A simple way to improve multivariate analyses of paleoecological data sets

John Alroy

Abstract.—Multivariate methods such as cluster analysis and ordination are basic to paleoecology, but the messy nature of fossil occurrence data often makes it difficult to recover clear patterns. A recently described faunal similarity index based on the Forbes coefficient improves results when its complement is employed as a distance metric. This index involves adding terms to the Forbes equation and ignoring one of the counts it employs (that of species found in neither of the samples under consideration). Analyses of simulated data matrices demonstrate its advantages. These matrices include large and small samples from two partially overlapping species pools. In a cluster analysis, the widely used Dice coefficient and the Euclidean distance metric both create groupings that reflect sample size, the Simpson index suggests large differences that do not exist, and the corrected Forbes index creates groupings based strictly on true faunal overlap. In a principal coordinates analysis (PCoA) the Forbes index almost removes the sample-size signal but other approaches create a second axis strongly dominated by sample size. Meanwhile, species lists of late Pleistocene mammals from the United States capture biogeographic signals that standard ordination methods do recover, but the adjusted Forbes coefficient spaces the points out more sensibly. Finally, when biome-scale lists for living mammals are added to the data set and extinct species are removed, correspondence analysis misleadingly separates out the biome lists, and PCoA based on the Dice coefficient places them to the edge of the cloud of fossil assemblage data points. PCoA based on the Forbes index places them in more reasonable positions. Thus, only the adjusted Forbes index is able to recover true biological patterns. These results suggest that the index may be useful in analyzing not only paleontological data sets but any data set that includes species lists having highly variable lengths.

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

Introduction

Multivariate ordination and cluster analysis were introduced to paleoecology half a century ago (Reyment 1963; Valentine and Peddicord 1967) and are now employed routinely. However, problems with applying multivariate methods to fossil data remain unresolved. For example, Bush and Brame (2010) only recently used simulation analyses to revisit the debate over whether detrended correspondence analysis (DCA; Hill and Gauch 1980) performs better or worse than nonmetric multidimensional scaling (NMDS; Shepard 1962), both methods having gained traction in paleoecology (e.g., Holland et al. 2001; Bonelli et al. 2006).

The very core of the matter is that paleontological data, and especially binary presence-absence data, are almost invariably biased in important ways. Samples are of different sizes, preservation quality varies from place to place, taxonomy is often problematic, specimens are

often unidentifiable, and so on. The unequal sample-size problem is particularly obvious, which may explain why it was addressed so very early on by the development of a binary faunal similarity coefficient intended for use with fossil data (Simpson 1943). The main purpose of the current paper is to address this particular issue in the specific context of multivariate analysis.

It is well known that other, more popular similarity measures such as the Dice coefficient are downward-biased when sampling is partial (Chao et al. 2005). This kind of bias could distort an ordination by exaggerating apparent differences between pairs of species lists, which should amplify what is called the arch effect (Hill 1973; Holland et al. 2001). An arch is present when the second-axis scores in an ordination are a quadratic function of the first-axis scores. To put this another way, the two ends of the primary gradient are pulled together on the second axis because the

dissimilarity between samples at the endpoints has reached or at least approached the maximum possible value.

In this paper I show how to ameliorate the arch effect and disentangle preservational biases from biological signals by using a better distance measure. It is the additive complement of a newly described similarity metric (Alroy 2015) that is an emendation of one going back to Forbes (1907). Forbes's measure F is:

$$F = a N / [(a + b)(a + c)] \qquad (1)$$

where a = the number found in both of the samples being compared, b and c = the numbers found only in one sample or the other, d = the number found in neither sample, and $N = a + b + c + d$.

Forbes's index has two major problems that are particularly serious in a paleoecological context. First, d is a meaningless number unless the data set captures effectively all of the species that existed in the biological system being studied. Such a thing is almost never true of fossil data. Second, Forbes's index is demonstrably upward-biased when sampling is very good. This bias is at its worst when true similarity, meaning the value $a'/(a' + b')$ or $a'/(a' + c')$ where the prime symbol indicates that there is no sampling bias, is around 50%.

The first problem is easily solved by ignoring the term d in computing N. Call this smaller sum $n = a + b + c$. Although this simple modification helps, it does not fully compensate for the upward bias. As has been shown (Alroy 2015), adding some mild, heuristic correction terms to the equation does:

$$F = a(n + \mathrm{sqrt}\ n)] / [a(n + \mathrm{sqrt}\ n) + 3/2\, b\, c)] \qquad (2)$$

This new index ranges strictly between 0 and 1, unlike equation (1) (which can exceed 1 when the number of species shared is greater than expected by chance). Specifically, it is equal to 0 whenever the two samples share no species and equal to 1 whenever they are identical or one is a subset of the other.

Meanwhile, Simpson's solution to the unequal sample-size problem (Simpson 1943) is to use the ratio $S = a/(a + b)$ if $b < c$ and otherwise $a/(a + c)$. In other words, his index is the number of shared species divided by the richness of the poorer sample. Simpson's index

is still sometimes used in paleontology (e.g., Tsubamoto et al. 2004).

