Periodicity of extinctions in the geologic past

(David M. Raup and J. John Sepkoski, Jr.)

Department of Geophysical Sciences, University of Chicago, Chicago, IL 60637

Contributed by David M. Raup, October 11, 1983

ABSTRACT The temporal distribution of the major extinctions over the past 250 million years has been investigated statistically using various forms of time series analysis. The analyzed record is based on variation in extinction intensity for fossil families of marine vertebrates, invertebrates, and protzoans and contains 12 extinction events. The 12 events show a statistically significant periodicity (P < 0.01) with a mean interval between events of 26 million years. Two of the events coincide with extinctions that have been previously linked to meteorite impacts (terminal Cretaceous and Late Eocene). Although the causes of the periodicity are unknown, it is possible that they are related to extraterrestrial forces (solar, solar system, or galactic).

Virtually all species of animals and plants that have ever lived are now extinct, and the known fossil record documents some 200,000 such extinctions. It has been generally assumed that extinction is a continuous process in the sense that species are always at risk and that mass extinctions simply reflect relatively short-term increases in that risk. Following this view, the extinction process is often described mathematically as a time homogeneous process using standard birth-death models (1–3). There is increasing evidence, however, that many extinctions are actually short-lived events of special stress, separated by periods of much lower, or even negligible, risk. Fischer and Arthur (4) departed from convention by arguing that the major extinction events of the past 250 million years (ma) occurred periodically at nearly constant intervals of 32 ma (see also ref 5). Their study used a limited data base, and no statistical testing was done. The purpose of this paper, therefore, is to test the proposition of periodicity in the record of marine extinctions over the past 250 ma (Late Permian to Recent) by using as rigorous a methodology as present data permit.

DATA BASE

The data for this study come from Sepkoski’s compilation (6) of the temporal ranges of =3,500 families of marine animals (vertebrate, invertebrate, and protzoan), but the subset of data and method of expressing extinction intensity differ from our previous analyses of these data (7, 8). For the present study, a culled subset of the total sample was used: all families with low-resolution ranges, not known to the level of the stratigraphic stage, were eliminated, as were families noted by Sepkoski (6) as having questionable taxonomic or stratigraphic designations. In addition, families still living today were ignored in order to avoid the damping effect of the “Pull of the Recent” (9). This culling process reduced the sample substantially (567 for the Late Permian to Recent) but in so doing removed much of the noise that characterizes data sets of this kind.

The Late Permian to Recent interval analyzed in this study contains several well-documented mass extinctions, has a comparatively accurate time scale, and is divided into relatively short stratigraphic stages. The interval comprises 39 international stages ranging in age from 253 ma B.P. (base of the Dzhulifan Stage of the Late Permian) to 11.3 ma B.P. (top of the Middle Miocene), using the Harland time scale (10). The mean duration of these 39 stages is 6.2 x 10^6 years, which makes it impossible to resolve extinction events separated by less than about 12 x 10^6 years (the Nyquist rate). Although finer resolution is possible in some parts of the geologic column and with some biologic groups, the resolution used in the present study is the best that can be achieved for comprehensive analysis in the present state of synoptic work.

One can debate the quality of the individual data. The family is a rather large and arbitrary taxonomic unit and as such tends to damp variation at the species level (11): even if 99% of the members of a family become extinct at one time, the persistence of a single species will prevent that family from reflecting the extinction event. However, in spite of this problem, patterns of familial diversification and extinction do seem to correlate well with data for fossil species (12), and the available data for families represent a far more uniform and less biased sampling of the fossil record than any species-level data set.

Age designations in the data set are also open to question, especially because geologic time scales are continually being revised. A given time scale represents a merger of a chronostratigraphic sequence (based primarily on fossils) and a scattering of radiometric dates (13). The present analysis uses the Harland and Odin time scales (10, 14), which have some substantial differences. The Harland scale is quite similar to that of Armstrong (15) and is nearly identical to the new time scale of the Geological Society of America (16). Thus, the Harland scale is probably the closest to a consensus, even though problems remain.

