geq s_2 \geq s_3$. They often do not, either because the ordering is entirely reversed ($s_1 < s_2 < s_3$), s_2 is the minimum ($s_2 < s_1$ and $s_2 < s_3$), or s_2 is the maximum ($s_2 > s_1$ and $s_2 > s_3$). In such cases I propose substituting the second-lowest of the three counts for s_3 in equation (4). No further manipulation is necessary. In practice, this will mean doing nothing if s_3 is already in between s_1 and s_2; computing a zero rate if s_1 is in between the other two, which could happen if the GF rate is either negative or positive; and otherwise substituting s_2 for s_3, which can only happen if the GF rate is negative and which will edge the rate closer to zero.

I have been unable to find any other corrections that meet reasonable criteria (as detailed later) and substantially increase precision while generating only small amounts of bias. However, whether the improvement is worthwhile is really a practical matter. We will therefore move on to analyses of simulated and empirical data involving the new algorithm, which I call second-for-third substitution because the second-highest count is substituted for the third-interval taxon count.

Simulation Analyses

Figure 1A presents an example of a Monte Carlo simulation trial in which there are 50 time intervals; a fixed-size sampling pool of 100 taxa; a per-taxon, per-interval sampling probability p_s of 0.3; and no turnover.

It is easy to see that the second-for-third correction primarily serves to ameliorate outlying negative values. This decrease in variance is also illustrated by trials in which there are 1000 intervals and the sampling probability is varied (Fig. 2A). Meanwhile, the corrected proportions are substantially biased only when sampling is very poor (Fig. 2B). In other words, even if sampling is as bad as $p_s = 0.15$ the algorithm generates no bias while very substantially increasing precision.

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press

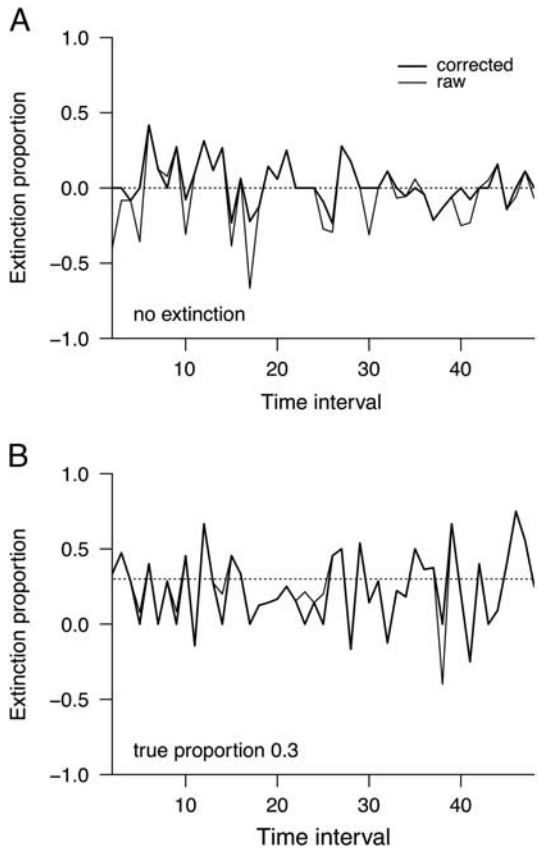


FIGURE 1. Two examples of 50-interval simulation trials. Parameters are described in the text. Thick lines = second-for-third proportions; thin lines = conventional gap-filler proportions (Alroy 2014). A, A simulation in which there is no actual turnover between intervals. B, A simulation in which there is 30% turnover.

