
Peer reviewed version

Link to published version (if available):
10.1098/rspb.2019.0091

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via The Royal Society at https://royalsocietypublishing.org/doi/10.1098/rspb.2019.0091. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
Does exceptional preservation distort our view of disparity in the fossil record?

Joseph T. Flannery Sutherland¹*, Benjamin C. Moon¹, Thomas L. Stubbs¹, Michael J. Benton¹.

Published as Joseph T. Flannery Sutherland, Benjamin C. Moon, Thomas L. Stubbs, Michael J. Benton. Does exceptional preservation distort our view of disparity in the fossil record?

286 Proceedings of the Royal Society B: Biological Sciences

Published version at https://royalsocietypublishing.org/doi/10.1098/rspb.2019.0091

ABSTRACT

How much of evolutionary history is lost because of the unevenness of the fossil record? Lagerstätten, sites which have historically yielded exceptionally preserved fossils, provide remarkable, yet distorting insights into past life. When examining macroevolutionary trends in the fossil record, they can generate an uneven sampling signal for taxonomic diversity; by comparison, their effect on morphological variety (disparity) is poorly understood. We show here that lagerstätten impact the disparity of ichthyosaurs, Mesozoic marine reptiles, by preserving higher diversity and more complete specimens. Elsewhere in the fossil record, undersampled diversity and more fragmentary specimens produce spurious results. We identify a novel effect, that a taxon moves towards the centroid of a Generalised Euclidean dataset as its proportion of missing data increases. We term this effect ‘centroid slippage’, as a disparity-based analogue of phylogenetic stemward slippage. Our results suggest that uneven sampling presents issues for our view of disparity in the fossil record, but that this is also dependent on the methodology used, especially true with widely used Generalised Euclidean distances. Mitigation of missing cladistic data is possible by phylogenetic gap filling, and heterogeneous effects of lagerstätten on disparity may be accounted for by understanding the factors affecting their spatiotemporal distribution.

Keywords: Lagerstätten, disparity, missing data, centroid slippage, Generalised Euclidean Distance
1 School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, UK. e-mails: jf15558@bristol.ac.uk, benjamin.moon@bristol.ac.uk, tom.stubbs@bristol.ac.uk, mike.benton@bristol.ac.uk.

1. Introduction

Palaeontologists express different attitudes to lagerstätten, deposits which have historically yielded exceptionally preserved fossils. Most see them as opportunities to encounter the true richness of past life, others as statistical outliers that distort the story presented by the fossil record [1]. The distortion is potentially huge; for example, only 10% of the Burgess Shale fauna would be seen under ‘normal’ conditions of fossilisation [2]. Alternatively, younger lagerstätten like Holzmaden or Jehol do not exhibit such high levels of biological novelty and many lagerstätte taxa co-occur in deposits lacking exceptional preservation. Understanding how lagerstätten bias the fossil record is therefore crucial to accurately reconstructing biodiversity through deep time.

The effects of lagerstätten sampling on taxonomic diversity are well documented [1,3,4], but do not necessarily extend to morphological variety (disparity), as these measures of biodiversity are generally decoupled [5,6,7]. Morphometric disparity analyses directly compare physical shapes between taxa, but their application in palaeontology is limited by the pervasive issue of missing anatomical data [8]. Cladistic disparity analyses use differential character states between taxa as a proxy for disparity, producing similar results to the morphometric approach [9] and readily accommodating missing character data. While that same missing data negatively affects measurement of disparity, lagerstätten positively affect character completeness [10,11,12]. Lagerstätte biases may therefore relate to the problem of missing data, introducing methodological issues to cladistic disparity analyses that are linked to biases in the geological record. Ciampaglio et al. [13] examined the effects of missing data on disparity by using artificial cladistic data sets that were up to 25% incomplete. Smith et al. [14] address the impacts of missing data at greater percentages, finding that disparity for a variety of metrics decreases as the proportion of missing cladistic data increases; this analysis, however, was performed on fossil data sets rather than artificial data sets where the true (complete data) disparity value can be measured. Consequently, the theoretical impacts of missing data at the highest percentages remain unexplored.