MEASUREMENT OF EXTINCTION RATES

Many metrics have been used to express variation in the intensity of extinction. Most commonly, the number of extinctions in an interval of geologic time is divided by standing diversity (total taxa present) and this in turn is divided by the estimated duration of the interval to yield a normalized, per capita rate of extinction per million years. In the present analysis, we have not normalized for time because of considerable uncertainty in the durations of stages (for example, the correlation of post-Paleozoic stage durations between the Harland and Odin time scales is only 0.48) and because of the growing suspicion, mentioned above, that extinction is not a continuous process (see also ref. 17). Furthermore, it should be noted that there is no statistical correlation between the raw extinction data and stage duration: for the stages of the Phanerozoic used by Sepkoski (6), numbers of extinctions have a correlation of −0.06 with stage durations in the Harland time scale. This is compatible with the hypothesis that extinction is an episodic process.

Abbreviation: ma, million year(s).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Evolution: Raup and Sepkoski

Fig. 1. Extinction record for the past 250 ma. Letter codes (bottom) identify stratigraphic stages. The best-fit 26-ma cycle is shown along the top. The relative heights of extinction peaks should not be taken as literal expressions of extinction intensity because the absence of extant taxa exaggerates the heights of younger peaks.

In the analysis that follows, the data are normalized for standing diversity by expressing the number of families becoming extinct as a percentage of the families present. Because standing diversity changes relatively slowly, the effect of this normalization is minor.

Fig. 1 shows the basic time series for the Late Permian to Middle Miocene interval using the Harland time scale. One point for extinction intensity is plotted at the end of each stage. The use of the end point is not entirely arbitrary. Many of the stratigraphic units were originally established on the basis of major faunal turnover (extinction) and it is no accident that the two largest mass extinctions (Late Permian and Late Cretaceous) were used to mark the boundaries between the Paleozoic, Mesozoic, and Cenozoic Eras. The plotting of extinctions at the ends of stages does not deny that some extinctions occur between stage boundaries, but this convention is used as the best general inference in the present state of knowledge: (Moving the points to the middle or even beginning of the stages changes the statistical results in no substantial way.)

For the purpose of this analysis, each peak in Fig. 1 is an extinction "event," a peak being any point flanked by lower points. Three of the four highest peaks correspond to the Late Permian (Guadalupian–Dzhulfian), Late Triassic (Norian), and terminal Cretaceous (Maestrichtian) events, all of which have long been recognized as major mass extinctions. The most questionable peak is probably the one at the extreme right of Fig. 1 (Middle Miocene at 11.3 ma). Because average family extinction rate declines through the Phanerozoic (7), sampling error becomes a problem as the recent is approached. We have included the Middle Miocene peak largely on the basis of other evidence (18, 19).

STATISTICAL ANALYSIS OF THE TIME SERIES

Fig. 1 gives a visual impression of a rather regular spacing of peaks of extinction through the Mesozoic and Cenozoic Eras. However, because the data represent interval estimates of the times of extinction and because no two peaks of extinction can be recognized if they are separated by less than one stage, qualitative impressions may be misleading. For this reason, we have applied a variety of standard and nonstandard tests of periodicity to the time series. In each case, special attention was given to the problems imposed by the irregular and scattered distribution of data points.

Fourier Analysis. In an initial, exploratory pass, the 39 extinction points in Fig. 1 were subjected to a simple Fourier analysis. The resulting smoothed power spectrum is illustrated in Fig. 2. This spectrum shows a peak in power near the first harmonic (reflecting the high extinction percentages near the two ends of the time series) and a pronounced peak at the eighth harmonic, suggesting a periodicity in the neighborhood of 30 million years. However, though promising, the second peak in the power spectrum should not be taken as proof of a persistent periodicity. It can be argued that the necessary minimal spacing of $12 \times 10^3$ years between observed extinction peaks can make random (Poisson) data appear periodic to Fourier analysis.

The Fourier results were corroborated by a standard autocorrelation analysis. The correlogram showed statistically significant autocorrelation ($P < 0.01$) for cycles between 27 and 35 ma. However, we do not consider this conclusive because autocorrelation gives undue weight to the long intervals of background extinction between peaks.

Nonparametric Testing. In order to accommodate the irregular distribution of data points in the time series and the interval nature of the geologic time scale, we developed a nonparametric test based on comparing the observed time series with an empirically developed distribution of all possible time series constructed from the same data. Test distributions were assembled by "brute force" from computer simulations involving Monte Carlo randomization of the actual data. By this means, we preserved the basic fabric of the data and avoided assumptions about the distribution. The test procedure involved the following steps:

(i) A perfectly periodic impulse function with a given cycle length between 12 and 60 ma was placed arbitrarily on the time series to establish a set of predicted times of extinction. For each extinction event (peak in Fig. 1), the distance to the closest predicted time was recorded as a positive or negative error. The mean of these errors was then subtracted from each predicted peak position to find the best-fit position of the function. (This procedure was iterated several times to ensure that no error exceeded half a cycle length.) The errors were then recalculated and their standard deviation was computed to be used as a metric of goodness of fit. (A standard deviation of zero would thus represent a perfect fit to a given cycle.) This approach has the advantage that each peak is considered independently and intervals between peaks are never measured explicitly so that we avoid problems that would be caused by extra or missing peaks.

