

Tools for identifying unexpectedly low microfossil count sums

Richard J. Telford^{*,a}

^a*Department of Biological Sciences, University of Bergen and Bjerknes Centre for Climate Research, Post Box 7803, N-5020 Bergen, Norway*

Abstract

Microfossil counts are a key data type in palaeoecology. Recent work has raised the possibility that some authors might misreport an important quality control parameter, the counts sums, occasionally dramatically so. This paper introduces methods that can flag assemblages with potentially misreported count sums and finds that some assemblage datasets fail these tests.

Keywords: Microfossil counts; quantitative methods

1. Introduction

Six percent of papers published in *Molecular and Cellular Biology* show evidence of inappropriate image duplication (Bik et al. 2018) either due to error or, more rarely, misconduct. It would be reckless to assume that the palaeoecological literature does not have equivalent problems: the low rate of retractions in the ecological and geological literature (Grieneisen and Zhang 2012) may partially reflect our limited ability to detect errors and misconduct rather than their prevalence.

Several numerical tools have been developed to identify questionable data. Deviations from the distribution of digits expected from the Newcomb-Benford law has been used to detect issues with scientific and financial data (Barabesi et al. 2018). Carlisle (2017) identified papers where the baseline differences in means for different treatments were surprisingly high or low given the variance of the data. Brown et al. (2017) developed the granularity-related inconsistency of means (GRIM) test which assesses whether means of integer data are consistent with the reported sample size. However none of these tests are directly applicable to microfossil assemblage data, one of the most common types of palaeoecological data.

A common assertion in papers reporting microfossil assemblage data is that a minimum of N microfossils were counted, where N is often fifty for chironomids and several hundred for pollen and diatoms. This is important because larger count sums are associated with smaller uncertainties, both in relative abundance of taxa and derived statistics such as transfer function reconstructions (Heiri and Lotter 2001; Payne and Mitchell 2009). However, mistakes happen and metadata such as count sums can be forgotten once

^{*}Corresponding Author
Email address: richard.telford@iuh.no (Richard J. Telford)

percentages are calculated. In addition, given the time-consuming nature of microfossil counting, especially when preservation is poor or concentrations are low, there may be an incentive to misreport the minimum count sum. The risk that count sums will be misreported is not just theoretical: Larocque-Tobler et al (2015) reported that their chironomid count sums were at least fifty head capsules; a subsequent corrigendum (Larocque-Tobler et al. 2016) acknowledged that count sums were actually as low as nineteen. Telford (2019a) reports some other cases where count sums may be much lower than reported.

If the genuine assemblage counts are archived for all taxa, it is trivial to identify undercounts. Unfortunately, count data might be falsified so that it appears to meet the reported count sum. A more common problem is that, regrettably, many palaeoecologists archive percent data without an indication of the count sum. This paper develops simple tests that can flag if the count data might have been misrepresented, or if the count sum of percent data is perhaps lower than reported.

The key insight that allows inference about the count sum is that assemblage data are expected to follow the typical features of a rank abundance curve. In particular, because there are many rare taxa in most communities (Darwin 1859), most community or assemblage samples will include taxa represented by a single individual, hereafter singletons, unless the sampling effort is high relative to the taxonomic richness (Coddington et al. 2009). In the occasional assemblages without singletons, the greatest common divisor should usually be one, i.e. few assemblages should have all counts divisible by an integer larger than one, especially if species richness is high.

In general, it is not possible to determine the sum from which percent were calculated, but the properties of community and assemblage counts make it possible to estimate it. Given that we expect the rarest taxon to be a singleton, the count sum N can be estimated as $1/p_{min} \times 100$ where p_{min} is the percent abundance of the rarest taxon. This method will fail for assemblages without singletons. We also expect the percent p to be calculated from integer counts, therefore there should be a count sum N such that $p_i/100 \times N$ is, within rounding error, an integer for all i taxa. Possible values for N can be found by a direct search algorithm over the range of plausible values of N . An infinite number of possible count sums that are consistent with the percent exist, but the lowest will give the correct value of N except in cases where the greatest common divisor is greater than one. These tests are closely related to GRIM (Brown and Heathers 2017), as all rely on the granularity of percentages calculated from integer data.