Finally, another metric worthy of attention is the Dice (also known as Sorenson) coefficient $D = 2a/(2a + b + c)$. Many others exist (Choi et al. 2010), but the Dice coefficient has robust theoretical properties (Hubálek 1982) and is very widely used in ecology, so it is the focus of much discussion in this paper. All of the criticisms leveled against it apply equally well to other widely known metrics such as the Jaccard index (which is even more downward-biased), the Ochiai index, the Kulczynski index, and so on.

After showing by simulation that the adjusted Forbes index works well even when sample sizes are uneven, I then present two simple empirical cases in which a conventional ordination either collapses biogeographic signals or produces an arch, but an ordination employing F' removes both problems. No claim is made that this approach is a panacea, but the results suggest that use of F' by paleoecologists and others should be considered a serious option.

Simulation Analysis

Analyses of simulated data matrices are used here to illustrate the different properties of the indices, and in particular to demonstrate that the Forbes index solves the crucial paleoecological problem of uneven sampling. The matrices were created using a Monte Carlo procedure. Species richness in each of two partially overlapping species pools was set to 100. The pools had 50% overlap, meaning that 50 species were found in both, 50 were found only in the first, and 50 were found only in the second. To construct the sampling pool, relative species abundances were generated by drawing numbers at random from a normal distribution with a mean of zero and a standard deviation of 2.5, exponentiating them, and dividing them through by their sum. Two sampling levels were assumed: either 100 or 1000 specimens. Specimens were drawn randomly to create 40 samples representing the four possible combinations of sample sizes and pool identities and the similarity measures were then computed. The similarities were converted to distances
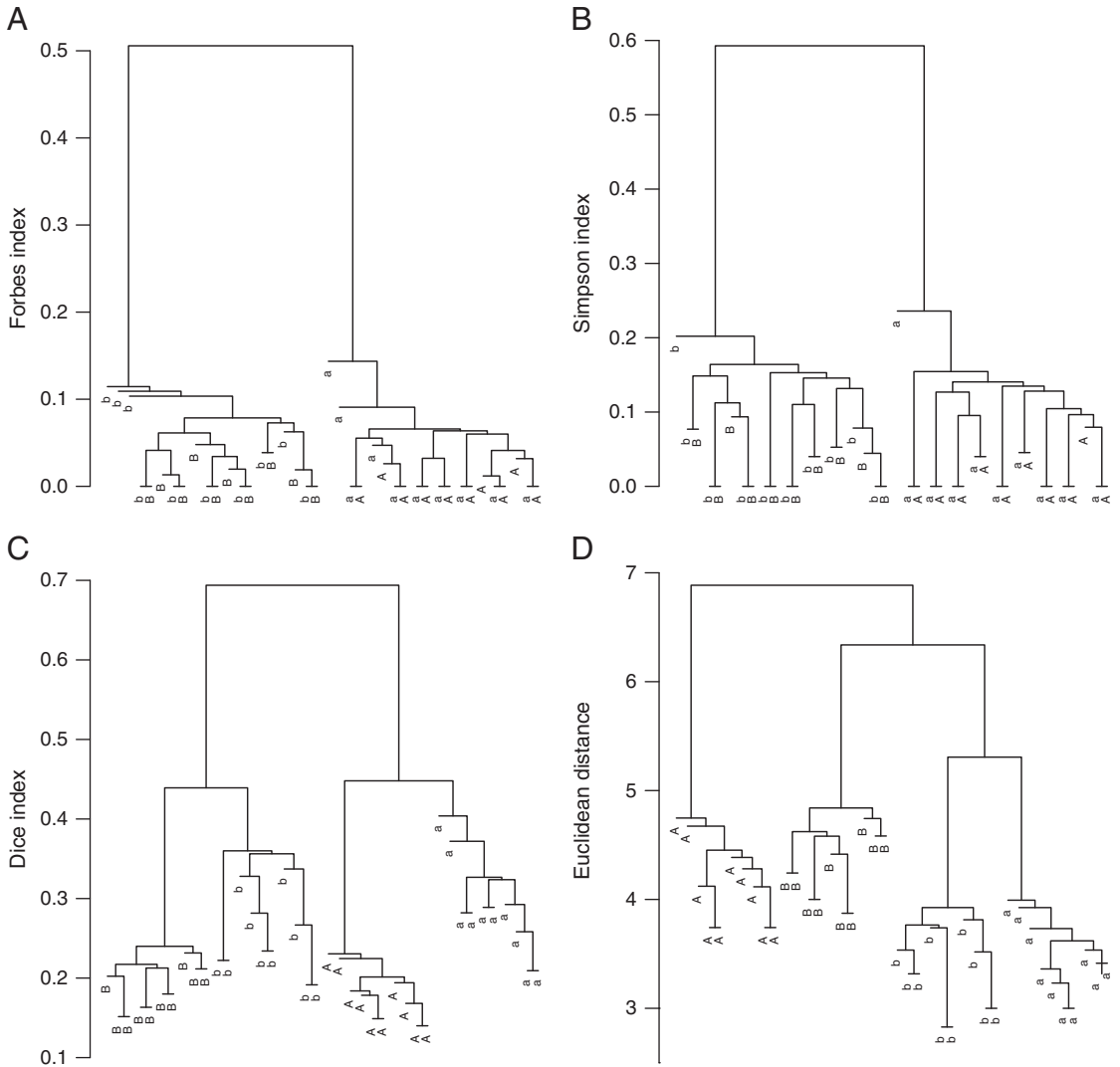
FIGURE 1. UPGMA (unweighted pair group with arithmetic mean) cluster dendrograms based on analyses of a simulated 40-sample data set using four different distance metrics. Computations were carried out using the R function hclust. "A" and "B" indicate 1000-specimen samples and "a" and "b" indicate 100-specimen samples. "A" and "a" samples are drawn from one pool and "B" and "b" samples from another. A, Distances are based on the Forbes index. B, Simpson index. C, Dice index. D, Euclidean distances.