Several morphometric disparity studies have excluded lagerstätten taxa from their analyses [15,16,17,18]. Each found little evidence for lagerstätten biases, but these conclusions cannot
necessarily be extended to cladistic disparity analyses. Lagerstätten bias the cladistic disparity of pterosaurs, flying reptiles with delicate, hollow bones [4]. While diversity and disparity may be decoupled, their fossil record is strongly dependent on lagerstätten [19]; such deposits undoubtedly record a sizeable proportion of their disparity. Butler et al. [4] also used phylogenetic correction to infill missing character states in their cladistic matrix, preventing investigation of the potential interrelationship between lagerstätten, missing data and disparity. Brusatte et al. [20] found that phylogenetic correction for missing data in pterosaurs alters the results of disparity analysis in diverse ways depending on the data sets, phylogenetic trees and disparity comparisons used, and does not simply inflate disparity estimates. However, they caution against phylogenetic correction, except for clades with a large amount of missing data or patchy fossil records. It is, therefore, worth exploring the fossil records of clades not composed predominantly of lagerstätten to understand the interlinked effects of methodology, missing data and exceptional preservation on disparity.

Ichthyosaurs, swimming reptiles with large, robust skeletons, occupied shallow and deep marine environments with high preservation potentials, giving them a rich fossil record. Ichthyosaurs were among the most significant groups to radiate in the oceans after the Permian–Triassic mass extinction. New taxa described from the Panxian, Luoping, Xingyi, and Guanling lagerstätten of China [21], alongside longer-known materials from the Besano lagerstätte of Italy and Switzerland and the well-established global record of Triassic ichthyosaurs [22], document rapid accumulation of disparity and anatomical innovations through the Triassic [23]. Despite reduction in ecomorphospace by the Triassic–Jurassic extinction [24], ichthyosaurs remained ecologically prominent throughout most of the Mesozoic [25]. With their rich fossil record, comprising lagerstätten and non-lagerstätten deposits, ichthyosaurs make a suitable case study to investigate the impacts of exceptional preservation on our view of disparity.

We also examine disparity in artificial cladistic data sets, allowing for measurement of the true disparity signal, followed by the random removal of 0 to 100% of characters to ascertain the impact of higher levels of missing data. Ichthyosaur disparity is calculated here from a moderately incomplete data set (52% missing characters) without any phylogenetic correction method, and so our analysis of the effects of missing data provides a means of critiquing the signal that emerges.

2. Materials and methods

(a) Character and distance matrices
In cladistic disparity analyses, character-taxon cladistic matrices are converted to taxon-taxon distance matrices and then ordinated, allowing disparity metrics to be calculated as summary statistics from the ordination space. Most studies use either Generalised Euclidean Distance (GED) or Gower Coefficient Distance (GCD) matrices. The former has proved popular because incomparable distances are mathematically infilled; complete data sets allow for a greater range of subsequent analytical techniques. Lloyd [26], however, cautioned that this approach may result in illogical data that could negatively impact a disparity study. GCD has become increasingly popular as it has the advantage of accounting for the effects of missing data at the cost of removing taxa which do not share any characters and so are incomparable. Maximum Observable Rescaled Distance (MORD) operates in a similar way to GCD [26]; however, this has not yet been widely utilised as it has only recently been developed. Here, we use disparity metrics calculated from all three distances to ascertain how the choice of distance matrix may affect measurements of disparity as GCD and MORD matrices are designed to more readily accommodate missing data.

Our chosen cladistic matrix is modified from Moon [27] and comprises 112 ichthyosauriforms scored for 287 discrete characters, where characters 100, 204, 205 and 218 are ordered. Phalarodon major was removed as it is a nomen dubium [28] and the character states for Suevoleviathan integer and S. disinteger were combined as the latter is a junior synonym of the former [29]. GED, GCD and MORD distance matrices were generated from this using the MorphDistMatrix function of the Claddis package [26] in R (v3.2.2) [30]. In the GCD and MORD matrices, 14 taxa had incomparable distances and were removed prior to ordination. While GED matrices are commonly employed for the specific reason that they do not require taxon removal, those same taxa were then removed from the GED matrix to give taxon parity in subsequent statistical tests.

(b) Ordination

Each distance matrix was subjected to principal coordinates analysis (PCO) in R (v3.2.2) [30]. PCO may introduce negative eigenvalues into the resultant data; thus, their corresponding axis lengths, as the square roots of those eigenvalues, are imaginary. Negative eigenvalue correction, utilising either the Lingoes [31] or Cailliez [32] methods, are used to avoid this issue and are commonly employed in PCO analyses. Such methods, however, reduce the amount of variation represented by lower ordination axes and so erode their disparity signal [33]. As there is no consensus on the number of PCO axes to use in subsequent analyses [26], this may be problematic for studies that focus on few axes, such as
the first three. To ascertain the impacts of negative eigenvalue correction, the squared Pearson correlation between the pairwise distances in PCO space and the original distance matrix was plotted, following the method of Kotrc and Knoll [33], along with the percentage variance summarised by increasing numbers of PCO axes. This was conducted for each type of distance matrix, with and without Cailliez negative eigenvalue correction.