This paper aims to test the methods presented above and present some cases with unexpected results. Some complications and caveats are discussed.

2. Methods

Publicly available datasets were downloaded from the Palaeodata Center, Neotoma, Pangaea and other sources. A range of ecological and palaeoecological data were sought to allow for differences in typical count sums and species richness. Fossil pollen assemblage in the Neotoma database (Williams et al.

Table 1: Percent of bird counts with singletons and of greatest common divisor (GCD) of one, at different taxonomic levels.

Taxonomic level	Percent with singletons	Percent GCD = 1	Median richness
species	99.95	100.00	54
genus	99.85	100.00	46
family	95.69	100.00	25
order	82.62	99.42	9

2018) (downloaded 2019-05-03) with small count sums (< 50), that appeared to be percentages, or were documented as back-transformed from digitised data were excluded. Datasets with percent data where the minimum count sum was not reported in the associated paper were excluded. Datasets discussed by Telford (2019a) as possibly having under-reported count sums were also excluded to avoid double reporting. Datasets with possible misreporting are anonymised, but no attempt is made to diagnose whether errors or misconduct are responsible. This paper does not attempt to be an exhaustive survey of all the data available.

All analyses were done in R version 3.4.4 (R Core Team 2018) and used the packages extraDistr version 1.8.10 (Wolodzko 2018), numbers version 0.7.1 (Borchers 2018), and countSum 0.0.3 (Telford 2019b). Code to replicate all the analyses shown above is archived at <https://github.com/richardjtelford/count.check.ms>.

3. Results

3.1. Prevalence of singletons

The vast majority of the over 65,000 bird counts from the North American breeding bird survey (Pardieck et al. 2018) have singletons at the species level (Table 1). To explore the effect of taxonomic richness on the prevalence of singletons, I aggregate the birds counts to progressively lower taxonomic resolutions. As richness declines, the proportion of counts having singletons declines (Table 1), reaching a moderate proportion at the order level, where most of the counts lacking singletons have fewer than five taxa. The vast majority of counts have a greatest common divisor of one, except when taxonomic richness is low.

To test the sensitivity of the prevalence of counts without singletons to the count sum, I resample the counts with 400 or more observations to smaller count sums using a multivariate hypergeometric distribution. At the order level, the proportion of counts with singletons increases steeply until it reaches a maximum at about 200, thereafter it declines slowly (Fig. 1). The prevalence of counts with singletons is low for small counts because it is possible that only the common taxa are counted; with larger counts, the chance of counting a single individual of a rare taxon increases. With even larger counts the prevalence of counts with singletons decreases due to saturation as singletons become doubletons and few new taxa are found.

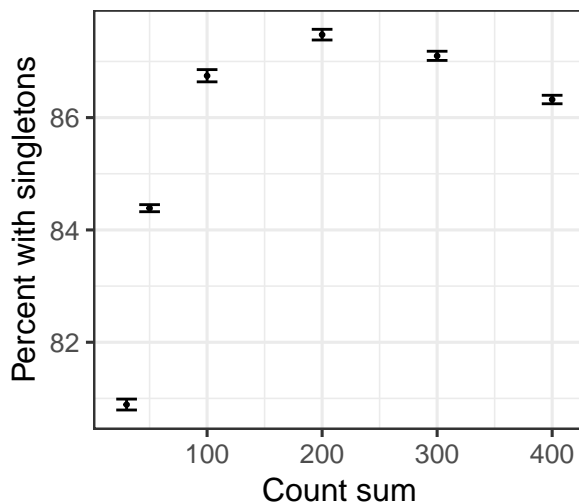


Figure 1: Effect of count sum on the proportion of bird counts without singleton at the order level. Results are the mean of ten trials, error bars are two standard errors.