by subtracting them from 1, which is necessary to make cluster analysis and ordination possible.

An unweighted pair-group (UPGMA) cluster analysis shows that the corrected Forbes index is almost completely unaffected by sample size (Fig. 1A). It breaks the samples down by species pool ("a" plus "A" or "b" plus "B" in the figure) and fully intermixes the large and small samples. Indeed, it resolves the hierarchy down into pairs that consistently

include one large sample and one small sample. The apparent reason is that each small sample is usually more similar to some of the large ones than to most or any of the small ones, thanks to high random error in its presence-absence pattern. The cluster dendrogram also puts almost all the tips at a roughly equal distance from the root (i.e., it is nearly ultrametric) and correctly shows that differences between samples in the same class are small.

By contrast, Simpson dissimilarities suggest that there are fairly substantial differences within the a + A and b + B clusters (Fig. 1B). The dendrogram is also not nearly as ultrametric. Worst, Dice dissimilarities and Euclidean distances are completely fooled by sampling bias: each of these metrics creates four clusters that correspond to combinations of species pools and sample sizes (Figs. 1C,D). The dendrograms are again not nearly as close to being ultrametric, and they suggest even larger differences between samples that are illusory. In a word, these methods reflect sampling just as strongly as they do the underlying biological pattern.

Multivariate ordination shows exactly the same thing. Principal coordinates analysis (PCoA; Gower 1966) is the parametric equivalent of NMDS and is therefore appropriate for showing how distance measures perform without obscuring their scaling properties relative to one another. A standard correspondence analysis (CA; Hill 1973) is also illustrated because this method has a strong reputation in ecology (Gauch 1982; Digby and Kempton 1987; Legendre and Gallagher 2001). All of the treatments recover the difference between the two species pools on the first axis. However, the PCoA plot for the corrected Forbes index (Fig. 2A) shows that it minimizes the sampling signal on the second axis. By contrast, the other methods pull apart the small and large samples and only the Simpson index comes close to removing the bias (Fig. 2B–D). These results suggest that any analysis of paleoecological data could be distorted simply because studies presenting order-of-magnitude variation in sample size are routine. Bonelli et al. (2006) and Bush and Brame 2010) are just two examples out of many.

Pleistocene Faunal Analysis

The Pleistocene record of North American mammals is a good test case for methods because there are strong biogeographic signals (Graham et al. 1996) in addition to an overriding taphonomic bias: most assemblages are dominated either by extinct megafaunal species or by extant small mammals. Thus, patterns in ordination plots are easily interpretable and a best-case scenario is easy to define: the first two axes should capture geography and the samples should not segregate strongly on the basis of body mass.

Forty seven late Pleistocene species lists were downloaded from the Fossilworks data portal (http://fossilworks.org) on 10 March 2014. Almost all of the lists were compiled by myself and Mark Uhen; many derive from the North American Mammalian Paleofaunal Database (Alroy 1999) and others from the Paleobiology Database. The data were spatially restricted to the United States because the fossil record of Mexico and Canada is relatively poor in this interval. Specifically indeterminate records and records of marine mammals and bats were excluded, and collections with fewer than ten species were discarded in order to make sure that similarity indices could be computed with a fair degree of precision. This treatment also minimizes the effect of sample-size variation, making the analysis more conservative. Average log body-mass values were computed for each collection by matching the species lists to the comprehensive global data set of Smith et al. (2003), which includes estimates for extinct megafauna in addition to Recent species.

Similarities were computed using the adjusted Forbes F', Simpson S, and Dice D coefficients and subtracted from 1 to produce distances. The distance matrices were then ordinated using PCoA, which again is the metric equivalent of NMDS. Because correct scaling is an apparent characteristic of F' (Fig. 1) (Alroy 2015), use of PCoA was considered to be a more relevant test of its properties than use of NMDS. As with the simulation, the data were subjected to a correspondence analysis (CA). CA was used instead of DCA because removing the arch effect in these data with such a brute-force method would make it harder to test the hypothesis that the F' does minimize sampling biases.