(ii) The real data were randomized (shuffled) with a random number generator, peaks of extinction were identified as before, and the goodness-of-fit metric was calculated for the distances between predicted and observed peaks in the randomized data. This was repeated 500 times for each cycle to estimate an empirical probability density distribution of standard deviations for the randomized data.

(iii) The standard deviation of distances (errors) from the cycle for the real data was then compared with the distribution formed by the 500 simulations. If the standard deviation of errors for the real time series was lower than 99% of the
randomized results, the cycle being investigated was considered to be statistically significant ($P < 0.01$).

The results of this test procedure using the Harland time scale are illustrated in Fig. 3. The axes of this graph are the fitted cycle lengths from 12 to 60 ma (abscissa) and the standard deviations of the distances between observed and predicted peaks (ordinate). The solid line in the graph plots the standard deviations for the real data, and the crossed and dotted lines show the 50th and 99th percentiles for the randomized data, respectively. As evident, the real data follow the mean simulation fairly closely over most cycle lengths. There are, however, two major excursions toward lower standard deviations. The larger excursion centers on 26 ma; where the standard deviation of distances of observed from predicted peaks is less than 95% of those computed from randomized data. The second excursion at 30 ma is less substantial, with the standard deviation for the real data being less than 95% of the randomized version.

Fig. 3 Inset shows detail for the result at 26 ma. Because the standard deviation for the real data at this cycle length was actually less than all 500 randomizations, we computed an additional 8,000 simulations. The frequency distribution of standard deviations for the additional simulations is illustrated in the Inset. As evident, none of these simulations had a standard deviation as low as the real time series. This is clearly a highly significant result ($P < 0.0001$).

Two special problems have to be considered in assessing this result. First, the time series of extinction data shows somewhat fewer peaks than would be predicted if it were a random walk—that is, the time series appears to have some memory (Markov property), as also indicated by the auto-correlation analysis. In view of this, the computer program was written so that only simulations with the same number of peaks as the real time series were used. This problem is not profound because when the computations were repeated without the Markov constraint, the results were virtually identical.

The other problem is potentially more serious and has to do with the nature of the time scale. The computer program that produced Fig. 3 caused both the extinction data and the time scale to be randomized. In randomizing the time scale, durations of stages were preserved but their sequence was changed. It could be argued that the statistical significance of the results could have been generated by periodic elements in the time scale itself rather than from the distribution of extinctions. But even if this were the case, periodicity of extinctions could still be claimed because the record of extinctions has been used in part to establish boundaries in the geologic time scale. Nevertheless, the analyses were also performed without randomization of the time scale. For the 26-ma cycle, the statistical confidence drops from $P < 0.0001$ to $P = 0.0045$, the latter being based on 4,000 simulations. This latter probability value should be taken as the most conservative estimate, although not necessarily the correct one.

Table 1 summarizes the results of all test runs using the nonparametric procedure with both the Harland and Odin time scales. Cycles that are significant at the 95% level ($P < 0.05$) or at the 99% level ($P < 0.01$) are indicated. Runs were also made separately on the first and second halves of the time series. The first half provided significant results for cycle lengths of 24–27 (P < 0.01) and 28 ma (P < 0.05); the second half gave significant results at 28 and 29 (P < 0.01) and at 27 and 30 ma (P < 0.05). The difference between these results is minor and probably reflects some “wobble” in the real data.

It is important to note that each row in Table 1 is the result of 49 separate tests, one for each cycle between 12 and 60 ma. Because many statistical tests were performed simultaneously, there is a nontrivial probability that one or more will indicate a “significant” $P$ value just by chance. One could calculate the binomial probabilities of such outcomes but this would require an assumption of the independence of the tests. In fact, the tests are not independent: a low $P$ value for one cycle is often associated with low values for adjacent cycles. Therefore, we have developed the probability distribution of multiple successes empirically by using a separate set of randomized versions of the time series as input for the basic program for the nonparametric test. This provides a good estimate of the likelihood of clusters of low $P$ values comparable to those in Table 1.