111 The 862 diatom counts from Owen’s Lake span a wide range of count sums (1
112 – 701) because, if concentrations were low, the requirement that at least 300 valves
113 were counted was replaced by a minimum area of slide to count (Bradbury 1997).
114 Discounting the 20 mono-specific assemblages (all with very low count sums),
115 97.39% of assemblages have singletons with a median number of singletons per
116 assemblage of five. Assemblages with count sums above 50 have a higher chance
117 of having singletons (98.18% vs 89.04%). 99.52% of assemblages have a greatest
118 common divisor of one, the remainder have a greatest common divisor of two.
119 Assemblages with a greatest common divisor above one all have low taxonomic
120 richness.

121 The 1775 assemblages in the North American testate amoeba training set
122 with available count data (Amesbury et al. 2018) have count sums between 52
123 and 456 tests (median = 151), and taxonomic richness ranges between 2 and
124 31 (median = 15) taxa. 94.87% of the assemblages have at least one singleton
125 (median 4) and 99.72% of assemblages have a greatest common divisor of one.
126 Of the five assemblages with a greatest common divisor above one, three have
127 fewer than five taxa; four have a greatest common divisor of two, and the other
128 one has a greatest common divisor of three.

129 The pollen data from the Neotoma database included over 182,000 assem-
130 blages in 4130 datasets. Nearly all assemblages have singletons (97.04%) and
131 a greatest common divisor of one (99.64%). The assemblages with a greatest
132 common divisor above one are not randomly distributed among the datasets:
133 only 3.37% of the datasets have any such assemblages. Some datasets with a
134 high proportion of assemblages with a greatest common divisor above one are
135 older datasets which may, despite the available metadata, have been digitised
136 with the loss of rare taxa and precision. For example, one dataset includes 84

137 assemblages with a median richness of 16 taxa. Count sums vary between 69
138 and 3201. The greatest common divisor is three for all assemblages, that is
139 all 1300 counts in the dataset are divisible by three. Excluding such datasets
140 would further increase the proportion of assemblages with singletons. Some other
141 Neotoma datasets are discussed in a subsequent section.

142 3.2. Estimating the percent sum

143 The Owen’s Lake dataset (Bradbury 1997) includes percentages for each
144 taxa, allowing the count sums estimated from the percent data to be verified.
145 Excluding mono-specific assemblages, for which no meaningful estimate of the
146 count sum is possible, the minimum percent method and the direct search
147 method correctly estimate the count sum for, respectively, 97.39% and 99.52%
148 of the assemblages. The few assemblages that fail the direct search method are
149 species poor and have low count sums.

150 The 22 chironomid head capsule assemblage dataset from Last Chance Lake
151 (Axford et al. 2017) also includes both the count sums and the percentage of each
152 taxa. The percentages are given to two decimal places: to account for rounding
153 errors, the countSum package uses the smallest and largest values consistent
154 with the reported percentage. With the minimum percent method the estimated
155 count sums are, within error, either identical to the reported count sum or twice
156 as much (Fig. 2). With the direct search method, the estimated count sums are
157 all exactly twice the reported count sum. The factor of two difference is because
158 all the counts include half head capsules (but only sometimes is the rarest taxon
159 represented by a half head capsule). Reporting half microfossils is common
160 for several microfossil groups, including chironomids and pollen (especially for
161 bisaccate conifers); occasionally other fractions are reported. Half counts make
162 the estimated count too high by a factor of two, so do not risk incorrectly flagging
163 count sums as being too small. The direct search method is more precise than
164 the minimum percent method as the rounding error is relatively smaller on the
165 larger percent values the method uses.