All of the PCoA ordinations clearly suggest that there are three large and well-spaced groupings of assemblages, each of which corresponds to a geographic region (Fig. 3A–C). These clusters are of samples from the southeastern United States (most of which are from Florida), the northeastern United States, and the western United States. This pattern
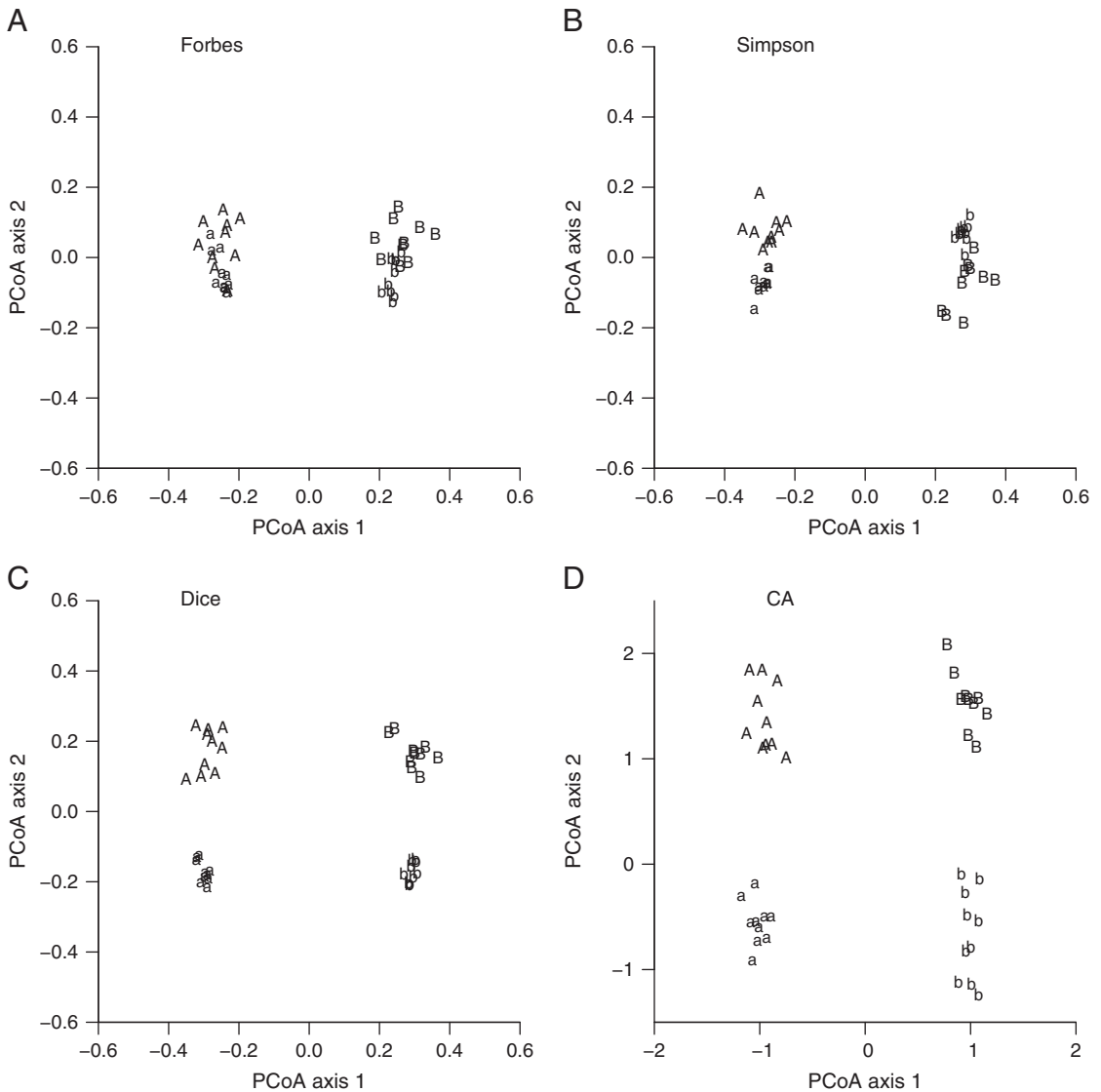
FIGURE 2.   Multivariate ordinations of the same simulated data set that was subjected to cluster analysis (Fig. 1). Letters correspond to groups of samples also illustrated in Fig. 1. A, Principal coordinates analysis (PCoA) with distances based on the adjusted Forbes index. B, PCoA based on the Simpson index. C, PCoA based on the Dice index. D, Correspondence analysis.

is unsurprising. For example, Graham et al. (1996) performed a TWINSPAN cluster analysis of similar FAUNMAP data and found an almost identical three-way primary split in the data. Simpson (1964) showed that the east-west biogeographic split corresponds to a sharp drop in species richness at the boundary between the Rocky Mountains and the Great Plains. Finally, Hagmeier and Stults (1964) proposed a North American mammalian

biogeographic classification that would place most of the Florida and northeastern U.S. fossil samples within just two provinces (respectively, the Austroriparian and Illinoian). Thus, the current results and those of Graham et al. (1996) may simply reflect the coincidence between sampling gaps and provincial boundaries.

