

A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness

Graeme T. Lloyd*

Department of Palaeontology, Natural History Museum,
Cromwell Road, London SW7 5BD, UK
*g.lloyd@nhm.ac.uk

Modelling has been underdeveloped with respect to constructing palaeobiodiversity curves, but it offers an additional tool for removing sampling from their estimation. Here, an alternative to subsampling approaches, which often require large sample sizes, is explored by the extension and refinement of a pre-existing modelling technique that uses a geological proxy for sampling. Application of the model to the three main clades of dinosaurs suggests that much of their diversity fluctuations cannot be explained by sampling alone. Furthermore, there is new support for a long-term decline in their diversity leading up to the Cretaceous–Paleogene (K–Pg) extinction event. At present, use of this method with data that includes either Lagerstätten or ‘Pull of the Recent’ biases is inappropriate, although partial solutions are offered.

Keywords: modelling; sampling bias; palaeobiodiversity; subsampling; dinosaurs

1. INTRODUCTION

In a groundbreaking and far-sighted paper, Raup [1] identified a number of problems with obtaining estimates of palaeobiodiversity. In addition, he offered two ways in which sampling bias could be addressed, namely subsampling and modelling. The former eventually led to the major research effort encapsulated by the Palaeobiology Database (<http://paleodb.org/>; [2]). However, the latter remained substantially undeveloped until Smith & McGowan [3] introduced a novel approach that corrected for rock availability by using a rock record proxy (number of maps with outcropping rock). The model assumes that true diversity is actually constant and observed diversity is purely a product of the sampling proxy. By comparing the predictions of such a model to actual values, we can identify portions of a palaeobiodiversity curve that are genuine excursions from the proxy-biased model and hence require other explanations.

One contribution of 12 to a Special Feature on ‘Models in palaeontology’.

However, the Smith & McGowan [3] approach has a number of limitations at present: (i) it assumes a linear relationship between logged diversity and sampling proxy data, (ii) it does not offer a significance test for any excursions, and (iii) it does not offer an easily applicable modelling approach to search for any remaining medium-term trends (their hinge regression—their fig. 5*b*—was fitted by eye). Barrett *et al.* [4] presented a possible solution to the second limitation by using the standard deviation of the residuals. However, this led to the seemingly contradictory situation, whereby the group that had the poorest fit to the model (sauropodomorphs) also had the fewest statistically significant excursions. Here, I develop an improved model-based approach that offers solutions to all of these limitations, and discuss modelling as a tool for removing sampling signal from palaeobiodiversity curves.

2. MATERIAL AND METHODS

The method outlined here requires just three items of data for input: the diversity values, the sampling proxy values and the numerical dates (in millions of years). The rest follows a step-by-step protocol that builds upon that of Smith & McGowan [3, pp. 766–767] as follows:

- The diversity measure and sampling proxy are sorted independently from lowest to highest.
- A model is now fitted to this data. (Smith & McGowan [3] used just a linear model, but here nonlinearity is catered for by additionally fitting logarithmic, exponential, hyperbolic, sigmoidal and polynomial models.)
- The ‘best’ model is chosen by calculating the sample size-corrected Akaike Information Criterion, the AIC_c [5], and the standard errors and deviations of this model are stored for later reference (see below).
- This model is then used to calculate predicted values of diversity for each sampling value in their correct time-series order.
- Residuals are created by subtracting these predicted values from the actual observed values for a sampling-corrected palaeobiodiversity estimate.
- Residuals may then be plotted alongside 1.96 standard errors or deviations using the values stored in step 3 as 95% confidence intervals. These thus provide more appropriate error bars than those of Barrett *et al.* [4] as they more accurately reflect significant excursions from the sampling-driven model.
- Medium-term (multi-time bin) trends are recovered by using the Multivariate Adaptive Regression Splines (MARS) approach of Friedman [6]. This is a more statistically robust method for identifying hinge points in a time series that automatically minimizes the residual sum of squares (RSS). (When applied to the Smith & McGowan [3] data, not shown, this approach was essentially congruent, although additional hinge points were also recognized.)
- Finally, a simple linear model, effectively a MARS with only one spline, is also fit to the data and the best multi-hinge point model is compared with this using the AIC to ensure its optimality. In other words, the fit must offer a sufficient improvement to be worth the extra complexity of the multiple hinge points.

The entire process has been automated in R [7] and is made freely available for use via the author’s website (<http://www.graemetlloyd.com/>) and with the data at DRYAD (doi:10.5061/dryad.8949).