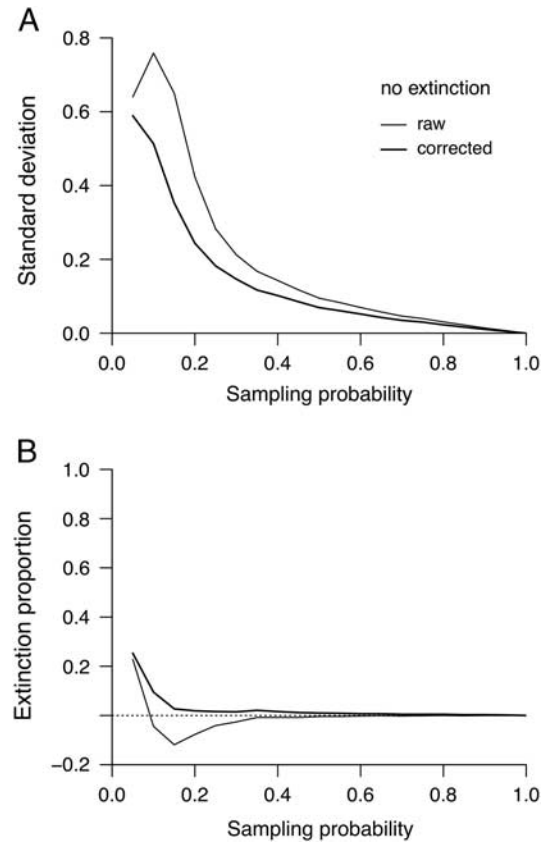


FIGURE 2. Results based on 1000-interval simulation trials in which the sampling probability is varied between 0.05 and 1 and there is no turnover. Thick and thin lines are as in Figure 1. A, Relationship between sampling probabilities and standard deviations of estimates across trials. B, Relationship between sampling probabilities and average estimated extinction proportions.

By contrast, the raw GF proportions are just as upwards-biased when $p_s = 0.05$ but transiently downwards-biased at modest levels.

Similar results are seen (Figs. 1B, 3) when one imposes a probability of turnover at each interval boundary of 0.3 (this being here the chance that a taxon will be removed and immediately replaced with a new one). The difference in variance between the raw and corrected GF estimates is not as extreme but is still considerable (Fig. 3A). Meanwhile, there is now no large or consistent difference in accuracy (Fig. 3B).

Finally, varying the turnover rate between 0 and 1 while holding p_s to 0.3 demonstrates that the correction is helpful even at relatively good sampling levels of around $p_s = 0.5$ (Fig. 4A). It is also apparent that both equations are in most cases slightly downwards-biased (Fig. 4B). The reason is that trials have been excluded when $t_2 + p = 0$ because this sum is the denominator of equation (4). In all other cases, by definition the denominator is on average farther above zero than the true expected mean. The alternative of including such trials but assuming a proportion of 1 seems unreasonable given that it would mean the taxonomic group went entirely extinct over and over. The fact that the equations produce virtually the same expected values however fast turnover might be (Fig. 4B) while producing quite different amounts of variance (Fig. 4A) is actually more relevant.

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press

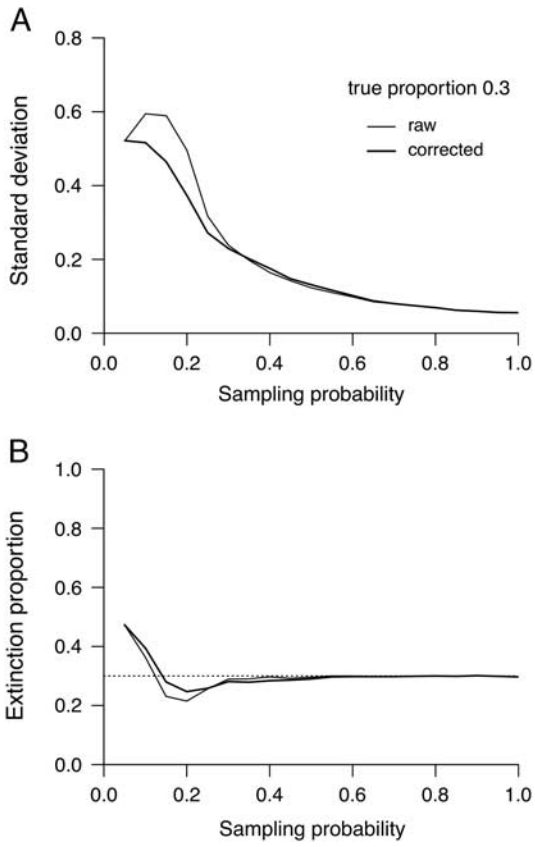


FIGURE 3. Results based on 1000 interval simulation trials in which the sampling probability is varied between 0.05 and 1 and the per interval extinction proportion is always 0.3. Thick and thin lines are as in Fig. 1. A, Relationship between sampling probabilities and standard deviations of estimates across trials. B, Relationship between sampling probabilities and average estimated extinction proportions.