It turns out that almost all clusters of “significant” results are concentrated among the longer cycles (>30 ma). In fact, using the Harland time scale, including its randomization, no cluster of three or more values with $P < 0.01$ occurred in cycles <30 ma in 500 trials. Thus, the probability of a cluster of three occurring by chance in the 12- to 29-ma interval is <0.002. The statistical significance of the 26- to 28-ma cycles in the first row of Table 1 (Harland time scale with time randomization) is sustained at $P < 0.01$.

For the second row in Table 1 (Harland time scale without time randomization), the highest value in the cluster of three is $P = 0.038$, and the expected distribution of clusters at least as strong as this was also determined by the procedure just described. Again, the frequency of clusters increases sharply with increasing cycle length. For cycles of <30 ma, only 4 of the 500 simulations showed clusters of at least three $P$ values of 0.038 or less. This corresponds to a probability of 0.008 and amply sustains the statistical significance of the cluster ($P < 0.01$). We thus conclude that both indications of periodicity between 25 and 27 ma shown for the Harland time scale.
is corroborative when range indications are significant. The analysis just described lessens confidence in the scattered indications of periodicity outside the 25- to 27-ma range in Table 1. The significant cycles at 19, 29, and 30 should be interpreted only as suggestions for future exploration when stronger data bases are available. The cycle at 30 ma may be real but cannot be confirmed with the present time series.

Best-Fit Cycle. The results of a final test of the extinction record are illustrated in Fig. 4. This figure represents a "composite cycle" constructed by averaging together 26-ma segments of the time series in Fig. 1. To compute the composite, the time series was marked off at 26-ma intervals in such a way that the Cretaceous–Tertiary boundary peak (at 65 ma B.P.) was at the center of one interval. The extinction data in each interval (interpolated to every 1 ma) were then rescaled so that the maximal value was 100%. These values were collected with their counterparts in other intervals and averages and dispersions were calculated.

This procedure was performed for a variety of interval lengths (i.e., cycle lengths) on either side of 26 ma. The illustrated composite cycle represents a best-fit to the data in the sense that the composite peak has maximal amplitude and minimal dispersion about the median values. Tests of the medians show highly significant differences: a Friedmann's test for points separated by 6 ma (the approximate average stage duration) results in an S value of 250 (P < 0.05 with four degrees of freedom). Other cycle lengths give less significant results, and the composite curve deteriorates into an irregular, non-unimodal curve (noise) when the cycle departs substantially from the best-fit cycle.

It should be noted that the composite curve in Fig. 4 rises to a fairly sharp peak, which is compatible with (although does not prove) the proposition that mass extinctions are discrete events. If extinctions resulted from continuous fluctuations in background rate, a composite curve approaching a sine-cosine function would be expected. The curve in Fig. 4 may in fact represent a rather conservative illustration of the abruptness of the average extinction peak; much of its breadth may result from the interval nature of the data and the fact that sampling error (i.e., failure to locate a family's precise interval of extinction) tends to smear the record of mass extinction backward in time (20). The slight asymmetry of the curve in Fig. 4 is consistent with this last proposition.

CONCLUSIONS

The time series in Fig. 1 has the 26-ma cycle superimposed in its best-fit position. The deviations from this best-fit position are listed in Table 2. The 26-ma cycle predicts 10 extinction events in the Permian–Miocene interval, whereas the actual time series contains 12 peaks. Of the 12 peaks, the smallest is in the Early Triassic (Olenekian); this peak reflects only ammonoid extinctions and may be spurious, but it was included in the analysis for the sake of consistency. The poorest fits are for 2 peaks in the Middle Jurassic and the one in the Early Cretaceous. These three peaks are low and may not be significantly higher than the surrounding background extinction.

It seems inescapable that the post-Late Permian extinction record contains a 26-ma periodicity, assuming that the Harland time scale (with its Geological Society of America coun-

![Figure 4](https://example.com/fig4.png)

**Figure 4.** Composite best-fit curve for extinction intensity using a 26-ma cycle. The range between the 25th and 75th percentiles of the data is shown by the vertical lines.