Nonetheless, the ability of the methods to recover these partially artifactual patterns is of keen methodological interest. Only F' (Fig. 3A)
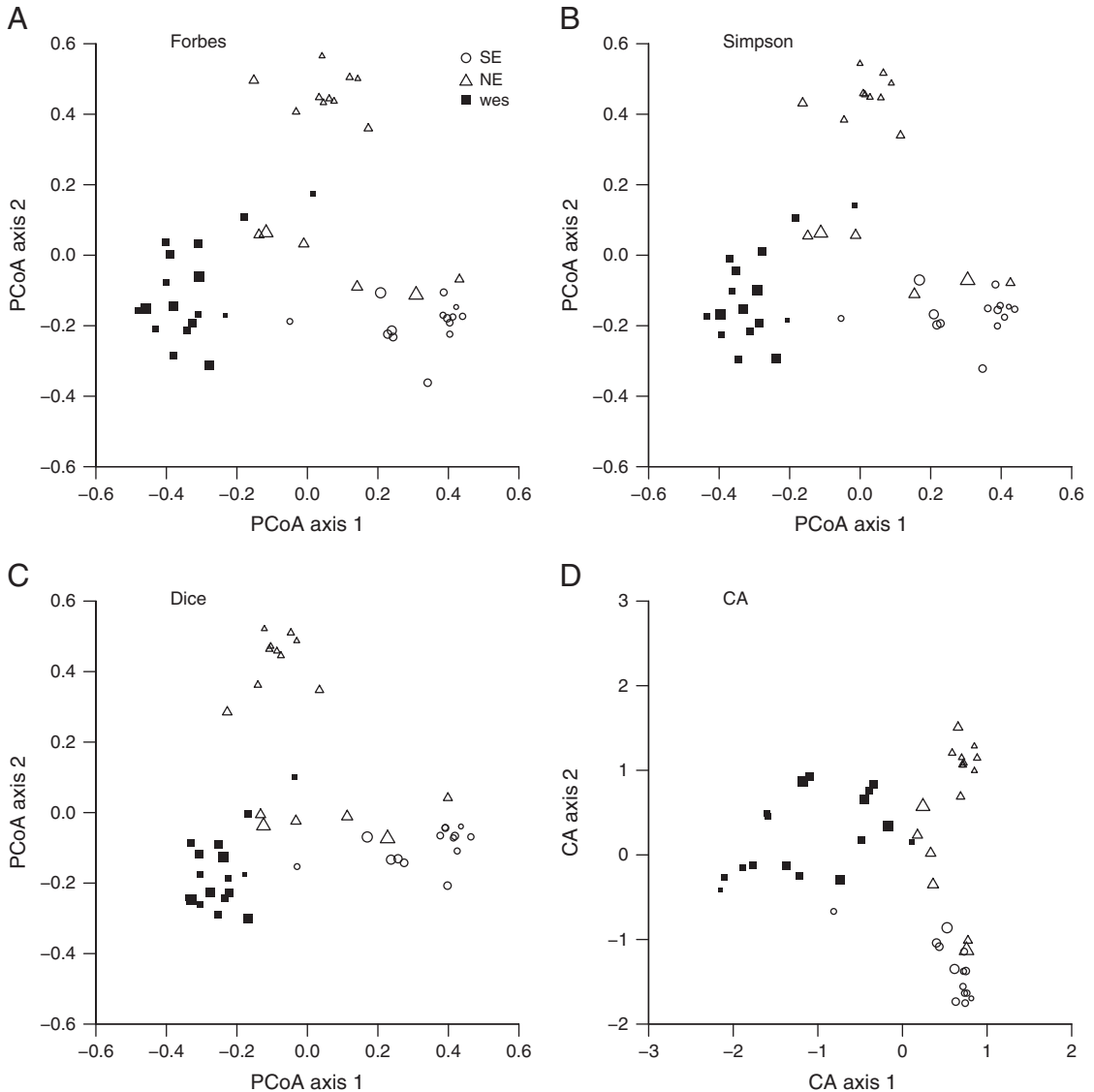
FIGURE 3. Multivariate ordinations of Pleistocene mammal data based on four different methodologies. NE = northeastern U.S. (open triangles); SE = southeastern U.S. (open circles); west = western U.S. (solid squares). Point sizes are directly proportional to the log mean body mass of the species in a given sample. A, Principal coordinates analysis (PCoA) with distances based on the adjusted Forbes index. B, PCoA based on the Simpson index. C, PCoA based on the Dice index. D, Correspondence analysis.

and the Simpson index (Fig. 3B) generate an entirely clean separation on axis 1 between samples from the east and west. By contrast, the CA ordination (Fig. 3D) tightly clusters all of the eastern U.S. samples into one end of the first axis; it also suggests relatively much greater variation amongst the western samples and obscures the gaps between the three groupings. Meanwhile, the Dice coefficient

(Fig. 3C) causes samples from the northeast and west to overlap on axis 1.