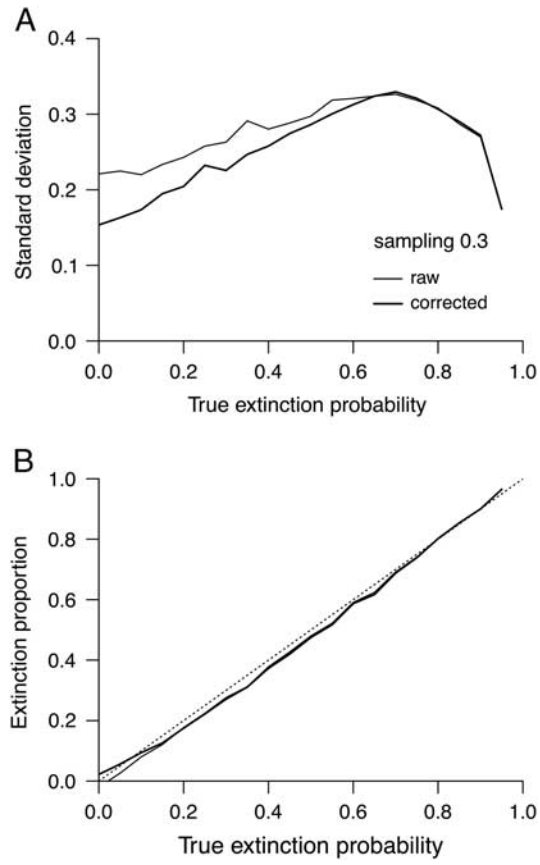


FIGURE 4. Results based on 1000 interval simulation trials in which extinction proportions are varied but the sampling probability is always 0.3. Thick and thin lines are as in Figure 1. A, Relationship between assumed extinction proportions and standard deviations of estimates across trials. B, Relationship between assumed proportions and average estimated extinction proportions.

Analysis of Phanerozoic Marine Data

The problem of reducing error in rate estimates is of more than mere academic interest because the paleobiological literature is awash with studies of individual taxonomic groups whose fossil records are simply not that good. Small and variable sample sizes are seen even in data sets pertaining to major phyla and classes. However, obtaining reliable data for such groups is essential for such undertakings as defining evolutionary faunas (Sepkoski 1981), showing whether higher taxa replace each other deterministically (Gould and Calloway 1980), and looking for evidence of logistic dynamics (Miller and Sepkoski 1988).

The GF and second-for-third extinction rate curves are computed here for four of the very most major marine taxonomic groups (Anthozoa, Brachiopoda, Bivalvia, and Gastropoda) to show whether the choice of equations does matter. The data were downloaded using the Fossilworks portal to the Paleobiology Database on 23 October 2014 using standard vetting criteria and a time scale with ca. 10 Myr resolution that are described elsewhere (Alroy 2014). Contributors have been thanked previously. Raw counts were used instead of sampling-standardized counts to make the problem more challenging.

The four groups were selected because they are highly abundant and have good fossil