Largely, as the number of uncorrected PCO axes increases, the degree of correlation with the original distance matrix increases smoothly, plateauing as higher axes are included (Fig. S1). By contrast, while inclusion of all corrected PCO axes produces a perfect correlation with the original distance matrix, it fluctuates for lower numbers of axes. Even for uncorrected GCD and MORD matrices, the correlation declines sharply over the first three axes, those that are plottable and commonly used in the calculation of disparity metrics.

On the basis of these results, PCO without negative eigenvalue correction was performed on each type of distance matrix. We used all positive PC axes in subsequent analyses to maximise the correlation with its corresponding distance matrix and summarising the greatest possible amount of variance on the lower axes, while excluding axes with negative eigenvalues: 107 for GED, 62 for GCD and 62 for MORD.

The different degrees of correlation between the original distance matrices, and corrected and uncorrected PCO matrices, warn against arbitrary selection of axes or selection based on certain criteria, such as summarising a particular percentage of variance, as the chosen number of axes may be poorly correlated with the original distance matrix. The sharp drop in correlation over the first few axes even in the uncorrected PCO matrices further warns against selection of axes without first determining how well they represent the original data.

(c) Disparity calculations

The selected taxa were sorted into 10 Ma time bins, spanning 250 Ma to 90 Ma, and the PCO scores for each distance matrix subsampled for the taxa in each bin. Following the recommendations of previous studies [34,35], a 10 Ma time bin length was selected as a suitable trade-off between resolution and sample size. This approach was also selected, rather than division into epochs, to avoid confounding effects associated with uneven time bin lengths. Disparity metrics were then calculated using disparity function of the R package dispRity [36]. The chosen metrics were the sums of ranges and variances, and the mean distance from the centroid; 95% confidence intervals were calculated from 1000 bootstrapped pseudoreplicates with replacement. PCO plots of morphospace for each bin were created.
using the first two PCO axes. The same metrics and confidence intervals were then calculated for a second data set pruned of taxa that occur in lagerstätte deposits.

(d) Lagerstätten taxa

The definition of a lagerstätte remains controversial. Sedimentological, geochemical and taphonomic criteria have been described [37], but no consensus exists. The definition of a lagerstätte may also include a historical component. Many classical lagerstatten have been repeatedly sampled for decades or even centuries, partly due to the exceptional scientific and aesthetic quality of their specimens, partly due to high specimen availability resulting from systematic excavation in quarries or rapid erosional exposure (for example coastal settings).

Other factors may compound this historical component, for example the initial accessibility of a locality or the effort invested in the subsequent preparation and description of any collected specimens; all these elements are likely linked to some degree by the ‘bonanza effect’ where palaeontologists preferentially sample and describe fossiliferous units of high scientific potential over others [38], leading to sampling spikes for those units. Furthermore, as lagerstätten show heterogeneous fidelity of preservation between clades, localities that yield taxa used in our analysis may be considered lagerstätten for ichthyosaurs, but not for other clades. Maisch [22] identified the Triassic Pardonet Formation and Hosselkus Limestone, and the Middle Jurassic Los Molles Formation, amongst others, as ichthyosaur lagerstätten, based on their relatively complete, well-preserved and well-sampled skeletons despite the lack of classical indicators such as soft tissue preservation. Other ichthyosaur-bearing horizons, particularly the Oxford Clay and Blue Lias formations, contain beds that show exceptional preservation, and so are referred to as lagerstätte-grade by some authors; however, other authors do not class them as such.

Ultimately the issue under consideration is how lagerstatten impact our picture of ichthyosaur disparity in terms of their contribution to character state codings. Thus, we define a conservative list of lagerstätten which may have induced spikes in the availability of highly codable specimens due to the balance of the historical (well sampled and documented) and taphonomic (exceptionally preserved) factors outlined above. Any significant differences in disparity metrics should therefore remain valid even with the elevation of existing deposits to lagerstätte grade, the future discovery of new ichthyosaur taxa from lagerstätten, or the discovery of new lagerstätten themselves. The chosen lagerstätten are the Early Triassic Nanlinghu and Jialingjiang formations of China, the Middle and Late Triassic Guanling and Xiaowa formations of China, the Middle Triassic Zhuganpo Member of the Falang Formation...
of China, the Middle Triassic Besano Formation of Switzerland and Italy, the Early Jurassic Posidonia Shale of Germany, the Early Jurassic Strawberry Bank Limestone, and Blue Lias and Charmouth Mudstone formations of the UK, the Middle to Late Jurassic Oxford Clay of the UK, the Late Jurassic Solnhofen Limestone of Germany, the Late Jurassic Kimmeridge Clay Formation of the UK, the Early Cretaceous Vaca Muerta Formation of Argentina, and the Early Cretaceous Toolebuc Formation of Australia. Out of the 112 taxa in this analysis, 37 are exclusively from lagerstätten (33%) and occur in bins 2–4, 6–8, 11–13, 15–16. As there was no way to assess whether lagerstätte or non-lagerstätte specimens of the same taxon provided particular character codings in the data set, we included co-occurring taxa in both partitions of the data set. This represents a compromise reflecting the certain (lagerstätte-exclusive taxa) and potential (co-occurring taxa) influences of lagerstatten on ichthyosaur disparity, rather than the worst case scenario where lagerstätten also dominate the proportion of known character states for co-occurring taxa. We also ran the following analyses where only lagerstatten-exclusive taxa were removed from disparity calculations, thus representing the best-case scenario.