All three PCoA plots pull out the northeastern U.S. samples on axis 2 (Figs. 3A–C). These assemblages happen to be dominated by small mammals, and other northeastern samples dominated by large mammals fall closer to the centers of the plots. The likely reason is not that the large mammal samples

are either particularly rich or poor, but rather that large mammals tend to have large geographic ranges in the Recent and this pattern also held during the Pleistocene. Thus, large mammal-dominated assemblages are unlikely to capture regional differences. Axis 2 therefore seems to present a mixture of geographic and taphonomic signals.

Of more interest, the axis 2 pattern suggests that CA in fact does suffer from an arch effect, with the western and southeastern points making up the ends of the arch and the northeastern points forming the midpoint (Fig. 3D). Thus, it appears that using any similarity index combined with PCoA does ameliorate the arch effect in this data set.

The final difference between the methods is that D and to some extent S compress the points within each grouping. One way to quantify this difference is use the axis 1 and 2 scores to compute nearest neighbor differences for the samples. The median values are respectively 0.0510, 0.0470, and 0.0406 in the F', S, and D plots. The F'-based values are significantly higher than the S- and D-based values according to a paired two-sample Wilcoxon test ($P = 0.0534$, 0.0036).

This compression within the PCoA space cannot be attributed in a simple way to compression of similarity values that is demonstrated by simulations (Alroy 2015). The reason is that D should overestimate the dissimilarity of neighboring points in the ordinations, which should spread them out. A better explanation is that D is more badly fooled by uneven sample sizes (Figs. 1, 2) and therefore pushes the northeastern small mammal assemblages farther out, which in turn compresses the rest of the ordination space (given the way PCoA works). If correct, this explanation again suggests that D has bigger problems than simple compression.

Combined Pleistocene and Recent Faunal Analysis

Given that unequal sample size is the main topic of concern here, a very strong test of the Forbes coefficient would be to add large samples to the analysis. If the method is robust in the way that is suggested by the simulations (Figs. 1, 2), then large samples should fall in biologically reasonable locations. If not, then they should cluster together. Just such an analysis is made possible by the fact that virtually complete mammalian species lists for North American biomes have been presented by Brown and Nicoletto (1991).

Adding the biome lists and subtracting the extinct species from the fossil lists shows that as with the simulated data (Fig. 2) the Dice coefficient and CA are entirely misled by this newly synthesized data set's huge variation in sample size (Fig. 4). D segregates the biome lists into an arc at one edge of the plot (near the western U.S. samples: compare Figs. 3C and 4C) whereas CA produces a straightforward arch that pushes most of the fossil data points onto one arm (Fig. 4D). Note that the biome lists have a median length of 68 species whereas the fossil lists have a median of 16. Thus, the poor performance of D and CA is indeed easily explained as an artifact of sampling.

In principle, the Pleistocene samples and the biome lists might not mix together cleanly because some fossil assemblages bring together species not found today in the same places (i.e., are "disharmonious"). However, these species form a small minority and most Pleistocene assemblages are thought to represent current biomes (Graham et al. 1996). Indeed, the positions of the biome lists in the F' plot (Fig. 4A) do make biogeographic sense: the grasslands biome falls squarely in the middle of the plot, the eastern forest biome is next to the northeastern samples, the western biome and fossil sample points are mixed together, the taiga and tundra samples fall between the western and northeastern U.S. samples, and the Mexican biomes fall in between the western and southeastern samples at the bottom of the plot.