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press

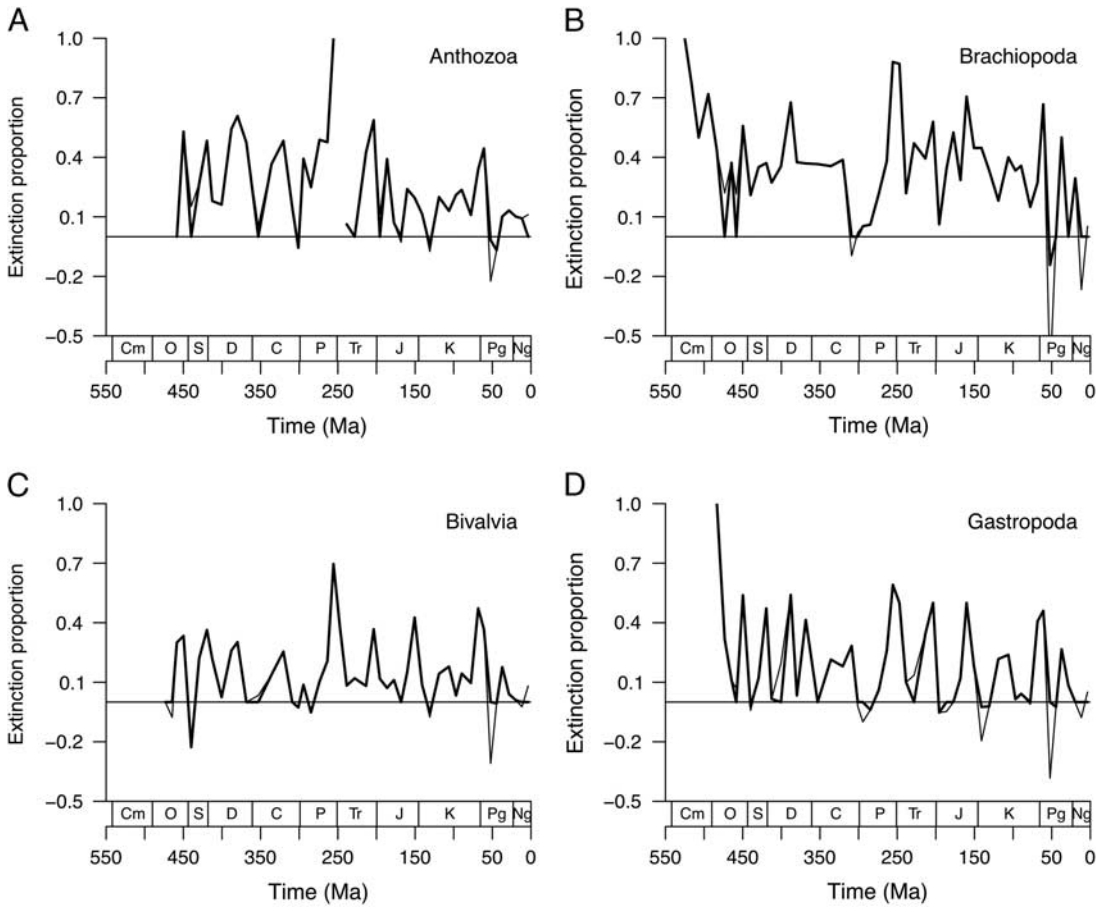


FIGURE 5. Genus level Phanerozoic extinction proportions for four major taxonomic groups. Thick and thin lines are as in Figure 1. A, Anthozoa. B, Brachiopoda. The truncated outlier in the Paleogene has a value of −0.714. C, Bivalvia. D, Gastropoda.

records spanning the entire Ordovician to Quaternary interval. Thus, obtaining a long time series of precise rates should be easy. By contrast, groups such as ostracods, bryozoans, crinoids, and echinoids have erratically sampled records and others such as trilobites, graptolites, and conodonts went extinct too early to allow computing long time series. Cephalopods would also be worth examining, but are omitted here to save space.

In fact, the GF rates for the four groups are often negative even though they do capture obviously meaningful features such as peaks representing major mass extinctions (Fig. 5). Second-for-third substitution reduces the number of negative rates from five to four for Anthozoa, three to one for Brachiopoda, eight to five for Bivalvia, and 13 to six for Gastropoda. Of particular note, substitution ameliorates the pronounced outliers in all four data sets that correspond with the second Paleogene bin (equivalent to the Early Eocene). The effects are not helpful in every single case: for example, substitution creates two substantial wobbles in the Ordovician part of the Brachiopoda curve (Fig. 5B). Nonetheless, the general pattern is most fairly interpreted as a step in the right direction.

Putting aside subjective impressions, it would help to quantify the improvement in precision. One easy approach is to compute standard deviations of the proportions (instantaneous rates are not dealt with here because a few values would be infinite in cases where turnover is complete). The Anthozoa, Brachiopoda, Bivalvia, and Gastropoda standard

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press

deviations for the GF data are respectively 0.231, 0.289, 0.216, and 0.247. After substitution, the four statistics decline to 0.230, 0.256, 0.206, and 0.229. All of the pairs of standard deviations save the one for Anthozoa are significantly different at $P < 0.001$ based on a bootstrap resampling analysis that involved randomly drawing the same number of rates from each time series with replacement and then computing the standard deviation. The procedure was iterated 1000 times and a paired Wilcoxon rank-sum test was performed to see if the resulting distributions differed.