166 3.3. Unexpected count data

167 Some of the pollen datasets in the Neotoma database have an inexplicably
168 high proportion of assemblages with a greatest common divisor greater than
169 one. Here I focus on four datasets produced by the same research group. All the
170 datasets, which are relatively recent, include information on the spike used to
171 calculate pollen concentration, so cannot have been digitised. The four datasets
172 have some assemblages with a greatest common divisor of two interspersed,
173 sometimes regularly, amongst assemblages with the expected greatest common
174 divisor of one. The assemblages with a greatest common divisor of one typically
175 have several singletons (Table 2).

176 Assemblages with a greatest common divisor of one typically have a higher
177 taxonomic richness than assemblages with a greatest common divisor of two
178 (Fig. 3). In the datasets with a wide range of count sums, the richness of
179 the assemblages with a greatest common divisor of two is comparable with the

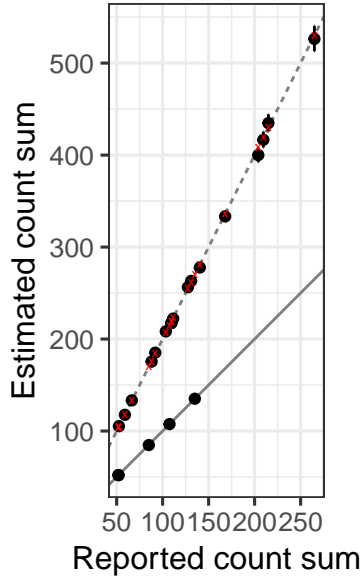


Figure 2: Estimated and reported chironomid count sums by the minimum percent method (solid symbols with error bars) and the direct search method (red crosses) from Last Chance Lake (Axford et al. 2017). Lines show the 1:1 (solid) and 2:1 (dashed) relationships.

Table 2: The number of assemblages, median number of taxa, percent of assemblages with a greatest common divisor (GCD) of one, and the median number of singletons in assemblages with a greatest common divisor of one, in four pollen datasets

Dataset	No. assemblages	Median no. taxa	% GCD = 1	Median no. singletons (GCD = 1)
1	53	36	49	16
2	83	36	61	17
3	27	19	70	7
4	41	19	93	10

richness of the assemblages with a greatest common divisor of one that have a count sum that is half as large.

One assemblage with a greatest common divisor of two has 73 taxa. Assuming the distribution of counts is (at least locally) uniform, the probability of n counts being divisible by k is $1/k^n$: for a single assemblage with this many taxa, the probability is 1 in 10^{21} . The probability that these counts reflect the true nature of the pollen assemblages is exceedingly low; it is far more likely that the data have been mishandled at some stage in some way.

3.4. Unexpected percent data

Dataset diatom1 includes 35 diatom assemblages with a median richness of 19 taxa. The associated paper reports that at least 400 valves were counted from each assemblage, which means that singletons should have a relative abundance

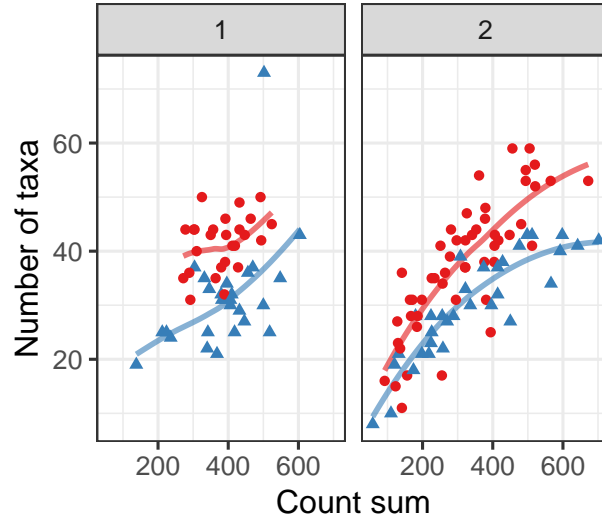


Figure 3: Number of taxa in against count sum for assemblages in two pollen datasets with a greatest common divisor of one (red circles) or two (blue triangles). Lines are loess with a span of 1. All four pollen datasets show the same pattern, but it is clearest in datasets 1 and 2 which have the highest proportion of assemblages with a greatest common divisor of two.