<table>
<thead>
<tr>
<th>Harland time scale</th>
<th>Odin time scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak observed, ma B.P.</td>
<td>Closest peak predicted, ma B.P.</td>
</tr>
<tr>
<td>Tertiary</td>
<td></td>
</tr>
<tr>
<td>Middle Miocene</td>
<td>11.3</td>
</tr>
<tr>
<td>Late Eocene</td>
<td>38</td>
</tr>
<tr>
<td>Cretaceous</td>
<td></td>
</tr>
<tr>
<td>Maestrichtian</td>
<td>65</td>
</tr>
<tr>
<td>Cenomanian</td>
<td>91</td>
</tr>
<tr>
<td>Hauterivian</td>
<td>125</td>
</tr>
<tr>
<td>Jurassic</td>
<td></td>
</tr>
<tr>
<td>Tithonian</td>
<td>144</td>
</tr>
<tr>
<td>Callovian</td>
<td>163</td>
</tr>
<tr>
<td>Bajocian</td>
<td>175</td>
</tr>
<tr>
<td>Piensbachian</td>
<td>194</td>
</tr>
<tr>
<td>Triassic</td>
<td></td>
</tr>
<tr>
<td>Norian</td>
<td>219</td>
</tr>
<tr>
<td>Olenekian</td>
<td>243</td>
</tr>
<tr>
<td>Permian</td>
<td></td>
</tr>
<tr>
<td>Dzhulfian</td>
<td>248</td>
</tr>
</tbody>
</table>

Standard deviation of errors = 3.85 Standard deviation of errors = 5.65
terpart) is a reasonable approximation of reality. This conclusion is based primarily on the nonparametric test procedure described in this paper, but the other, less rigorous tests are largely confirmatory, especially the best-fit composite cycle (at 26 ma. Fig. 4). Of particular importance is the fact that the nonparametric test gives approximately the same results when applied independently to the two halves of the time series. The Fourier and autocorrelation analyses yield cycles reasonably close to 26 ma. In view of the various sources of distortion in these analyses, we do not see these discrepancies as significant. It is likely, however, that as the quality of the time scale and paleontological data improve, the length of the estimated cycle may shift somewhat.

It is possible that the appearance of a 26-ma cycle actually results from a longer cycle of, say, 52 ma in combination with a scattering of random events. This model has been tested and found to be a weaker description of the data than the simple 26-ma cycle.

Also, with more and better data, studies of periodicity can be extended to the Paleozoic. At present, the Sepkoski data set shows no evidence of Paleozoic periodicity that can survive the test procedures used here for the younger record. This may well be due to the relatively weak state of the Paleozoic time scale but nothing can be said unequivocally at this time.

**IMPLICATIONS**

If periodicity of extinctions in the geologic past can be demonstrated, the implications are broad and fundamental. A first question is whether we are seeing the effects of a purely biological phenomenon or whether periodic extinction results from recurrent events or cycles in the physical environment. If the forcing agent is in the physical environment, does this reflect an earthbound process or something in space? If the latter, are the extraterrestrial influences solar, solar system, or galactic? Although none of these alternatives can be ruled out now, we favor extraterrestrial causes for the reason that purely biological or earthbound physical cycles seem incredible, where the cycles are of fixed length and measured on a time scale of tens of millions of years. By contrast, astronomical and astrophysical cycles of this order are plausible even though candidates for the particular cycle observed in the extinction data are few. One possibility is the passage of our solar system through the spiral arms of the Milky Way Galaxy, which has been estimated to occur on the order of 10² years (21). Shoemaker has argued (21) that passage through galactic arms should increase the comet flux and this could, following the Alvarez hypothesis (22), provide an explanation for the biological extinctions. Two of the extinction events being considered here (Late Cretaceous and Late Eocene) are associated with evidence for meteorite impact (23, 24). However, much more information is needed before definitive statements about causes can be made. It may turn out that the biological extinction record is sensitive to periodic phenomena that other indicators have failed to recognize.

The implications of periodicity for evolutionary biology are profound. The most obvious is that the evolutionary system is not “alone” in the sense that it is partially dependent upon external influences more profound than the local and regional environmental changes normally considered. Much has been written about the “bottlenecking” effect of mass extinction. With kill rates for species estimated to have been as high as 77% and 96% for the largest extinctions (11, 25), the biosphere is forced through narrow bottlenecks and the recovery from these events is usually accompanied by fundamental changes in biotic composition (26). Without these perturbations, the general course of macroevolution could have been very different.

We thank Eric W. Holman, Ronald Thisted, and D. S. Simberloff for discussions of statistical procedures. This research was supported by National Aeronautics and Space Administration Grant NAG 2-237.