Patterns in the S-based ordination (Fig. 4B) are highly similar to those in the F' plot (Fig. 4A), raising the question of whether S is just as good as F'. This question is easily dismissed by noting five things. First, the simulated cluster analyses show that S exaggerates differences between nearly identical samples (Fig. 2). Second, two-sample simulations (Alroy 2015) suggest that S is categorically downward-biased, so regardless of its
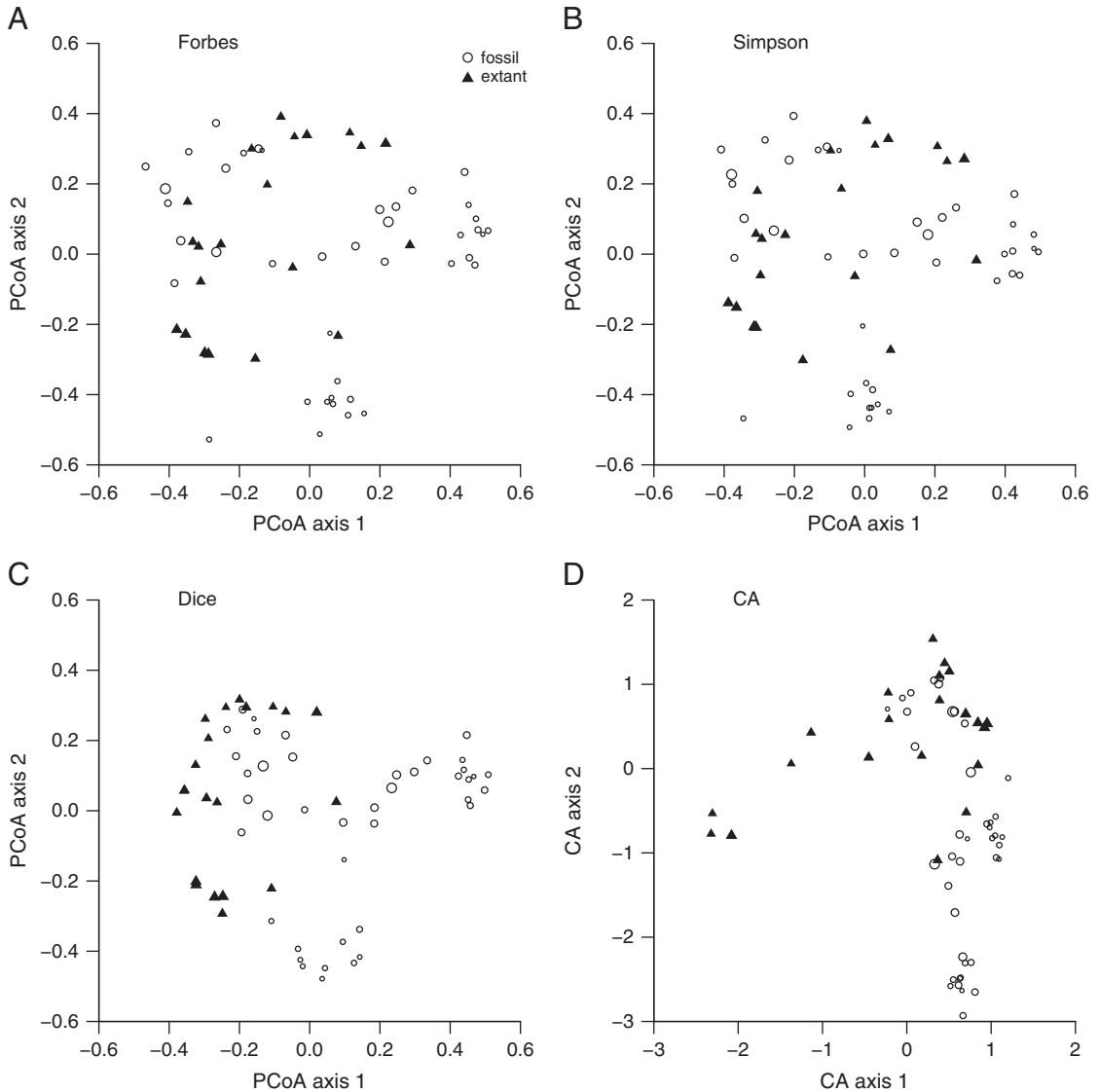
FIGURE 4. Multivariate ordinations of combined data for Recent mammals found in Pleistocene fossil assemblages (open circles) and for mammals living today in North American biomes (closed triangles). Point sizes are proportional to body mass, as in Figure 3. A, PCoA with distances based on the adjusted Forbes index. B, PCoA based on the Simpson index. C, PCoA based on the Dice index. D, Correspondence analysis.

performance in a cluster analysis or ordination the actual raw values are suspect. Third, F' includes terms explicitly intended to account for sample size (eq. 2) and S does not. Fourth, S ignores the richness of the larger sample, and it seems counterintuitive to assume that this value includes no biologically useful information at all. Finally and most importantly, S is a ratio of just two numbers and is therefore subject to high binomial error.

Discussion

The choice of analytical methods in paleoecology is extremely complex, and no single recommendation could or should be taken to heart by all practitioners. Nonetheless, the difference in performance between the adjusted Forbes index and its rivals is rather pronounced in the current analysis (Figs. 1–4). For example, although the Simpson index is

clearly adequate for recovering patterns in the Pleistocene data set (Figs. 3, 4) it is clearly inadequate when patterns are more challenging (Figs. 1, 2). The Forbes index's theoretical properties also seem to be very robust (Alroy 2015), whereas the Simpson index rather uncomfortably throws out one of the three underlying counts used by all other metrics.

In sum, it is not really clear how the ordination and cluster analysis results could be more intuitive or seem to be less biased: use of the modified Forbes index really does seem to solve major problems with sampling. Nonetheless, some options need to be mentioned. One of them is DCA. There is good agreement among ecologists (Legendre and Gallagher 2001) that sample data should not be subjected to principal components analysis (PCA), that the family of correspondence analysis methods is more appropriate, and that DCA is an improvement on basic CA (Gauch 1982; Holland et al. 2001). Indeed, regardless of how the underlying data are transformed PCA in particular will routinely produce a strong arch, and it often produces an arch so severe that the endpoints fold back into the middle of the first axis (i.e., a horseshoe effect [see Legendre and Gallagher 2001]). The arch effect is certainly the main concern of theorists working in this area, and DCA does remove it.

There are, however, reasons to believe that metric multidimensional scaling (i.e., PCoA) and NMDS are even better than DCA (Bush and Brame 2010). PCoA certainly does seem to work when it is applied to the simulated data and based on Forbes distances (Fig. 2A). However, it must be emphasized that scaling methods are only as good as the underlying similarity metrics they use. To be specific, PCoA based on the Euclidean distance replicates PCA whereas PCoA based on the chi-square distance replicates CA (Digby and Kempton 1987). These two metrics and all of the older binary similarity coefficients fail exactly because they compress similarity values and therefore distance values (Alroy 2015). Far from being ad hoc or heuristic, using F' in combination with PCoA (Fig. 2A) is therefore entirely reasonable on theoretical grounds and preferable on practical grounds.