Several interesting and important macroevolutionary patterns are visible in these curves. For example, the Anthozoa data suggest very high volatility that is consistent with the tendency of reefal organisms to suffer disproportionately during major mass extinctions (Kiessling and Simpson 2011). The Brachiopoda curve indicates persistently high background rates, a potential double pulse of extinction in the Late Permian, and an apparent shutdown in turnover in the latest Carboniferous and Early Permian that is highly peculiar. The Bivalvia curve depicts very low background rates punctuated by spikes that generally correspond to the major mass extinctions. And finally, the much more volatile Gastropoda curve agrees with it in suggesting that a fairly large extinction pulse might have characterized the Jurassic/Cretaceous boundary. All of these patterns are visible in the raw GF rates, but removing outliers and negative rates helps to cast them into clearer light.

In sum, the corrected GF equation produces figures that are still noisy—some negative rates are still present in all cases—but much more plausible. It stands to reason that similar noise reduction will be seen in other data sets, even in relatively large ones such as these.

Discussion

The new second-for-third algorithm is obviously heuristic. However, it meets a variety of sensible criteria for an estimator that corrects for sample size:

1. The procedure is so simple it can be computed by hand.
2. The GF equation's only clear-cut sources of information about the balance between extinction and sampling are the s_1 and s_3 counts (eq. 4). The correction involves the latter variable alone, so it avoids relying on counts such as t_2, p, and g that present mixed signals.
3. As mentioned, sampling error can cause s_1 to be less than s_3 even though there is no biological reason to ever see such a pattern. Suppose s_1 is zero and $s_2 < s_3$. The rate will still be negative, but the correction will bring it closer to zero despite not generally creating a strong upwards bias in the estimator (as was shown previously). There is no a priori reason to employ a correction that might have the opposite effect.
4. Finally, the algorithm has good limit properties in the sense of only having an effect when sampling is poor, turnover rates are not high, and the species pool is modest. First, if sampling is perfect then s_2 and s_3 are zero and the counts therefore cannot be out of order, which means that no substitution is made. Second, if extinction is very rapid then the same thing is true because s_2 and s_3 both limit on zero. Third, if the number of taxa is very large then the ranking of s_1, s_2, and s_3 is very unlikely to be out of order, again making a correction unnecessary.

In sum, the worst-case scenario is that the algorithmically corrected version of GF might generate a small bias given very poor sampling (Figs. 2B, 3B). In many other circumstances it is likely to be highly accurate. We would expect no visible bias and substantially better accuracy given a realistic case in which a data set has fine stratigraphic resolution, which means that the extinction rate is well below 0.3, and in which p_s is 0.1 to 0.5 (Figs. 2, 4). Furthermore, when sampling probabilities are clearly below 0.1 there are serious questions about whether turnover statistics should be computed in the first place.

Acknowledgments

I thank G. Hunt for a helpful review and A. Kocsis for raising my level of anxiety over the gap-filler equation's imprecision.

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press

Literature Cited

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. Palaeogeography, Palaeoclimatology, Palaeoecology 127:285– 311.

——. 2008. Dynamics of origination and extinction in the marine fossil record. Proceedings of the National Academy of Sciences USA 105:11,536–11,542.

——. 2014. Accurate and precise estimates of origination and extinction rates. Paleobiology 40:374–397.

Foote, M. 1994. Temporal variation in extinction risk and temporal scaling of extinction metrics. Paleobiology 20:424–444.

——. 1999. Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. Paleobiology Memoirs No. 1. Paleobiology 25(Suppl. to No. 2).

——. 2000. Origination and extinction components of taxonomic diversity: general problems. *In* D. H. Erwin and S. L. Wing, eds. Deep time: Paleobiology's perspective. Paleobiology 26(Suppl. to No. 4), 74–102.

Gould, S. J., and C. B. Calloway. 1980. Clams and brachiopods–ships that pass in the night. Paleobiology 6:383–396.

Kiessling, W., and C. Simpson. 2011. On the potential for ocean acidification to be a general cause of ancient reef crises. Global Change Biology 17:56–67.

Miller, A. I., and J. J. Sepkoski, Jr. 1988. Modelling bivalve diversification: the effect of interaction on a macroevolutionary system. Paleobiology 14:364–369.

Sepkoski, J. J., Jr. 1981. A factor analytic description of the Phanerozoic marine fossil record. Paleobiology 7:36–53.

https://doi.org/10.1017/pab.2015.26 Published online by Cambridge University Press