To test the effects of this extended method, I apply the new approach to the dinosaur data of Barrett *et al.* [4], an occurrence-level list based on an older database [8]. Other published datasets were considered, but for reasons covered in the discussion these were deemed inappropriate.

3. RESULTS

When the modelling approach outlined above is applied to Barrett *et al.* [4] species-level dinosaur data, sauropodomorphs show a poor fit to the sampling-driven model and ornithischians a good fit, as noted by Barrett *et al.* [4]. However, now theropods show a considerable number of points outside the (standard deviation)

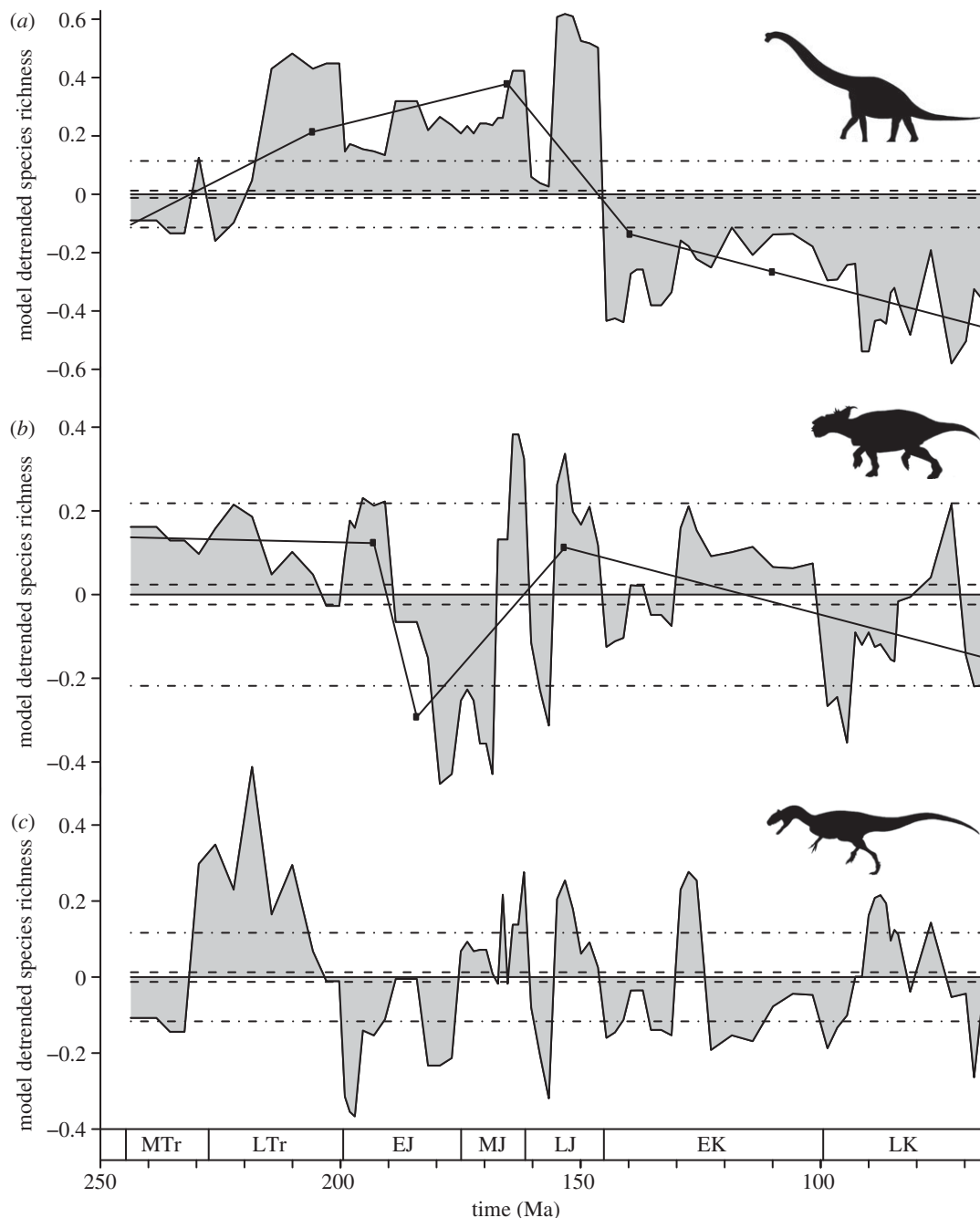


Figure 1. Time series of residuals (grey polygon) from a modelling approach that assumes true taxonomic richness is constant and apparent richness is driven purely by sampling. Dinosaur data from Barrett *et al.* [4] for (a) sauropodomorphs, (b) ornithischians and (c) theropods. Dashed line indicates 1.96 standard errors and dashed-dotted line 1.96 standard deviations of the model. Solid lines with closed circles (hinge points) in (a) and (b) are the results of a MARS analysis [6]. Geological epoch abbreviations are as follows: MTr, Middle Triassic; LTr, Late Triassic; EJ, Early Jurassic; MJ, Middle Jurassic; LJ, Late Jurassic; EK, Early Cretaceous and LK, Late Cretaceous.

95% confidence interval (figure 1). The MARS results also show that these clades exhibit different medium-term trends. Sauropodomorphs show an initial rising trend up to around the Jurassic–Cretaceous boundary followed by a decline, consistent with previous interpretations for the group. Ornithischians, on the other hand, initially show level diversity, followed by a major trough in the Early–Middle Jurassic and a slower decline through the Cretaceous. However, most of these medium-term trends are safely contained within the confidence intervals of the sampling-driven, i.e. constant diversity model. Theropods are unusual

here in that they show no clear medium-term trends, with a single linear model considered optimal for the group, although there are many short-term fluctuations.

Using the same data, Barrett *et al.* [4] argued that the results showed a ‘diminution of ornithischian and theropod dinosaur lineages prior to the K–P extinction event’ (p. 2671). There is support for that contention here. For both sauropodomorphs and ornithischians, their medium-term trends show a decline leading up to their extinction at the K–Pg, and all three clades exhibit lower than predicted richness in the bins preceding the K–Pg (figure 1).

4. DISCUSSION