(e) Lagerstätten effects

Statistically significant differences between time bins with and without lagerstätten taxa were assessed in two ways. Firstly, disparity metrics from bins which contain lagerstätten and those from the corresponding bins with lagerstätten taxa removed were extracted from the overall time series and tested for significant differences using paired sample two-tailed T tests and Wilcoxon Signed Ranks tests. Secondly, two-tailed Pearson and Spearman tests of correlation with Holm correction [39] for false discovery rates were used to test whether the removal of lagerstätten taxa affected the trend in each metric through time. Generalised differencing [40] using Graeme Lloyd’s gen.diff function [41], was performed on the raw disparity metrics to detrend and remove short-term autocorrelation from the data. Correlations were then tested for in the same manner. To relate missing data (see below) and sample size biases to possible lagerstätte effects, the same tests as above were conducted on diversity and mean character completeness time series, and an independent samples, two-tailed T-test was used to test for a significant difference in completeness between lagerstätte and non-lagerstätte taxa.

(f) Missing data
To assess how the proportion of missing data for a taxon in the original cladistic matrix might affect its position in morphospace, Pearson and Spearman tests of correlation were used to test for a relationship between the percentage of missing data from each taxon and their Pythagorean distance from the origin of the PCO plots in \( n \) dimensions, where \( n \) is the number of PCO axes. A linear model was then fitted to the data, showing that as the percentage of missing data increases, the Pythagorean distance to the origin decreases (see below). The procedure was repeated using just the first two axes to provide a means of critiquing the PCO morphospace plots. Disparity metrics for each level of missing data were also calculated, using the methodology described above (Fig. S2).

To ascertain whether the relationship is simply a quirk of the data set, the same procedure was conducted on two other cladistic data sets taken from the literature: 81 non-avian theropod dinosaurs coded for 413 characters taken from Nesbitt and Ezcurra [42], and 80 sauropterygians coded for 270 characters taken from Benson and Bowdler [43], as well as the ichthyosaur cladistic matrix used in this study. Both data sets were downloaded from Graeme Lloyd’s website (http://www.graemetlloyd.com/matr.html) on 22/07/17, converted to distance matrices and subjected to PCO without negative eigenvalue correction in R (v3.2.2) prior to the calculation of Pythagorean distances for each taxon and correlation with the percentages of missing data (Fig. S3, S4, S5, S6).

The relationship between missing data and disparity was investigated further using simulations. A single data set comprising 20 operational taxonomic units (OTU) coded for 300 unordered, binary, equally weighted characters with randomly assigned states generated in R (v3.2.2) [30]. Within the same cladistic matrix, the 20 OTUs were duplicated ten times and 10% through to 90% of character states were deleted for successive groups of 10 OTUs. The cladistic matrix was converted to a GED matrix and subjected to PCO without negative eigenvalue correction. The same relationship between missing data and Pythagorean distance was tested for using the method above.

Additional simulations were conducted to specifically focus on the relationship between missing data and distance from the centroid. To ensure the accuracy of these simulations, 20 empirical cladistic matrices were downloaded from Graeme Lloyd’s website on 01/05/18 and the proportion of different character states assessed for each; the proportion of higher character states was seen to decline in an approximate geometric relationship (Fig. S7). Normal, Poisson and negative binomial distributions were then fitted to the distribution of missing data in each matrix (Fig. S8) and the model fit assessed using the Akaike Information Criterion. Negative binomial distributions performed the best, but only marginally compared
to normal distributions (Table S1). Simulated cladistic matrices covering a range of
dimensions were generated in R (v3.2.2). Each matrix was then assigned a different normal
distribution of missing data (this strategy was chosen for simplicity as a negative binomial
distribution tends towards normality for large sample sizes), and equal or geometric
distributions of unordered, randomly assigned character states for binary, ternary, quaternary,
quinary, and senary coding strategies. Each matrix was transformed, ordinated and analysed
using the methodology above.