of 0.25% or less. However, the rarest taxon in two assemblages have a relative abundance of 3.23%, and the relative abundance of all other taxa in these assemblages are, within rounding error, integer multiples of this. If the count sums actually are at least 400, this would imply that each species count in these assemblages is an integer multiple of at least 13. This seems unlikely. Alternatively, the actual count sum for these assemblages could be as low as 31 valves. In total, six assemblages appear to have a count sum below 400 valves.

The archived data include 56 taxa that have a maximum abundance of at least 1.25%; this is about half of the 109 taxa reported in the paper. This pruning of rare taxa will increase the risk of counts without singletons, which would cause the minimum percent method to fail, but will have little impact on the direct search method.

Dataset diatom2 has 71 assemblages with a median richness of 57 taxa. The associated paper reports that the diatom counts included 400–500 valves per assemblage except for four diatom-poor assemblages with at least 100 valves. The direct search method estimates that twelve assemblages have a count sum of less than 400, and three have less than 100. The direct search and minimum percent methods agree, within rounding error, on the estimated count sums, implying that either all assemblages have singletons and count sums are low in some assemblages, or that the greatest common divisor of the raw counts is greater than one. This means that for the assemblage where the minimum percent is 7.14, which has seven taxa, all counts would need to be multiples of 8 for the count sum to be at least 100.

215 Dataset chironomid1 has 55 assemblages. Although the associated paper
216 reports that assemblages with count sums below 50 were discarded, the direct
217 search method estimates that three assemblages have count sums between 38
218 and 40. These assemblages are diverse (richness between 15 and 17 taxa), so it
219 is unlikely that the true counts are double that estimated here and the greatest
220 common divisor is two.

221 Palaeoceanographic dataset marine1 has 160 assemblages with a median
222 taxonomic richness of 34.5 taxa. The associated paper reports that assemblages
223 with count sums below 100 were omitted. The direct search method estimates
224 that five assemblages have count sums below 100 (this excludes one mono-specific
225 assemblage), with count sums as low as 34. With this dataset, the direct search
226 and the minimum percent method give identical results for some assemblages but
227 highly divergent estimates for others, with the direct search method sometimes
228 giving estimates of several thousand, in one case over two orders of magnitude
229 higher than the minimum percent method. Some of the divergence appears to
230 be because some, mainly low diversity, assemblages lack singletons. The most
231 extreme divergences appear to be data entry errors with incorrect rounding,
232 perhaps during taxonomic revisions of the percent data rather than the raw
233 counts.

234 4. Discussion

235 Analysis of the breeding bird, Owen's Lake diatom, testate amoeba and
236 pollen datasets has shown that, as expected, the vast majority of community
237 and assemblage counts have singletons, and an even larger proportion have a
238 greatest common divisor of one. Therefore, these characteristics are potentially
239 useful for identifying data where the count sum is misreported.

240 My search through archived microfossil data found several datasets where
241 count sums are, or appear to be, smaller than that reported. Some of the count
242 sums appear to be an order of magnitude below what was reported. The non-
243 pollen datasets examined here are a convenience sample of available data that
244 met the inclusion criteria. As such, this analysis cannot be used to accurately
245 determine the prevalence of datasets with possible undercounts, but it appears
246 that a non-negligible fraction of the literature is affected. It is possible in
247 some cases that the assemblages with low counts were omitted or merged prior
248 to analysis, but the number of assemblages reported was not updated. Care
249 was taken to identify any such cases by, for example, examining stratigraphic
250 diagrams to see if the low count assemblages were included. In some cases, the
251 assemblages with low apparent counts can be seen in the published stratigraphic
252 diagrams.