The use of modelling as a tool for use in palaeobiodiversity analysis has been relatively unexploited despite offering some clear advantages. Perhaps the most obvious of these is that modelling is less demanding of a large sample size and has hence become popular among tetrapod workers (e.g. [4,9]). However, it is also much more flexible in the number of sampling biases that can be considered. For example, it is not straightforward how you could subsample with respect to map area or rock volume. A more pragmatic advantage is that modelling is effectively instantaneous with modern computing, whereas some subsampling methods, particularly at large sample sizes can take a considerable time to run. Finally, it is always more enlightening to use multiple methods that purport to perform the same task in order to reinforce a shared result or identify weaknesses where there is conflict. Modelling thus adds a useful alternative to subsampling that enables for comparative interpretation (e.g. [10]).

Application of the specific modelling approach used here gives results that have some important implications for studying sampling-biased diversity curves. Firstly, it is notable that for all three clades analysed here different results are obtained from those of Barrett *et al.* [4]. As the data are identical, the only explanation is that a nonlinear (polynomial) relationship between sampling and diversity is preferred in every case, justifying this inclusion in the modelling process. Secondly, it is clear that in the majority of test cases analysed, the null model of constant diversity is a poor or at least incomplete one. This is perhaps a more encouraging result than the high correlations commonly found between sampling and diversity, because it implies that more than simple sampling bias drives palaeobiodiversity curves. Finally, medium-term trends are a common feature of the sampling-corrected time series, which can be interpreted as (probably logistic [11]) rising or falling changes in palaeobiodiversity. However, it should be noted that the current picture of declining dinosaur diversity may just be a feature of an out-of-date dataset, for example, a more up-to-date sauropodomorph dataset [10] suggests their Cretaceous diversity may not be as depressed as shown here.

An outstanding issue with this approach is whether the results can really be considered as removing sampling bias alone and whether the remaining signal can be considered biological. Under the 'common cause' model of Peters [12] some measures of sampling would be expected to correlate with diversity for biological reasons, therefore to remove sampling signal would also mean removing biological signal. Unfortunately, at present both the sampling-biased and common cause interpretations make broadly similar predictions. However, methods are being introduced that help distinguish between competing signals [12,13], although Butler *et al.* [14] demonstrated statistically that continental flooding driven by sea-level change was not a plausible mechanism for common cause in dinosaurs. At present, this question remains unresolved and must be considered separately from the approach used here.