3. Results

(a) Lagerstätten biases

Time bins with lagerstätten taxa show significantly lower values for each disparity metric
(Table 1; Fig. 1), when lagerstätte taxa are excluded. Morphospace occupation is similarly
reduced when lagerstätte taxa are excluded (Fig. 2). After generalised differencing, centroid
distances from GCD and MORD fail to correlate between lagerstätte and non-lagerstätte time
series using Pearson correlation; centroid distance from all three distances, and the sums of
variances from GED also fail using Spearman correlation (Table 2, 3). This shows that the
preservational biases of lagerstätten impact the long-term trend in disparity observed over
hundreds of millions of years. It should be noted, however, that while there is a statistical
difference between the time series, they broadly recover the same peaks and troughs in
disparity, aside from the late Early to Late Jurassic where lagerstätten dominate their fossil
record. An independent-samples, two-tailed T test showed a statistically significant
difference (T = -3.47, df = 54.69, p < 0.005, n = 25, 40) in the percentage completeness of
cladistic characters between the lagerstätte (57.4% complete) and non-lagerstätte taxa
(37.2.0% complete), and taxonomic diversity-per-bin shows significant differences within
lagerstätte bins when lagerstätte taxa are excluded (Table 1). Together, these results suggest
that lagerstätten affect disparity in two ways: through the preservation of more taxa; and
through the preservation of more complete specimens. Biases in individual disparity metrics
and morphospace plots, however, are partially dependent on the choice of distance matrix.

Reanalysis, excluding only lagerstätte-exclusive taxa from disparity calculations, had no
effect on the statistical significance or otherwise of our results, aside from significant
Spearman correlations for the GED-derived centroid distance after generalised differencing.
This demonstrates that the impacts of lagerstätten remain prevalent even for the best-case
scenario. The greater completeness of lagerstätte specimens, however, suggests that the true
impacts of lagerstätten on ichthyosaur disparity may lie somewhere between our comprise and the worst-case scenario.

(b) Missing data and centroid slippage
A highly significant, negative relationship was found between the percentage of missing characters for a taxon in a cladistic matrix and its Pythagorean distance to the origin of \(n\)-dimensional principal coordinate (PCO) space derived from GED, where \(n\) is the number of PCO axes examined (Table 4; Fig. 3). This trend holds true across simulated cladistic matrices spanning a wide variety of dimensions, coding strategies and levels of missing data (Fig. 4). For convenience, we term this effect ‘centroid slippage’ (PCO performed on a GED matrix is mathematically equivalent to principal components analysis, where the centroid of the dataset lies at zero on all axes). The effect of centroid slippage is nullified when GCD and MORD are used. Together with the reduced impact of lagerstätten effects when alternative distances are used, this highlights a key methodological issue for disparity analyses. In other words, taxa move from their true positions in morphospace towards the centroid.

Generalised Euclidean Distance produces the most severe biases in disparity metrics and is especially subject to centroid slippage, as gaps in the distance matrix are infilled with a weighted mean fractional univariate distance based on the calculable distances [26]. The limits of morphospace size are affected by the proportion of missing data for taxa comprising the hull of the morphospace (Fig. 5), and trends in morphospace occupation by the dispersal of taxa within that morphospace. Lagerstätten effects are part of this, with the higher character completeness of their taxa often determining the bounds of morphospace in the bins in which they occur, while their position relative to more incomplete taxa is warped as the latter slip towards the centroid (Fig. 2).

(c) Ichthyosaur disparity through time
Despite their different responses to geological sampling biases, bivariate morphospace plots and disparity metric time series from all three distance types converge upon trends in disparity concordant with the findings of previous authors. Morphospace occupation was greatest in the first 8.7 myr of the Triassic (Fig. 1, 2), during which time early ichthyosaurs gained great ecomorphological diversity by radiating into ecological niches left vacant by the Permian–Triassic mass extinction, plus possibly new, previously unoccupied ecospace [23]. Disparity rapidly decreased during the loss of circa-littoral and shallow water forms during the Late Triassic and remained reduced in accordance with the conservative morphology of
Early Jurassic ichthyosaurs, despite their high diversity [24]. Overall, there was a shift in morphospace occupation throughout the Mesozoic (Fig. 2), corresponding to the shift from early elongate taxa to later thunniform taxa [41].