253 It is also possible that the apparently low count sums are a false positive.
254 However, as taxonomically diverse assemblages with a greatest common divisor
255 greater than one are rare and all the datasets with apparently low count sums
256 contained several such assemblages, this is unlikely.

257 These tests require the assemblage data which are often not archived. How-
258 ever, if the count sums are very low it is possible to recognise this from a

259 stratigraphic diagram. For example, a chironomid stratigraphy had an assem-
260 blage where all five taxa had a relative abundance of about 20%, suggesting a
261 count sum of five, far below the reported fifty chironomids. Another assemblage
262 had six taxa with a relative abundance of just over 10% and a seventh with
263 twice as much, suggesting a count sum of eight chironomids. On request, the
264 authors sent the data. None of the 15 assemblages in the dataset had more than
265 fifty chironomids: the count sums varied between 5 and 40.

266 4.1. Challenges

267 Microfossil percent data are often given to two decimal places. This is
268 sufficient precision for the direct search method to have utility with counts sums
269 of several thousand. Some data are only given to the nearest percent, which
270 means that neither method has utility with count sums above 100.

271 The direct search method will fail if the dataset includes taxa calculated
272 with different count sums, for example pollen sums of trees, shrubs and upland
273 herbs and a pollen and spore sum that also includes pteridiophytes. It should
274 be possible to identify such cases from the meta-data and knowledge of usual
275 practice with different proxies. It will also fail, as shown by dataset marine1, if
276 percentages are incorrectly calculated or rounded.

277 Some taxonomic groups have microfossils that often come in groups of
278 attached individuals. For example, each diatom cell has two valves which are
279 the counting unit. These valves usually separate during processing, but paired
280 valves are often found, and some taxa such as *Aulacoseira* produce long chains
281 of strongly bound valves. This might make singletons slightly less likely. The
282 results from Owen’s Lake suggest that this is not an important problem, as most
283 assemblages have many singletons.

284 4.2. Consequences

285 Small undercounts in a few assemblages will have minimal impact on the
286 precision of any palaeoenvironmental reconstruction or other statistics derived
287 from the assemblage. Substantial undercounts will potentially seriously effect
288 the precision of the results. Such undercounts might constitute a data handling
289 error, for example, if samples with low counts were supposed to be merged but
290 that step was forgotten. Substantial and pervasive under-counting could be
291 construed as scientific misconduct due to either negligence or falsification, and
292 action to correct the literature is probably required.

293 Some papers do not report count sums. When this important quality metric
294 is omitted, the reader should be able to assume that the standard minimum
295 count sum for the taxonomic group has been used (i.e. 50 for chironomids, several
296 hundred for pollen and diatoms). If the actual count sums are materially below
297 this, then this potentially constitutes falsification by omission (Fanelli 2013).

298 5. Conclusions

299 A non-negligible fraction of the archived assemblage data includes counts
300 that appear to have a count sum below that reported in the paper. Some of the

apparent counts are small enough that the uncertainty of the count and derived statistics will be substantially larger than expected. These results highlight the importance of archiving both the raw data and the code required to process the data.

Acknowledgements

Some data were obtained from the Neotoma Paleoecology Database (<http://www.neotomadb.org>), and the work of the data contributors and the Neotoma community is gratefully acknowledged.