In proposing this method originally, Smith & McGowan [3] noted that the resulting residuals may represent either biological signal or some other unknown bias(es). One potential confounding bias may be 'Lagerstätten effects' [9], where a particular area, locality or collection yields far more taxonomic diversity than average owing to either exceptional preservation or exceptional palaeontological interest. The 'Pull of the Recent' [15] may also be considered a special case of this problem. (These two issues are why other datasets were not considered here.) Along with the range-through approach in general, such data can artificially inflate diversity by unfairly separating richness and sampling in the first step of the modelling process (see §2). Consequently, this can lead to the appearance of a record that is relatively unbiased by sampling when in reality this is not the case. There are partial solutions to these problems. For example, the Pull of the Recent is simply avoided by using a sampled-in-bin rather than a range-through palaeobiodiversity curve. However, even if a criterion for recognizing a dominant influence of Lagerstätten in certain time bins is applied—such as when greater than 50 per cent of taxonomic richness within a time bin comes from a single formation or locality [9]—at present there is no method for incorporating this data into the modelling approach presented here beyond simply excluding them and hence a subsampling approach [11] may be preferable for such data.

In summary, the refined modelling approach developed here offers a new and simple means for subtracting sampling signal from diversity curves. Although it is not appropriate for every dataset, initial results suggest that once a specific sampling proxy is removed a greater degree of biological signal may be present in palaeobiodiversity data than previously thought.

I would like to thank Al McGowan for discussions of methodology and Andrew Smith and three anonymous reviewers for helpful comments on an earlier draft of this article. This research was supported by NERC grant NE/F016905/1 to Andrew Smith, Jeremy Young and Paul Pearson.

- 1 Raup, D. M. 1972 Taxonomic diversity during the Phanerozoic. *Science* **177**, 1065–1071. (doi:10.1126/science.177.4054.1065)
- 2 Alroy, J. *et al.* 2001 Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266. (doi:10.1073/pnas.111144698)
- 3 Smith, A. B. & McGowan, A. J. 2007 The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of western Europe? *Palaeontology* **50**, 765–774. (doi:10.1111/j.1475-4983.2007.00693.x)
- 4 Barrett, P. M., McGowan, A. J. & Page, V. 2009 Dinosaur diversity and the rock record. *Proc. R. Soc. B* **276**, 2667–2674. (doi:10.1098/rspb.2009.0352)
- 5 Johnson, J. B. & Omland, K. S. 2004 Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108. (doi:10.1016/j.tree.2003.10.013)
- 6 Friedman, J. H. 1991 Multivariate adaptive regression splines. *Ann. Stat.* **19**, 1–67. (doi:10.1214/aos/1176347963)

- 7 R Development Core Team. 2010 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See <http://www.R-project.org>.
- 8 Weishampel, D. B., Dodson, P. & Osmólska, H. (eds) 2004 *The Dinosauria*. Berkeley, CA: University of California Press.
- 9 Benson, R. B. J., Butler, R. J., Lindgren, J. & Smith, A. S. 2010 Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proc. R. Soc. B* **277**, 829–834. (doi:10.1098/rspb.2009.1845)
- 10 Mannion, P. D., Upchurch, P., Carrano, M. T. & Barrett, P. M. 2011 Testing the effect of the rock record on diversity: a multidisciplinary approach to elucidating the generic richness of sauropodomorph dinosaur through time. *Biol. Rev.* **86**, 157–181. (doi:10.1111/j.1469-185X.2010.00139.x)
- 11 Alroy, J. 2010 The shifting balance of diversity among major marine animal groups. *Science* **329**, 1191–1194. (doi:10.1126/science.1189910)
- 12 Peters, S. E. 2005 Geologic constraints on the macroevolutionary history of marine animals. *Proc. Natl Acad. Sci. USA* **102**, 12 326–12 331. (doi:10.1073/pnas.0502616102)
- 13 Marx, F. G. & Uhen, M. D. 2010 Climate, critters, and cetaceans: Cenozoic drivers of the evolution of modern whales. *Science* **327**, 993–996. (doi:10.1126/science.1185581)
- 14 Butler, R. J., Benson, R. B. J., Carrano, M. T., Mannion, P. D. & Upchurch, P. 2011 Sea level, dinosaur diversity and sampling biases: investigating the ‘common cause’ hypothesis in the terrestrial realm. *Proc. R. Soc. B* **278**, 1165–1170. (doi:10.1098/rspb.2010.1754)
- 15 Raup, D. M. 1979 Biases in the fossil record of species and genera. *Carn. Mus. Nat. Hist. Bull.* **13**, 85–91.