4. Discussion

Compared to the morphometric approach to disparity, the cladistic approach is an excellent way of examining disparity in an incomplete fossil record, yet it is sensitive to the choice of distance matrix and disparity metric. Gower Coefficient Distance is increasingly recommended as it can account for the effect of missing data [44]; Lloyd [26] introduced MORD which may perform even more strongly. While these latter distances will likely prove preferable in future studies, two problems associated with GED may yet cause problems with our current view of disparity in the fossil record, as this distance type is still commonly employed in studies using the cladistic approach.

Firstly, centroid slippage makes a taxon appear morphologically average, analogous to the way in which missing data in phylogenetic studies causes stemward slippage [45], where the incomplete taxon appears more plesiomorphic than is actually the case. By extension, just as stemward slippage may confound the determination of phylogenetic relationships among a set of taxa, centroid slippage will confound the morphological relationships among taxa, affecting their relative positions in morphospace. Lagerstätte biases relate to this first issue due to the greater completeness of their taxa.

The second issue is that, as PCO is a method used to visualise variance in a data set using similarities and differences between groups of objects, centroid slippage may be problematic for studies that attempt to use separation in morphospace as a proxy for ecomorphological separation between clades. Highly incomplete taxa may be removed from the clusters to which they are truly affiliated and pushed into another, or groups that are truly morphologically separate may end up overlapping due to centroid slippage. More broadly, the changing positions of taxa in all dimensions will alter the size and position of occupied morphospace, strongly altering the disparity signal.

Focusing on the complete signal contained within the fossil record, the high number of Triassic and Early Jurassic lagerstätten suggests that this part of ichthyosaur evolutionary history can be interpreted with some confidence. In particular, their Jurassic fossil record is composed predominantly of lagerstätten, giving high fidelity, yet also indicating the potential severity of lagerstätte biases were these deposits absent. Ichthyosaur lagerstätten are fewer in number from the Cretaceous onwards, and correspondingly interpretation of their
evolutionary history is less confident [24]. This is particularly problematic in the few million years prior to their extinction during the mid-Cretaceous, where their fossil record is highly fragmentary [25].

Butler et al. [4] demonstrated that the peaks and troughs in the record of pterosaur disparity are largely artificial due to lagerstätte sampling. While the ichthyosaur fossil record is less dependent on lagerstätten deposits, there is nevertheless a biased signal between bins with and without lagerstätten. Lagerstätten vary in their depositional environments and are geographically and temporally localised; thus, they only give a truer record of particular ecosystems, but not necessarily a more complete global signal. Despite this, the disparity and diversity signals they preserve are more complete and so are inherently more faithful than bins without lagerstätte deposits; these latter must therefore be interpreted with caution.

Lagerstätten improve the fidelity of the disparity signal for a time bin by preserving a greater range of species, and by preserving more complete specimens. Thus, while the bias of missing species and missing data may affect bins containing lagerstätten less, the relative increase in disparity in lagerstätte-containing bins will affect attempts to draw broad conclusions about patterns of disparity through time; for example if linear models are fitted to time series to try to discern an overall trend. An important point to consider is that while lagerstätten provide more faithful preservation, they have differing impacts on different groups. To an extent, this is dependent on the depositional environment of the ecosystem and the ecological biases of the organisms within that ecosystem; for example, infaunal organisms are typically more likely to be preserved than nektonic or pelagic species, and shelf environments more likely to receive deposition than forests. As such, the distribution and type of lagerstätten in time and space may have heterogeneous effects on the disparity of different clades. Here we document a clear impact of lagerstätten on ichthyosaur disparity, whereas a study of Palaeozoic echinoderms [46] found that lagerstätten had virtually no effect on their overall disparity (strong effects were noted because of taphonomic degradation, however, highlighting the problem of missing data). In turn, this poses difficulties for associating changes in disparity within distinct groups to external causes or to evolutionary interactions between those groups.

Together, these two points highlight the need to find ways to correct the biases imposed by lagerstätten. The removal of highly incomplete taxa prior to analysis does not solve the problem of missing data, merely sidestepping it at the cost of a straight reduction in disparity. Phylogenetic correction is an alternative, but its utility is dependent on how well resolved the chosen phylogenetic hypothesis is. As this is itself affected by the amount of missing data in
a cladistic matrix, the instability of phylogenetic and disparity analyses may trend together; thus, the utility of phylogenetic correction will decline even as the disparity analysis calls more strongly for correction. This is also the case for the use of gap weighting to infill missing data. Smith et al. [14] note that these methods all affect the disparity signal differently. They present a novel method of correction, which identifies linkage in the presence or absence of characters in a data set, and then selectively removes additional data to equalise the distribution of missing characters between the taxa, minimising the degree of disparity reduction at varying levels of missing data. In turn, while the true disparity values cannot be measured, relative changes in the size and occupation of morphospace through time can be identified with greater confidence. Application of this method to future disparity studies may help reduce the impact of the greater completeness of lagerstätte specimens.