The arch effect does have more to do with the presence of zero similarities than with the compression of non-zero similarities by biased metrics (Gauch 1982). The test data sets used in this paper do produce an arch effect when a conventional method is used (Figs. 3D, 4D) and do include a good number of zero similarities, so there are reasons to believe that combining F' with PCoA will in general often eliminate the need to take drastic measures to remove the arch effect. However, if there are still concerns about zero similarities then a reasonable option would be to alter the data by using either of the related step-across and extended similarity methods (Williamson 1978; De'ath 1999). I have not done so here strictly in order to keep the analyses simple and thereby guarantee that any differences in performance among the methods really do stem from their intrinsic properties.

Acknowledgments

Literature Cited

Alroy, J. 1999. Putting North America's end-Pleistocene megafaunal extinction in context: large scale analyses of spatial patterns, extinction rates, and size distributions. Pp. 105–143 *in* R. D. E. MacPhee, ed. Plenum, New York.

——. 2015. A new twist on a very old binary similarity coefficient. Ecology (in press).

Bonelli, J. R., Jr., C. E. Brett, A. I. Miller, and J. B. Bennington. 2006. Testing for faunal stability across a regional biotic transition: quantifying stasis and variation among recurring coral-rich biofacies in the Middle Devonian Appalachian Basin. Paleobiology 32:20–37.

Brown, J. H., and P. F. Nicoletto. 1991. Spatial scaling of species composition: body masses of North American land mammals. American Naturalist 138:1478–1512.

Bush, A. M., and R. I. Brame. 2010. Multiple paleoecological controls on the composition of marine fossil assemblages from the Frasnian (Late Devonian) of Virginia, with a comparison of ordination methods. Paleobiology 36:573–591.

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters 8:148–159.

Choi, S.-S., S.-H. Cha, and C. C. Tappert. 2010. A survey of binary similarity and distance measures. Systemics, Cybernetics and Informatics 8:43–48.

De'ath, G. 1999. Extended similarity: a method of robust estimation of ecology distances from high beta diversity data. Plant Ecology 144:191–190.

Digby, P. G. N., and R. A. Kempton. 1987. Multivariate analysis of ecological communities. Chapman and Hall, London.

Gauch, H. G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Forbes, S. A. 1907. On the local distribution of certain Illinois fishes: an essay in statistical ecology. Bulletin of the Illinois State Laboratory of Natural History 7:272–303.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.

Graham, R. W., E. L. Lundelius, Jr., M. A. Graham, E. K. Schroeder, R. S. Toomey, III, E. Anderson, A. D. Barnosky, J. A. Burns, C. S. Churcher, D. K. Grayson, R. D. Guthrie, C. R. Harington, G. T. Jefferson, L. D. Martin, H. G. McDonald, R. E. Morlan, H. A. Semken, Jr., S. D. Webb, L. Werdelin, and M. C. Wilson. 1996. Spatial response of mammals to Late Quaternary environmental fluctuations. Science 272:1601–1606.

Hagmeier, E. M., and C. D. Stults. 1964. A numerical analysis of the distributional patterns of North American mammals. Systematic Zoology 13:125–155.

Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. Journal of Ecology 61:237–249.

Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis, an improved ordination technique. Vegetatio 42:47–58.

Holland, S. M., A. I. Miller, D. L. Meyer, and B. F. Dattilo. 2001. The detection and importance of subtle biofacies within a single lithofacies: the Upper Ordovician Kope Formation of the Cincinnati, Ohio region. Palaios 16:205–217.

Hubálek, Z. 1982. Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. Biological Reviews 57:669–689.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280.

Reyment, R. A. 1963. Multivariate analytical treatment of quantitative species associations: an example from palaeoecology. Journal of Animal Ecology 32:535–547.

Shepard, R. N. 1962. The analysis of proximities: multidimensional scaling with an unknown distance function. II. Psychometrika 27:219–246.

Simpson, G. G. 1943. Mammals and the nature of continents. American Journal of Science 241:1–31.

——. 1964. Species density of North American Recent mammals. Systematic Zoology 13:57–73.

Smith, F. A., K. Lyons, S. K. M. Ernest, K. E. Jones, D. Kaufman, T. Dayan, P. A. Marquet, J. H. Brown, and J. P. Haskell. 2003. Body mass of late Quaternary mammals. Ecological Archives E084–E094.

Tsubamoto, T., M. Takai, and N. Egi. 2004. Quantitative analyses of biogeography and faunal evolution of middle to late Eocene mammals in East Asia. Journal of Vertebrate Paleontology 24:657–667.

Valentine, J. W., and R. G. Peddicord. 1967. Evaluation of fossil assemblages by cluster analysis. Journal of Paleontology 41:502–507.

Williamson, M. H. 1978. The ordination of incidence data. Journal of Ecology 66:911–920.