References

- Amesbury MJ, Booth RK, Roland TP et al (2018) Towards a Holarctic synthesis of peatland testate amoeba ecology: Development of a new continental-scale palaeohydrological transfer function for North America and comparison to European data. *Quaternary Science Reviews* 201:483–500. doi: 10.1016/j.quascirev.2018.10.034
- Axford Y, Levy LB, Kelly MA et al (2017) Timing and magnitude of early to middle Holocene warming in East Greenland inferred from chironomids. *Boreas* 46:678–687. doi: 10.1111/bor.12247
- Barabesi L, Cerasa A, Cerioli A, Perrotta D (2018) Goodness-of-fit testing for the Newcomb-Benford Law with application to the detection of customs fraud. *Journal of Business & Economic Statistics* 36:346–358. doi: 10.1080/07350015.2016.1172014
- Bik EM, Fang FC, Kullas AL et al (2018) Analysis and correction of inappropriate image duplication: The Molecular and Cellular Biology Experience. *Molecular and Cellular Biology* 38:e00309–18. doi: 10.1128/MCB.00309-18
- Borchers HW (2018) Numbers: Number-theoretic functions
- Bradbury JP (1997) A diatom-based paleohydrologic record of climate change for the past 800 k.y. from Owens Lake, California. In: An 800,000-year paleoclimatic record from core OL-92, Owens Lake, Southeast California. Geological Society of America
- Brown NJL, Heathers JAJ (2017) The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science* 8:363–369. doi: 10.1177/1948550616673876
- Carlisle JB (2017) Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia* 72:944–952. doi: 10.1111/anae.13938
- Coddington JA, Agnarsson I, Miller JA et al (2009) Undersampling bias: The null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology* 78:573–584. doi: 10.1111/j.1365-2656.2009.01525.x
- Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London
- Fanelli D (2013) Redefine misconduct as distorted reporting. *Nature* 494:149. doi: 10.1038/494149a
- Grieneisen ML, Zhang M (2012) A comprehensive survey of retracted ar-

343 ticles from the scholarly literature. PLOS ONE 7:1–15. doi: 10.1371/jour-
 344 nal.pone.0044118
 345 Heiri O, Lotter AF (2001) Effect of low count sums on quantitative envi-
 346 ronmental reconstructions: an example using subfossil chironomids. Journal of
 347 Paleolimnology 26:343–350. doi: 10.1023/a:1017568913302
 348 Larocque-Tobler I, Filipiak J, Tylmann W et al (2015) Comparison be-
 349 tween chironomid-inferred mean-August temperature from varved Lake Żabińskie
 350 (Poland) and instrumental data since 1896 AD. Quaternary Science Reviews
 351 111:35–50. doi: 10.1016/j.quascirev.2015.01.001
 352 Larocque-Tobler I, Filipiak J, Tylmann W et al (2016) Corrigendum to “Com-
 353 parison between chironomid-inferred mean-August temperature from varved
 354 Lake Żabińskie (Poland) and instrumental data since 1896 AD” [Quat. Sci.
 355 Rev. 111 (2015) 35–50]. Quaternary Science Reviews 140:163–167. doi:
 356 10.1016/j.quascirev.2016.01.020
 357 Pardieck K, Ziolkowski D, Lutmerding M, Hudson M-A (2018) North Ameri-
 358 can breeding bird survey dataset 1966 - 2017, version 2017.0. [https://doi.org/10.](https://doi.org/10.5066/F76972V8)
 359 [5066/F76972V8](https://doi.org/10.5066/F76972V8)
 360 Payne RJ, Mitchell EAD (2009) How many is enough? Determining optimal
 361 count totals for ecological and palaeoecological studies of testate amoebae.
 362 Journal of Paleolimnology 42:483–495. doi: 10.1007/s10933-008-9299-y
 363 R Core Team (2018) R: A language and environment for statistical computing.
 364 R Foundation for Statistical Computing, Vienna, Austria
 365 Telford RJ (2019a) Review and test of reproducibility of sub-decadal resolu-
 366 tion palaeoenvironmental reconstructions from microfossil assemblages
 367 Telford RJ (2019b) CountSum: Check assemblage count sums and percent
 368 Williams JW, Grimm EC, Blois JL et al (2018) The neotoma paleoecol-
 369 ogy database, a multiproxy, international, community-curated data resource.
 370 Quaternary Research 89:156–177. doi: 10.1017/qua.2017.105
 371 Wolodzko T (2018) ExtraDistr: Additional univariate and multivariate dis-
 372 tributions