The bias from preservation of a greater number of taxa is harder to mitigate, but may be tackled through a better understanding of the factors that control the spatiotemporal distribution of lagerstätten. Lagerstätten of similar ages, representing similar facies environments, may occur in clusters resulting from deposition in a broad geographical region where conditions were generally conducive to exceptional preservation [47]. Trends in the size and distribution of these clusters through time result from fluctuations in the environmental constraints required for their formation at local, regional and global scales [47]. Preservational biases resulting in lagerstätten dominate in the marine environment, while terrestrial lagerstätten are more strongly controlled by sampling biases [47]. Thus, it may be possible to demonstrate more general links between these biases and the disparity signal from the fossil record. The spatiotemporal distribution of lagerstätten may also provide useful insight into the problem of missing data. Tutin and Butler [48] found a significant relationship between skeletal completeness and character completeness metrics for plesiosaurs, and between plesiosaur and ichthyosaur completeness, suggesting common controls on skeletal preservation in both groups; this could feasibly include a lagerstätten bias. Lagerstätte biases in completeness may not be present for all groups; however, Verriere et al. [49] found no evidence for lagerstätte effects on parareptile diversity and in turn no correlation between their diversity and skeletal or character completeness metrics.

**Competing interests.**

We declare we have no competing interests.

**Authors’ contributions.**
J.F.S, T.L.S and B.C.M designed the analysis. J.F.S carried out the analysis. B.C.M provided the cladistic data set. J.F.S and M.J.B wrote the paper, and all authors commented on the paper.

**Funding.**

B.C.M. is funded by Leverhulme Research Project Grant RPG-2015-126 to M.J.B.

**Data accessibility.**

All additional data for this article are provided in the electronic supplementary material: Table S1, Figs S1–S7, along with all R scripts and data used in the analyses.

**Acknowledgements.**

We thank Dr Steve Donovan, and two other anonymous referees, whose comments helped us to greatly improve the quality of this manuscript.

This manuscript has not been published in whole or in part elsewhere and is not currently being considered for publication in another journal.

**References**


Table 1. Paired samples two-tailed T tests and Wilcoxon Signed Ranks (W) tests between disparity metrics (Sum V = sum of variances, Sum R = sum of ranges, Cent = mean distance from the centroid), and character completeness (Compl) and taxonomic diversity (Div), from time bins with lagerstätte taxa, and with those taxa removed (df = 9, p-values are italicised, T and W statistics are plain text).

<table>
<thead>
<tr>
<th></th>
<th>Sum V</th>
<th>Sum R</th>
<th>Cent</th>
<th>Compl</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>GED</td>
<td>4.67, &lt;0.005</td>
<td>11.01, &lt;0.0001</td>
<td>5.80, &lt;0.0005</td>
<td>-1.25, 0.24</td>
</tr>
<tr>
<td></td>
<td>GCD</td>
<td>5.08, &lt;0.001</td>
<td>10.42, &lt;0.0001</td>
<td>6.94, &lt;0.0001</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MORD</td>
<td>4.94, &lt;0.001</td>
<td>10.15, &lt;0.0001</td>
<td>6.56, &lt;0.0005</td>
<td>-</td>
</tr>
<tr>
<td>W</td>
<td>GED</td>
<td>55, &lt;0.005</td>
<td>55, &lt;0.005</td>
<td>55, &lt;0.005</td>
<td>10, 0.16</td>
</tr>
<tr>
<td></td>
<td>GCD</td>
<td>55, &lt;0.005</td>
<td>55, &lt;0.005</td>
<td>55, &lt;0.005</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MORD</td>
<td>55, &lt;0.005</td>
<td>55, &lt;0.005</td>
<td>55, &lt;0.005</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. Two-tailed tests of Pearson correlation between raw and generalised differenced (GD) disparity metric time series, and character completeness and diversity, with lagerstätte taxa present and removed (p-values are italicised). See Table 1 for column title abbreviations.

<table>
<thead>
<tr>
<th></th>
<th>Sum V</th>
<th>Sum R</th>
<th>Cent</th>
<th>Compl</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>GED</td>
<td>0.79, &lt;0.005</td>
<td>0.80, &lt;0.005</td>
<td>0.78, &lt;0.005</td>
<td>0.94, &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>GCD</td>
<td>0.92, &lt;0.0001</td>
<td>0.82, &lt;0.001</td>
<td>0.767, &lt;0.01</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MORD</td>
<td>0.90, &lt;0.0001</td>
<td>0.81, &lt;0.001</td>
<td>0.65 &lt;0.05</td>
<td>-</td>
</tr>
<tr>
<td>GD</td>
<td>GED</td>
<td>0.67, &lt;0.05</td>
<td>0.74, &lt;0.05</td>
<td>0.67, &lt;0.05</td>
<td>0.93, &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>GCD</td>
<td>0.92, &lt;0.0001</td>
<td>0.76, &lt;0.01</td>
<td>0.45, 0.12</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MORD</td>
<td>0.89, &lt;0.0005</td>
<td>0.76, &lt;0.005</td>
<td>0.44, 0.13</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3. Two-tailed tests of Spearman correlation between raw and generalised differenced (GD) disparity metric time series, and character completeness and diversity, with lagerstätte taxa present and removed (p-values are italicised). See Table 1 for column title abbreviations.

<table>
<thead>
<tr>
<th></th>
<th>Sum V</th>
<th>Sum R</th>
<th>Cent</th>
<th>Compl</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>GED</td>
<td>0.85, &lt;0.005</td>
<td>0.81, &lt;0.001</td>
<td>0.781 &lt;0.005</td>
<td>0.94, &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>GCD</td>
<td>0.80, &lt;0.005</td>
<td>0.75, &lt;0.005</td>
<td>0.36, 0.20</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MORD</td>
<td>0.87, &lt;0.0005</td>
<td>0.77, &lt;0.005</td>
<td>0.32, 0.26</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sum V</th>
<th>Sum R</th>
<th>Cent</th>
<th>Compl</th>
<th>Div</th>
</tr>
</thead>
<tbody>
<tr>
<td>GD</td>
<td>GED</td>
<td>0.42, 0.14</td>
<td>0.66, &lt;0.05</td>
<td>0.54, 0.12</td>
<td>0.98, &lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>GCD</td>
<td>0.71, &lt;0.05</td>
<td>0.65, &lt;0.05</td>
<td>0.52, 0.07</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MORD</td>
<td>0.75, &lt;0.01</td>
<td>0.64, &lt;0.05</td>
<td>0.42, 0.16</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4. Two-tailed tests of Pearson and Spearman correlation between a taxon’s percentage missing data and its Pythagorean distance from the origin of PCO space in n dimensions where n is the total number of PCO axes for each data set (p-values are italicised).

<table>
<thead>
<tr>
<th></th>
<th>Ichthyosauria</th>
<th>Sauropterygia</th>
<th>Dinosauria</th>
<th>Artificial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>-0.87, &lt;0.0001</td>
<td>-0.91, &lt;0.0001</td>
<td>-0.84, &lt;0.0001</td>
<td>-0.98, &lt;0.0001</td>
</tr>
<tr>
<td>Spearman</td>
<td>-0.89, &lt;0.0001</td>
<td>-0.90, &lt;0.0001</td>
<td>-0.88, &lt;0.0001</td>
<td>-0.97, &lt;0.0001</td>
</tr>
</tbody>
</table>
Figure 1. Comparisons of ichthyosaur diversity disparity through time, with and without lagerstätte taxa. Time series of disparity metrics, along with taxonomic diversity, for each time bin from 250 – 90 Ma. A. Taxonomic diversity; B. Sum of variances; C. Sum of ranges; D. Mean distance from the centroid. Error bars represent the bootstrapped 95% confidence interval.

Figure 2. Changing morphospace occupation of ichthyosaurs through geological time. 10 myr morphospace plots of the first two PCO axes of the ichthyosaur data set from 250 – 90 Ma.

Figure 3. The effect of missing data on distance from the morphospace origin. Linear regressions between a taxon’s percentage missing data and its Pythagorean distance from the origin of PCO space on the first two PCO axes.

Figure 4. Simulation of the proportion of missing data versus distance from the morphospace origin. Heatmaps of the relationship between a taxon’s percentage missing data and its Pythagorean distance from the origin of PCO space on the first two PCO axes, derived from 1,100 matrices and 100,000 taxa. The shift from blue to red indicates an increasing density of points at a given coordinate.

Figure 5. Impact of missing data on morphospace. Morphospace plots of the first two PCO axes of a simulated data set, artificially degraded for varying levels of missing data, compared to the true morphospace for the complete simulation data set.